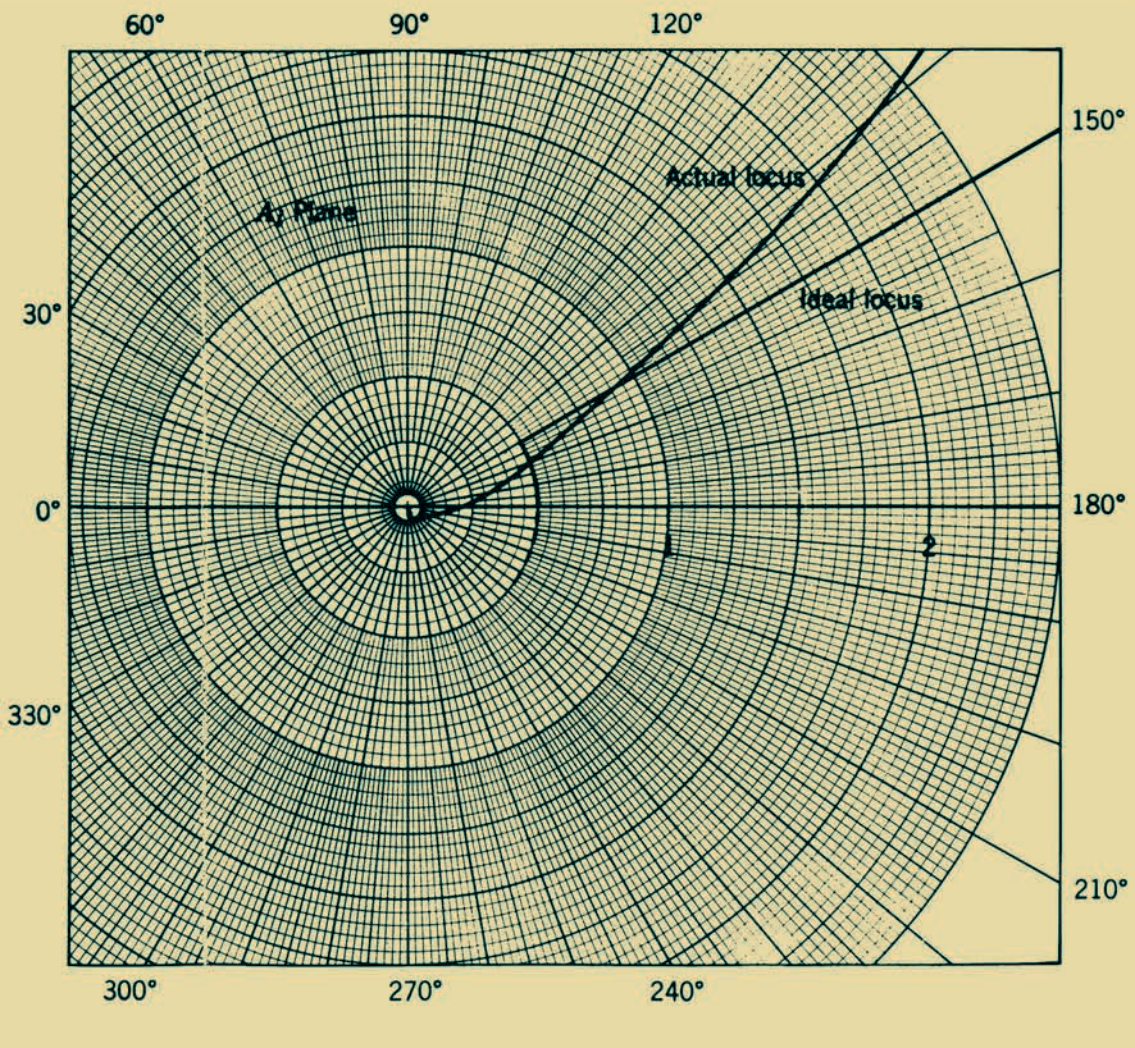
Amplifying Devices and Low-Pass Amplifier Design

E. M. CHERRY D. E. HOOPER



About the book

This book presents methods for designing amplifiers that are useful, reproducible in quantity and reliable. There is discussion of the values of vacuum tube and transistor parameters, and their variations with operating conditions, temperature and age. A special feature of the book is the unified, parallel treatment of tubes, transistors and f.e.t.s. In addition, there are more than a dozen amplifier designs to illustrate principles.

In the authors' view, modern engineering education tends to concentrate on the fundamental background of knowledge or "pure science" to the neglect of practical aspects of engineering. With this in mind they indicate how mathematics, physics, and specialized electrical subjects can be used as tools in the essentially practical task of designing amplifiers.

Intended primarily as a university level text, it provides proved rules of circuit design and is therefore of immediate use not only to students but also to practicing engineers and scientists.

HB-DS–2510

Cherry & Hooper

Amplifying Devices and Low-Pass Amplifier Design

Wiley

About the Authors

E. M. CHERRY is a senior lecturer in Electrical Engineering at Monash University, Victoria, Australia. University of Melbourne granted him a Bachelor's degree in 1957, a Master's degree in 1959 and a Ph.D. in 1962.

Dr. Cherry was on the staff of Bell Telephone Laboratories Inc., Murray Hill, New Jersey during 1963-1965. Prior to joining Monash University in 1967, he worked for one year with the U. K. Atomic Energy Authority at Harwell.

He has experience in the fields of television deflection circuitry, high-speed transistor amplifiers and pulse circuits, semiconductor physics, transmission line theory, and low-noise nuclear pulse amplifiers.

D. E. HOOPER is currently on the staff of Plessey Pacific Pty. Ltd. He received both the Bachelor's and Master's degrees in Electrical Engineering from the University of Melbourne in 1953 and 1962 respectively.

Before assuming his present position in 1967, Mr. Hooper spent 10 years in the Electrical Engineering Department of the University of Melbourne, first as a lecturer and later a senior lecturer. For a year in 1963 he was a Visiting Research Associate at the U. K. Atomic Energy Authority at Harwell.

Mr. Hooper has research and teaching experience in the areas of transistor physics and characterization, pulse forming circuits, wide-band amplifiers, active filters, and integrated circuit design.

Amplifying Devices and Low-Pass Amplifier Design

E. M. Cherry
Monash University

D. E. Hooper
Plessey Pacific Pty. Ltd.
(formerly University of Melbourne)

John Wiley and Sons, Inc., New York · London · Sydney

Library of Congress Catalog Card Number: 67-29933

GB 471 15345 1

Printed in the United States of America

WILEY

Preface

Many fine books have been written on the subject of electronic amplifiers. Nevertheless we believe that there is still a need for a book that at once provides the physical basis for understanding yet emphasizes the nonideal or engineering aspects of design.

Modern engineering education is tending more and more to concentrate on the fundamental background of knowledge with the intention that the new graduate can enter any field and, after a shaking-down period, do something useful. This is an admirable intention, but we wonder whether there is not too much concentration on the "pure science" aspects of engineering, to the neglect of the practical aspects. The student should at some stage be confronted with the fact that physically realizable systems are not ideal. More important, he should be shown how to design nonideal but still useful systems. In short, he should be taught engineering as well as science. It is one thing to study the idealized Otto cycle in thermodynamics and the differential equations of hydrodynamics; it is quite another thing to design an internal combustion engine, including a detailed design of the carburetor.

We are not alone in our view that there is too little regard for practice in modern engineering education. R. W. Johnson has touched on the problem when he writes:*

"One only needs to attend any of the engineering conventions, symposia, section or professional group meetings to discover that there is an appalling lack of understanding among many of our younger engineers as to the things really important to the art of designing. At these affairs, intelligent and sincere young men proudly describe their analyses, developments, and achievements. In many cases, original work has been done, and its author truly displays a good background of prior work. But too often the dissertation is a rediscovery of work already done, or describes hardware of marginal usefulness, and clearly demonstrates a lack of understanding of many things basic to engineering—the art of designing!"

* R. W. Johnson, "Are electronics engineers educated?" *Proc. Inst. Radio Engrs.*, **49**, 1319, August 1961.

If we accept the premise that electrical engineers should receive some "engineering" training in their undergraduate course, the question remains how much? Because it is clearly impossible to treat all subjects in detail, one subject is probably sufficient provided that the techniques used are representative of others. A good choice (but by no means the only one) is the design of low-pass amplifiers. These are basic; they occur in some form in almost all items of electronic equipment and a knowledge of their design is thus potentially useful to workers in many fields. In addition, low-pass amplifier design is easy to integrate into existing courses, particularly into the laboratory sessions.

The subject matter in a course of amplifier engineering falls into the same three broad categories as other branches of engineering—materials, formal analytical techniques, and approximations. The materials of amplifier engineering are active devices and passive components, and knowledge of them is based on physical principles. The active devices must be thoroughly understood; typical parameter values, their variation with operating conditions and age, and likely manufacturing tolerances should be known. A familiarity with resistor and capacitor types is necessary also. The basic analytical tool in amplifier engineering is linear circuit theory, and this too is a fundamental branch of knowledge. The part of amplifier design that is essentially engineering is in making approximations. When an analytical expression becomes too unwieldy for convenience, it is not enough just to neglect high-order terms; it is necessary to determine what constraints should be placed on the system or what design rules should be observed to ensure that these terms are in fact small. Engineering has aptly been described as "the art of making approximations."

With this in mind, our aim has been to write a book that gives proven rules for circuit design and is of immediate use to practicing engineers and scientists. We discuss the practical steps in designing amplifiers whose performance remains within specified limits despite the initial production tolerances on device parameters and component values, and their subsequent variations with environment and age. A university textbook, however, should be much more than a collection of rules. We have therefore emphasized their physical basis and in so doing have brought together mathematics, physics, and specialized electrical subjects to form a textbook on engineering.

STUDENT ASSIGNMENTS IN AMPLIFIER DESIGN

Unlike analytical problems, there is seldom a unique solution to a design problem. It may be thought that a course in amplifier design will result in an increased load on the teacher, but in our experience this is not

so. In contrast to structural and mechanical designs, amplifiers can be built readily and tested experimentally. We have encountered such remarkably good response from students that in many cases they have actually built and tested the amplifiers in their own time. Students realize that a design will work if it is correct and that a marginal design must be considered a failure. As a result, there are no really hopeless submissions of the type that inevitably appear among solutions to analytical problems. Design problems are almost self-correcting. In certain cases the normal laboratory sessions have been used by students for building and testing their designs. A healthy spirit of competition has developed between students attempting these projects and has led to an energetic comparison of final circuits and, of course, a deep understanding of the particular topic and the general philosophy of circuit design. A limited number of assignments are given at the end of the book.

LAYOUT OF THE BOOK

The book divides naturally into two parts. Part 1, containing Chapters 1–9, is intended for a first course in electronic circuits, whereas Part 2, containing Chapters 11–16, is intended for a second course. Feedback theory is not required in Part 1, but it is used extensively in Part 2; this arrangement gives students a "feel" for design by confronting them with simple problems before they are burdened with the extra background information needed in designing feedback circuits. Chapter 10 is a formal introduction to feedback theory that provides the link between the two parts, whereas Chapter 17 is a rounding off. Figure 0.1 shows the relation between the book and preceding or partly collateral courses in linear network theory and physical electronics.

It is not essential that all the material covered in Chapters 2–4 be assimilated before proceeding with Chapters 5–7 and, to a lesser extent, Chapters 9, 10, 11, 12, 15, and 16. In our own universities, the material in Chapters 2–4 is covered first in the third year of the undergraduate course and again in more detail in the final (fourth) year. This is not to suggest that some of the material is unimportant in everyday amplifier design. Rather it is intended to balance the course; amplifier circuits are first encountered about one third of the way through the third year.

SOME COMMENTS ON THE TREATMENT OF SUBJECTS

We have attempted to provide fairly complete discussion of only some methods of designing amplifiers. The design philosophy developed by way of these examples is not restrictive and may be applied to many other cases.

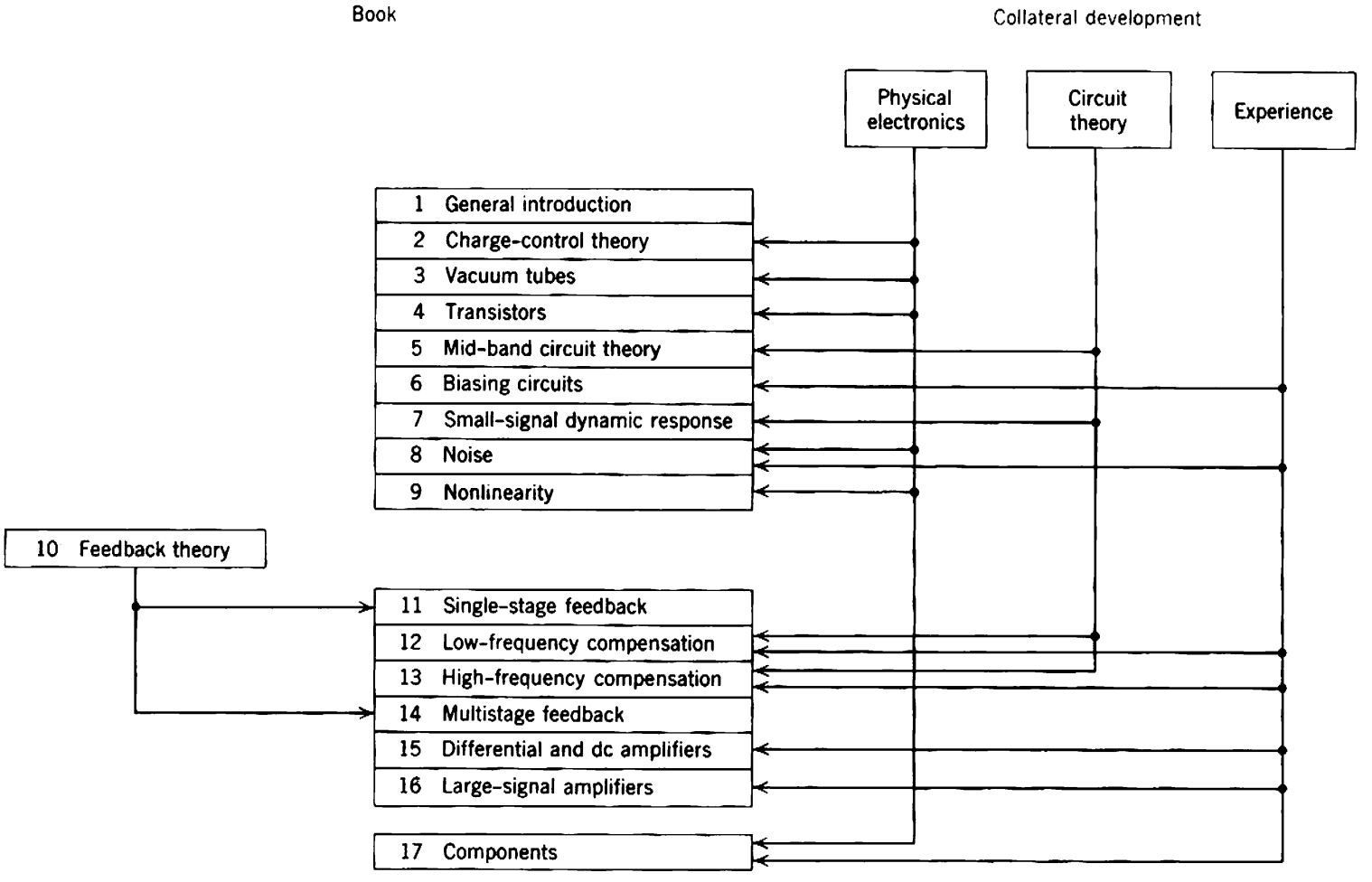


Fig. 0.1 Position of the book in an undergraduate course in electronic engineering.

Our guide has been the desire to tackle problems as simply and directly as possible, and this has led to a number of unusual features in the book:

1. Vacuum tubes, bipolar transistors, and unipolar (field-effect) transistors are considered in Chapters 2, 3, and 4 as approximations to the charge-control model. This provides a common ground for comparing and contrasting the devices.

2. The charge-control approach results in the universal use of physical equivalent circuits for amplifying devices rather than the abstract network equivalent circuits that are so often referred to in the literature.

3. The controlled generator in these physical equivalent circuits is in all cases a mutual conductance. It is concluded that the bipolar (conventional) transistor should be considered as a voltage amplifier (like the vacuum tube and field-effect transistor) rather than a current amplifier (as is customary). The treatment of vacuum tubes and transistors can therefore be integrated and runs parallel at each stage of development.

4. An extensive discussion is given of parameter variations with operating conditions, temperature, age, and device replacement.

5. Noise and distortion are considered throughout the book, not relegated to a final chapter.

6. Feedback amplifiers receive extensive treatment. Most of the "advanced" amplifiers discussed in Part 2 of the book employ negative feedback.

It might be felt that inclusion of material on vacuum tubes is out of character with a book published in 1967. We have retained this material for two reasons:

1. It emphasizes the continuing tradition in electronic circuit design, which goes back for more than 30 years. This could be valuable to the increasing number of young engineers who seem unaware of the significant circuit developments of the vacuum tube era (particularly by the radar development teams of the 1940s), and who continue to make "new" discoveries of established techniques.

2. Despite the developments in transistors and microelectronics, vacuum-tube equipment is still being designed and much existing equipment will remain in service for many years to come.

The impact of microelectronic or integrated circuit techniques calls for some comment. Economical design of silicon integrated circuits differs from that of discrete-component circuits principally in that the former aims at minimizing the total area of silicon used whereas the latter aims at minimizing the number of active devices. Consequently, a few of the performance yardsticks adopted in this book (e.g., stage gain and stage gain-bandwidth product) are more appropriate to discrete-component

than integrated circuits. Nevertheless, the bulk of the theoretical material is directly applicable to both and has been applied with success. Over-all, the basic philosophy of the pen-and-paper stage of design is the same, namely, the production of circuits whose performance can be predicted before they are built; the integrated-circuit designer is just not able to make post-production changes in his design.

We do, of course, hope that this book will be of immediate use to undergraduates and practicing engineers. We will, however, consider it a success if it can play even a small part in developing a stronger sense of "engineering" among workers in the electronics field.

ACKNOWLEDGEMENTS

This book has been five years in the writing, and for four of these years we have been in different countries. We take this opportunity of expressing our gratitude to the Bell Telephone Laboratories and the United Kingdom Atomic Energy Authority who provided us with temporary employment and gave us very valuable experience.

Over the years a number of people have influenced our thinking about electronic circuits: our co-workers at B.T.L. and U.K.A.E.A., the post-graduate students at Melbourne University, and the undergraduate students who asked questions. To all of them, especially those who contributed directly to the book by reading and commenting on parts of the manuscript, we extend our thanks. C. F. Hempstead, C. I. Sach, M. A. Schapper and A. R. T. Turnbull were of particular help.

Susan Jackson, Pat Lardiere, Mrs. S. R. Lewis and Lenore Macdonald have had the unenviable task of typing and retyping our rough manuscripts. Without the efficient and understanding cooperation of these ladies the book would have been much longer in production.

The Bell Telephone Laboratories bore the brunt of preparing the final manuscript and the accurate diagrams. C. C. Cutler, Head of Communications Systems Research, made this possible and also gave us a very frank criticism of an early version of the manuscript. Our publishers, too, have made a number of suggestions through their review board and their editorial consultant R. F. Shea.

For the most part this book has been written at home as a spare-time project. We thank our wives Diana Cherry and Jennifer Hooper for providing us with happy homes in which to work and for so long putting up with silent, working husbands, five, six, and seven nights a week.

September 1967

E. M. CHERRY
Monash University
Victoria, Australia

D. E. HOOPER
Plessey Pacific Pty. Ltd.
Victoria, Australia

List of Principal Symbols

A	(i) area (usually subscripted)	
	(ii) flicker noise constant	Eq. 8.12*b*
	(iii) availability factor	Eq. 17.15
A_I, A_V	current gain, voltage gain	Eqs. 5.24 and 5.16
A_l	loop gain	Eq. 10.37
a	(i) voltage exponent of depletion layer width	Eq. 4.8
	(ii) temperature exponent of I_{CO}	Table 4.6
	(iii) superposition constant	Eqs. 10.28 and 10.72
B	susceptance (usually subscripted)	
b	(i) generalized feedback factor	Eqs. 10.28 and 10.72
	(ii) temperature exponent of mobility	Eq. 4.23
	(iii) generalized feedback factor	Eqs. 10.28 and 10.72
$\mathscr{B}$	bandwidth	Fig. 1.3
$\mathscr{B}_N$	noise bandwidth	Eq. 8.69
C	capacitance (usually subscripted)	
c	(i) incremental capacitance (usually subscripted)	
	(ii) flicker noise constant	Eq. 2.65
	(iii) temperature exponent of β_N	Table 4.6
	(iv) superposition constant	Eqs. 10.28 and 10.72

c_{AG}	extrinsic anode-grid capacitance	Section 3.4.2
c_{AK}, c_{GK}	intrinsic anode-cathode and grid-cathode capacitances	Eqs. 3.39 and 3.38
c_B	base-charging capacitance	Eq. 4.55
c_C	effective internal input capacitance of a transistor	Eq. 4.111
c_{in}, c_{out}	total grid-cathode and anode-cathode capacitances	Section 3.4.2
c_{tC}, c_{tE}	collector and emitter transition capacitances	Eqs. 4.65 and 4.64
c_{t0}	capacitance constant	Eq. 4.11
c_1, c_2, c_{11}, c_{12}, c_{21}, c_{22}	parameters of the charge-control model	Eqs. 2.39–2.44
D_e, D_h	diffusion constants for electrons and holes	
D_I	intermodulation distortion	Eqs. 9.11 and 9.12
D_r	rth harmonic distortion	Eqs. 9.2
D_T	total harmonic distortion	Eq. 9.3
DF	general driving-point function	
DF_D	direct component of DF	Eq. 10.72
$d(i_{NF}{}^2)$	mean-square spot flicker noise current	Eq. 2.65
$d(i_{NI}{}^2)$	mean-square spot induced noise current	Eq. 3.62
$d(i_{NS}{}^2)$	mean-square spot shot noise current	Eq. 2.63
$d(i_{NT}{}^2)$, $d(v_{NT}{}^2)$	mean-square spot thermal noise current and voltage	Eqs. 2.62 and 2.61
$d(i_N')^2$, $d(i_N'')^2$	mean-square spot noise current generator pair at the input and output	Fig. 8.1*c*
$d(i_N{}^2)$, $d(v_N{}^2)$	mean-square spot noise current and voltage generator pair at the input	Fig. 8.1*g*
$d(i_{Ni}{}^2)$, $d(v_{Ni}{}^2)$	mean-square spot noise current and voltage present in the input circuit	Eqs. 8.57
dN_{Ai}/dN_{Si}	ratio of spot noise powers at the input due respectively to the amplifier and source	Eqs. 8.83
E	energy	
$\mathscr{E}$	electric field	

F	(i) return difference	Eq. 10.40
	(ii) spot noise factor	Eqs. 8.82 and 8.116
$\bar{F}$	average noise factor	Eq. 8.121
F_V	voltage conversion factor	Eq. 3.72
f	frequency in hertz or cycles per second (usually subscripted)	See ω
f_{co}	cutoff frequency	
$f_C, f_T, f_1, f_\alpha, f_\beta$	characteristic frequencies of a transistor	See ω_C, etc.
G	(i) conductance (usually subscripted)	
	(ii) perveance	Eq. 3.34
G_T	transfer conductance	Eq. 5.13
g	(i) incremental conductance (usually subscripted)	
	(ii) rate of carrier generation	
	(iii) generalized reference variable	Eqs. 10.28 and 10.72
	(iv) loop gain advantage through peaking	Eq. 14.21
g_m	mutual conductance	Eq. 2.53
$\bar{g}_m$	large-signal g_m	Eq. 6.16
g_{mA}	average or nominal value of g_m	Table 3.1
g_f, g_1, g_2	parameters of the charge-control model	Eqs. 2.54–2.56
$\mathscr{GB}$	gain-bandwidth product	Eq. 2.60
h	loop gain advantage through peaking	Eq. 14.27
I	dc or quiescent current (usually subscripted)	
I_{CO}, I'_{CO}	collector saturation currents with emitter and base open-circuited	Eqs. 4.44 and 4.47
I_{DO}	drain saturation current	Eqs. 4.123 and 4.136
i	incremental current (usually subscripted)	
$i_{Ni}{}^2{}_{eq}$	equivalent mean-square noise current referred to the input	Eq. 8.65*a*
J	current density (usually subscripted)	

K	flicker noise constant	Eq. 2.65
K_1, K_2	parameters associated with input and output resistances	Eqs. 5.21 and 5.35
k	(i) Boltzmann's constant (1.38×10^{-23} joule/°K)	
	(ii) ratio of screen to anode current	Eq. 3.46
	(iii) coefficient of coupling	Eq. 7.79
	(iv) number of identical groups of stages in an n-stage amplifier	
L	(i) inductance (usually subscripted)	
	(ii) channel length	Figs. 4.25 and 4.30
M	mutual inductance	
m	(i) mass	
	(ii) mass of electron (9.11×10^{-31} kg)	
	(iii) field factor	Eq. 4.27
	(iv) shunt peaking parameter	Eq. 13.37
	(v) mean time between failures	Eq. 17.10
m_R	mean time to repair	Eq. 17.14
N	(i) turns ratio of a transformer	
	(ii) number of units under test	Eq. 17.4
N_A, N_D	acceptor and donor concentrations	
N_F	number of failures	Eq. 17.4
n	(i) electron concentration	
	(ii) number of stages	
n_{eff}	number of effective poles	Eq. 13.120*b*
n_i	intrinsic concentration	Eq. 4.2
n_p, n_z	numbers of poles and zeros	
n.i.p.	number of ion pairs	Eqs. 8.126 and 8.131
P	power (usually subscripted)	
P_D	distortion power	Eq. 9.14
p	(i) hole concentration	
	(ii) general pole in the complex frequency plane	
p_{eff}	effective distance of poles from origin	Eq. 13.120*a*
Q	controlled mobile charge	Eq. 2.1
Q_C	controlling charge	Eq. 2.1

Q_{C1}, Q_{C2}	components of Q_C	Eq. 2.1
Q_{tot}	total mobile charge	Eq. 4.34
q	(i) charge of electron (1.60×10^{-19} coulomb)	
	(ii) incremental or signal charge (usually subscripted)	
R	(i) resistance (usually subscripted)	
	(ii) reliability	Eq. 17.7
R_{ext}	external component of load resistance	Section 5.1.6
R_L	total load resistance	Section 5.1.6
R_N	noise resistance	Eq. 8.125
R_P	parallel combination of all biasing resistors	Section 5.1.6
R_T	transfer resistance	Eq. 5.25
r	(i) incremental resistance (usually subscripted)	
	(ii) rate of carrier recombination	
	(iii) exponent in bandwidth reduction factor	Eq. 13.124
r_A	anode resistance	Eq. 3.37
$\bar{r}_A$	large-signal r_A	Eq. 6.16
r_B	base resistance	Fig. 4.2
r_C	collector resistance	Eq. 4.59
r_D	drain resistance	Eq. 4.129
r_E	emitter resistance	Eq. 4.38
S	sensitivity	Eq. 10.46
S/dN	spot signal-to-noise ratio (usually subscripted)	Eqs. 8.55
S/N	signal-to-noise ratio (usually subscripted)	Eqs. 8.58
s	complex frequency: $s = \sigma + j\omega$	
T	absolute temperature	
T_a	ambient temperature	
T_E	noise temperature	Eq. 8.122
T_K	cathode temperature	
T_S	effective source temperature	Section 8.4.1
TF	general transfer function	
TF_D	direct component of TF	Eq. 10.28

t	time	
t_r	rise time	Fig. 1.2
u_1, u_2	constants of charge-control theory	Eqs. 2.47 and 2.49
V	(i) dc or quiescent voltage (usually subscripted) (ii) electrostatic potential	
V_a	avalanche breakdown voltage	Eq. 4.81
V_{AB}	"knee" voltage	Fig. 3.12
V_e	anode voltage of equivalent diode	Eq. 3.27
V_i	contact or "built-in" voltage	Eq. 4.5
V_m	potential at barrier plane	Section 3.1.1.3
V_P	pinch-off voltage	Eqs. 4.116 and 4.133
v	(i) incremental or signal voltage (usually subscripted) (ii) velocity	
${v_{Ni}}^2_{eq}$	equivalent mean-square noise voltage referred to the input	Eq. 8.65*b*
W	width (usually subscripted)	
W_e	width of equivalent diode	Section 3.1.2.1
W_I	insulating barrier width	Fig. 4.25
W_n, W_p	depletion layer widths in *n*- and *p*-type material	Eqs. 4.10
W_0	width constant	Eq. 4.9
X	reactance (usually subscripted)	
x	(i) distance (ii) gain margin	Fig. 14.3
x_m	position of barrier plane	Section 3.1.1.3
Y	admittance (usually subscripted)	
Y_T	transfer admittance	Table 1.1
y	(i) incremental admittance (usually subscripted) (ii) phase margin	Fig. 14.3

Z	impedance (usually subscripted)	
Z_T	transfer impedance	Table 1.1
Z_0	characteristic impedance	
z	(i) incremental impedance (usually subscripted) (ii) general zero in the complex frequency plane	
α_B, α_C, α_E	components of α_N	Eqs. 4.75–4.78
α_I	counterpart of α_N for inverted operation	Section 4.2.2
α_N	common-base current amplification factor of a transistor in normal operation	Eq. 4.45
$\bar{\alpha}_N$	large-signal α_N	Eq. 4.83*a*
β	current amplification factor	Eqs. 2.57 and 8.4
β	feedback factor	Eq. 10.36
β_I	counterpart of β_N for inverted operation	Section 4.2.2
β_N	common-emitter current amplification factor of a transistor in normal operation	Eq. 4.46
$\bar{\beta}_N$	large-signal β_N	Eq. 4.83*b*
β_{NA}	average or nominal value of β_N	Table 4.7
Γ	space-charge smoothing factor	Eq. 2.64
γ	(i) differential error	Eq. 9.1
	(ii) charge-control parameter	Eq. 5.12
γ', γ''	differential errors at the signal peaks when the input is a sinusoid	Eqs. 9.17
ε	permitivity	
ζ	damping ratio for complex poles	Eq. 13.29
η	(i) coupling efficiency	Eq. 11.58
	(ii) power conversion efficiency	Eq. 16.17
	(iii) overshoot in step response	Fig. 1.2
	(iv) tilt in square wave response	Fig. 1.2

θ	thermal resistance	
λ	(i) common-mode rejection	Eq. 15.1
	(ii) failure rate	Eq. 17.4
μ	voltage amplification factor	Eqs. 2.58, 3.40, and 4.74
μ	reference variable	Eq. 10.36
$\bar{\mu}$	large-signal μ	Eq. 6.16
μ_e, μ_h	electron and hole mobilities	
μ_S	screen amplification factor	Eq. 3.42
ν	constant in feedback theory	Eq. 10.59
ξ	constant in high-frequency peaking	Eqs. 13.119
ρ	resistivity	
$\rho(t)$	time response function	
σ	(i) real part of complex frequency	
	(ii) specific heat	
τ	time constant (usually subscripted)	See ω
τ, τ_1	life time and mean transit time of carriers	Eqs. 2.3 and 2.2
τ_{th}	thermal time constant	Eq. 2.69
τ_C, τ_T, τ_1, τ_α, τ_β	characteristic time constants of a transistor	See ω_C, etc.
Υ	ratio of mean pole distances for four-terminal and two-terminal peaking	Eq. 13.131
φ_I, φ_V	mean-square spot noise current and voltage generators per hertz at the input	Eqs. 8.56
χ	bandwidth reduction factor	Eq. 13.122
$\psi(f)$, $\psi(j\omega)$, $\psi(s)$	frequency-dependent functions	Eq. 7.1
ψ_k	factor of ψ function	Eq. 7.2

ω	frequency in radians per second (usually subscripted)	
	Note. f, ω, and τ are used interchangably with $2\pi f_k \equiv \omega_k \equiv 1/\tau_k$	
ω_C	current gain-bandwidth product of a transistor	Eq. 4.110
ω_T	short-circuit current gain-bandwidth product of a transistor	Eqs. 4.71 and 4.73
ω_1	gain-bandwidth product of an intrinsic transistor	Eqs. 4.66 and 4.72
ω_α	characteristic frequency of a transistor, at which $\lvert i_C/i_E\rvert_{sc} = \alpha_N/\sqrt{2}$	Eq. 4.93
ω_β	characteristic frequency of a transistor, at which $\lvert i_C/i_B\rvert_{sc} = \beta_N/\sqrt{2}$	Eq. 4.108

List of Principal Subscripts

A	(i) anode (ii) indicates the average or nominal value given in manufacturers' data sheets (as in Tables 3.1 and 4.7)
B	base
B'	internal base
C	(i) collector (ii) coupling (iii) channel
com	common-mode
D	(i) drain (ii) decoupling (iii) indicates specifically that the parameter belongs to a device [as in $d(v_N^2)_D$]
dif	differential-mode
E	emitter
e	electron
eff	indicates the effective value [as in $R_{L(\text{eff})}$]
ext	indicates the external component [as in $R_{L(\text{ext})}$]
F	feedback
G	(i) grid (ii) gate
h	hole
i	(i) input (ii) indicates a parameter of an intrinsic device {as in $[Y_T(s)]_i$}

in	input; used when some aspect of the terminology is unusual
K	cathode
L	load
m	indicates the mid-band value (as in A_{Vm})
max	indicates the maximum value [as in $P_{C(\max)}$]
min	indicates the minimum value [as in $V_{G(\min)}$]
N	indicates noise; used to differentiate between signal and noise quantities (as in v_{Ni})
n	n-type semiconductor
o	output
out	output; used when some aspect of the terminology is unusual
oc	indicates the open-circuit value (as in μ_{oc})
opt	indicates the optimum value (as in n_{opt})
p	p-type semiconductor
pp	indicates the peak-to-peak value (as in I_{pp})
Q	indicates the quiescent value (as in V_{iQ})
S	(i) source of signal (ii) screen (iii) source electrode of f.e.t. (iv) indicates signal; used to differentiate between signal and noise (as in v_{Si})
sc	indicates the short-circuit value (as in μ_{sc})
tot	indicates the total value [as in $P_{D(tot)}$]
0	(i) indicates a specific value (as in ω_0) (ii) indicates the equilibrium value [as in $n_0(x)$] (iii) indicates the initial value (as in γ_0)
1, 2, 3	control, collecting and emitting electrodes of the general charge-control model

RULES FOR SUBSCRIPTS

VOLTAGES

(i) single subscript = voltage at electrode
e.g.:

V_A = dc voltage at anode
v_E = signal voltage at emitter

(ii) repeated subscript = electrode supply voltage
e.g.:

V_{AA} = anode supply voltage
V_{EE} = emitter supply voltage

(iii) double subscript = voltage difference between electrodes
e.g.:

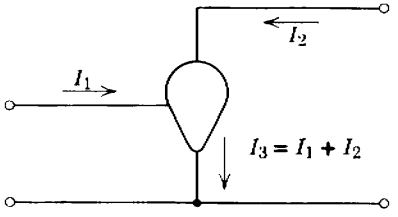
V_{AK} = dc anode-to-cathode voltage
v_{BE} = signal voltage between base and emitter

CURRENTS

(i) single subscript = electrode current
e.g.:

I_A = anode dc current
i_B = base signal current

(ii) reference directions for electrode currents thus:



ELECTRODE SUPPLY RESISTORS

single subscript = resistor between electrode and its dc supply
e.g.:

R_A = anode supply resistor
R_E = emitter supply resistor

ELEMENTS OF SMALL-SIGNAL EQUIVALENT CIRCUIT

double subscripts are used on unnamed incremental resistances and capacitances in the equivalent circuit of a device
e.g.:

c_{GK} = intrinsic grid-to-cathode capacitance
r_{GS} = gate-to-source resistance

Contents

Amplifying Devices and Low-Pass Amplifier Design

Chapter 1

Introduction to Amplifiers

Amplifiers exist in one form or another in nearly every item of electronic equipment. A knowledge of amplifier theory and design thus is basic to a comprehensive understanding of most electronic circuits. Often the presence of the amplifier sections is obvious; at other times it is disguised by the circuit arrangement. Obvious examples are such items as audio-frequency amplifiers, radio-frequency amplifiers, and direct-current amplifiers. Less obvious are the amplifier sections of oscillators, multivibrators, and regulated power supplies.

This chapter is a general introduction to amplifier theory and design; it emphasises the need for understanding a variety of background concepts and indicates the layout of the book.

1.1 BASIS OF AMPLIFICATION

The process of amplification is well-known intuitively but rather difficult to define in a complete and yet meaningful way. The elementary idea of amplification is based on normal English usage as suggested by the following extracts from the *Concise Oxford Dictionary*:

amplification (noun): extension, enlargement, making the most of a thing
amplify (verb): enhance, enlarge, expatiate
amplifier (noun): appliance for increasing the loudness of sounds, strength of [radio] signals, etc.

Thus amplification is commonly associated with an increase in the level of signals; for example, the small output signal from a microphone may be increased to the larger magnitude necessary for the operation of a loudspeaker. For this idea of amplification to be meaningful, however, care must be exercised in defining what is meant by "signal."

If signal refers to energy or power, there must be an increase in signal level between the input of an amplifying device and the output. Amplifying devices such as vacuum tubes and transistors may be thought of as energy-conversion devices in which a small amount of input signal power stimulates the extraction of a larger amount of signal power from a power reservoir—the power supply. The signal power extracted from the power supply is made available at the output and power amplification is achieved. The ratio of the output signal power to the input signal power is the *power gain*, which may be expressed as a numerical ratio but is more commonly expressed logarithmically in decibels:

$$(\text{power gain})_{\text{dB}} = 10 \log_{10} (\text{power gain})_{\text{numeric}}. \tag{1.1}$$

This is probably the most general picture of the amplification process and may be applied either to conventional amplifiers that use a dc power supply or to maser and parametric amplifiers that use a time-dependent power supply.

If signal refers to time-varying voltages or currents, the signal level does not necessarily increase between the input and output of an amplifying device; to be useful an amplifier need not necessarily have a voltage gain greater than unity. Indeed, if the input signal is a voltage and the output signal is a current, the idea of an increase in level is meaningless. It is necessary to think of an amplifier as a circuit that preserves a certain functional relationship between the input and output signals. The fact that power gain is achieved is not always obvious from the functional relationship, but, of course, the functional relationships that are realized with amplifiers cannot be realized with purely passive circuits. Table 1.1 lists the more important *transfer functions*—that is, the combinations of input and output signal types.

Both signal power and signal voltage or current are useful concepts, but the second is particularly convenient for amplifiers of the kind considered in this book. Ratios of currents and/or voltages are more simply controlled and calculated than power ratios. Consider, for example, the ideal voltage amplifier shown in Fig. 1.1*a*. For a constant input voltage $v_i(t)$ the output voltage $v_o(t)$ remains constant, irrespective of the value of the load resistance R_L; the power gain, on the other hand, is given by

$$\text{power gain} = \frac{v_o^2}{R_L} \frac{1}{P_i}, \tag{1.2}$$

where P_i is the input signal power. For constant input power (which often corresponds to constant input voltage) the power gain is a function of the load resistance. Voltage gain is an invariant quantity, whereas

Table 1.1 Most-Used Transfer Functions

Input Signal	Output Signal	Transfer Function
Voltage	Voltage	Voltage gain $A_V = v_o/v_i$
Current	Current	Current gain $A_I = i_o/i_i$
Voltage	Current	Transfer admittance $Y_T = i_o/v_i$
Current	Voltage	Transfer impedance $Z_T = v_o/i_i$
Charge	Voltage	Transfer elastance $S_T = v_o/q_i$

power gain is not. The remaining parts of Fig. 1.1 illustrate three other useful examples the invariance of current gain, transfer admittance, and transfer impedance. The amplifier is represented in these four diagrams by a triangular-shaped box that indicates the direction of signal flow. To be useful all of these circuits must amplify in the sense that their power gain is greater than unity, but the performance is most conveniently specified by the relevant ratio of voltages and/or currents. It is customary

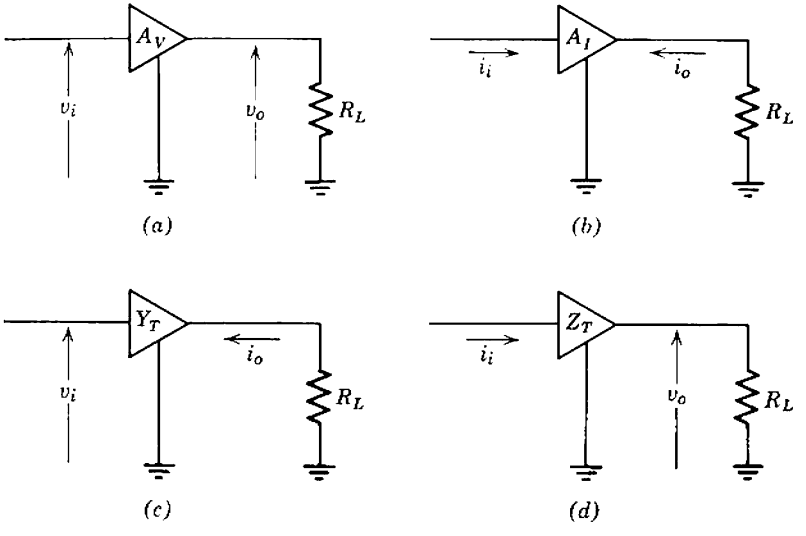


Fig. 1.1 Transfer functions: (*a*) voltage gain; (*b*) current gain; (*c*) transfer admittance; (*d*) transfer impedance.

to specify voltage and current gain as a numerical ratio, although logarithmic units such as decibels are sometimes used:

$$\begin{pmatrix}\text{voltage or}\\ \text{current gain}\end{pmatrix}_{\text{dB}} = 20 \log_{10} \begin{pmatrix}\text{voltage or}\\ \text{current gain}\end{pmatrix}_{\text{numeric}} \tag{1.3}$$

Transfer admittance and transfer impedance are dimensional quantities, usually expressed as so many milliamperes per volt or volts per milliampere. It is left as an exercise for the reader to verify that the presence of the multiplying factor 20 in Eq. 1.3 enables logarithmic power gain to be determined directly from logarithmic voltage or current gain, provided that the ratio of input and load resistances is known.

The ideal amplifiers with constant transfer functions shown in Fig. 1.1 cannot be realized in practice. As R_L approaches zero, A_V and Z_T must also approach zero, for no physical device can provide the infinite output current required to maintain the output voltages. Similarly, as R_L approaches infinity, A_I and Y_T must approach zero, for no physical device can provide an infinite output voltage. Consequently, the transfer function A_V, A_I, Y_T, or Z_T of a practical amplifier is approximately invariant only for a limited range of R_L; beyond this range the transfer function depends on the load. However, the analysis and design of a complete amplifier will be so much simplified if the transfer functions of the component building blocks can be made independent of their loads that every effort should be made to have the building blocks approach this ideal. The use of building blocks with a constant transfer function is the cornerstone of the design philosophy set out in this book.

1.2 FIDELITY OF AMPLIFICATION

Ideally, the output signal waveform from an amplifier should be an exact replica of the input signal or be related to it in some precise functional way such as ideal integration. All practical amplifiers depart from this ideal; the accuracy of performance is usually known as the fidelity of the amplification. As might be guessed from the excessive use of the term "high fidelity" in advertising, there is no single quantitative measure of fidelity. However, the individual factors that limit fidelity are capable of fairly precise definition and should be discussed. These factors are dynamic response, noise, hum, and microphony.

1.2.1 Dynamic Response

Dynamic response may be defined as the response to steady or time-varying input signals of large or small amplitude. There is a fundamental limitation to dynamic response, and further limitations may be

introduced artificially. As the rate of change of an input signal is increased, the inherent inertia of the charge in the amplifying devices does not allow the corresponding rate of change to be reached by the output. Therefore the ideal functional relationship between the input and output is not preserved for signals with rates of change greater than a certain limiting value, and the dynamic response can be represented by a constant transfer function only for signals whose rate of change is less than this critical value. Alternatively, because time functions can be synthesized from a Fourier spectrum of sinusoids, the dynamic response is represented by a constant only for signals whose Fourier components lie below a certain maximum frequency. Artificial limitations to the dynamic response of an amplifier are often introduced to simplify some aspects of the design. Coupling capacitors and transformers (Chapter 6) are used to limit the dynamic response for signals with very low rates of change or, alternatively, signals whose Fourier components lie below a critical frequency. In most cases some of the Fourier components of the input signal lie outside the critical band of frequencies, and consequently there is some loss of signal information in any amplifier. Minimum acceptable levels of performance for different applications are discussed in Chapter 7.

1.2.1.1 Small-Signal Dynamic Response: Fundamental Limits

For signals of very small amplitude—the so-called *small-signal conditions*—the dynamic response is independent of signal amplitude; the amplifier behaves as a linear network. The dynamic response of a linear network may be specified in a number of ways, the two most common being its *time response* and *frequency response*.

Time response relates the shape of the output waveform to a time waveform at the input. The two most often used input waveforms are the step function and the square wave. Figure 1.2*a* shows a typical response

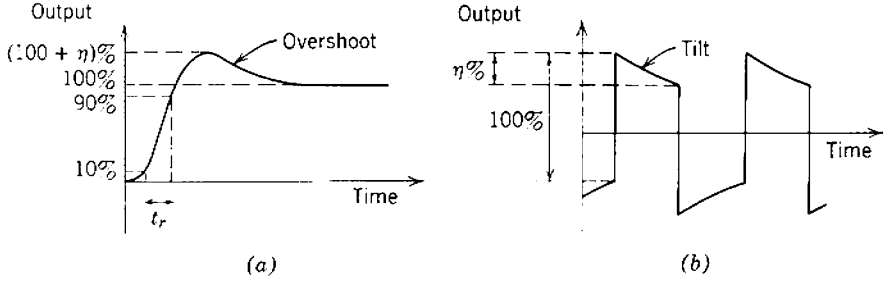


Fig. 1.2 Time response: (*a*) rise time and overshoot; (*b*) percentage tilt (referred to in Section 1.2.1.2). The time scale is compressed in (*b*) relative to (*a*) by several orders of magnitude; (*a*) is the detail of the vertical edge in (*b*).

waveform for a step-function input. The shape of this output waveform is often defined incompletely but usefully by quoting the 10 to 90% *rise time* t_r and the maximum *overshoot* as a percentage η of the final amplitude of the waveform.

If a sinusoidal signal of constant amplitude but variable frequency is applied to the amplifier, it is found that the output signal is approximately constant in amplitude for frequencies below a certain value, whereas above this frequency the output falls away. The *cutoff frequency* is the frequency at which the output voltage or current (and therefore the gain) falls below some specified fraction. This fraction is usually (but not always) taken as

$$\frac{1}{\sqrt{2}} = 0.707.$$

On a decibel scale it corresponds to a fall in gain of 3 dB; the power gain is halved. The *bandwidth* $\mathscr{B}$ of an amplifier is the frequency range over which its gain is constant within the specified limits, and for a low-pass amplifier which has no artificial restrictions on its dynamic response the bandwidth is numerically equal to the cutoff frequency. Figure 1.3*a* illustrates cutoff frequency and bandwidth; the significance of the generic name *low-pass amplifier* is apparent.

The time and frequency response of a linear network are interdependent, and the complete way of specifying both is by means of the *complex frequency plane*. The complete specification of the small-signal transfer function of an amplifier involves a knowledge of the position of its poles and zeros in the complex frequency plane, together with a multiplying constant ν:

$$TF(s) = \nu\left[\frac{(1 - s/z_1)(1 - s/z_2)\cdots}{(1 - s/p_1)(1 - s/p_2)\cdots}\right], \tag{1.4}$$

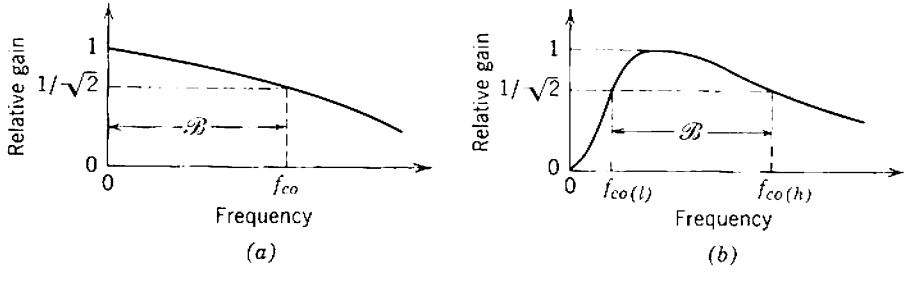


Fig. 1.3 Frequency response: cutoff frequency and bandwidth. Part (*b*) is referred to in Section 1.2.1.2.

where $TF(s)$ is the transfer function (A_V, A_I, Y_T, Z_T, etc.),
$z_1, z_2 \cdots$ are the zeros,
$p_1, p_2 \cdots$ are the poles.

Except for the special cases discussed at the end of Section 7.3.1, the constant ν is the *mid-band transfer function* TF_m of the amplifier (A_V, A_I, G_T, R_T, etc.). For an amplifier with no artificial restrictions on its dynamic response the true transfer function approaches the ideal constant mid-band transfer function only for small signals whose complex frequency approaches zero. Therefore the dynamic response can be represented exactly by a constant only for signals whose rate of change (for a time waveform) or whose frequency (for a sinusoid) approaches zero. However, the dynamic response is approximately constant for signals whose modulus of complex frequency is somewhat less than the modulus of the innermost (nearest to the origin) singularity. It is a good approximation to represent the dynamic response by a constant over this working range of complex frequency; from Eq. 1.4

$$TF(s) \approx \nu = TF_m. \tag{1.5}$$

Because of its ready application to small-signal problems, linear circuit theory is an extremely important tool in amplifier design, and a knowledge of manipulation in the complex frequency plane is assumed in this book.

1.2.1.2 Small-Signal Dynamic Response: Artificial Limits

Artificial limiting of dynamic response by means of coupling and bypass capacitors can be specified in the same general ways as the fundamental limitation due to inertia. Incomplete but useful specifications are the deformation of a time waveform and the cutoff frequency for sinusoids; complete specification is by means of a singularity plot in the complex frequency plane. Figure 1.2*b* illustrates the most common specification of waveform deformation, the *percentage tilt* or *sag* on square wave, and Fig. 1.3*b* illustrates the *lower 3 dB cutoff frequency*. The *bandwidth* of an amplifier is the difference between the upper and lower cutoff frequencies, and the range of frequencies between the cutoff points is called the *passband*.

An amplifier with restricted low-frequency response might be called a band-pass amplifier, for it passes only the band of frequencies between two limits. However, the name low-pass amplifier is retained, for the low-frequency restriction is introduced artificially. The name *band-pass amplifier* is reserved for the class of tuned amplifier in which the fundamentally limited passband does not extend to zero frequency. Low-pass amplifiers are sometimes called *base-band amplifiers*. Special low-pass

amplifiers whose low-frequency response is not restricted are called *dc amplifiers* because their useful response extends to zero frequency.

1.2.1.3 Large-Signal Dynamic Response

The complex frequency plane representation of dynamic response is applicable only to linear systems. Active devices such as vacuum tubes and transistors always have nonlinear electrical characteristics, and the s-plane representation is formally invalid except for the special case of infinitesimally small signals assumed throughout Sections 1.2.1.1 and 1.2.1.2. As the signal input to a practical amplifier is increased in magnitude, two effects occur:

1. There is some failure of the proportionality between input and output signals; successive equal increments of the input signal do not produce corresponding equal increments of the output signal.
2. The bandwidth, rise time, and tilt become functions (usually complicated functions) of the signal amplitude.

The signal amplitude at which nonlinearity in the devices manifests itself as an amplitude-dependence of the response is called the *maximum of dynamic range* or, more simply, the *dynamic range*. A common specification of dynamic range is the largest amplitude of input voltage or current for which the response is independent of signal amplitude, within specified limits.

An important simplification occurs in the special case of slowly varying signals for which inertia effects are negligible and the transfer function $TF(s)$ can be approximated by TF_m. Only the time-independent nonlinearities are important and their effect can be deduced from a knowledge of the dc characteristics of the devices. For example, the dc characteristic relating anode voltage V_A to grid voltage V_G for the vacuum-tube amplifier shown in Fig. 1.4*a* has the general form shown in Fig. 1.4*b*. For normal amplifier operation the vacuum tube must be operated within the limits set by zero anode current (the *cutoff* limit) and anode current maximum (the *saturation* or *bottoming* limit). Typically, the quiescent grid potential V_{GG} is set as shown in Fig. 1.4*c*. If a slowly varying (i.e., low-frequency) input sinusoid of very small amplitude is superimposed on this quiescent grid voltage, the resulting signal component of anode voltage is very close to a perfect sinusoid—waveform (i) in Fig. 1.4*c*. If the input is increased to the order of the amplitude shown by waveform (ii), the output waveform departs from a true sinusoid; the negative excursion of the anode voltage exceeds the positive excursion. This type of *distortion* is considered in Chapter 9. Finally, if the input sinusoid is increased to the order of

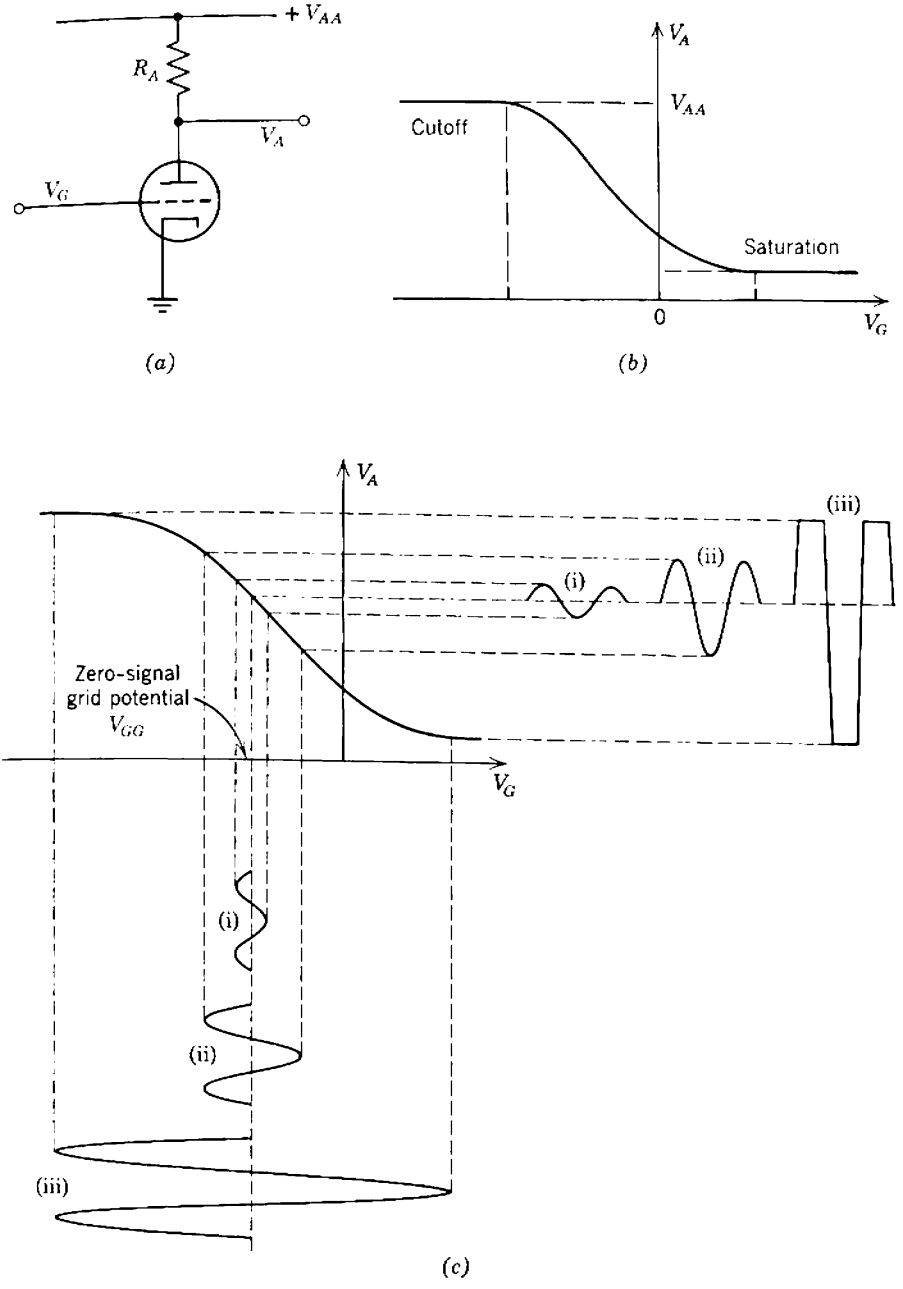


Fig. 1.4 Distortion: (*a*) circuit diagram; (*b*) transfer function; (*c*) input versus output for various signal amplitudes.

amplitude indicated by waveform (iii), the tube is driven out of its normal operating range on both positive and negative signal excursions and gross distortion results. Gross distortion must always be avoided in low-pass amplifiers; the tolerable level of distortion due to the nonlinearity within the normal range of operation depends on particular applications.

If the input sinusoid is increased in frequency to such an extent that $TF(s)$ cannot be approximated by TF_m, the anode and grid voltage sinusoids in (i) do not remain 180° out of phase. Furthermore, the output waveforms in (ii) and (iii) do not remain symmetrical about their peaks and cannot be determined from the dc transfer characteristic alone. The determination of accurate output waveforms is extremely tedious. Often, high-frequency (or fast transient) problems can be investigated by assuming that the dynamic range is set only by the gross nonlinearities of saturation and current cutoff.

Broadly, distortion may be thought of as the addition of corrupting Fourier components to the signal as it passes through an amplifier.

1.2.2 Noise

Noise is a completely different mechanism which adds corrupting components to the signal; it arises from the randomness in the motion of the charged carriers in electrical and electronic processes. Components of noise are associated with the flow of electrons and holes in a semiconductor and the flow of electrons in a vacuum or conductor. Unlike distortion, noise is important when the signal amplitude is small; noise in a device sets a lower limit to the amplitude of signals for which it is a satisfactory amplifier—that is, the minimum of dynamic range. Noise sources are discussed in Chapters 2 to 4, and means for minimizing the total noise of an amplifier are considered in Chapter 8.

1.2.3 Hum and Microphony

Hum and microphony also are mechanisms for producing corrupting components which add to the signal. Like noise, hum and microphony are important when the signal levels are small and may set a lower limit to the signal input for which an amplifier is satisfactory. Unlike noise, these two man-made corrupting mechanisms can theoretically be reduced to negligible proportions. Hum is caused by the leakage of power-mains-frequency voltages or currents into the signal path of an amplifier and may be introduced via a conduction path or by electrostatic or electromagnetic induction. Microphony is caused by internal vibration when an amplifying device is subjected to mechanical shock.

1.3 THE PHYSICAL REALIZATION OF AMPLIFIERS

Amplifiers use one or more amplifying devices, usually vacuum tubes, transistors, or both. The combination of one amplifying device and its associated passive circuitry is generally known as an *amplifier stage*; multistage amplifiers are interconnections of a number of stages. It should be noted that this definition of stage should not be taken too literally and that in some advanced circuits it is convenient to group two or more devices together to constitute one stage. Nevertheless, it is convenient to assume that this arbitrary definition is valid and to consider the problems in realizing single- and multistage amplifiers.

1.3.1 Single-Stage Amplifiers

The small-signal behavior of an isolated amplifying device can be represented by a small-signal equivalent circuit that involves resistors, capacitors, and voltage or current sources. For example, it is well known that the equivalent circuit shown in Fig. 1.5 represents the small-signal behavior of a vacuum tube at low frequencies. The apparent voltage gain of this circuit is given by

$$\frac{v_{AK}}{v_{GK}} = -g_m r_A \tag{1.6}$$

and its magnitude is 100 for a typical triode, type ECC83/12AX7. (Lowercase letters v and i represent signal quantities, i.e., changes about the steady-state or quiescent dc values; dc voltages and currents are represented by capital letters V and I.) This apparent voltage gain can never be achieved in practice because the equivalent circuit cannot exist in isolation. The equivalent circuit is a valid representation of performance only when the tube is biased so that

(i) the anode is positive with respect to the cathode,
(ii) the grid is negative with respect to the cathode,
(iii) a finite anode current flows.

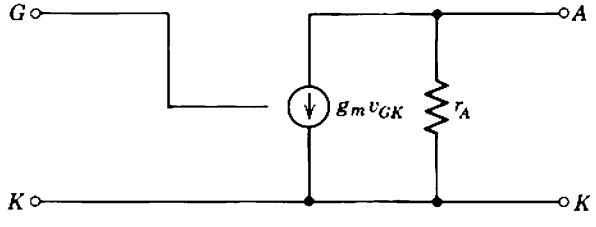


Fig. 1.5 Small-signal equivalent circuit for a vacuum tube at low frequencies.

Some form of *biasing circuit* must be used to ensure that these *quiescent conditions* exist, and this biasing circuit invariably has some effect on the small-signal performance. The small-signal equivalent circuit of an amplifier stage consists of the device equivalent circuit in combination with other elements. Usually the performance of an amplifier stage is inferior to the performance inherent in the device itself.

Consider the influence of the biasing circuit on small-signal performance in the three simple examples of Fig. 1.6. In the trivial case of Fig. 1.6*a* two batteries set the necessary quiescent conditions. This is absurd from the signal viewpoint, however, for no signal voltage can be applied to the grid or developed at the anode. In Fig. 1.6*b* a signal voltage source is placed in series with the grid bias battery. The circuit has a finite transfer conductance given by

$$\frac{i_o}{v_S} = \frac{i_A}{v_{GK}} = g_m, \tag{1.7}$$

but, because no signal voltage can be developed at the anode, the voltage gain is zero. The final step of adding a resistor R_A in series with the

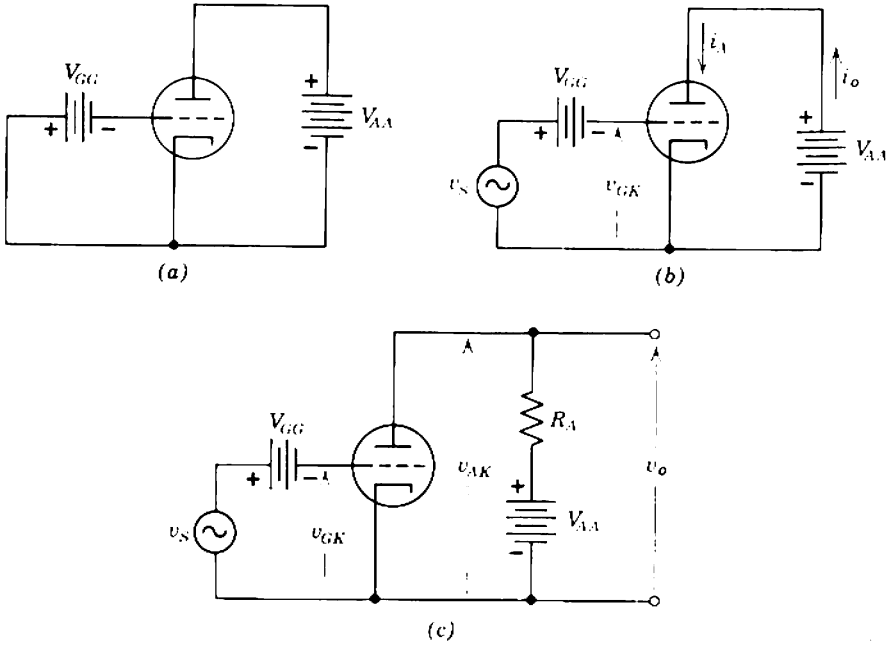


Fig. 1.6 Evolution of a voltage amplifier: (*a*) biasing batteries; (*b*) introduction of input signal; (*c*) load resistor and output voltage.

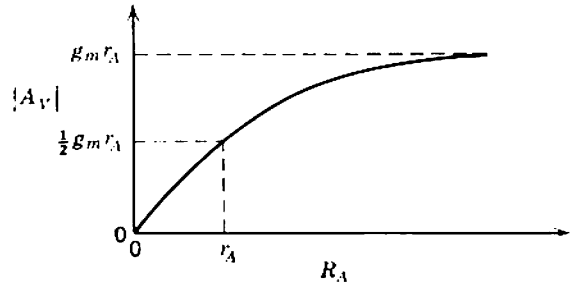


Fig. 1.7 Variation of gain with load resistance.

anode supply battery as in Fig. 1.6*c* allows a finite voltage gain to be achieved:

$$\frac{v_o}{v_S} = \frac{v_{AK}}{v_{GK}} = -g_m r_A \left(\frac{1}{1 + r_A/R_A} \right). \tag{1.8}$$

The variation of voltage gain with anode load resistor is shown in Fig. 1.7; the gain increases monotonically with R_A toward the limiting value $g_m r_A$. However, the quiescent anode current flows through R_A, and a significant dc potential is developed across it. This sets an upper limit to the value of R_A which can be used with a given supply potential. Suppose an ECC83/12AX7 is to be operated at 0.5 mA anode current and that the anode-to-cathode voltage is to be not less than 150 V. If a 200-V power supply is available, the maximum value for R_A is 100 kΩ. But r_A for the tube is 100 kΩ at 0.5 mA anode current. Therefore the maximum attainable gain is given by Eq. 1.8 as 50; this is only half the value of 100 inherent in the tube itself.

The problem of achieving a satisfactory compromise between signal and bias circuit requirements is important with all devices. In principle, the use of an inductor or transformer as shown in Figs. 1.8*a* and *b* provides an

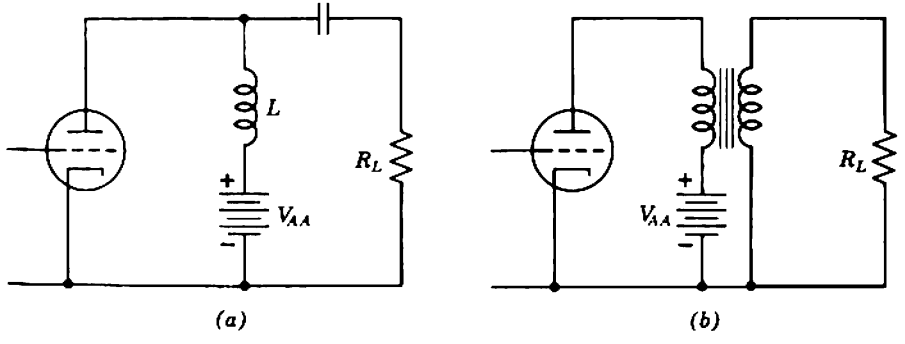


Fig. 1.8 Biasing with inductors or transformers.

elegant solution. Ideally, there is negligible dc voltage drop across the inductor or the primary winding of the transformer. Because no dc current flows in the load resistor R_L, R_L can be made large and a high voltage gain is obtained. In practice, these circuits are often unsatisfactory because of the difficulty in realizing inductors and transformers that preserve their characteristics over a large bandwidth. However, these circuits find extensive use in band-pass or tuned amplifiers.

It might be expected from the foregoing remarks that the function of a biasing circuit is merely to ensure that the amplifying device will operate somewhere within its range of quiescent conditions. In fact, reasonable precision is required in the biasing circuit because of the strong dependence of many critical parameters in the small-signal equivalent circuit on the quiescent conditions. Furthermore, there are definite limitations to the range of quiescent conditions that can be used with a particular device because of the effects of excessive power dissipation and heating or other mechanisms of device deterioration; these limits are specified by *device ratings*.

Thus, to design small-signal amplifier stages, a knowledge of the following is required:

(i) the small-signal equivalent circuit and the laws of variation of the small-signal parameters,
(ii) linear circuit theory,
(iii) the large-signal (nonlinear) performance as it affects biasing circuit design,
(iv) device ratings,
(v) noise performance of amplifying devices.

Topics (i), (iii), (iv), and (v) are most logically introduced by way of device physics and are covered in Chapters 2, 3, and 4; biasing circuits are considered in Chapter 6, and the circuit theory of simple amplifier stages is given in Chapters 5 and 7. More advanced amplifiers usually employ a circuit technique known as *feedback*, which is introduced in Chapter 10.

1.3.2 Multistage Amplifiers

The magnitude of the voltage gain, current gain, transfer admittance, or transfer impedance that can be achieved with a single-stage amplifier has definite limits. For these limits to be exceeded a number of stages must be joined together to produce a multistage amplifier. One or both of the following interconnections may be used.

In *multiplicative amplification* the signal passes successively through a line of cascaded single stages, as indicated in Fig. 1.9*a*. Each stage, in

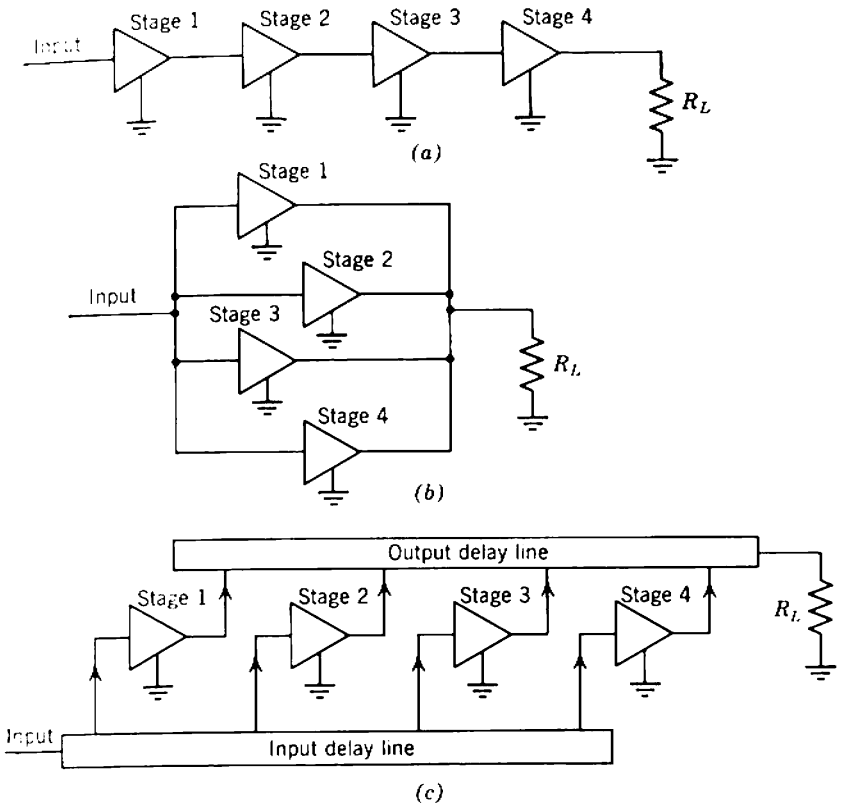


Fig. 1.9 Multistage amplifiers: (*a*) multiplicative amplification; (*b*) split-band amplification; (*c*) distributed amplification.

turn, adds to the over-all power gain. The transfer function of the multistage amplifier is the product of the individual stage transfer functions, provided that this multiplication is dimensionally consistent; the individual stage transfer functions must be such that the output signal type (voltage or current) from one stage is the input signal type to the following stage. An important simplification results if the transfer function chosen for any stage has the same value both when the stage exists in isolation and when it is interconnected with the other stages. The transfer functions of the stages can then be calculated individually; the over-all transfer function is the product of these independent quantities. If this simplification is not achieved, the transfer function of any stage within a multistage amplifier depends in part on the following stage; *interaction* is said to occur, and the design of one stage affects the performance of those adjacent. The approaches to design considered in this book aim at minimizing or controlling interaction between stages, thereby simplifying amplifier design.

In *additive amplification* the signal outputs from a line of parallel single stages are added. At least two methods are used.

1. In the *split-band* method the individual stages are arranged as shown in Fig. 1.9*b* and each accepts only those signals that lie within a certain narrow frequency range. For reasons that are discussed in later chapters these narrow-band stages may have fairly high gains; the product of the stage gain and bandwidth is a constant. The passbands of the individual stages are arranged so that the complete amplifier covers the required over-all frequency range, and their gains are made equal.

2. In a second method of additive amplification the individual stages are arranged with their inputs and outputs connected by separate delay lines, as shown in Fig. 1.9*c*. If the two delay lines have equal propagation velocities, the input signal is applied to each stage in turn and their outputs add. This extremely important mode of operation for amplifiers is known as *distributed amplification.*

1.4 COMPONENTS FOR AMPLIFIERS

The most important components in amplifiers are the active devices, and good circuit design must be based on a knowledge of their physical properties. Active device parameters change with quiescent operating voltage and current and usually change with temperature and time; they may also change with pressure, humidity, and under conditions of mechanical shock. In addition, because modern amplifying devices are mass produced, significant unit-to-unit variations arise from finite manufacturing tolerances. If the effects of all these possible sources of variation are estimated by the designer, certain expected ranges of device-parameter variation may be estimated. Some parameters vary only slightly, perhaps $\pm 2\%$, whereas others may vary over the range from half to three times the expected average value.

Amplifiers use components other than active devices. The two most common additional circuit element types are resistors and capacitors, with inductors, transformers, and certain nonlinear elements such as thermistors occurring less frequently. These *passive circuit elements* are available over a wide range of values and in a number of different grades. The following ranges for the values of resistors and capacitors might be considered as normal, although values outside these ranges can be obtained.

Resistors:

1 Ω to 10 MΩ.

Capacitors:

1 pF to 1000 μF.

The normal grade of resistor or capacitor has a tolerance of $\pm 10\%$ on its value, although some obsolescent types have $\pm 20\%$. This standard of $\pm 10\%$ is based partly on the economics of manufacture but also has been established because more precise components could not be justified in the light of vacuum-tube and transistor parameter tolerances. Better-grade components have $\pm 5\%$ tolerance and often rather superior characteristics—lower excess noise for a resistor, lower leakage for a capacitor, and so on. Precision-grade components with tolerances as close as $\pm 1\%$ and decidedly superior characteristics are fairly readily available and still closer tolerances can be obtained if required. Some types of capacitor, notably electrolytics and high-k ceramics, have guaranteed minimum values: typically their tolerance is -0 to $+100\%$.

The foregoing comments apply specifically to amplifiers in which discrete, individually made active devices and passive components are assembled to produce a complete circuit. An increasing number of modern circuits are so fabricated that the devices and components are produced together within a single *chip* of silicon. Section 17.3 contains an introduction to the technology of these *integrated circuits*. Their design may differ in some points of detail from the design of discrete-element circuits because a smaller range of passive component values is available and because manufacturing tolerances are different; the general principles, however, are unaltered. Realizable integrated resistor and capacitor values usually lie below 100 kΩ and 1 nF, respectively. Absolute tolerances on component values are poor, but tolerances on ratios tend to be good.

The designer must constantly be aware of the existence of these parameter tolerances and use circuit techniques that minimize their effect on over-all performance. It is then possible to obtain extremely precise performance from amplifiers that use mainly unselected devices and components. Furthermore, it is possible to design these circuits to any degree of precision completely on paper, without the necessity of resorting to experimental fiddling with components. This is by far the most satisfactory approach to circuit design, but strangely enough it has not found universal acceptance. Indeed, the following quotation from a paper by F. C. Williams* is just as valid today as it was when written in 1946:

"In approaching the problems of precision, reliability and reproducibility, the author and his colleagues have rightly or wrongly assumed that these three requirements reduce to a single requirement, 'designability,' that is, the development of circuits whose operation can be predicted accurately

* F. C. Williams, "Introduction to circuit techniques for radiolocation," *J. Inst. Elec. Engrs.* (*London*), **93**, Part III A, 289, 1946.

before they are built. This is an accepted principle in most branches of engineering but, owing to the many imperfections of [vacuum tubes] and components, the specialized nature of the necessary design work, and the extreme ease of assembly and modification of experimental equipment, it has not found universal acceptance in the field of electronics. A circuit which is designed to give a certain result and which is shown to give that result by experiment with a single sample will probably be reproducible in bulk and will probably work reliably in service, provided the design has taken care of component tolerances and the conditions of service as regards temperature range, humidity, and so on."

1.5 STANDARD PRACTICE AND CREATIVITY

The following extract from the *Report on Evaluation of Engineering Education, 1952–1955*, presents a pithy general picture of what is involved in engineering design:

"The capacity to design includes more than mere technical competence. It involves a willingness to attack a situation never seen or studied before and for which data are often incomplete; it also includes an acceptance of full responsibility for solving the problems on a professional basis."

At almost all levels of design the designer must be aware of the conflict existing between standard practice and creativity. Standard practice is not necessarily correct nor is creativity necessarily desirable in a particular problem. Thus the ideal designer should be aware of standard practices and their limitations and at the same time should be able to think creatively when necessary. There are plenty of so-called designers who have found their own niche in an organization and spend their time in mechanical repetition of particular design procedure for, say, a filter or transformer. This is certainly not design in its highest sense and may be more economically performed by a computer. Mere concentration on standard practice is not in itself rewarding.

This book discusses specific methods of design that have been found useful, and in this sense they are standard. However, the background philosophy is capable of application in completely new situations and, it is hoped, will assist readers who wish to go beyond a professional life of mechanical repetition.

Chapter 2

Characteristics of Amplifying Devices I: General Introduction

There have been two major inventions in the field of amplifying devices. The first was the vacuum triode invented by de Forest in 1905 and the second was the point-contact transistor invented by Shockley, Brattain, and Bardeen in 1948. Higher gain tetrode and pentode vacuum tubes have been developed from the basic triode and an entire family of semiconductor devices from the point-contact transistor. This book is concerned principally with the circuit applications of triode and pentode vacuum tubes and bipolar transistors, for these are the standard amplifying devices in use at present. Most of the concepts can be reapplied to field-effect transistors (these are mentioned briefly) and to other three-terminal devices that may be developed in the future. Two-terminal devices such as the tunnel diode are not considered.

The basic objectives of a circuit designer are to choose the device type best suited to a given application and to use it to best advantage. The initial choice may depend on many factors—system requirements, environmental conditions, economics—but in every case it will be strongly influenced by the general electrical characteristics of the competing devices. The final circuit design requires a more detailed knowledge of these characteristics. Some scheme is therefore required from which device behavior can be predicted, both in outline and detail, under any conditions of operation. Information is required on

(i) the *dynamic response* to applied signals of large or small amplitude,
(ii) the corrupting *noise* produced within the device,
(iii) the power, voltage, and current limitations or *ratings* of the device.

These aspects of device performance are introduced in the following sections; dynamic response is considered in Sections 2.1 to 2.5, noise in Section 2.6, and ratings in Section 2.7.

2.1 CHARACTERIZATION OF DYNAMIC RESPONSE

Many models have been suggested for characterizing the dynamic response of electronic devices—graphical characteristics, accurate and approximate analytical expressions, linear approximations of various types—and still there are others. In deciding on the model that is most likely to be useful in circuit design, the following points should be noted:

1. The same model should be applicable to a number of electronic devices and to a number of modes of operation of a given device. This provides a basis for comparing and contrasting the performance of competing devices.
2. The model should be directly related to fundamental physical processes, but for the particular case of signals that are small enough for the device parameters to be considered as constants the model should reduce to a form suited to direct use in linear circuit theory. This removes the gap that would otherwise exist between "device" and "circuit" concepts.
3. The model should have parameters that can be measured easily and have known laws of variation with operating conditions.
4. The model should be as accurate as necessary for normal use and for some slightly abnormal uses. It should not be more accurate than warranted by the numerical data available if this increased accuracy is achieved by increased complexity.
5. Although the basic form of the model should be simple, it should lend itself to expansion to allow for those second-order effects that are sometimes significant.

One model that satisfies these requirements is based on the charge-control theory of amplification.

2.2 CHARGE-CONTROL THEORY OF AMPLIFICATION*

A number of electronic devices, including the vacuum tube and transistor, consist of four basic parts:

(i) an *emitting electrode* which emits charged carriers,

* The following references are important and will repay study: R. BEAUFOY and J. SPARKES, "The junction transistor as a charge-controlled device," *ATE J.*, 13, 310,

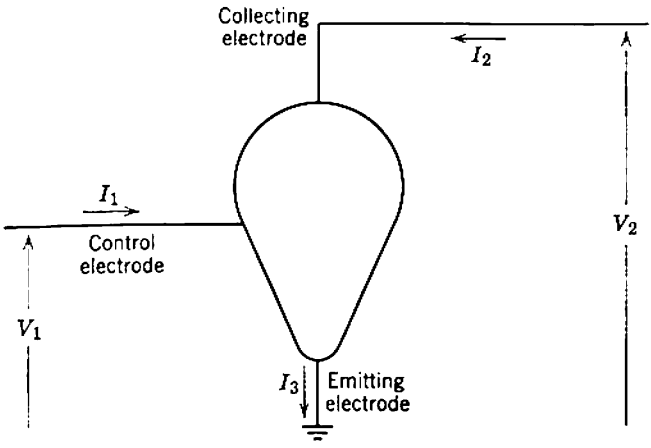


Fig. 2.1 The charge-control model.

(ii) a conducting channel or *control region* through which the carriers move,
(iii) a *collecting electrode*,
(iv) a gate or *control electrode* which (largely) controls the carrier flow.

Such devices can be represented by the charge-control model (Fig. 2.1).

The variable from which the charge-control model takes its name is the mobile carrier charge in the control region, for charge-control theory recognizes that the electrode currents are uniquely related to this charge and its time derivatives. Even so, the mobile carrier charge is known as the *controlled charge* Q because it is under the influence of another *controlling charge* Q_C. This controlling charge exists principally as a component Q_{C1} on the control electrode, although a small portion Q_{C2} can exist on the collecting electrode. In general, therefore, Q_C is a function of both the input voltage V_1 and the output voltage V_2. Because there can be no net accumulation of charge in the model, the steady-state values of Q and Q_C must be equal and opposite:

$$-Q = Q_C = Q_{C1} + Q_{C2}. \tag{2.1}$$

In a vacuum tube the two sets of charge are quite separate. The controlled charge Q is the excess electron space charge that is stored principally in the cathode-grid space when the tube conducts, whereas

---

October 1957. E. JOHNSON and A. ROSE, "Simple general analysis of amplifier devices with emitter control and collector functions," *Proc. Inst. Radio Engrs.*, **47**, 407, March 1959. R. D. MIDDLEBROOK, "A modern approach to semiconductor and vacuum device theory," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 887, Suppl. 17, 1959.

the positive controlling charge Q_C exists principally on the control electrode (the grid) and raises its potential above the very negative value necessary to cut off the anode current. A small but significant component of Q_C also exists on the anode (or screen of a pentode) so that the controlling charge Q_{C1} on the grid is not equal to $-Q$. Therefore Q is not entirely under the control of the grid voltage. The unipolar (field-effect) transistor is similar in that the two sets of charge are well separated but differs in that ideally there is no component of Q_C associated with the collecting electrode, the drain.

The two sets of charge in a bipolar transistor coexist in the base region and are very close together. The controlling charge is the majority carrier charge introduced into the base region via the base contact, whereas the controlled charge is the minority carrier charge injected into the base from the emitter. There is no component Q_{C2} associated with the collector. In an *n-p-n* transistor Q_C is a hole charge and Q is an electron charge; in a *p-n-p* transistor Q_C is an electron charge and Q is a hole charge.

The process of amplification (or control) consists of changing the controlling charge Q_C by means of the input voltage, thereby changing the controlled charge Q. The resulting change in output current can be used to produce a change in output voltage across a finite load impedance. However, such a change in V_2 tends to reduce the initial changes in Q_C and Q, so a useful voltage gain will be achieved only if V_1 has a greater effect on Q_C than does V_2. This implies that

$$\left(\frac{\partial Q_C}{\partial V_1}\right)_{V_2} > \left(\frac{\partial Q_C}{\partial V_2}\right)_{V_1}$$

for a useful device.

2.2.1 Static Conditions

Under conditions of constant electrode voltages the electrode currents and the charges Q and Q_C are constant also. The relation between the collecting-electrode current I_2 and the charge can be written as

$$I_2 = -\frac{Q}{\tau_1} = \frac{Q_C}{\tau_1}, \tag{2.2}$$

where τ_1 is the *mean transit time per carrier* through the conducting channel. In general, τ_1 depends on the particular values of the electrode voltages.

The controlling and controlled charges tend to recombine with each other. Therefore, if the output current is to remain constant, a current

must flow around the input circuit to replenish Q_C and Q. Usually recombination can be represented by a constant *lifetime* τ which is the time constant of the exponential decay of excess charge. In these circumstances

$$I_1 = -\frac{Q}{\tau} = \frac{Q_C}{\tau} \tag{2.3}$$

and the emitting-electrode current of the device is

$$I_3 = I_1 + I_2 = -Q\left(\frac{1}{\tau} + \frac{1}{\tau_1}\right) = Q_C\left(\frac{1}{\tau} + \frac{1}{\tau_1}\right). \tag{2.4}$$

From Eqs. 2.2 and 2.3 the *dc current gain* of the device is

$$\frac{I_2}{I_1} = \frac{\tau}{\tau_1}. \tag{2.5}$$

It is apparent that τ_1 should be minimized and τ maximized if a high value of current gain is to be achieved. In a vacuum tube or field-effect transistor the physical separation of the two charges minimizes the possibility of recombination; both τ and the current gain are very large indeed. In comparison, recombination is significant for a bipolar transistor, and the current gain is finite.

The essential postulate of charge-control theory is that *the electrode currents at constant electrode voltages are given at all times by Eqs. 2.2 to 2.4.* No account is taken of the time immediately following a change in electrode voltages during which the charge rearranges itself and the electrode currents may not have their steady-state values. In other words, charge-control theory assumes that the electrode currents depend on the total value Q of the controlled mobile charge within a device but not on the spatial distribution of Q in the conducting channel.

Although charge-control theory asserts that the electrode currents depend only on the total charge, a knowledge of the equilibrium charge distribution is required in a calculation of τ_1 for particular devices. Consider, for example, the charge distribution shown for the plane-parallel device in Fig. 2.2. An expression for the transit time can be written in the integral form

$$\tau_1 = \frac{q}{I_2}\int_0^W A(x)n(x)\,dx, \tag{2.6}$$

where W = the distance between emitter and collector,
$n(x)$ = the concentration of mobile carriers,
$A(x)$ = cross-sectional area of the conducting channel,
q = the electronic charge, 1.60×10^{-19} coulomb.

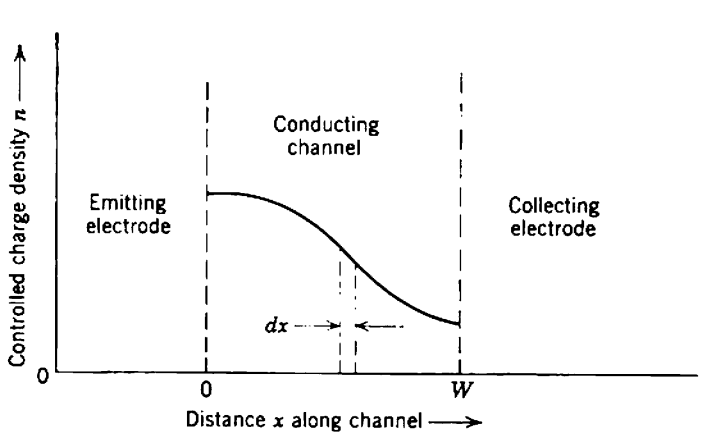


Fig. 2.2 Mobile charge density for a device having plane-parallel geometry.

It is important to realize that τ_1 relates the output current to the *total* mobile charge in transit, whereas τ relates the input current to the *excess* charge in the control region. Usually these two charges are one and the same, as evidenced by the use of Q for both throughout this simplified development. Formally, however, the total charge consists of the excess charge plus any constant equilibrium charge that may exist; as the equilibrium charge is constant, all derivatives of total and excess charge are identical. Section 4.2.1.3 contains a discussion of the transistor, a device in which the total charge differs marginally from the excess charge.

2.2.2 Dynamic Conditions

The foregoing development shows that under conditions of constant voltage it is possible to relate the input and output currents to the controlled charge Q. If the controlling charge Q_C and, correspondingly, the controlled charge Q change with time, additional charging components of current must flow into the device. The total charging current is

$$I_{\text{charging}} = -\frac{dQ}{dt} = \frac{dQ_C}{dt} = \frac{dQ_{C1}}{dt} + \frac{dQ_{C2}}{dt} \tag{2.7}$$

and is divided between the input and output circuits. If the charges change slowly with time, it is reasonable to suppose that the charge-control assumption is true, so that the total electrode currents at any instant are the sums of the charging components with the steady-state values deter-

mined by the instantaneous value of the charge. Thus, from Eqs. 2.2, 2.3, and 2.4,

$$I_1(t) = \frac{Q_C(t)}{\tau} + \frac{dQ_{C1}(t)}{dt}, \tag{2.8}$$

$$I_2(t) = \frac{Q_C(t)}{\tau_1} + \frac{dQ_{C2}(t)}{dt}, \tag{2.9}$$

$$I_3(t) = Q_C(t)\left(\frac{1}{\tau} + \frac{1}{\tau_1}\right) + \frac{dQ_C(t)}{dt}. \tag{2.10}$$

The main difficulty in making this assumption lies in the definition of just what constitutes a "slow" change. An alternative statement of the basic charge-control assumption is that the time taken for the charge to rearrange itself after a change in electrode voltages is small compared with the time taken for the charge increments to be supplied from the external circuits. Effectively, the charge may always be assumed to be in its steady-state distribution. An ideal charge-controlled device is one that obeys Eqs. 2.8–2.10 exactly. The formal approximation in representing practical devices by the charge-control model is considered in Section 2.4, and the general accuracy of the charge-control representation of vacuum tubes and transistors is confirmed in Sections 3.1 and 4.3.

If the values of Q, τ, and τ_1 can be found from the physics of a particular device. the charge-control equations can, in principle, be solved to find the relations between the time-varying electrode voltages and currents, that is, to find the dynamic response of the device. In general, Eqs. 2.8–2.10 are nonlinear: τ_1 is not constant and device physics are such that Q is not a linear function of electrode voltages. Exact solution for the dynamic response is therefore very tedious. The situation of greatest interest in this book, however, is that in which small time-varying signals are superimposed on constant or *quiescent* electrode voltages and currents. The quiescent components can be found easily, despite the nonlinearity of the equations, because all time-varying terms vanish. The time response to infinitesimal signals can then be found by assuming that the charge-control equations are linear, with constant coefficients determined by the quiescent conditions. In the language of Section 1.2.1 this response is the *small-signal dynamic response* of the device. Although formally incorrect, this idealized response is a valid approximation to the actual response of the nonlinear system to signals below a certain finite amplitude. Section 2.5 considers in general terms the small-signal response of the charge-control model.

2.3 ELEMENTARY EXAMPLES OF CHARGE-CONTROLLED DEVICES

In this section the general concepts of charge-control theory are illustrated by reference to three idealized electronic devices.

2.3.1 Vacuum Triode

Consider as a first example the simplified operation of a vacuum triode shown in Fig. 2.3. Figure 2.3*a* represents the cutoff condition in which no anode current I_A flows, despite the fact that a positive voltage is applied to the anode and the cathode is heated to emit electrons. The tube is held in its nonconducting state by the application of a negative voltage to the grid; this voltage is of sufficient magnitude to prevent the field of the anode (shown as lines of force) from reaching the electrons in the vicinity of the cathode.

If the magnitude of the negative grid voltage is reduced (Fig. 2.3*b*), the field from the anode passes through the grid wires and allows a portion of the electronic space charge to become free for conduction purposes. The positive increment in grid voltage is produced by an increase in the positive charge Q_{C1} on the grid. In consequence, the number of free electrons constituting the space charge Q in the vicinity of the cathode increases. Figure 2.3*b* represents an equilibrium state in which the individual electrons constituting the mobile charge Q move through the grid structure toward the anode and produce an anode current. A cathode current must flow, both to preserve current continuity and to maintain the mobile charge.

It is shown in Section 3.1 that a linear relationship exists between voltage and charge. Thus the total controlling charge Q_C varies with the grid and anode voltages V_G and V_A as shown in Fig. 2.4*a*. Provided that the anode current is finite, any change in grid voltage changes Q_C and consequently changes Q and I_A. In addition, if the anode voltage is changed, Q_C, Q, and, consequently, I_A will vary. The essence of triode amplification lies in the fact that changes in V_G have a greater effect on the value of Q than have changes in V_A.

2.3.2 Field-Effect Transistor

Consider as a second example the simplified operation of a field-effect (unipolar) transistor shown in Fig. 2.5; an *n*-channel device is shown, but the complementary *p*-channel device also exists. Figure 2.5*a* represents the cutoff condition in which no drain current I_D flows, despite the fact

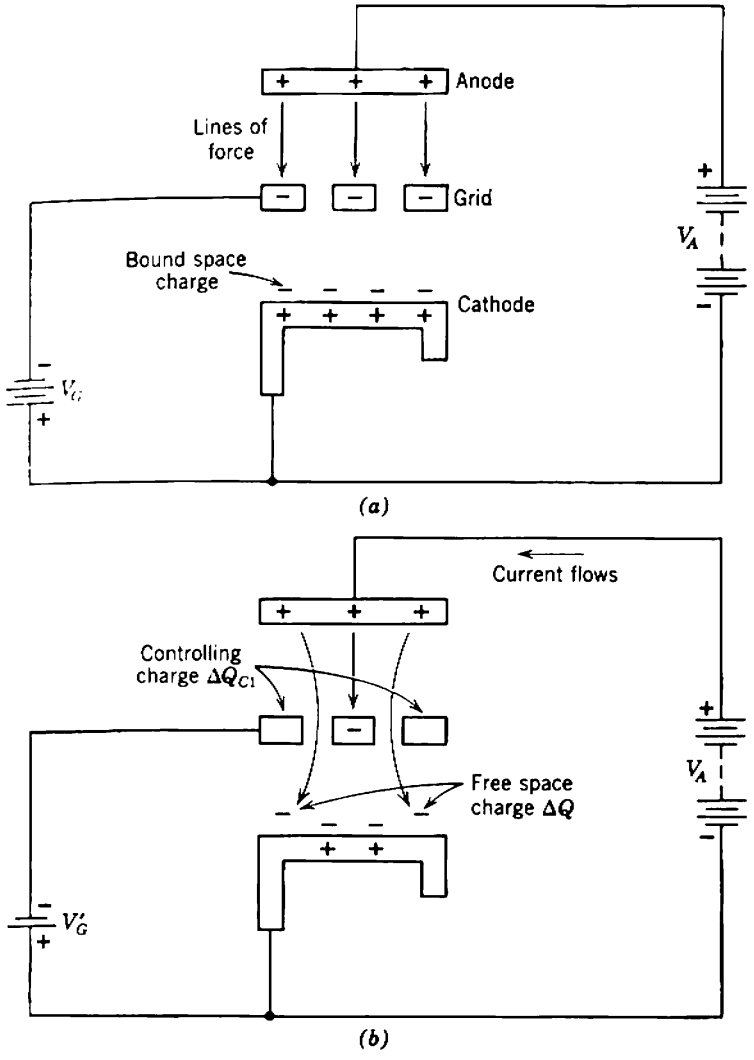


Fig. 2.3 The vacuum triode: (*a*) nonconducting; (*b*) conducting.

that a voltage V_D exists across the *n*-type conducting channel. The device is held in its nonconducting state by the application of a negative voltage to the *p*-type gate electrode; this voltage is of just sufficient magnitude to cause the depletion layer of the reverse biased *p-n* junction to extend throughout that part of the conducting channel which is enclosed by the gate. The conducting channel is effectively open-circuited in this region, for it is depleted of free carriers.

If the magnitude of negative gate voltage is reduced (Fig. 2.5*b*), the

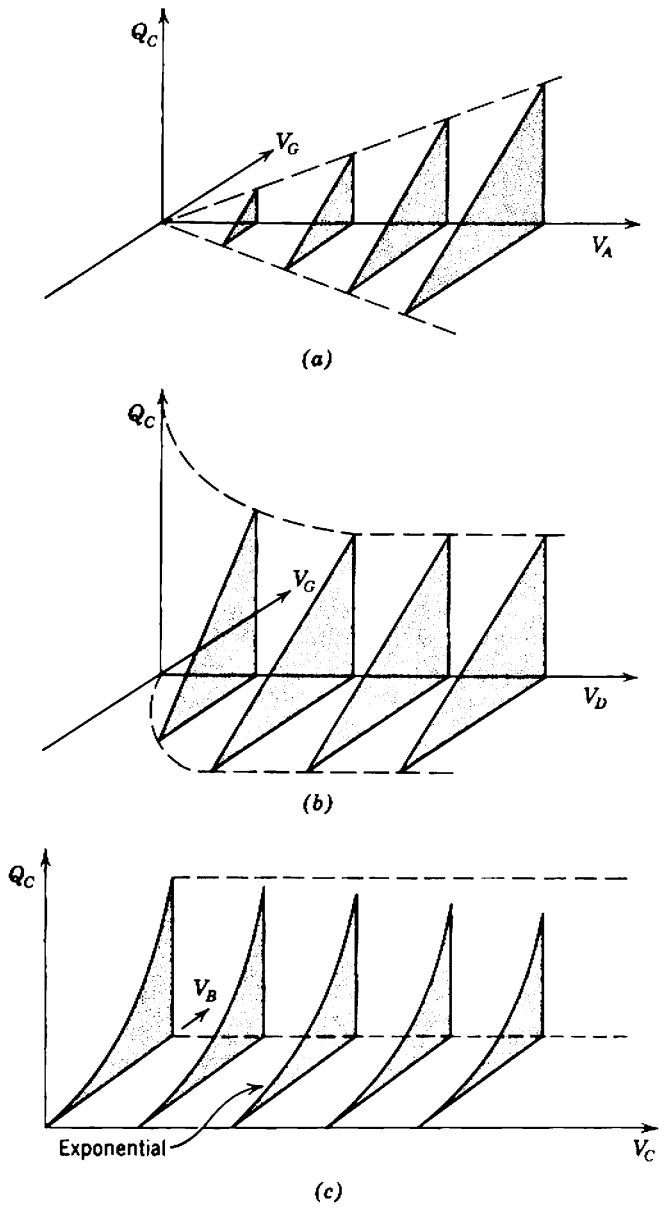


Fig. 2.4 Variation of Q_C with V_1 and V_2: (*a*) vacuum triode; (*b*) field-effect transistor (referred to in Section 2.3.2); (*c*) bipolar transistor (referred to in Section 2.3.3).

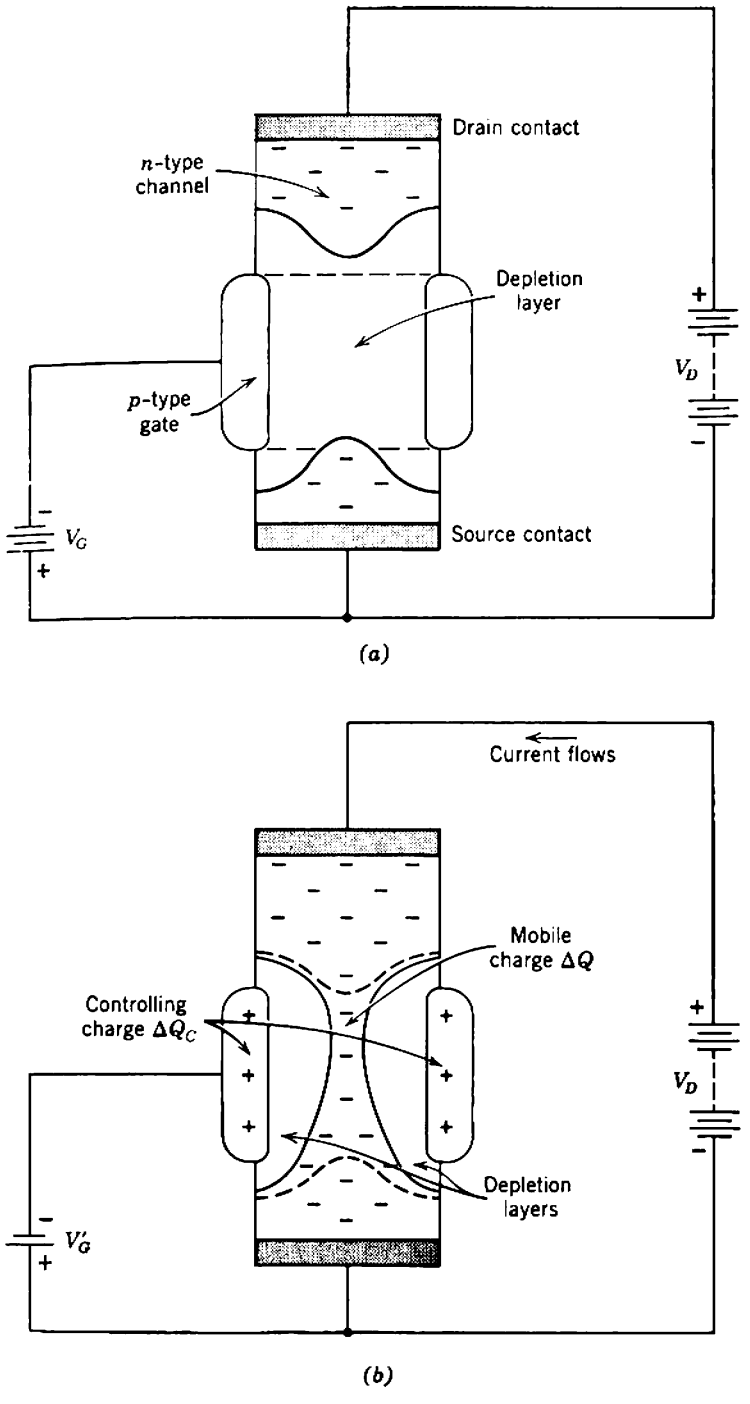


Fig. 2.5 The field-effect transistor: (*a*) nonconducting; (*b*) conducting.

depletion layer no longer extends right through the conducting channel. Electrons (majority carriers) return to the portion of the conducting channel that was formerly part of the depletion layer and are free to move to the drain. This increase in mobile charge Q is accompanied by an increase in controlling gate charge Q_C; except at low drain voltages, no component of Q_C is associated with the drain. It is shown in Sections 4.8.2 and 4.8.3 that Q_C varies linearly with gate voltage, so that the total variation of Q_C with V_G and V_D is as shown in Fig. 2.4*b*.

Aside from differences in the actual mechanism of current flow, the vacuum triode and field-effect transistor are basically similar in behavior. The two sets of charge exist in completely different sections of the device, and recombination between them is usually negligible; Q in the f.e.t. exists in the conducting channel, whereas Q_C exists on the gate electrode, and the two charges are separated by a depletion layer. The alternative name *unipolar transistor* stems from the dominant presence of a single type of carrier in the conducting channel. This contrasts with the bipolar transistor which is considered as a third example.

2.3.3 Bipolar Transistor

Consider the simplified operation of a bipolar transistor shown in Fig. 2.6; an *n-p-n* transistor is shown, but the complementary *p-n-p* device also exists. Figure 2.6*a* represents the cutoff condition in which no collector current I_C flows. The collector-to-base junction is reverse-biased by the voltage V_C. (In this simple discussion the collector saturation current, which is due to the presence of thermally produced minority carriers in the base region, is assumed to be negligible.) The transistor is held in the nonconducting state because there is no forward bias voltage across the emitter-to-base junction. Under these conditions there is no injection of minority carriers into the base region, and consequently, in the absence of both drift fields and concentration gradients, no emitter-to-collector current can flow.

Suppose that the base electrode is taken positive with respect to the emitter by introducing extra majority carriers (holes) into the base region via the base lead. For charge neutrality extra minority carriers (electrons) are injected into the base region from the emitter junction, as shown in Fig. 2.6*b*. For the simple case of a uniform-base transistor the majority and minority carrier concentrations in the base region vary linearly with distance. The majority carriers are held in equilibrium by an electric field, but the minority carriers are free to diffuse from emitter to collector and thereby produce an output current I_C. The injected electrons constitute the mobile charge Q, and the excess holes constitute the controlling

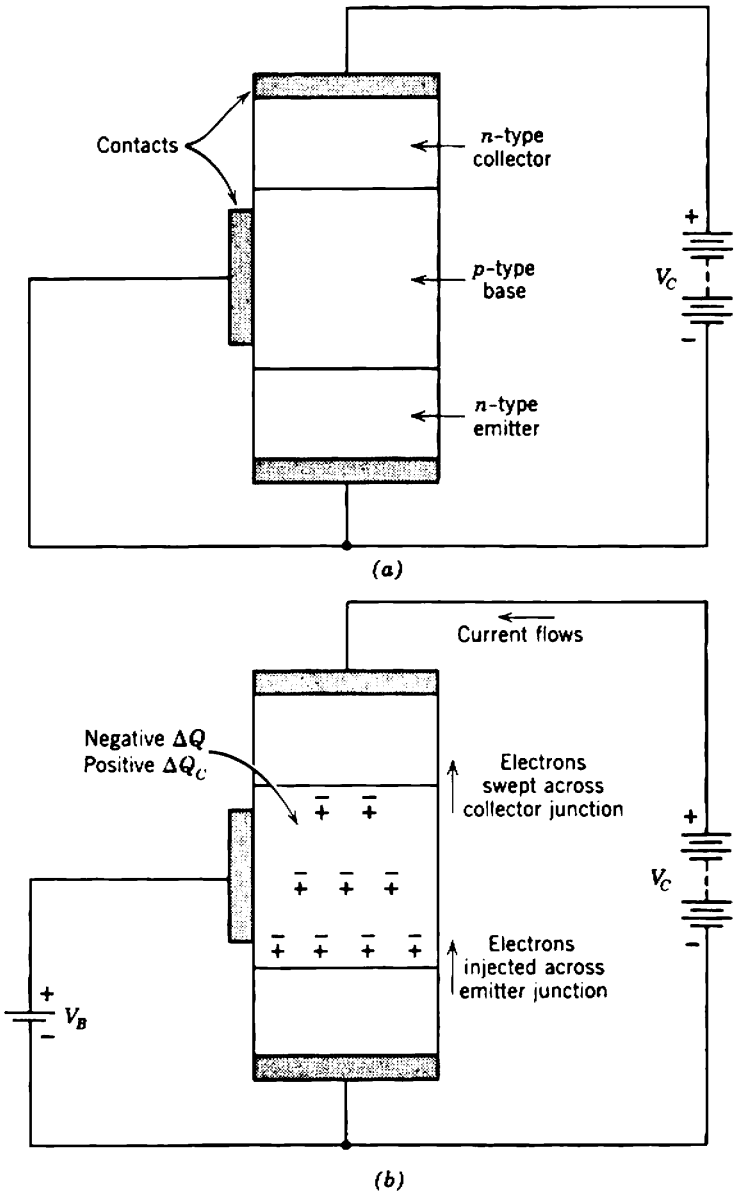


Fig. 2.6 The bipolar transistor: (*a*) nonconducting; (*b*) conducting.

charge Q_C. The two charges coexist in the base region, and consequently a relatively high probability of recombination exists. Therefore equal and opposite recombination components of emitter and base currents must flow if Q is to remain constant, and a change of input voltage will change both the output and input currents. Notice that no component of Q_C is

associated with the collector. It is shown in Section 4.2 that Q_C varies exponentially with base voltage but that the collector voltage influences the charge associated with the base; the total variation of Q_C with V_B and V_C is shown in Fig. 2.4*c*.

2.4 THE FORMAL APPROACH TO CHARGE-CONTROL THEORY*

It is instructive to consider in general terms a set of partial differential equations that accurately specifies the time and distance behavior of mobile charges in a control region. These equations are significant in two respects: they highlight the general simplifying assumptions on which the charge-control theory for time-varying signals is based, and they are useful in comparing the results of charge-control and more accurate theories for specific devices. A consistent set of equations relates the following quantities, all of which may be functions of time and position:

current density **J**,
mobile carrier concentration p (holes) or n (electrons),
mobile carrier velocity **v**,
electric field $\mathcal{E}$,
electric potential V.

The five equations relating these variables can then be written as follows:

1. Maxwell's equation for the total electron current density becomes

$$\mathbf{J} = -qn\mathbf{v} + \varepsilon \frac{\partial \mathcal{E}}{\partial t}, \tag{2.11}$$

where the first and second terms represent the convection and displacement components, respectively, q is the electronic charge, and ε is the permittivity.

2. Poisson's equation can be used to relate the field to the total charge density:

$$\nabla \cdot \mathcal{E} = \frac{q}{\varepsilon}(p - n + N_D - N_A), \tag{2.12}$$

where N_D and N_A are the donor and acceptor densities, respectively.

3. The continuity equation for electrons becomes

$$\frac{\partial n}{\partial t} = (g - r) - \nabla \cdot (n\mathbf{v}), \tag{2.13}$$

* Section 2.4 may be omitted on a first reading.

where g and r are the rates of carrier generation and recombination, respectively.

4. The equation of motion for mobile carriers depends on the physics of the conduction process. In general, the velocity depends on both field and mobile carrier concentration; for electrons, the velocity can be written in functional form as

$$\mathbf{v} = \mathbf{v}(V, n) \tag{2.14}$$

and the functional relation can be evaluated in particular cases.

5. The potential and field are related by

$$\mathcal{E} = -\nabla V. \tag{2.15}$$

Equations similar to (2.11), (2.13), and (2.14) can be written for the behavior of holes of concentration p in semiconductors.

Ideally, this set of partial differential equations should be solved for the given boundary conditions to determine the behavior of carriers in the control region of any device. The boundary conditions involve the time-varying electrode voltages, so the solutions are the general volt/ampere characteristics of the device, that is, its dynamic response. Although some simplifying approximations (discussed in Section 2.4.1 following) are valid for most device types, the general solution of Eqs. 2.11–2.15 remains tedious.

If it is assumed that all signal waveforms change sufficiently slowly for Eqs. 2.8–2.10 to be valid approximations, the general equations (2.11–2.14) for carrier behavior can be simplified further. Determination of Q and τ_1 as functions of electrode voltages becomes relatively straightforward, and these parameters can be substituted into the charge-control equations to predict the small-signal dynamic response of practical devices. There is, of course, the formal error that Eqs. 2.8–2.10 are not strictly accurate for any practical device, but the general accuracy of the charge-control predictions is confirmed in Chapters 3 and 4.

2.4.1 Accurate Predictions

For high-vacuum devices it is a valid approximation to neglect generation and recombination of electrons in the control region so that

$$(g - r) = 0 \tag{2.16}$$

may be substituted into Eq. 2.13. In addition, it is well known that the acceleration of free electrons satisfies

$$\frac{d\mathbf{v}}{dt} = -\frac{q}{m}\mathcal{E}. \tag{2.17}$$

A simplified, but nevertheless accurate, set of equations corresponding to Eqs. 2.11–2.15 for a vacuum device with plane-parallel geometry is

$$J = -qnv + \varepsilon \frac{\partial \mathscr{E}}{\partial t}, \tag{2.18}$$

$$\frac{\partial \mathscr{E}}{dx} = -\frac{q}{\varepsilon} n, \tag{2.19}$$

$$\frac{\partial n}{\partial t} = -\frac{\partial (nv)}{\partial x}, \tag{2.20}$$

$$\frac{dv}{dt} = -\frac{q}{m} \mathscr{E}, \tag{2.21}$$

$$\mathscr{E} = -\frac{\partial V}{\partial x}. \tag{2.22}$$

Combination of Eqs. 2.18 and 2.19 leads to

$$J = \varepsilon \left(\frac{\partial \mathscr{E}}{\partial x} \frac{dx}{dt} + \frac{\partial \mathscr{E}}{\partial t} \right);$$

that is,

$$J = \varepsilon \frac{d\mathscr{E}}{dt}. \tag{2.23}$$

This last equation relates the current density to the total time derivative of the electric field. Then, from Eq. 2.21,

$$J = -\varepsilon \frac{m}{q} \frac{d^3x}{dt^3}, \tag{2.24}$$

and, if any of the time derivatives of x are known (velocity, acceleration, or rate of change of acceleration), the current density can be determined. Consequently Eq. 2.24 provides the basis for determining the dynamic properties of all plane-parallel high-vacuum devices. The results yielded by this accurate approach are compared with those predicted from charge-control theory in Section 3.1.

Compared with vacuum devices, semiconductor devices have high densities of mobile charge and low electric fields. The convection component of current density therefore greatly exceeds the displacement component, and a valid approximation in Eq. 2.11 is

$$\frac{\partial \mathscr{E}}{\partial t} = 0. \tag{2.25}$$

A second valid approximation for semiconductors is that the net rate of carrier recombination is proportional to the excess carrier density, so that

$$(g - r) = -\frac{n - n_0}{\tau}, \tag{2.26}$$

where n_0 is the equilibrium concentration and τ is the lifetime. Finally, the velocity of an electron in a semiconductor with modility μ and diffusion constant D is

$$\mathbf{v} = -\mu\mathcal{E} - \frac{D}{n}\nabla n. \tag{2.27}$$

Hence the accurate set of equations for semiconductor devices with plane-parallel geometry is

$$J = -qnv, \tag{2.28}$$

$$\frac{\partial \mathcal{E}}{\partial x} = 0, \tag{2.29}$$

$$\frac{\partial n}{\partial t} = -\frac{n - n_0}{\tau} - \frac{\partial (nv)}{\partial x}, \tag{2.30}$$

$$v = -\mu\mathcal{E} - \frac{D}{n}\frac{\partial n}{\partial x}, \tag{2.31}$$

$$\mathcal{E} = -\frac{\partial V}{\partial x}. \tag{2.22}$$

From Eqs. 2.29–2.31 it follows that

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial x^2} + \mu\mathcal{E}\frac{\partial n}{\partial x} - \frac{n - n_0}{\tau}. \tag{2.32}$$

Equation 2.32 can be used, at least in principle, for determining the dynamic properties of plane-parallel semiconductor devices. The results yielded by this accurate approach are compared with those predicted from charge-control theory in Section 4.3.2.

2.4.2 Charge-Control Predictions

The basic assumption of charge-control theory is that the time taken for the charge to rearrange itself after a change in electrode voltages is negligible compared with the time taken to change these voltages. Effectively, it is assumed that there is no variation in the mobile charge density under conditions of constant electrode voltages, no matter how recently

these voltages may have been changed. Mathematically, the charge-control assumption is that

$$\frac{\partial \mathscr{E}}{\partial t} = \frac{\partial n}{\partial t} = 0 \tag{2.33}$$

in Eqs. 2.11 and 2.13. Only if this assumption is true can the output current of a device be written as

$$I_2 = -\frac{Q}{\tau_1} + \frac{dQ_{C2}}{dt} = \frac{Q_C}{\tau_1} + \frac{dQ_{C2}}{dt}$$

for all time. If there is any redistribution of charge (although no change in the total charge) after a change in electrode voltages, the output current following the change becomes

$$I_2 = \frac{Q_C}{\tau_1} + \frac{dQ_{C2}}{dt} - I_\varepsilon(t), \tag{2.34}$$

where the error current I_ε decays to zero with increasing time. Notice that the input current I_1 is not ordinarily dependent on the charge-control assumption; recombination in practical devices is proportional to the excess charge but largely independent of the charge distribution so that

$$I_1 = \frac{Q_C}{\tau} + \frac{dQ_{C1}}{dt} \tag{2.35}$$

applies for all time. This distinction between input and output currents is discussed further in Sections 3.1.1.3 and 3.1.2.3.

Incorporating the charge-control assumption (Eq. 2.33) into the results of Section 2.4.1, we find that the simplified equations for predicting the charge-control parameters Q and τ_1 become

vacuum devices

$$J = -qnv, \tag{2.36}$$

$$\frac{\partial \mathscr{E}}{\partial x} = -\frac{q}{\varepsilon} n, \tag{2.19}$$

$$\frac{\partial (nv)}{\partial x} = 0, \tag{2.37}$$

$$\frac{\partial v}{\partial t} = -\frac{q}{m} \mathscr{E}, \tag{2.21}$$

$$\mathscr{E} = -\frac{\partial V}{\partial x}; \tag{2.22}$$

semiconductor devices

$$J = -qnv, \tag{2.28}$$

$$\frac{\partial \mathscr{E}}{\partial x} = 0, \tag{2.29}$$

$$\frac{\partial(nv)}{\partial x} + \frac{n - n_0}{\tau} = 0, \tag{2.38}$$

$$v = -\mu\mathscr{E} - \frac{D}{n}\frac{\partial n}{\partial x}, \tag{2.31}$$

$$\mathscr{E} = -\frac{\partial V}{\partial x}. \tag{2.22}$$

Notice that Eq. 2.37 for vacuum devices is merely the derivative of Eq. 2.36 and as such contains no new information.

2.5 SMALL-SIGNAL REPRESENTATION OF THE CHARGE-CONTROL MODEL

The relations between time-varying electrode currents and the total controlling charge are given by the basic equations (2.8), (2.9), and (2.10). The electrode voltages can be introduced into calculations (and charge terms eliminated) if the relation between voltages and controlling charge is known. In general, Q_C of a practical device varies nonlinearly with voltage, and the set of differential equations linking the electrode voltages and currents is nonlinear. However, in the special case of changes in voltage and current that are small compared with their steady (quiescent) values the equations may be assumed to be linear and a general small-signal model can be developed. This linear model must be used with caution, for it is valid formally only for signals of infinitesimal amplitude. For signals of finite amplitude the relationship between input and output is nonlinear and the output signal is distorted. The maximum signal amplitude for which a linear model is valid depends on the fidelity required; a detailed discussion is given in Chapter 9.

The controlling charge in a general three-electrode model is a function of both the input and output voltages, V_1 and V_2:

$$Q_C = Q_C(V_1, V_2).$$

Increments ΔV_1 and ΔV_2 produce a total increment ΔQ_C in the controlling charge, and for small increments

$$\Delta Q_C = c_1 \Delta V_1 + c_2 \Delta V_2,$$

where

$$c_1 = \left(\frac{\partial Q_C}{\partial V_1}\right)_{V_2}, \tag{2.39}$$

$$c_2 = \left(\frac{\partial Q_C}{\partial V_2}\right)_{V_1}. \tag{2.40}$$

The constants c_1 and c_2 have the dimensions of capacitance. Similarly, changes in Q_{C1} and Q_{C2} may be written as

$$\Delta Q_{C1} = c_{11}\,\Delta V_1 + c_{12}\,\Delta V_2,$$

$$\Delta Q_{C2} = c_{21}\,\Delta V_1 + c_{22}\,\Delta V_2,$$

where

$$c_{11} = \left(\frac{\partial Q_{C1}}{\partial V_1}\right)_{V_2}, \tag{2.41}$$

$$c_{12} = \left(\frac{\partial Q_{C1}}{\partial V_2}\right)_{V_1}, \tag{2.42}$$

$$c_{21} = \left(\frac{\partial Q_{C2}}{\partial V_1}\right)_{V_2}, \tag{2.43}$$

$$c_{22} = \left(\frac{\partial Q_{C2}}{\partial V_2}\right)_{V_1}. \tag{2.44}$$

Notice that

$$c_1 = c_{11} + c_{21},$$

$$c_2 = c_{12} + c_{22}.$$

When charge-control theory is applied to four- or five-electrode devices, changes in Q_C depend on changes in three or four electrode voltages. Although not attempted in this section, the theory can be extended readily to such cases; a discussion of the pentode is given in Section 3.1.3.

The charge-control equations (2.8 and 2.9) may be combined with (2.39) to (2.44) to give the transfer and driving-point admittances of the linear two-port representation of the charge-control model. The input-circuit equation is derived here completely; the reader may verify the other three.

SHORT-CIRCUIT INPUT ADMITTANCE. Provided that the lifetime τ is assumed to be constant, differentiation of Eq. 2.8 gives

$$\Delta I_1 = \frac{1}{\tau}\left(\frac{\partial Q_C}{\partial V_1}\right)_{V_2} \Delta V_1 + \frac{d}{dt}\left[\left(\frac{\partial Q_{C1}}{\partial V_1}\right)_{V_2} \Delta V_1\right],$$

and by substituting from Eqs. 2.39 and 2.41 we obtain

$$\Delta I_1 = \frac{c_1}{\tau}\,\Delta V_1 + c_{11}\frac{d}{dt}(\Delta V_1).$$

Therefore

$$\left(\frac{\partial I_1}{\partial V_1}\right)_{V_2} = \frac{c_1}{\tau} + c_{11}\frac{d}{dt}, \tag{2.45}$$

where d/dt operates on V_1, the denominator of the left-hand side.

FORWARD TRANSFER ADMITTANCE. For greater generality τ_1 is allowed to vary with V_1 and V_2, and

$$\left(\frac{\partial I_2}{\partial V_1}\right)_{V_2} = \frac{c_1}{\tau_1}(1 + u_1) + c_{21}\frac{d}{dt}, \tag{2.46}$$

where

$$u_1 = -\frac{I_2}{c_1}\left(\frac{\partial \tau_1}{\partial V_1}\right)_{V_2}. \tag{2.47}$$

The parameter u_1 can be determined for specific devices; it is a constant, independent of V_1 and I_2.

SHORT-CIRCUIT OUTPUT ADMITTANCE.

$$\left(\frac{\partial I_2}{\partial V_2}\right)_{V_1} = \frac{c_2}{\tau_1}(1 + u_2) + c_{22}\frac{d}{dt}, \tag{2.48}$$

where

$$u_2 = -\frac{I_2}{c_2}\left(\frac{\partial \tau_1}{\partial V_2}\right)_{V_1}. \tag{2.49}$$

REVERSE TRANSFER ADMITTANCE.

$$\left(\frac{\partial I_1}{\partial V_2}\right)_{V_1} = \frac{c_2}{\tau} + c_{12}\frac{d}{dt}. \tag{2.50}$$

Because the model is linear, superposition can be applied and the total increments in I_1 and I_2 are given by

$$\Delta I_1 = \left(\frac{c_1}{\tau} + c_{11}\frac{d}{dt}\right)\Delta V_1 + \left(\frac{c_2}{\tau} + c_{12}\frac{d}{dt}\right)\Delta V_2, \tag{2.51a}$$

$$\Delta I_2 = \left[\frac{c_1}{\tau_1}(1 + u_1) + c_{21}\frac{d}{dt}\right]\Delta V_1 + \left[\frac{c_2}{\tau_1}(1 + u_2) + c_{22}\frac{d}{dt}\right]\Delta V_2. \tag{2.52a}$$

These equations may be written more compactly by using lowercase letters for incremental (signal) voltages and currents and transforming to the complex frequency plane:

$$i_1(s) = \left(\frac{c_1}{\tau} + sc_{11}\right)v_1(s) + \left(\frac{c_2}{\tau} + sc_{12}\right)v_2(s), \tag{2.51b}$$

$$i_2(s) = \left[\frac{c_1}{\tau_1}(1 + u_1) + sc_{21}\right]v_1(s) + \left[\frac{c_2}{\tau_1}(1 + u_2) + sc_{22}\right]v_2(s). \tag{2.52b}$$

These equations correspond to the small-signal equivalent circuit of Fig. 2.7*a*.

Figure 2.7*a* is the most general equivalent circuit for the charge-control model. It consists of two self-conductance elements g_1 and g_2, two capacitive self-susceptance elements c_{11} and c_{22}, two mutual-conductance elements $g_m v_1$ and $g_f v_2$, and two capacitive mutual-susceptance elements $sc_{21}v_1$ and $sc_{12}v_2$; the mutual admittances are shown as controlled generators. Notice the symmetry of this circuit; if τ is allowed to vary with V_1 and V_2, the parameters involving τ take on the more complex forms of those involving τ_1 and the symmetry becomes more perfect. It is unrealistic, however, to emphasize the symmetry of this equivalent circuit, for in order to achieve useful gain the transmission from input to output should far exceed that from output to input. This implies that gross asymmetry occurs in the parameter magnitudes. In the most ideal circumstances all the controlling charge is associated with the control electrode and is a function of V_1 alone. Then

$$c_{11} = c_1 = \left(\frac{\partial Q_C}{\partial V_1}\right)_{V_2},$$

$$c_{12} = c_{21} = c_{22} = c_2 = 0,$$

and the equivalent circuit simplifies to the nonsymmetrical form of Fig. 2.7*b*. The current generator that remains is essential to amplifying devices and is called the *mutual conductance* g_m.

No practical device is quite so simple as Fig. 2.7*b*, but a field-effect transistor operated at moderate drain voltage approaches the ideal. This can be verified by noting from Fig. 2.4*b* that c_2 $[= (\partial Q_C/\partial V_D)_{V_G}]$ is zero over a substantial range of drain voltage. Furthermore, τ is very large for this device; it can often be assumed that g_1 $(= c_1/\tau)$ vanishes and the equivalent circuit reduces still further to Fig. 2.7*c*.

For a vacuum triode it is usually permissible to assume that τ is infinite. In addition, c_{12} and c_{21} are zero, as there is no mechanism involving the mobile charge by which a change in either the grid or anode voltage can change the controlling charge on the other electrode. Thus

$$c_{11} = c_1 = \left(\frac{\partial Q_C}{\partial V_G}\right)_{V_A},$$

$$c_{22} = c_2 = \left(\frac{\partial Q_C}{\partial V_A}\right)_{V_G},$$

and the equivalent circuit reduces to Fig. 2.7*d*.

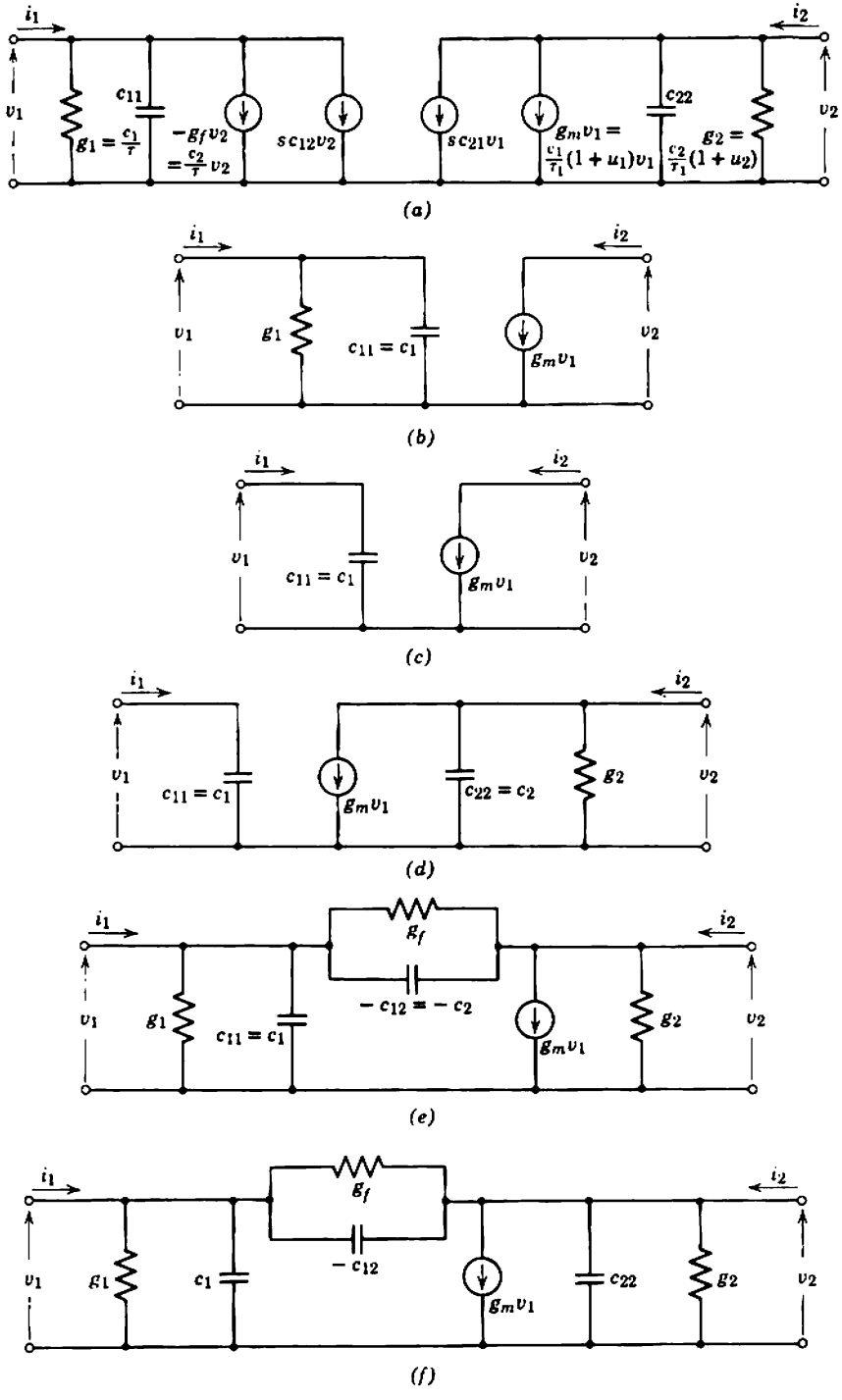


Fig. 2.7 Small-signal equivalent circuit for a charge-controlled device: (*a*) general model; (*b*) ideal model; (*c*) field-effect transistor; (*d*) vacuum triode; (*e*) bipolar transistor; (*f*) simplified general model.

For a bipolar transistor all the controlled charge is inside the base region. Therefore

$$c_{11} = c_1 = \left(\frac{\partial Q_C}{\partial V_B}\right)_{V_C},$$

$$c_{12} = c_2 = \left(\frac{\partial Q_C}{\partial V_C}\right)_{V_B},$$

$$c_{21} = c_{22} = 0.$$

As far as the input circuit is concerned, the generators $g_f v_2$ and $sc_{12}v_2$ can be approximated by the bridging or *feedback elements* shown in Fig. 2.7*e*. The error introduced by this transformation is negligible, provided that c_{11} is much greater than c_{12}. Reference to Fig. 2.4*c* shows that c_2 has a small negative value, but it is shown in Section 4.2.3 that

$$u_2 \approx -2.$$

Therefore all parameters in the equivalent circuit are, in fact, positive.

The elements g_m, c_1, and g_1 are the really basic constituents of the charge-control model. The other parameters account for the second-order dependence of Q_C on V_2, and although these elements may not vanish for particular devices, invariably they are small if the device is useful as an amplifier. In every case, therefore, the general equivalent circuit of Fig. 2.7*a* can be manipulated into the single generator form shown in Fig. 2.7*f*. Figure 2.7*f* is the basic equivalent circuit used throughout the remainder of this book.

Practical amplifying devices are not represented exactly by this basic or *intrinsic equivalent circuit* (or the appropriate simplified form). Parasitic or *extrinsic elements* must be added to account for direct interelectrode capacitances, *p-n* junction transition capacitances, interelectrode leakage conductances, and resistances in series with electrodes. In addition, practical devices are not truly charge-controlled; the electrode currents depend on the charge distribution as well as the total charge. More complicated models with more complicated equivalent circuits can be developed to account for these charge distribution effects but are not justified from an engineering point of view. The charge-control model, Fig. 2.7*f* (with the addition of the parasitic elements noted above), satisfies all the conditions set out in Section 2.1.

2.5.1 Theoretical Limits to Performance

It is instructive to consider three theoretical limits to the performance of the general charge-control model as an amplifier—the short-circuit current gain, open-circuit voltage gain, and gain-bandwidth product. All three

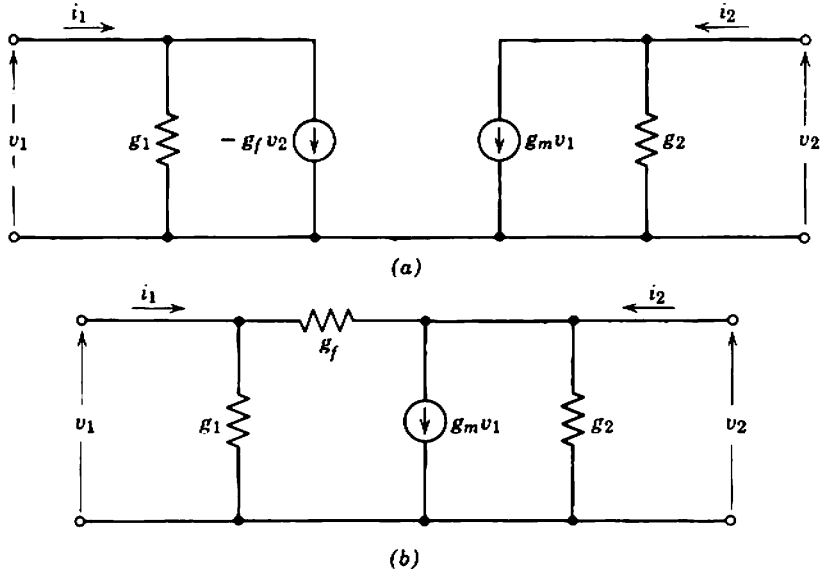


Fig. 2.8 Mid-band equivalent circuit for the charge-control model.

quantities are significant in practical amplifier design, for they represent upper bounds to the gain that can be achieved with one device.

At frequencies that are not too high, the capacitive susceptances in the general equivalent circuit of Fig. 2.7*a* are very small compared with the conductances. Below a certain maximum frequency all capacitances may be neglected and the equivalent circuit reduces to Fig. 2.8*a*. In the alternative version derived from Fig. 2.7*f* the generator g_f is replaced by a conductance as in Fig. 2.8*b*. For reasons given in Section 7.3 the frequency range over which Fig. 2.8 is an accurate representation of the charge-control model is called the *mid-band range*; Fig. 2.8 is the *mid-band equivalent circuit*. The elements g_m, g_1, g_2, and g_f have simple physical interpretations and can be measured easily for practical devices. From Eqs. 2.51 and 2.52

$$g_m = \left(\frac{\partial I_2}{\partial V_1}\right)_{V_2} = \frac{c_1}{\tau_1}(1 + u_1), \tag{2.53}$$

$$g_1 = \left(\frac{\partial I_1}{\partial V_1}\right)_{V_2} = \frac{c_1}{\tau}, \tag{2.54}$$

$$g_2 = \left(\frac{\partial I_2}{\partial V_2}\right)_{V_1} = \frac{c_2}{\tau_1}(1 + u_2), \tag{2.55}$$

$$g_f = -\left(\frac{\partial I_1}{\partial V_2}\right)_{V_1} = -\frac{c_2}{\tau}. \tag{2.56}$$

If the output terminals of the charge-control model are short-circuited for signals (i.e., if V_2 is held constant), the *small-signal short-circuit current gain* β is

$$\beta = \left(\frac{\partial I_2}{\partial I_1}\right)_{V_2} \equiv \left(\frac{\partial I_2}{\partial V_1}\right)_{V_2}\left(\frac{\partial V_1}{\partial I_1}\right)_{V_2} = \frac{g_m}{g_1} = \frac{\tau}{\tau_1}(1 + u_1); \qquad (2.57)$$

β is the largest current gain of which the model is capable. In general, this small-signal current gain differs from the dc current gain of the model, given by Eq. 2.5. An alternative name for β is the *current amplification factor* of the model.

If the output terminals are open-circuited for signals (i.e., if I_2 is held constant), the *small-signal open-circuit voltage gain* μ of the model is

$$\mu = -\left(\frac{\partial V_2}{\partial V_1}\right)_{I_2} \equiv +\left(\frac{\partial V_2}{\partial I_2}\right)_{V_1}\left(\frac{\partial I_2}{\partial V_1}\right)_{V_2} = \frac{g_m}{g_2} = \frac{c_1}{c_2}\left(\frac{1 + u_1}{1 + u_2}\right); \qquad (2.58)$$

μ is the largest voltage gain of which the model is capable. Physically, μ is the ratio of the changes in V_1 and V_2 which together maintain I_2 constant. These changes must be in opposite directions, for increments in the same direction produce aiding components of ΔQ_C and hence ΔI_2; a negative sign is therefore included in the definition to give μ a positive value. A common alternative name for μ is the *voltage amplification factor*. Notice that if the voltage gain is to be greater than unity c_1 must be greater than c_2; that is, as already discussed, changes in V_1 must have a greater effect on Q_C than changes in V_2. For an ideal charge-controlled device in which V_2 has no effect on Q_C the elements c_2, g_2, and g_f all vanish as in Fig. 2.7*b*. However, c_1 must be finite if g_m is to be finite; a charge-controlled device must have a finite input capacitance.

If the output terminals of the model are short-circuited for signals, a number of elements in the complete equivalent circuit may be neglected. Subject only to the assumption that c_{21} is small, so that

$$c_{11} \approx c_1,$$

Fig. 2.7*a* reduces to Fig. 2.9*a*. The short-circuit current gain is given as a function of frequency by

$$\frac{i_2(j\omega)}{i_1(j\omega)} = \frac{g_m}{g_1}\left[\frac{1}{1 + j\omega(c_1/g_1)}\right] = (1 + u_1)\frac{\tau}{\tau_1}\left[\frac{1}{1 + j\omega\tau}\right]. \qquad (2.59)$$

This function is plotted on logarithmic scales in Fig. 2.9*b*. The product of the low-frequency gain and the 3-dB bandwidth is a useful measure of the model's capability in wide-band amplifier circuits and is defined as the *gain-bandwidth product* $\mathscr{GB}$:

$$\mathscr{GB} = \frac{g_m}{c_1} = \frac{1 + u_1}{\tau_1}. \qquad (2.60)$$

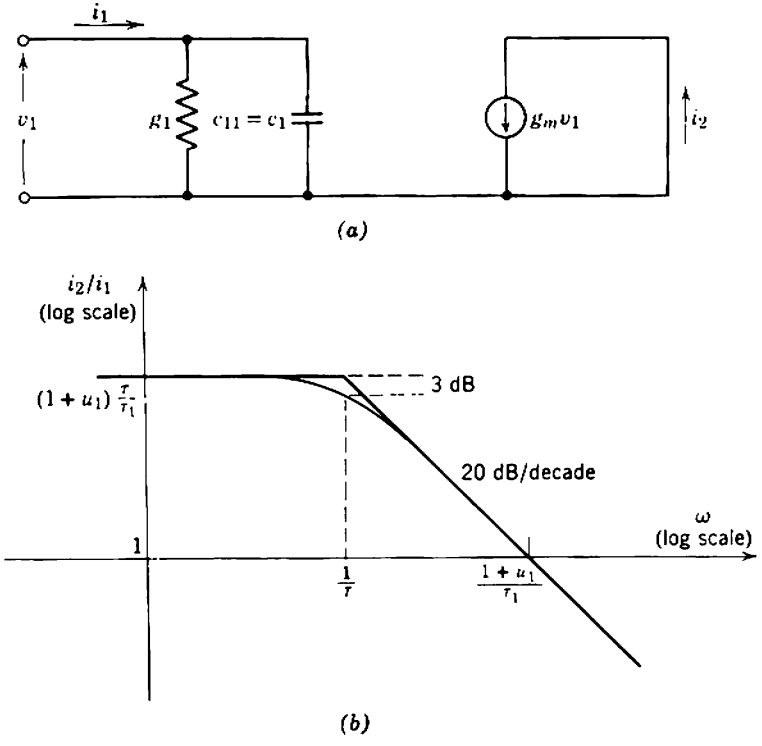


Fig. 2.9 Gain-bandwidth product.

It transpires that the product of the current gain and 3-dB bandwidth of the model in any circuit configuration whatsoever can never exceed $\mathscr{GB}$ by more than a factor of two and in many instances falls below $\mathscr{GB}$. Gain-bandwidth product is therefore one of the most significant parameters of an active device and is discussed extensively throughout this book, particularly in Chapter 13.

2.6 INTRODUCTION TO NOISE IN AMPLIFYING DEVICES

The electrode currents of amplifying devices arise from the motion of individual charge carriers. The flow of current is therefore corpuscular rather than fluid in nature, and random fluctuations due to changes in the number of carriers involved are superimposed on the signal currents. These corrupting components of current are known as *noise*.

The purpose of this section is to discuss only two aspects of the theoretically unavoidable random noise—the properties of the physical sources of noise and the circuit representation of noise in an amplifying device—both of which provide the basis for the specific investigations of later

chapters. It must be emphasized that the treatment given here is incomplete and not always formally correct. The results quoted, however, are of sufficient accuracy for most amplifier applications. The reader is directed to the specialized books available* for a complete and formally correct development of noise theory.

2.6.1 Types of Noise

Noise can be grouped into three main types—thermal, shot, and flicker.

Thermal noise in conductors is produced by the random motion of current carriers and is present regardless of the presence of signal currents. Nyquist† has shown that the mean-square noise voltage appearing across a resistance R is

$$d(v_{NT}^2) = 4kTR\,df, \tag{2.61}$$

where k = Boltzmann's constant (1.37×10^{-23} joule/°K),
T = absolute temperature,
df = incremental frequency range over which noise voltage is summed (the unit is hertz or cycles per second, not radians per second).

Thermal noise has a uniform power spectrum; that is, the noise power in any band of frequencies of given width is constant, independent of the actual frequencies. Noise having this characteristic is described as *white*.

In calculations involving both signals and noise a practical resistor can be represented as a noiseless resistor in series with a noise voltage generator $d(v_{NT})$ as in Fig. 2.10*a*. Alternatively, the corresponding shunt combination of a noiseless resistor and a noise current generator $d(i_{NT})$ can be used (Fig. 2.10*b*). Obviously,

$$d(i_{NT}) = \frac{d(v_{NT})}{R}.$$

---

* The following books may be found useful: W. R. BENNETT, *Electrical Noise* (McGraw-Hill, New York, 1960); F. N. H. ROBINSON, *Noise in Electrical Circuits* (University Press, Oxford, 1962); A. VAN DER ZIEL, *Noise* (Prentice-Hall, Englewood Cliffs, N.J., 1954); L. D. SMULLIN and H. A. HAUS (eds.), *Noise in Electron Devices* (Technology Press and Wiley, New York, 1959); C. A. HOGARTH (ed.), *Noise in Electronic Devices* (published on behalf of the Physical Society by Chapman and Hall, London, 1961).

† H. NYQUIST, "Thermal agitation of electric charge in conductors," *Phys. Rev.*, **32**, 110, July 1928. See also W. SCHOTTKY, "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern," *Ann. Physik*, **57**, 541, December 1918; J. B. JOHNSON, "Thermal agitation of electricity in conductors," *Phys. Rev.*, **32**, 97, July 1928.

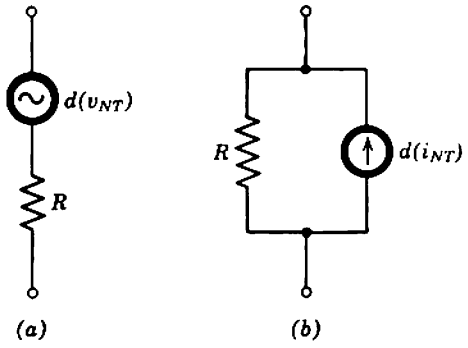


Fig. 2.10 Alternative representations of thermal noise.

Therefore

$$d(i_{NT}^2) = \frac{4kT}{R}\,df. \tag{2.62}$$

Shot noise in amplifying devices originates during carrier migration through a control region. Because the region is virtually free of carriers when the electrode currents are zero, the noise must depend on the current level. Schottky* has shown that the mean-square value of the noise current due to a carrier stream is

$$d(i_{NS}^2) = 2qI\,df, \tag{2.63}$$

where q = the electronic charge (1.60×10^{-19} coulomb),
I = the dc value of the carrier stream.

The noise contribution of any direct current I flowing between two terminals of a device may be represented by a noise current generator $d(i_{NS})$ linking the two terminals. Like thermal noise, shot noise is white.

Experimental results agree with predictions based on Eq. 2.63 for carrier streams in a semiconductor and emission-limited electron streams in a vacuum, but the noise current for a space-charge-limited electron stream in a vacuum is less than predicted. To accommodate this and possibly other cases it is customary to introduce a correcting factor Γ into the simple expression so that

$$d(i_{NS}^2) = 2qI\Gamma^2\,df. \tag{2.64}$$

Usually Γ (the *space-charge reduction factor*) lies in the range 0.1 to 1.0. The value of Γ depends on the emission velocity of the carriers, the

* W. SCHOTTKY, *loc. cit.*

electric field within the control region, and any random diversion of carriers to an adjacent stream. The last-mentioned effect results in an increase in I' in multielectrode tubes; the extra component is known as *partition noise* and is quite important in pentodes.

Flicker noise is associated with imperfections in manufacturing technology and has become less significant as production techniques have improved. Flicker noise in a vacuum tube arises from low-frequency fluctuations in the cathode emission, whereas flicker noise in a semiconductor device appears to be caused by imperfections of the semiconducting crystal. Unlike thermal and shot noise, which have uniform power spectra, flicker noise has a power spectrum that varies approximately inversely with frequency. Consequently, this type of noise is often known as $1/f$ noise; another name in use is *excess noise*. Often the flicker noise current between two electrodes can be related to the direct current I flowing between them:

$$d(i_{NF}^2) = K\frac{I^c}{f}\,df, \tag{2.65}$$

where K and c are empirical constants.

2.6.2 Representation of Noise Performance

In a given amplifying device whose physical mode of operation is understood, the origins of the significant components of noise are known. Consequently, a noise model of the device can be produced by superimposing a number of noise current or voltage generators on the small-signal equivalent circuit; each generator represents one component of noise. The total noise voltage or current at the output can be determined by applying the principle of superposition to sum the contributions of individual noise generators. It is obvious from Fig. 2.11 that the only consistent way of summing the thermal noise of two resistors is to add the squares of appropriate noise voltages or currents. This simple argument has general validity; the total mean-square output noise from a device is obtained by summing the individual mean-square components.

There may be as many as a dozen components of noise in an amplifying device, and the basic noise model can become unwieldy. A more convenient noise model is obtained if the physical noise generators are replaced by two new generators situated at the input terminals of the equivalent circuit (Fig. 2.12). The equivalent circuit itself is taken as noiseless, and the magnitudes of these two generators are chosen so that together they produce the same output noise as all the physical sources of noise in the device. It is unfortunate that the generators in this alternative

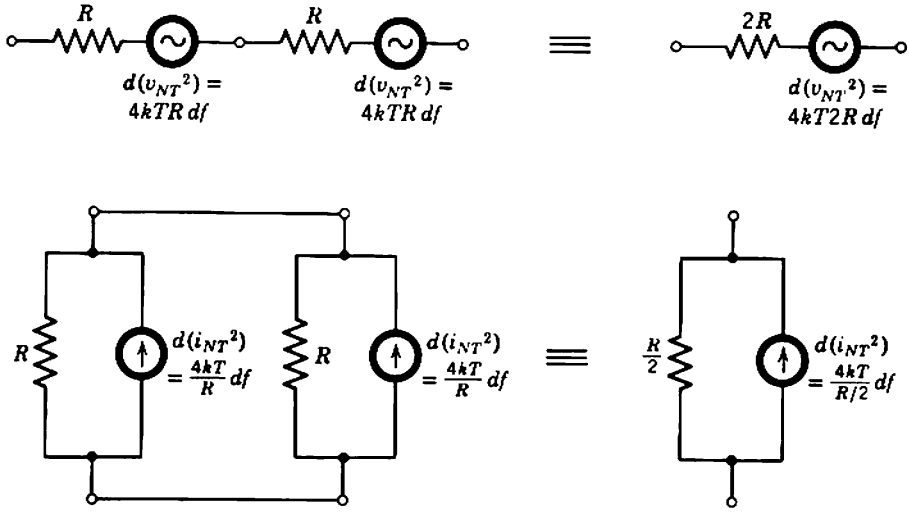


Fig. 2.11 Mean-square summation rule for noise generators.

noise model are not related to single physical noise processes. However, the values of $d(v_N^2)$ and $d(i_N^2)$ can be determined from manufacturer's data or (as shown in Chapter 8) measured or expressed in terms of the component physical processes. A formal difficulty is that the noise processes giving rise to $d(v_N^2)$ and $d(i_N^2)$ are not independent. The generators are said to be *correlated* and the mean-square rule for summation does not apply rigorously. With few exceptions, this formal complication can be ignored, for experimental evidence suggests that correlation is small in commercial amplifying devices. Furthermore, the mismatched technique for circuit design developed in later chapters reduces the contribution of either $d(v_N^2)$ or $d(i_N^2)$ almost to zero. Consequently, it can be assumed that $d(v_N^2)$ and $d(i_N^2)$ are independent, and the over-all noise contribution is obtained by adding the squares of the individual contributions of $d(v_N^2)$ and $d(i_N^2)$.

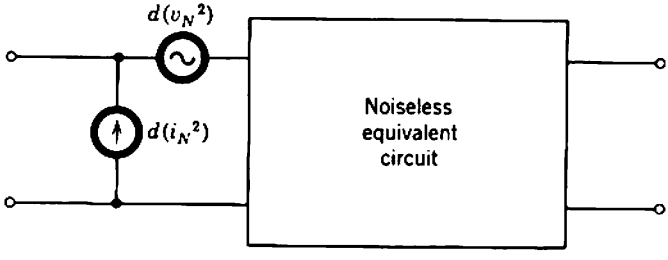


Fig. 2.12 Two-generator noise model of a device.

2.7 RATINGS OF AMPLIFYING DEVICES

To achieve a satisfactory dynamic response from an amplifying device lower and upper limits must be placed on the operating voltage and current. For example, the mutual conductance of a vacuum tube falls away to zero at low currents approaching cutoff and at high currents approaching the temperature-saturated value for the cathode. Additional upper limits are imposed by the manufacturer or user. The life of a device may be reduced if these *ratings* are exceeded, and rapid failure may result if the ratings are grossly exceeded. The factor underlying most ratings is the maximum safe operating temperature for the parts of a device, beyond which deterioration is rapid. Usually it is the ratings that set the practical upper limits to operating voltages and currents rather than considerations of dynamic response.

Because a temperature rise can be associated with the dissipation of energy, a common specification is the maximum power that can be dissipated at any electrode; the power dissipated at any instant is the product of the instantaneous voltage and current. Furthermore, the voltage and current may be separately limited. For example, the maximum allowed anode voltage for a tube is set by the localized rise in temperature due to electrolysis of the insulating supports. The emitter current of a transistor may be limited by the current-carrying capacity of its lead wire.

2.7.1 Thermal Time Constants

One significant difficulty associated with relating a maximum temperature limit to power (or voltage or current) is that the relation depends on the time for which the power is applied. Consider the simple thermal system shown in Fig. 2.13*a*. A body of mass m and specific heat σ is at temperature T in an ambient of temperature T_a. Newton's law of cooling applies approximately, and the rate of loss of heat is

$$\left.\frac{dH}{dt}\right|_{\text{loss}} = -\frac{T - T_a}{\theta},$$

where θ is the thermal resistance. Suppose that power P is dissipated in the device; that is, heat is supplied at a rate P joule per second. The net rate of change in heat is

$$\frac{dH}{dt} = P - \frac{T - T_a}{\theta}.$$

The rate of change of temperature of the body is given by

$$\frac{dH}{dt} = m\sigma \frac{dT}{dt},$$

and elimination of H yields

$$\frac{dT}{dt} + \frac{T - T_a}{\theta m\sigma} = \frac{P}{m\sigma}. \tag{2.66}$$

If power P_{max} is first applied when the mass is at ambient temperature, solution of Eq. 2.66 shows that the temperature increases exponentially with time:

$$T = T_a + \theta P_{max}\left[1 - \exp\left(\frac{-t}{\theta m\sigma}\right)\right]. \tag{2.67}$$

The maximum temperature reached after an infinite time is

$$T_{max} = T_a + \theta P_{max} \tag{2.68}$$

and the time constant in the exponential is known as the *thermal time constant* τ_{th}:

$$\tau_{th} = \theta m\sigma. \tag{2.69}$$

The form of this thermal transient is shown in Fig. 2.13*b*. If the power is supplied for a period greater than a few times τ_{th}, the maximum temperature T_{max} will virtually be reached. On the other hand, if the power P_{max} is supplied for a period t_0 which is considerably smaller than τ_{th}, the peak temperature becomes

$$T_0 \approx T_a + \theta P_{max}\frac{t_0}{\tau_{th}}. \tag{2.70}$$

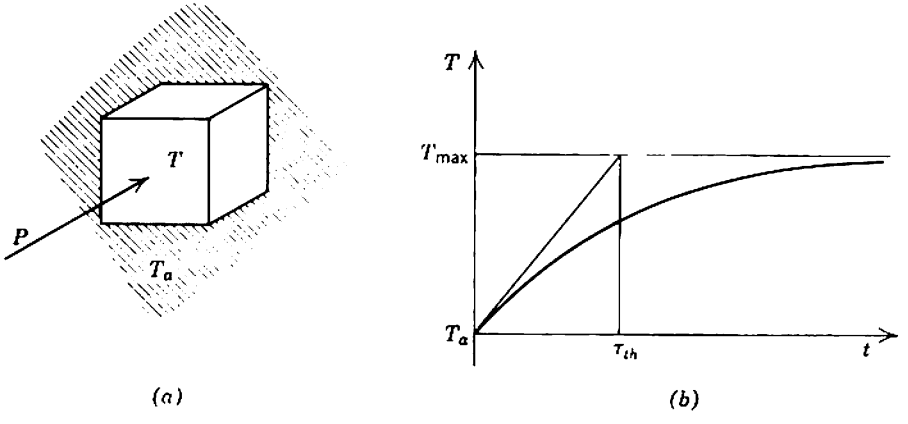


Fig. 2.13 Single-time-constant thermal system.

Hence the power is not supplied long enough for thermal equilibrium to be established. Alternatively, for this short period t_0 a peak power P_0 greater than P_{max} can be supplied without the temperature exceeding T_{max}:

$$P_0 \approx P_{max} \frac{\tau_{th}}{t_0}. \tag{2.71}$$

These results can be extended to the more general case in which the power is a continually varying time function. If the period of the time function is greater than a few times τ_{th} the temperature follows the time variation of the power. Therefore, if T_{max} is not to be exceeded, the peak power must not exceed P_{max}. On the other hand, if the period is much less than τ_{th}, the temperature depends only on the average value of power. Thus the power rating for a device may be taken to refer to average power for devices having a large value of τ_{th}, and peak power for devices having a small value of τ_{th}. Typical examples of these two extremes are vacuum tubes operating at normal signal frequencies and small transistors operating at low audio frequencies.

In most practical devices the energy lost by radiation, conduction, and convection passes through a number of sections of the thermal system before reaching the ultimate thermal sink (usually the atmosphere) which is at ambient temperature. Each of these sections has a different thermal

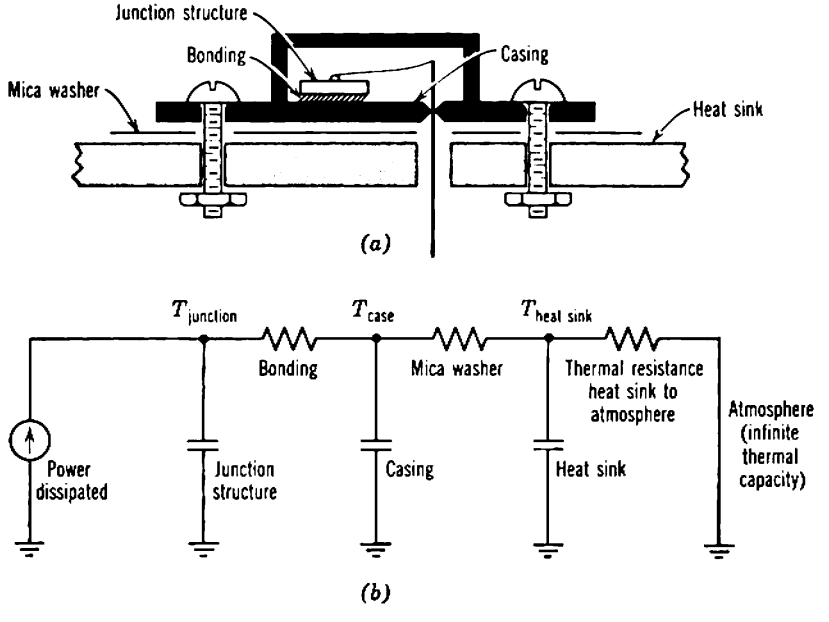


Fig. 2.14 Example of a multi-time-constant thermal system: the transistor.

resistance and thermal capacitance and consequently a different thermal time constant. In an investigation of the behavior of such complex systems a family of linear differential equations must be solved. An alternative approach is to build an electrical analogue of the thermal system and perform direct measurements. Figure 2.14 shows a multi-stage thermal system and its electrical analogue; current represents power, voltage represents temperature, electrical resistance represents thermal resistance, and electrical capacitance represents thermal energy storage. The use of analogues is particularly helpful in determining a maximum temperature when the power varies transiently with time.

2.7.2 Manufacturers' Ratings

Manufacturers' quoted ratings are of two kinds. *Absolute ratings* are the definite limiting values of power, current, or voltage that should never be exceeded. In abnormal applications absolute ratings of devices can be used to predict accurately the safe conditions of operation; electrical analogues are useful in such predictions. *Design center ratings* allow a safety factor for the normally encountered component tolerances and variations in supply voltage. They are very convenient if a device is used as the manufacturer intends, but their lack of precision may prove an embarrassment in unusual applications.

In general, the various ratings quoted for a device are not mutually compensating; it is not permissible to exceed one rating and make a corresponding reduction in another. Unfortunately, manufacturers show little agreement in their methods for specifying device ratings; the imprecise design center ratings are more commonly given and often there is confusion between true ratings and the optimum operating conditions for a specific application. For these reasons device users are sometimes forced to determine absolute ratings for themselves. These ratings may be based on calculations or measurements of maximum temperature in thermally critical sections of the device, experience with similar devices, or life-test data. All ratings should be based ultimately on the results of life tests; typical failure rates at absolute maximum ratings are listed in Table 17.2.

Chapter 3

Characteristics of Amplifying Devices II: Vacuum Tubes

The purpose of this and the next chapter is to apply the general principles of Chapter 2 to the particular cases of vacuum tubes and transistors. Dynamic response, noise, and ratings are considered, together with numerical values for the parameters of typical devices.

In determining the dynamic response, the first step is to derive the fundamental charge-control parameters Q and τ_1 from device physics. The small-signal parameters follow immediately. For simplicity, the initial development of Q and τ_1 is for devices having plane-parallel geometry and other appropriate idealizations. Nevertheless, the final results require only simple, obvious modifications before they can be applied to practical devices. In addition to the charge-control elements, extrinsic elements must appear in the equivalent circuit of practical devices to account for such factors as direct interelectrode capacitances. Because the entire development is founded on device physics, it follows that all parameters in the equivalent circuit vary in simple, predictable ways with operating conditions and temperature. Furthermore, the physical approach allows the effects of manufacturing tolerances and device aging to be estimated.

The noise model of a device is produced by adding the pertinent physical noise generators to the small-signal equivalent circuit. The values of these generators are considered, but the manipulation of the basic noise model into a more convenient practical form is deferred until Chapter 8.

The ratings of both vacuum tubes and transistors are based on maximum safe temperatures for various parts of the devices, but the heating mechanisms are completely different. The physical basis for the ratings is therefore important in interpreting the information supplied by the manufacturer.

Finally, perusal of the specifications for commercial amplifying devices shows that many types bearing different identifying numbers have nearly identical characteristics. Despite the enormous number of device types presently in production, there are, in fact, only a few distinct classes. This point is perhaps the most important single contribution of Chapters 3 and 4.

3.1 THEORY OF VACUUM TUBES

The theory of amplifying vacuum tubes is most conveniently introduced by way of a nonamplifying type, the diode. In this section the characteristics of a diode are determined from the charge-control model, and the triode and pentode are regarded as extensions of the diode. The validity of the resulting small-signal equivalent circuits is assessed by a comparison with quoted results for more accurate models.

3.1.1 The Vacuum Diode

Consider the ideal vacuum diode, shown in Fig. 3.1, which has plane-parallel geometry of area A with the anode and cathode spaced a distance W apart. The variable x represents the distance from the cathode to any

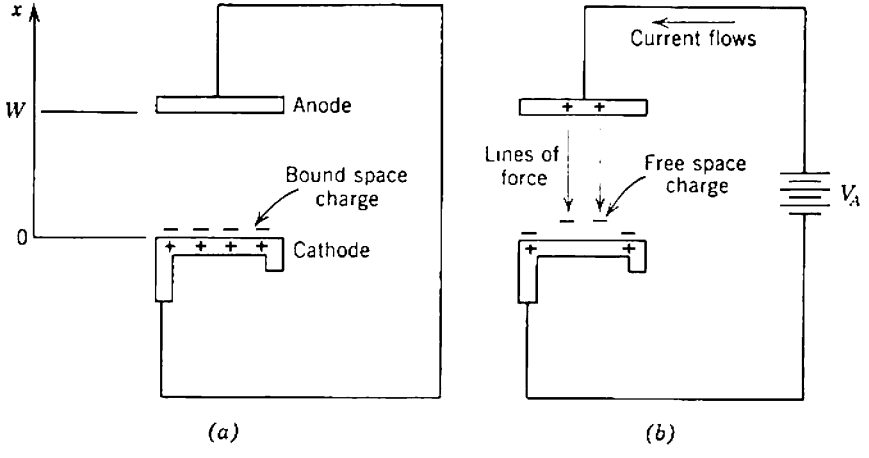


Fig. 3.1 The vacuum diode: (*a*) nonconducting; (*b*) conducting.

point in the conducting channel. For simplicity it is assumed that all electrons are emitted from the heated cathode with zero velocity.

If the cathode temperature is high enough, there will be a large space charge of stationary electrons lying close to the cathode surface and bound to it by the electrostatic attraction of the positive ions (Fig. 3.1*a*). No current flows. However, if a positive charge Q_C is placed on the anode by holding it at a positive potential V_A with respect to the cathode, a portion $Q\,(=-Q_C)$ of the bound space charge will become free. Because the electric field (shown as lines of force in Fig. 3.1*b*) is negative throughout the cathode-anode space, the free electrons drift toward the anode and current flows around the circuit; electrons collected at the anode are replaced by further emission from the cathode. For moderate values of V_A a component of the space charge remains bound to the cathode, which behaves like an infinite source of stationary electrons. A boundary condition in the analysis, therefore, is that the electron velocity is zero at the cathode. An increase in V_A results in an increase in Q and a corresponding increase in I. At a critical value of V_A the bound space charge falls to zero and the assumption of a supply of stationary electrons is invalidated. In other words, the current becomes so large that electrons drift away from the cathode as fast as they are generated thermally. The current remains constant for further increases in V_A and is called the *temperature saturated* value for the particular cathode temperature.

3.1.1.1 Derivation of Q and τ_1

The charge-control parameters Q and τ_1 can be derived as functions of V_A from the equations given in Section 2.4.2. As these equations neglect displacement current, errors will be introduced for signals with rapid rates of change, but the steady values of Q and τ_1 will be predicted correctly. The simplified equations describing the behavior of an electron in an evacuated control region are

$$\frac{I}{A} = +qnv, \tag{3.1}$$

$$\frac{\partial \mathscr{E}}{\partial x} = -\frac{q}{\varepsilon} n, \tag{3.2}$$

$$\mathscr{E} = -\frac{\partial V}{\partial x}, \tag{3.3}$$

$$\frac{dv}{dt} = -\frac{q}{m} \mathscr{E}. \tag{3.4}$$

Notice that a plus sign occurs in Eq. 3.1 because of the reference direction for I in the charge-control model.

In the following development τ_1 and Q are first expressed as functions of I (Eqs. 3.9 and 3.11); I is then expressed as a function of V_A (Eq. 3.13) and eliminated (Eqs. 3.14 and 3.15). It is assumed that the current remains less than the temperature-saturated value for the cathode and the boundary condition $v = 0$ at $x = 0$ applies.

Elimination of n between Eqs. 3.1 and 3.2 gives

$$\frac{\partial \mathscr{E}}{\partial x} = -\frac{I}{\varepsilon A}\frac{dt}{dx}$$

and, integrating with respect to x,

$$\mathscr{E}(x) = -\frac{I}{\varepsilon A}\int_0^x \frac{dt}{dx}\,dx;$$

that is,

$$\mathscr{E}(x) = -\frac{I}{\varepsilon A}\,t(x) \tag{3.5}$$

where $t(x)$ is the transit time from the cathode to the plane at x. Substitution of (3.5) into (3.4) gives

$$\frac{dv}{dt} = \frac{q}{m}\frac{I}{\varepsilon A}\,t. \tag{3.6}$$

By integrating twice, with the boundary conditions $x = v = 0$ at $t = 0$, we obtain

$$v = \frac{dx}{dt} = \frac{q}{2m}\frac{I}{\varepsilon A}\,t^2, \tag{3.7}$$

$$x = \frac{q}{6m}\frac{I}{\varepsilon A}\,t^3. \tag{3.8}$$

For the particular case $x = W$, t is the transit time τ_1 and Eq. 3.8 becomes

$$\tau_1 = \left(\frac{6m\varepsilon A W}{qI}\right)^{1/3} \tag{3.9}$$

The mobile charge density n can now be evaluated. From Eq. 3.1

$$qAn = \frac{I}{v},$$

and by substituting from Eqs. 3.7 and 3.8 we have

$$qAn = \frac{1}{3}\left(\frac{6m\varepsilon A}{q}\right)^{1/3}\left(\frac{I}{x}\right)^{2/3}. \tag{3.10}$$

This variation of n with x is shown in Fig. 3.2. The total charge is

$$Q = -qA \int_0^W n \, dx;$$

that is,

$$Q = -\left(\frac{6m\varepsilon A W}{q}\right)^{1/3} I^{2/3}. \tag{3.11}$$

Finally, I is expressed as a function of V_A. From Eqs. 3.3 and 3.5

$$\frac{\partial V}{\partial x} = \frac{I}{\varepsilon A} t(x).$$

Integration yields

$$V(x) = \frac{I}{\varepsilon A} \int_0^x t(x) \, dx,$$

and, using Eq. 3.8,

$$V(x) = \frac{3}{4\varepsilon A} \left(\frac{6m\varepsilon A}{q}\right)^{1/3} I^{2/3} x^{4/3}. \tag{3.12}$$

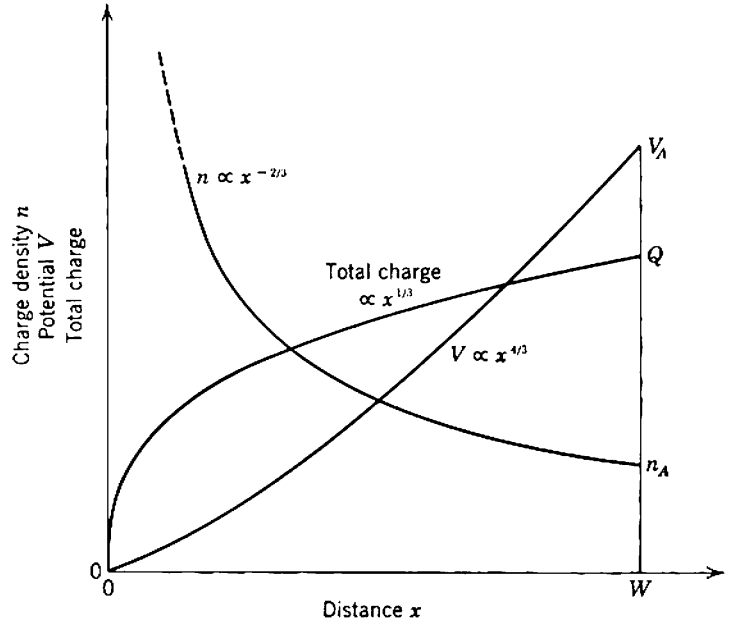


Fig. 3.2 The vacuum diode: charge density, potential, and total charge as functions of distance.

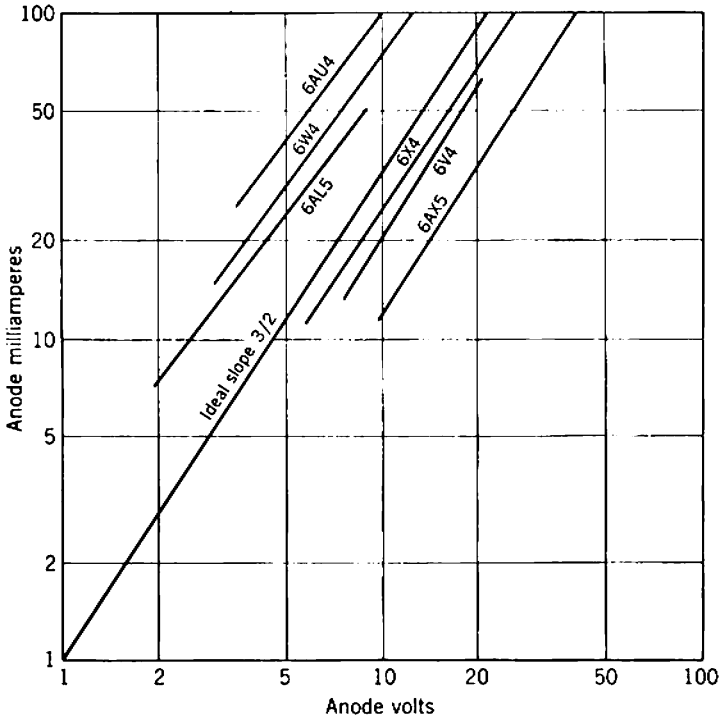


Fig. 3.3 Volt-ampere characteristics of commercial diodes.

This variation of potential with distance is shown in Fig. 3.2. At the anode $V = V_A$ and $x = W$; rearrangement of Eq. 3.12 therefore leads to

$$I = \frac{4\varepsilon A}{9W^2}\left(\frac{2q}{m}\right)^{1/2} V_A^{3/2}. \tag{3.13}$$

Substitution into Eqs. 3.11 and 3.9 gives

$$Q = -\frac{4\varepsilon A}{3W} V_A, \tag{3.14}$$

$$\tau_1 = 3W\left(\frac{m}{2q}\right)^{1/2} V_A^{-1/2}. \tag{3.15}$$

Equation 3.13 is the well-known *three-halves power law* of Langmuir and Child, which is followed by most diodes. Figure 3.3 shows the measured volt-ampere characteristics of a number of commercial diodes on log scales; in each case the characteristic is very nearly a straight line of the ideal slope.

3.1.1.2 Small-Signal Equivalent Circuit

A diode is a two-terminal passive device rather than the three-terminal active device considered in Sections 2.2 and 2.5. By analogy with Eq. 2.9 its controlling equation is

$$I = \frac{Q_C}{\tau_1} + \frac{dQ_C}{dt} = -\frac{Q}{\tau_1} - \frac{dQ}{dt}, \tag{3.16}$$

and a simplified development parallel to Section 2.5 may be undertaken for the two-terminal device. It follows that the equivalent circuit of a diode is the parallel RC combination shown in Fig. 3.4*a*, for which

$$c = \frac{dQ_C}{dV_A} = -\frac{dQ}{dV_A}, \tag{3.17}$$

$$g = \frac{dI}{dV_A} = \frac{c}{\tau_1}(1 + u), \tag{3.18}$$

where

$$u = -\frac{I}{c}\frac{d\tau_1}{dV_A}. \tag{3.19}$$

The capacitance c follows from Eqs. 3.14 and 3.17 as

$$c = \frac{4\varepsilon A}{3W} = \tfrac{4}{3}c_0, \tag{3.20}$$

where c_0 $(= \varepsilon A/W)$ is the direct electrostatic capacitance between the planar electrodes. This change in capacitance is due to the presence of the space charge, which in turn is due to the heated cathode; c is commonly called the *hot capacitance* of the diode to distinguish it from the electrostatic or *cold capacitance* c_0.

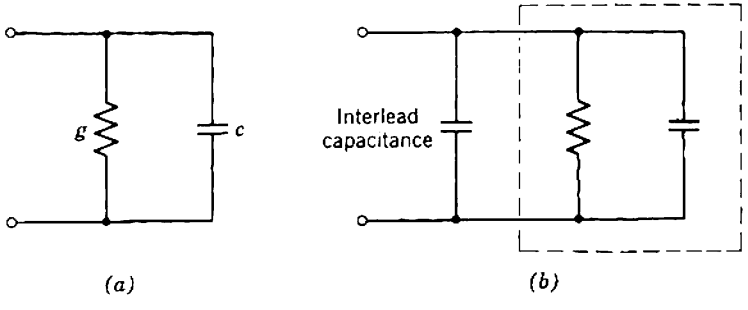


Fig. 3.4 Small-signal equivalent circuit of a diode: (*a*) intrinsic elements; (*b*) complete.

The parameter u corresponds to u_1 and u_2 in Section 2.5, and substitution into Eq. 3.19 shows that

$$u = 0.5. \tag{3.21}$$

All parameters u_k in charge-control theory have this value when the current flow in the control region is space-charge-limited acceleration.*

Finally, substitution into Eq. 3.18 yields

$$g = \frac{3I}{2V_A}. \tag{3.22}$$

This value can be obtained directly by differentiating the Langmuir-Child equation (3.13).

The equivalent circuit is completed by adding the extrinsic interlead capacitance shown in Fig. 3.4*b*.

3.1.1.3 Discussion

The electrons emitted from the cathode of a practical diode have finite emission velocities, the distribution in the x-direction being Maxwellian. Therefore, if the electric field were negative everywhere in the cathode-anode space, all electrons emitted from the cathode would travel to the anode and the current would be temperature-saturated. This behavior is not observed at moderate values of V_A in practical diodes, and consequently the electric field near the cathode must be positive and oppose the initial motion of the electrons. This reversal of the field implies that there is a region of zero field somewhere between the anode and cathode and that the potential has a negative minimum in this zero-field region. If the Maxwellian distribution is approximated by assuming that all electrons have the same emission velocity, the zero-field region reduces to a plane. Electrons give up their kinetic energy to the opposing electric field between the cathode and this *barrier plane* and on the average have zero kinetic energy at the plane. Thus the barrier plane may be considered as a virtual cathode that provides zero emission velocity, and the controlled space charge Q exists between it and the anode. The theory in Section 3.1.1.1 can be applied to practical diodes with the following substitutions:

$$V \rightarrow V' + V_m,$$

$$x \rightarrow x' - x_m,$$

where $-V_m$ is the minimum potential and x_m is the position of the barrier

* R. D. MIDDLEBROOK, "A modern approach to semiconductor and vacuum device theory," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 887, Suppl. 17, 1959.

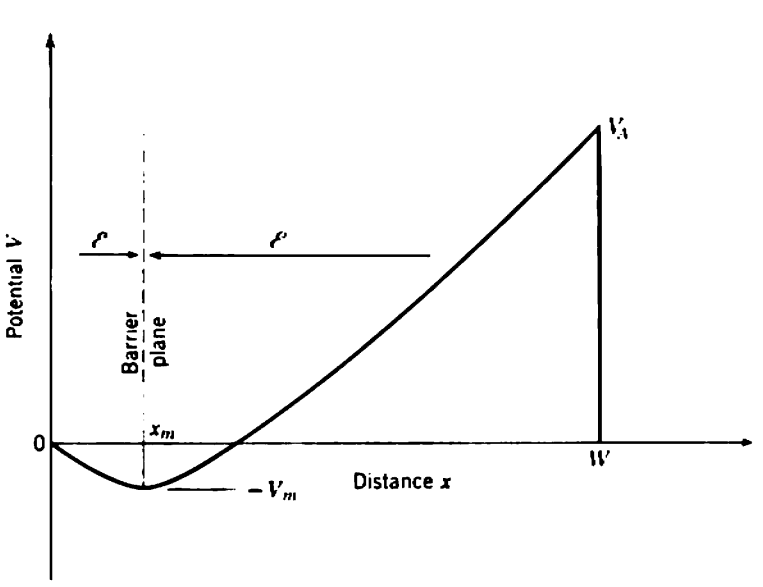


Fig. 3.5 Potential in a diode with finite emission velocity. Notice the reversal of the electric field.

plane (Fig. 3.5). Expressions for V_m and x_m are available in specialized books.* Both depend slightly on the space-charge density and emission velocity, and therefore second-order changes occur in g and c with changes in current and heater voltage (cathode temperature).

Most practical diodes have nonplanar geometry. However, the Langmuir-Child law

$$I = \text{const } V^{3/2}$$

has been shown to be universally true in at least three ways:

(i) by the use of dimensional analysis,
(ii) by carrying out theoretical developments for other simple geometries, for example, cylindrical electrodes,
(iii) by observing the volt-ampere characteristics of practical diodes.

Changes from the simple planar geometry affect only the value of the constant; the same conclusion applies for all other relations between Q, τ_1, V_A, and I.

The conductance g predicted from Eq. 3.18 agrees with measurements on practical diodes at low frequencies, but not at high frequencies. As

* For example, K. R. SPANGENBERG, *Vacuum Tubes*, Chapter 8 (McGraw-Hill, New York, 1948).

noted in Section 3.1.1.1, this error is expected because the displacement component of current is neglected in Eq. 3.1. Accurate derivations show that g varies with frequency, as in Fig. 3.6; the frequency at which g departs appreciably from its low-frequency value is of the order of 100 MHz for typical diodes.

The capacitance c predicted from Eq. 3.17 is in error at all frequencies. The error at high frequencies is expected, being caused by the approximation in Eq. 3.1; the error at low frequencies is perhaps surprising but can be traced back to the fundamental assumption of all charge-control theory (Eq. 2.33). In the development of Eq. 2.9 for the general three-electrode charge-control model it is assumed that the mobile charge takes up its final spatial distribution immediately on a change in electrode voltages. This assumption is not true for any practical device, and the change in output current lags behind a change in electrode voltage. Formally, Eq. 2.9 should be replaced by Eq. 2.34:

$$I_2 = \frac{Q_c}{\tau_1} + \frac{dQ_{c2}}{dt} - I_\epsilon(t),$$

where $I_\epsilon(t)$ is the error current. It follows that the capacitance c_{22} is not given by Eq. 2.44 but by

$$c_{22} = \left(\frac{\partial Q_{c2}}{\partial V_2}\right)_{V_1} - \left(\frac{\partial}{\partial V_2}\right)_{V_1} \int I_\epsilon(t)\, dt. \tag{3.23}$$

For the special case of the two-electrode diode the capacitance becomes

$$c = \frac{dQ_c}{dV_A} - \frac{d}{dV_A} \int I_\epsilon(t)\, dt. \tag{3.24}$$

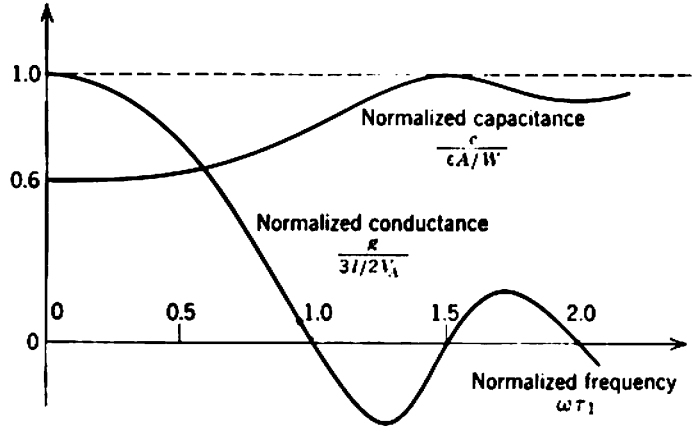


Fig. 3.6 Conductance and capacitance of a diode as functions of frequency.

The variation of c with frequency is derived in specialized books* and is plotted in Fig. 3.6. In many practical diodes the extrinsic interlead capacitance is so large that it completely masks the intrinsic capacitance and the error in Eq. 3.17 is not significant.

3.1.2 The Vacuum Triode

Consider the ideal vacuum triode shown in Fig. 3.7, which has plane-parallel geometry of area A and electrode spacings W_G and W_A. As in the case of the diode, assume that the cathode emits electrons with zero initial velocity and that the current is less than the temperature-saturated value.

3.1.2.1 The Equivalent Diode

A triode can be analyzed by means of an equivalent diode whose anode lies in the plane at W_e and is at the potential V_e that occurs at this plane in the triode. Conditions between the cathode and plane are therefore the same in both the actual triode and the hypothetical diode. Equations

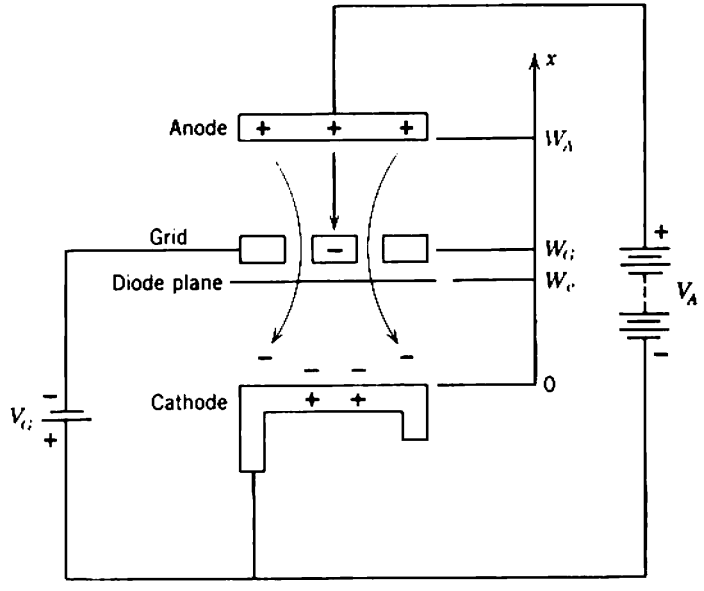


Fig. 3.7 The vacuum triode.

* For example, W. E. BENHAM and I. A. HARRIS, *The Ultra High Frequency Performance of Receiving Tubes*, Chapter 3 (McGraw-Hill, New York, 1957). For additional references and further discussion see K. R. SPANGENBERG, *op. cit.*, Chapter 16.

3.14 and 3.15 give the controlled space charge Q and the transit time τ_1 of the diode as

$$Q = -\frac{4\varepsilon A}{3W_e} V_e = -\tfrac{4}{3} V_e c_{eK0} \tag{3.25}$$

$$\tau_1 = 3W_e\left(\frac{m}{2q}\right)^{1/2} V_e^{-1/2} \tag{3.26}$$

where c_{eK0} is the electrostatic or cold capacitance between the plane and cathode. The controlling charge Q_C must be equal to the sum of all charges above the plane, namely, the positive charges Q_{C1} and Q_{C2} on the grid and anode, respectively, and the negative space charge between W_e and W_A. If

$$W_e = W_G,$$

so that the equivalent anode lies immediately below the grid, the space charge between W_e and W_A is very small. Therefore

$$Q_C = (V_G - V_e)c_{eG0} + (V_A - V_e)c_{eA0},$$

where c_{eG0} and c_{eA0} are cold capacitances between the plane and the triode grid and anode, respectively. Because Q and Q_C are equal and opposite,

$$(V_G - V_e)c_{eG0} + (V_A - V_e)c_{eA0} = \tfrac{4}{3} V_e c_{eK0},$$

from which

$$V_e = \frac{V_G + V_A \dfrac{c_{eA0}}{c_{eG0}}}{1 + \dfrac{c_{eA0}}{c_{eG0}} + \dfrac{4}{3}\dfrac{c_{eK0}}{c_{eG0}}}. \tag{3.27}$$

The capacitances in Eq. 3.27 are the direct electrostatic capacitances in the absence of space charge. The values of c_{eA0} and c_{eK0} can be written down

$$c_{eA0} = \frac{\varepsilon A}{W_A - W_e} = \frac{\varepsilon A}{W_A - W_G},$$

$$c_{eK0} = \frac{\varepsilon A}{W_e} = \frac{\varepsilon A}{W_G},$$

but the value of c_{eG0} cannot, for it depends on the geometry of the grid. It is therefore an advantage to perform the star-delta transformation shown in Fig. 3.8 to replace ratios of hypothetical capacitances by ratios of the measurable capacitances between anode, grid, and cathode:

$$\frac{c_{eA0}}{c_{eG0}} = \frac{c_{AK0}}{c_{GK0}},$$

$$\frac{c_{eK0}}{c_{eG0}} = \frac{c_{AK0}}{c_{AG0}}.$$

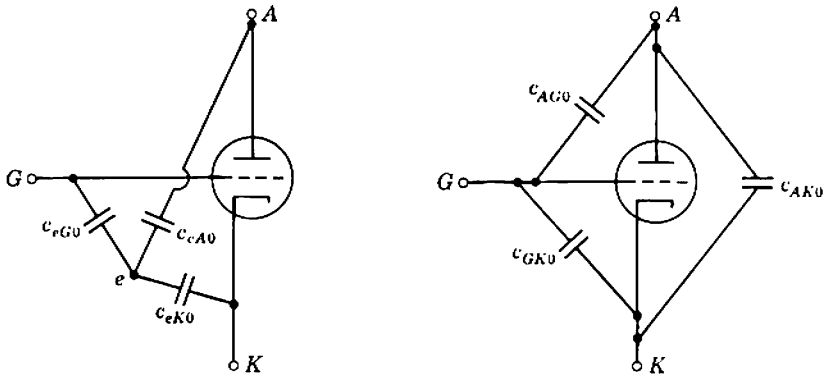


Fig. 3.8 Star-delta transformation of the cold capacitances of a triode.

It is a further advantage to define

$$\mu = \frac{c_{GK0}}{c_{AK0}}. \tag{3.28}$$

The value of μ depends only on a tube's geometry and lies between 10 and 100 for typical modern triodes. Thus

$$V_e = \frac{V_G + V_A/\mu}{1 + \dfrac{1}{\mu}\left(1 + \dfrac{4}{3}\dfrac{c_{GK0}}{c_{AG0}}\right)}, \tag{3.29}$$

and because μ is moderately large

$$V_e \approx V_G + \frac{V_A}{\mu}. \tag{3.30}$$

In the ideal case of infinite μ

$$V_e = V_G$$

and V_A has no effect at all; μ gives a measure of the effectiveness of the grid as a shield between the anode and space charge.

Finally, the charge-control parameters for the equivalent diode whose anode lies in the plane of the triode grid follow from Eqs. 3.25, 3.26, and 3.30 as

$$Q_C = -Q = \frac{4\varepsilon A}{3W_G} V_e \approx \frac{4\varepsilon A}{3W_G}\left(V_G + \frac{V_A}{\mu}\right), \tag{3.31}$$

$$\tau_1 = 3W_G\left(\frac{m}{2q}\right)^{1/2} V_e^{-1/2} \approx 3W_G\left(\frac{m}{2q}\right)^{1/2}\left(V_G + \frac{V_A}{\mu}\right)^{-1/2}. \tag{3.32}$$

3.1.2.2 Small-Signal Equivalent Circuit

The cathode currents of the triode and equivalent diode are equal, being given under static conditions by

$$I_K = \frac{Q_C}{\tau_1} = GV_e^{3/2} \approx G\left(V_G + \frac{V_A}{\mu}\right)^{3/2}, \tag{3.33}$$

where the constant G is the *perveance* of the tube:

$$G = \frac{4\varepsilon A}{9W_G^2}\left(\frac{2q}{m}\right)^{1/2}. \tag{3.34}$$

For a vacuum tube to be used as an amplifier, cathode current must flow and therefore V_e must be positive. The anode voltage V_A is made positive and large (of the order of 100 V), but V_G is made a few volts negative so that the grid structure will repel the space charge. The probability of an electron striking the grid wires is therefore small, and the grid current approaches zero. In other words, the effective lifetime τ of electrons in the control region is very large, the dc current gain approaches infinity, and

$$I_A = I_K \approx G\left(V_G + \frac{V_A}{\mu}\right)^{3/2} \tag{3.35}$$

provided that

$$\frac{-V_A}{\mu} < V_G < 0.$$

The small-signal equivalent circuit derived in Section 2.5 (Fig. 2.7*d*) is redrawn in Fig. 3.9*a* with elements labeled as the specific parameters of a tube. The parameter r_A $(= 1/g_2)$ is called the *anode resistance*. Substitution of Eqs. 3.31 and 3.32 into the relevant equations* of Section 2.5 results in

$$g_m = \left(\frac{\partial I_A}{\partial V_G}\right)_{V_A} = \frac{3G^{2/3}}{2} I_K^{1/3}, \tag{3.36}$$

$$\frac{1}{r_A} = \left(\frac{\partial I_A}{\partial V_A}\right)_{V_G} = \frac{3G^{2/3}}{2\mu} I_K^{1/3}, \tag{3.37}$$

$$c_{GK} = \left(\frac{\partial Q_C}{\partial V_G}\right)_{V_A} = \frac{4\varepsilon A}{3W_G}, \tag{3.38}$$

$$c_{AK} = \left(\frac{\partial Q_C}{\partial V_A}\right)_{V_G} = \frac{4\varepsilon A}{3\mu W_G}. \tag{3.39}$$

* Equations 3.38 and 3.39 are derived first by substitution into Eqs. 2.39 and 2.40. The parameters u_1 and u_2 follow from Eqs. 2.47 and 2.49 as 0.5, and substitution into Eqs. 2.53 and 2.55 yields Eqs. 3.36 and 3.37.

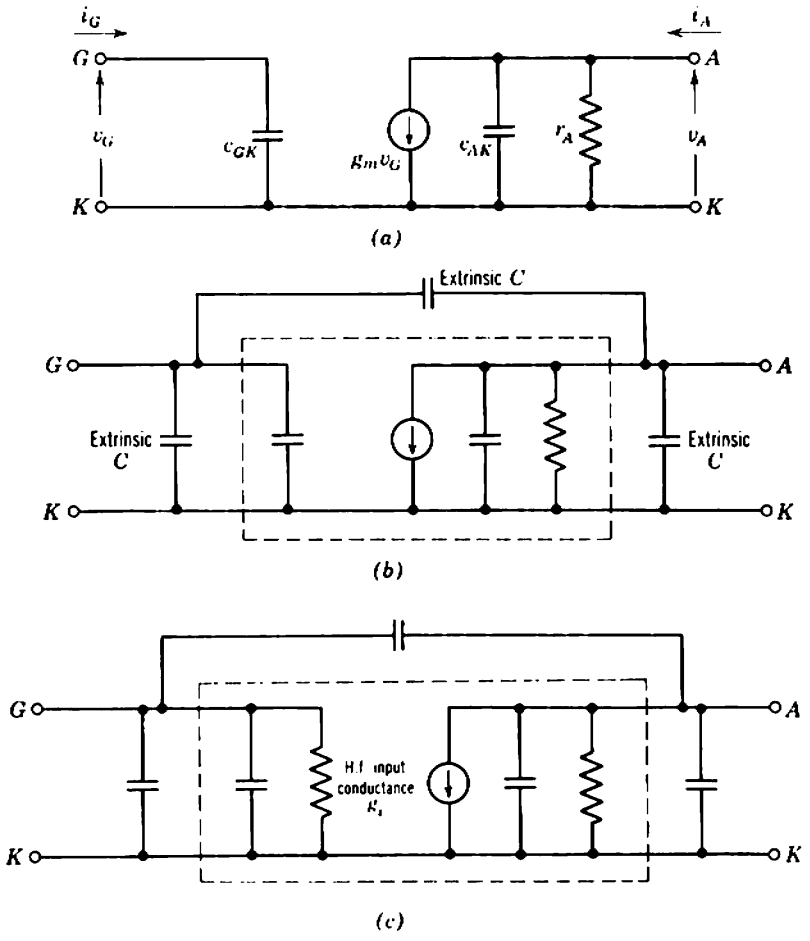


Fig. 3.9 Small-signal equivalent circuit of a triode: (*a*) intrinsic elements; (*b*) complete at low frequencies; (*c*) complete at high frequencies.

Notice that c_{GK} and c_{AK} are the hot capacitances of the tube—the capacitances in the presence of space charge. Rearrangement of these equations shows that the geometrical parameter μ is given by any of the following expressions:

$$\mu = -\left(\frac{\partial V_A}{\partial V_G}\right)_{I_A} = g_m r_A = \frac{c_{GK0}}{c_{AK0}} = \frac{c_{GK}}{c_{AK}}. \tag{3.40}$$

Thus μ is identified as the voltage amplification factor* introduced in

* In common usage the word *voltage* is omitted because the current amplification factor of a tube is not meaningful. In the interests of clarity and consistency the word *voltage* is retained throughout this book.

Section 2.5.1. Equations 3.36 and 3.37 can be derived without reference to charge-control theory by differentiating Eq. 3.35.

Addition of the extrinsic interelectrode capacitances shown in Fig. 3.9*b* completes the small-signal equivalent circuit.

3.1.2.3 Discussion

The simplifying assumptions for the diode and triode introduce similar inaccuracies. Because of the finite emission velocities of electrons in a triode, a barrier plane forms between the cathode and grid. The distances W_G and W_A ought to be referred to this plane, and the potentials V_G and V_A to the attendant potential minimum. Changes in the barrier plane with current and cathode temperature cause second-order perturbations about the cube-root laws predicted for g_m and r_A and second-order changes in c_{GK} and c_{AK}. Most practical triodes have nonplanar geometry and the constant multipliers in the equations have different values.

With these modifications, charge-control theory correctly predicts the mutual conductance g_m and anode resistance r_A for a triode at low frequencies. The output capacitance c_{AK} is wrong for the same reason that the diode capacitance is wrong, but the input capacitance c_{GK} is correct. Formally, c_{GK} is derived from Eq. 2.35, which is exact and does not depend on the basic assumption of all charge-control theory. It should be pointed out that c_{AK} is so small compared with the extrinsic capacitance of practical triodes that the error in charge-control theory is difficult to detect experimentally.

Further errors are introduced at high frequencies by the omission of displacement current from Eq. 3.1. All parameters* become functions of signal frequency, but over the frequency range for which a triode is useful in low-pass amplifier circuits the more important elements g_m and c_{GK} remain fairly constant. In addition, a triode develops the input conductance† shown in Fig. 3.9*c*:

$$g_i \approx \frac{g_m \omega^2 \tau_1^2}{20} \left(1 + \frac{44}{9} \frac{\tau_2}{\tau_1}\right), \tag{3.41}$$

where τ_2 is the transit time between grid and anode.

Extrinsic elements not shown in Fig. 3.9*b* become significant at high frequencies, the two most likely elements being the inductance of the cathode lead and the interface resistance that occurs just below the emitting surface of an oxide-coated cathode. These elements can be

* The parameters are derived in BENHAM and HARRIS, *op. cit.*, Chapter 4. Discussion without derivation is in SPANGENBERG, *loc. cit.*

† The origin of this conductance is discussed in BENHAM and HARRIS, *loc. cit.*

added in the cathode leg of the equivalent circuit. However, their major effect is to increase the input conductance of a tube by a component that varies as ω^2. Therefore, it is more convenient to combine this conductance with the component given by Eq. 3.41 as a single effective value. Cathode interface resistance may build up to a few tens of ohms and reduce the effective g_m as a tube approaches the end of its life.

In summary, the equivalent circuit of Fig. 3.9*c* represents the dynamic response of a vacuum triode over the useful frequency range.

3.1.2.4 Grid Current

It is assumed in the argument leading to Eq. 3.35 that grid current is zero provided V_G is negative. The grid current is indeed negligible for many practical purposes, but it is finite and may be as large as 10^{-8} A. Although small, this current has important consequences in high-input-impedance amplifiers and some low-noise applications.

Components of grid current can originate in the following ways:

1. Some electrons collide with the grid structure to produce a component of current directed into the tube. This positive component falls as the grid is taken more negative (Fig. 3.10) but rises sharply if the grid becomes positive with respect to the barrier potential.

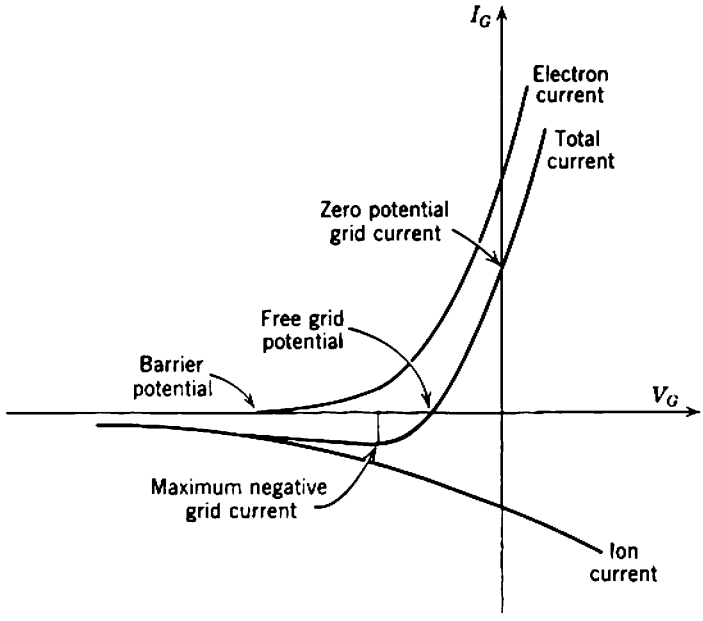


Fig. 3.10 Grid current in a triode.

2. Some electrons collide with residual gas molecules in the tube to produce positive ions. Most of these ions are collected by the grid and produce a component of grid current directed out of the tube. If the rate of ionization were constant, this component would be substantially independent of grid voltage. However, the ionization rate is roughly proportional to the anode current and therefore decreases as the grid is taken more negative.

3. In high g_m tubes (for which the grid is very close to the cathode) the grid may be so heated by radiation that it emits thermionic electrons. This results in another negative component of grid current (usually far smaller than the ion component), whose magnitude depends on the cathode temperature. Grids are often gold plated to reduce thermionic emission.

4. Leakage paths may exist between the grid and other electrodes, both inside and outside the tube envelope.

There are other possible mechanisms, but in most cases (1) and (2) predominate and the over-all grid volt-ampere characteristic has the form shown in Fig. 3.10. Notice the existence of a *free grid potential* at which the current is zero but the slope resistance is relatively low (perhaps 10^8 Ω) and a point of *maximum negative current* at which the slope resistance is infinite.

3.1.3 The Vacuum Pentode

The vacuum tetrode and pentode shown in Fig. 3.11 were developed originally to overcome the circuit disadvantages of the large extrinsic capacitance c_{AG} in a triode. This capacitance is reduced in a tetrode by inserting a shielding electrode, the *screen grid*, between the normal or *control grid* and the anode. A fifth electrode, the *suppressor grid*, is placed between the screen and anode of a pentode to prevent impact-produced secondary electrons moving from the anode to the screen. The undesirable consequences of the large anode-to-grid capacitance in a triode and secondary emission in a tetrode are discussed in many books* and are not repeated here. In normal operation the screen is connected to a positive supply, typically 100 V; the suppressor is connected to the cathode.

Because the controlled space charge Q exists principally in the cathode-grid space, a pentode can be investigated by means of its equivalent diode. Extension of the triode analysis shows that

$$V_e \approx V_G + \frac{V_S}{\mu_S} + \frac{V_A}{\mu}, \tag{3.42}$$

* For example, F. E. TERMAN, *Electronic and Radio Engineering*, 4th ed., Chapters 6.9 and 12.10 (McGraw-Hill, New York, 1955); see also Section 7.5.2.1.

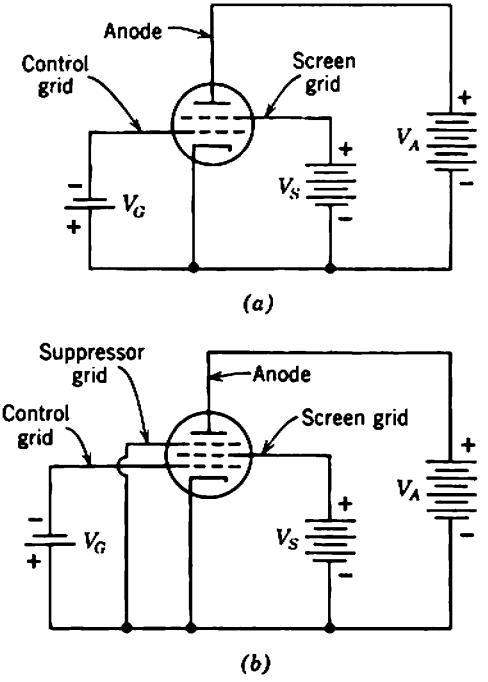


Fig. 3.11 The vacuum tetrode and pentode.

where μ_S is the *screen amplification factor*, defined, by comparison with the triode, as

$$\mu_S = -\left(\frac{\partial V_S}{\partial V_G}\right)_{I_A, V_A} \tag{3.43}$$

The controlling charge follows as

$$Q_C = -Q = \frac{4\varepsilon A}{3W_G} V_e \approx \frac{4\varepsilon A}{3W_G}\left(V_G + \frac{V_S}{\mu_S} + \frac{V_A}{\mu}\right), \tag{3.44}$$

from which

$$I_K = GV_e^{3/2} \approx G\left(V_G + \frac{V_S}{\mu_S} + \frac{V_A}{\mu}\right)^{3/2}, \tag{3.45}$$

where the perveance G is given by Eq. 3.34.

The grid acts as a partial shield between the screen and space charge and, like μ for a triode, μ_S gives a measure of this shielding. With similar grid structures, μ for a triode and μ_S for a pentode are comparable in magnitude. The anode of a pentode is triply shielded from the space charge by the grid, screen, and suppressor. Accordingly, V_A has almost no control over Q, and μ for a pentode is very large indeed. The V_A terms in Eqs. 3.42, 3.44, and 3.45 are often omitted.

If V_G is negative, the cathode current divides between the positive screen and anode. The ratio depends almost entirely on the geometry of the screen structure and is independent of the electrode voltages, provided V_A is greater than a value about 20% of V_S known as the *bottoming voltage* V_{AB}. Thus

$$I_S = kI_A, \tag{3.46}$$

where k is a constant, typically 0.2, and

$$I_A = \frac{1}{1+k} I_K \approx \frac{G}{1+k}\left(V_G + \frac{V_S}{\mu_S} + \frac{V_A}{\mu}\right)^{3/2}, \tag{3.47}$$

$$I_S = \frac{k}{1+k} I_K \approx \frac{kG}{1+k}\left(V_G + \frac{V_S}{\mu_S} + \frac{V_A}{\mu}\right)^{3/2}. \tag{3.48}$$

Normally a pentode is operated with V_A greater than the bottoming or *knee* voltage V_{AB}, and I_A is given by Eq. 3.47. For values of V_A less than V_{AB}, I_A varies roughly as $V_A^{3/2}$.* Electrons that do not strike the screen structure enter the screen-anode space and form a space charge in the vicinity of the suppressor. This space charge behaves as a virtual cathode and in conjunction with the anode forms a space-charge-limited diode. As V_A is increased, the space charge falls and I_A increases until all available electrons are swept to the anode. Typical curves of anode, screen, and suppressor currents versus V_A are shown in Fig. 3.12.

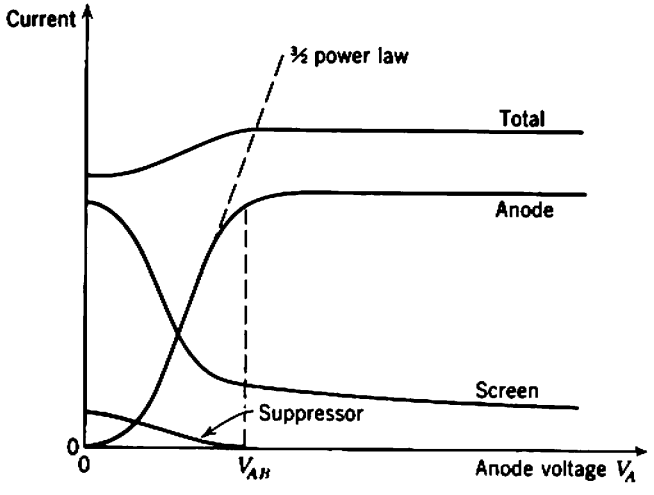


Fig. 3.12 Electrode currents in a pentode as a function of V_A; V_G and V_S are held constant.

* SPANGENBERG, *op. cit.*, Chapter 11.

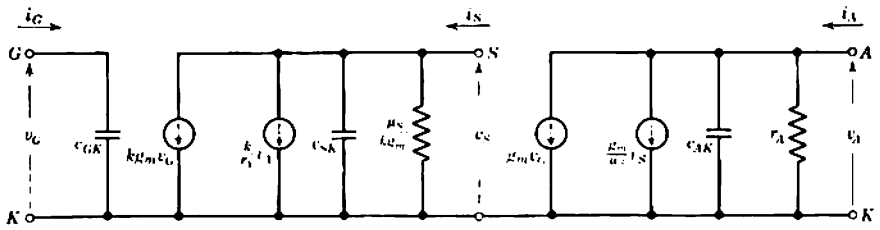


Fig. 3.13 Small-signal intrinsic equivalent circuit of a pentode.

The intrinsic small-signal equivalent circuit of a pentode contains the nine elements shown in Fig. 3.13. The terminal equations are

$$\Delta Q_C = c_{GK}\,\Delta V_G + c_{SK}\,\Delta V_S + c_{AK}\,\Delta V_A, \tag{3.49}$$

$$\Delta I_A = g_m\,\Delta V_G + \frac{g_m}{\mu_S}\Delta V_S + \frac{1}{r_A}\Delta V_A, \tag{3.50}$$

$$\Delta I_S = kg_m\,\Delta V_G + \frac{kg_m}{\mu_S}\Delta V_S + \frac{k}{r_A}\Delta V_A, \tag{3.51}$$

where Eqs. 3.44, 3.47, and 3.48 give

$$g_m = \left(\frac{\partial I_A}{\partial V_G}\right)_{V_S,V_A} = \frac{3G^{3/2}}{2(1+k)}\,I_K^{1/2}, \tag{3.52}$$

$$\frac{1}{r_A} = \left(\frac{\partial I_A}{\partial V_A}\right)_{V_G,V_S} = \frac{3G^{3/2}}{2\mu(1+k)}\,I_K^{1/2}, \tag{3.53}$$

$$c_{GK} = \left(\frac{\partial Q_C}{\partial V_G}\right)_{V_S,V_A} = \frac{4\varepsilon A}{3W_G}, \tag{3.54}$$

$$c_{SK} = \left(\frac{\partial Q_C}{\partial V_S}\right)_{V_G,V_A} = \frac{4\varepsilon A}{3\mu_S W_G}, \tag{3.55}$$

$$c_{AK} = \left(\frac{\partial Q_C}{\partial V_A}\right)_{V_G,V_S} = \frac{4\varepsilon A}{3\mu W_G}. \tag{3.56}$$

A complete equivalent circuit would include extrinsic interlead capacitances and a frequency-dependent input conductance.

3.2 NOISE SOURCES IN VACUUM TUBES

Over the frequency range from a few kilohertz to a few megahertz the dominant noise in a vacuum tube is the *shot noise* associated with each electron stream. There is only one electron stream in a diode or negative-grid triode, from the cathode to the anode, and there is only one component of shot noise. There are two streams in a negative-grid pentode,

from the cathode to both anode and screen, but the total noise exceeds the shot noise expected of them in isolation. The excess component stems from the random division of cathode current between the conducting electrodes and is termed *partition noise*. Both shot and partition noise have uniform power spectra, the mean-square noise current being proportional to bandwidth. A specification of noise current is therefore meaningless unless the bandwidth is given also.

Other components of noise become significant at the ends of the frequency range; *flicker noise* appears at low frequencies, *induced grid noise* at high frequencies. Because both have power spectra that vary with frequency, it is necessary to specify the center frequency as well as the bandwidth over which the noise is measured.

3.2.1 Vacuum Triode

The noise performance of a vacuum triode can be represented over a wide frequency range by the model shown in Fig. 3.14. This model is a combination of the small-signal equivalent circuit (Fig. 3.9*c*) with three noise generators.

The shot noise on the anode current is represented over a bandwidth *df* by the current generator

$$d(i_{NS}^2)_A = 2q\Gamma^2 I_A\, df. \tag{3.57}$$

It is shown in specialized books* that the reduction factor for a space-charge-limited electron stream is

$$\Gamma^2 = \frac{3k(0.644T_K)}{qV_c}, \tag{3.58}$$

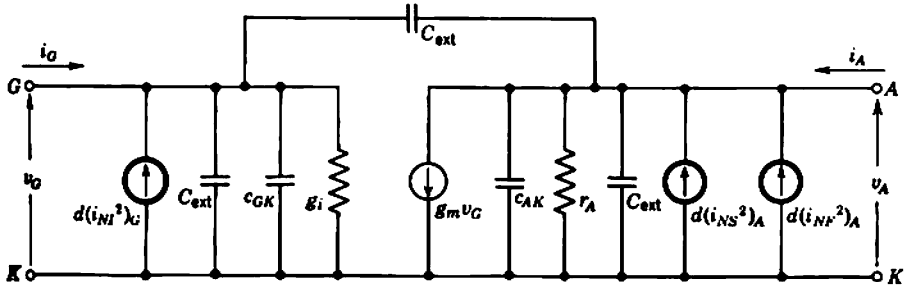


Fig. 3.14 Noise model of a triode.

* For example, C. F. QUATE, "Shot noise from thermionic cathodes," Chapter 1, in L. D. SMULLIN and H. A. HAUS (eds.), *Noise in Electron Devices* (Technology Press and Wiley, New York, 1959). The factor 0.644 is $3(4 - \pi)/4$.

where k = Boltzmann's constant,
T_K = the cathode temperature, typically 1000°K for small tubes,
V_e = the anode potential of the equivalent diode.

Equations 3.33 and 3.36 give

$$V_e = \frac{3I_K}{2g_m} = \frac{3I_A}{2g_m}, \tag{3.59}$$

so that

$$d(i_{NS}^2)_A = 4k(0.644T_K)g_m\,df. \tag{3.60}$$

Thus shot noise can be given another interpretation—the thermal noise of g_m at temperature $(0.644T_K)$.

The flicker noise on the anode current is represented over a bandwidth df by the current generator

$$d(i_{NF}^2)_A = K\frac{I_A^c}{f}\,df. \tag{3.61}$$

The exponent c is about 2 for most vacuum tubes; K is a few times 10^{-13} for a low-noise triode but may be one or two orders of magnitude larger for triodes with poor flicker-noise performance. For typical tubes the values of g_m, T_K, and the empirical constant K are such that flicker noise exceeds shot noise at frequencies below a few kilohertz. Flicker noise in a given tube can be reduced only by reducing the anode current.

Individual electrons induce a charge on the grid as they pass it *en route* to the anode. This charge varies randomly with time, and it is shown in specialized books* that the induced grid noise can be represented over a bandwidth df by a current generator

$$d(i_{NI}^2)_G = 4k(1.431T_K)g_i\,df, \tag{3.62}$$

where g_i is the high-frequency input conductance of the tube. Because g_i varies as f^2, induced grid noise vanishes at low frequencies.

3.2.2 Vacuum Pentode

Shot noise is present on both the anode and screen electron streams of a pentode, and Zieglerf has shown that

$$d(i_{NS}^2)_A = 2q\left(\Gamma^2 I_A + \frac{I_A I_S}{I_A + I_S}\right)df, \tag{3.63}$$

$$d(i_{NS}^2)_S = 2q\left(\Gamma^2 I_S + \frac{I_A I_S}{I_A + I_S}\right)df. \tag{3.64}$$

---

* For example, BENHAM and HARRIS, *op. cit.*, Chapter 10. The factor 1.431 is $5(4 - \pi)/3$.

† M. ZIEGLER, "Noise in amplifiers contributed by the valves," *Philips Tech. Rev.*, **2**, 329, November 1937.

In these equations the first term in brackets represents the shot noise on the space-charge-limited streams in isolation:

$$\Gamma^2 = \frac{3k(0.644T_K)}{qV_e},$$

where Eqs. 3.45 and 3.52 show that

$$V_e = \frac{3I_A}{2g_m} \tag{3.65}$$

for a pentode as well as a triode. The second terms in Eqs. 3.63 and 3.64 represent the partition noise. Notice that the contribution is the same for each electron stream; an impulsive change in anode current necessitates an equal and opposite change in screen current. This applies strictly only if the electron stream in the vicinity of the screen is free from fluctuations. In practice, the space charge smooths out the fluctuations in the cathode current stream and Ziegler's simple equations are obeyed with fair precision.* The shot noise of a pentode (including partition) can be represented by the three mean-square noise current generators in Fig. 3.15*a*, and superposition can be used to determine the mean-square noise voltage or current anywhere in the circuit.

If the screen is fed from a dc supply of finite internal impedance, a noise voltage is developed at the screen by the noise current. This noise voltage excites the screen-anode mutual conductance $(g_m/\mu_S)v_S$ in Fig. 3.13 and consequently increases the mean-square anode noise current. However, if the screen is connected by a low (ideally zero) impedance path to the cathode, no screen noise voltage can be developed. The model simplifies to Fig. 3.15*b* and the anode noise current is reduced. In this case it is convenient to combine the uncorrelated mean-square anode noise current generators (Fig. 3.15*c*) and express the total shot noise in the form of Eq. 2.64:

$$d(i_{NS}^2)_A = 2q\Gamma_A^2 I_A\, df. \tag{3.66}$$

The effective space-charge reduction factor for the anode is

$$\Gamma_A^2 = \Gamma^2 + \frac{k}{1+k}, \tag{3.67}$$

where k is the ratio of dc screen current to dc anode current. The practical conclusion is that the screen should be bypassed to the cathode with a capacitor.

---

* For accurate equations and more involved theoretical arguments, see W. SCHOTTKY, "Small-shot effect and flicker effect," *Phys. Rev.*, **28**, 74, July 1926; C. J. BAKKER, "Current distribution fluctuation in multielectrode radio valves," *Physica*, **5**, 58, July 1938; D. O. NORTH, "Fluctuations in space-charge-limited currents at moderately high frequencies," *RCA Rev.*, **5**, 244, October 1940.

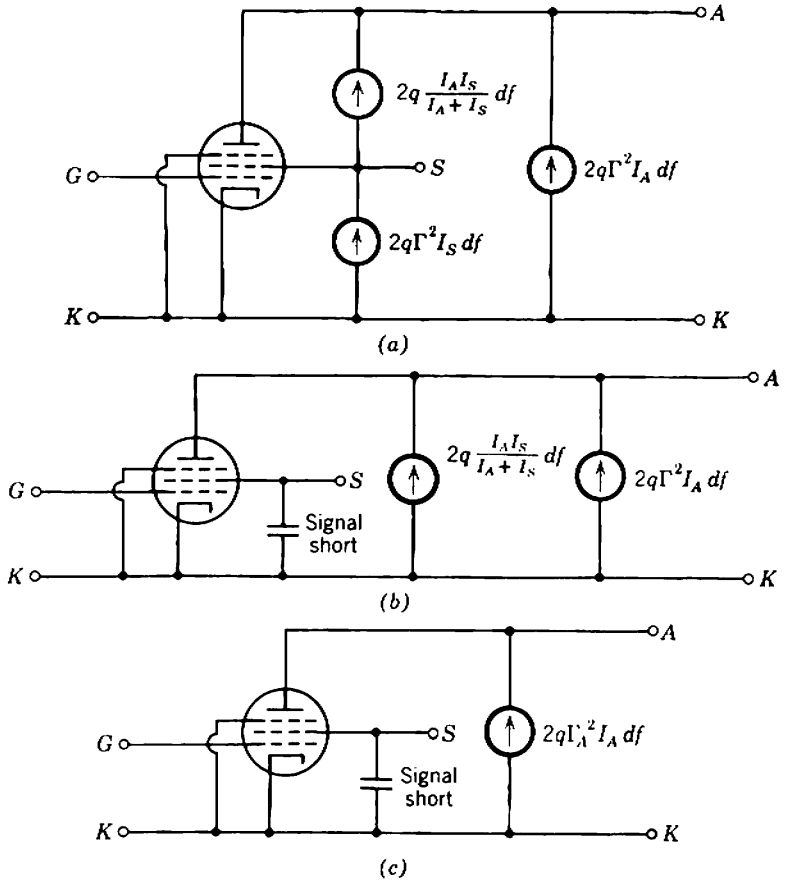


Fig. 3.15 Shot-noise model of a pentode.

Flicker noise is present on the cathode current stream of a pentode, and in addition is generated by the partition process. The partition component of flicker is relatively smaller than the partition component of shot noise.* In the general case flicker noise can be represented by three current generators situated as shown in Fig. 3.16*a*. For the special case in which the screen is bypassed to the cathode, so that no screen noise voltage can be developed, the anode noise current is reduced and the model simplifies to Fig. 3.16*b*. It is convenient to combine the two anode flicker noise generators (Fig. 3.16*c*) and write

$$d(i_{NF}^2)_A = K_A \frac{I_A^c}{f}\,df, \tag{3.68}$$

* T. B. TOMLINSON, "Partition components of flicker noise," *J. Brit. Inst. Radio Engrs.*, **14**, 515, November 1954.

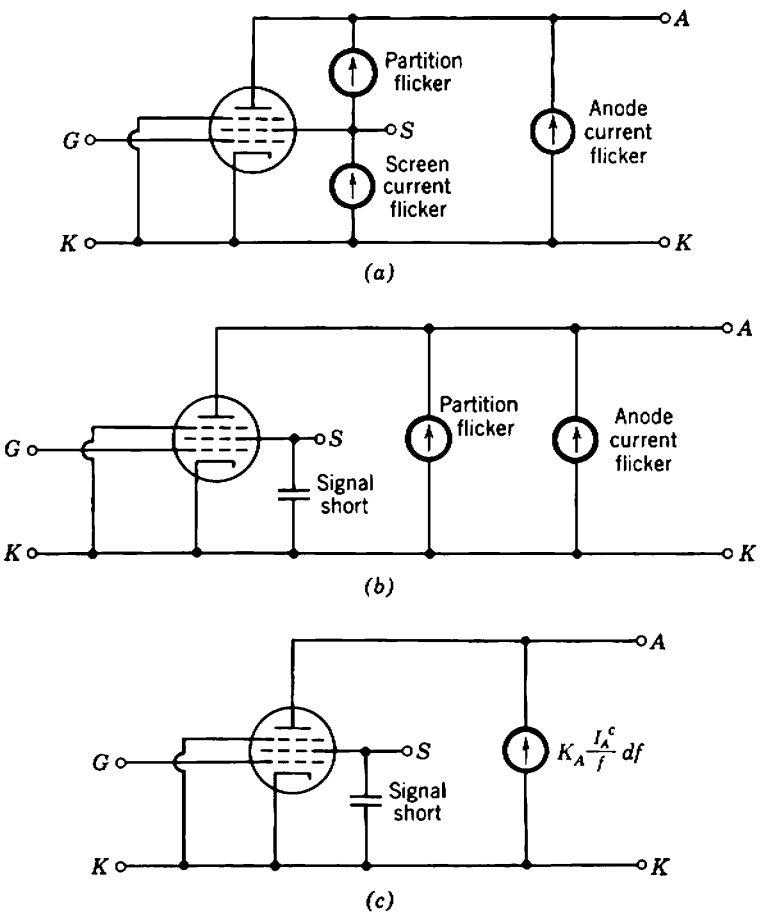


Fig. 3.16 Flicker-noise model of a pentode.

where the effective flicker noise constant K_A for the anode must be determined experimentally.

Induced grid noise in pentodes occurs in precisely the same way as in triodes and may be represented by a mean-square noise current generator $d(i_{NI}{}^2)_G$ across the input terminals.

3.2.3 General Comments on Tube Noise

Shot noise and induced grid noise are well-understood physical phenomena. Unless a vacuum tube has some gross defect, its shot and induced grid noise can be predicted accurately from the equations given. Moreover, the shot and induced grid noise do not vary more than a few percent

from tube to tube of a given type. Flicker noise, however, is not a well-understood phenomenon. It is known that flicker noise on the cathode current stream is caused by minute imperfections in the cathode, but these imperfections are not manifested in any other way. The flicker noise equation (2.65), is essentially empirical, and the constant K may vary by several orders of magnitude between tubes of one type. Some tubes show excessive flicker noise which does not follow Eq. 2.65. In low-noise applications, particularly at audio frequencies, it may be necessary to select tubes to achieve satisfactory flicker-noise performance. So-called low-noise tubes are available commercially; on the average, these tubes have less flicker noise than other types, but isolated specimens of ordinary tubes may have the least noise of all.

In addition to flicker noise on the cathode current stream, imperfections in vacuum tubes cause *grid current noise*. All components of dc grid current contribute additive components of shot (and possibly flicker) noise; likely sources of grid current are listed in Section 3.1.2.4. Like flicker noise on the cathode current stream, grid current noise varies widely from tube to tube and is amenable only to empirical treatment. Nevertheless, there are a few common-sense precautions. Grid current noise caused by electrons striking the grid can be minimized by suitable choice of quiescent point; the grid should be appreciably negative with respect to the barrier potential. Gas noise is a function of the tube and can be reduced by selection or by operating a tube continuously for a long period of time. However, brief periods of overheating, or switching the heater power off and on, may release gas adsorbed inside a tube and change a quiet tube into a noisy one. Noise due to thermionic emission from the grid tends to be less in low-g_m tubes because of their large cathode-grid spacing. Often some improvement can be achieved in high-g_m tubes by lowering the heater voltage. Noise due to interelectrode leakage currents is a function of the tube and its socket; a high-quality tube socket is mandatory in low-noise applications.

There are two distinct steps in designing a low-noise vacuum-tube amplifier. The first, which forms the subject matter of Chapter 8, is to choose quiescent points and other circuit details that minimize the well-understood shot and induced grid noise. The second is to use tubes for which the flicker noise and grid current noise are known to be small. For these quiet tubes Eqs. 3.61 and 3.68 give reasonable estimates of the flicker noise, and grid current noise may be represented over a bandwidth df by a generator

$$d(i_{NS}^2)_G = 2q \sum (I_G)\, df, \tag{3.69}$$

connected in shunt with $d(i_{NI}^2)_G$. The summation sign $\sum$ indicates that the

magnitudes of the individual components of I_G must be added. There is no simple way of predicting the noise in an excessively noisy tube, nor is there any point in doing so.

3.3 RATINGS OF VACUUM TUBES

From a practical point of view the most significant rating of a vacuum tube is its maximum safe *anode power dissipation* P_A. Unless the circuit configuration is unusual, no other rating is likely to be exceeded if P_A remains within safe bounds. The value of $P_{A(\max)}$ is usually limited by the associated rise in temperature of the envelope and internal insulating supports. The highest envelope temperature occurs at the point closest to the anode and for glass tubes should not exceed about 250°C. Above this temperature the glass between electrode lead wires or base pins begins to conduct electrolytically; the result is excessive noise and rapid failure. Unless the contrary is stated, a manufacturer's quoted value of $P_{A(\max)}$ applies for 25°C ambient. If a tube is operated in an elevated ambient temperature (for example, because of heating by adjacent tubes or components), the safe value of P_A is reduced. There is no simple relation between ambient temperature and the derated $P_{A(\max)}$; Fig. 3.17 is one

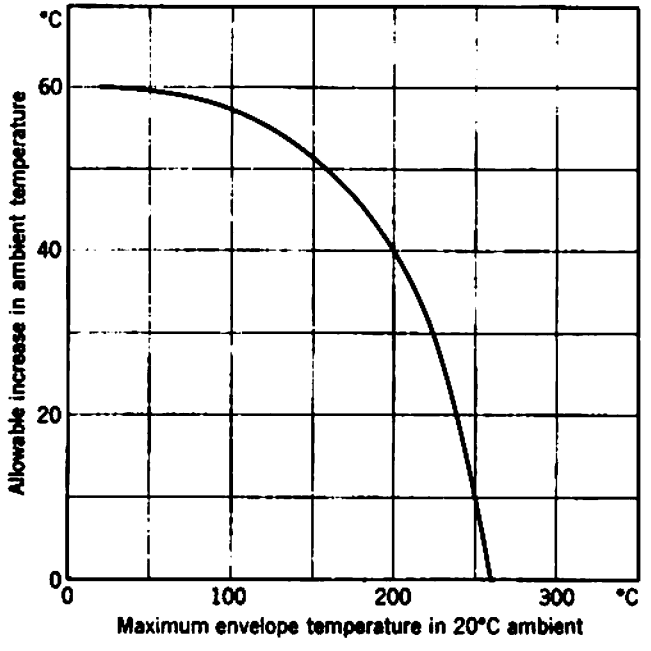


Fig. 3.17 Anode power derating curve. (Courtesy Mullard Ltd., London.)

manufacturer's curve for allowed increase in ambient temperature versus the maximum envelope temperature in a 20°C ambient.

The maximum safe *anode voltage* V_A is set by breakdown of the insulating supports between the anode and other electrodes. Because breakdown voltage is a function of temperature, $V_{A(\max)}$ depends on the anode power dissipation and ambient temperature. Normally, a manufacturer's quoted value of $V_{A(\max)}$ applies for a tube operating at full rated power in a 25°C ambient; higher voltages are permissible at reduced power dissipation. Figure 3.18 is one manufacturer's voltage-uprating curve for tubes operating at a fraction of their safe anode dissipation in a particular ambient temperature. The *anode current* I_A is limited ultimately by the steady emission current available without deterioration of the cathode. However, in most circuits with normal supply voltages the rated anode power dissipation will be exceeded long before the current reaches $I_{A(\max)}$.

The maximum *screen power dissipation* P_S is set by the allowable temperature rise in the screen grid wires and their supports. The *screen voltage* V_S is limited by breakdown of the insulation and depends on the temperature. For most purposes the total device dissipation may be taken as the anode power, and the screen and anode voltage uprating factors are the same. The screen power need not be derated in an elevated ambient temperature because the screen is totally enclosed within the anode. The *screen current* I_S is nearly always limited by the safe screen power dissipation rather than by the safe emission from the cathode.

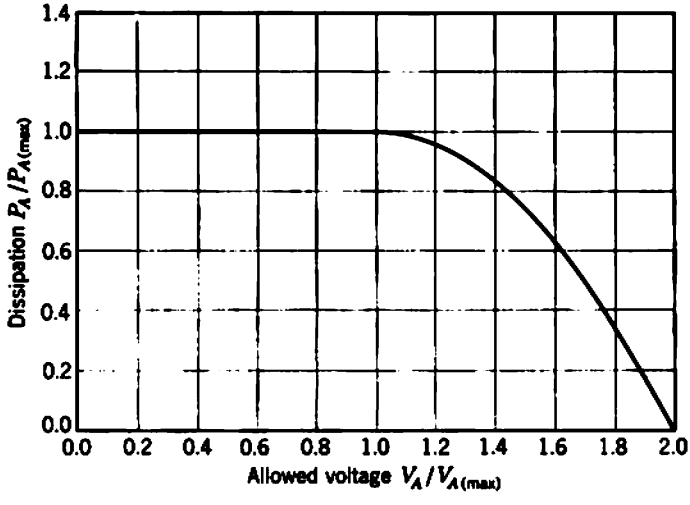


Fig. 3.18 Anode voltage uprating curve. (Courtesy Mullard Ltd., London.)

The maximum *grid power dissipation* P_G is set by the allowable temperature rise in the fine grid wires. Manufacturers rarely quote a safe value for $P_{G(max)}$, but 100 mW is reasonable for small tubes. Often a minimum negative grid voltage is specified for amplifying tubes. This is not a true rating; rather, it is based on the onset of grid current as the grid becomes positive with respect to the barrier potential. European manufacturers commonly specify the voltage corresponding to $+0.3$ μA grid current. In addition, a maximum negative grid voltage may be specified, set by breakdown of the insulation. This rating has no application to amplifiers, for it implies cutoff of the cathode current.

The maximum *cathode-to-heater voltage* V_{KH} is set by breakdown of the insulation between these two elements. In the absence of specific information 100 V is a safe rating for most tubes. However, induction of mains-frequency hum from the heater onto the cathode increases at high values of V_{KH}; therefore, quite apart from considerations of insulation breakdown, it is seldom desirable to operate tubes with V_{KH} large. Some manufacturers give different ratings for V_{KH} positive and negative and instantaneous as well as steady values.

Manufacturer's ratings assume that the heater voltage is held to about $\pm 15\%$ of its nominal value, that the ambient pressure is not less than 50 mm of mercury (corresponding to 60,000 ft altitude), and that the relative humidity is not greater than 80%.

It is important to note that the over-all thermal time constant for a small glass tube is of the order of five minutes. The thermal time constants of individual electrodes—the anode, screen, and grid—are shorter but not less than some seconds. Therefore, unless specific information is given to the contrary, it is valid to assume that absolute power ratings refer to average power in applications involving signals of frequency greater than a few hertz, not peak power. Absolute voltage ratings, however, refer to peak values.

3.4 VARIATION OF VACUUM-TUBE SMALL-SIGNAL PARAMETERS

The values of parameters in the small-signal equivalent circuits of triodes (Fig. 3.9) and pentodes (Fig. 3.13) depend mainly on the type of tube and the chosen quiescent values of electrode currents and voltages. However, because of tolerances in manufacturing processes and aging of the cathode, tubes bearing the same type number may show significant unit-to-unit parameter variations. Thus the circuit designer requires a knowledge of the probable magnitude of parameter variations with

changes in quiescent currents and voltages and with tube replacement. Some of this information can be deduced from the theoretical treatment of vacuum tubes in Section 3.1; the rest is provided by manufacturer's data, acceptance specifications of device users, measurements of actual tube parameters, and experience. The aim of this section is to use these sources to establish design data that indicate the general trends and magnitudes of small-signal parameter variations.

3.4.1 Nominal Small-Signal Parameters of Representative Vacuum Tubes

Vacuum tubes may be classified initially according to their maximum permissible anode dissipation. Directly related to this classification is their physical size; small tubes can be safely dissipate 1 or 2 W, whereas the largest tubes normally encountered (i.e., excluding transmitting types) can dissipate about 30 W. Tubes may be classified further according to their nominal g_m at a particular cathode current; those with large values of g_m are usually referred to as *high slope* types. In addition, triodes may be classified according to their voltage amplification factor μ; pentodes have values of μ which are so large that this classification becomes meaningless. Typical parameters for small tubes are

triodes and pentodes	
normal slope	$g_m = 2$ mA/V
high slope	$g_m = 5$ mA/V
very high slope	$g_m = 10$ mA/V or more
triodes only	
low μ	$\mu = 5$
medium μ	$\mu = 20$
high μ	$\mu = 80$

It is interesting that g_m for tubes is of the order of a few milliamperes per volt, in contrast with bipolar transistors which typically have g_m some 10 times greater (see Section 4.6.1). The restriction of g_m to such small values in vacuum tubes is due basically to the large separation between the controlled charge (in the cathode-grid space) and controlling charge (mainly on the grid). This separation of the charges results in a low input capacitance. A large voltage increment must therefore be applied to the grid to change the space charge and anode current. Formally, the small value of g_m can be ascribed to a limitation on the maximum value of the perveance. Mutual conductance depends on the perveance (Eq. 3.36 or 3.52), and from Eq. 3.34 the maximum value of perveance is limited by the maximum attainable cathode area and the minimum attainable cathode-grid spacing.

Power tubes have larger electrode areas, and consequently larger values of g_m, than small tubes. As a rule of thumb, typical values of g_m increase as the square root of rated anode dissipation. Power triodes are obsolescent; when a power triode is required, it is often necessary to connect a pentode as a triode by strapping its screen and anode together. Power triodes and triode-connected pentodes tend to have low values of μ.

3.4.1.1 Remote Cutoff Tubes

The tubes considered so far are known collectively as *sharp cutoff* types. Their grid is a uniform structure, which consists of a number of equally spaced turns of wire. Such tubes satisfy the condition

$$V_G \approx \frac{-V_A}{\mu}$$

when the entire cathode current is cut off. In other tubes known as *remote cutoff* types the grid wires are unequally spaced and sometimes have different diameters. The result is effectively a number of tubes in parallel, each having a different value of μ; an alternative name for this type of tube is *variable* μ. Figure 3.19 shows the I_A–V_G characteristic of a remote cutoff tube that is a composite of three different sharp cutoff types; g_m is given by the slope of the I_A–V_G characteristic at constant V_A and the g_m–V_G characteristic has a relatively low gradient in the vicinity of cutoff. Consequently, large signal voltages can be applied to the grid of a remote cutoff tube without excessive distortion in the anode current waveform, even if the tube is biased near cutoff. Remote cutoff tubes are commonly used in amplifiers that have *automatic gain control*, but such applications are beyond the scope of this book.

3.4.2 Variation of Small-Signal Parameters with Quiescent Conditions

The first-order trends in the variation of small-signal parameters with quiescent voltages and current can be predicted from the theory in Sections 3.1.2 and 3.1.3. For both triodes and pentodes

$$g_m \propto I_K^{1/2}, \tag{3.70}$$

$$r_A \propto I_K^{-1/2}. \tag{3.71}$$

The capacitances, amplification factors μ and μ_S, and, for a pentode, the ratio k of screen to anode current are all constants. This first-order theory is adequate for normal design purposes.

Second-order theory shows that small departures from these laws are expected because of changes in the barrier plane. At moderately high currents the barrier plane remains fairly constant and the ideal laws are

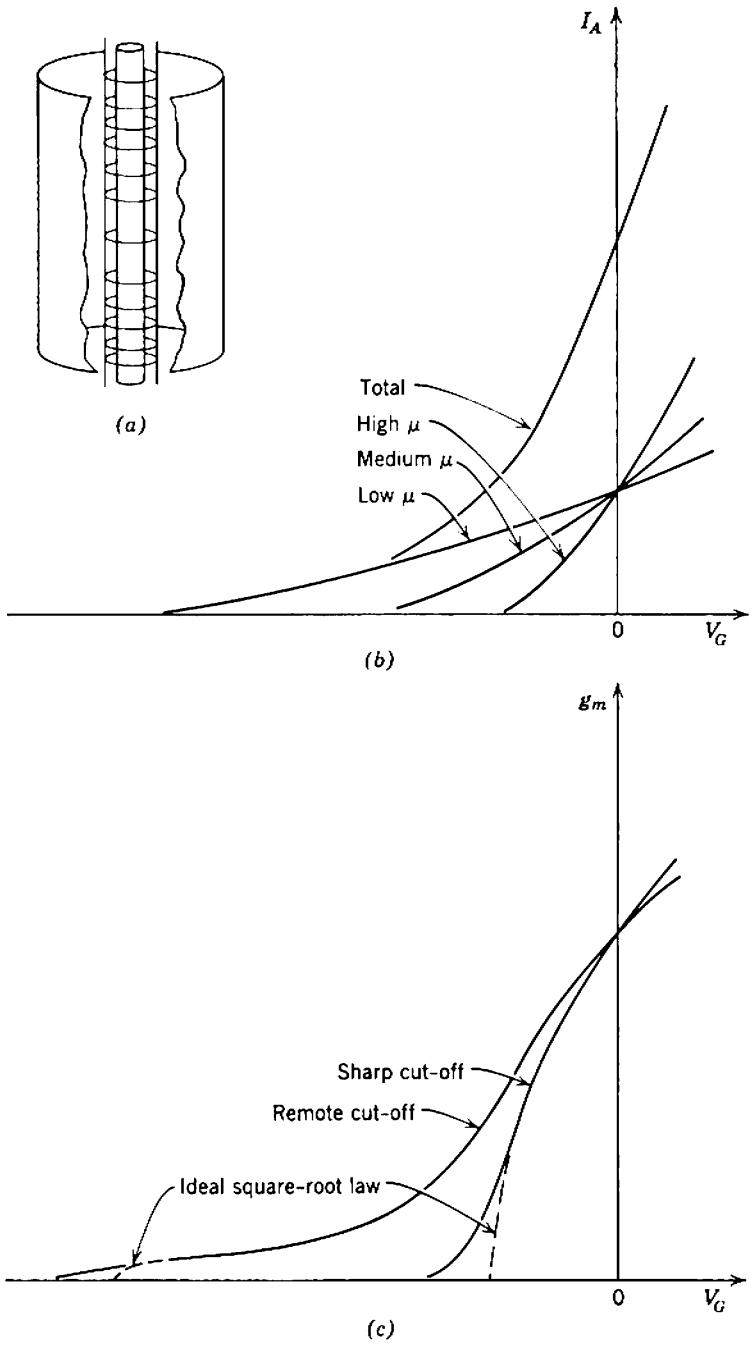


Fig. 3.19 Remote cutoff tubes: (*a*) grid structure; (*b*) I_A–V_G for three-section tube; (*c*) g_m–V_G.

obeyed quite closely. However, the bound space charge increases at low currents and the barrier potential rises. It follows that μ and μ_S will fall, that g_m will vary rather more rapidly than the predicted cube-root law, and that r_A will vary less rapidly. Figure 3.20 shows the measured and predicted curves for a typical triode, type ECC83/12AX7. The capacitances vary more or less linearly over the entire range of cathode current, but at most the variation is only one or two pF. Therefore the intrinsic capacitances of a tube can be lumped with the extrinsic interlead capacitances and the total regarded as constant. Tube manufacturers specify only these total capacitances—c_{in} and c_{out} for the total grid-cathode and anode-cathode capacitances, respectively, and c_{AG} for the extrinsic anode-grid capacitance; there is no simple way of separating the intrinsic and

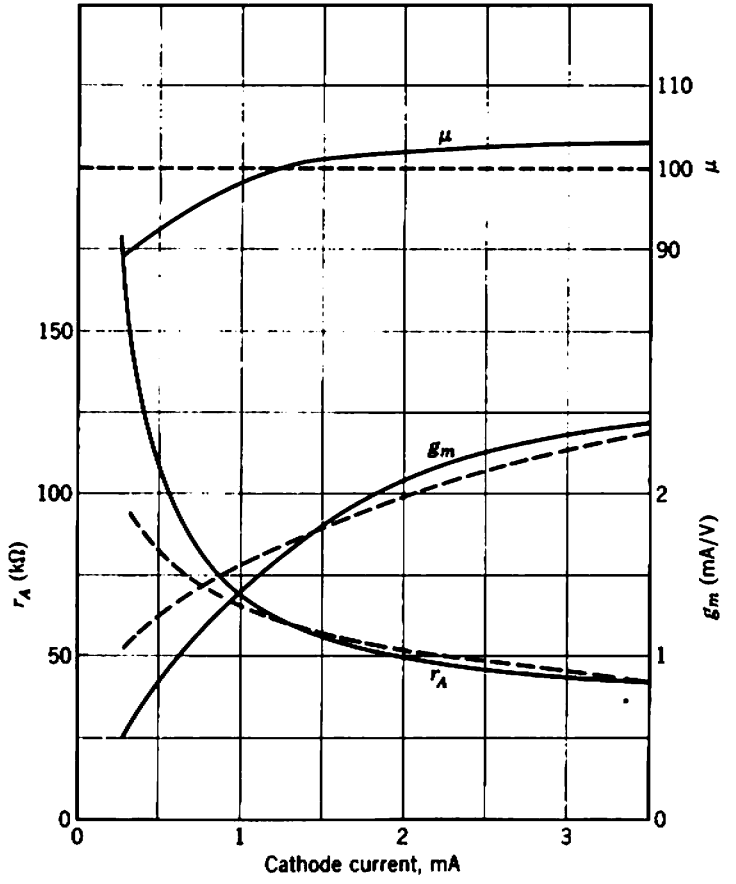


Fig. 3.20 Variation of small-signal parameters with cathode current. Tube type ECC83/12AX7, V_A = 250 V: --- predicted; —— measured. (Courtesy Mullard Ltd., London.)

extrinsic components. This contrasts with the bipolar transistor for which the intrinsic and extrinsic capacitances have quite different laws of variation with quiescent point.

In addition to their variation with cathode current, the small-signal parameters of a tube depend to a small extent on the combination of electrode voltages used to produce a given current. At constant cathode current g_m, μ, μ_S, and k fall slightly with increasing V_A or V_S.

Moderate changes in the heater voltage of a tube operating at constant cathode current have very little effect on the small-signal parameters, for the emission current available from an oxide-coated cathode is far in excess of normal requirements. However, a change in grid voltage will be required to maintain the cathode current constant.

3.4.2.1 Special Considerations for Pentodes

A problem of some practical importance for pentodes is the effect of a change in electrode potentials (particularly V_S) on known values of small-signal parameters (particularly g_m). Usually the small-signal parameters are known from manufacturer's data or measurements over a range of grid voltage only at one screen voltage V_{S1}, and often it is necessary to find g_m at some particular voltages V_{G2} and V_{S2}. The general formal solution involves first finding I_K at the desired quiescent voltages, a laborious process involving Eq. 3.45 and the known data, and then calculating g_m from Eq. 3.70. The labor in the general solution can be reduced by using a *voltage conversion factor* F_V. In this method the grid voltage V_{G1} which satisfies

$$\frac{V_{G2}}{V_{G1}} = \frac{V_{S2}}{V_{S1}} = F_V \tag{3.72}$$

is found. Then from Eq. 3.45 (neglecting the V_A term)

$$\frac{I_{K2}}{I_{K1}} = F_V^{3/2},$$

and, using Eq. 3.70,

$$\frac{g_{m2}}{g_{m1}} = F_V^{1/2}.$$

But g_{m1}, the value of g_m at V_{G1} and V_{S1}, is known from the data. Therefore the desired value g_{m2} at V_{G2} and V_{S2} is

$$g_{m2} = g_{m1}F_V^{1/2} = g_{m1}\left(\frac{V_{S2}}{V_{S1}}\right)^{1/2}. \tag{3.73}$$

When a pentode is connected as a triode by strapping the anode and

screen together, the anode resistance $r_{A(\text{triode})}$ is given in terms of the pentode parameters by

$$\frac{1}{r_{A(\text{triode})}} = \frac{\Delta I_A + \Delta I_S}{\Delta V_S}$$

$$= (1 + k)\frac{\Delta I_A}{\Delta V_S}$$

$$= (1 + k)\frac{g_m}{\mu_S}.$$

Therefore

$$r_{A(\text{triode})} = \frac{\mu_S}{(1 + k)g_m}. \tag{3.74}$$

3.4.3. Variation of Small-Signal Parameters with Tube Replacement

The parameters of tubes of a given type operating under identical quiescent conditions show unit-to-unit variations between individual tubes manufactured in one batch, between average tubes from different production batches, and most important of all, with tube age.

Variations in the parameters of new tubes stem from the inherent inaccuracies associated with production techniques, principally in

(i) electrode geometry, particularly the spacing of grid and screen wires,
(ii) resistance of heater elements,
(iii) work function of the cathode material,
(iv) concentration of positive gas ions in the tube envelope.

The spread in parameters varies with the type of tube and the acceptance specification set by the manufacturer or user. The user is seldom able to distinguish between intra- and interbatch variations and must therefore assume all-inclusive tolerances. The following parameter spreads are typical for new receiving tubes tested at fixed values of cathode current, anode voltage, and screen voltage:

$\pm 25\%$ for mutual conductances and self resistances,
$\pm 10\%$ for amplification factors and inter-electrode capacitances.

In addition to these manufacturing tolerances on the parameters of new tubes, the circuit designer must consider the effect of aging. Oxide-coated cathodes exhibit a gradual reduction in emission capabilities throughout their lives, with a consequent reduction in g_m. Deterioration of the cathode is due mainly to bombardment by positive gas ions and to evaporation of the coating. If the cathode is heated, deterioration occurs whether or not a cathode current is flowing, although the mechanism

differs. The percentage rate of reduction in g_m depends on the tube type and its time of operation; for standard receiving tubes a 20% reduction is likely over the first 1000 hours of their lives.

Prager* has produced curves that relate the survival rate of *premium quality tubes* to age for various allowed percentage changes in g_m. Obviously, the greater the allowed percentage change, the greater the survival rate over a given period. The typical curves reproduced in Fig. 3.21 indicate that 80% survival can be achieved after 9000 hours if a 40% change in g_m is acceptable. However, if the acceptable change in g_m is reduced to 10%, 80% survival corresponds to only 1000 hours of operation. Consequently, to achieve a long useful life for tubes in electronic equipment it is highly desirable to allow for changes in g_m up to about 40%. Standard tube types have much shorter lives than premium quality types.

In this book it is assumed arbitrarily that the limits set out in Table 3.1 account for both the production tolerances on the major parameters and deterioration with age. These limits apply at constant values of cathode current, anode voltage, and screen voltage. The quantities μ_A, μ_{SA}, g_{mA}, r_{AA}, and k_A are the values of the parameters μ, μ_S, g_m, r_A, and k expected for "average" vacuum tubes of a given type and are the values normally listed in manufacturers' data. New vacuum tubes tend to show values of g_m that lie above the average and values of r_A that lie below. As a vacuum tube approaches the end of its useful life, g_m falls below average and r_A

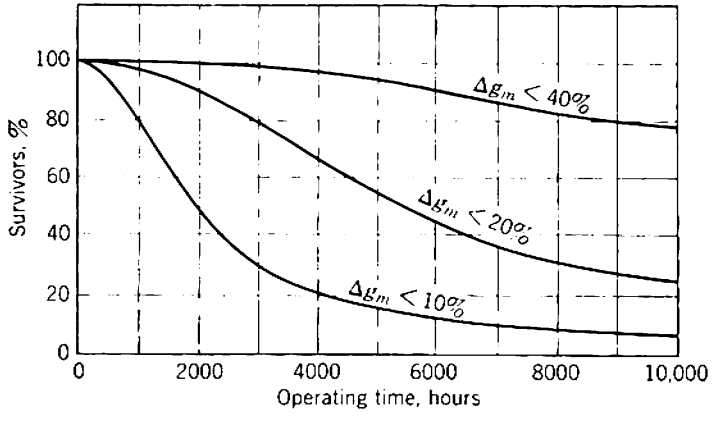


Fig. 3.21 Survival rate for premium quality tube type 5692 as a function of allowed change in g_m. (Courtesy Radio Corporation of America.)

* H. J. PRAGER, "Performance evaluation of 'special red' tubes," *RCA Rev.*, **14** 413, September 1953.

Table 3.1 Tolerances on Vacuum Tube Parameters, Referred to Their Average Values

Parameter	Tolerance
Triode	
Mutual conductance g_m	$0.6\, g_{mA} \le g_m \le 1.5\, g_{mA}$
Voltage amplification factor μ	$0.9\, \mu_A \le \mu \le 1.1\, \mu_A$
Anode resistance r_A	$0.6\, r_{AA} \le r_A \le 1.5\, r_{AA}$
Pentode	
Mutual conductance g_m	$0.6\, g_{mA} \le g_m \le 1.5\, g_{mA}$
Voltage amplification factor μ	Unspecified, but large
Screen amplification factor μ_S	$0.8\, \mu_{SA} \le \mu_S \le 1.25\, \mu_{SA}$
Anode resistance r_A	Unspecified, but large
$k = I_S/I_A$	$0.8\, k_A \le k \le 1.25\, k_A$

rises. The values of μ, μ_S, and k remain fairly constant. These tolerances may be reviewed in the light of information on specific tube types, but it is unlikely that modifications will be significant. Data on parameter tolerances are available from manufacturers and certain specialized user organizations.

3.5 CONCLUDING COMMENTS ON VACUUM TUBES

Inspection of vacuum-tube manufacturers' data books shows that there are thousands of types bearing different identifying numbers. However, the tolerances on vacuum tube parameters are so large that for design purposes many supposedly different types are electrically indistinguishable.

Consider, for example, the classification of pentodes into normal, high, and very high slope types with nominal g_m of 2, 5, and 10 mA/V, respectively. If the tolerances suggested by Table 3.1 are applied, the expected ranges of g_m will be as shown in Table 3.2; for only three basic pentode

Table 3.2 Range of g_m at Constant Cathode Current for Typical Pentodes

Type	Nominal g_m	Range of g_m
Normal	2 mA/V	1.2–3 mA/V
High slope	5	3–7.5
Very high slope	10	6–15

types operating at constant current is there likely to be a continuous spread in g_m from 1.2 to 15 mA/V. On this basis the inclusion of further basic types with intermediate values of g_{mA} becomes quite unnecessary. The majority of vacuum-tube types at present in production are redundant.

There are good reasons for part of this redundancy. Some of the many type numbers belong to different groupings of a few basic electrode structures inside multiple-unit tubes. For example, the triode sections of the following types are nominally identical:

EBC91/6AV6	double-diode triode
ECC83/12AX7	twin triode
ECL86/6GW8	triode power-pentode.

Other vacuum-tube types are made with a number of different heater voltages (though the same power) for use in different environments. An example is a certain power pentode:

EL84/6BQ5	6.3 V at 750 mA for parallel operation
PL84	15 V at 300 mA for series operation
UL84/45B5	45 V at 100 mA for series operation.

Nevertheless, most of the vacuum-tube types in production are redundant; they differ only in minor electrical detail or in the connections to the base pins.

Chapter 4

Characteristics of Amplifying Devices III: Bipolar and Field-Effect Transistors

This chapter is similar in scope to Chapter 3; the general principles of Chapter 2 are applied to the particular cases of bipolar and field-effect (unipolar) transistors.

The theory of the bipolar junction transistor is readily developed from semiconductor physics. The principles of carrier statistics and carrier ballistics are summarized in Section 4.1, and from them the basic charge-control parameters Q_C and τ_1 are found. Relations between the dc terminal voltages and currents follow immediately, and the parameters of a small-signal equivalent circuit follow from substitution into the general theory in Section 2.5. The accuracy of the charge-control approach is assessed by a comparison with results obtained from more accurate models of a transistor.

The theory of the field-effect transistor is similarly developed from device physics. However, the f.e.t. is at present a less versatile amplifying device than the bipolar transistor and does not warrant so extensive a treatment. For the same reason it is assumed in accord with common usage that the term *transistor* applies to the bipolar device; field-effect transistors are always referred to explicitly.

4.1 A REVIEW OF SEMICONDUCTOR PHYSICS*

The important materials in semiconductor amplifying devices are the tetravalent elements germanium and silicon. Pure or *intrinsic* crystals of these elements are doped with controlled amounts of trivalent *acceptor* or pentavalent *donor* atoms to produce mechanically stable, near-perfect *extrinsic* crystals that have an excess of holes or electrons. Typical impurity concentrations for the extrinsic germanium used in transistors range from 10^{14} to 10^{19} atoms/cm^3. Because crystalline germanium contains about 4×10^{22} atoms/cm^3, the proportion of impurity atoms ranges from about 2 parts in 10^9 to 2 parts in 10^4.

Both holes and electrons exist in any semiconducting crystal and are called *carriers* of charge. The concentrations are equal in intrinsic materials, whereas holes or electrons predominate in *p*-type or *n*-type material respectively. *Majority carriers* are those in excess; the others are *minority carriers*. The concentrations of carriers in a crystal are determined by carrier statistics, and the currents that result from carrier motion are determined by carrier ballistics.

4.1.1 Carrier Statistics

Under conditions of energy equilibrium the hole concentration p and electron concentration n at any point in a crystal are related to the concentrations of donor impurities N_D and acceptor impurities N_A by two general equations:

$$(N_D + p) - (N_A + n) = 0, \tag{4.1}$$

$$pn = n_i^2 = \nu T^3 \exp\left(-\frac{E_{gap}}{kT}\right), \tag{4.2}$$

where n_i = the concentration of hole-electron pairs in intrinsic material at the same temperature, often referred to as the *intrinsic concentration*,

E_{gap} = the energy difference between the valence and conduction bands,

* This review is intended only to introduce the notation and to indicate the general level of the treatment that follows. Readers should refer to specialized books for a more ordered development, e.g: R. B. ADLER, A. C. SMITH, and R. L. LONGINI, *An Introduction to Semiconductor Physics*, S.E.E.C., Vol. 1 (Wiley, New York, 1964). P. E. GRAY, D. DE WITT, A. R. BOOTHROYD, and J. F. GIBBONS, *Physical Electronics and Circuit Models of Transistors*, S.E.E.C., Vol. 2 (Wiley, New York, 1964). R. H. MATTSON, *Basic Junction Devices and Circuits* (Wiley, New York, 1963). J. L. MOLL, *Physics of Semiconductors* (McGraw-Hill, New York, 1964). R. P. NANAVATI, *Introduction to Semiconductor Electronics* (McGraw-Hill, New York, 1963).

k = Boltzmann's constant,
T = the absolute temperature,
ν = a constant.

Equation 4.1 is merely a statement of the principle of conservation of charge, whereas Eq. 4.2 is known as the *equilibrium equation* and is based on Fermi-Dirac statistics.

For heavily doped p-type material N_A is much greater than N_D. It follows from Eq. 4.1 that p is much greater than n and that

$$p \approx N_A, \tag{4.3}$$

which shows that the equilibrium majority carrier concentration is largely determined by the doping and is almost independent of temperature. Substitution in Eq. 4.2 gives

$$n = \frac{n_i^2}{p} \approx \frac{n_i^2}{N_A}, \tag{4.4}$$

which shows that the minority carrier concentration varies as n_i^2. At room temperature the exponential term in Eq. 4.2 varies much more rapidly than the T^3 term, and the total variation of n_i^2 is approximately exponential with temperature. Expressions corresponding to Eqs. 4.3 and 4.4 apply for heavily doped n-type material.

The equilibrium condition can be disturbed by some energy stimulus. Examples are in increase in temperature, irradiation with light or other electromagnetic radiation, irradiation with nuclear radiation, and injection or extraction of carriers (usually by means of a p-n junction). As long as the stimulus is maintained the carrier concentrations remain in excess of their equilibrium values, but once the stimulus is removed the excess holes and electrons recombine and the concentrations decay with time to their equilibrium values. Recombination occurs at different rates in the body of a crystal and at its surface, and the decay of carrier concentrations can be expressed as a sum of exponentials. However, it is a reasonable engineering approximation to represent the decay by a single exponential whose time constant τ is the *effective lifetime* of carriers in the crystal. The value of τ varies widely between otherwise similar crystals.

In transistor electronics the most important way of disturbing the carrier equilibrium is by means of p-n junctions on one or more faces of the crystal. If a p-n junction is formed as shown in Fig. 4.1 and a voltage V is applied to it externally, an opposing voltage drop $(V_i - V)$ exists across the transition between the two regions, or *depletion layer*. The depletion layer, as its name implies, is substantially depleted of mobile charged carriers and is a region of high electric field; only the fixed ionized acceptor

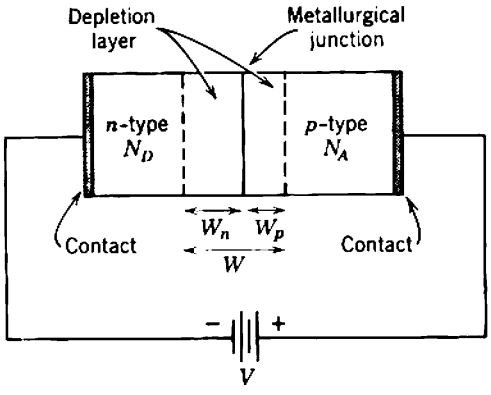


Fig. 4.1 A *p-n* junction.

and donor atoms of the crystal lattice remain. The *contact potential* V_i is related to the equilibrium carrier concentrations:

$$V_i = \frac{kT}{q} \ln\left(\frac{n_{n0}}{n_{p0}}\right) = \frac{kT}{q} \ln\left(\frac{p_{p0}}{p_{n0}}\right), \tag{4.5}$$

where p_{p0} and n_{p0} = the equilibrium concentrations in the *p*-type material,
p_{n0} and n_{n0} = the equilibrium concentrations in the *n*-type material.

(The subscript zero is added to distinguish an equilibrium value from a nonequilibrium value.) Equation 4.5 is derived from a Boltzmann approximation to the Fermi-Dirac distribution function. For V in the range

$$-\infty < V < V_i$$

the minority carrier concentrations at the boundaries of the depletion layer are given by

$$n_p(0) = n_{p0}(0) \exp\left(\frac{qV}{kT}\right), \tag{4.6a}$$

$$p_n(0) = p_{n0}(0) \exp\left(\frac{qV}{kT}\right), \tag{4.6b}$$

where the argument (0) indicates a value at the boundary. These equations suggest the possibility of varying the minority carrier concentrations over several orders of magnitude by means of the applied voltage V. They are known as the *Boltzmann relations* and are extremely important. The nonequilibrium majority carrier concentrations are

$$p_p(x) = N_A(x) + n_p(x), \tag{4.7a}$$

$$n_n(x) = N_D(x) + p_n(x). \tag{4.7b}$$

Equations 4.7 are a statement of charge neutrality and apply throughout the regions rather than at the boundaries alone.

An important consequence of changing the voltage applied to a *p-n* junction is that the depletion layer width (W in Fig. 4.1) varies according to the law

$$W = W_0(V_t - V)^a. \tag{4.8}$$

The constants a and W_0 depend on the junction structure. The exponent a varies from 1/2 (for an abrupt change from *p*-type to *n*-type material) to 1/3 (for a linear change). For an abrupt planar junction the *width constant* W_0 is

$$W_0 \approx \left[\frac{2\varepsilon}{q}\left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{1/2}, \tag{4.9}$$

where N_A and N_D are the doping densities of the *p*- and *n*-regions, respectively, and ε is the permittivity. It also follows that the penetrations into the *p*- and *n*-regions are

$$W_p = \frac{W}{1 + N_A/N_D}, \tag{4.10a}$$

$$W_n = \frac{W}{1 + N_D/N_A}, \tag{4.10b}$$

which shows that the penetration is predominantly into the region of lesser impurity concentration. Because the depletion layer width changes with applied voltage, the numbers of ionized donor and acceptor atoms change. Thus a change in voltage involves a change in charge, and a *transition capacitance* is associated with a *p-n* junction. The capacitance per unit area varies according to the law

$$c_t = c_{t0}(V_t - V)^{-a}, \tag{4.11}$$

where a is the same exponent as in Eq. 4.8, and the *capacitance constant* for an abrupt junction is

$$c_{t0} = \left[\left(\frac{q\varepsilon}{2}\right)\left(\frac{N_A N_D}{N_A + N_D}\right)\right]^{1/2}. \tag{4.12}$$

4.1.2 Carrier Ballistics

Only one carrier type, the electron, is of first-order importance in vacuum tubes. The residual gas in a tube ensures that some positive ions are present, but they are due to technological imperfections and play no useful part in the operation. The electrons move only in the presence of an electric field, and their acceleration is proportional to it. Semiconductor devices are more complex in that there are two carrier types

(holes and electrons) and motion occurs in two ways; carriers *drift* in the presence of an electric field and *diffuse* in the presence of a concentration gradient.

Because of repeated collisions with the fixed atoms of the crystal lattice, the drift velocity of carriers is proportional to the electric field:

$$\mathbf{v}_{h(\text{drift})} = \mu_h \mathcal{E}, \tag{4.13a}$$

$$\mathbf{v}_{e(\text{drift})} = -\mu_e \mathcal{E}, \tag{4.13b}$$

where μ_h and μ_e are the hole and electron *mobilities*. The resulting current densities are proportional to the carrier concentrations:

$$\mathbf{J}_{h(\text{drift})} = q\mu_h p \mathcal{E}, \tag{4.14a}$$

$$\mathbf{J}_{e(\text{drift})} = q\mu_e n \mathcal{E}. \tag{4.14b}$$

The current density due to a concentration gradient is proportional to this gradient but independent of the carrier concentration:

$$\mathbf{J}_{h(\text{diffusion})} = -qD_h \nabla p, \tag{4.15a}$$

$$\mathbf{J}_{e(\text{diffusion})} = qD_e \nabla n, \tag{4.15b}$$

where D_h and D_e are the *diffusion constants*. The corresponding diffusion velocities are

$$\mathbf{v}_{h(\text{diffusion})} = \frac{\mathbf{J}_h}{qp} = -\frac{D_h}{p} \nabla p, \tag{4.16a}$$

$$\mathbf{v}_{e(\text{diffusion})} = -\frac{\mathbf{J}_e}{qn} = -\frac{D_e}{n} \nabla n. \tag{4.16b}$$

Notice that the mobilities and diffusion constants are related by the Einstein equation:

$$\frac{D_h}{\mu_h} = \frac{D_e}{\mu_e} = \frac{kT}{q}. \tag{4.17}$$

Summing, the general equations for carrier motion and current density are

$$\mathbf{v}_h = \mu_h \mathcal{E} - \frac{D_h}{p} \nabla p, \tag{4.18a}$$

$$\mathbf{v}_e = -\mu_e \mathcal{E} - \frac{D_e}{n} \nabla n, \tag{4.18b}$$

$$\mathbf{J}_h = q(\mu_h p \mathcal{E} - D_h \nabla p), \tag{4.19a}$$

$$\mathbf{J}_e = q(\mu_e n \mathcal{E} + D_e \nabla n). \tag{4.19b}$$

The drift component of current density vanishes at points where the carrier concentration falls to zero, but the diffusion component remains finite as long as there is a concentration gradient.

The resistivity ρ of a semiconducting material is the ratio of electric field to total drift component of current density:

$$\rho = \frac{1}{q(\mu_h p + \mu_e n)}. \tag{4.20}$$

For intrinsic material the carrier concentrations are equal to one another and to the intrinsic concentration n_i. Thus

$$\rho_i = \frac{1}{qn_i(\mu_h + \mu_e)} \tag{4.21}$$

and ρ_i decreases (roughly) exponentially with temperature because n_i increases. For highly doped p-type material p is much greater than n and

$$\rho_p = \frac{1}{q\mu_h p} = \frac{1}{q\mu_h N_A}. \tag{4.22a}$$

Similarly, for highly doped n-type material

$$\rho_n = \frac{1}{q\mu_e n} = \frac{1}{q\mu_e N_D}. \tag{4.22b}$$

If it were not for the temperature dependence of mobility, ρ_p and ρ_n would be independent of temperature. Over the useful range of doping densities and temperatures an approximation is

$$\mu \propto T^{-b}, \tag{4.23}$$

Table 4.1 Properties of Germanium and Silicon

Property	Germanium	Silicon	Unit
Atomic weight	72.60	28.08	--
Atomic number	32	14	--
Lattice constant	5.66	5.43	Angstrom units
Relative permittivity (dielectric constant)	15.8	11.7	—
Energy gap E_{gap} at 300°K	0.665	1.12	Electron volts
n_i^2 at 300°K	5.61×10^{26}	1.9×10^{20}	cm^{-6}
Hole mobility μ_h	1900	480	cm^2/volt-sec
Electron mobility μ_e	3900	1350	cm^2/volt-sec
Hole diffusion constant D_h	49	12	cm^2/sec
Electron diffusion constant D_e	102	35	cm^2/sec

where $b \approx 2$, and it follows from Eq. 4.17 that

$$D \propto T^{1-b}. \tag{4.24}$$

At high temperatures ρ_p and ρ_n fall sharply because n_i becomes so large that both holes and electrons contribute to the current; Eq. 4.21 applies.

Table 4.1 lists some physical parameters of germanium and silicon.

4.2 ELEMENTARY THEORY OF THE TRANSISTOR

Junction transistors are manufactured by a number of different processes, but in most cases the resulting structure has the same essential features:

(i) the junction faces are planar and parallel,
(ii) the area of the emitter junction is smaller than the area of the collector junction,
(iii) the separation between the junctions is less than their linear dimensions.

The result is that the carrier trajectories run predominantly parallel to one another but perpendicular to the junction faces, and the plane-parallel representation in Fig. 4.2 is realistic. Because the carrier trajectories run across the base, the effective cross-sectional area of the control region is approximately that of the emitter, namely, A_E. The base lead wire, however, is necessarily attached to a point well away from the control

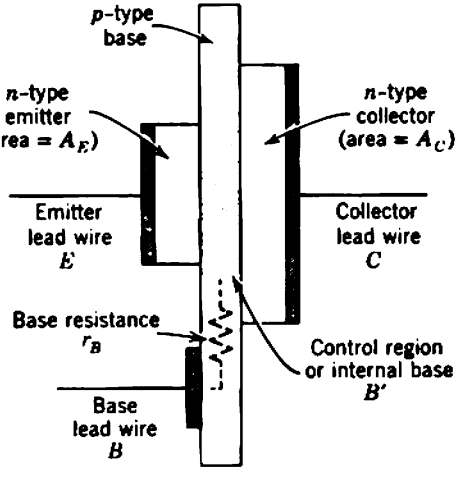


Fig. 4.2 Idealized model of an *n-p-n* transistor.

region, and consequently a *base (spreading) resistance* r_B appears between the base terminal B and the control region, or *internal base* B'.*

Although Fig. 4.2 depicts an n-p-n transistor for which holes are the majority carriers in the base and electrons are the minority carriers, the following theory applies also to p-n-p types with the roles of holes and electrons interchanged.

The emitter and usually the collector regions of a transistor are much more highly doped than the base. This has four important consequences:

1. Because of the finite resistivities of the emitter, base, and collector, the voltages applied to the junctions are not the terminal voltage differences V_{BE} and V_{BC}. However, the resistivities of the emitter and collector are low because of their high doping densities, and the only significant voltage drop is across the base resistance r_B. The junction voltages are therefore given quite accurately by $V_{B'E}$ and $V_{B'C}$. (Exceptions to this statement occur with some obsolescent grown junction transistors and nonepitaxial diffused types which have quite a high collector resistivity, and a resistance of a few hundred ohms appears between the collector lead wire and the collector junction. Exceptions also occur in those integrated circuits in which a low-resistivity collector region is buried inside a silicon chip and the lead wire is attached at some distant point on the surface of the chip. Emitter spreading resistance is discussed in Section 4.7.1.)

2. The depletion layers extend predominantly into the base region, rather than into the emitter and collector. For a transistor with abrupt junctions and constant base doping N_A the emitter and collector depletion-layer widths follow from Eqs. 4.8, 4.9, and 4.10 as

$$W_E = \left[\frac{2\epsilon}{qN_A}(V_i - V_{B'E})\right]^{1/2}, \tag{4.25a}$$

$$W_C = \left[\frac{2\epsilon}{qN_A}(V_i - V_{B'C})\right]^{1/2}. \tag{4.25b}$$

Figure 4.3 shows the relation between the effective base width W_B and the metallurgical base width.

3. When either junction is forward biased, the dominant component of current is the stream of electrons injected into the base against the electric field in the depletion layer. It is a good approximation to neglect all other components at typical operating current densities but not at very high or very low current densities (see Section 4.3.1.2).

* For alternative symbols see Section 4.7.1.

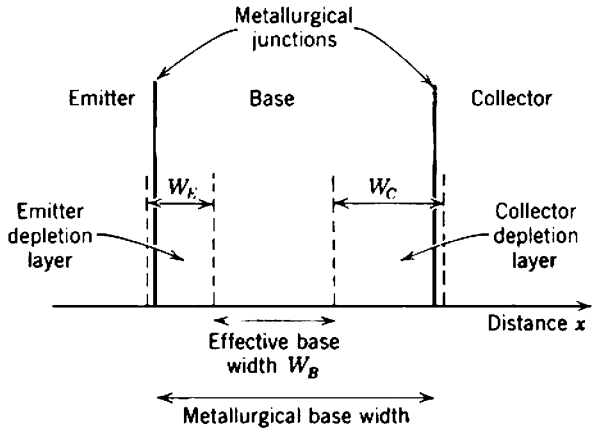


Fig. 4.3 Relations between metallurgical base width, effective base width W_B, and the emitter and collector depletion layer widths W_E and W_C.

4. When either junction is reverse biased, the dominant component of current is the stream of thermally generated electrons swept out of the base by the aiding field in the depletion layer.

In normal operation as an amplifier, the emitter junction of a transistor is forward biased whereas the collector junction is reverse biased. Electrons are injected into the base across the emitter junction and swept out at the collector junction. No other carrier types are able to cross the junctions in significant numbers; therefore current flow is due predominantly to the motion of electrons across the base, and electrons, the minority carriers in the base, must constitute the mobile controlled charge of charge-control theory. If the electron distribution in the base can be found for given junction voltages, the total mobile charge can be evaluated by integration and the collector current I_C will follow from the equations of carrier ballistics. Division of the charge by the current gives the transit time τ_1, and the parameters of a small-signal equivalent circuit can be derived from charge-control theory.

4.2.1 Carrier Distribution in the Base Region

The acceptor concentration is constant throughout the *p*-type base region of a *diffusion* or *uniform-base transistor*, but it varies with distance in *drift* or *graded-base* types. In the latter case the concentration is graded from a high value near the emitter to a low value near the collector, and as a result there is an electric field across the base. Carrier motion is due partly to drift in this electric field, which leads to the name *drift*

transistor. The diffusion transistor can be considered as a special case in which the field vanishes and carrier motion is entirely due to diffusion.

The grading of the acceptor concentration can be discrete or continuous. Discrete grading is used in the obsolescent *n-p-i-n* transistor types, but most modern high-frequency transistors use a continuous grading. The analysis is simplified by assuming that the acceptor concentration varies exponentially with distance; this is a sound engineering approximation because the grading realized with standard production processes is very nearly exponential. The acceptor concentration can be written as

$$N_A(x) = N_A(0) \exp\left(-\frac{mx}{W_B}\right), \tag{4.26}$$

where m is called the *field factor*, being given by

$$m = \ln\left[\frac{N_A(0)}{N_A(W_B)}\right], \tag{4.27}$$

and $N_A(0)$ and $N_A(W_B)$ are the acceptor concentrations at edges of the emitter and collector depletion layers (Fig. 4.4). The value of m is limited by the maximum and minimum attainable values of $N_A(0)$ and $N_A(W_B)$. The maximum value of $N_A(0)$ is restricted to about 2×10^{17} atoms/cm^3 (corresponding to about 0.04 Ω-cm resistivity) because the emitter region must be much more highly doped. The minimum useful value of $N_A(W_B)$ is about 1×10^{14} atoms/cm^3 (corresponding to about 30 Ω-cm), beyond which the germanium is intrinsic. Thus their maximum ratio is about 2000:1, and the maximum value of m is about 8. For the special case of a uniform-base transistor the inbuilt field and m are zero.

Under equilibrium conditions there can be no significant flow of holes in the x-direction because holes cannot cross the collector depletion layer

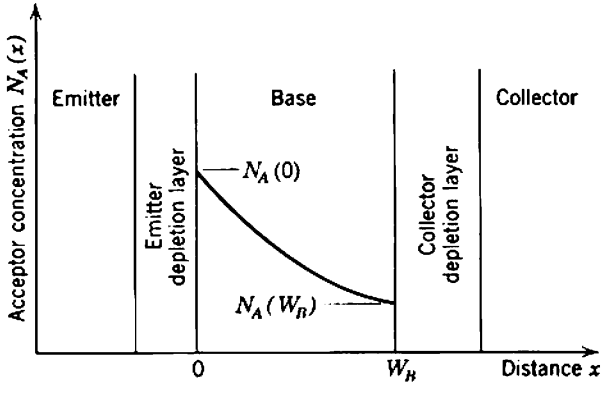


Fig. 4.4 Acceptor concentration in the base of an *n-p-n* transistor.

against the high opposing field. Consequently, the total hole current density is zero, and from Eq. 4.19*a*

$$J_h = q\left(\mu_h p\mathscr{E} - D_h \frac{dp}{dx}\right) = 0.$$

Thus the field that must be built into the base region by grading the acceptor concentration is

$$\mathscr{E} = \left(\frac{D_h}{\mu_h p}\right)\frac{dp}{dx}.$$

Provided the injected electron concentration $n(x)$ is small (implying low-current operation), Eq. 4.7*a* shows that the hole and acceptor concentrations are equal:

$$p(x) \approx N_A(x). \tag{4.28}$$

Substitution from Eqs. 4.17, 4.26, and 4.27 therefore gives

$$\mathscr{E} = -\left(\frac{D_h}{\mu_h}\right)\left(\frac{m}{W_B}\right) = -m\left(\frac{kT}{qW_B}\right). \tag{4.29}$$

Thus the field is constant throughout the base region, and is in the negative x-direction so that it assists electrons from emitter to collector. The magnitude of the field increases with the field factor m.

4.2.1.1 Total Mobile Charge, Q_{tot}

The electron concentrations at the boundaries of the base region are given by the Boltzmann relation (Eq. 4.6). At the edge of the forward-biased emitter junction

$$n(0) = n_0(0)\exp\left(\frac{qV_{B'E}}{kT}\right) \gg n_0(0), \tag{4.30a}$$

where the equilibrium electron concentration at any point follows from Eq. 4.4 as

$$n_0(x) = \frac{n_i^2}{p_0(x)} \approx \frac{n_i^2}{N_A(x)}. \tag{4.31}$$

Similarly, at the edge of the reverse-biased collector junction

$$n(W_B) = n_0(W_B)\exp\left(\frac{qV_{B'C}}{kT}\right) \to 0. \tag{4.30b}$$

From Eq. 4.19*b* the electron current density across the base is given at any point by

$$J_e(x) = q\left\{\mu_e n(x)\mathscr{E} + D_e\frac{d[n(x)]}{dx}\right\},$$

and by substituting from Eqs. 4.17 and 4.29,

$$J_e(x) = qD_e\left\{-\frac{m}{W_B}n(x) + \frac{d[n(x)]}{dx}\right\}. \tag{4.32}$$

By making the approximation that $J_e(x)$ is constant throughout the control region we can integrate Eq. 4.32 with Eq. 4.30*b* as a boundary condition to give the electron concentration $n(x)$:

$$n(x) = n(0)\left[\frac{\exp(m) - \exp(mx/W_B)}{\exp(m) - 1}\right]. \tag{4.33}$$

The assumption of constant J_e implies that recombination is zero, and although this is certainly not true for a transistor the recombination (base) current is very small compared with the collector current. The error in Eq. 4.33 is therefore not more than a few percent for the charge distribution near the collector and is quite negligible near the emitter. Figure 4.5 shows the general form of $n(x)$ for various values of m.

The total electron charge in the base region is

$$Q_{\text{tot}} = -qA_E\int_0^{W_B} n(x)\,dx$$

and substitution from Eq. 4.33 gives

$$Q_{\text{tot}} = -qA_E W_B\, n(0)\left\{\frac{m - 1 + \exp(-m)}{m[1 - \exp(-m)]}\right\}. \tag{4.34}$$

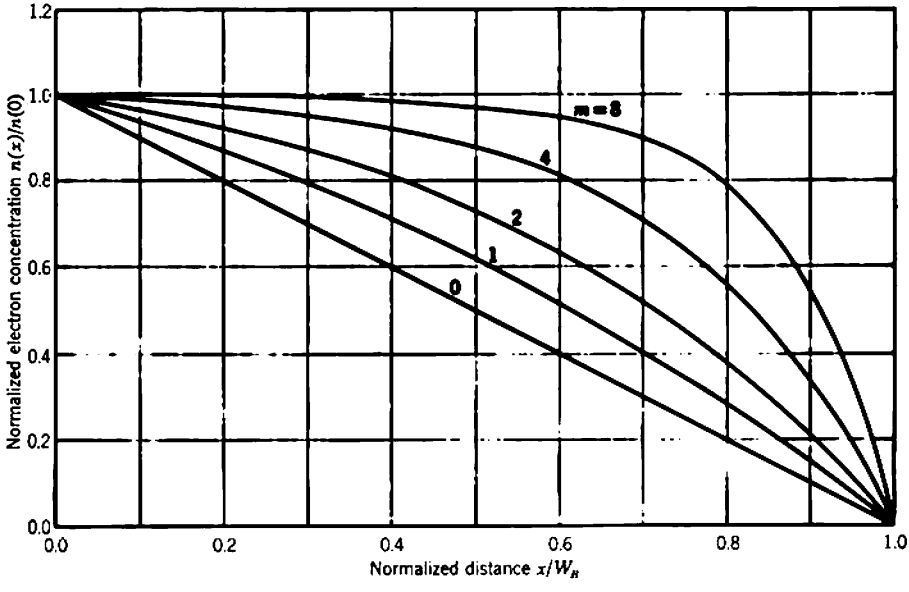


Fig. 4.5 Electron (minority carrier) concentration in the base of an *n-p-n* transistor.

In the special case of a uniform-base transistor ($m = 0$) l'Hospital's rule can be applied to Eqs. 4.33 and 4.34 to give

$$n(x) = n(0)\left(1 - \frac{x}{W_B}\right), \tag{4.35}$$

$$Q_{\text{tot}} = -qA_E W_B\, n(0)\left(\frac{1}{2}\right). \tag{4.36}$$

4.2.1.2 Emitter Current and Emitter Resistance

The electron current density in the base of an *n-p-n* transistor can be calculated from Eq. 4.32 and the electron concentration $n(x)$. In general, J_e at any point has drift and diffusion components, except at the collector junction where the drift component vanishes because the electron concentration is zero. Equation 4.33 for the electron distribution is an approximation based on the assumption of constant J_e. Because the approximation is best at the emitter, the current predicted from Eq. 4.33 should be equated with the emitter current I_E rather than the collector current I_C. At the emitter boundary of the base region

$$I_E = -A_E J_e(0)$$

(the minus sign because the reference direction for I_E is out of the device) and substituting from Eqs. 4.32, 4.33, and 4.34 we have

$$I_E = \left[\frac{m}{1 - \exp(-m)}\right]\left[\frac{qA_E D_e\, n(0)}{W_B}\right] = -\frac{Q_{\text{tot}} D_e}{W_B{}^2}\left[\frac{m^2}{m - 1 + \exp(-m)}\right]. \tag{4.37}$$

For later use it is convenient to define the *emitter resistance* r_E as the slope resistance of the base-emitter diode. This quantity appears in the definitions of many of the small-signal parameters of a transistor. In taking the derivative of I_E, two terms in Eq. 4.37 should formally be considered as functions of $V_{B'E}$:

(i) the injected minority carrier concentration $n(0)$ varies exponentially with $V_{B'E}$ (Eq. 4.6),
(ii) the base width W_B depends on $V_{B'E}$ because the emitter depletion layer width is a function of $V_{B'E}$ (Eq. 4.25*a*).

The latter effect is insignificant compared with the former.* Therefore

* Actually, Eq. 4.37 cannot be differentiated with respect to W_B to find the total change in current because W_B has been assumed constant at various points in the derivation (Eqs. 4.26, 4.27, 4.29, and 4.32). As evidence, note that the first form of Eq. 4.37 apparently varies as $W_B{}^{-1}$, whereas the second form varies as $W_B{}^{-2}$.

$$\frac{1}{r_E} = \left(\frac{\partial I_E}{\partial V_{B'E}}\right)_{V_{CE}} = \left[\frac{m}{1 - \exp(-m)}\right]\left[\frac{qA_E D_e n(0)}{W_B}\right]\left(\frac{q}{kT}\right) = \left(\frac{q}{kT}\right) I_E. \tag{4.38}$$

Thus the slope resistance is the same for all transistors, and its value depends only on the dc emitter current.

4.2.1.3 Effect of Thermal Generation

The total mobile charge Q_{tot} in the base of a transistor consists of two components. One is the excess minority carrier charge injected at the emitter junction, the other is the thermally generated equilibrium minority carrier charge Q_0:

$$Q_0 = -qA_E \int_0^{W_B} n_0(x)\, dx = -qA_E n_i^2 \int_0^{W_B} \frac{dx}{N_A(x)}. \tag{4.39}$$

Because the value of Q_0 is independent of the voltage $V_{B'E}$ applied to the emitter junction, the controlled charge Q of charge-control theory must be the excess injected charge obtained by subtracting Q_0 from Q_{tot}. The controlling charge Q_C is the excess majority carrier charge introduced via the base contact:

$$Q_C = -Q = -(Q_{tot} - Q_0). \tag{4.40}$$

Notice that Q can be positive or negative, depending on the relative magnitudes of Q_{tot} and Q_0. In contrast, Q cannot change sign in devices such as the vacuum tube for which there is no appreciable carrier generation in the control region.

The base current of a transistor makes up the difference between the rates of recombination and thermal generation of carriers in the control region. At constant temperature the rate of generation is constant but the rate of recombination is proportional to the total minority carrier charge present in the base. The base current is zero when the total charge is equal to the equilibrium charge, and therefore

$$I_B = -\frac{Q_{tot} - Q_0}{\tau} = -\frac{Q}{\tau} = \frac{Q_C}{\tau}, \tag{4.41}$$

where τ is the lifetime of carriers in the base.

The collector current can now be evaluated with very small error by subtracting the base current predicted by Eq. 4.41 from the accurately known and much larger emitter current (Eq. 4.37):

$$I_C = I_E - I_B = -Q_{tot}\left\{\frac{D_e}{W_B^2}\left[\frac{m^2}{m - 1 + \exp(-m)}\right] - \frac{1}{\tau}\right\} - \frac{Q_0}{\tau}. \tag{4.42}$$

Thus I_C has two components, one constant and the other proportional to Q_{tot}. Because the whole of Q_{tot} is a mobile charge (in the sense that the whole of Q_{tot} can move), the coefficient of Q_{tot} must be the reciprocal of the transit time τ_1; this is quite obvious from considering changes in Q_{tot} and the resultant changes in I_C:

$$\frac{1}{\tau_1} = \frac{D_e}{W_B^2}\left[\frac{m^2}{m - 1 + \exp(-m)}\right] - \frac{1}{\tau}. \tag{4.43}$$

The constant component of I_C is the thermally generated *collector saturation current* I_{CO} which flows in the absence of emitter current:

$$I_{CO} = -\frac{Q_0}{\tau} = \frac{qA_E n_i^2}{\tau}\int_0^{W_B}\frac{dx}{N_A(x)}. \tag{4.44}$$

Because it is proportional to the intrinsic concentration n_i^2, I_{CO} is extremely temperature-sensitive.

Three new symbols are introduced to simplify the notation:*

$$\alpha_N = \frac{\tau}{\tau + \tau_1} = \frac{\beta_N}{\beta_N + 1}, \tag{4.45}$$

$$\beta_N = \frac{\tau}{\tau_1} = \frac{\alpha_N}{1 - \alpha_N}, \tag{4.46}$$

$$I'_{CO} = \frac{I_{CO}}{1 - \alpha_N} = Q_0\left(\frac{\tau + \tau_1}{\tau\tau_1}\right); \tag{4.47}$$

β_N is recognized as the dc current gain of charge-control theory (Eq. 2.5) and α_N approaches unity if β_N is large. With this notation, the transit time becomes

$$\tau_1 = \frac{W_B^2}{\alpha_N D_e}\left[\frac{m - 1 + \exp(-m)}{m^2}\right] \tag{4.48}$$

and the collector current is given by any of the three expressions

$$I_C - I'_{CO} = -\frac{Q_{tot} - Q_0}{\tau_1} = -\frac{Q}{\tau_1} = \frac{Q_C}{\tau_1}, \tag{4.49a}$$

$$I_C - I'_{CO} = \beta_N I_B, \tag{4.49b}$$

$$I_C - I'_{CO} = \alpha_N(I_E - I'_{CO}). \tag{4.49c}$$

All the above equations are consistent with Fig. 4.6; a transistor is the parallel combination of a charge-control model with a constant current generator. The collecting-electrode current and the emitting-electrode

* For alternative symbols see Section 4.7.1.

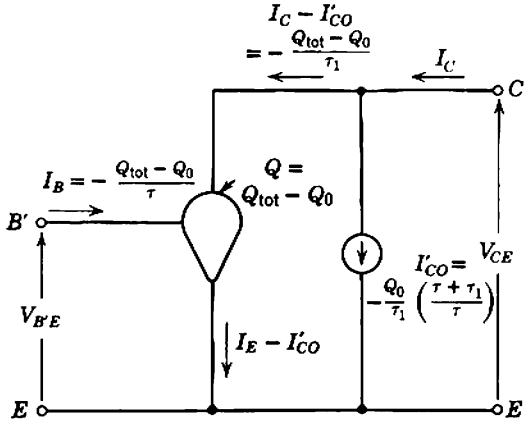


Fig. 4.6 The transistor as a charge-controlled device; I'_{CO} must be connected in shunt with the ideal model to account for thermal generation of minority carriers in the base.

current of the charge-control model are $(I_C - I'_{CO})$ and $(I_E - I'_{CO})$ respectively, so that I'_{CO} takes on the significance of an uncontrolled part of the emitter-to-collector current stream that flows when the controlled charge Q, the controlling charge Q_C, and the base current I_B are all zero. Evidently, spontaneous generation of mobile carriers in the control region of a charge-controlled device can be represented by a constant current generator across the output terminals of the charge-control model; because this generator is constant, it will not appear in the small-signal equivalent circuit of the device.

The charge-control representation of a transistor (Fig. 4.6) gives no indication of the actual paths followed by carriers; the useful control mechanism represented by the charge-control model in Fig. 4.6 is inseparably mixed with the processes that give rise to I'_{CO}. There are, in fact, three identifiable carrier streams in the base of a transistor:

(i) The useful component of current injected at the emitter flows across the base and reaches the collector; this component, $\alpha_N I_E$, follows from Eqs. 4.37, 4.42, 4.43, and 4.48.

(ii) The remainder of the injected current, $(1 - \alpha_N)I_E$, recombines and so an equal current must flow in the base lead wire.

(iii) Thermal generation of minority carriers in the base results in a carrier stream I_{CO} flowing across the collector junction, that is, flowing between base and collector.

Thus the total electrode currents in a transistor can be written as

$$I_C = \alpha_N I_E + I_{CO}, \tag{4.50a}$$

$$I_B = (1 - \alpha_N)I_E - I_{CO}. \tag{4.50b}$$

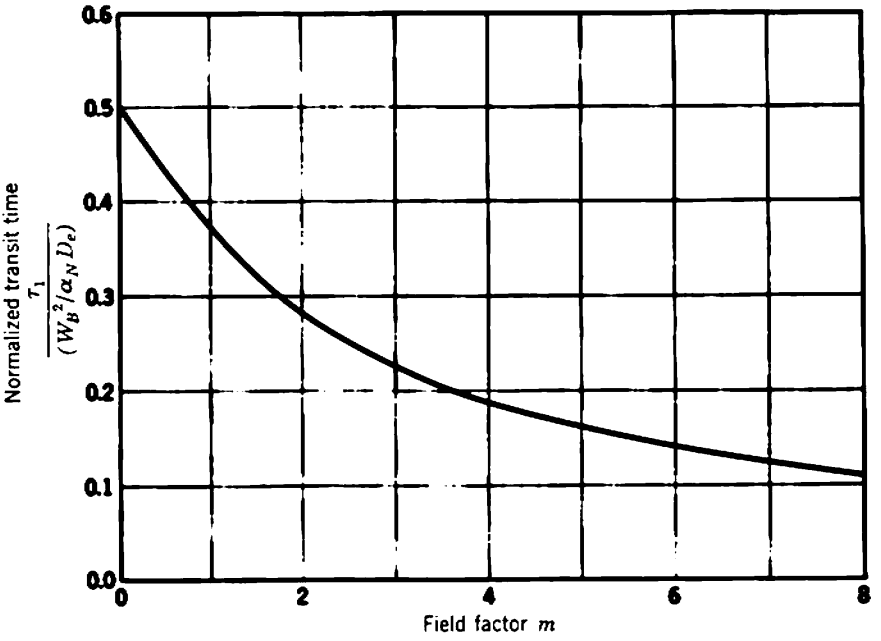


Fig. 4.7 Transit time as a function of field factor.

From Eq. 4.50*a*, I_{CO} takes on the significance of the collector current that flows when the emitter is open-circuited; I_{CO} is the saturation current of the collector junction. From Eq. 4.49*b*, I'_{CO} is the collector current that flows when the base is open-circuited.

Figure 4.7 is a normalized plot of τ_1 against m; τ_1 falls as m and the electric field built into the base increase. In the special case of a uniform-base transistor Eq. 4.48 becomes

$$\tau_1 = \frac{W_B^2}{\alpha_N D_e}\left(\frac{1}{2}\right). \tag{4.51}$$

4.2.2 Large-Signal Equivalent Circuit

The foregoing theory enables the relations between the dc terminal voltages and currents to be written down and suggests a *large-signal* or *dc equivalent circuit* which satisfies these equations. There are four important concepts:

1. Equation 4.30*a* shows that the injected minority carrier concentration $n(0)$ at the edge of the emitter depletion layer varies exponentially with $V_{B'E}$. To achieve a useful level of injection $V_{B'E}$ must be about 200 mV for germanium transistors and 700 mV for silicon types.

2. The emitter current varies linearly with $n(0)$ (Eq. 4.37), provided $n(x)$ at any point does not become so large that Eq. 4.28 is invalidated.

3. When thermal generation is considered, it follows that I_C is made up of two components; one is proportional to I_E and the other is a temperature-sensitive saturation current (Eq. 4.50*a*).

4. There is an appreciable resistance r_B between the base terminal B and the internal base B'.

When the voltage drops in the external circuit are large compared with a few hundred millivolts, it is reasonable to neglect $V_{B'E}$ and represent a transistor by the model shown in Fig. 4.8*a*. A better approximation is obtained by adding a constant voltage drop (200 or 700 mV for germanium or silicon transistors, respectively) in the emitter leg of the circuit (Fig. 4.8*b*). A still higher order of approximation results from adding a controlled current source, as in Fig. 4.8*c*:

$$I_E = (I_E|_{V_{B'E}=0}) \exp\left(\frac{qV_{B'E}}{kT}\right), \tag{4.52}$$

so that electrode currents increase exponentially with $V_{B'E}$. At 25°C, $V_{B'E}$ changes by 59.2 mV for each decade of current. In view of the production tolerances on transistor parameters, Fig. 4.8*b* is an adequate representation for most purposes.

Any junction transistor is a symmetrical device in the sense that it is a "sandwich" structure in which the two outer layers are the same type of

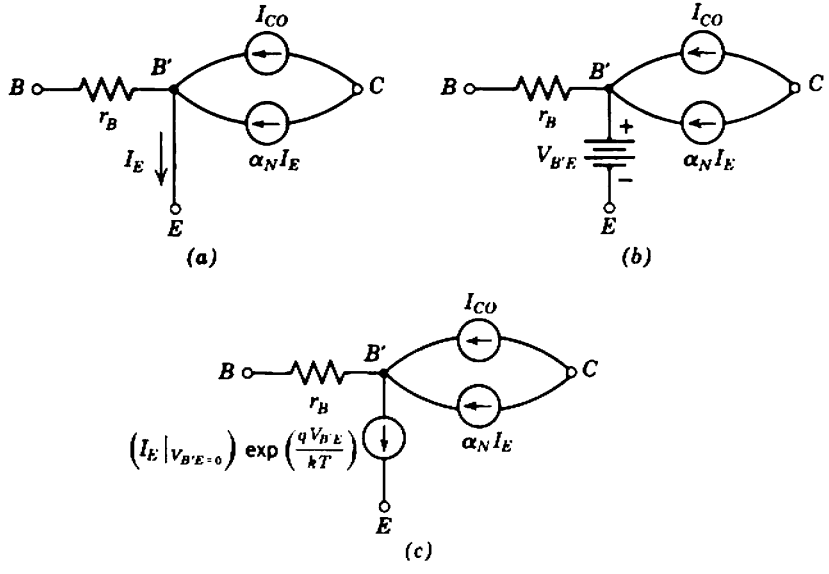


Fig. 4.8 Large-signal equivalent circuits of a transistor.

material. Although the emitter and collector may be made grossly different in structure to optimize performance, any transistor will operate with the roles of these two electrodes interchanged. This mode of operation is said to be *inverted* to distinguish it from the *normal mode*, and a new set of parameters α_I, β_I, and I_{EO} are defined for inverted operation. Knowledge of these parameters as well as the normal parameters α_N, β_N, and I_{CO} allows the large-signal performance to be calculated in switching circuits in which both the emitter and collector junctions become forward biased.*

4.2.3 Small-Signal Equivalent Circuit

In considering the small-signal performance of a transistor, the current generator I'_{CO} in Fig. 4.6 can be neglected because it is constant. Figure 4.9*a* shows the small-signal equivalent circuit for an intrinsic transistor,

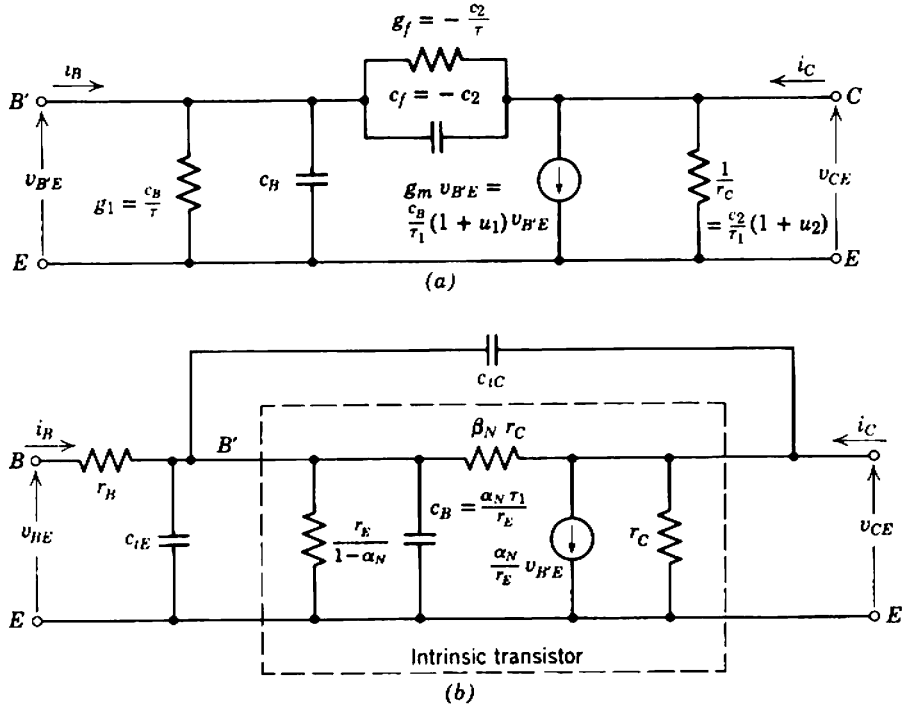


Fig. 4.9 Small-signal equivalent circuit of a transistor: (*a*) intrinsic elements only; (*b*) with the usual extrinsic elements (the collector and emitter spreading resistances may sometimes be required).

* J. J. EBERS and J. L. MOLL, "Large-signal behavior of junction transistors," *Proc. Inst. Radio Engrs.*, **42**, 1761, 1954.

derived in Section 2.5 (Fig. 2.7*e*) from charge-control theory. Two of the elements are given specific names:

(i) the *base-charging capacitance* c_B replaces the general capacitance c_1,

(ii) the *collector resistance* r_C replaces the general resistance $1/g_2$.

If the charge-control parameters Q_C and τ_1 given by Eqs. 4.40 (in conjunction with 4.34 and 4.39) and 4.48 are substituted into the relevant equations of Section 2.5, the values of all parameters can be found. Notice that Q_0 is a constant and its derivatives are zero.

The parameters involving changes in $V_{B'E}$ are obtained by straightforward substitution.* The values are

$$g_m = \left(\frac{\partial I_C}{\partial V_{B'E}}\right)_{V_{CE}} = \left(\frac{\tau}{\tau + \tau_1}\right)\left(\frac{q}{kT}\right)I_E = \frac{\alpha_N}{r_E}, \tag{4.53}$$

$$g_1 = \left(\frac{\partial I_B}{\partial V_{B'E}}\right)_{V_{CE}} = \left(\frac{\tau_1}{\tau + \tau_1}\right)\left(\frac{q}{kT}\right)I_E = \frac{1 - \alpha_N}{r_E}, \tag{4.54}$$

$$c_B = \left(\frac{\partial Q_C}{\partial V_{B'E}}\right)_{V_{CE}} = \left(\frac{\tau\tau_1}{\tau + \tau_1}\right)\left(\frac{q}{kT}\right)I_E = \frac{\alpha_N\tau_1}{r_E}. \tag{4.55}$$

Notice that the parameter u_1 defined in Eq. 2.47 is zero because τ_1 is independent of $V_{B'E}$:

$$u_1 = -\frac{(I_C - I'_{CO})}{c_B}\left(\frac{\partial \tau_1}{\partial V_{B'E}}\right)_{V_{CE}} = 0. \tag{4.56}$$

The parameter u_1 is, in fact, zero for all devices in which carrier motion is (or can be ascribed to) diffusion.† Notice, too, the compactness that results from expressing g_m, g_1, and c_B in terms of r_E and α_N.

The derivation of the parameters involving changes in V_{CE} is rather more complicated. If the collector voltage is increased, the collector depletion layer will widen, thereby reducing the base width W_B. The derivative

$$\left(\frac{\partial W_B}{\partial V_{CB'}}\right)_{V_{B'E}} = \left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}}$$

is therefore finite and negative. Therefore from Eqs. 4.40, 4.34, and 4.39

$$c_2 = \left(\frac{\partial Q_C}{\partial V_{CE}}\right)_{V_{B'E}} \approx qA_E\, n(0)\left\{\frac{m - 1 + \exp(-m)}{m[1 - \exp(-m)]}\right\}\left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}}$$

and, using Eqs. 4.37 and 4.48,

$$c_2 = \left(\frac{\partial Q_C}{\partial V_{CE}}\right)_{V_{B'E}} = \tau_1\left(\frac{\alpha_N I_E}{W_B}\right)\left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}}. \tag{4.57}$$

* Equation 4.55 is derived first by substitution into Eq. 2.39; Eqs. 4.30*a*, 4.38, and 4.48 are used in the subsequent algebraic simplifications. Equations 4.53 and 4.54 then follow by substitution into Eqs. 2.53 and 2.54.

† R. D. MIDDLEBROOK, "A modern approach to vacuum and semiconductor device theory," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 887, Suppl. 17, 1959.

Notice that c_2 is negative (because of the sign of the derivative), as expected from the discussion in Section 2.3.3. The parameter u_2 defined in Eq. 2.49 is

$$u_2 = -\frac{I_C - I'_{CO}}{c_2}\left(\frac{\partial \tau_1}{\partial V_{CE}}\right)_{V_{B'E}}.$$

From Eq. 4.48

$$\left(\frac{\partial \tau_1}{\partial V_{CE}}\right)_{V_{B'E}} = \left(\frac{2\tau_1}{W_B}\right)\left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}}$$

and therefore

$$u_2 = -2\left(\frac{I_C - I'_{CO}}{\alpha_N I_E}\right).$$

Under typical operating conditions $I_C \gg I'_{CO}$ and $I_C \approx \alpha_N I_E$, so that

$$u_2 \approx -2. \tag{4.58}$$

Therefore the values of the parameters in Fig. 4.9 are*

$$\frac{1}{r_C} = \left(\frac{\partial I_C}{\partial V_{CE}}\right)_{V_{B'E}} = -\frac{c_2}{\tau_1} = -\left(\frac{\alpha_N I_E}{W_B}\right)\left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}}, \tag{4.59}$$

$$g_f = -\left(\frac{\partial I_B}{\partial V_{CE}}\right)_{V_{B'E}} = -\frac{c_2}{\tau} = \frac{1}{\beta_N r_C}, \tag{4.60}$$

$$c_f = -c_2. \tag{4.61}$$

All parameters have positive values, but c_f is so small that it is invariably neglected; values larger than 1 pF are exceptional—0.01 pF is typical.

The collector junction of a transistor is reverse biased by a large amount, so that

$$V_i - V_{B'C} \approx V_{CE}.$$

Moreover, the junction is usually abrupt. Equation 4.25*b* therefore gives

$$\left(\frac{\partial W_B}{\partial V_{CE}}\right)_{V_{B'E}} = -\left(\frac{\partial W_C}{\partial V_{CE}}\right)_{V_{B'E}} \approx -\left[\frac{\varepsilon}{2q\,N_A(W_B)\,V_{CE}}\right]^{1/2} \tag{4.62}$$

to show that r_C is proportional to $\sqrt{V_{CE}}$.

4.2.3.1 Extrinsic Elements

A number of extrinsic elements should be added to the small-signal equivalent circuit of a transistor.

As noted in Section 4.2, the finite spreading resistance of the semiconducting crystal appears in series with each of the electrode lead wires.

* Equations 4.59 and 4.60 follow directly from Eqs. 2.55 and 2.56. Equation 4.61 is based on the discussion of Fig. 2.7*e*.

The collector and emitter spreading resistances are often neglected, but the base spreading resistance r_B is significant and should be included in a complete equivalent circuit.

It is well known that *p-n* junction diodes do not obey the ideal law

$$I = I_0\left[\exp\left(\frac{qV}{kT}\right) - 1\right] \tag{4.63}$$

at reverse voltages. The excess current can be ascribed to a *leakage conductance* across the junction. A similar leakage conductance appears across the reverse-biased collector junction of a transistor, where it is in shunt with g_f in the equivalent circuit. Section 4.3.1.4 discusses the origin of this leakage conductance; it is so small that it can almost always be neglected.

Stray capacitances appear between all lead wires and between the electrodes and the encapsulation. These capacitances are of the order of 1 pF for small transistors and are neglected except possibly at very high frequencies. However, the transition capacitances of the junctions are significant and must be included in a small-signal equivalent circuit. The junctions of most transistors are abrupt, and the base is much less doped than the emitter or collector. Equations 4.11 and 4.12 therefore give the *emitter transition capacitance* as

$$c_{tE} = A_E\left[\frac{q\epsilon\, N_A(0)}{2(V_1 - V_{B'E})}\right]^{1/2}. \tag{4.64}$$

The collector junction is reverse biased by a large amount and the *collector transition capacitance* becomes

$$c_{tC} \approx A_C\left[\frac{q\epsilon\, N_A(W_B)}{2V_{CE}}\right]^{1/2}. \tag{4.65}$$

Table 4.2 Parameters that Completely Characterize the Small-Signal Performance of a Transistor

Independent Parameter	Related Parameter	Relation
Base resistance r_B	—	—
Collector resistance r_C	—	—
Emitter resistance r_E	Dc emitter current I_E	$r_E = \dfrac{kT}{qI_E}$
Current amplification factor β_N	α_N	$\beta_N = \dfrac{\alpha_N}{1 - \alpha_N}$
Base charging capacitance c_B	Transit time τ_1	$c_B = \dfrac{\alpha_N \tau_1}{r_E}$
Emitter transition capacitance c_{tE}	—	—
Collector transition capacitance c_{tC}	—	—

Figure 4.9*b* shows the small-signal equivalent circuit for a transistor with the important extrinsic elements. Apart from the fact that the shunt input capacitance is split into two components, this circuit is topologically identical with Giacoletto's *hybrid* π equivalent circuit.* The nomenclature, of course, is different. Table 4.2 lists the independent parameters in the equivalent circuit and their relations to other parameters.

4.2.3.2 Sinusoidal Response

Because the parameter u_1 is zero for a transistor, the current amplification factor β of charge-control theory is numerically equal to the dc current gain (Eqs. 2.57 and 2.5). Either is represented by β_N (or β_I for inverted operation). It also follows from Section 2.5.1 that the short-circuit current gain and the gain-bandwidth product of an intrinsic transistor are

$$\frac{i_C(j\omega)}{i_B(j\omega)} = \beta_N\left(\frac{1}{1 + j\omega\tau}\right) = \beta_N\left(\frac{1}{1 + j\omega\beta_N\tau_1}\right), \tag{4.66}$$

$$\mathscr{GB} = \frac{g_m}{c_B} = \frac{1}{\tau_1} = \frac{\alpha_N}{r_E c_B}. \tag{4.67}$$

In some applications a signal current is fed into the emitter of a transistor (rather than the base) and the output is taken from the collector. The short-circuit current gain of an intrinsic transistor in this configuration is

$$\frac{i_C(j\omega)}{i_E(j\omega)} = \alpha_N\left(\frac{1}{1 + j\omega\alpha_N\tau_1}\right). \tag{4.68}$$

Thus α_N takes on the significance of a current amplification factor. The gain-bandwidth product is again $1/\tau_1$, illustrating that $\mathscr{GB}$ is independent of the configuration in which a device is used as an amplifier.

A time constant τ_T is defined for an extrinsic transistor as

$$\tau_T = \frac{r_E(c_B + c_{tE})}{\alpha_N}.$$

If we neglect the collector transition capacitance (which is invariably small) the short-circuit current gains become

$$\frac{i_C(j\omega)}{i_B(j\omega)} = \beta_N\left(\frac{1}{1 + j\omega\beta_N\tau_T}\right), \tag{4.69}$$

$$\frac{i_C(j\omega)}{i_E(j\omega)} = \alpha_N\left(\frac{1}{1 + j\omega\alpha_N\tau_T}\right); \tag{4.70}$$

* L. J. GIACOLETTO, "Study of *p-n-p* alloy junction transistors from dc through medium frequencies," *RCA Rev.*, **15**, 506, December 1954.

the gain-bandwidth product is

$$\mathcal{GB} = \frac{g_m}{c_B + c_{tE}} = \frac{1}{\tau_T} = \frac{\alpha_N}{r_E(c_B + c_{tE})}. \tag{4.71}$$

Thus the effect of the emitter transition capacitance is to reduce the realizable gain-bandwidth product in the ratio τ_1/τ_T. Figure 4.10 compares the current gains of the intrinsic and extrinsic transistor.

In some of the subsequent work it is convenient to use frequency, radian frequency, and time constant interchangeably. The following nomenclature is introduced for the gain-bandwidth product of the intrinsic and extrinsic transistors:

$$2\pi f_1 = \omega_1 = \frac{1}{\tau_1}, \tag{4.72}$$

$$2\pi f_T = \omega_T = \frac{1}{\tau_T}. \tag{4.73}$$

Another symbol sometimes used is μ, the voltage amplification factor of the intrinsic transistor, defined for the general charge-control model in Eq. 2.58. From Eqs. 4.53 and 4.59

$$\mu = -\left(\frac{\partial V_C}{\partial V_{B'E}}\right)_{I_C} = \alpha_N\left(\frac{r_C}{r_E}\right). \tag{4.74}$$

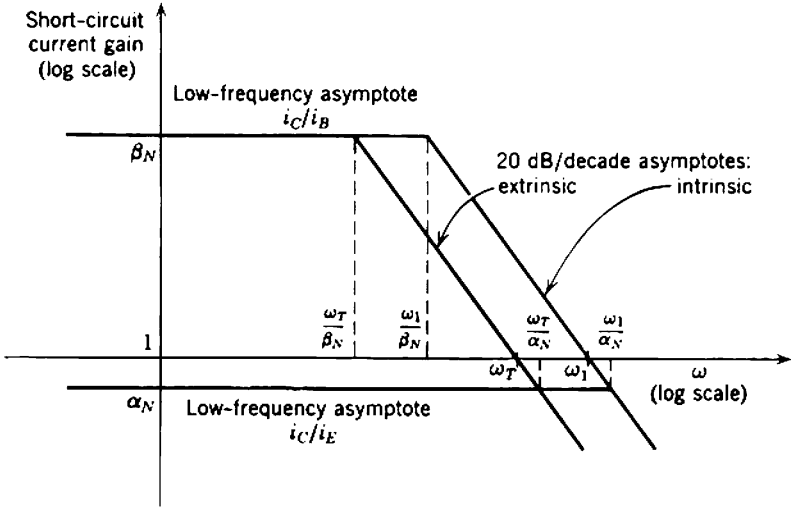


Fig. 4.10 Asymptotes of current gain versus frequency for intrinsic and extrinsic transistors.

4.3 LIMITATIONS OF ELEMENTARY TRANSISTOR THEORY

The charge-control representations of vacuum tubes and transistors suffer from much the same limitations. Some errors are introduced in the physical idealizations and others by the inadequacy of the charge-control approximation. This section is larger than the corresponding discussion of tubes because the errors are more numerous and more subtle. However, the total error is certainly no greater for transistors than for tubes; the large-signal representation of Fig. 4.8 and the small-signal representation of Fig. 4.9 are adequate for low-pass amplifier design.*

4.3.1 Physical Idealizations†

The idealized physical model assumed in Section 4.2 has three main inaccuracies: the assumption that the injected minority carrier concentration in the base is small compared with the acceptor concentration so that the minority carrier concentration is given by Eq. 4.28 rather than Eq. 4.7*a*, the assumption that I_E is equal to the minority carrier current injected into the base, and the assumption that I_{CO} arises entirely from thermal generation of minority carriers in the base. The parameters τ_1, β_N, r_E, and I_{CO} are therefore not given exactly by Eqs. 4.48, 4.46, 4.38, and 4.44.

4.3.1.1 Variation of τ_1

The transit time τ_1 is a function of the dc collector voltage and current. The first-order dependence on collector voltage is incorporated into the elementary theory. As the collector voltage is increased, the collector depletion layer widens and reduces the base width, and the transit time decreases because it is proportional to W_B^2 (Eq. 4.48). At very high voltages the collector depletion layer can extend throughout the entire control region and touch the emitter depletion layer. In theory, τ_1 falls to zero and the collector current would become infinite if it were not limited by external resistances. This effect is called *punch-through* and the corresponding collector voltage is the *punch-through voltage*. At very low collector voltages the field in the depletion layer becomes so small that the carrier velocity falls below its scatter-limited value; the transit

* The remainder of Section 4.3 may be omitted on a first reading, although isolated points are referred to in later parts of the chapter.

† Two classic papers on the variation of the small-signal parameters of a transistor are: J. M. EARLY, "Effects of space-charge-layer widening in junction transistors," *Proc. Inst. Radio Engrs.*, **40**, 1401, November 1952; W. M. WEBSTER, "On the variation of junction transistor current-amplification with emitter current," *Proc. Inst. Radio Engrs.*, **42**, 914, June 1954.

time through the depletion layer becomes comparable with the transit time through the base. Significant controlled charge is stored in the depletion layer and τ_1 ($= -Q/I_C$) is increased.

The dependence of τ_1 on emitter current disappears from the elementary theory because of the assumption implicit in Eq. 4.28. The majority carrier concentration in the base is equal to the doping density only if the minority carrier concentration is small in comparison; Eq. 4.7 applies when this is not the case. At high current densities, under *high-level injection conditions*, the injected minority carrier concentration $n(x)$ becomes comparable with, or even exceeds, the acceptor concentration $N_A(x)$, and Eq. 4.29 no longer gives the electric field in the base.

Consider first a uniform-base transistor for which $N_A(x)$ is constant throughout the base, so that $p(x)$ in the elementary theory is also sensibly constant. As the forward bias on the emitter junction is increased, the concentration gradient of injected minority carriers increases, and for charge neutrality the majority carrier gradient must also increase. There is a tendency, therefore, for majority carriers to diffuse toward the collector. However, there cannot be a net flow of majority carriers toward the collector junction because they cannot cross it against the high opposing field. Hence there must be an electric field in the base that holds the majority carriers in equilibrium, and this field (like the field actually built into a graded-base transistor) assists minority carriers across the base. Poisson's equation shows that the field increases with current density, and as a result τ_1 falls; the fall in τ_1 amounts to a factor of two at very high currents.*

Much of the same happens in a graded-base transistor. The built-in graded distribution of majority carriers is masked by the excess carriers, and in the limit of very high current the conditions are indistinguishable from those in a uniform-base transistor. The transit time rises or falls, depending on whether the initial advantage of the built-in field is less or more than the factor 2. For most graded-base transistors the initial advantage is about fourfold, and τ_1 therefore rises at high currents.

Another mechanism that may account for a rise in τ_1 at high currents is *base pulling*.† If the collector junction is formed between very lightly doped p and n regions (the usual situation for epitaxial transistors), the fixed donor and acceptor concentrations in the depletion layer are small. At high currents the mobile charge density in transit through the depletion layer is comparable with the fixed charge density and distorts the electric-field

* W. M. WEBSTER, *loc. cit.*

† C. J. KIRK, "A theory of transistor cut-off frequency (f_T) falloff at high current densities," *Trans. Inst. Radio Engrs.*, **ED-9**, 164, March 1962.

pattern. The depletion layer shrinks, and the residual base width W_B is increased.

4.3.1.2 Variation of α_N and β_N

Elementary transistor theory assumes that the emitter current is equal to the minority carrier current injected into the base and that the base current is related to the excess minority carrier charge in the base by the carrier lifetime τ. Neither assumption is quite true for practical transistors. A convenient way for displaying the inaccuracies is to consider the effects that determine the small-signal current amplication factor α_N. Once the value of α_N is established, a time constant τ_{eff} can be defined such that

$$\alpha_N = \frac{\tau_{\text{eff}}}{\tau_{\text{eff}} + \tau_1},$$

and charge-control theory applies with τ_{eff} replacing τ. However, it is not necessary to take this formal step, for the second forms of Eqs. 4.53, 4.54, and 4.55 express the small-signal parameters directly in terms of α_N and the slope resistance of the base-emitter diode. Notice that any variations of α_N near the point at which $\alpha_N = 1$ are magnified enormously in β_N.

The current amplification factor α_N is the product of three terms—the *emitter efficiency* α_E, the *transport factor* or *base efficiency* α_B, and the *collector efficiency* or *avalanche multiplication factor* α_C. For an *n-p-n* transistor the following definitions apply:

$$\alpha_E = \frac{d(\text{electron current injected into the base})}{d(\text{total emitter current})}, \tag{4.75}$$

$$\alpha_B = \frac{d(\text{electron current swept out of the base})}{d(\text{electron current injected into the base})}, \tag{4.76}$$

$$\alpha_C = \frac{d(\text{total collector current})}{d(\text{electron current swept out of the base})}, \tag{4.77}$$

$$\alpha_N = \alpha_E \alpha_B \alpha_C. \tag{4.78}$$

4.3.1.2.1. Emitter Efficiency

The emitter efficiency α_E of a transistor reaches a maximum value between 0.95 and unity at an optimum emitter current but falls for currents on either side of the optimum. This can be explained in terms of elementary *p-n* junction theory modified to account for carrier recombination in the emitter depletion layer. The value of α_E is independent of collector voltage because α_E depends only on happenings at the emitter junction.

The operation of an ideal junction diode is often explained in terms of four components of current. Majority carriers are injected from each side of the junction to the other against the electric field in the depletion layer, and minority carriers generated thermally on each side of the junction are swept out. The two thermal currents are independent of the voltage applied to the junction, although they are extremely sensitive to variations of temperature. In contrast, the injected currents vary exponentially with applied voltage and mask the thermal currents except at low forward biases or very high temperatures. Further, the ratio of the two injected currents (and, incidentally, the two thermal currents) is roughly the ratio of the majority carrier concentrations on the two sides of the junction. Either injected current can therefore be made to dominate by doping the two sides of the junction unequally.

Transistors have their emitter region much more highly doped than their base region. At moderate forward biases, therefore, the dominant component of emitter current is the electron current injected into the base and the emitter efficiency α_E approaches unity. Elementary transistor theory ignores all other components. However, α_E falls at low forward biases where the injected currents are of the same order as the thermal currents. Emitter efficiency also falls at high currents with the onset of high-level injection. The majority carrier concentration in the base rises to preserve charge neutrality, and correspondingly the current injected from the base back into the emitter increases in relation to (but remains less than) the useful current injected into the base. This phenomenon is known as *conductivity modulation* of the base.

Often a more significant mechanism which accounts for the fall in α_E at low currents is recombination in the emitter depletion layer. Recombination between holes and electrons in any semiconducting crystals takes place principally at localized points or *traps*, and some of these traps exist in the emitter depletion layer or at its surface. Electrons injected into the depletion layer from the emitter are stored in the traps and later recombine with holes injected from the base. The result is a majority carrier current flow through the junction. The recombination current per unit depletion-layer volume follows the law*

$$I = I_0\left[\exp\left(\frac{qV}{nkT}\right) - 1\right], \tag{4.79}$$

where n lies between 1 and 2, depending on the detail of the junction.

* C. T. SAH, R. N. NOYCE, and W. SHOCKLEY, "Carrier generation and recombination in *p-n* junctions, and *p-n* junction characteristics," *Proc. Inst. Radio Engrs.*, **45**, 1228, September 1957.

Because of the factor n in the denominator, the recombination current becomes comparable with the useful injected current at low forward biases and α_E falls. At a given bias the injected current decreases as the impurity concentration on the less doped side of the junction is increased. Therefore α_E falls more at low currents for graded-base transistors than for uniform-base transistors because of the heavy doping of the base region adjacent to the emitter. The recombination current increases with increasing trap density (i.e., decreasing carrier life time). Silicon crystals tend to have anomalously high trap densities (particularly at the surface), and α_E of most silicon transistors manufactured before 1962 therefore falls sharply at low currents. More recent technology has overcome this deficiency.

4.3.1.2.2 Transport Factor

The transport factor α_B is given by

$$\alpha_B = \frac{\tau}{\tau + \tau_1}, \tag{4.80}$$

where τ is the effective lifetime of minority carriers in the base and τ_1 is the transit time. The elementary theory of Section 4.2 assumes in effect that α_B is the only nonunity term in α_N. Because of changes in τ_1, α_B increases with increasing collector voltage and either increases or decreases with increasing emitter current. The effective lifetime τ remains fairly constant. Notice that τ is very different numerically from the life expectancy of a carrier in the control region. Recombination takes place at traps, and traps are so much denser at the base surface than in the control region that almost all recombination takes place at the surface. The surface behaves as a sink for carriers, and the transport factor is the proportion of carrier trajectories that finish on the collector junction rather than the base surface. For narrow-base transistors the trajectories run parallel across the control region and α_B is 0.99 or more. As the base width is increased, the trajectories diverge and α_B may be as low as 0.95.

4.3.1.2.3 Collector Efficiency

A very high electric field exists in the collector depletion layer, and carriers passing through it are accelerated to high velocities. If the collector voltage and field strength are great enough, carriers acquire sufficient energy to ionize the crystal lattice by collision and thereby increase the number of carriers in transit. The situation is analogous to

Townsend discharge in gases. McKay and Miller* have shown that the *avalanche multiplication factor* for minority carrier currents is

$$\alpha_C = \frac{1}{1 - (V_{CB}/V_a)^n}, \tag{4.81}$$

where V_a is the avalanche breakdown voltage and the exponent n varies from 3 to 6, depending on the junction structure. Amplifying transistors are rarely operated at collector voltages for which α_C is more than a few tenths of a percent greater than unity. The multiplication factor depends on the junction temperature but not on the current density.

4.3.1.2.4 Summary

In summary, the small-signal current amplification factor α_N rises initially with increasing emitter current, reaches a maximum value which (at normal collector voltages) is just less than unity, and then falls again. In addition, α_N increases monotonically with collector voltage and becomes infinite at the point of avalanche breakdown. Figure 4.11 shows the variation of α_N with I_E and V_{CE}, and Fig. 4.12 shows the corresponding curves for β_N. There is a wide range over which α_N and β_N are substantially constant, and transistors are normally operated within this range.

Because the small-signal current amplification factors are not strictly constant, Eqs. 4.50*a* and 4.49*b* should be rewritten:

$$I_C = \bar{\alpha}_N I_E + I_{CO}, \tag{4.82a}$$

$$I_C = \bar{\beta}_N I_B + I'_{CO}, \tag{4.82b}$$

where $\bar{\alpha}_N$ and $\bar{\beta}_N$ are suitably defined *average* or *large-signal current amplification factors.* The *small-signal current amplification factors* are defined for increments in the electrode currents:

$$\alpha_N = \left(\frac{\partial I_C}{\partial I_E}\right)_{V_{CE}} = \bar{\alpha}_N + I_E\left(\frac{\partial \bar{\alpha}_N}{\partial I_E}\right)_{V_{CE}}, \tag{4.83a}$$

$$\beta_N = \left(\frac{\partial I_C}{\partial I_B}\right)_{V_{CE}} = \bar{\beta}_N + I_B\left(\frac{\partial \bar{\beta}_N}{\partial I_B}\right)_{V_{CE}}. \tag{4.83b}$$

The large- and small-signal parameters are exactly equal at two points only; $\bar{\alpha}_N$ and α_N are equal at zero I_E and the turning point of $\bar{\alpha}_N$, whereas $\bar{\beta}_N$ and β_N are equal at zero I_B and the turning point of $\bar{\beta}_N$. However, the differences at all other points are small compared with the production

* K. G. McKay, "Avalanche breakdown in silicon," *Phys. Rev.*, **94**, 877, May 1954; S. L. Miller, "Avalanche breakdown in germanium," *Phys. Rev.*, **99**, 1234, August 1955.

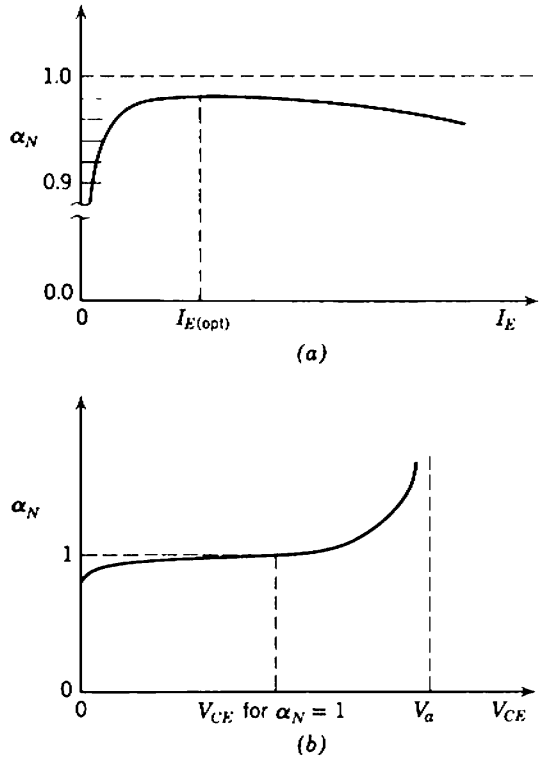


Fig. 4.11 Variation of α_N with I_E and V_{CE}: (*a*) α_N versus I_E at typical operating V_{CE}; (*b*) α_N versus V_{CE} at typical operating I_E.

tolerances; therefore the formal distinction between large- and small-signal parameters is ignored for most of this book.

4.3.1.3 Variation of r_E

Elementary transistor theory assumes only one component in the emitter current—minority carrier injection into the base. As a consequence of this assumption, the slope resistance of the base-emitter diode is inversely proportional to the emitter current (Eq. 4.38):

$$\frac{1}{r_E} = \left(\frac{\partial I_E}{\partial V_{B'E}}\right)_{V_{CE}} = \left(\frac{q}{kT}\right) I_E.$$

Over the useful operating range of emitter currents, minority carrier injection dominates and Eq. 4.38 is a good approximation. However, the additional components listed in Section 4.3.1.2.1 become significant at low and high currents, and r_E is given by an expression of the form

$$\frac{1}{r_E} = \left(\frac{\partial I_E}{\partial V_{B'E}}\right)_{V_{CE}} = \frac{q}{\sigma(I_E)\, kT}(I_E + I_{EO}). \tag{4.84}$$

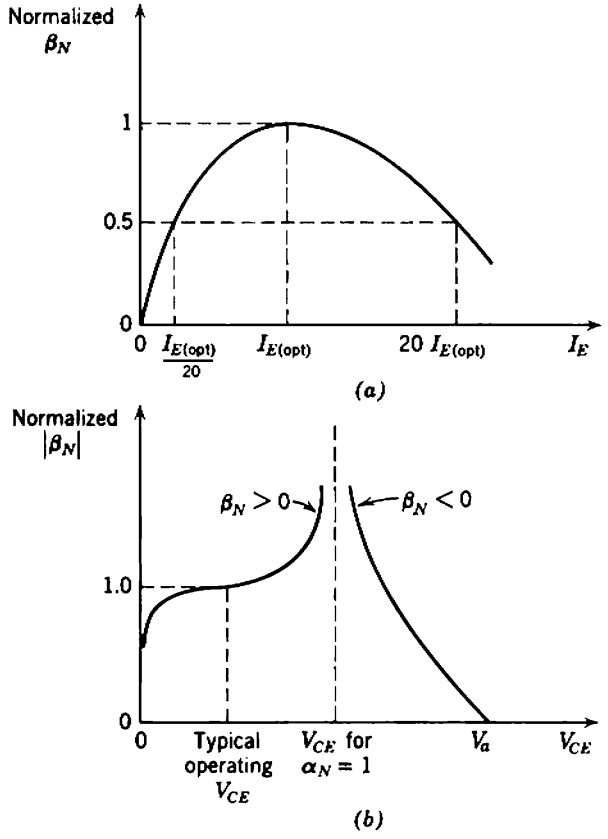


Fig. 4.12 Variation of β_N with I_E and V_{CE}: (*a*) typical normalized β_N versus I_E at typical operating V_{CE}; (*b*) normalized $|\beta_N|$ versus V_{CE} typical operating I_E.

At low currents the function $\sigma(I_E)$ rises from 1 to 2 to account for the factor n in the depletion-layer recombination current; I_{EO} accounts for the noninfinite slope resistance at zero current and is roughly the total thermal current of the junction. Although variations and irregularities occur in α_N and r_E separately, the electron current injected into the base behaves as predicted by elementary theory. It has been observed experimentally that the collector current increases exponentially with $V_{B'E}$ over eight decades.*

At high currents $\sigma(I_E)$ again rises because of a combination of two main effects. First, with the onset of high-level injection and conductivity modulation of the base, the slope resistance of an otherwise ideal junction diode doubles. The increase in majority carrier concentration depresses

* J. E. IWERSON, A. R. BRAY, and J. J. KLEIMACK, "Low-current alpha in silicon transistors," *Trans. Inst. Radio Engrs.*, ED-9, 474, November 1962.

the minority carrier concentration expected at equilibrium (Eq. 4.2) and therefore depresses the injected minority carrier concentration at a given forward bias (Eq. 4.6). Second, to represent the base spreading resistance by the single extrinsic element $r_{bb'}$ is an approximation. The base current produces a transverse voltage drop inside the control region itself, and as a result the emitter junction is not forward biased uniformly over its entire area. Injection concentrates at the junction periphery, thereby reducing the effective area and increasing the slope resistance.

4.3.1.4 Variation of I_{CO}

The collector saturation current I_{CO} of a practical transistor can be divided into two main components, each of which can be subdivided. The first main component is the current in the body of the transistor structure, and the second is the current on the collector junction surface. All components increase with increasing collector voltage.

The body component of I_{CO} has four subcomponents. Two are due to thermal generation of minority carriers in the base and collector; the component generated in the base is the larger when the base is less doped than the collector and is the only component of I_{CO} considered in elementary transistor theory. Spontaneous hole-electron pair generation in the collector depletion layer accounts for a third component. This mechanism in a reverse-biased junction is the counterpart of recombination in a forward-biased junction; the current generated per unit depletion-layer volume follows Eq. 4.79 and saturates, but because the depletion-layer width is a function of collector voltage the total current increases indefinitely with increasing collector voltage.* Finally, all three components can be multiplied by the avalanche process.

Surface currents arise from the presence of contamination on the surface. Qualitatively, the contamination on the collector junction surface of a *n-p-n* transistor almost always consists of positively charged ions. This positive charge induces a negative space charge just below the surface of the *p*-type base material, so that the majority carriers at this point are electrons. The *p*-type base material near the surface is *inverted* into *n*-type, and this *n*-type *inversion layer* links up with the *n*-type collector, as shown in Fig. 4.13. The junction area is increased, and the saturation current is increased accordingly. However, the inversion layer is extremely thin—typically 10^{-7} cm. It therefore has a high resistance, and even a minute current flowing in it produces an appreciable voltage drop. It follows that the area of the inversion region that remains reverse biased despite this voltage drop is a function of the applied collector voltage, and

---

* C. T. SAH, R. N. NOYCE, and W. SHOCKLEY, *loc. cit.*

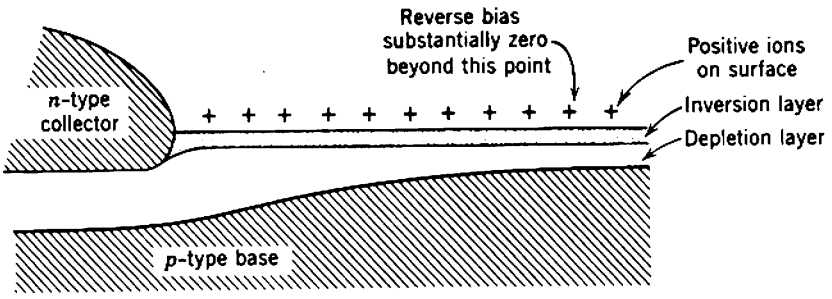


Fig. 4.13 Inversion layer on the base surface of an *n-p-n* transistor, showing the extension of the collector junction.

consequently the surface current is a function of voltage. Cutler and Bath* have shown that an ideal surface current should obey the law

$$I = I_0\left[\exp\left(\frac{qV}{kT}\right) - \left(\frac{qV}{kT}\right) - 1\right]^{1/2}, \tag{4.85}$$

and that at a large reverse bias

$$I \approx I_0\left(+\frac{q|V|}{kT}\right)^{1/2}. \tag{4.86}$$

Practical surface currents do not follow this parabolic law exactly for a number of reasons:

1. The ionic surface charge is not constant because it migrates with changes in the applied voltage.
2. In addition to thermal components, the surface junction saturation current contains components due to spontaneous carrier generation and avalanche multiplication.
3. The geometry is not ideal.

The first is perhaps the most significant because the time variation of surface current can give rise to a large component of flicker noise. The surface treatment of transistors is aimed at reducing the surface current, so that I_{CO} is predominantly the body junction saturation current. Transistors showing a large variation of I_{CO} with collector voltage usually have a poor surface and high flicker noise.

* M. Cutler and H. M. Bath, "Surface leakage current in silicon fused junction diodes," *Proc. Inst. Radio Engrs.*, **45**, 39, January 1957.

4.3.2 Limitations of Charge-Control Theory

With the changes to the physical model set out above, charge-control theory correctly predicts the values of all elements in the equivalent circuit at low frequencies. Errors occur at high frequencies because of the fundamental assumption of charge-control theory, and although these errors can be expressed in the same way as for tubes (as graphs of the variation of parameters with frequency) they are traditionally expressed in terms of an α *cutoff frequency*.

4.3.2.1 The Intrinsic Transistor

Charge-control theory assumes that the mobile charge in the control region takes up its final spatial distribution immediately on a change in electrode voltages. In fact, the electron distribution in the base of an *n-p-n* transistor is given by Eq. 2.32; neglecting recombination

$$\frac{\partial n}{\partial t} = D_e \frac{\partial^2 n}{\partial x^2} + \mu_e \mathscr{E} \frac{\partial n}{\partial x}. \tag{4.87}$$

The charge-control approximation assumes that $(\partial n/\partial t)$ is zero.

In principle, Eq. 4.87 can be solved for given boundary conditions to find the electron distribution as a function of distance and time, but in practice formal solution is seldom possible and either a numerical solution must be obtained or measurements made on an analogue. Probably the most convenient analogue is the lumped *RC* transmission line shown in Fig. 4.14, in which the $(k + 1)$th elements have the values

$$R_{k+1} = R_1 \exp\left(-m \frac{k}{N}\right), \tag{4.88a}$$

$$C_{k+1} = C_1 \exp\left(m \frac{k}{N}\right), \tag{4.88b}$$

and

$$R_1 C_1 = \frac{\tau_1}{N^2}\left[\frac{m^2}{m - 1 + \exp(-m)}\right] = \frac{1}{N^2}\left(\frac{W_B^2}{\alpha_N D_e}\right). \tag{4.89}$$

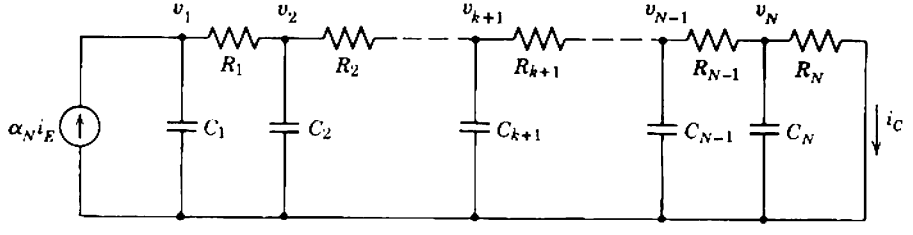


Fig. 4.14 *RC* transmission line analogue of the base region of a transistor. This model assumes that all recombination occurs in the emitter depletion layer; if desired, shunt conductances can be added to simulate recombination in the base.

Provided the number of sections N is large, the equation for charge in the analogue has the same form as Eq. 4.87. Measurement of the charge on the various capacitors of the analogue therefore yield values that can be interpreted as the time-distance variation of electron concentration in the base of a transistor that has the same excitation. The collector current is

$$i_C(t) = \frac{v_N(t)}{R_N}. \tag{4.90}$$

Because the line length (corresponding to base width W_B) is fixed, this simple analogue cannot be used to determine the effect of changes in collector voltage. It can, however, be used to determine the short-circuit collector current that flows in response to the following excitations:

(i) emitter current drive,
(ii) base current drive [determinable from (i) because $i_B(t) = i_E(t) - i_C(t)$],
(iii) base-to-emitter voltage drive (i.e., the response of g_m alone).

Emitter current drive is the most interesting and useful case. It allows an estimate to be made of the frequency range over which the charge-control representation is valid and provides one means for determining the field factor m from measurements on practical transistors.

Consider a small step ΔI_E in emitter current at time $t = 0$. Figure 4.15

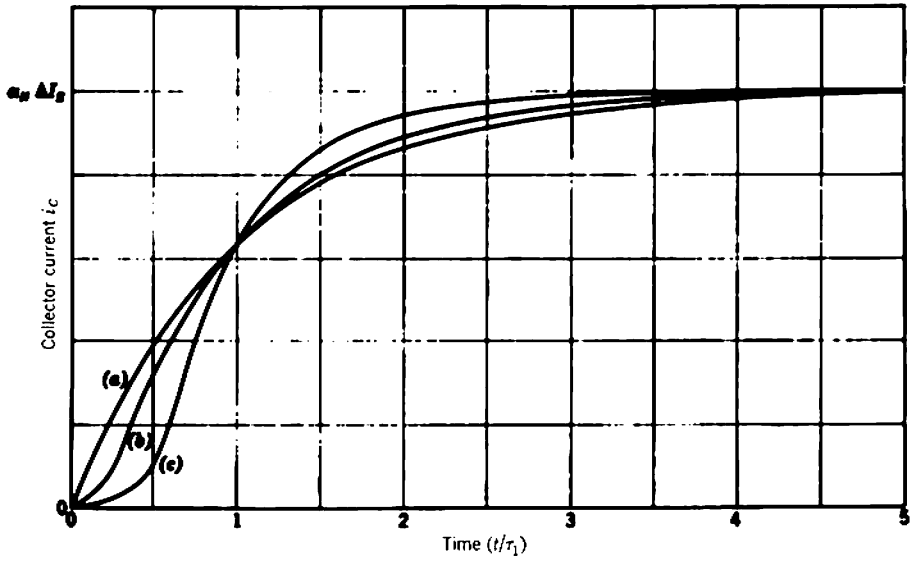


Fig. 4.15 Short-circuit collector current transients in response to an emitter current step: curve (*a*)—ideal charge-controlled transistor; curve (*b*)—uniform-base transistor ($m = 0$); curve (*c*)—graded-base transistor ($m = 8$).

shows the resulting collector current transients for a uniform-base transistor ($m = 0$), a highly-graded-base transistor ($m = 8$), and an ideal charge-controlled transistor. In the last case the transient follows from Eq. 4.68 as

$$i_C(t) = \alpha_N \Delta I_E \left[1 - \exp\left(\frac{-t}{\alpha_N \tau_1}\right)\right]. \tag{4.91}$$

The charge-control approximation is a poor fit to the early parts of the practical transients, but the error falls below 2% for times greater than about $2\tau_1$ (for the uniform-base transistor) or $4\tau_1$ (for the graded-base transistor). This suggests that the charge-control model is a valid representation for signals that do not change appreciably over the time interval $2\tau_1$ or $4\tau_1$. A working rule for circuit design is that the representation is valid provided none of the dominant poles in the complex frequency plane has a magnitude greater than $\omega_1/2$ for uniform-base transistors or $\omega_1/4$ for graded-base transistors.

A much better approximation to the collector current transient of a practical transistor is a delayed exponential:

$$[i_C(t)]_{t<t_D} = 0, \tag{4.92a}$$

$$[i_C(t)]_{t\geq t_D} = \alpha_N \Delta I_E \left[1 - \exp\left(\frac{t_D - t}{\tau_\alpha}\right)\right], \tag{4.92b}$$

where t_D is the *delay time* and the significance of the time constant τ_α can be established thus:

In complex frequency notation the emitter-to-collector current gain can be written as

$$\frac{i_C(s)}{i_E(s)} = \alpha_N \left[\frac{\exp(-st_D)}{1 + s\tau_\alpha}\right] \tag{4.93a}$$

or, in real frequency notation,

$$\frac{i_C(j\omega)}{i_E(j\omega)} = \alpha_N \left[\frac{\exp(-j\omega t_D)}{1 + j\omega\tau_\alpha}\right]. \tag{4.93b}$$

Writing

$$2\pi f_\alpha = \omega_\alpha = \frac{1}{\tau_\alpha}, \tag{4.94}$$

ω_α takes on the significance of a 3-dB cutoff frequency, the α *cutoff frequency*, at which the current gain falls to $1/\sqrt{2}$ of its low-frequency value. The exponential term contributes an *excess phase shift* over that of a single-pole function. Empirically, the relation between the transit time

τ_1 of a transistor, its measured α cutoff frequency, and the delay time that gives the best fit to the transient is*

$$t_D = \tau_1 - \tau_\alpha. \tag{4.95}$$

The base-to-collector current gain of an intrinsic transistor is not given exactly by Eq. 4.67 and Fig. 4.10, although the error is small at frequencies below $\omega_1/2$ or $\omega_1/4$ as appropriate for uniform-base and graded-base transistors. Curve (*a*) in Figure 4.16 is a typical plot. Notice the distinction between the gain-bandwidth product ω_1 at which the asymptote passes through unity gain and the true frequency of unity current gain.

Measurement of the ratio ω_α/ω_1 provides one means of determining the field factor m. Figure 4.17 shows polar plots of the emitter-to-collector current gain of uniform-base and graded-base transistors and the limiting case of an ideal charge-controlled transistor ($t_D = 0$ in Eqs. 4.93 and $\tau_1 = \tau_\alpha$). It is assumed that α_N is unity. Also shown (as a broken line) is the locus of ω_α:

$$\left|\frac{i_C(j\omega)}{i_E(j\omega)}\right| = \frac{\alpha_N}{\sqrt{2}}.$$

The ratio ω_α/ω_1 is unity for the ideal transistor, 1.22 for the uniform-base transistor, and increases monotonically with field factor to 1.93 at $m = 8$. For practical transistors in which α_N is not unity ω_α is larger by about the

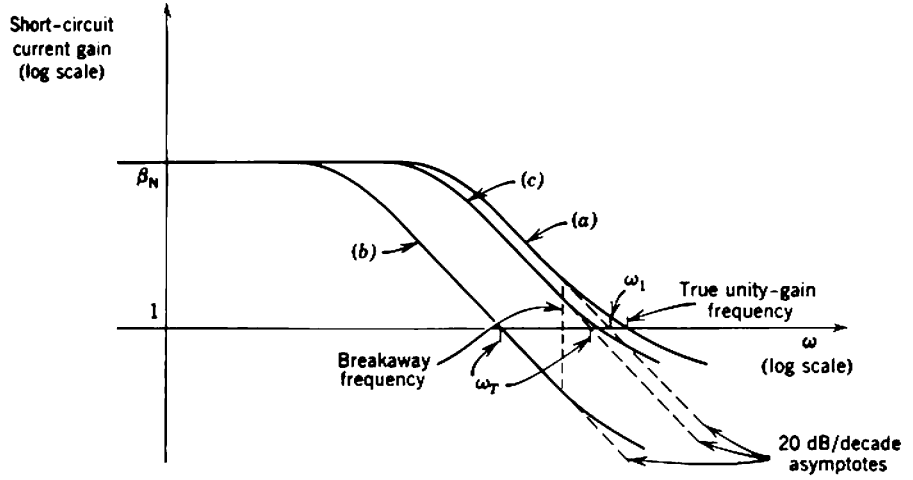


Fig. 4.16 Short-circuit current gain of a transistor: curve (*a*)—intrinsic transistor; curve (*b*)—$c_B \ll c_{tE}$; curve (*c*)—$c_B \gg c_{tE}$. The break-away frequency is about $\omega_1/2$ for a uniform-base transistor, and $\omega_1/4$ for a graded-base transistor.

* F. J. HYDE, "The current gains of diffusion and drift types of junction transistors," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 1046, Suppl. 17, 1959.

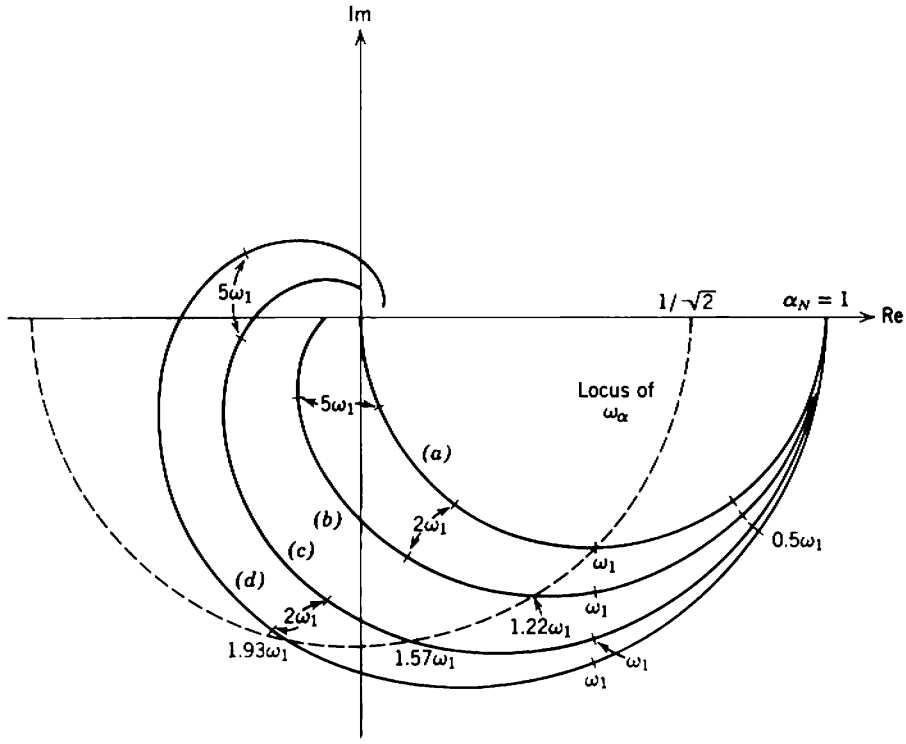


Fig. 4.17 Polar plots of $i_C(j\omega)/i_E(j\omega)$: curve (*a*)—charge-controlled device; curve (*b*) intrinsic uniform-base transistor ($m = 0$); curve (*c*) intrinsic graded-base transistor ($m = 4$); curve (*d*)—intrinsic graded-base transistor ($m = 8$).

factor $1/\alpha_N$. On the same graph, the locus of the true unity base-to-collector current gain magnitude is a straight line:

$$\text{Re}\left[\frac{i_C(j\omega)}{i_E(j\omega)}\right] = \frac{1}{2}.$$

Figure 4.18 is a plot of ω_α/ω_1 versus m; a good numerical approximation is

$$\frac{\omega_\alpha}{\omega_1} \approx 1.2 + \frac{m}{10}. \tag{4.96}$$

4.3.2.2 The Extrinsic Transistor

When extrinsic elements are added to the equivalent circuit of a transistor with emitter current drive, the emitter transition capacitance c_{tE} alone affects the response significantly. This capacitance appears in shunt with C_1 in the *RC* transmission-line representation of the base region, and its effect depends on the relative sizes of c_B and c_{tE}. If a transistor has c_{tE} much larger than c_B, the transmission line has little effect on the total response; Fig. 4.9*b* is an excellent equivalent circuit, and

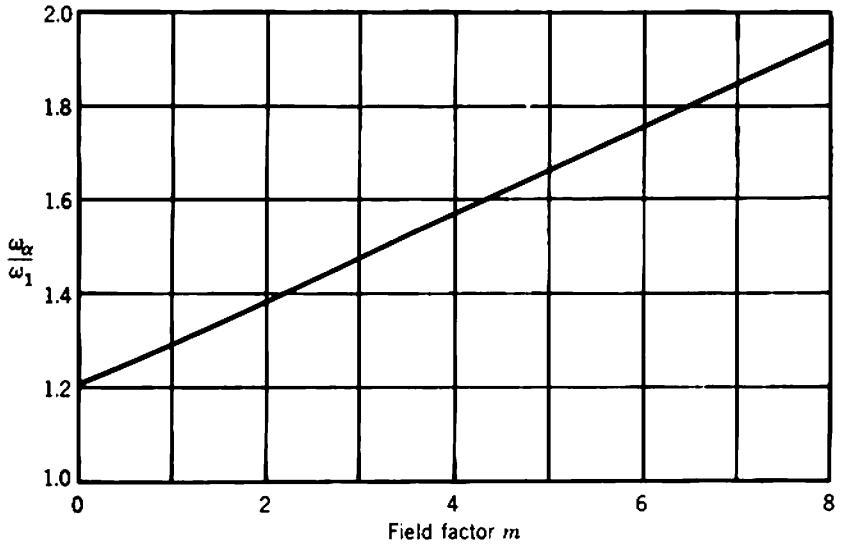


Fig. 4.18 Ratio of ω_α to ω_1 for an intrinsic transistor, as a function of the field factor.

the emitter-to-collector current gain approaches a semicircular locus similar to that of an ideal charge-controlled device. On the other hand, c_{tE} can be neglected if it is small compared with c_B, and the response approaches that of an intrinsic transistor. Most transistors at normal quiescent points approximate the latter condition.

Two characteristic frequencies are defined for an extrinsic transistor:

1. The *gain-bandwidth product* ω_T is introduced in Section 4.2.3.2 (Eqs. 4.71 and 4.73) in connection with the ideal charge-controlled transistor. At high frequencies the base-to-collector current gain of a practical transistor follows a 20-dB/decade asymptote (Eq. 4.69) out to $\omega_1/2$ or $\omega_1/4$ as appropriate for a uniform-base or graded-base transistor but then breaks away. The ratio between the breakaway frequency and ω_T depends on the ratio of c_{tE} to c_B [curves (*b*) and (*c*) in Fig. 4.16].

2. The α *cutoff frequency* corresponds to a 3-dB fall in emitter-to-collector current gain and is less than the α cutoff frequency of the intrinsic transistor by a factor that depends on the ratio of c_{tE} to c_B. The symbols f_α, ω_α, and τ_α are used for the extrinsic as well as the intrinsic transistor frequencies.

4.4 NOISE SOURCES IN TRANSISTORS

At frequencies above a few kilohertz the dominant components of noise in a transistor are the shot noise associated with each carrier stream and the thermal noise of the base resistance. All are white.

Figure 4.19 shows the three independent carrier streams in a transistor (Section 4.2.1.3). The useful stream flows from emitter to collector and constitutes the component of minority carrier current injected into the base that does not recombine. Associated with this carrier stream is a mean-square shot noise current

$$d(i_{NS}^2)_C = 2q\alpha_N I_E \, df = 4k(0.5T)g_m \, df. \tag{4.97}$$

The second form of Eq. 4.97 follows directly from Eq. 4.53 and makes an interesting comparison with Eq. 3.60 for a vacuum triode. A second carrier stream flows from emitter to base. Ideally, this carrier stream is simply the recombination current, but in practice there are also components due to reverse injection from base to emitter, thermal generation, and recombination in the emitter depletion layer. All mechanisms contribute additive components of shot noise and the total is

$$d(i_{NS}^2)_B = 2q(1 - \alpha_N)I_E \, df. \tag{4.98}$$

Finally, all components of the collector-to-base saturation current I_{CO} contribute shot noise:

$$d(i_{NS}^2)_S = 2qI_{CO} \, df. \tag{4.99}$$

Like a vacuum tube, an intrinsic transistor contains no dissipative elements, and, correspondingly, there are no sources of thermal noise. However, the extrinsic base resistance r_B is dissipative, and its thermal noise often degrades the total performance far below that of the intrinsic device. The mean-square noise voltage associated with r_B is

$$d(v_{NT}^2) = 4kTr_B \, df. \tag{4.100}$$

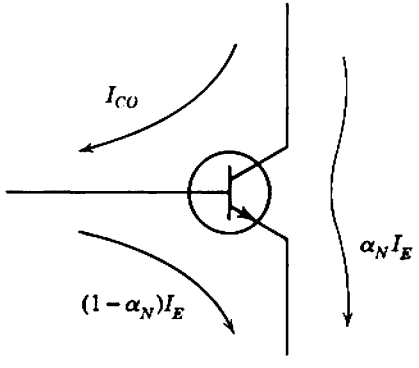


Fig. 4.19 Carrier current streams in an *n-p-n* transistor.

Figure 4.20 shows a noise model for use at mid and high frequencies obtained by superimposing these four physical noise generators on the hybrid-π small-signal equivalent circuit. The desirable characteristics of a low-noise transistor are

(i) α_N approaching unity, particularly at low emitter currents,
(ii) low saturation current I_{CO},
(iii) low base resistance r_B.

At frequencies below a few kilohertz the shot and thermal noise of a transistor are masked by flicker noise. Flicker noise in semiconductors appears to be associated with random changes in the rates of carrier generation and recombination, and any condition that results in a high concentration of traps or in high trap activity is likely to produce flicker noise. Empirically, the mean-square flicker noise current associated with a recombination current I is given by an expression of the form

$$d(i_{NF}^2) = \frac{KI^c}{f}\, df.$$

The exponent c is 1 or 2, depending on the dominant recombination mechanism. If recombination is due merely to a high concentration of traps (e.g., at the surface of the base region), c is about 2. However, if recombination is due to a low density of activated traps (e.g., in a depletion layer), c is about 1.

In a transistor the current stream from emitter to collector is the part of the injected minority carrier current that does not recombine. Accordingly, this current stream is substantially free from flicker noise. The

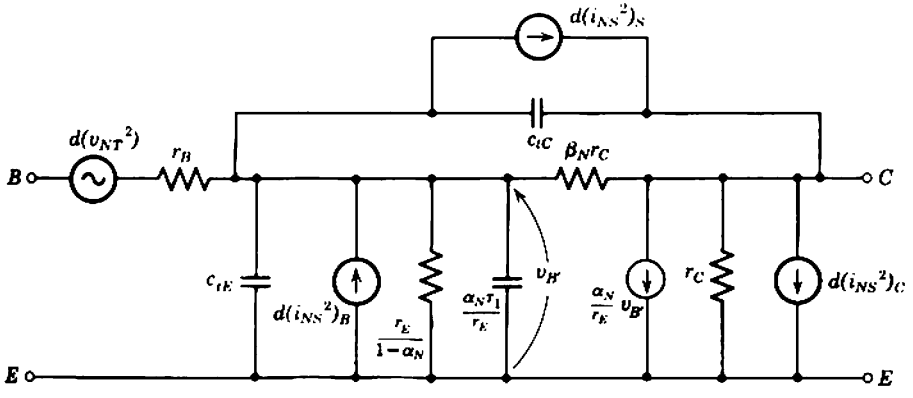


Fig. 4.20 Mid- and high-frequency noise model of a transistor.

emitter-to-base current stream, however, is almost entirely a recombination current, and the associated flicker noise is

$$d(i_{NF}^2)_B = \frac{KI_E^c}{f}\,df. \tag{4.101}$$

This noise current appears in shunt with $d(i_{NS}^2)_B$ in Fig. 4.20. The greatest part of the recombination in most modern transistors occurs in the emitter depletion layer, and the exponent c is therefore 1. Wide, uniform-base germanium transistors are the only common exceptions; recombination at the base surface dominates in these transistors, and c is about 2. The constant K varies over an extremely large range, even for transistors of the same type. For this reason transistors for low-noise applications at low frequencies must be selected either by the manufacturer or by the user.

The collector saturation current I_{CO} consists of a number of components (Section 4.3.1.4). Empirically, the components in the body of the junction are free from flicker noise, whereas the surface components are not. The surface component of flicker noise is somewhat akin to grid current noise in tubes; there is no simple way of predicting its value, nor is there any point in doing so because transistors for low-noise amplifiers are always selected to have good surfaces.

4.5 RATINGS OF TRANSISTORS

There is more uniformity in the methods used to specify the ratings of transistors than of vacuum tubes. Furthermore, the extensive use of absolute maximum ratings provides the user with precise information.

4.5.1 Power and Temperature Ratings

The most important limitation to the power dissipation in a transistor is imposed by the temperature rise of the crystal lattice near the junctions. The *maximum safe junction temperature* is in the range 85 to 100°C for germanium transistors and 150 to 200°C for silicon transistors. When a transistor is used as an amplifier, the power dissipated at the reverse-biased collector junction far exceeds the power dissipated at the forward-biased emitter junction. It is therefore usual for manufacturers to specify the maximum allowed *collector power dissipation* but not the maximum emitter dissipation. Because the thermal time constant of the junctions is low (often of the order of milliseconds), it is prudent to assume that a maximum power rating is a peak rating (unless, of course, the manufacturer provides information to the contrary).

The maximum safe collector dissipation for a transistor operating in an ambient temperature T_a is

$$P_{C(\max)} = \frac{T_{j(\max)} - T_a}{\theta}, \tag{4.102}$$

where $T_{j(\max)}$ is the safe junction temperature and θ is the thermal resistance. Manufacturers specify $T_{j(\max)}$ and either θ or $P_{C(\max)}$ for a particular ambient (usually 25°C). In the latter case

$$\theta = \frac{T_{j(\max)} - 25°\text{C}}{P_{C(\max)}}.$$

A quoted value of θ for a small transistor usually applies for operation in free air without any form of heat sink; typical values are in the vicinity of 0.5°C/mW. For larger high-power transistors the quoted value of θ applies when the device is used with a heat sink of specified size. In such circumstances the value of θ is determined principally by the surface area of the sink. Figure 4.21 shows the variation of θ with area for a square

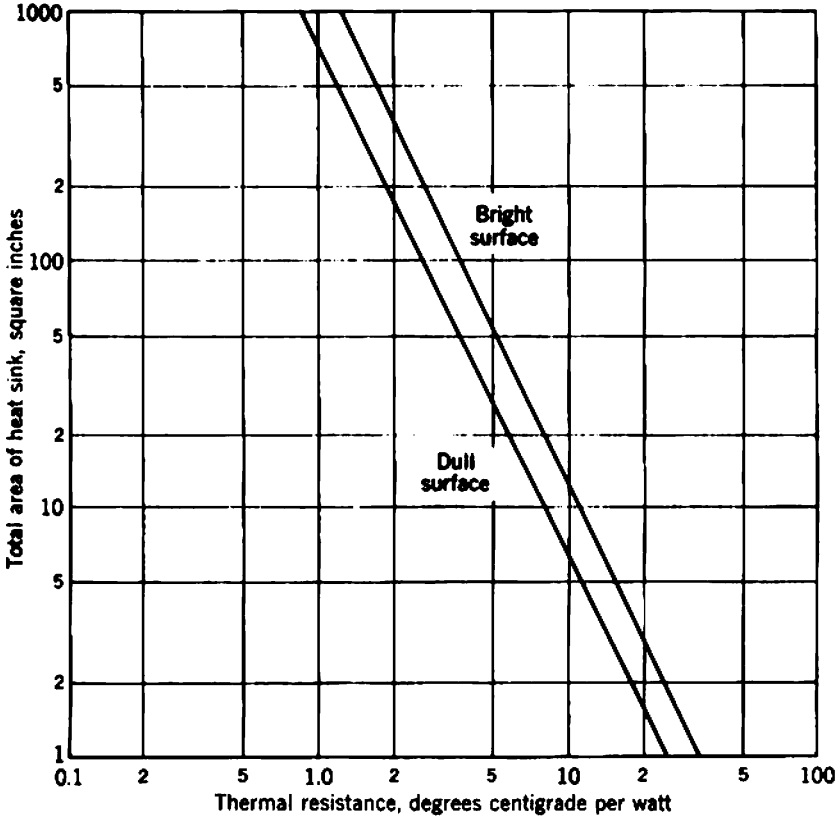


Fig. 4.21 Thermal resistance of a square aluminum heat sink about $\frac{1}{8}$-in. thick.

aluminum plate about $\frac{1}{8}$-in. thick, with both dull and bright surfaces. Changes in the shape of the plate, its material (i.e., thermal conductivity), and thickness have only small effects on θ, provided the changes are not so gross that they give a nonuniform temperature. The curves for dull and bright surfaces correspond to emissivities about 0.8 and 0.1, respectively, and apply for sink temperatures up to about 50 C. Radiation becomes appreciable at higher temperatures and θ falls; the change is greater for sinks with dull surfaces.

In addition to specifying a maximum safe junction temperature, manufacturers sometimes quote a minimum temperature rating. This rating is to avoid the onset of low-temperature deterioration of transistor materials. Furthermore, it should be noted that the uprating implied by Eq. 4.102 is usually restricted to ambient temperatures above 25°C. As the ambient temperature falls from $T_{j(\max)}$, the allowed power dissipation increases until T_a reaches 25°C. No further increase in power dissipation is allowed if T_a falls below 25°C.

4.5.2 Voltage Ratings

Manufacturers specify an upper limit to the *collector voltage* that should be applied to a transistor. This rating is based on considerations of one or more of the following phenomena; whichever occurs at the lowest voltage determines the rating.

1. *Avalanche multiplication* in the collector depletion layer (Section 4.3.1.2.3) causes I_C and α_N to increase with increasing collector voltage, and both become infinite at the avalanche breakdown voltage. The current amplification factor β_N is infinite at the collector voltage for which $\alpha_N = 1$ and becomes negative at higher voltages. Negative β_N causes the input resistance of a transistor to become negative, and oscillation can occur if the source resistance is high. A collector voltage rating based on avalanche multiplication therefore depends on the resistance in the base circuit. There is no inherent failure mechanism associated with avalanche multiplication, provided the multiplication and current do not become so large that the rated power dissipation is exceeded.

2. *Surface currents* are due to the presence of contamination on the collector junction surface (Section 4.3.1.4). These voltage-dependent components of I_{CO} are not associated with minority carrier currents in the base and so do not affect the values of α_N and β_N. However, the surface power dissipation increases with increasing collector voltage and at the surface breakdown voltage becomes large enough to destroy the transistor. The maximum safe surface power dissipation is somewhat less than the normal rated collector dissipation because the surface current tends to be

localized. Surface breakdown often occurs at a lower voltage than the body avalanche voltage, which makes it impossible to bias the collector into the region of high avalanche multiplication.

3. *Punch-through* occurs when the collector depletion layer extends throughout the entire control region (Section 4.3.1.1). Like avalanche multiplication, no inherent failure mechanism is associated with punch-through, provided the current and power dissipation are held within ratings. Punch-through does, of course, impose a limit to the normal mode of operation.

4. *Thermal runaway* is an unstable condition that occurs only with a combination of high collector voltage and careless design of the biasing circuit. The collector current depends on the parameters β_N, V_{BE}, and I_{CO}, and all are functions of temperature. If the temperature rises, β_N, V_{BE}, and I_{CO} change in the directions that increase the collector current. When the collector voltage exceeds a critical value, the resultant increase in collector dissipation is more than sufficient to maintain the initial rise in temperature and a regenerative condition is set up; the collector dissipation increases until the transistor is destroyed. The runaway voltage is as much a function of the biasing circuit as the transistor; Section 16.1.1 contains further information.

Manufacturers sometimes specify a maximum reverse base-to-emitter voltage. This rating is applicable to switching circuits and the class-B amplifiers discussed in Section 16.2.3 but not to the linear circuits that form the bulk of this book. The rating is based on avalanche or surface breakdown of the junction and may be as low as 1 or 2 V for graded-base transistors.

It is important to realize that there is no inherent failure mechanism associated with high voltages—only with the high temperatures that result from excessive power dissipation. If a circuit is such that the power dissipation can be guaranteed to remain within safe limits, manufacturers voltage ratings may be exceeded. The circuit designer must, of course, be prepared to accept the consequences of negative β_N, high I_{CO}, or an effective emitter-collector short circuit. However, manufacturers' voltage ratings are conservative, and these effects will not be observed unless the voltage rating is exceeded by a large amount.

4.5.3 Current Ratings

Maximum current ratings for transistors are of two types:

(i) based on considerations of excessive power dissipation and temperature rise, leading to destruction of the transistor;
(ii) based on the fall in β_N or rise in τ_1 at high currents.

Manufacturers seldom indicate whether a quoted maximum current is a true rating of type (i) or merely a limit of type (ii) to the normal range of operation. Specifications of type (ii) may be exceeded.

As with high voltages, there does not appear to be any inherent failure mechanism in germanium or silicon associated with high current densities, although this is not the case for some other semiconductors (notably gallium arsenide). Heating, either general or localized at a hot-spot, may damage the transistor if the current is maintained for any length of time, but so long as instantaneous junction temperatures remain within ratings, small transistors can safely be pulsed to currents as large as 10 A. If the collector voltage is so low that the junction power dissipation remains within safe bounds, a likely practical failure mechanism at high average currents is melting of the emitter lead wire near the point at which it is attached to the junction structure.

4.6 VARIATION OF TRANSISTOR PARAMETERS

The values of the parameters in the large- and small-signal equivalent circuits of a transistor (Figs. 4.8 and 4.9) depend mainly on the type of transistor, the chosen quiescent values of electrode currents and voltages, and the ambient temperature. However, because of the tolerances in manufacturing processes and aging effects (particularly at the surfaces), transistors bearing the same type number may show significant unit-to-unit parameter variations. Thus the circuit designer requires a knowledge of the probable magnitude of parameter variations with changes in quiescent currents and voltages and with transistor replacement. Some of this information can be deduced from the theoretical treatment in Section 4.2; the rest is provided by manufacturers' data, acceptance specifications of device users, measurements of actual transistor parameters, and experience. The aim of this section is to use these sources to establish design data that indicate the general trends and magnitudes of parameter variations.

4.6.1 Nominal Parameters of Representative Transistors

Transistors may be classified first according to their physical size. Related to this is their maximum permissible collector dissipation $P_{C(\max)}$. Typical small germanium transistors in free air at 25°C can safely dissipate about 100 mW, whereas the largest germanium transistors normally encountered can dissipate about 20 W. Silicon transistors can dissipate about three times as much power as physically similar germanium types because higher junction temperatures are allowed. Germanium transistors

are predominantly *p-n-p* structures because they are easier to manufacture, whereas silicon transistors are predominantly *n-p-n*. However, transistors of the opposite polarities are available.

Transistors can be further classified according to their intrinsic small-signal parameters τ_1 and β_N. [The other intrinsic parameters are not meaningful for classification purposes; r_E is known identically for all transistors (Eq. 4.38), r_C is so large that it is an insignificant parameter, and c_B is an equivalent specification to τ_1 if r_E and α_N are known (Eq. 4.55).] The transit time τ_1 depends on the base width and the electric field built into the base region (Eq. 4.48), and its value lies in the range 500 nsec–2 nsec for uniform-base transistors or 20 nsec–100 psec for graded-base transistors. The current amplification factor β_N depends on many factors (Section 4.3.1.2); typical values lie between 20 and 500, the corresponding values of α_N being 0.95 and 0.998. Notice that

$$\alpha_N \approx 1 \tag{4.103}$$

and

$$\frac{1}{1-\alpha_N} = \beta_N + 1 \approx \beta_N. \tag{4.104}$$

Transistors with small values of τ_1 are described as *high-frequency* types, and those with large values of β_N are described as *high-gain* types.

The intrinsic and extrinsic small-signal parameters of a transistor are functions of device geometry, and order-of-magnitude interrelations therefore exist between many of them.

First, high power dissipation capabilities demand large junction areas and therefore large extrinsic capacitances and small extrinsic resistances. The emitter and collector transition capacitances c_{tE} and c_{tC} are proportional to the junction area and, as a rule-of-thumb, vary in direct proportion to $P_{C(\max)}$; the base resistance r_B varies inversely with $P_{C(\max)}$. The intrinsic parameters are independent of junction area, but a small value of τ_1 requires a close emitter-to-collector spacing (i.e., base width W_B). It is difficult to manufacture large-area transistors with very close spacing; therefore the requirements of high power and high frequency are mutually incompatible. However, technological advances have allowed the manufacture of devices with complicated geometry (e.g., interdigited emitters), and the power limits for high-frequency transistors are continually being raised. Table 4.3 summarizes the relations between $P_{C(\max)}$ and the parameters of transistors with the same τ_1.

Second, because small base widths can be manufactured only in small-area structures, relations exist between τ_1 and the extrinsic elements r_B, c_{tE}, and c_{tC}. As a rule-of-thumb, r_B for transistors with the same $P_{C(\max)}$ varies as $\tau_1^{1/3}$, whereas c_{tE} and c_{tC} vary as $\tau_1^{1/2}$. Notice that c_{tE} is much

Table 4.3 Order-of-Magnitude Variation of Parameters with $P_{C(max)}$ (τ_1 being Constant)

r_B	$\propto P_{C(max)}^{-1}$
r_C	—
r_E	—
β_N	—
c_{tE}	$\propto P_{C(max)}$
c_{tC}	$\propto P_{C(max)}$
I_{CO}	$\propto P_{C(max)}$

larger for a graded-base transistor than a uniform-base transistor with the same τ_1; c_{tE} depends on the doping of the base region adjacent to the emitter (Eq. 4.64) and the doping is much heavier in a graded-base transistor. Table 4.4 lists typical values for the parameters of small transistors as functions of τ_1; it is most unlikely that the parameters of any transistor will differ from these typical values by more than a factor of 2. The base-charging capacitance c_B (which is an equivalent specification to τ_1) is given for comparison with the extrinsic capacitances.

The collector saturation current I_{CO} depends on the junction area and whether the transistor is a germanium or silicon type. For small germanium transistors at 25°C, I_{CO} is of the order of a few microamperes; for silicon transistors I_{CO} is a few nanoamperes.

4.6.2 Variation of Parameters with Quiescent Conditions

The first-order trends in the variation of transistor parameters with quiescent voltages and current can be predicted from the theory in Section 4.2. Table 4.5 summarizes the results and lists the reference equations. This first-order theory is adequate for normal design purposes.

Second-order theory shows that small perturbations about the ideal variation laws are to be expected, particularly at very low or very high currents and voltages. Table 4.5 gives references to the relevant paragraphs of Section 4.3.1. If a transistor manufacturer supplies a graph of any parameter versus quiescent conditions, this information should obviously be used in preference to the ideal laws in Table 4.5.

Two points require explanation. First, the base-to-emitter voltage drop varies as the logarithm of emitter current (Eq. 4.52), but over the range of I_E for which a transistor is useful in amplifier circuits V_{BE} is constant within about 100 mV. Equation 4.64 therefore shows that c_{tE} is approximately constant also. Second, the collector transition capacitance of a uhf transistor may be so small that it is masked by the constant interlead capacitance.

Table 4.4 Typical Parameters of Small Transistors as Functions of τ_1 ($P_{C(\max)}$ = 100 mW for Germanium, 300 mW for Silicon). Quiescent Conditions: I_E = 1 mA, V_{CE} = 6 V, T_a = 25°C

Parameter	Approximate Variation with τ_1	Uniform Base τ_1 = 100 nsec	Uniform Base τ_1 = 10 nsec	Graded Base τ_1 = 10 nsec	Graded Base τ_1 = 1 nsec	Graded Base τ_1 = 100 psec
r_B	$\propto \tau_1^{1/3}$	250 Ω	100 Ω	100 Ω	50 Ω	25 Ω
r_C	$\propto \tau_1^{1/2}$	150 kΩ	50 kΩ	500 kΩ	150 kΩ	50 kΩ
r_E	—	Identically $\frac{kT}{qI_E}$ for all transistors: $\frac{kT}{q}$ = 25 mV at 16°C; 26 mV at 28°C; 27 mV at 40°C				
β_N	—	20 to 500 for all transistors: low gain 20–50; medium gain 50–100; high gain 100–500				
c_{tE}	$\propto \tau_1^{1/2}$	150 pF	50 pF	150 pF	50 pF	15 pF
c_{tC}	$\propto \tau_1^{1/2}$	30 pF	10 pF	15 pF	5 pF	1.5 pF
c_B	$\propto \tau_1$	4000 pF	400 pF	400 pF	40 pF	4 pF

Table 4.5 Variation of Transistor Parameters with Quiescent Conditions

Parameter	Variation with I_E	Variation with V_{CE}	Reference
r_B	—	—	Section 4.3.1.3
r_C	$\propto I_E^{-1}$	$\propto V_{CE}^{a}$	Eqs. 4.59 and 4.62
r_E	$\propto I_E^{-1}$	—	Eq. 4.38; also Section 4.3.1.3
β_N	—	—	Section 4.3.1.2
τ_1	—	—	Section 4.3.1.1
c_B	$\propto I_E$	—	Eq. 4.55
c_{tE}	—	—	See text
c_{tC}	—	$\propto V_{CE}^{-a}$	Eq. 4.65—see text
V_{BE}	—	—	See text
I_{CO}	—	—	Section 4.3.1.4

The exponent a is 1/2 for transistors having an abrupt collector junction, but falls to 1/3 for linear-graded junctions.

Notice that, although c_{tE} is constant, c_B is directly proportional to I_E. At low currents, therefore, c_B is much smaller than c_{tE} and can be neglected in comparison. At high currents, however, c_{tE} can be neglected in comparison with c_B. The gain-bandwidth product ω_T for an extrinsic transistor is given by Eqs. 4.71 and 4.73, and by rearranging

$$\frac{1}{\omega_T} = \tau_T = \tau_1 + \frac{r_E c_{tE}}{\alpha_N} = \tau_1 + \left(\frac{c_{tE}}{\alpha_N}\right)\left(\frac{kT}{q}\right)\left(\frac{1}{I_E}\right). \tag{4.105}$$

The second term dominates at low currents so that

$$\omega_T \approx \left(\frac{\alpha_N}{c_{tE}}\right)\left(\frac{q}{kT}\right) I_E,$$

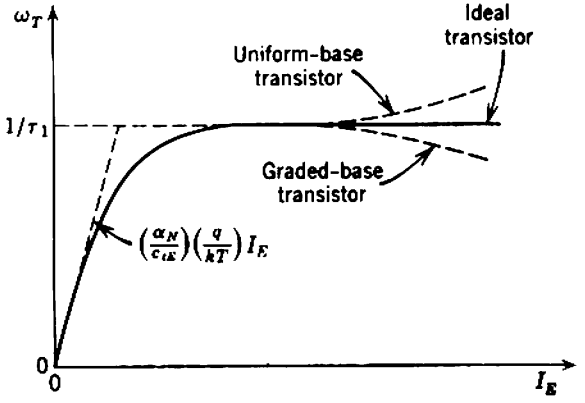


Fig. 4.22 Plot of ω_T against I_E, showing the asymptotes.

but the first term dominates at high currents so that

$$\omega_T \approx \frac{1}{\tau_1}.$$

Figure 4.22 shows a plot of ω_T versus I_E; the broken lines at high current indicate the behavior when the second-order variation of τ_1 with I_E is taken into account.

4.6.3 Variation of Parameters with Temperature*

Elementary transistor theory shows that some parameters exhibit a first-order dependence on temperature. Absolute temperature T occurs explicitly in the expressions for some, and is implicit in others through the mobility μ and diffusion constant D (Eqs. 4.23 and 4.24). Table 4.6 summarizes the dependence of transistor parameters on temperature and gives the reference equations.

The values of β_N, V_{BE}, and I_{CO} depend on many factors, all of which are functions of temperature. Invariably, β_N and I_{CO} rise with increasing temperature, whereas V_{BE} falls. Empirically, β_N varies as T^c, where the exponent c usually lies between 1 and 3; over the normal temperature range T^2 is a reasonable compromise. I_{CO} varies exponentially with

Table 4.6 Variation of Transistor Parameters with Absolute Temperature

Parameter	Variation with T	Reference
r_B	$\propto T^b$	Eq. 4.23 ($b \approx 2$)
r_C	—	Eqs. 4.59 and 4.62
r_E	$\propto T$	Eq. 4.38
β_N	$\propto T^c$	See text ($c \approx 2$)
τ_1	$\propto T$	Eqs. 4.24 and 4.48
c_B	—	Eq. 4.55
c_{tE}	—	Eq. 4.64
c_{tC}	—	Eq. 4.65
V_{BE}	$\left(\frac{\partial V_{BE}}{\partial T}\right)_{I_E} \approx -2.5$ mV/°C	See text
I_{CO}	$\propto \exp(T/a)$	See text: $a \approx 14$°C for Ge, $a \approx 8.5$°C for Si

* A thorough theoretical treatment of temperature effects is contained in W. W. GÄRTNER, *Transistors—Principles, Design, and Applications* (Van Nostrand, Princeton, 1960).

temperature, doubling every 10°C temperature rise for germanium devices and 6°C for silicon. V_{BE} varies linearly with temperature at a rate about -2.5 mV/°C.

4.6.4 Variation of Parameters with Transistor Replacement

The parameters of transistors of a given type operating under identical quiescent conditions show unit-to-unit variations. These variations occur between individual transistors manufactured in one batch, between average transistors from different production batches, and, most important of all, with transistor age.

Variations in the parameters of new transistors arise from the inherent inaccuracies associated with production techniques, principally in

(i) surface condition,
(ii) trap density in the emitter depletion layer,
(iii) electrode geometry, particularly the base width,
(iv) resistivity of the base region.

The spread in parameters varies with the type of transistor and the acceptance specification set by either the manufacturer or user. The user is seldom able to distinguish between intra- and interbatch variations and must therefore assume all-inclusive tolerances. In addition to manufacturing tolerances on the parameters of new transistors, the circuit designer must consider the effects of aging. The principal failure mechanism for transistors is a deterioration of the crystal surface, which manifests itself electrically as a fall in β_N and a rise in I_{CO}. Over-all, the tolerances listed in Table 4.7 provide a useful guide for long-life designs that use unselected transistors; the added subscript A indicates an "average" value, the value normally listed in manufacturer's data sheets. These tolerances apply at fixed values of emitter current and collector-to-emitter voltage.

When a number of transistors are manufactured together in a single chip of silicon (as in an integrated circuit), the tolerances on the absolute values of their parameters are of the order suggested by Table 4.7. The matching, however, tends to be very good. Parameters such as β_N and V_{BE} are closely equal for all transistors on the same chip, and they track together with variations in operating point and temperature.

Transistors from a single production run are sometimes separated into a number of different groups, each having a narrow spread in one parameter (usually β_N). Although such selection reduces the spread in this one parameter, the circuit designer still has to accommodate large spreads in the others. Furthermore, the tolerance on the low-spread parameter

Table 4.7 Tolerances on Transistor Parameters, Referred to Their "Average" Values

Parameter	Tolerance
Base resistance r_B	$0.7\, r_{BA} \leq r_B \leq 1.4\, r_{BA}$
Collector resistance r_C	$0.7\, r_{CA} \leq r_C \leq 1.4\, r_{CA}$
Emitter resistance r_E	$0.98\, r_{EA} \leq r_E \leq 1.02\, r_{EA}$
Current amplification factor α_N	$0.98\, \alpha_{NA} \leq \alpha_N \leq 1.02\, \alpha_{NA}$
Current amplification factor β_N	$0.7\, \beta_{NA} \leq \beta_N \leq 2.0\, \beta_{NA}$
Transit time τ_1	$0.7\, \tau_{1A} \leq \tau_1 < 1.4\, \tau_{1A}$
Base-charging capacitance c_B	$0.7\, c_{BA} \leq c_B \leq 1.4\, c_{BA}$
Emitter transition capacitance c_{tE}	$0.7\, c_{tEA} \leq c_{tE} \leq 1.4\, c_{tEA}$
Collector transition capacitance c_{tC}	$0.7\, c_{tCA} \leq c_{tC} \leq 1.4\, c_{tCA}$
Base-to-emitter voltage drop V_{BE}	± 50 mV
Collector saturation current I_{CO}	$0.2\, I_{COA} \leq I_{CO} \leq 5.0\, I_{COA}$

often increases with time to the order suggested in Table 4.7. It is therefore prudent to design amplifiers that will accommodate large spreads in transistor parameters; this leads to greater reliability, to a reduction in the number of basic device types, and to an improvement in over-all economy.

4.7 DATA SHEETS AND ALTERNATIVE PARAMETER SPECIFICATIONS

In view of the tolerance on transistor parameters, the classification of Section 4.6.1 suggests that there are not more than a few dozens of basically different transistor types, yet manufacturers' catalogs list thousands of type numbers. Some of this duplication of basic types occurs because electrically similar transistors may have different ratings. A little more occurs because the same junction structure may be mounted in several different encapsulations. Nevertheless, the fact remains that many transistor types at present in production are redundant.

Many virtually identical transistors are made to appear dissimilar by presenting their data for different equivalent circuits or data taken at different operating conditions. A very real problem confronting the circuit designer is the sorting out of information provided by manufacturers and calculating from it some meaningful data that will allow comparison of transistors. The advantages of the charge-control or hybrid-π equivalent circuit that justify its use in this book are set out in Section 2.1. Perhaps the most important is that its parameters vary in simple and predictable ways with quiescent conditions and temperature. Once the parameters are known for a reference set of conditions they can be converted to any other conditions by means of Tables 4.5 and 4.6.

The following paragraphs list some methods for computing the hybrid-π parameters from data that are likely to be supplied by manufacturers. The task is not easy. If the general nature of a transistor is known, an intelligent guess is often the best way for arriving at its parameters. Even measurement of half a dozen transistors may be quicker and more rewarding than a computation. Low-frequency (resistive) and high-frequency (capacitive) parameters may be considered separately.

4.7.1 Low-Frequency Parameters

At least five low-frequency equivalent circuits are in common use:

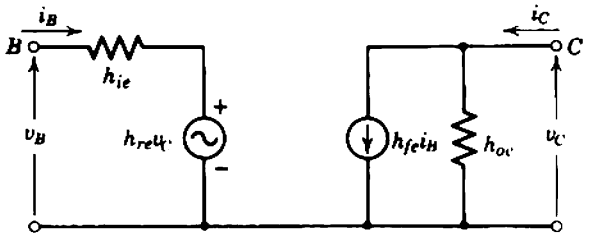
(i) hybrid-π,
(ii) common-emitter h,
(iii) common-base h,
(iv) T,
(v) y (or g).

Figure 4.23 shows the last four. The common-emitter and common-base h-parameters are formal models of 4-terminal network theory and bear no relation to device physics; the y-parameters are also a formal model, but they and the T-parameters have a physical basis. Table 4.8 lists the useful relations between the hybrid-π parameters and the parameters of the other models. Notice that Table 4.8 gives the hybrid-π parameters at the quiescent point and temperature applicable to the original data.

Strictly, Table 4.8 is meaningful only for the parameters of a particular transistor, whereas the data supplied by manufacturers are the averages of the parameters of a number of transistors. It is possible for errors to occur in the value of r_B due to inconsistencies introduced by this averaging process; r_B is obtained by subtracting two nearly equal quantities, and a small error in either one will produce a gross error in r_B. The typical values in Table 4.4 can be used as a check. Some manufacturers quote the value of r_B as well as the main data in h-, T-, or y-parameters; a variety of notation is used—base spreading resistance, ohmic base resistance, r'_b, and $r_{bb'}$ are the most common.

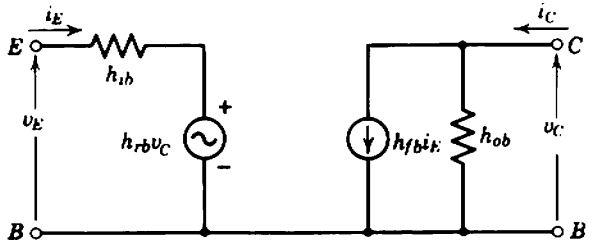
Errors can occur in Table 4.8 if the original data are taken at a very high or very low emitter current. First, Section 4.3.1.3 shows that the slope resistance of the intrinsic base-emitter diode is not given exactly by Eq. 4.38. In addition, the extrinsic emitter spreading resistance introduced in Section 4.2.3.1 is in series with the diode and increases its effective slope resistance:

$$r_E = \frac{\sigma(I_E)\, kT}{q(I_E - I_{EO})} + r_{E(\text{spread})}. \tag{4.106}$$



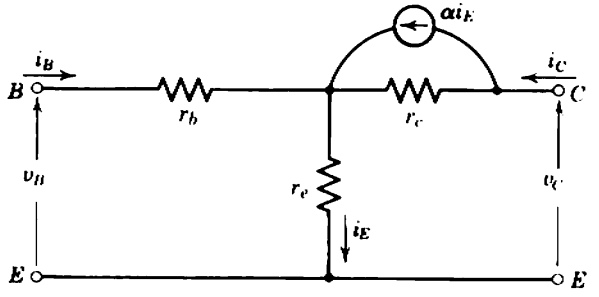
Common-Emitter h:

$$v_B = h_{ie}i_B + h_{re}v_C,$$
$$i_C = h_{fe}i_B + h_{oe}v_C.$$



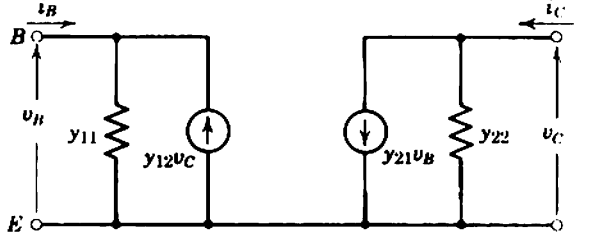
Common-Base h:

$$v_E = h_{ib}i_E + h_{rb}v_C,$$
$$i_C = h_{fb}i_E + h_{ob}v_C.$$



T:

$$v_B \approx i_B\left(r_b + \frac{r_e}{1-\alpha}\right) + v_C\left[\frac{r_e}{(1-\alpha)r_c}\right],$$
$$i_C = i_B\left(\frac{\alpha}{1-\alpha}\right) + v_C\left[\frac{1}{(1-\alpha)r_c}\right].$$



y (or g):

$$i_B = y_{11}v_B - y_{12}v_C,$$
$$i_C = y_{21}v_B + y_{22}v_C.$$

Fig. 4.23 Alternative low-frequency equivalent circuits of a transistor.

Table 4.8 Values of Hybrid-π Low-Frequency Parameters in Terms of Other Common Data Systems

Hybrid π Parameter	Common-Emitter h Parameters	Common-Base h Parameters	T-Parameters	y-Parameters
r_B	$h_{ie} - (h_{fe} + 1)\dfrac{kT}{qI_E}$	$\left(h_{ib} - \dfrac{kT}{qI_E}\right)\left(\dfrac{1}{1 - h_{fb}}\right)$	$r_b - \dfrac{\alpha}{2(1 - \alpha)}\left(\dfrac{kT}{qI_E}\right)$	$\dfrac{1}{y_{11}} - \dfrac{kT}{qI_E}\left(\dfrac{y_{21}}{y_{11}} + 1\right)$
r_C	$\dfrac{2}{h_{oe}}$	$\dfrac{2(1 - h_{fb})}{h_{ob}}$	$2r_c(1 - \alpha)$	$\approx \dfrac{1}{y_{22}}$
r_E		Identically $\left(\dfrac{kT}{qI_E}\right)$ for all transistors:	$\dfrac{kT}{q} = \begin{cases} 25 \text{ mV at } 16°C \\ 26 \text{ mV at } 28°C \\ 27 \text{ mV at } 40°C \end{cases}$	
β_N	h_{fe}	$\dfrac{h_{fb}}{1 - h_{fb}}$	$\dfrac{\alpha}{1 - \alpha}$	$\dfrac{y_{21}}{y_{11}}$

Note. Some manufacturers give some parameters a negative sign. All quantities substituted into this table should be positive.

The contribution of $r_{E(\text{spread})}$ is significant only at very high currents for which the slope resistance of the intrinsic diode is small. In the expressions for r_B in Table 4.8 (kT/qI_E) should be replaced by the total slope resistance r_E. Second, the extrinsic collector leakage conductance (also introduced in Section 4.2.3.1) is in shunt with g_f in the small-signal equivalent circuit (Fig. 4.9*a*) and raises the total conductance above $(1/\beta_N r_C)$. This effect is most noticeable at low emitter currents for which g_f is small. If the leakage conductance is significant, Table 4.8 will yield a low value for r_C, but this error can almost always be neglected in amplifier design (see Section 5.3.2, for example).

The collector saturation current I_{CO} is usually specified in manufacturers' data sheets but may be given a negative sign; I'_{CO} is also specified sometimes. Common alternative notations for I_{CO} and I'_{CO} are I_{CBO} and I_{CEO}, respectively. The base-to-emitter voltage drop V_{BE} may be specified at a particular emitter current, but the values 200 and 700 mV for germanium and silicon transistors, respectively, are accurate enough for most purposes.

4.7.2 High-Frequency Parameters

The elements required to complete the hybrid-π equivalent circuit at high frequencies are the base-charging capacitance, emitter transition capacitance, and collector transition capacitance. With the exception of c_{tC}, the values of these elements are almost never quoted explicitly by manufacturers and must be deduced from other data.

The *collector transition capacitance* c_{tC} is numerically equal to the output capacitance in the common-base configuration (but not in common-emitter), and this is usually specified on transistor data sheets. The most usual notation is C_{ob}.

The *emitter transition capacitance* c_{tE} is sometimes quoted for zero or some reverse bias on the base-emitter junction for which the transistor is cut off. This value of c_{tE} is quite useless for amplifier design; Eq. 4.64 shows that it is different by at least an order of magnitude from the value under forward bias conditions.

The *sum of* c_B *and* c_{tE} can be found from the gain-bandwidth product ω_T at a particular dc emitter current I_E. (The gain-bandwidth product f_T in hertz is more commonly quoted than ω_T in radians per second.) From Eq. 4.71

$$c_B + c_{tE} = \frac{\alpha_N \tau_T}{r_E} = \alpha_N \left(\frac{q}{kT}\right) \frac{I_E}{\omega_T}. \tag{4.107}$$

If ω_T is quoted at several values of I_E, then c_B and c_{tE} can be separated by plotting $(1/\omega_T)$ against $(1/I_E)$. Equation 4.105 shows that such a plot is a

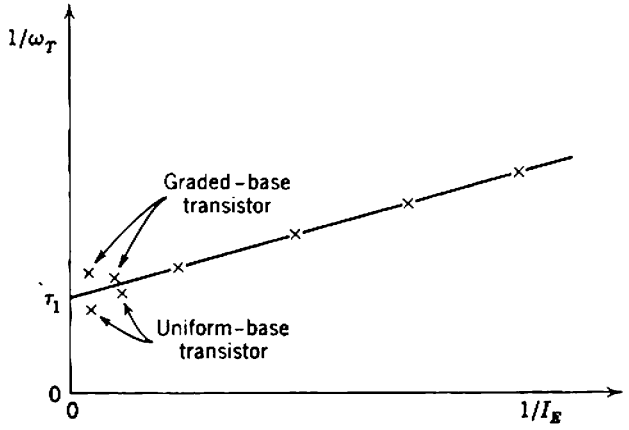


Fig. 4.24 Plot of $(1/\omega_T)$ against $(1/I_E)$, to find τ_1 and c_{tE}.

straight line; the intercept is τ_1 and the gradient is proportional to c_{tE} (Fig. 4.24). The departure from a straight line close to the origin is due to the variation of τ_1 at high currents. Plotting $(1/\omega_T)$ against $(1/I_E)$ is probably the most useful method for finding the high-frequency parameters of a transistor.

4.7.2.1 Characteristic Frequencies

In addition to the gain-bandwidth product ω_T, three other frequencies may be specified for the extrinsic transistor:

ω_α defined and discussed in Section 4.3.2.2,
ω_β and ω_C defined below.

(Characteristic frequencies for the intrinsic transistor are almost never specified.) The general procedure is to find ω_T from these frequencies and then deduce c_B and c_{tE} from ω_T. Notice that ω_T and ω_α are often confused in the literature and may even be equated. In determining the high-frequency parameters from a characteristic frequency, the circuit designer must first ascertain exactly which frequency is meant by the notation adopted in the data sheet.

The ratio of ω_α to ω_T lies between 1 and 2 and depends on the value of the field factor m and the ratio of c_B to c_{tE}. At low emitter currents c_B is much smaller than c_{tE} and

$$\omega_\alpha \approx \omega_T,$$

independent of m. At high emitter currents c_B becomes much larger than c_{tE}, so that $\omega_T \approx \omega_1$. Therefore for a uniform-base transistor ($m = 0$)

$$\omega_\alpha \approx 1.2\,\omega_T,$$

whereas for a graded-base transistor ($m = 8$)

$$\omega_\alpha \approx 1.9\,\omega_T.$$

The circuit designer must make an intelligent guess at the ratio appropriate for a given transistor type and quiescent conditions.

The *β cutoff frequency* is the frequency at which the short-circuit base-to-collector current gain falls by 3 dB from its low-frequency value. From Eq. 4.69

$$\omega_\beta = \frac{1}{\beta_N \tau_T} = \frac{\omega_T}{\beta_N}. \tag{4.108}$$

The following nomenclature is introduced for use in later chapters:

$$2\pi f_\beta = \omega_\beta = \frac{1}{\tau_\beta}. \tag{4.109}$$

The characteristic frequency ω_C is defined as

$$\omega_C = \frac{\alpha_N}{r_E c_C}, \tag{4.110}$$

where

$$c_C = c_B + c_{tE} + c_{tC}(1 + |A_V|) \tag{4.111}$$

and A_V is the voltage gain of the transistor. For reasons discussed in Section 7.5.2.2, c_C is the total effective input capacitance of a transistor and ω_C appears in the expression for realizable gain-bandwidth product. Additional nomenclature is introduced:

$$2\pi f_C = \omega_C = \frac{1}{\tau_C}. \tag{4.112}$$

4.8 THE FIELD-EFFECT TRANSISTOR

The use of the field-effect (or unipolar) transistor as an amplifier was proposed* some three years after the initial proposal of the bipolar transistor. However, field-effect transistors (abbreviated to f.e.t.) did not become available in production quantities until about 10 years after the bipolar transistor was introduced. In part, this production delay reflects on the fact that the f.e.t. is not a serious competitor with the bipolar transistor in the great majority of applications. The superiority of the bipolar device stems from the small spacing between the controlled and controlling charges; the two coexist in the base region, and consequently the capacitance c_1 and mutual conductance g_m ($\propto c_1/\tau_1$) are large. The two sets of charge in a unipolar device are separated by a high-field region,

* W. SHOCKLEY, "A unipolar 'field-effect' transistor," *Proc. Inst. Radio Engrs.*, **40**, 1365, November 1952.

and c_1 and g_m are both relatively small unless the high-field region is reduced to a few atomic thicknesses. This reduction leads to very fragile devices. Nevertheless, the f.e.t. is useful as a special-purpose device for amplifiers that require a very high input impedance (comparable with that of a vacuum tube) or a good noise performance when fed from a high-impedance source.

A number of different types of f.e.t. structure have been developed in which the controlling and controlled charges are separated by different types of high-field region. As examples, the charges in the insulated-gate type are separated by an insulating barrier, whereas the charges in the junction type are separated by the depletion layer of a reverse-biased *p-n* junction. Bockemuehl* has presented an elegant general derivation of the characteristics of all f.e.t. structures, but it is perhaps preferable to base an introductory treatment on particular types with simple geometry. The insulated-gate f.e.t. is a good choice for the initial discussion of Section 4.8.1 because it is simple to analyze, somewhat more of a general case than the junction f.e.t., and has great potential as a circuit element. However, the *p-n* junction f.e.t. is at present more readily available; its marginally more complex analysis follows in Section 4.8.2. It eventuates that these two physically dissimilar devices have nearly identical transfer characteristics, so that elementary f.e.t. theory can be generalized without appreciable loss of accuracy (Section 4.8.3). The treatment concludes with discussions of noise, temperature effects, and ratings.

4.8.1 Elementary Theory of the Insulated-Gate F.e.t.

Figure 4.25 depicts an *n*-channel insulated-gate f.e.t. with plane-parallel geometry. The channel is of length L and cross-sectional area A_C and has nonrectifying contacts at the ends for the source and drain. Separated from the channel by an insulating barrier of thickness W_I is the metallic gate electrode.

The channel of an insulated-gate f.e.t. is made extremely thin. If its width is designated by δ, the depth (into the page) is (A_C/δ). In practice the channel usually takes the form of a layer of *n*-type material at the interface between an intrinsic silicon substrate and an insulating layer of silicon dioxide. Because of their "sandwich" nature, insulated-gate f.e.t.s are often known as metal-oxide-semiconductor transistors (abbreviated to m.o.s.t.). These structures can be fabricated by solid-state diffusion or thin-film deposition techniques; devices made by the thin-film process are sometimes known as thin-film transistors (t.f.t.).

* R. R. BOCKEMUEHL, "Analysis of field-effect transistors with arbitrary charge distribution," *Inst. Elec. Electronics Engrs. Trans.*, **ED-10**, 31, January 1963.

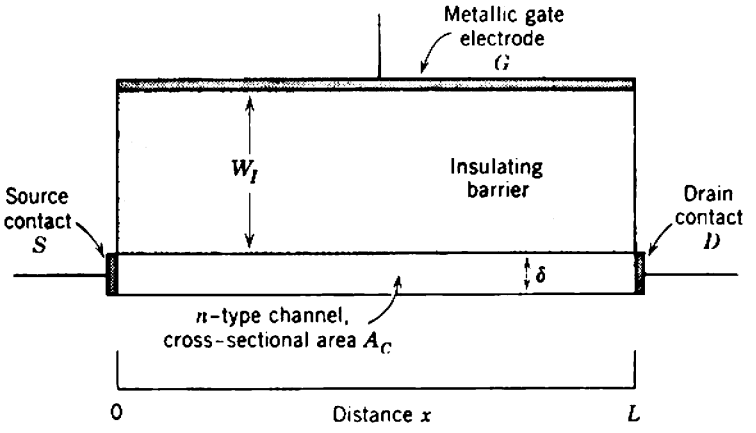


Fig. 4.25 Idealized model of an *n*-channel insulated-gate f.e.t.

4.8.1.1 Carrier Distribution in the Channel

In the absence of electrode voltages the electron (majority carrier) concentration at any point x in the channel is determined by the doping density. If we make the assumption that the donor concentration N_D is constant throughout the channel, the electron concentration is constant also:

$$n_n(x) = N_D. \tag{4.113}$$

However, when voltages V_G and V_D are applied to the gate and drain electrodes, additional charge is induced in the channel. The gate-to-channel capacitance per unit area is

$$C = \frac{\varepsilon}{W_I}$$

and the charge induced per unit area is

$$\sigma_I(x) = C\,V(x),$$

where $V(x)$ is the voltage drop across the insulating barrier. Hence the average electron concentration induced in the channel at x is

$$\bar{n}_I(x) = -\frac{\sigma_I(x)}{q\delta} = -\frac{\varepsilon\,V(x)}{q\,\delta W_I}. \tag{4.114}$$

The total electron concentration is

$$\bar{n}(x) = n_n(x) + \bar{n}_I(x)$$

and, using Eqs. 4.113 and 4.114,

$$\bar{n}(x) = N_D\left[1 - \frac{V(x)}{V_P}\right], \tag{4.115}$$

where

$$V_P = \frac{q\,\delta N_D W_I}{\varepsilon}. \tag{4.116}$$

The voltage V_P is known as the *pinch-off voltage* and is the voltage drop required across the insulating barrier to reduce the average electron concentration to zero.

Notice particularly that $\bar{n}_i(x)$ and $\bar{n}(x)$ represent the average concentrations of electrons in the thin-film channel at distance x from the source electrode. The channel is thin in comparison with the insulator, but not infinitely thin, and the actual carrier concentration varies across the channel as shown in Fig. 4.26. As $V(x)$ is taken toward the pinch-off voltage V_P, the channel begins to deplete from the gate side; this constriction of the conducting channel extends further into the channel until complete when $V(x) = V_P$.

The drain voltage V_D produces an electric field along the channel (in the negative x-direction for an n-channel device for which V_D is positive), and carrier motion is predominantly electron drift in this electric field. Equation 4.14*b* gives the mean current density across a plane at x as

$$\bar{J}_e(x) = q\mu_e \bar{n}(x)\, \mathscr{E}(x);$$

that is,

$$\bar{J}_e(x) = -q\mu_e \bar{n}(x) \frac{d[V_C(x)]}{dx},$$

where $V_C(x)$ is the potential in the channel. Now the voltage drop across the insulating barrier is

$$V(x) = V_C(x) - V_G;$$

hence

$$\frac{d[V(x)]}{dx} = \frac{d[V_C(x)]}{dx}.$$

Thus the mean electron current density in the channel is

$$\bar{J}_e(x) = -q\mu_e \bar{n}(x) \frac{d[V(x)]}{dx}$$

and, using Eq. 4.115,

$$\bar{J}_e(x) = \left(\frac{q\mu_e V_P}{N_D}\right) \bar{n}(x) \frac{d[\bar{n}(x)]}{dx}. \tag{4.117}$$

Equation 4.117 may be compared with Eq. 4.32 for the electron current density in an *n-p-n* bipolar transistor. Notice that the current density and electron concentration are constants across any plane normal to the x-direction of the bipolar transistor; $J_e(x)$ and $n(x)$ are actual values in the bipolar transistor, whereas $\bar{J}_e(x)$ and $\bar{n}(x)$ are mean values in the insulated-

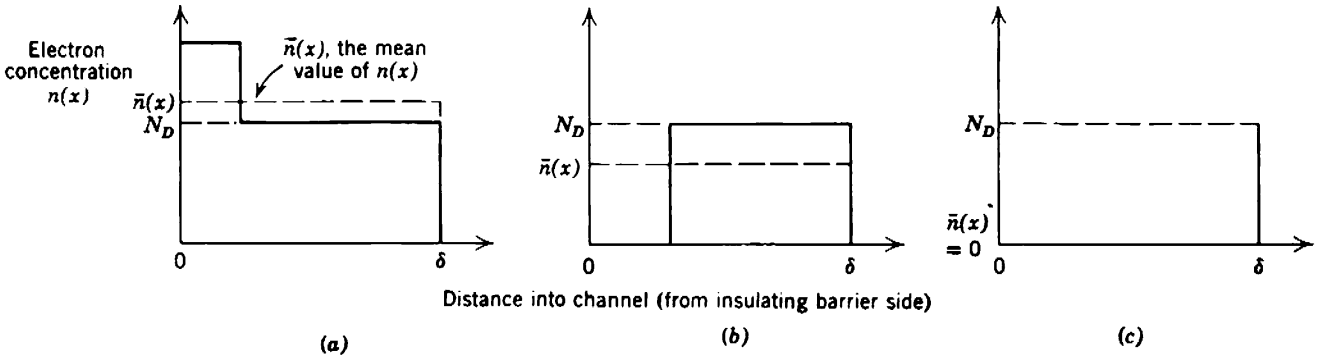


Fig. 4.26 Electron distribution in the channel: (*a*) enhancement, $V_G > 0$; (*b*) depletion, $V_G < 0$; (*c*) pinch-off, $V_G = -V_P$.

gate f.e.t. The total electron current across any plane in the channel of the f.e.t. is

$$I_e(x) = \bar{J}_e(x)A_C. \tag{4.118}$$

This current is sensibly constant at all points along the channel and equal to the drain current I_D because no carriers can penetrate the insulating barrier:

$$I_D = -I_e(x) = -\left(\frac{qA_C\mu_e V_P}{N_D}\right)\bar{n}(x)\frac{d[\bar{n}(x)]}{dx}. \tag{4.119}$$

The minus sign appears in Eq. 4.119 because the reference direction chosen for I_D is into the device. With this sign convention, I_D is positive for an n-channel f.e.t.

At the source end of the channel the voltage drop across the insulating barrier is

$$V(0) = -V_G$$

and the mean electron concentration at the source follows from Eq. 4.115 as

$$\bar{n}(0) = N_D\left(1 + \frac{V_G}{V_P}\right). \tag{4.120}$$

Integration of Eq. 4.119 with Eq. 4.120 as a boundary condition yields the mean electron concentration as

$$[\bar{n}(x)]^2 = N_D{}^2\left[\left(1 + \frac{V_G}{V_P}\right)^2 - x\left(\frac{2I_D}{qA_C\mu_e N_D V_P}\right)\right]. \tag{4.121}$$

4.8.1.2 Constriction of the Channel

Provided I_D is finite, Eq. 4.121 shows that the average electron concentration decreases from the source ($x = 0$) to the drain ($x = L$). Figure 4.27*a* indicates the effect of increasing drain current on $\bar{n}(x)$ when the gate voltage V_G is positive. The concentration is enhanced near the source at all current levels, and operation with positive V_G is therefore known as *enhancement operation*. At low currents the enhancement persists throughout the length of the channel, but at high currents the enhancement changes to depletion near the drain. Figure 4.27*b* indicates the effect of increasing drain current when V_G is negative. Here the concentration is depleted at all points along the channel, and operation with negative V_G is known as *depletion operation*. Finally, Fig. 4.27*c* depicts full depletion at the source, obtained by setting

$$V_G = -V_P.$$

Study of Fig. 4.27 shows that full depletion of the channel can be obtained in two distinct ways. In either case full depletion represents

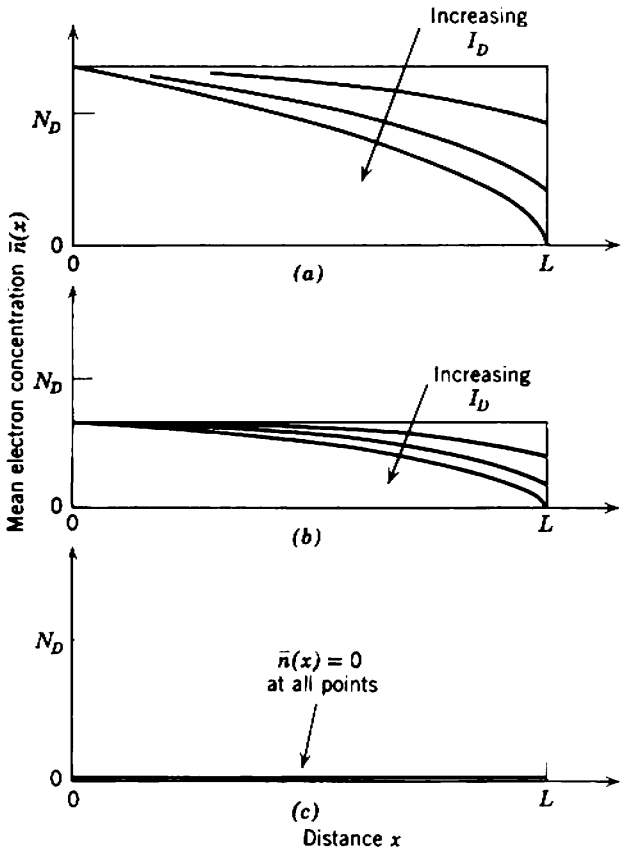


Fig. 4.27 Mean electron concentration along the channel: (*a*) $V_G > 0$; (*b*) $V_G < 0$; (*c*) $V_G = -V_P$.

both a change in the mode of operation of the device and a limit to the validity of the derivation of Eq. 4.121.

1. *Cutoff* occurs when the gate voltage is taken to $-V_P$. The channel is depleted of mobile carriers throughout its entire length (Fig. 4.27*c*) and no drain current can flow.

2. *Pinch-off* occurs when the drain current is increased to the point at which the channel is fully depleted at the drain ($x = L$). If $\bar{n}(L) = 0$, Eq. 4.115 will require that the voltage across the insulating barrier be

$$V(L) = V_P;$$

hence the drain voltage must be

$$V_D = V_P + V_G.$$

Moreover, if $\bar{n}(x) = 0$ and $x = L$ are substituted into Eq. 4.121,

$$I_D = I_{DO}\left(1 + \frac{V_G}{V_P}\right)^2, \tag{4.122}$$

where

$$I_{DO} = \frac{qA_C\mu_e N_D V_P}{2L}. \tag{4.123}$$

The current I_{DO} is the current that flows at pinch-off for zero gate voltage.

There is a formal inconsistency between the finite current predicted at pinch-off by Eq. 4.122 and the zero electron concentration at the drain predicted by Eq. 4.115. The inconsistency arises from extrapolating the theory for finite $\bar{n}(x)$ to the limiting case of zero $\bar{n}(x)$; it suggests that the potential gradient must be infinitely large over the infinitely small length of pinched-off channel. Despite this inconsistency, the pinched-off current predicted by Eq. 4.122 agrees well with practice. Indeed, if the drain voltage is increased beyond the value necessary for pinch-off, the drain current remains almost constant at the value predicted by Eq. 4.122. A formal investigation of the charge and potential distributions in the pinched-off region involves a two-dimensional solution of Poisson's equation, which is certainly not warranted in an elementary discussion because changes in charge and potential within the small constricted region of the channel have almost no effect on the distributions elsewhere. The distribution of $\bar{n}(x)$ is given by Eq. 4.121 except in the immediate vicinity of the constriction at $x = L$. Near this point $A_C \int \bar{n}(x)\, dx$ is small but finite over a small but finite distance in the x-direction, and any drain voltage applied in excess of the pinch-off value $(V_P + V_G)$ is absorbed by small changes in $\bar{n}(x)$ over a very small fraction of the channel length. The bulk of the mobile charge is quite unaffected.

4.8.1.3 Drain Volt-Ampere Characteristics, Q and τ_1

Within the region bounded at cutoff by

$$V_G \geq -V_P$$

and at pinch-off by

$$V_D \leq V_P + V_G$$

the drain volt-ampere characteristics can be obtained by integrating Eq. 4.119. Replacement of $\bar{n}(x)$ from Eq. 4.115 yields

$$I_D \int_0^L dx = qA_C\mu_e N_D \int_{-V_G}^{V_D - V_G} \left[1 - \frac{V(x)}{V_P}\right] d[V(x)].$$

Carrying out the integration and using Eq. 4.123 in the subsequent algebra, we obtain

$$I_D = I_{DO}\left[\frac{2V_D(V_P + V_G) - V_D^2}{V_P^2}\right]. \tag{4.124}$$

The left-hand portion of the drain characteristics in Fig. 4.28 (corresponding to operation below pinch-off) follows Eq. 4.124. The right-hand portion (corresponding to operation above pinch-off) is a set of constant current lines based on the extrapolation of Eq. 4.122.

The mobile charge in the channel (the controlled charge Q of charge-control theory) is

$$Q = -qA_C \int_0^L \bar{n}(x)\, dx.$$

From Eqs. 4.121 and 4.123 the electron distribution is

$$\bar{n}(x) = N_D\left[\left(1 + \frac{V_G}{V_P}\right)^2 - \frac{x}{L}\left(\frac{I_D}{I_{DO}}\right)\right]^{1/2};$$

hence

$$Q = -\frac{2qA_CLN_DI_{DO}}{3I_D}\left\{\left(1 + \frac{V_G}{V_P}\right)^3 - \left[\left(1 + \frac{V_G}{V_P}\right)^2 - \frac{I_D}{I_{DO}}\right]^{3/2}\right\}. \tag{4.125}$$

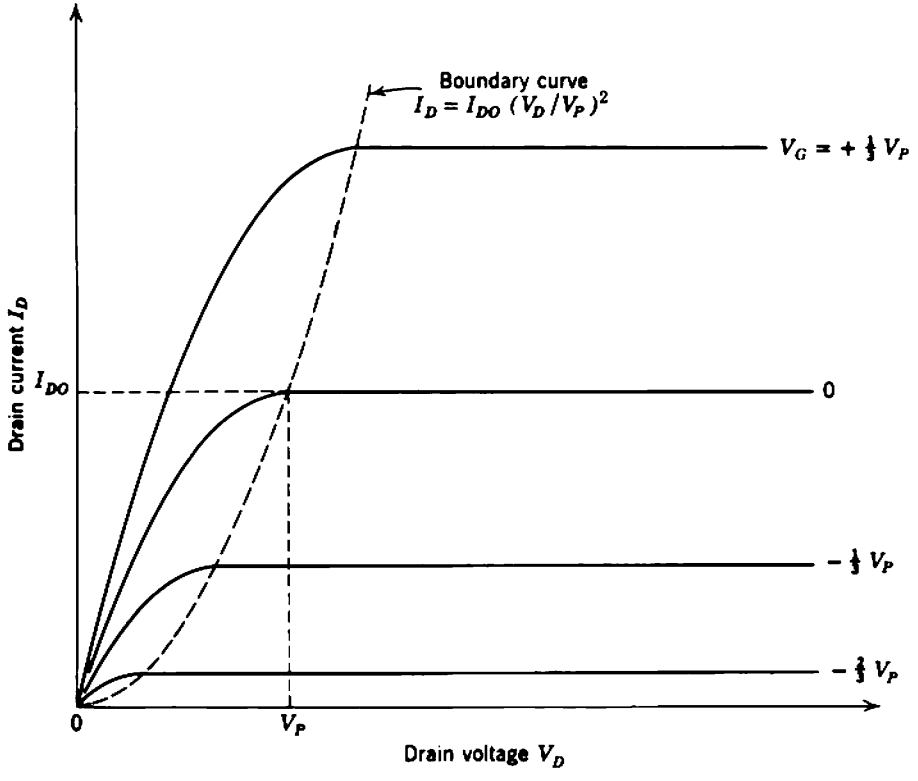


Fig. 4.28 Drain characteristics of an ideal insulated-gate f.e.t.

Because the drain current I_D is a function of both V_G and V_D (Eq. 4.124), Q is, in general, a complicated function of the electrode voltages. However, a field-effect transistor, like a pentode vacuum tube, is normally operated in the region of its characteristics in which the current is substantially independent of collecting-electrode voltage. In other words, an f.e.t. is normally operated above pinch-off. Equation 4.122 therefore applies and

$$\frac{I_D}{I_{DO}} = \left(1 + \frac{V_G}{V_P}\right)^2,$$

so that Q becomes

$$Q = -\frac{2qA_CLN_D}{3}\left(1 + \frac{V_G}{V_P}\right). \tag{4.126}$$

The mean transit time of electrons through the channel of an f.e.t. is

$$\tau_1 = -\frac{Q}{I_D}.$$

Substituting from Eqs. 4.122 and 4.126 for operation above pinch-off and using Eq. 4.123 in the subsequent algebraic manipulations, we find that the transit time becomes

$$\tau_1 = \frac{4L^2}{3\mu_e}\left(\frac{1}{V_P + V_G}\right). \tag{4.127}$$

For the special case of $V_G = 0$,

$$\tau_1|_{V_G=0} = \frac{4L^2}{3\mu_e V_P}. \tag{4.128}$$

The two basic charge-control parameters $Q_C(= -Q)$ and τ_1 having now been evaluated for operation above pinch-off, the small-signal parameters can be found from the general theory of Section 2.5. These substitutions are deferred, however, until the junction f.e.t. has been discussed.

The drain volt-ampere characteristics of a practical f.e.t. may differ from Fig. 4.28 in two significant respects:

1. Above pinch-off the drain current is not quite independent of drain voltage. The *drain resistance* r_D may be as small as a few tens of kilohms:

$$\frac{1}{r_D} = \left(\frac{\partial I_D}{\partial V_D}\right)_{V_G}. \tag{4.129}$$

2. An ideal insulated-gate f.e.t. can be operated in either the depletion mode or the enhancement mode. However, the channel in some n-type devices takes the form of an n-type inversion layer on a p-type substrate. This inversion layer appears only when the gate voltage exceeds a positive threshold value, and the I_D versus V_G characteristic is shifted horizontally

(Fig. 4.29). The equation for drain current above pinch-off (corresponding to Eq. 4.122) is

$$I_D = I_{DO}\left(\frac{V_G}{V_P} - 1\right)^2. \tag{4.130}$$

Provided the insulator voltage rating is not exceeded, the gate current in an insulated-gate f.e.t. is extremely small (of the order of picoamperes) and may be dominated by direct leakage between lead-wires.

4.8.2 Outline Theory of the *p-n* Junction F.e.t.

Figure 4.30 shows an *n*-channel junction f.e.t. with plane-parallel geometry. The channel is of length L, width $2a$, and cross-sectional area A_C, and has nonrectifying contacts at the ends for the source and drain. Highly doped *p*-type regions are formed on two parallel faces of the channel, and in combination these two electrically connected *p*-type regions constitute the gate structure. [Some practical junction f.e.t.s are effectively a bisection of Fig. 4.30 down its axis of symmetry; one gate electrode and half the channel are omitted. This form more closely parallels the insulated-gate device (Fig. 4.25).]

There are two basic differences between the insulated-gate and *p-n* junction f.e.t.:

1. Enhancement operation is not practicable with the junction device because application of a forward bias to the gate results in minority

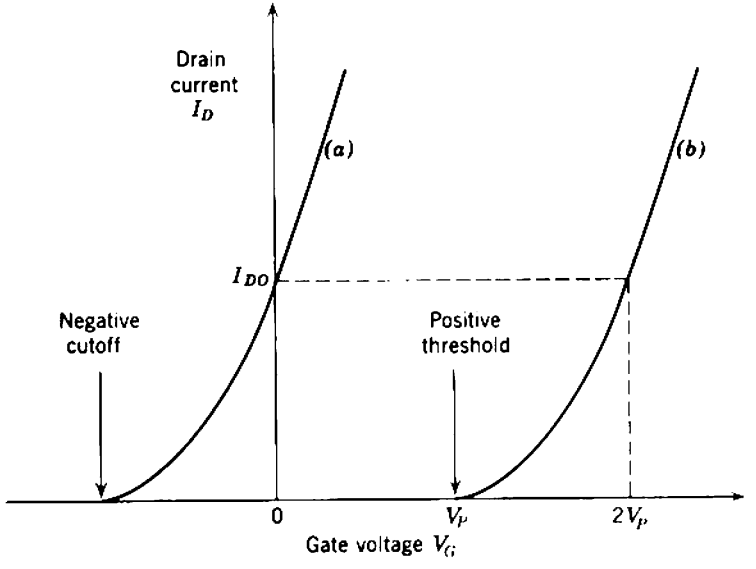


Fig. 4.29 Mutual characteristics of insulated-gate f.e.t.s: curve (*a*)—usual type; curve (*b*)—enhancement-only type.

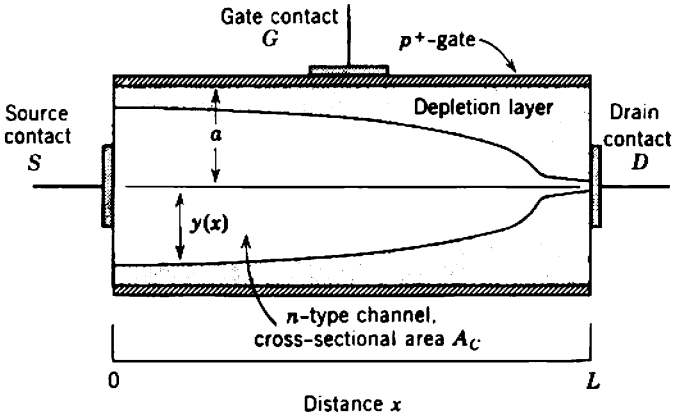


Fig. 4.30 Idealized model of an *n*-channel junction f.e.t.

carrier injection across the junction. A large gate current flows and the input resistance becomes small.

2. The conducting channel is relatively thick, and an analysis of the depletion process must take into account the geometrical position of the depletion layer.

The first step in analyzing the *p-n* junction f.e.t. is to determine the depletion layer profile as a function of voltage. Because the *p*-type gate is much more highly doped than the *n*-type channel, the depletion layer width $W(x)$ at any point follows from Eqs. 4.8 to 4.10 as

$$W(x) = \left[\left(\frac{2\varepsilon}{qN_D}\right)V(x)\right]^{1/2}. \tag{4.131}$$

where $V(x)$ is the magnitude of reverse voltage across the junction. It is assumed in Eq. 4.131 that the built-in contact potential V_i may be neglected or, if not neglected, included in the value of V. This assumption simplifies the algebra, but the existence of a contact potential does affect the temperature-dependence of device characteristics (Section 4.8.6). The depletion layer width W is related to the undepleted channel width y by

$$W(x) = a - y(x).$$

Therefore

$$y(x) = a\left\{1 - \left[\frac{V(x)}{V_P}\right]^{1/2}\right\}, \tag{4.132}$$

where the pinch-off voltage for a junction f.e.t. is

$$V_P = \frac{qN_D a^2}{2\varepsilon}. \tag{4.133}$$

The electron current density at any point in the channel is

$$J_e(x) = -q\mu_e n(x)\frac{d[V_C(x)]}{dx} = -q\mu_e N_D \frac{d[V(x)]}{dx}.$$

Therefore the total current along the undepleted portion of the channel (which is equal to the drain current) is

$$I_D = -I_e(x) = -J_e(x)A_C\left[\frac{y(x)}{a}\right];$$

that is,

$$I_D = qA_C\mu_e N_D\left\{1 - \left[\frac{V(x)}{V_P}\right]^{1/2}\right\}\frac{d[V(x)]}{dx}.$$

Integration over the channel length gives

$$I_D\int_0^L dx = qA_C\mu_e N_D\int_{-V_G}^{V_D - V_G}\left\{1 - \left[\frac{V(x)}{V_P}\right]^{1/2}\right\}d[V(x)]. \quad (4.134)$$

In the general case the drain current is a complicated function of electrode voltages. However, for operation at the point of pinch-off, the drain voltage is

$$V_D = V_P + V_G.$$

Evaluation of Eq. 4.134 then yields

$$I_D = I_{DO}\left[1 - 3\left(\frac{-V_G}{V_P}\right) + 2\left(\frac{-V_G}{V_P}\right)^{3/2}\right], \quad (4.135)$$

where I_{DO} is the pinched-off drain current at zero gate voltage:

$$I_{DO} = \frac{qA_C\mu_e N_D V_P}{3L}. \quad (4.136)$$

As with the insulated-gate f.e.t., the drain current predicted by Eq. 4.135 is extrapolated into the region beyond pinch-off. Equations 4.133, 4.135, and 4.136 may be compared with Eqs. 4.116, 4.122, and 4.123 for the insulated-gate f.e.t.

The mobile charge Q can be found by integrating N_D over the undepleted portion of the channel:

$$Q = -qA_C N_D\int_0^L\left[\frac{y(x)}{a}\right]dx.$$

For the simplest possible case, operation above pinch-off, the value of Q reduces to

$$Q = -qA_C L N_D\left[1 - \left(\frac{-V_G}{V_P}\right)^{1/2}\right]\left[\frac{1 + 3(-V_G/V_P)^{1/2}}{2 + 4(-V_G/V_P)^{1/2}}\right]. \quad (4.137)$$

The last term in Eq. 4.137 lies between the limits of 2/3 at $V_G = -V_P$

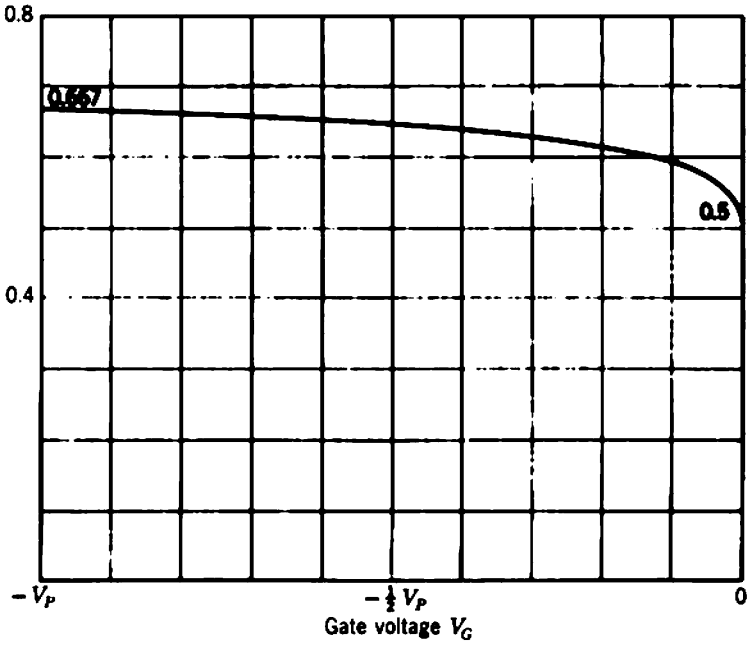


Fig. 4.31 The last term in Eq. 4.137.

and 1/2 at $V_G = 0$; Fig. 4.31 shows a plot. Therefore Q for operation above pinch-off can be approximated by

$$Q \approx -\frac{2qA_CLN_D}{3}\left[1 - \left(\frac{-V_G}{V_P}\right)^{1/2}\right], \tag{4.138}$$

which may be compared with Eq. 4.126. The transit time τ_1 for operation above pinch-off can be found by dividing Eq. 4.137 by Eq. 4.135:

$$\tau_1 = \frac{3L^2}{2\mu_e V_P}\left\{\frac{[1 + 3(-V_G/V_P)^{1/2}]}{[1 - (-V_G/V_P)^{1/2}][1 + 2(-V_G/V_P)^{1/2}]^2}\right\}. \tag{4.139}$$

In the special case of $V_G = 0$

$$\tau_1|_{V_G=0} = \frac{3L^2}{2\mu_e V_P}, \tag{4.140}$$

which may be compared with Eq. 4.128.

4.8.3 Generalizations

Equations 4.122 and 4.135 establish the forms of the mutual characteristics of ideal insulated-gate and abrupt-junction f.e.t.s operating above pinch-off. These mutual characteristics are plotted together in Fig. 4.32 and their similarity (for negative V_G) is apparent. Similar analyses can be

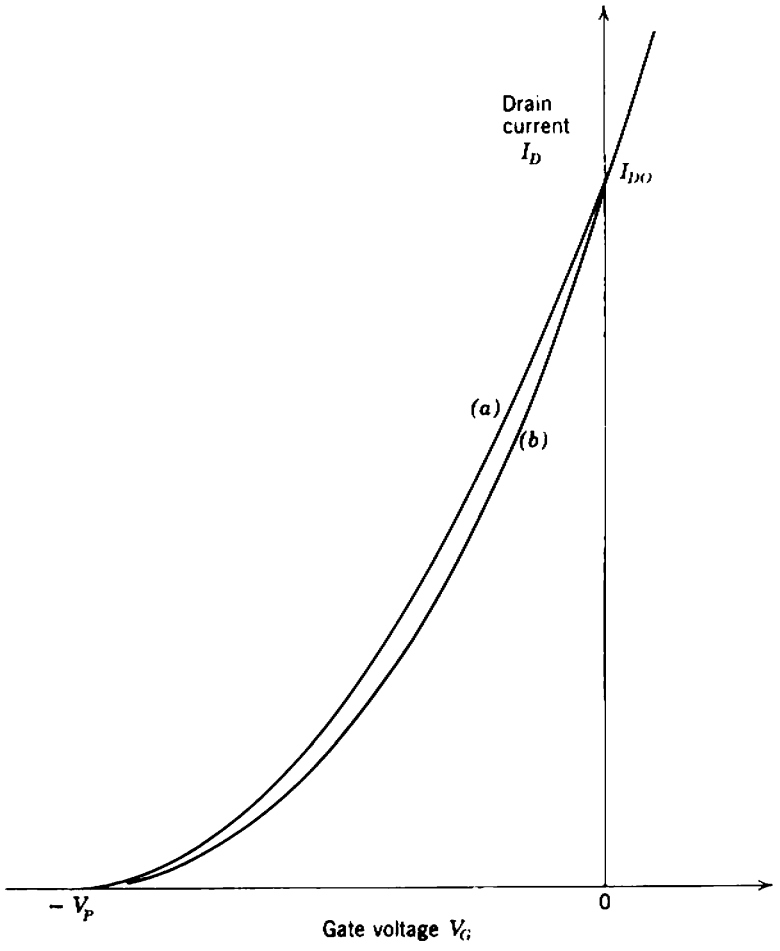


Fig. 4.32 Comparison of the mutual characteristics of insulated-gate and junction f.e.t.s: curve (*a*)—insulated-gate f.e.t. (Eq. 4.122); curve (*b*)—junction f.e.t. (Eq. 4.135)

performed for other types of f.e.t. (e.g., a *p-n* junction type with a graded junction) or for more precise physical models (e.g., one that takes into account the variation of mobility with carrier concentration). Such analyses usually lead to mutual characteristics which lie between those plotted in Fig. 4.32.* Thus different physical models for f.e.t. structures yield mutual characteristics which are, for most practical purposes, identical. Furthermore, measurements on many different commercial and experimental f.e.t. types verify the validity of the simple square law between drain current and gate voltage suggested by Eq. 4.122. Finally,

* I. RICHER, "Basic limits on the properties of field-effect transistors," *Solid State Electronics*, **6**, 539, September–October 1963.

Middlebrook* has presented a plausible general derivation of square law behavior. Consequently, the drain current for any f.e.t. operating above pinch-off can be written as

$$I_D = I_{DO}\left(1 + \frac{V_G}{V_P}\right)^2. \tag{4.141}$$

The two constants V_P and I_{DO} are sufficient to define the dc performance of any f.e.t., regardless of its structure. In practice, I_{DO} can be measured directly but V_P (ideally the gate voltage required for drain current cutoff) cannot because of "tailing" of the characteristic near cutoff. However, V_P can be deduced by extrapolating the gradient of the characteristic at zero V_G; the intercept is $-V_P/2$ (Fig. 4.33).

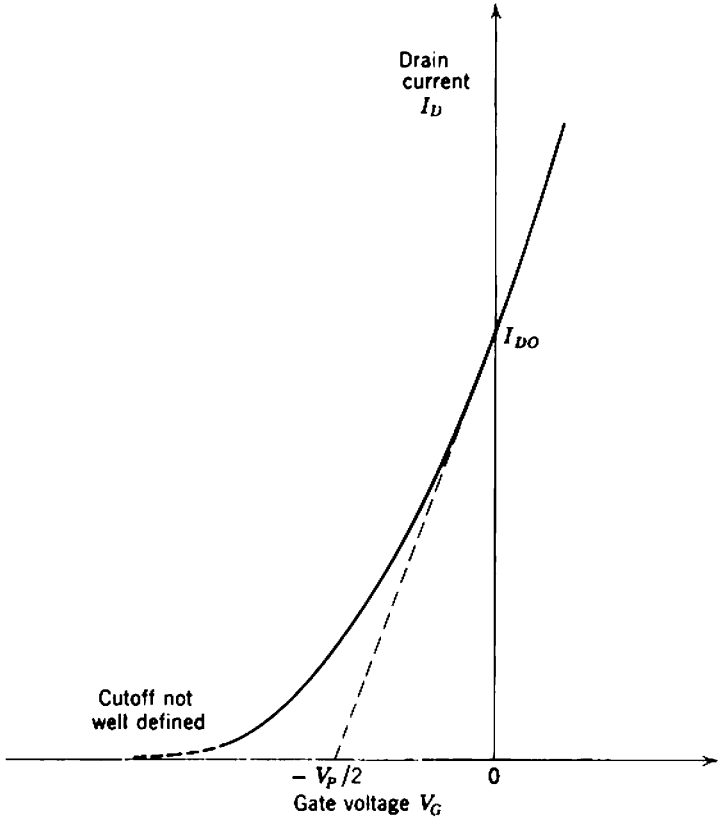


Fig. 4.33 Interpretation of V_P and I_{DO} for a practical f.e.t.

* R. D. MIDDLEBROOK, "A simple derivation of field-effect transistor characteristics," *Proc. Inst. Elec. Electronics Engrs.*, **51**, 1146, August 1963.

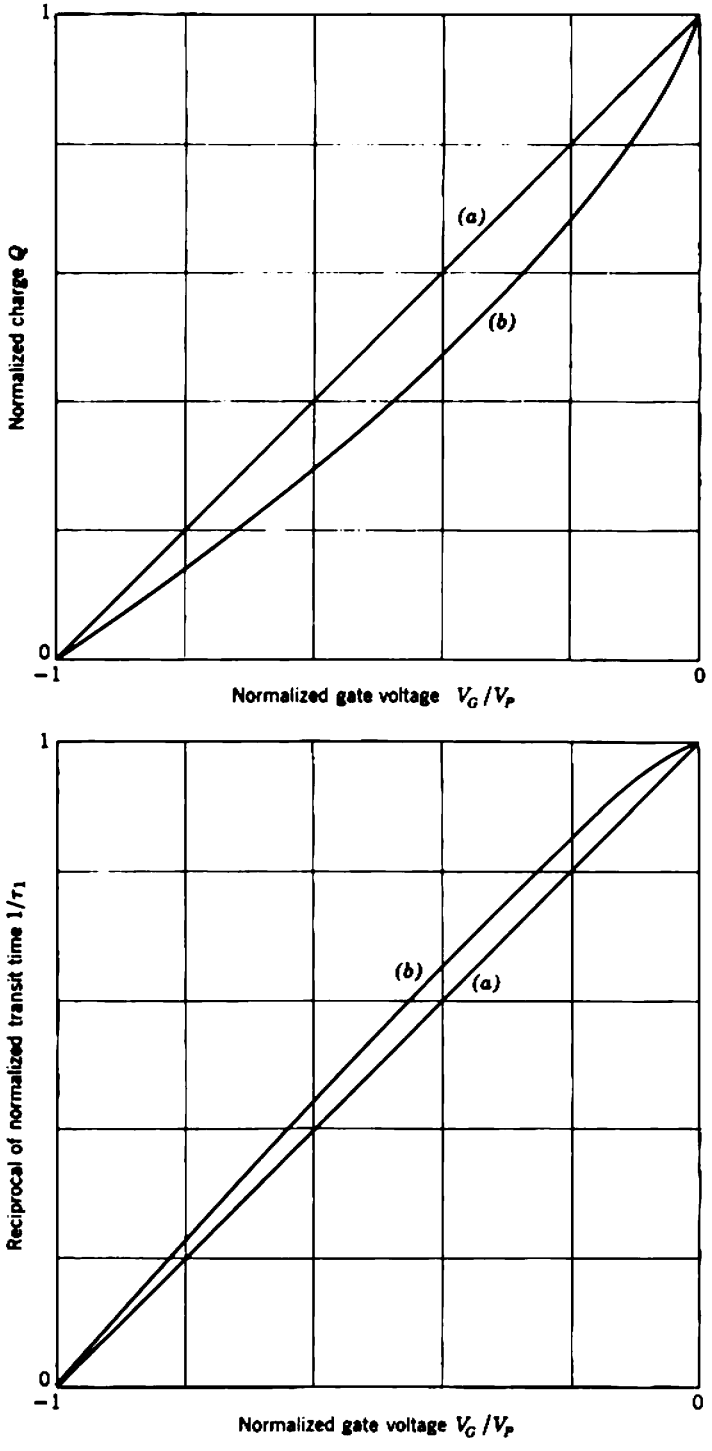


Fig. 4.34 Comparison of Q and τ_1 for insulated-gate and abrupt-junction f.e.t.s: curve (*a*)—insulated-gate f.e.t.; curve (*b*)—junction f.e.t.

To complete the comparison of abrupt-junction and insulated-gate f.e.t.s, Fig. 4.34 shows normalized plots of the charge-control parameters Q and τ_1 versus V_G, obtained from Eqs. 4.126, 4.127, 4.137, and 4.139. The similarity of the two device types is obvious.

4.8.4 Small-Signal Equivalent Circuit

Figure 4.35 shows the small-signal equivalent circuit for an f.e.t., first derived in Section 2.5 (Fig. 2.7*c*). The parameters can be found by substituting the known values of Q_C $(= -Q)$ and τ_1 into the general equations of Section 2.5. For an ideal insulated-gate device the results are*

$$g_m = \left(\frac{\partial I_D}{\partial V_G}\right)_{V_D} = \frac{2I_{DO}}{V_P}\left(1 + \frac{V_G}{V_P}\right) = \left[2\left(\frac{\varepsilon}{W_I}\right)\left(\frac{A_C}{\delta}\right)\left(\frac{\mu_e}{L}\right)\right]^{1/2} I_S^{1/2}, \quad (4.142)$$

$$c_1 = \left(\frac{\partial Q_C}{\partial V_G}\right)_{V_D} = \frac{2}{3}\left(\frac{\varepsilon}{W_I}\right)\left(\frac{A_C}{\delta}\right)L. \quad (4.143)$$

Notice that the source and drain currents I_S and I_D are equal. Equation 4.142 is expressed in terms of the emitting-electrode current I_S for consistency with the equations for g_m of the vacuum tube and bipolar transistor (3.36 and 4.53). The general forms of Eqs. 4.142 and 4.143 are typical of all f.e.t. types; g_m is proportional to $\sqrt{I_S}$, whereas c_1 is (substantially) constant.

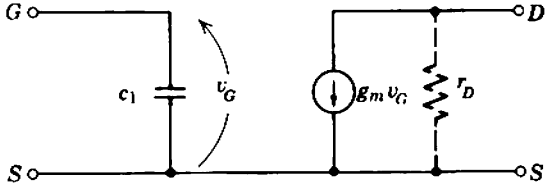


Fig. 4.35 Small-signal equivalent circuit for an intrinsic f.e.t.

* Equation 4.143 is derived first by direct substitution of (4.126) into (2.39). Equation 4.142 then follows from (2.53) and (2.47); (4.116), (4.122), (4.123), and (4.127) are used in the algebra.

The parameter u_1 defined in Eq. 2.47 is unity for all f.e.t.s and, incidentally, for all other device types in which the carrier motion is space-charge-limited drift.* The relation between g_m, c_1, and τ_1 therefore follows from Eq. 2.53 as

$$c_1 = \tfrac{1}{2} g_m \tau_1. \tag{4.144}$$

The mutual conductance can be obtained directly by differentiating the mutual characteristic [Eq. 4.122 for the insulated-gate f.e.t., (4.135) for the junction f.e.t., and (4.141) in general]. Notice that the value of g_m at $V_G = 0$ predicted for the generalized case is $(2I_{DO}/V_P)$, compared with $(3I_{DO}/V_P)$ for the junction device. The fact that experimental observations on practical junction f.e.t.s agree more closely with the first value suggests that these devices are not represented accurately by the physical model assumed in Section 4.8.2.

The input capacitance c_1 of an insulated-gate f.e.t.† is 2/3 the gate-to-channel capacitance in the absence of drain voltage and drain current; the capacitance per unit area is (ε/W_t), the channel depth is (A_C/δ), and the channel length is L. That the capacitance above pinch-off should be less than the capacitance without current is expected intuitively. When the gate voltage is changed, the full voltage increment appears across the insulating barrier only at the points at which the channel voltage is fixed, namely, the source and drain. At other points the charge rearranges itself to minimize the voltage change.

To complete the equivalent circuit the drain resistance defined in Eq. 4.129 must be added to the equivalent circuit, together with extrinsic capacitances and leakage conductances between all electrodes.

4.8.5 Noise Sources in Field-Effect Transistors

Ideally, the only noise in an f.e.t. is the thermal noise of the resistive channel, which can be represented by a current generator $d(i_{NT}{}^2)_D$ across

* R. D. MIDDLEBROOK, "A modern approach to semiconductor and vacuum device theory," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 887, Suppl. 17, 1959.

† The input capacitance of a junction f.e.t. can be found by differentiating Eq. 4.137. Over the useful operating range c_1 is substantially constant at 2/3 the gate-to-channel capacitance in the absence of drain voltage and with V_G set equal to $-V_P$ so that the depletion layer fills the channel. Over the same range u_1 is unity and the results for junction f.e.t.s closely parallel those for insulated-gate types. As V_G approaches zero, however, c_1 rises sharply (by a factor 9/2) and u_1 falls to zero. These variations are of little practical consequence; strictly, V_G in the analysis should include the built-in contact potential, so that when the external gate-to-source voltage is zero V_G will still be negative. Setting V_G of the analysis to zero corresponds to infinite forward injection current into the gate.

the output terminals (Fig. 4.36). Van der Ziel* has evaluated $d(i_{NT}^2)_D$ for an abrupt junction f.e.t. operating below pinch-off; the analysis can readily be extended to operation above pinch-off and to other f.e.t. types, and the general result is

$$d(i_{NT}^2)_D = 4k(\nu T)g_m\, df. \tag{4.145}$$

The numerical factor ν is rather more than one half; for an insulated-gate device it is 2/3, and for an abrupt-junction device it is the function of V_G plotted in Fig. 4.31. It is interesting that the expressions for thermal noise in an f.e.t. and shot noise in a triode or bipolar transistor (Eqs. 3.60 and 4.97) are the same.

Other and often dominant sources of noise are the flicker noise of the channel and the noise associated with the gate current. Flicker noise in field-effect transistors (like flicker noise in bipolar transistors) appears to be associated with random changes in the rates of carrier generation and recombination.† Flicker noise can be represented by a current generator $d(i_{NF}^2)_D$ across the output terminals, and

$$d(i_{NF}^2)_D = \frac{KI_D^{\,c}}{f}\, df. \tag{4.146}$$

where the exponent c (like the exponent in Eq. 4.101) is about 1. The corner frequency at which the thermal and flicker components of channel noise are equal is quite high, typically 100 kilohertz.

The noise current generator $d(i_{NS}^2)_G$ across the input terminals in Fig. 4.36 represents the shot noise associated with the gate current:

$$d(i_{NS}^2)_G = 2q\,\Sigma(I_G)\, df, \tag{4.147}$$

where $\Sigma(I_G)$ is the sum of the magnitudes of the individual components of gate current. Flicker components of gate-current noise may exist in

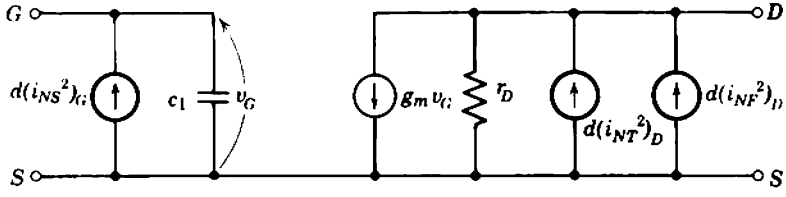


Fig. 4.36 Noise model for an f.e.t.

* A. VAN DER ZIEL, "Thermal noise in field-effect transistors," *Proc. Inst. Radio Engrs.*, **50**, 1080, August 1962.

† C.-T. SAH, "Theory of low-frequency generation noise in junction-gate field-effect transistors," *Proc. Inst. Elec. Electronics Engrs.*, **52**, 795, July 1964.

addition to the shot components. Field-effect transistors are often used with high-impedance sources, and gate-current noise is significant in these circumstances. It is therefore somewhat paradoxical that little theoretical or practical information on gate current is available to circuit designers. Experimentally it is observed that gate current for a particular f.e.t. depends somewhat on gate voltage and device temperature, but varies widely from one f.e.t. to another of the same type. The following comments serve merely to suggest some mechanisms which can be of importance.

An obvious comment concerns the essential difference between junction and insulated-gate device types. The *p-n* junction f.e.t. is restricted to operation with a reverse bias on the gate junction, because forward bias results in a large injection component of gate current. In contrast, no such current can flow in an insulated-gate f.e.t., so the gate voltage may be positive or negative. In the absence of injection components, a likely source of gate current is hole-electron pair generation in the depletion layer or insulating barrier.* This component of current increases exponentially with temperature (Eq. 4.79) and in the case of the junction f.e.t. it is proportional to the depletion layer volume. Reference to Section 4.8.2 therefore suggests that the generation component of gate current at $V_G = 0$ should be roughly half the value at cutoff. Measurements on some commercial devices are in agreement with this prediction, and it appears that carrier generation is the dominant source of gate current in junction f.e.t.s. However, other components do exist; for example, some junction devices exhibit unusually large values of gate current which may be due to surface defects. The components of gate current due to carrier generation, leakage, and even the normal saturation current of a reverse-biased *p-n* junction are all additive and contribute a small input conductance to the device.

4.8.6 Effects of Temperature on F.e.t. Characteristics

Bipolar transistors depend on the motion of minority carriers, and consequently their electrical characteristics depend strongly on temperature. In contrast, field-effect transistors depend on the motion of majority carriers through the control region and it might be expected that f.e.t. characteristics would not show any such strong temperature-dependence. Nevertheless, quite substantial changes do occur in f.e.t.

* C.-T. Sah, R. N. Noyce, and W. Shockley, "Carrier generation and recombination in *p-n* junctions, and *p-n* junction characteristics," *Proc. Inst. Radio Engrs.*, **45**, 1228, September 1957.

characteristics over the usable temperature range— -55 to $+150$°C for silicon devices. There are three major effects:

(i) The *gate current* increases substantially with temperature rise and, for *p-n* junction devices, I_G doubles for about every 10°C rise in temperature.
(ii) The *mobility* of carriers in the control region decreases with temperature rise (Eq. 4.23) and reduces the drain current.
(iii) The *contact potential* between the gate and the control region of a junction f.e.t. decreases with temperature rise (Eq. 4.5); hence the depletion-layer width falls and the drain current at constant gate voltage increases.

The last two effects oppose each other so that conditions of perfect temperature compensation can, in principle, be achieved.

To be specific, consider a *p-n* junction f.e.t. The effective gate voltage for this device (V_G in the analysis of Section 4.8.2) is the difference between the actual gate voltage and the junction contact potential V_i:

$$V_G = V_{G(\text{actual})} - V_i. \tag{4.148}$$

As shown in the generalizations of Section 4.8.3, the dc characteristics for operation above pinch-off are defined by the actual gate voltage required for drain current cutoff (designated $V_{G(\text{cutoff})}$) and the drain current at zero actual gate voltage (designated $I_{D(\text{zero})}$). The relations between these and the ideal parameters V_P and I_{DO} follow by inspection as

$$V_{G(\text{cutoff})} = -V_P + V_i \tag{4.149}$$

and from the generalized dc characteristic (Eq. 4.141)

$$I_{D(\text{zero})} = I_{DO}\left(1 - \frac{V_i}{V_P}\right)^2. \tag{4.150}$$

The pinch-off voltage V_P is defined by Eq. 4.133, and all terms in V_P are substantially independent of temperature. Therefore, since the contact potential V_i falls with increasing temperature, $V_{G(\text{cutoff})}$ for an *n*-channel device must become more negative. $I_{D(\text{zero})}$ can either increase or decrease with increasing temperature, depending on the relative magnitudes of two effects; I_{DO} falls (because of the temperature dependence of μ_e in Eq. 4.136) but the second term in Eq. 4.150 rises. $I_{D(\text{zero})}$ has a positive temperature coefficient if V_P is small, but a negative coefficient if V_P is large. Figure 4.37 shows the three possible effects of a temperature increase on the characteristics of a junction f.e.t. Section 15.3.2.3 contains further discussion.

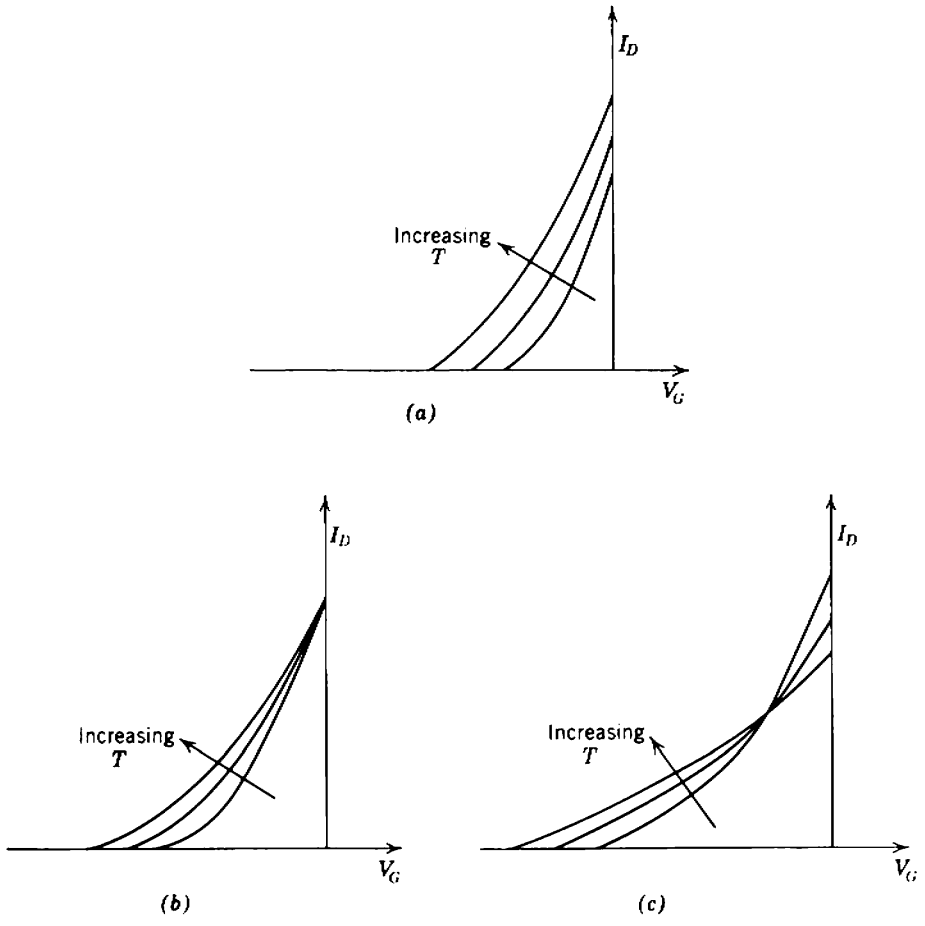


Fig. 4.37 Effect of temperature on f.e.t. mutual characteristics: (*a*) small V_P; (*b*) medium V_P; (*c*) large V_P.

4.8.7 Ratings of Field-Effect Transistors

Field-effect transistors are operated at rather higher voltages than bipolar transistors (often more than 10 volts) in order to reach the constant-current part of their drain characteristics. Operating current levels are of the order of a few milliamperes, and consequently the power dissipation is of the order of tens of milliwatts. Values of thermal resistance are quoted by manufacturers, and the device internal temperature can be calculated as for bipolar transistors.

The safe value of I_D is limited by power dissipation. Usually a maximum value of V_D is specified, based on breakdown between the drain and

gate. For similar reasons there is often a maximum reverse voltage rating between the gate and source. These gate voltage ratings are very important in some types of insulated-gate f.e.t., and care must be taken to avoid damage from surges. At least one manufacturer warns that the electrostatic voltages produced by pushing the leads into foam polystyrene or similar materials can permanently damage the insulating barrier.

Chapter 5

Small-Signal Mid-Band Circuit Theory of Elementary Amplifiers

The early sections of Chapter 2 are concerned with the development of a small-signal equivalent circuit for a hypothetical charge-controlled device. In Chapters 3 and 4 it is shown that some of the elements in this equivalent circuit may vanish when practical vacuum tubes or transistors are considered, and that it may be necessary to add extrinsic elements to complete the equivalent circuit of such devices. The purpose of the present chapter is to consider in theory how a charge-controlled device can be used as an amplifier. Moreover, this chapter is concerned only with *small-signal* amplification at *mid-band* frequencies. The significance of these terms is:

1. The parameters of an active device depend on the particular levels of voltage and current present in it. These parameters therefore fluctuate from point to point on a signal waveform. Small-signal operation implies that the signal fluctuations about the zero-signal quiescent point are small, so that the parameters may be assumed constant. In other words, the equivalent circuit is assumed to be linear.

2. The reactance of a capacitor increases as the frequency is reduced, and becomes very large indeed at low frequencies. Therefore, provided that the frequency of a signal which is to be amplified is not too high, the shunting capacitors in the equivalent circuit of the charge-control model can be neglected. Often artificial limits are imposed on low-frequency performance by the finite reactances of coupling, bypass, and decoupling capacitors. The purpose of these capacitors is discussed in Chapters 6 and 7; they can be neglected at all but the lowest frequencies. In the

mid-band range of frequencies the reactance of all capacitors is neglected; low-pass capacitors are assumed to be signal open-circuits and high-pass capacitors to be signal short-circuits.

Section 5.1 is concerned with the formal development of the transfer and driving-point functions of an amplifier stage using the hypothetical charge-control model. Sections 5.2 and 5.3 interpret this development for two practical devices—the vacuum tube and the transistor. The problems in cascading a number of stages to form a complete amplifier are considered in Sections 5.4 and 5.5, and Section 5.7 lists the transfer and driving-point functions of practical devices when they are used in unusual configurations.

The transfer and driving-point functions derived for practical devices in Sections 5.2 and 5.3 differ from those derived in Section 5.1 for the hypothetical charge-control model. These differences arise in two quite separate ways. First, extrinsic elements are introduced to complete the equivalent circuits for the practical devices, and this increases the complexity of the equations. Second, a number of approximations are made to simplify practical design work; this introduces no significant errors because even the approximate expressions are much more accurate than the numerical data available for practical devices. In general, the small-signal parameters of a vacuum tube are not reproducible with a tolerance closer than about $\pm 50\%$ of their nominal value. There is considerable variation in the tolerance of transistor parameters; a few are reproducible to within 1 or 2%, the majority lie within the range 70 to 140% of their nominal value, and some have even greater tolerance limits. Other circuit elements such as resistors, capacitors, and inductors have production tolerances also.

The types of approximation are quite different in the two cases of the vacuum tube and the transistor. For vacuum tubes, the elements associated with recombination of carriers may be neglected; therefore the equations developed in Section 5.1 simplify considerably, and these simplified equations apply accurately almost irrespective of how a vacuum tube is used as a circuit element. Similar comments apply to the field-effect transistor, but no general simplification is valid for the bipolar transistor. Instead, a design philosophy is developed that minimizes the effects of certain elements; these elements may be omitted from the analysis *provided the circuit is designed so that the simplifying assumptions do in fact apply.* New amplifying devices of the charge-control family may be developed in the future. The general equations developed in Section 5.1 will apply to these devices, but the design approximations corresponding to Section 5.2 or 5.3 may need modification to account for

their specific physical properties. It is hoped that the treatment of two different types of approximation in this book will suggest a design philosophy suitable for any new device.

The approximations in the circuit equations are not introduced to simplify the algebraic computations, though this is a very useful consequence. Rather the approximations become an aid to circuit design through their high-lighting the basic factors in the operation of a circu it In this way it becomes clear which are the important first-order factors that determine the gain of an amplifier, and which are merely second-order corrections that can be minimized by suitable design techniques. This concentration on approximations involving the parameters of a physically based equivalent circuit has the following advantages:

1. It suggests a design philosophy for which the gain depends primarily on the more reproducible device parameters. The circuit is therefore more *designable*. In a designable circuit the performance remains within specified tolerance limits when the circuit is mass produced, without requiring selection of components. The source of nondesignability in amplifier circuits is principally in the tolerance on the parameters of the active devices, that is, the vacuum tubes or the transistors.

2. The visualization of just how a circuit works is an aid to logical circuit design. Every element in the circuit can be associated with one or more specific purposes. The alternative is a mathematical proof that a complete circuit has a certain input-to-output relation, with very little insight to the purpose of any element.

3. If the purpose of each circuit element is clearly understood, it is simple to predict the effect of any change. It is therefore relatively simple to optimize a circuit design. As a general rule, the more completely a circuit can be designed on paper without recourse to experimental adjustment, the more likely it is to be reproducible. Alternatively, it is possible to foresee the consequences of an unintentional change due to the tolerance on device characteristics.

5.1 THE GENERAL CHARGE-CONTROL MODEL

Figure 5.1 is the small-signal equivalent circuit for the charge-control model at mid-band frequencies, with a resistive load R_L connected across its output terminals. Such an arrangement is called an *amplifier stage*. Figure 5.1*a* is the exact mid-band equivalent circuit first shown in Fig. 2.8*a*; its equations are

$$i_1 = g_1 v_1 - g_f v_2, \tag{5.1}$$

$$i_2 = g_m v_1 + g_2 v_2, \tag{5.2}$$

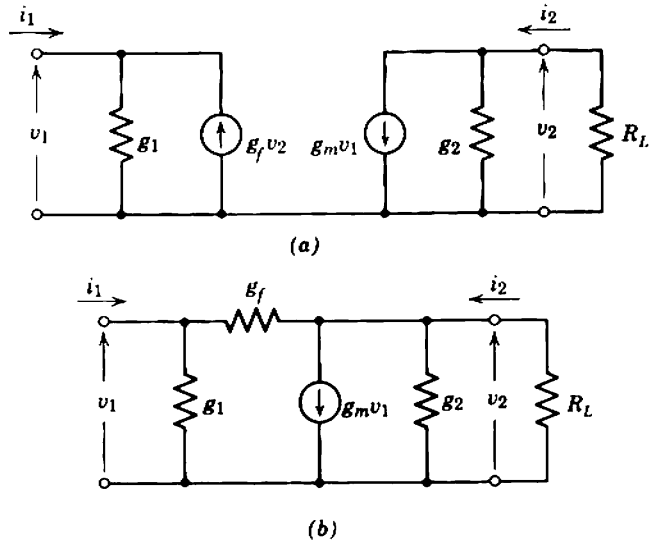


Fig. 5.1 Mid-band equivalent circuits for a charge-controlled device with a resistive load.

where, in this chapter and for the remainder of the book, increments in voltage or current (i.e., signal quantities) are represented by lowercase letters. Figure 5.1*b* is the approximate mid-band equivalent circuit introduced in Fig. 2.8*b*, its equations being

$$i_1 = (g_1 + g_f)v_1 - g_f v_2, \tag{5.3}$$

$$i_2 = (g_m + g_f)v_1 + (g_f + g_2)v_2. \tag{5.4}$$

However, in any useful amplifying device

$$g_f \ll g_m, \tag{5.5}$$

$$g_f \ll g_1, \tag{5.6}$$

$$g_f \ll g_2, \tag{5.7}$$

so that Eqs. 5.3 and 5.4 reduce to Eqs. 5.1 and 5.2. The single-generator equivalent circuit (Fig. 5.1*b*) is used almost exclusively throughout this book, as it emphasizes the difference between the desired forward transmission (through the mutual conductance) and the undesired reverse transmission.

The notation can be simplified by introducing a number of new symbols. First, r_1, r_2, and r_f are introduced as the reciprocals of g_1, g_2, and g_f:

$$r_1 = \frac{1}{g_1}, \tag{5.8a}$$

$$r_2 = \frac{1}{g_2}, \tag{5.8b}$$

$$r_f = \frac{1}{g_f}. \tag{5.8c}$$

These symbols are used to remove a number of fractions within fractions in the algebraic development. Second, it follows from Section 2.5.1 that the elements in the charge-control model are interrelated:

$$\frac{g_m}{g_1} = g_m r_1 = (1 + u_1)\frac{\tau}{\tau_1} = \beta, \tag{5.9}$$

$$\frac{g_2}{g_f} = \frac{r_f}{r_2} = -(1 + u_2)\frac{\tau}{\tau_1} = \frac{\beta}{\gamma}, \tag{5.10}$$

where

$$\beta = (1 + u_1)\frac{\tau}{\tau_1}, \tag{5.11}$$

$$\gamma = -\frac{1 + u_1}{1 + u_2}. \tag{5.12}$$

β is identified as the current amplification factor introduced in Eq. 2.57. The parameters g_m and g_2 are more reproducible than g_1 and g_f, because they are directly related to the fundamental mode of operation of practical devices. Therefore g_1 and g_f (which merely account for recombination) are eliminated in the final forms of all equations derived in this chapter.

5.1.1 Transfer Conductance and Voltage Gain

Equation 5.2 represents the device shown in Fig. 5.2. When a load resistance R_L is connected to the output terminals of the device, the current in this load resistance can be found from

$$G_T = \frac{i_2}{v_1} = g_m\left(\frac{r_2}{r_2 + R_L}\right). \tag{5.13}$$

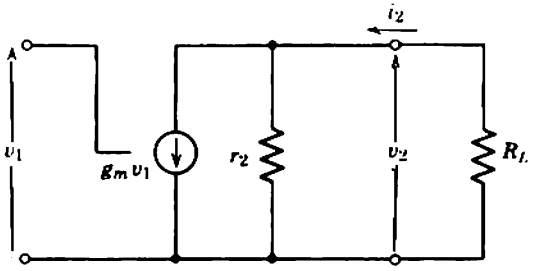


Fig. 5.2 The form of the circuit equivalent to Eq. 5.2.

The quantity G_T is the *transfer conductance* of the stage. The significance of the terms in the expression for G_T is the following:

(i) g_m is the mutual conductance of the device; it represents the internal current generator of the device and hence the current that would flow into a short-circuit external load:

$$g_m = \left(\frac{i_2}{v_1}\right)_{sc}. \tag{5.14}$$

(ii) $r_2/(r_2 + R_L)$ represents the fraction of the current from the g_m generator that flows into a finite load resistor R_L, rather than into r_2.

The output voltage is given by the product of the output current with the load resistance. A minus sign is required because of the reference direction chosen for i_2:

$$v_2 = -i_2 \times R_L. \tag{5.15}$$

Therefore the *voltage gain* of the stage is

$$A_V = \frac{v_2}{v_1} = -G_T \times R_L, \tag{5.16}$$

and substituting,

$$A_V = \frac{v_2}{v_1} = -g_m\left(\frac{r_2}{r_2 + R_L}\right)R_L. \tag{5.17}$$

The significance of the minus sign in the voltage gain is that there is a phase shift of 180°; a positive increment of input voltage results in a negative increment of output voltage.

5.1.2 Input Resistance

When an input voltage is applied to a device, an input current flows because of the recombination of carriers in the control region. This

input current gives rise to an *input conductance* of an amplifier stage. The input current can be evaluated from Eq. 5.1, and division by the input voltage gives the input conductance:

$$G_i = \frac{i_1}{v_1} = g_1 - \frac{v_2}{v_1} g_f. \tag{5.18}$$

Therefore, from Eq. 5.17,

$$G_i = g_1 + g_m g_f \left(\frac{r_2}{r_2 + R_L}\right) R_L$$

and using Eqs. 5.8 to 5.10 the *input resistance* is

$$R_i = \frac{1}{G_i} = \frac{v_1}{i_1} = r_1 \left[\frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L}\right], \tag{5.19}$$

or, in terms of the more fundamental quantity g_m,

$$R_i = \frac{v_1}{i_1} = \frac{\beta}{g_m}\left[\frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L}\right] = \frac{K_1 \beta}{g_m}, \tag{5.20}$$

where, for later use, we write

$$K_1 = \frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L}. \tag{5.21}$$

Notice that R_L must lie between zero and infinity, so that

$$\frac{1}{1 + \gamma} \leq K_1 \leq 1. \tag{5.22}$$

5.1.3 Current Gain and Transfer Resistance

The *current gain* of a charge-controlled device is the ratio of the output and input currents. Both currents are directly proportional to the excess mobile charge in the control region. For the output current the constant of proportionality is the transit time τ_1, and this is a fairly reproducible physical parameter. For the input current, however, the constant is the life time τ, and this is the physical parameter that has the least reproducibility of all. The current gain of a charge-controlled device cannot therefore be relied upon to have a particular value quoted by a manufacturer.

The current gain of the stage shown in Fig. 5.1 is

$$A_I = \frac{i_2}{i_1} = \left(\frac{v_1}{i_1}\right)\left(\frac{i_2}{v_1}\right) = R_i \times G_T, \tag{5.23}$$

and substitution from Eqs. 5.13, and 5.20 gives

$$A_I = \frac{i_2}{i_1} = \beta\left[\frac{r_2}{r_2 + (1 + \gamma)R_L}\right]. \tag{5.24}$$

The ratio of the output voltage to the input current, or *transfer resistance* of the stage is

$$R_T = \frac{v_2}{i_1} = -A_I \times R_L \tag{5.25}$$

and substituting

$$R_T = \frac{v_2}{i_1} = -\beta\left[\frac{r_2}{r_2 + (1 + \gamma)R_L}\right]R_L. \tag{5.26}$$

It should be emphasized that the transfer conductance and transfer resistance of a stage, G_T and R_T, respectively, are not reciprocals.

5.1.4 Finite Source Resistance

The transfer functions G_T, A_V, A_I, and R_T relate the output signal of an amplifier to the input signal; v_1 is the signal voltage actually present between the input terminals and i_1 is the current that actually flows into the amplifier. The input signal of an amplifier stage can be produced in any of several ways:

1. Within a multistage amplifier, the input signal to one stage is the output signal from the preceding stage. This signal depends on the value of the transfer function of the first stage, hence on its load resistance. The input resistance of the second stage constitutes a load on the first stage and must be taken into account.
2. In rare cases, an amplifier may be fed from a true voltage or current source so that the appropriate input signal is constrained to a particular value, quite independent of the input resistance.
3. The input stage of an amplifier is usually fed from some transducer or other source whose output resistance is finite. The ratio of the output signal from the amplifier to the input signal actually present is given by the appropriate transfer function G_T, A_V, A_I, or R_T. The input signal, however, depends on the input resistance of the amplifier.

In this third case it is useful to have the output signal expressed in terms of the generator within the source, because this generator is the signal to be amplified. Figures 5.3*a* and *b* show equivalent voltage-generator and current-generator representations of the source; notice that

$$v_S = i_S R_S. \tag{5.27}$$

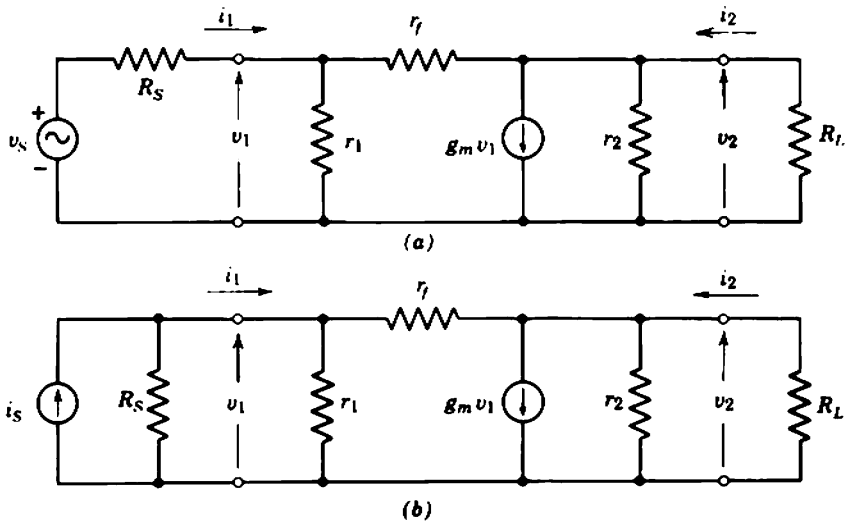


Fig. 5.3 Mid-band equivalent circuit for a charge-controlled device with a resistive load, fed from a source of finite output resistance: (*a*) voltage-generator representation of source; (*b*) current-generator representation of source.

In the voltage-generator representation, the source resistance R_S forms a voltage divider with the input resistance R_i of the device such that

$$\frac{v_1}{v_S} = \frac{R_i}{R_S + R_i}. \tag{5.28}$$

Then

$$\frac{i_2}{v_S} = \left(\frac{v_1}{v_S}\right)\left(\frac{i_2}{v_1}\right) = \left(\frac{R_i}{R_S + R_i}\right)G_T,$$

and substitution from Eqs. 5.13, 5.20, and 5.28 gives

$$\frac{i_2}{v_S} = \left(\frac{g_m}{1 + g_m R_S/K_1\beta}\right)\left(\frac{r_2}{r_2 + R_L}\right). \tag{5.29}$$

In the current-generator representation, R_S shunts the input resistance of the device, and the input voltage is

$$\frac{v_1}{i_S} = \frac{R_S \times R_i}{R_S + R_i}. \tag{5.30}$$

Then

$$\frac{i_2}{i_S} = \left(\frac{v_1}{i_S}\right)\left(\frac{i_2}{v_1}\right) = \left(\frac{R_S \times R_i}{R_S + R_i}\right)G_T,$$

and substitution from Eqs. 5.13, 5.20, and 5.30 gives

$$\frac{i_2}{i_S} = \left(\frac{g_m R_S}{1 + g_m R_S/K_1\beta}\right)\left(\frac{r_2}{r_2 + R_L}\right) \tag{5.31}$$

or, as expected from Eq. 5.27,

$$\frac{i_2}{i_S} = R_S \times \left(\frac{i_2}{v_S}\right). \tag{5.32}$$

5.1.5 Output Resistance

In the same way that the source for any stage may be considered either as a voltage generator with a series resistor or a current generator with a shunt resistor, so the output of a stage may be considered as an ideal generator with a resistor. Figure 5.4 shows the current-generator representation. The output current in the load is

$$i_2 = i_g\left(\frac{R_o}{R_o + R_L}\right). \tag{5.33}$$

In terms of the equations derived above

$$i_2 = v_S \times \left(\frac{i_2}{v_S}\right)$$

and substitution from Eqs. 5.21 and 5.29 gives

$$i_2 = v_S\left(\frac{g_m}{1 + g_m R_S/\beta}\right)\left(\frac{K_2 r_2}{K_2 r_2 + R_L}\right), \tag{5.34}$$

where

$$K_2 = \frac{R_S + \beta/g_m}{(1 + \gamma)R_S + \beta/g_m}. \tag{5.35}$$

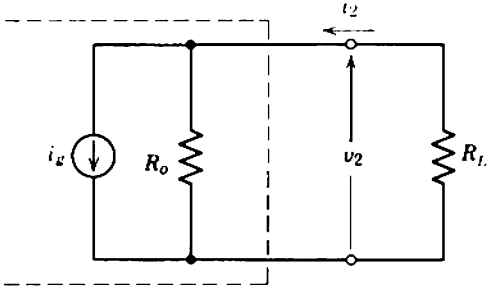


Fig. 5.4 Current-generator representation of the output circuit of an active device.

Notice that R_S must lie between zero and infinity, so that

$$\frac{1}{1+\gamma} \leq K_2 \leq 1. \tag{5.36}$$

By comparison of Eq. 5.33 with Eq. 5.34, it follows that

$$i_g = v_S\left(\frac{g_m}{1 + g_m R_S/\beta}\right) \tag{5.37}$$

and

$$R_o = r_2\left[\frac{R_S + \beta/g_m}{(1+\gamma)R_S + \beta/g_m}\right] = K_2 r_2. \tag{5.38}$$

The significance of the terms is the following:

(i) i_g represents the current output into a short-circuit load,
(ii) R_o is the *output resistance* of the stage for the given value of R_S.

In the same way that the input resistance R_i is a function of the load R_L, so the output resistance R_o is a function of the source R_S. In the special cases of voltage or current drive in which R_S becomes zero or infinite, the output resistance becomes r_2 or $r_2/(1+\gamma)$ respectively; these results are expected from the forms of Eqs. 5.13 and 5.24. A certain amount of confusion surrounds the concept of output resistance of an amplifier, so it is worth making the following points:

1. The input resistance of an amplifier is the quotient of the input voltage and current. The output resistance certainly is not the quotient of the output voltage and current; this quotient is the load resistance R_L (Eq. 5.15).
2. The output resistance is the resistance seen looking back into the amplifier from the load. It is the resistance that would be measured between the output terminals of the amplifier if the source generator were reduced to zero and load were disconnected. (With the source reduced to zero but the load connected, the resistance measured is obviously the parallel combination of R_o and R_L.)
3. If maximum power transfer to the load is required with a given combination of source resistance and amplifier, the output resistance should be matched by the load. If maximum power transfer from source to load is required, the amplifier should be designed to have input and output resistances R_i and R_o that match the source and load respectively; this will require a relatively involved calculation because R_i depends on R_L, and R_o depends on R_S. In passing, it may be remarked that the technique of impedance matching for maximum power transfer has almost

no place in the production of designable circuits, and is scarcely mentioned in this book. This is due in part to the complexity of the design calculations involved but, more importantly, the gain of matched amplifiers is extremely dependent on device parameter variations.

4. In normal mismatched amplifier design the output resistance need not be known precisely nor need it be calculated explicitly. The load resistance R_L usually is known, either as a precise magnitude or as an estimated range of values. For an amplifier that is to deliver a defined signal current to R_L, it is necessary merely to ensure that R_o is sufficiently large in comparison with R_L to remove any uncertainty in the division of current between R_o and R_L. Equally, for an amplifier that is to deliver a defined signal voltage to R_L, it is necessary merely to ensure that R_o is sufficiently small in comparison with R_L so that there is negligible uncertainty in the division of voltage.

5. The output resistance can be deduced from the transfer function thus: R_o is equal to the value of R_L that reduces the output voltage to half its open-circuit value or that reduces the output current to half its short-circuit value:

$$v_o|_{R_L = R_o} = \tfrac{1}{2} v_o|_{oc}, \tag{5.39a}$$

$$i_o|_{R_L = R_o} = \tfrac{1}{2} i_o|_{sc}. \tag{5.39b}$$

Equation 5.38 can be transformed using Eq. 5.9 to give

$$R_o = r_2 \left[\frac{R_S + r_1}{(1 + \gamma)R_S + r_1} \right]. \tag{5.40}$$

This form of R_o is symmetrical with one form of R_i in Eq. 5.19:

$$R_i = r_1 \left[\frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L} \right].$$

This symmetry arises from the symmetry of the charge-control model (Fig. 5.1*a*), but it is only of passing interest. The alternative forms given in Eqs. 5.20 and 5.38 are more useful.

5.1.6 Load Resistance

The load resistance R_L for the device in a practical amplifier stage is made up of two components. The first is the external load R_{ext} to which the output signal is applied to produce some useful result. In a cascade of stages, the external load for one stage is the input resistance of the next stage. The second component of R_L arises from the need for some element(s) to supply the quiescent voltages and currents to the device.

Often this element is a single resistor, but transformers or parallel combinations of resistors are used sometimes. Figure 5.5 shows typical circuit arrangements for vacuum tubes and transistors; the reasons for these circuit arrangements are given in the discussion of biasing circuits in Chapter 6. For the present, the important point is that some dc supply elements are required. Each of these elements provides a shunting path for the signal currents and, at mid-band frequencies, the total shunting resistance is the parallel combination of all elements. This

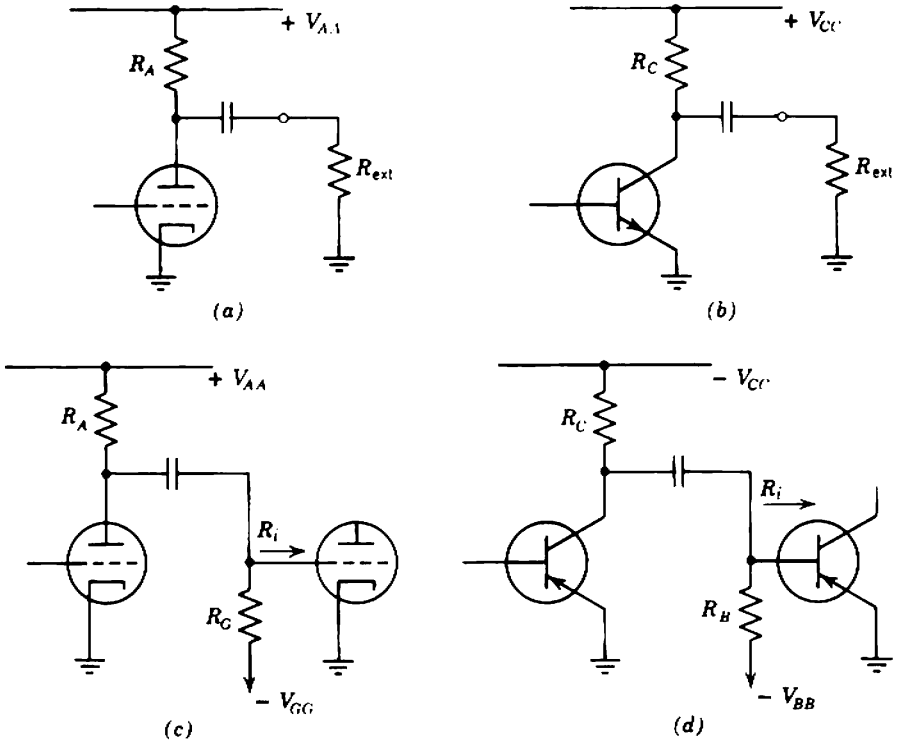


Fig. 5.5 The load resistance for the device is

$$R_L = R_P \parallel R_{ext},$$

where

(*a*) $R_P = R_A$.

(*b*) $R_P = R_C$.

(*c*) $R_P = R_A \parallel R_G$,
$R_{ext} = R_i \ (\approx \infty)$.

(*d*) $R_P = R_C \parallel R_B$,
$R_{ext} = R_i$.

shunting resistance will be referred to as the *supply resistor* R_P. The load resistance is

$$R_L = \frac{R_{ext}R_P}{R_{ext} + R_P}. \tag{5.41}$$

It is important in some of the following sections to know the largest value the load resistance can take. The load resistance is as large as possible when the external load is infinite, and as a corollary, when there is no following amplifier stage. Thus the load for a vacuum tube must satisfy the inequality

$$R_L \leq R_A, \tag{5.42}$$

while for a transistor

$$R_L \leq R_C. \tag{5.43}$$

5.1.7 Summary

Table 5.1 Summary of Transfer and Driving-Point Functions for an Amplifier Stage Using the Hypothetical Charge-Control Model (Fig. 5.1) with $g_1 = g_m/\beta$, $g_2 = 1r_2$, and $g_l = \gamma/\beta r_2$

$$G_T = \frac{i_2}{v_1} = g_m\left(\frac{r_2}{r_2 + R_L}\right), \tag{5.13}$$

$$A_V = \frac{v_2}{v_1} = -G_T \times R_L, \tag{5.16}$$

$$A_I = \frac{i_2}{i_1} = \beta\left[\frac{r_2}{r_2 + (1 + \gamma)R_L}\right], \tag{5.24}$$

$$R_T = \frac{v_2}{i_1} = -A_I \times R_L, \tag{5.25}$$

$$R_i = \frac{v_1}{i_1} = \frac{\beta}{g_m}\left[\frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L}\right]. \tag{5.20}$$

For a source v_S or i_S, of resistance R_S,

$$R_o = K_2 r_2 \tag{5.38}$$

where

$$K_2 = \frac{R_S + \beta/g_m}{(1 + \gamma)R_S + \beta/g_m}, \tag{5.35}$$

$$\frac{i_2}{v_S} = \left(\frac{g_m}{1 + g_m R_S/K_1\beta}\right)\left(\frac{r_2}{r_2 + R_L}\right) \tag{5.29}$$

where

$$K_1 = \frac{r_2 + R_L}{r_2 + (1 + \gamma)R_L}, \tag{5.21}$$

$$\frac{i_2}{i_S} = R_S \times \left(\frac{i_2}{v_S}\right). \tag{5.32}$$

5.2 THE VACUUM TUBE

This section is concerned with triode vacuum tubes or with pentodes in which there is no screen circuit degeneration. It is assumed that for signals the screen of a pentode is connected to the cathode so there are only three accessible electrodes—the cathode, grid, and anode. In practice, the screen is usually connected to the cathode via a capacitor whose reactance at signal frequencies is so small as to constitute a virtual short circuit. Pentodes with screen degeneration are considered in Section 7.4.1.2. The theory presented in this section is also applicable to field-effect transistors but these are not mentioned explicitly.

The equivalent circuit for a vacuum tube is simpler than that of the charge-control model considered in Section 5.1, because recombination is virtually nonexistent and, for most applications, grid current is negligible. Further, no extrinsic elements need be taken into account at mid-band frequencies. Accordingly, there is no input current, β becomes infinite, g_1 and g_f vanish, and the mid-band transfer functions involving an input current to the device become meaningless. The mid-band equivalent circuit for a vacuum tube is shown in Fig. 5.6; the general elements are relabeled as the specific elements of a tube (Table 5.2) and the general formulas are listed in Table 5.3. Notice that the voltage amplification factor μ of a vacuum tube is

$$\mu = g_m r_A. \tag{5.44}$$

Table 5.2 General Elements and Vacuum-Tube Parameters

Hypothetical Model	Vacuum Tube
g_m	g_m mutual conductance
r_2	r_A anode resistance
v_1	v_G grid voltage
v_2	v_A anode voltage
i_2	i_A anode current

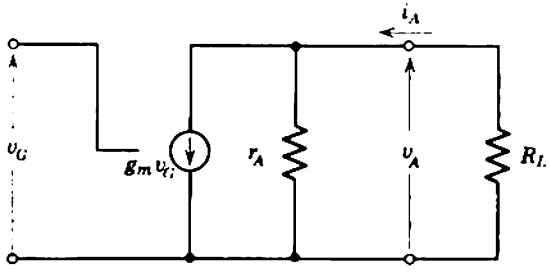


Fig. 5.6 Mid-band equivalent circuit for a vacuum tube with a resistive load.

Table 5.3 Transfer and Driving-Point Functions of a Vacuum-Tube Stage

$$G_T = \frac{i_A}{v_G} = g_m\left(\frac{r_A}{r_A + R_L}\right) = \frac{\mu}{r_A + R_L}, \tag{5.45}$$

$$A_V = \frac{v_A}{v_G} = -g_m\left(\frac{r_A}{r_A + R_L}\right)R_L = -\frac{\mu R_L}{r_A + R_L}. \tag{5.46}$$

For a source v_S or i_S, of resistance R_S,

$$R_o = r_A, \tag{5.47}$$

$$\frac{i_A}{v_S} = g_m\left(\frac{r_A}{r_A + r_L}\right) = \frac{\mu}{r_A + R_L}, \tag{5.48}$$

$$\frac{i_A}{i_S} = R_S g_m\left(\frac{r_A}{r_A + r_L}\right) = \frac{\mu R_S}{r_A + R_L}. \tag{5.49}$$

5.2.1 Approximations, Practical Values, and Tolerances

An important conclusion from Eqs. 5.45 to 5.49 is that the transfer and driving-point functions of a vacuum tube depend critically on the operating conditions and the type of tube. Both g_m and r_A vary widely with quiescent point and with tube type, and even at the same quiescent point they vary significantly from unit to unit of the same type due to manufacturing tolerances. Therefore, if an amplifier stage using a given vacuum-tube type is to yield a gain anywhere near the expected value, it is essential that the tube be biased at the quiescent point for which the assumed nominal values of g_m and r_A apply. Methods for biasing a vacuum tube at any required quiescent point are discussed in Chapter 6.

If R_L can be made so large that r_A is negligible in comparison, the equations simplify somewhat. From Eq. 5.46

$$A_V \approx -g_m r_A = -\mu \tag{5.50}$$

provided that

$$R_L \gg r_A.$$

Similarly, from Eq. 5.45,

$$G_T \approx \frac{g_m r_A}{R_L} = \frac{\mu}{R_L}. \tag{5.51}$$

Provided that the approximation in neglecting r_A can be made valid, this manipulation of the equations is significant for the following reason.

As discussed in Section 3.1.2.1, the value of μ is determined largely by the geometry of a vacuum tube's structure. Therefore the value of μ is reproducible from unit to unit of a given tube type and is almost independent of the quiescent conditions. In comparison, the values of g_m and r_A

vary widely with operating point; further, their values are dependent on the age of a tube and its cathode temperature and on a very close quality control of the materials used in fabricating the cathode. Although g_m and r_A are subject to wide variation, their product μ remains relatively constant and is by far the most accurately predictable parameter of an unselected tube. Therefore a transfer function such as Eq. 5.50 or 5.51 which depends on μ (but not on g_m and r_A individually) is more accurately predictable than one that depends on g_m and r_A. Typical maximum tolerances on the values of μ, g_m, and r_A at a given anode voltage and current are

$$0.9\,\mu_A \leq \mu \leq 1.1\,\mu_A, \tag{5.52}$$

$$0.6\,g_{mA} \leq g_m \leq 1.5\,g_{mA}, \tag{5.53}$$

$$0.6\,r_{AA} \leq r_A \leq 1.5\,r_{AA}, \tag{5.54}$$

where μ_A, g_{mA}, and r_{AA} are the values expected for an "average" tube.

It is difficult in practice to realize a circuit in which R_L is large compared with r_A. This is most easily seen by considering some typical parameter values. The value of r_A for a medium-μ triode operating at 10 mA cathode current is of the order of 10 kΩ; a load resistance large in comparison would be 100 kΩ or more. Therefore, according to Eq. 5.42, an anode supply resistor at least 100 kΩ is required, even if the external load is infinite. Because 10 mA anode current flowing through a 100 kΩ resistor produces a 1000-V drop, the supply must be at least 1000 V if this resistor is to be used. While a 1000-V supply is not impossible to produce, it is undesirable in most electronic equipment. High-μ triodes and pentodes, with typical r_A values of 50 kΩ and 500 kΩ respectively at 10 mA cathode current, require supplies of 5 kV and 50 kV as orders of magnitude if R_L is to be large compared with r_A.

A rather poor approximation to high-load conditions for a triode can be obtained by operating the tube at reduced anode current; I_A falls more rapidly than r_A rises (Eq. 3.71) so that the required supply voltage falls. A practical limit is

$$R_L \approx 3\,r_A,$$

although

$$R_L \approx 1.5\,r_A$$

is a more representative design value. The gain lies between outside limits 70 to 150% of the nominal value. However, the great majority of new vacuum tubes lie well inside the tolerance limits of g_m and r_A, so a more likely tolerance on the gain is $\pm 15\%$. Values of load resistor other than $1.5\,r_A$ give only small changes in gain tolerance. Smaller load resistors give slightly greater tolerance and, incidentally, reduce the gain; larger

load resistors have the opposite effects. Low-current operation of a vacuum tube is sometimes referred to as *starvation operation.*

In an amplifier stage using a pentode, it is virtually impossible to realize a load that is other than very small compared with r_A. Equations 5.45 and 5.46 become

$$G_T = \frac{i_A}{v_G} = g_m, \tag{5.55}$$

$$A_V = \frac{v_A}{v_G} = -g_m R_L, \tag{5.56}$$

provided that

$$R_L \ll r_A.$$

The absolute tolerance on the gain of a pentode is the tolerance on g_m, namely 60 to 150% of the nominal value. The great majority of new pentodes, however, yield a gain within $\pm 20\%$ of nominal.

5.2.2 Practical Amplifier Design

The exact equations for a vacuum-tube stage are simple and, consequently, the theoretical analysis of a vacuum-tube amplifier is relatively simple also. However, the production tolerances on vacuum-tube parameters are such that there is typically a spread of $\pm 15\%$ in the gain of a triode and $\pm 20\%$ for a pentode, with wider limits occurring sometimes. Therefore, the fact that the simple equations are exact is of little significance in practical amplifier design.

As an example in the use of the equations for a vacuum tube, consider the design of a single-stage amplifier using one triode section of the type ECC83/12AX7. Type ECC83/12AX7 consists of two nominally identical triodes in one envelope. The voltage gain of the amplifier is to be 50, and the anode supply resistor (which constitutes the only load) is to be 100 kΩ.

The value of μ for a nominal tube is 100, relatively independent of the operating point. Therefore, from Eq. 5.46, designing an amplifier stage with nominal gain 50 involves finding an operating point at which r_A is 100 kΩ. An infinite number of such points exist, but a convenient point, read from manufacturer's published characteristics, is

$$V_A = 150 \text{ V},$$

$$I_K = 0.5 \text{ mA}.$$

Thus the circuit with the supply voltage and quiescent currents and voltages marked is shown in Fig. 5.7. The tolerance on r_A at the given

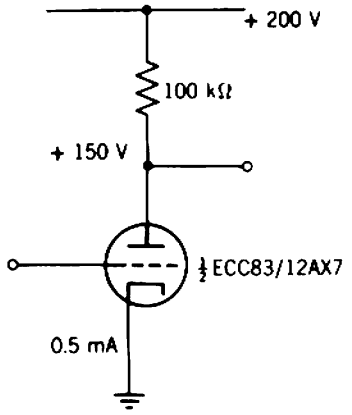


Fig. 5.7 Quiescent currents and voltages required in a single-stage amplifier employing one section of type ECC83/12AX7 double triode to give a nominal voltage gain of 50.

quiescent point is given by Eq. 5.54, and the resultant tolerance on the gain is from 80 to 125%. Added to this is the tolerance on μ, so that the total gain tolerance is approximately 70 to 140%, that is, the gain will lie between 35 and 70. The majority of new ECC83/12AX7 tubes yield gains between 40 and 60.

5.3 THE JUNCTION TRANSISTOR

The mid-band equivalent circuit for a transistor is more complicated than that for the charge-control model assumed in Section 5.1, because the control electrode (the internal base) is not accessible at a device terminal. The internal base is isolated from the so-called base terminal by the base resistance r_B as shown in the equivalent circuit of Fig. 5.8. The general elements of the hypothetical model and the corresponding specific elements of a transistor are listed in Table 5.4.

The base resistance r_B forms a voltage divider at the input with the input resistance of the intrinsic transistor, and consequently the equations involving input voltage will be more complicated than those of the ideal charge-control model. However, r_B is identical in its effect to the source resistance R_S used in the analysis of the ideal model. Therefore the equations for the transistor are as for the ideal model, with r_B replacing R_S, and v_B replacing v_S; for example, from Eq. 5.29,

$$\frac{i_C}{v_B} = \left(\frac{\alpha_N/r_E}{1 + \alpha_N r_B/K_1\beta_N r_E}\right)\left(\frac{r_C}{r_C + R_L}\right),$$

Table 5.4 General Elements and Junction Transistor Parameters

Hypothetical Model	Junction Transistor
r_1	$r_E/(1 - \alpha_N)$
r_f	$\beta_N r_C$
g_m	α_N/r_E mutual conductance
r_2	r_C collector resistance
v_1	$v_{B'}$ internal base voltage
v_2	v_C collector voltage
i_1	i_B base current
i_2	i_C collector current
β	β_N current amplification factor
γ	1 (Eq. 5.12, with $u_1 = 0$ and $u_2 = -2$, from Section 4.2.3)

which can be manipulated into the form

$$G_T = \frac{i_C}{v_B} = \left[\frac{\alpha_N}{(1 - \alpha_N)r_B/K_1 + r_E}\right]\left(\frac{r_C}{r_C + R_L}\right). \tag{5.57}$$

The complete set of transfer and driving-point functions for a transistor stage is given in Table 5.5 (Eqs. 5.57 to 5.66).

Table 5.5 shows that the gain of a transistor stage may depend on many factors, and that in general the calculation of gain is a very tedious process. In particular, the gain depends explicitly or implicitly on β_N. Now, β_N is the transistor parameter that has the largest production tolerance of all; usually β_N is not specified with a tolerance closer than

$$0.7\,\beta_{NA} \leq \beta_N \leq 2\,\beta_{NA}, \tag{5.67}$$

where β_{NA} is the value of β_N expected for an "average" transistor. Because of this large tolerance, the gain of a designable transistor circuit must not

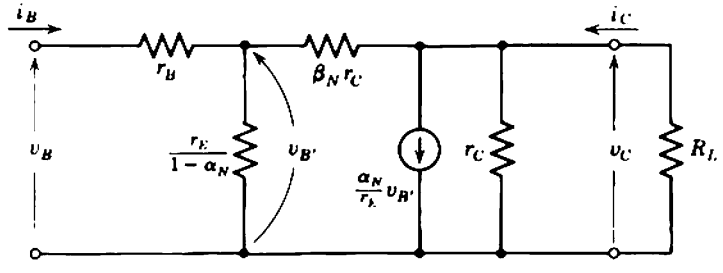


Fig. 5.8 Mid-band equivalent circuit for a transistor with a resistive load.

Table 5.5 Exact Transfer and Driving-Point Functions of a Transistor Stage

$$G_T = \frac{i_C}{v_B} = \left[\frac{\alpha_N}{(1 - \alpha_N)r_B/K_1 + r_E}\right]\left(\frac{r_C}{r_C + R_L}\right) \tag{5.57}$$

where

$$K_1 = \frac{r_C + R_L}{r_C + 2R_L}, \tag{5.58}$$

$$A_V = \frac{v_C}{v_B} = -G_T \times R_L, \tag{5.59}$$

$$A_I = \frac{i_C}{i_B} = \beta_N\left(\frac{r_C}{r_C + 2R_L}\right), \tag{5.60}$$

$$R_T = \frac{v_C}{i_B} = -A_I \times R_L, \tag{5.61}$$

$$R_I = \frac{v_B}{i_B} = r_B + K_1\left(\frac{r_E}{1 - \alpha_N}\right). \tag{5.62}$$

For a source v_S or i_S, of resistance R_S,

$$R_o = K_2 r_C \tag{5.63}$$

where

$$K_2 = \frac{(R_S + r_B) + r_E/(1 - \alpha_N)}{2(R_S + r_B) + r_E/(1 - \alpha_N)}, \tag{5.64}$$

$$\frac{i_C}{v_S} = \left[\frac{\alpha_N}{(1 - \alpha_N)(R_S + r_B)/K_1 + r_E}\right]\left(\frac{r_C}{r_C + R_L}\right), \tag{5.65}$$

$$\frac{i_C}{i_S} = R_S \times \left(\frac{i_C}{v_S}\right). \tag{5.66}$$

depend on β_N except as a second-order correction; a design philosophy that achieves this result is outlined in Section 5.3.1. When circuits are designed on this basis, many of the terms in the equations of Table 5.5 make only small contributions to the transfer and driving-point functions, and may be omitted. Table 5.5 can therefore be simplified very greatly.

Even if a transistor amplifier has been produced without regard to its designability, equations simpler than those of Table 5.5 are adequate to describe its performance. If unselected transistors are used in such an amplifier, the possible error due to the uncertainty in β_N far outweighs that due to neglecting many terms in the equations. Probably the only application for the equations in their exact form is in the accurate calculation of the gain of a nondesignable amplifier that uses selected transistors having accurately known parameters; it is difficult to see any real point in such an exercise.

5.3.1 Outline of Design Philosophy

Consider Eq. 5.57 which gives the transfer conductance. To a very good approximation

$$\beta_N = \frac{\alpha_N}{1 - \alpha_N} \approx \frac{1}{1 - \alpha_N}, \tag{5.68}$$

so that

$$G_T = \frac{i_C}{v_B} \approx \left(\frac{\alpha_N}{r_B/K_1\beta_N + r_E}\right)\left(\frac{r_C}{r_C + R_L}\right). \tag{5.69}$$

The transfer conductance can be made almost entirely independent of β_N by reducing I_E and hence increasing r_E to a point at which

$$r_E \gg \frac{r_B}{K_1\beta_N}$$

when

$$\frac{i_C}{v_B} \approx \frac{\alpha_N}{r_E}\left(\frac{r_C}{r_C + R_L}\right). \tag{5.70}$$

A small dependence on β_N must always remain because α_N occurs in Eq. 5.70 and β_N is implicit in α_N. With the normal order of magnitude of β_N, however, the tolerance on α_N is only a few percent:

$$\alpha_N = \frac{\beta_N}{1 + \beta_N} \approx 1. \tag{5.71}$$

If the tolerance on β_N is as given by Eq. 5.67, the relation between $(r_B/K_1\beta_{NA})$ and r_E required for typical tolerances on transfer conductance is given in Table 5.6.

Table 5.6 Tolerance on Transfer Conductance

Required Tolerance		Necessary Value of $\frac{r_B/K_1\beta_{NA}}{r_E}$
Exact Percentage	Approx. dB	
95 to 105%	±0.5 dB	$\frac{1}{9}$
91 to 111%	±1.0 dB	$\frac{1}{4}$
80 to 133%	±2.5 dB	1

Clearly, the process of reducing I_E so as to make G_T independent of β_N can be used to reduce the dependence on r_B and K_1 to any desired extent. Equally, the dependence of G_T on r_C can be reduced by making

$$R_L \ll r_C,$$

and this will also reduce the uncertainty in K_1 by causing K_1 to approach unity. Typical production tolerances on r_B and r_C are

$$0.7\, r_{BA} \leq r_B \leq 1.4\, r_{BA}, \tag{5.72}$$

$$0.7\, r_{CA} \leq r_C \leq 1.4\, r_{CA}, \tag{5.73}$$

where r_{BA} and r_{CA} are the values of r_B and r_C expected for an "average" transistor. Ultimately, the transfer conductance can be made to depend on r_E alone. But r_E is reproducible, being given by

$$r_E = \frac{kT}{qI_E}. \tag{5.74}$$

Thus the transfer conductance can be made to depend only on the operating point and temperature, and designable circuits can be produced which make use of the predictable transfer conductance.

In comparison, no choice of quiescent point and load resistance can eliminate the direct proportionality between current gain and β_N. The tolerance on current gain per stage is thus the 3:1 tolerance of β_N. Transistors, however, can be (and often are) used as current amplifiers in applications where a large tolerance on gain is acceptable. Examples are amplifiers in which negative feedback is applied around several stages to stabilize their combined gain (see Chapters 10 and 14) and amplifiers employing automatic gain control (which are beyond the scope of this book).

5.3.2 Practical Simplifications

It is very difficult to achieve a load resistance R_L for a transistor that is at all comparable in magnitude with r_C. As discussed in Section 5.1.5, the largest possible load is the collector supply resistor R_C, corresponding to an infinite external load. The value of r_C for a transistor operating at 1 mA emitter current is unlikely to be less than 50 kΩ (100 to 200 kΩ being more typical). If the supply is 20 V, the largest supply resistor that can possibly be used is 20 kΩ (10 kΩ being a more likely maximum). Reducing the current allows a larger R_C, but also raises r_C of the transistor in direct proportion. The extreme range of values for K_1, $r_C/(r_C + R_L)$, and $r_C/(r_C + 2R_L)$ are thus

$$0.86 \leq K_1 \leq 1$$

$$0.83 \leq \frac{r_C}{r_C + R_L} \leq 1$$

$$0.71 \leq \frac{r_C}{r_C + 2R_L} \leq 1$$

Continuing the discussion of Section 5.3.1, it is reasonable to assume that, in a designable amplifier, these three terms are very close to unity. This implies that the elements r_C and $\beta_N r_C$ can be omitted from the equivalent circuit, to give the simplified form shown in Fig. 5.9.

The equivalent circuit of Fig. 5.9 is used almost without exception for mid-band calculations throughout the remainder of this book. Even if a circuit has not been designed with the object of minimizing the effects of r_C and $\beta_N r_C$, the error in neglecting them is small—only 1 or 2% in all but the most abnormal circumstances. Numerical examples given later in this chapter compare the results obtained from the complete and simplified circuits. The simplified circuit does overestimate the gain when the voltage gain approaches the voltage amplification factor (open-circuit voltage gain) of the transistor:

$$\mu = \alpha_N\left(\frac{r_C}{r_E}\right). \tag{5.75}$$

This occurs only when R_L is abnormally large. For almost all transistors μ is greater than 1000, and values as high as 10^4 are not uncommon.

The transfer conductance, current gain, and input resistance calculated from the simplified equivalent circuit are

$$G_T = \frac{i_C}{v_B} = \frac{\alpha_N}{(1-\alpha_N)r_B + r_E} \approx \frac{\alpha_N}{r_B/\beta_N + r_E}, \tag{5.76}$$

$$A_I = \frac{i_C}{i_B} = \beta_N, \tag{5.77}$$

$$R_I = \frac{v_B}{i_B} = r_B + \frac{r_E}{1-\alpha_N} \approx r_B + \beta_N r_E. \tag{5.78}$$

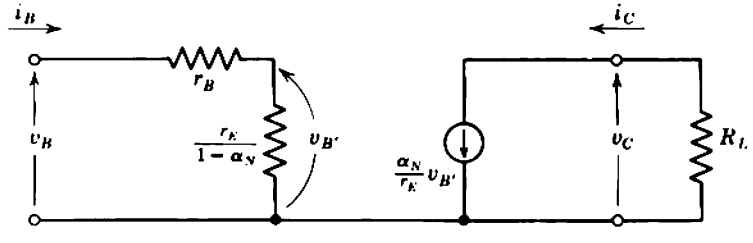


Fig. 5.9 Simplified mid-band equivalent circuit for a transistor with a resistive load.

If the stage has been designed so that its transfer conductance is independent of β_N (i.e., if I_E is reduced to the point at which r_E is much greater than r_B/β_N), Eq. 5.76 reduces to

$$G_T = \frac{i_C}{v_B} = \frac{\alpha_N}{r_E}. \tag{5.79}$$

This expression, however, always overestimates the gain of practical stages. It has been found by experience that a working approximation for G_T is the empirical relation

$$G_T = \frac{i_C}{v_B} \approx \frac{0.8}{r_E}. \tag{5.80a}$$

But r_E is given by Eq. 5.74; substituting values gives

$$G_T = \frac{i_C}{v_B} \approx 32 I_E \text{ mA/V} \tag{5.80b}$$

at room temperature, for I_E in mA.

The gain of a transistor fed from a source of finite resistance is of considerable importance in connection with cascaded stages. Using the current-generator representation of the source, the output current follows from Eq. 5.66 and the simplified equivalent circuit as

$$\frac{i_C}{i_S} = \frac{\alpha_N R_S}{(1 - \alpha_N)(R_S + r_B) + r_E}. \tag{5.81}$$

Two algebraic manipulations are useful. First, if the circuit is designed with

$$r_E \gg (1 - \alpha_N)(R_S + r_B),$$

it is convenient to transform Eq. 5.81 into

$$\frac{i_C}{i_S} \approx \frac{\alpha_N R_S}{r_E}\left[\frac{1}{1 + (R_S + r_B)/\beta_N r_E}\right]. \tag{5.82}$$

This corresponds to feeding the transistor from a low-resistance or approximate-voltage source of magnitude $i_S R_S$. The first term in Eq. 5.82 results if the ideal expression for G_T is used (Eq. 5.79), whereas the second term is a near-unity correction to account for the finite value of β_N. The ratio i_C/i_S is the current gain of the complete system and can be made stable by choosing r_E sufficiently large. The system current gain i_C/i_S is quite different from the stage current gain i_C/i_B; the latter is equal to β_N and is

unstable. The second manipulation is useful if R_S is much greater than R_i of the transistor, that is, if

$$R_S \gg r_B + \beta_N r_E,$$

when

$$\frac{i_C}{i_S} \approx \beta_N \left[\frac{1}{1 + (r_B + \beta_N r_E)/R_S} \right]. \tag{5.83}$$

This corresponds to feeding the transistor from a high-resistance or approximate-current source of magnitude i_S. In this case substantially all the current of the source generator flows into the stage and the current gain

Table 5.7 Transfer and Driving-Point Functions of a Transistor Stage

(a) *Transfer Conductance*

(i) exact—seldom used:

$$G_T = \frac{i_C}{v_B} = \left[\frac{\alpha_N}{(1 - \alpha_N) r_B / K_1 + r_E} \right] \left(\frac{r_C}{r_C + R_L} \right), \tag{5.57}$$

where

$$K_1 = \frac{r_C + R_L}{r_C + 2R_L}; \tag{5.58}$$

(ii) for small R_L—designable circuits:

$$G_T = \frac{i_C}{v_B} = \frac{\alpha_N}{(1 - \alpha_N) r_B + r_E} \approx \frac{\alpha_N}{r_B/\beta_N + r_E}; \tag{5.76}$$

(iii) order of magnitude ($\pm 20\%$):

$$G_T = \frac{i_C}{v_B} \approx \frac{0.8}{r_E}. \tag{5.80}$$

(b) *Current Gain*

(i) exact—seldom used:

$$A_I = \frac{i_C}{i_B} = \beta_N \left(\frac{r_C}{r_C + 2R_L} \right); \tag{5.60}$$

(ii) for small R_L—designable circuits:

$$A_I = \frac{i_C}{i_B} = \beta_N. \tag{5.77}$$

(c) *Input Resistance*

(i) exact—seldom used:

$$R_i = \frac{v_B}{i_B} = r_B + K_1 \left(\frac{r_E}{1 - \alpha_N} \right); \tag{5.62}$$

(ii) for small R_L—designable circuits:

$$R_i = \frac{v_B}{i_B} = r_B + \frac{r_E}{1 - \alpha_N} \approx r_B + \beta_N r_E. \tag{5.78}$$

of the system approaches the current gain β_N of the transistor. The second term in Eq. 5.83 is a near-unity correction to account for the current in R_S.

The simplified equivalent circuit of Fig. 5.9 suggests that the output resistance of a transistor is infinite and independent of the source resistance. Although this is obviously not exact, it is a good working approximation as the output resistance of a transistor is very large indeed compared with all other circuit impedances. For the rare cases in which the value of R_o is of interest, the complete equivalent circuit and Eqs. 5.63 and 5.64 must be used.

Table 5.7 summarizes the formulas for G_T, A_I, and R_i of a transistor amplifier stage.

5.3.3 Practical Amplifier Design

The exact equations in Table 5.5 for a transistor are much more complex than the corresponding equations in Table 5.3 for a vacuum tube. A reasonable comment would be that analysis of a transistor amplifier can be a formidable undertaking, whereas analysis of a vacuum-tube amplifier is quite simple. This section shows how circuit design can be simpler for transistors than for tubes.

Consider the design of a transistor amplifier to give a voltage gain of 50 into a load resistance of 4.7 kΩ. Equation 5.59 shows that the transfer conductance must be 10.64 mA/V. Equation 5.80 gives an order of magnitude for G_T and substitution of the data gives

$$r_E = 75.2\ \Omega.$$

But

$$r_E = \frac{kT}{qI_E}.$$

Therefore at room temperature

$$I_E = 0.33\ \text{mA}.$$

Thus *any transistor whatsoever will give approximately the required voltage gain, provided it is operated at* 0.33 mA *emitter current.* For this application all transistors are to a first order identical.

Before the more accurate equations can be used, it is necessary to choose a transistor type. From Table 4.4 a uniform-base audio frequency transistor operating at $\frac{1}{3}$ mA might have typical parameters

$$\beta_{NA} = 40,$$

$$r_B = 300\ \Omega,$$

$$r_C = 400\ \text{k}\Omega.$$

Substitution of these data into the equations of Table 5.7 gives the voltage gain as:

(i) exact	54.8,
(ii) for small R_L	55.4,
(iii) order-of-magnitude	50.0.

Notice that the inaccuracies in the approximate calculations are only 1.2 and 10.4% for (ii) and (iii), respectively. A much greater source of inaccuracy is the tolerances on β_N and r_B; (the value of r_C is so large that even a 2:1 change makes only 1% difference to the gain). Typical outside limits for β_N and r_B might be

$$25 \leq \beta_N \leq 85,$$

$$200\ \Omega \leq r_B \leq 500\ \Omega.$$

Substitution of the extreme values into the small-load expression for voltage gain gives

$$\left.\begin{aligned} \beta_N &= 85 \\ r_B &= 200\ \Omega \end{aligned}\right\} \quad A_{V(\max)} = 59.8,$$

$$\left.\begin{aligned} \beta_N &= 25 \\ r_B &= 500\ \Omega \end{aligned}\right\} \quad A_{V(\min)} = 45.7.$$

The variation of gain with transistor parameters is much greater than the errors due to the approximation in the equations. From an engineering point of view, the approximations are therefore completely justified.

The gain of the amplifier is about 10% higher than required, and estimated by the order-of-magnitude equation. The required gain 50 can be

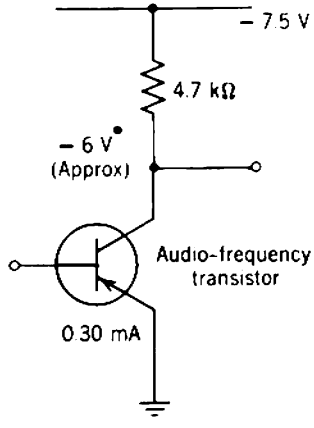


Fig. 5.10 Quiescent currents and voltages required in a single-stage amplifier employing an audio transistor to give a nominal voltage gain of 50.

obtained from "average" transistors to within 1 to 2% if the emitter current is reduced by 10%; the required emitter current is 0.30 mA, and the quiescent voltages are shown in Fig. 5.10. The maximum expected spread in gain is 40 to 55, with the great majority of transistors giving a gain between 47 and 52. Thus, the design of a transistor amplifier with specified voltage gain is simpler than that of a vacuum-tube amplifier, as it is not necessary to consult characteristic curves to obtain small-signal data.

5.4 CASCADED VACUUM-TUBE STAGES

No complications arise at mid-band frequencies when a number of vacuum-tube stages are connected together in cascade to form a multistage amplifier. In theory, the load resistance for any vacuum tube in a cascade is the parallel combination of the resistance R_P (the parallel combination of all biasing resistors) and the input resistance R_i for the following vacuum tube. Since recombination in a vacuum tube is negligible, its input resistance becomes infinite. In practice, therefore, the load for a vacuum tube in a cascade is its own anode supply resistor in shunt with the grid supply resistor for the following tube, and the voltage gains of the individual stages multiply directly to give the over-all gain of the amplifier:

$$A_V = \left\{\frac{g_{m1}r_{A1}R_{P1}}{r_{A1} + R_{P1}}\right\} \times \left\{\frac{g_{m2}r_{A2}R_{P2}}{r_{A2} + R_{P2}}\right\} \times \cdots. \tag{5.84}$$

The uncertainty in the gain of any one stage is discussed in Section 5.2.1. When a number of stages are cascaded, the possible percentage uncertainty in gain increases as the number of stages is increased. Against this, it is unlikely with a random selection of tubes that all errors will be in the same direction, so that the probable error decreases. The error distribution curve becomes a sharper peak with broader skirts as the number of stages increases, provided that the tube selection is random. It is likely, however, that the tube selection in mass produced amplifiers will not be random; all the tubes of any one manufacturing batch tend to have parameter values with a bias in the same direction.

5.4.1 Practical Example

As an example, two amplifier stages of the type designed in Section 5.2.2 were connected together in cascade. Figure 5.11 shows the elemental circuit diagram; the components for biasing the tubes at the required

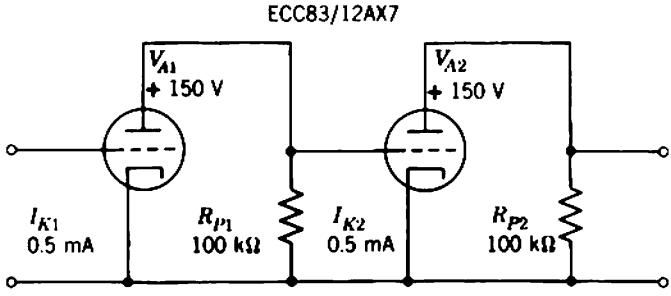


Fig. 5.11 Elemental circuit diagram for a two-stage vacuum-tube amplifier. The nominal gain is 2500.

quiescent point are omitted. The calculated gain per stage is 50, so the gain of the two stages is 2500.

Figure 5.12 is a histogram showing the gain of 89 new tubes from one production batch. The gain of two stages spreads over about ± 10%, with the center about 20% above the nominal value. Measurements on limited numbers of tubes from different batches suggest that ± 10% is a typical spread per batch; for the greater part, batch centers of new tubes lie above the nominal value, the highest observed being +40%. If tubes approaching the end of their life are included, the absolute limits suggested by Eqs. 5.52 to 5.54 appear reasonable.

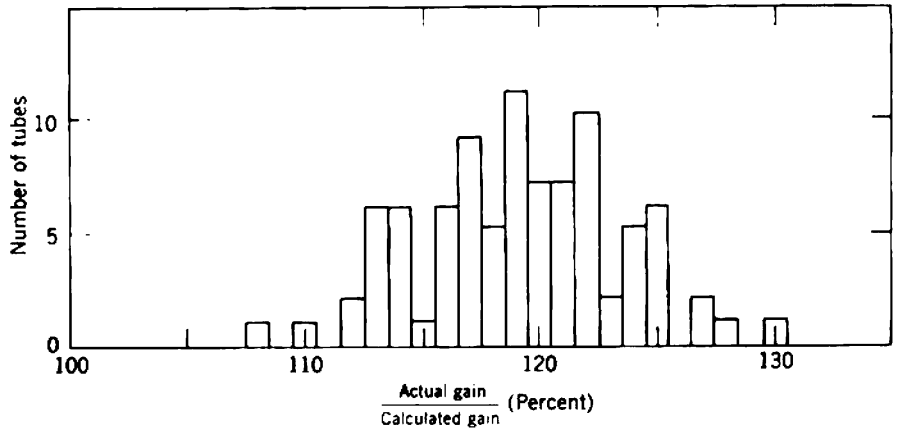


Fig. 5.12 Histogram showing the gain of 89 new tubes from one production batch in the circuit of Fig. 5.11.

5.5 CASCADED TRANSISTOR STAGES

In direct contrast with a vacuum tube, a transistor has a finite and relatively small input resistance, so that the load resistance for any transistor (except the last) in a multistage amplifier includes the input resistance of the next transistor. In the general case, the voltage gain of a multistage transistor amplifier can be calculated in the following way, provided the transistor parameters are known:

1. The voltage gain of the last stage can be calculated from Eq. 5.59. The output dividing factor $r_C/(r_C + R_L)$ and K_1 may or may not be significant, depending on the order-of-magnitude of the load; the approximate expression (Eq. 5.76) for G_T may be of adequate accuracy.
2. The input resistance of the last stage can be calculated from Eq. 5.62. Again, the factor K_1 may or may not be significant, so that the approximate form (Eq. 5.78) may be adequate.
3. The load resistance for the penultimate stage can now be calculated. This load is the parallel combination of the supply resistor R_P and the input resistance of the last stage.
4. The voltage gain of the penultimate stage can now be calculated as in step 1 above.

This cycle repeats as many times as there are stages. It is necessary to begin a calculation of gain with the output stage, and work back one stage at a time toward the input. It is not possible, without very great algebraic complexity, to write down a general equation (corresponding to Eq. 5.84 for vacuum-tube circuits) into which the data for the individual transistors can be substituted to give the over-all gain. The gain of any stage depends to some extent on the values of all parameters for all later stages in the amplifier.

Because the input resistance of a transistor stage depends critically on β_N, its value is not known accurately unless a selected transistor is used. Therefore, even though the transfer conductance of a stage can be designed to be independent of β_N (by making r_E very much greater than $r_B/K_1\beta_N$), its voltage gain in a cascade may be quite unpredictable because its load depends on the value of β_N for the following transistor. Expressed another way, there is interaction between stages; the voltage gain of one transistor depends on β_N of the following transistor.

5.5.1 Technique for Designability

A multistage transistor amplifier can be designed so that its gain is independent of β_N and interaction between stages by making the supply

resistors for all stages sufficiently small. If R_P is much less than the expected "average" input resistance of the following stage, the total load for any transistor is very nearly equal to R_P, independent of β_N for the following transistor. The voltage gain is therefore

$$A_V \approx -G_T \times R_P. \tag{5.85}$$

However, G_T can be designed to be stable; A_V can therefore be stable also.

This technique for stabilizing voltage gain v_o/v_i also stabilizes the other transfer functions (i_o/v_i, i_o/i_i, and v_o/i_i) between all points within the amplifier. Consider the amplifier shown in Fig. 5.13, where it is assumed

$$R_P \ll R_i,$$

so that the voltage gain per stage is given by Eq. 5.85. Each "stage" is regarded as consisting of a transistor in conjunction with all the supply resistors constituting R_P across its output terminals. A stage, however, might equally well (though unconventionally) be regarded as consisting of a transistor in conjunction with R_P at its input terminals. Such a stage has a stable current gain; the input current develops a known input voltage across R_P (in shunt with the much larger R_i), and this input voltage excites the stable G_T of the transistor to give a known output current:

$$v_i = i_i \times R_P,$$

$$i_o = v_i \times G_T;$$

therefore

$$A_I = \frac{i_o}{i_i} = R_P \times G_T. \tag{5.86}$$

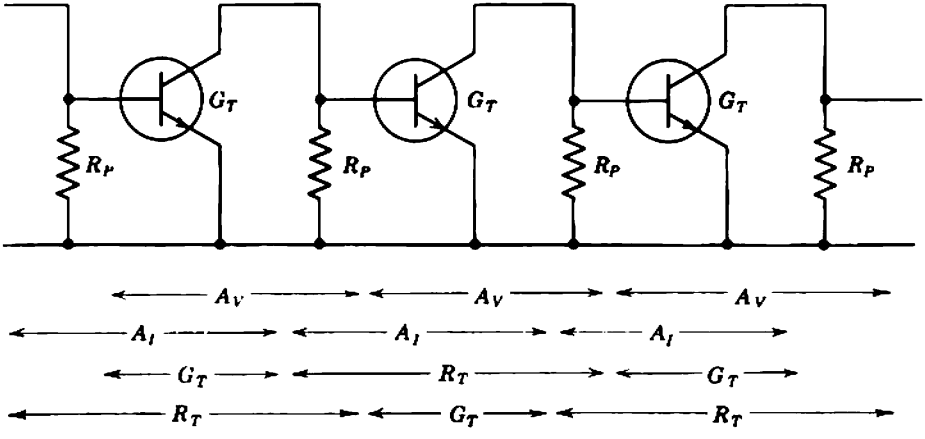


Fig. 5.13 Break-up of a multistage transistor amplifier, showing that all transfer functions of all stages are stable provided that $R_P \ll R_i$.

Notice that the current gain of this unconventional stage is quite different from the current gain of a normal stage; the latter is approximately β_N, and can never be stable.

Two other breakups of the amplifier into stages are possible. In each case the stages have alternately stable transfer conductance and transfer resistance. In a G_T stage the transistor is considered alone and its transfer conductance is stable as in Section 5.3.1. In an R_T stage the transistor is associated with the supply resistors at both its input and output; as in the current-gain representation, the input current develops a known voltage across R_P at the transistor's input terminals, this voltage excites the stable G_T of the transistor to give a current output and, finally, this output current develops a known voltage across R_P at the transistor's output terminals:

$$v_i = i_i \times R_P,$$

$$i_o = v_i \times G_T,$$

$$v_o = -i_o \times R_P;$$

therefore,

$$R_T = \frac{v_o}{i_i} = -R_P \times G_T \times R_P. \tag{5.87}$$

Clearly, the way in which the stages within an amplifier are broken up cannot affect a calculation of over-all gain. In certain circumstances, however, one of the representations can lead to a much simpler calculation than the others. For example, the limiting gain per stage with a given accuracy is calculated in the next section on the current-gain basis; the feedback amplifiers discussed in Chapter 11 are best represented by alternate transfer conductance and transfer resistance. Notice that vacuum-tube amplifiers can also be broken up into these four possible representations; this is left as an exercise for the reader.

One important physical concept results from these alternative representations of multistage amplifiers. The breakup into stages must be dimensionally consistent; that is, the output signal type (current or voltage) from one stage must be the same as the input signal type for the following stage. It is meaningless, for example, to multiply the voltage gain of one stage by the current gain of the next. For a two-stage amplifier, 16 combinations of transfer function are possible, but only 8 of them are meaningful:

$G_{T1} \times G_{T2}$ is meaningless,

$G_{T1} \times A_{V2}$ is meaningless,

$G_{T1} \times A_{I2}$ = over-all G_T,

$$G_{T1} \times R_{T2} = \text{over-all } A_V,$$

$$A_{V1} \times G_{T2} = \text{over-all } G_T,$$

$$A_{V1} \times A_{V2} = \text{over-all } A_V,$$

$$A_{V1} \times A_{I2} \text{ is meaningless,}$$

$$A_{V1} \times R_{T2} \text{ is meaningless,}$$

$$A_{I1} \times G_{T2} \text{ is meaningless,}$$

$$A_{I1} \times A_{V2} \text{ is meaningless,}$$

$$A_{I1} \times A_{I2} = \text{over-all } A_I,$$

$$A_{I1} \times R_{T2} = \text{over-all } R_T,$$

$$R_{T1} \times G_{T2} = \text{over-all } A_I,$$

$$R_{T1} \times A_{V2} = \text{over-all } R_T,$$

$$R_{T1} \times A_{I2} \text{ is meaningless,}$$

$$R_{T1} \times R_{T2} \text{ is meaningless.}$$

5.5.2 Gain and Stability

While the technique of reducing the value of R_P can be used to stabilize the gain of a transistor amplifier, it has the disadvantage of reducing the gain per stage. As R_P is reduced in comparison with R_i of the following transistor, the gain becomes more stable but also becomes smaller. It follows that an upper limit exists to the gain that can be realized per stage for a given percentage uncertainty.

Consider the equivalent circuit shown in Fig. 5.14 for two stages in a cascade. As was shown, the amplifier can be broken up in a number of

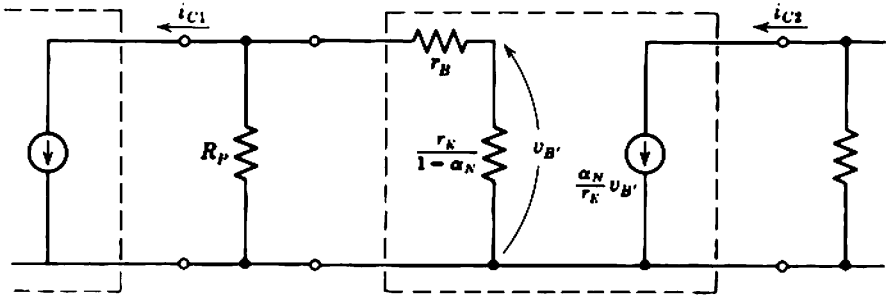


Fig. 5.14 Equivalent circuit for two stages within a cascade.

different ways for the purpose of analysis. The most convenient for the present application is the current-gain representation in which each transistor is associated with the supply resistor at its input. The current gain from one collector to the collector of the following transistor is

$$\frac{i_{C2}}{i_{C1}} = -\frac{\alpha_N R_P}{(1-\alpha_N)(R_P + r_B) + r_E} \approx -\frac{\alpha_N R_P}{(R_P + r_B)/\beta_N + r_E}. \qquad (5.88)$$

This equation is identical in form to Eq. 5.81 for a transistor fed from a source of finite internal resistance. The first transistor may be considered as the source current generator for the second transistor and R_P is the source resistance.

If the amplifier is designed so that

$$R_P \gg r_B + \beta_N r_E,$$

Eq. 5.88 can be manipulated into the form

$$\frac{i_{C2}}{i_{C1}} = -\beta_N\left[\frac{1}{1 + (r_B + \beta_N r_E)/R_P}\right] \approx -\beta_N, \qquad (5.89)$$

where the approximate form applies if R_P is very large. This corresponds to using the transistor as a current amplifier. Most of the (signal) collector current from one transistor flows into the base of the next transistor and, in the limit, the current gain per stage (or voltage gain if the normal representation is used) approaches β_N. However, the tolerance on the gain is large and, in the limit, it approaches the tolerance on β_N. Transistors are often used in the current mode, but the 3:1 tolerance on gain per stage must be accepted.

If the amplifier is designed so that

$$\frac{R_P + r_B}{\beta_N} \ll r_E,$$

the gain becomes almost independent of β_N. This is the mode of operation that makes use of the stable G_T and was discussed qualitatively in Section 5.5.1. If the tolerance on β_N is as suggested by Eq. 5.67, then Table 5.8 gives the value of $(R_P + r_B)/\beta_{NA} r_E$ required for various degrees of stability. In addition to requiring $(R_P + r_B)$ to be less than a certain amount, it is desirable that

$$R_P \gg r_B.$$

This increases the gain available with any specified tolerance because R_P appears in the numerator of Eq. 5.88. Table 5.8 also lists the available gains in terms of average β_N. Notice that Table 5.8 lists the conditions for a certain tolerance on gain even with worst-case transistors. The

Table 5.8 Tolerances and Limiting Values of Gain per Stage (Based on the Assumption $0.7\beta_{NA} \leq \beta_N \leq 2\beta_{NA}$)

Required Tolerance		Necessary Value of $\frac{R_P + r_B}{\beta_{NA} r_E}$	Available Gain
Exact Percentage	Approx. dB		
95 to 105%	±0.5 dB	$\frac{1}{9}$	$\frac{\beta_{NA}}{10}\left(\frac{R_P}{R_P + r_B}\right)$
91 to 111%	±1 dB	$\frac{1}{4}$	$\frac{\beta_{NA}}{5}\left(\frac{R_P}{R_P + r_B}\right)$
80 to 133%	±2.5 dB	1	$\frac{\beta_{NA}}{2}\left(\frac{R_P}{R_P + r_B}\right)$
70 to 200%	±5 dB	∞	$\beta_{NA}\left(\frac{R_P}{R_P + r_B}\right)$

majority of production transistors lie well within these worst-case tolerance limits. A sound rule of thumb is to design everyday amplifiers with

$$R_P \approx \tfrac{1}{3}\beta_N r_E,$$
$$R_P \geq 3r_B.$$

For 90% of production transistors, the spread in gain per stage (current or voltage) is ±5% about

$$A \approx \tfrac{1}{4}\beta_{NA}.$$

5.5.3 Input and Output Stages

The gain limits per stage set out in Table 5.8 do not apply for the input and output stage of an amplifier.

The transfer conductance of the output stage cannot be made greater than that of any stage within an amplifier by increasing its emitter current, because its input resistance would fall and react on the penultimate stage. Its load resistance, however, does not involve the input resistance of any following stage and consequently can be made relatively large without becoming unpredictable. The stable voltage gain can therefore be larger than the limiting value given in Table 5.8.

The load resistance of an input stage cannot be increased because it involves the input resistance of the second stage, and would therefore become unpredictable. If the input stage is fed from a low resistance source, however, its emitter current and hence transfer conductance may be increased and the voltage gain can be increased beyond the limiting value suggested in Table 5.8.

5.5.4 Practical Example

A number of different methods and formulas for calculating the gain of a transistor amplifier are given in the previous sections. The most appropriate for amplifier design is largely a matter of personal choice.

The most direct, though not the quickest method for calculating the gain is to use the expressions for G_T and R_i of an isolated transistor (Eqs. 5.76 and 5.78) together with the voltage-gain approach set out in the introduction to Section 5.5:

$$A_V = (G_{T1} \times R_{L1}) \times (G_{T2} \times R_{L2}) \times \cdots \times (G_{TN} \times R_{LN}), \quad (5.90)$$

remembering that, in general, R_L for any stage involves the input resistance of the following transistor. The current-gain-between-collectors approach based on Eq. 5.88 is rather more elegant in that it requires less steps in a numerical calculation:

$$A_V = G_{T1} \times A'_{I2} \times A'_{I3} \times \cdots \times A'_{IN} \times R_{LN}. \quad (5.91)$$

Against this, the approach involves equations other than those for an isolated transistor, and it is necessary to avoid confusion between the normal stages and the unconventional stages whose gain is indicated by a prime in Eq. 5.91. Other methods may suggest themselves as more suitable for special applications. In each case, however, the equivalent circuit applicable for small values of load resistance (Fig. 5.9) should be used.

Figure 5.15 is the outline of a practical circuit for two cascaded transistor stages. The components for biasing the transistors at the desired operating points are omitted. As much gain as possible with $\pm 10\%$ absolute tolerance per stage is required into a 4.7-kΩ load resistor. The transistors are general purpose audio types with the average parameters

$$\beta_N \approx 40,$$
$$r_B \approx 300\ \Omega.$$

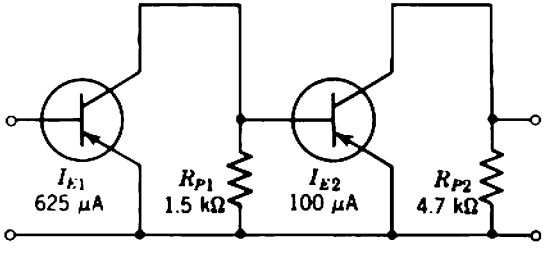


Fig. 5.15 Elemental circuit diagram for a two-stage transistor amplifier. The nominal gain is 480.

For the first stage, from Table 5.6, it is required that

$$\frac{r_{B1}}{r_{E1}} = 0.2\,\beta_{NA} \approx 8$$

for $\pm 10\%$ stability in G_T. Therefore, choose

$$r_{E1} = 40\ \Omega$$

and thus

$$I_{E1} = 625\ \mu\text{A}.$$

The transfer conductance follows from Eq. 5.76 as

$$G_{T1} = \frac{\alpha_{N1}}{r_{B1}/\beta_{N1} + r_{E1}} = 20.6\ \text{mA/V}.$$

For the second stage, from Table 5.8, it is required that

$$\frac{R_{P1} + r_{B2}}{r_{E2}} = 0.2\,\beta_{NA} \approx 8.$$

Choose R_{P1} as 1.5 kΩ (an intelligent guess) and therefore

$$R_{P1} + r_{B2} = 1.8\ \text{k}\Omega.$$

Thus choose

$$r_{E2} = 250\ \Omega,$$

$$I_{E2} = 100\ \mu\text{A}.$$

The transfer conductance and input resistance follow from Eqs. 5.76 and 5.78 as

$$G_{T2} = \frac{\alpha_{N2}}{r_{B2}/\beta_{N2} + r_{E2}} = 3.81\ \text{mA/V},$$

$$R_{i2} = r_{B2} + \beta_{N2} r_{E2} = 10.3\ \text{k}\Omega.$$

The load resistance for the first stage is

$$R_{L1} = R_{P1} \parallel R_{i2} = 1.3\ \text{k}\Omega.$$

Therefore

$$A_{V1} = G_{T1} \times R_{L1}$$
$$= 26.8.$$

The load for the second transistor is its 4.7-kΩ collector supply resistor and therefore

$$A_{V2} = 17.9.$$

Finally, the total gain is

$$A_V = A_{V1} \times A_{V2} = 480,$$

at a temperature (16°C) for which

$$\frac{kT}{q} = 25\ \text{mV}.$$

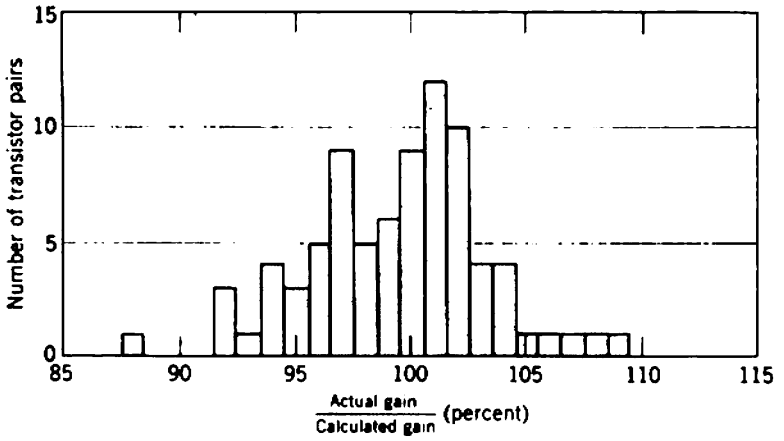


Fig. 5.16 Histogram showing the gain of 82 random pairs of audio transistors in the circuit of Fig. 5.15.

Changing the value of R_{P1} from 1.5 kΩ does not appreciably change the available gain. Increasing its value very slightly increases the available gain, but this necessitates a smaller emitter current in the second transistor which in turn reduces the available signal output swing. Reducing the value of R_{P1} has the opposite effects. Figure 5.16 is a histogram showing the spread in gain for 82 random pairs of transistors from various different manufacturers' general purpose audio-frequency types. The results show that the voltage gain is within ± 5% of the expected value for 85% of audio transistors. This tolerance is slightly closer than that for new vacuum tubes of one batch and much closer than that for a random selection of vacuum tubes of any age.

5.6 A COMPARISON OF VACUUM TUBES WITH TRANSISTORS

If precision of gain is the prime requirement of an amplifier then, at least from the discussion in this chapter, the transistor is superior to the vacuum tube. Little can be done to stabilize the gain of simple vacuum-tube amplifiers of the type which have been considered in this chapter, whereas the gain of a transistor amplifier can be stabilized almost as accurately as required. The process of stabilizing the gain of a transistor, however, greatly reduces the gain per stage. Furthermore, the dc emitter current must be held within very close limits by the biasing circuits. Finally, the emitter current must be small and only a small output signal can be obtained before the distortion becomes gross; this limitation is discussed in detail in Chapter 9. Thus too much importance should not

be attached to the possibility of achieving a highly stable gain from simple transistor amplifiers. Other preferable means, based on the use of negative feedback, exist and are applicable to both vacuum tubes and transistors; the second half of this book deals mainly with the design of feedback amplifiers.

If the magnitude of the gain per stage is the prime requirement of an amplifier without regard to stability, then the pentode is in a class by itself. Typical high-slope pentodes (g_m = 5 mA/V at I_K = 10 mA) can achieve a gain per stage of 300 to 400 when operated at much reduced anode current from a 300-V supply with a load of 0.5 MΩ. The more exotic types (g_m = 15 mA/V at I_K = 10 mA) can achieve a gain in excess of 1000. Typically, the tolerance on gain is ±30%. Triodes or present transistors cannot approach these values of gain per stage.

In applications for which ±15% tolerance on gain per stage is acceptable the choice between tubes and transistors is primarily one of convenience, as roughly the same number of stages is required. From a practical point of view the lower supply voltages and power requirements, and the complete absence of a heater supply may swing the decision in favor of the transistor. In special environments, such as at high temperatures or in high radiation, the use of vacuum tubes may be mandatory. Transistors are preferred when equipment is to be subjected to mechanical shocks. One theoretical distinction between vacuum-tube and transistor circuits is the difference in impedance level; the impedance of a normal circuit using vacuum tubes is higher by one or two orders of magnitude. A transistor amplifier may therefore be more suitable if the source and load impedances are low, whereas vacuum tubes may be more suitable if the impedances are high. As a very rough guide, 10 kΩ might be taken as a line of demarkation between high and low impedances, but it is entirely practicable to use transistors with source and load impedances above this figure or to use vacuum tubes with impedances less than 10 kΩ.

5.7 OTHER AMPLIFIER CONFIGURATIONS

The amplifier circuits considered so far are said to be of the common-emitting-electrode configuration, because the emitting electrode (the cathode of a tube or emitter of a transistor) is the device terminal common to both input and output circuit meshes. Charge-controlled devices are capable of power gain in two other configurations, however, the common-control-electrode and common-collecting-electrode. Although these configurations are occasionally useful, they are very much abused. There are few instances in low-pass amplifier circuits when one of the common-emitting

electrode feedback stages considered in Chapter 11 will not prove more satisfactory. Nevertheless, common-control- and common-collecting-electrode arrangements are so widely used that they should be mentioned. The transfer and driving-point functions are listed in Tables 5.9 to 5.13, and the circuit characteristics are discussed in Sections 5.7.1 and 5.7.2 below. It is left as an exercise for the reader to verify the formulas listed in the tables, and consider their physical interpretation.

5.7.1 Common-Control-Electrode

The *common-grid* and *common-base* amplifier circuits are shown in outline in Fig. 5.17. Note that, although a triode is shown in Fig. 5.17*a*, a pentode may be used equally well. These circuits are characterized by a low input resistance and a high output resistance, where the terms high and low are relative to the common-cathode or common-emitter configurations. The transfer conductance and voltage gain are very nearly the same as for common-cathode or common-emitter stages, whereas the current gain is close to unity.

Tables 5.9 and 5.10 list the transfer and driving-point functions. Table 5.9 gives three expressions for G_T and R_i. The first of these is an exact expression. The second is an approximate expression, whose accuracy is

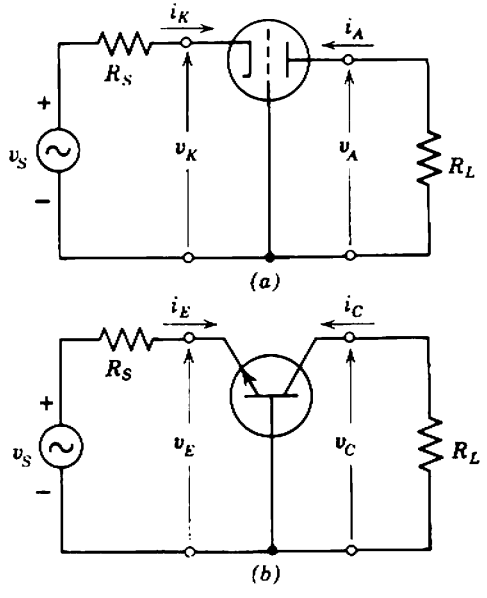


Fig. 5.17 Elemental circuit diagrams for common-control-electrode amplifier stages: (*a*) common grid; (*b*) common base.

Table 5.9 Common-Grid Amplifier Stage

(a) *Transfer Conductance*

(i) exact:

$$G_T = \frac{i_A}{v_K} = -\frac{\mu + 1}{r_A + R_L}; \tag{5.92}$$

(ii) approximate, for $\mu \gg 1$:

$$G_T = \frac{i_A}{v_K} = -g_m\left(\frac{r_A}{r_A + R_L}\right); \tag{5.93}$$

(iii) approximate, for $\mu \gg 1$, $R_L \ll r_A$:

$$G_T = \frac{i_A}{v_K} = -g_m. \tag{5.94}$$

(b) *Voltage Gain*

$$A_V = \frac{v_A}{v_K} = -\left(\frac{i_A}{v_K}\right) \times R_L = -G_T \times R_L. \tag{5.95}$$

(c) *Current Gain*

$$A_I = \frac{i_A}{i_K} = -1. \tag{5.96}$$

(d) *Transfer Resistance*

$$R_T = \frac{v_A}{i_K} = -\left(\frac{i_A}{i_K}\right) \times R_L = -A_I \times R_L. \tag{5.97}$$

(e) *Input Resistance*

(i) exact:

$$R_i = \frac{v_K}{i_K} = \frac{r_A + R_L}{\mu + 1}; \tag{5.98}$$

(ii) approximate, for $\mu \gg 1$:

$$R_i = \frac{v_K}{i_K} = \frac{1}{g_m} + \frac{R_L}{\mu}; \tag{5.99}$$

(iii) approximate, for $\mu \gg 1$, $R_L \ll r_A$:

$$R_i = \frac{v_K}{i_K} = \frac{1}{g_m}. \tag{5.100}$$

(f) *Output Resistance*

$$R_o = r_A + \mu R_S. \tag{5.101}$$

dependent on a large value of μ; if μ is 10 or more, the error is less than 10%. Finally, the third expression gives an order of magnitude, which is useful for design purposes and is based on the assumption that the load resistance is much less than the anode resistance. Table 5.10 gives expressions corresponding to the approximate equivalent circuit of Fig. 5.9.

Table 5.10 Common-Base Amplifier Stage

$$G_T = \frac{i_C}{v_E} = -\frac{\alpha_N}{(1 - \alpha_N)r_B + r_E} \approx -\frac{\alpha_N}{r_B/\beta_N + r_E}, \tag{5.102}$$

$$A_V = \frac{v_C}{v_E} = -\left(\frac{i_C}{v_E}\right) \times R_L = -G_T \times R_L, \tag{5.103}$$

$$A_I = \frac{i_C}{i_E} = -\alpha_N, \tag{5.104}$$

$$R_T = \frac{v_C}{i_E} = -\left(\frac{i_C}{i_E}\right) \times R_L = -A_I \times R_L, \tag{5.105}$$

$$R_i = \frac{v_E}{i_E} = (1 - \alpha_N)r_B + r_E \approx \frac{r_B}{\beta_N} + r_E. \tag{5.106}$$

5.7.2 Common-Collecting-Electrode

The *common-anode* and *common-collector* amplifier circuits are shown in outline in Fig. 5.18. These circuits are often called the *cathode follower*

Table 5.11 Common-Anode Amplifier Stage (Cathode Follower)

(a) *Voltage Gain*

(i) exact:

$$A_V = \frac{v_K}{v_G} = \frac{\mu R_L}{r_A + (\mu + 1)R_L}; \tag{5.107}$$

(ii) approximate, for $\mu \gg 1$:

$$A_V = \frac{v_K}{v_G} = \frac{g_m R_L}{1 + g_m R_L}; \tag{5.108}$$

(iii) approximate, for $\mu \gg 1$, $g_m R_L \gg 1$:

$$A_V = \frac{v_K}{v_G} = 1. \tag{5.109}$$

(b) *Input Resistance*

(i) exact:

$$R_i = \frac{v_G}{i_i} = R_G\left[\frac{r_A + (\mu + 1)R_L}{r_A + R_L}\right]; \tag{5.110}$$

(ii) approximate, for $\mu \gg 1$:

$$R_i = \frac{v_G}{i_i} = R_G\left(\frac{1 + g_m R_L}{1 + R_L/r_A}\right); \tag{5.111}$$

(iii) approximate, for $\mu \gg 1$, $g_m R_L \gg 1$, $R_L \ll r_A$:

$$R_i = \frac{v_G}{i_i} = R_G(g_m R_L). \tag{5.112}$$

(continued)

Table 5.11 (*continued*)

(c) *Output Resistance*

(i) exact:

$$R_o = r_A\left[\frac{R_S + R_G}{R_S + (\mu + 1)R_G}\right]; \tag{5.113}$$

(ii) approximate, for $\mu \gg 1$:

$$R_o = r_A\left(\frac{R_S + R_G}{R_S + \mu R_G}\right); \tag{5.114}$$

(iii) approximate, for $\mu \gg 1$, $R_G \gg R_S$:

$$R_o = \frac{1}{g_m}. \tag{5.115}$$

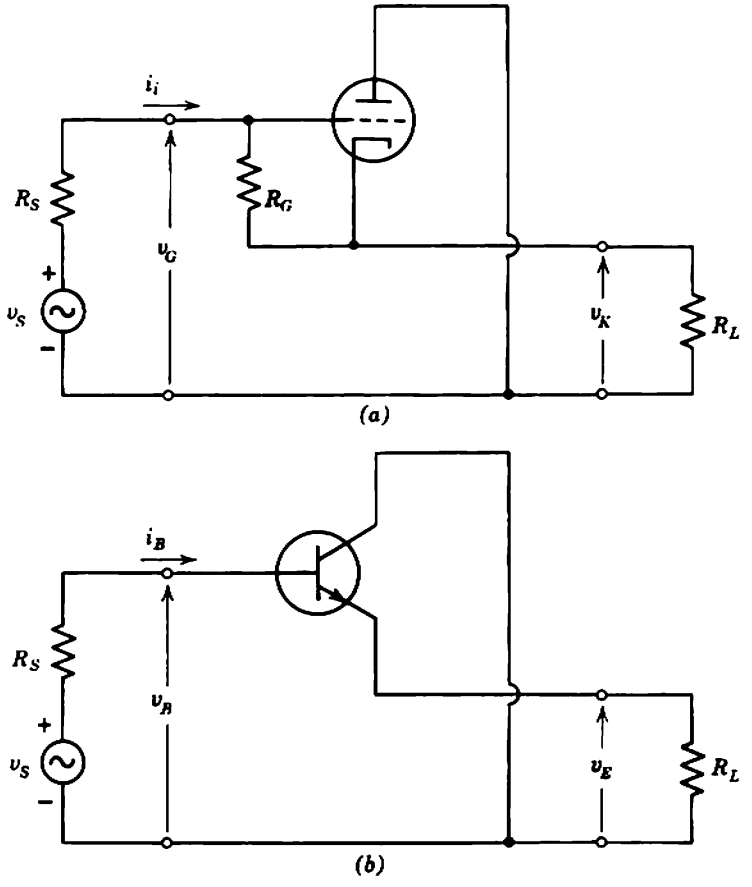


Fig. 5.18 Elemental circuit diagrams for common-collecting-electrode amplifier stages: (*a*) common anode (cathode follower); (*b*) common collector (emitter follower).

and *emitter follower*, respectively. A pentode may be used as a cathode follower but gives very little improvement over a triode. These circuits are characterized by a high input resistance and a low output resistance. Both have a voltage gain slightly less than unity. Table 5.11 lists the relevant transfer and driving-point functions of a cathode follower. Three forms are given; the first is exact, the second and third are approximate. Note that a resistor R_G is shown connected between the grid and cathode; this is required for biasing. Table 5.12 lists the parameters of an emitter follower.

Table 5.12 Common-Collector Amplifier Stage (Emitter Follower)

(a) *Voltage Gain*

(i) exact:

$$A_V = \frac{v_E}{v_B} = \frac{R_L}{(1 - \alpha)r_B + r_E + R_L} \approx \frac{R_L}{r_B/\beta_N + r_E + R_L}; \quad (5.116)$$

(ii) approximate, for $R_L \gg (r_B/\beta_N + r_E)$:

$$A_V = \frac{v_E}{v_B} = 1. \quad (5.117)$$

(b) *Input Resistance*

(i) exact:

$$R_i = \frac{v_B}{i_B} = r_B + \frac{r_E + R_L}{1 - \alpha_N} \approx r_B + \beta_N(r_E + R_L); \quad (5.118)$$

(ii) approximate, for $R_L \gg (r_B/\beta_N + r_E)$:

$$R_i = \frac{v_B}{i_B} = \frac{R_L}{1 - \alpha_N} \approx \beta_N R_L. \quad (5.119)$$

(c) *Output Resistance*

(i) exact:

$$R_o = (1 - \alpha_N)(R_S + r_B) + r_E \approx \frac{R_S + r_B}{\beta_N} + r_E; \quad (5.120)$$

(ii) approximate, for $(R_S + r_B)/\beta_N \ll r_E$:

$$R_o = r_E. \quad (5.121)$$

Chapter 6

Biasing Circuits

The analysis of transistor circuits in Sections 6.1–6.3 is in terms of *n-p-n* devices. This has the merit that all polarities are the same as in the analogous vacuum-tube circuit. The analyses apply without any change whatsoever to *p-n-p* devices provided all quantities are interpreted as magnitudes.

Chapters 2, 3, and 4 discuss the characteristics of charge-controlled amplifying devices. It is shown that the small-signal parameters of these devices depend critically on the quiescent dc operating current and, to a small extent, on the particular combination of electrode voltages used to produce a specified current. The mutual conductance varies as $I_K^{1/2}$ in a vacuum tube, as I_E in a bipolar transistor, and as $I_S^{1/2}$ in a pinched-off field-effect transistor. Chapter 5 contains the circuit theory of amplification at mid-band frequencies. Naturally, the gain of an amplifier depends on the small-signal parameters of its active devices; in many cases it is proportional to g_m. Therefore, the quiescent points for the devices must be fixed with an accuracy that depends on the gain stability required and on the detail of the circuit.

This chapter considers linear circuit arrangements for biasing vacuum tubes and bipolar transistors at any desired quiescent point. Biasing circuits for field-effect transistors follow vacuum-tube principles almost exactly, and are mentioned only briefly. It is assumed that one or more dc supply voltages are available, and that the stability of these supplies is at least as great as the required stability of the quiescent point. Linear biasing circuits are adequate for all small-signal amplifiers where dc power wastage is not an important consideration, and for many large-signal amplifiers. Nonlinear circuits are preferable when dc power wastage is a consideration. They are discussed in Chapter 16 in the context of large-signal amplifiers. In these nonlinear circuits temperature- or voltage-

sensitive elements such as thermistors are used to achieve some measure of compensation.

Most linear biasing circuits are in fact feedback circuits. While these circuits can be analyzed and designed without any reference to feedback, a feedback-oriented approach is considered more sophisticated and accordingly, the word "feedback" appears in many of their names. Application of feedback principles to the circuit arrangements in this chapter is left as an exercise for the reader.

The primary aim of a biasing circuit is to hold the emitting-electrode current I_3 constant despite the production tolerance and effect of aging on the characteristics of the device. A secondary aim is to hold constant the potential difference V_2 between the emitting electrode and the collecting electrode (and any auxiliary electrodes such as the screen of a pentode). At constant I_3 and V_2, the voltage V_1 that must be applied to the control electrode and the current I_1 that flows depend on the characteristics of the device; V_1 and I_1 are thus dependent variables and are subject to production tolerances and aging. A biasing circuit must accommodate reasonable spreads in the values of V_1 and I_1 at the desired I_3 and V_2.

Figure 6.1 shows the outline of one form of biasing system; no provision is made for applying signals to the device. The current-sensing circuit applies to the control electrode a voltage which tends to correct any difference between the actual and desired emitting-electrode currents. If I_3 is known, then I_2 is known also, the relations between the emitting- and collecting-electrode currents for practical devices being

$$I_A = I_K \text{ for a triode,}$$

$$I_A = I_K/(1 + k) \text{ for a pentode,}$$

$$I_C \approx \alpha_N I_E \text{ for a transistor,}$$

$$I_D = I_S \text{ for an f.e.t.,}$$

and the parameters k and α_N are reproducible. Therefore, the collecting-electrode voltage is known and fixed, satisfying the second requirement of a biasing circuit:

$$V_2 = V_{22} - I_2 R_2.$$

This system forms the basis of the collecting-electrode-feedback circuits discussed in Section 6.3.

Figure 6.2 shows the outline of another biasing system from which the emitting-electrode-feedback circuits of Section 6.1 are derived. Normally, $|V_2|$ is much greater than $|V_1|$ for a charge-controlled device. It is therefore satisfactory for a biasing system to hold I_3 and $(V_2 - V_1)$, rather than I_3 and V_2, constant; the uncertainty in V_1 is an insignificant

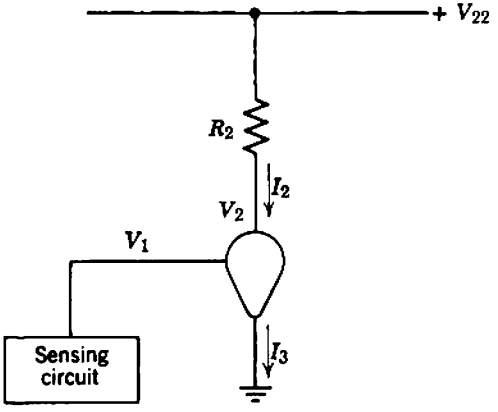


Fig. 6.1 Biasing system of the first kind.

fraction of $(V_2 - V_1)$. In this biasing system, the dc current source sets I_3, and I_2 and V_2 $(\approx V_2 - V_1)$ follow as in the preceding circuit.

The majority of amplifier stages use some form of the emitting-electrode-feedback biasing system (Fig. 6.2). This circuit and its modified forms are therefore discussed extensively in Sections 6.1 and 6.2; the physical mode of operation is emphasized, accurate and approximate expressions for the currents and voltages are derived, and a number of design rules are given. Section 6.3 discusses the collecting-electrode-feedback biasing system; less detail is given, the intention being that the reader should apply the physical principles of Sections 6.1 and 6.2 to this circuit. Sections 6.5 and 6.6 are catalogs of more complicated biasing systems that may prove useful under special circumstances. Little more is given than a circuit diagram and an approximate expression for the emitting-electrode current.

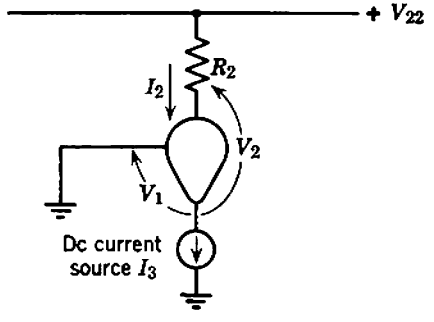


Fig. 6.2 Biasing system of the second kind.

It must be emphasized that biasing-circuit design is largely a matter of common sense based on physical principles. There is great scope for originality in circuit designs.

6.1 EMITTING-ELECTRODE FEEDBACK

Figure 6.3 shows practical realizations of the *emitting-electrode-feedback biasing circuit* for the general device, the vacuum triode, and the bipolar transistor. The dc current source in the emitting-electrode circuit is approximated by a resistor connected to a dc supply voltage.

Provision is made in Fig. 6.3 for applying a signal to the device. First, a capacitor is connected from the emitting electrode to ground. It is assumed in the small-signal analysis of Chapter 5 that no signal voltage exists at the emitting electrode, despite the presence of signal currents in

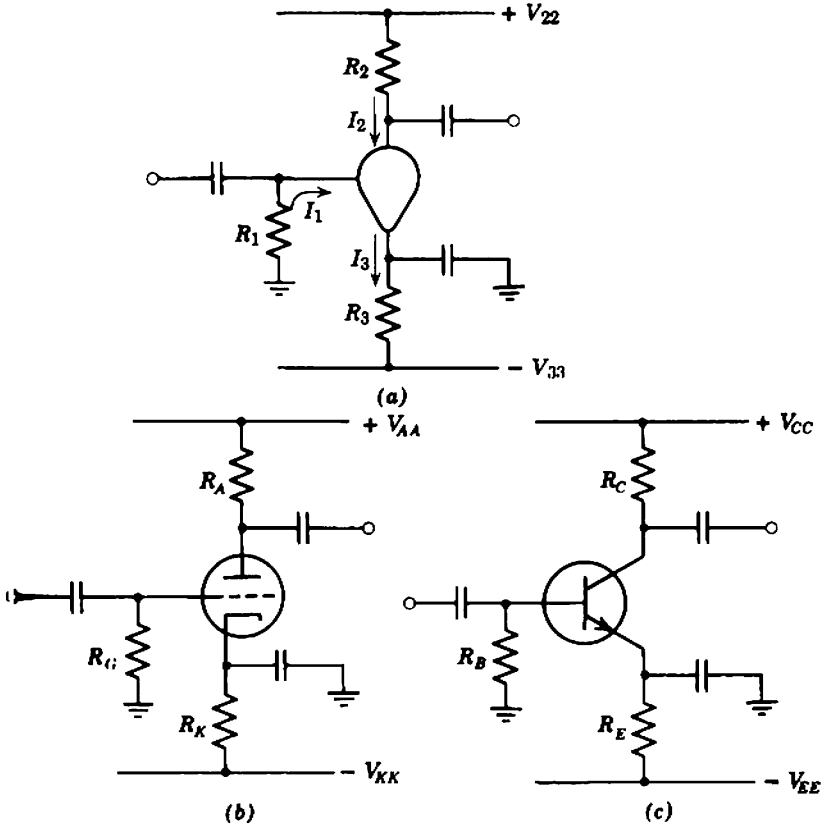


Fig. 6.3 Emitting-electrode-feedback biasing circuit: (*a*) general charge-controlled device; (*b*) vacuum triode; (*c*) *n-p-n* transistor.

the electrode. The emitting electrode is taken as the datum for signals. Therefore, the emitting electrode must be connected to ground through a low-impedance signal path; the *bypass capacitor* in Fig. 6.3 provides this path. Bypassing is discussed more fully in Section 6.4, and again in Section 7.4 in the context of the low-frequency response of an amplifier. Second, the control electrode is connected to ground through a resistor R_1, R_G, or R_B as appropriate—rather than directly to ground as in Fig. 6.2. An ac signal voltage can therefore be coupled onto the control electrode via a *coupling capacitor*. However, the control-electrode dc current I_1 produces a voltage drop across R_1, and because I_1 is subject to production tolerance and aging there is uncertainty in the control-electrode voltage. Thus the value of R_1 must be a compromise:

(i) R_1 should be large, because it shunts the input resistance of the device and increases the loading on the ac signal source;
(ii) R_1 should be small so as to minimize the uncertainty in control-electrode dc voltage, and therefore the dc voltages at all other points in the circuit.

The form of the circuits in Fig. 6.3 is sometimes inconvenient, as two dc supplies of opposite polarity are required. Figure 6.4 shows two modified forms. In the first form two supplies of the same polarity are required. In the second form the control-electrode supply is replaced by a potential divider across the collecting-electrode supply. The relations between Figs. 6.3 and 6.4 are trivial and are listed in Table 6.1.

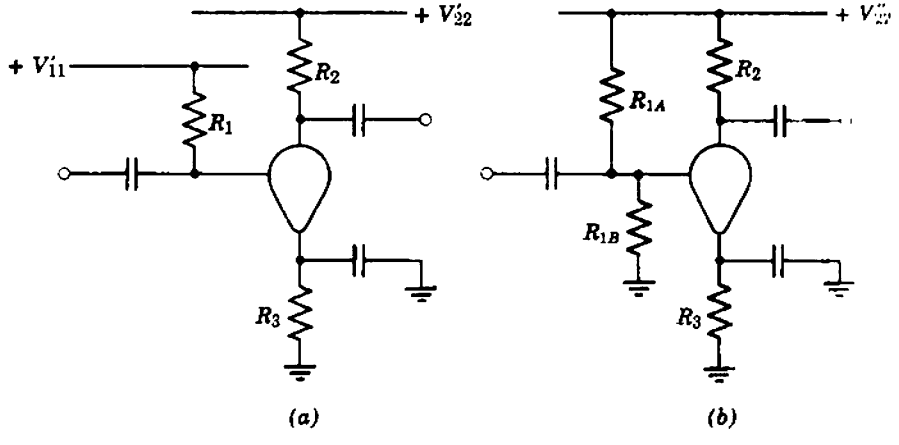


Fig. 6.4 Alternative arrangements of the emitting-electrode-feedback biasing circuit: (*a*) two supplies of the same polarity; (*b*) single supply.

Table 6.1 Relations Between Figs. 6.3 and 6.4

Parameter in Fig. 6.3*a*	Parameter in Fig. 6.4*a*	Parameter in Fig. 6.4*b*
V_{33}	V'_{11}	$V''_{22}\left(\frac{R_{1B}}{R_{1A} + R_{1B}}\right)$
$V_{22} + V_{33}$	V'_{22}	V''_{22}
R_1	R_1	$R_{1A} \parallel R_{1B}$

The emitting electrode in Fig. 6.3 is near ground potential, differing from it only by the sum of the voltage drop across R_1 and the voltage drop between the control and emitting electrodes. Ideally these voltage drops are much less than V_{33}, and the emitting-electrode current is

$$I_3 = \frac{V_{33}}{R_3}. \tag{6.1}$$

The current in a practical biasing system is not given exactly by Eq. 6.1, because the voltage drops are a significant fraction of V_{33}. Being dependent on the device characteristics, these voltage drops introduce uncertainties into the emitting-electrode current, but any finite degree of biasing stability can be achieved by making V_{33} sufficiently large.

Of the two practical circuits in Fig. 6.3 the emitter-feedback biasing system has the more general analysis. This analysis is given in Section 6.1.1 and the results for the more restricted cathode-feedback system are deduced in Section 6.1.2.

6.1.1 • Emitter-Feedback Biasing

Figure 6.5 shows an *emitter-feedback biasing circuit*, arranged in the form of Fig. 6.3 with all currents and voltages marked. Provided the collector is positive with respect to the emitter so that minority carriers cannot be injected from the collector into the base and the transistor is operating normally, the following relations hold between the electrode currents:

$$I_B = (1 - \alpha_N)I_E - I_{CO},$$

$$I_C = \alpha_N I_E + I_{CO}.$$

The parameters α_N, V_{BE}, and I_{CO} are, respectively, the (large signal) emitter-to-collector current gain (Section 4.3.1.2.4), the base-to-emitter voltage drop, and the collector saturation current. All are functions of

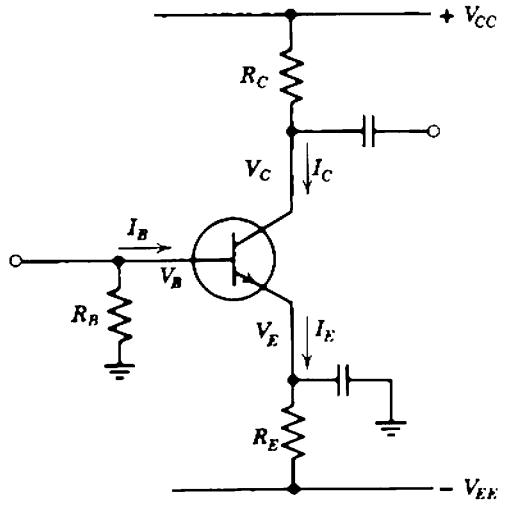


Fig. 6.5 Basic form of emitter-feedback biasing with two supplies of opposite polarity.

the transistor type, its temperature, and the quiescent point; their production tolerances and variations with age cause uncertainty in the quiescent point. The emitter current is

$$I_E = \left(\frac{V_{EE} - V_{BE} + I_{CO}R_B}{R_E}\right)\left[\frac{1}{1 + (1 - \alpha_N)R_B/R_E}\right] \tag{6.2}$$

and because

$$\beta_N = \frac{\alpha_N}{1 - \alpha_N} \approx \frac{1}{1 - \alpha_N}, \tag{6.3}$$

it follows that

$$I_E \approx \left(\frac{V_{EE} - V_{BE} + I_{CO}R_B}{R_E}\right)\left(\frac{1}{1 + R_B/\beta_N R_E}\right). \tag{6.4}$$

A satisfactory biasing system must define the quiescent currents and voltages independent of the transistor parameters. Typically, a tolerance from ± 10 to $\pm 25\%$ is required for I_E in an amplifier circuit. Therefore the correcting terms in Eq. 6.4 due to β_N, V_{BE}, and I_{CO} should be small. Equation 6.4 reduces to

$$I_E \approx \frac{V_{EE}}{R_E} \tag{6.5}$$

as suggested by Eq. 6.1. The existence of the finite parameters β_N, V_{BE}, and I_{CO} causes a perturbation about this ideal quiescent point.

The dependence of I_E on V_{BE} and I_{CO} can be made as small as desired by choosing V_{EE} sufficiently large in comparison with the expected "average"

values of V_{BE} and $I_{CO}R_B$. If I_E is required to be stable to within $\pm 10\%$ against variation of V_{BE} alone, the expected value of $(V_{EE} - V_{BE} + I_{CO}R_B)$ must be designed as ten times the maximum peak variation of V_{BE}. For germanium transistors at normal temperatures, V_{BE} lies in the range 0.1 to 0.4 V; for silicon transistors the range is 0.6 to 0.9 V. Therefore an emitter-feedback biasing system using a germanium transistor will have $\pm 10\%$ tolerance on I_E due to variation of V_{BE} alone if

(i) V_{BE} is assumed to be 0.25 V so that the peak variation is ± 0.15 V,
(ii) V_{EE} is designed to be greater than 1.75 V so that the expected value of $(V_{EE} - V_{BE} + I_{CO}R_B) \geq 1.50$ V.

Table 6.2 is constructed from a series of similar calculations.

Table 6.2 Value of V_{EE} for Various Degrees of Stability Against V_{BE}

Required Tolerance on I_E	Germanium Transistors	Silicon Transistors
$\pm 5\%$	3.25 V	3.75 V
$\pm 10\%$	1.50 V	2.00 V
$\pm 25\%$	0.85 V	1.35 V

Because the variation of I_E with I_{CO} depends on the value of R_B, no general rule can be formulated corresponding to Table 6.2 but the calculation of V_{EE} is straightforward in specific cases. A rule of thumb is to choose R_B such that the maximum possible value of $I_{CO}R_B$ is about 0.4 V and to assume that its average value is 0.2 V. If V_{EE} is designed as about 2 V (as required for 10% stability against V_{BE}), the tolerance due to I_{CO} is also $\pm 10\%$.

The variation of I_E with β_N can be reduced to any desired extent by designing the ratio of R_B to R_E sufficiently small, for the second term in Eq. 6.4 then approaches unity. Figure 6.6 is a normalized plot of this term, in which β_N is normalized to its average value β_{NA}; the running parameter is $R_B/\beta_{NA}R_E$. If β_N has the tolerance assumed throughout this book (from 70 to 200% of its nominal value) and I_E is required to have $\pm 10\%$ tolerance, then $R_B/\beta_{NA}R_E$ must be designed as about 0.2. Supposing further that the average value β_{NA} is 50, the ratio R_B/R_E must not exceed 10.

In the most precise work it may be worthwhile to make use of the known temperature coefficients of V_{BE}, I_{CO}, and β_N. These temperature

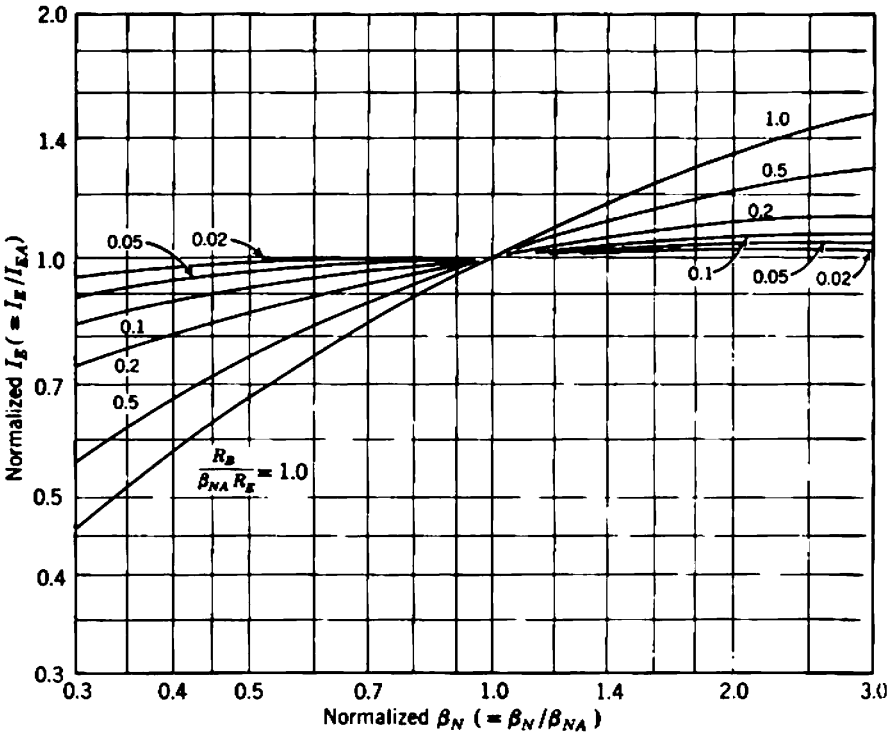


Fig. 6.6 Design nomogram for emitter-feedback biasing circuits.

coefficients are discussed in Section 4.6.3. Near room temperature the values are:

$$\frac{dV_{BE}}{dT} = -2.5 \text{ mV/°C}; \tag{6.6}$$

$$I_{CO} \text{ doubles for every } 8°\text{C rise in temperature}; \tag{6.7}$$

$$\beta_N \propto T^2. \tag{6.8}$$

All three effects tend to increase the emitter current as the temperature rises. There is no mutual compensation.

6.1.1.1 Numerical Example

Consider biasing a germanium *p-n-p* audio-frequency transistor to about 0.3 mA emitter current with maximum possible variation ±15%. There is a single supply of −9 V, the collector supply resistor R_C is to be 4.7 kΩ, and the biasing resistance R_B shunting the input is to be not less than 10 kΩ. (A *p-n-p* transistor is specified rather than an *n-p-n* to illustrate that the same equations can be applied to either type, provided all currents

and voltages are interpreted as magnitudes.) The parameters of the transistor over the temperature range 0 to 50°C are

$$0.1 \leq V_{BE} \leq 0.4 \text{ V},$$

$$0 \leq I_{CO} \leq 50\ \mu\text{A},$$

$$\beta_{NA} = 40,$$

$$25 \leq \beta_N \leq 85.$$

Preferred value resistors are to be used throughout. (This example corresponds to the amplifier discussed in Section 5.3.3.)

Because only one supply is available, the circuit must be of the form shown in Fig. 6.4*b*. Figure 6.7 shows one satisfactory combination of resistor values. The equivalent base supply potential V_{BB} is chosen as 4.5 V to mask the variation of V_{BE} and I_{CO}. The values of R_B and R_E are 11 kΩ and 15 kΩ, respectively, so that $(R_B/\beta_{NA}R_E) = 0.018$ and the dependence on β_N is small. Substitution into Eq. 6.3 shows that I_E is 0.265 mA with all parameters at the low-current limit and 0.329 mA at the high-current limit. The variation is ±11% about the mean.

6.1.1.2 Fixed Biasing

Fixed biasing is a degenerate form of emitter-feedback biasing in which R_E is reduced to zero. Almost invariably the circuit takes the practical

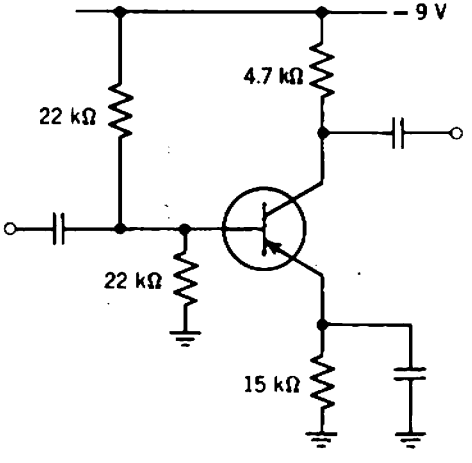


Fig. 6.7 Example of an emitter-feedback biasing circuit for a *p-n-p* germanium transistor. The circuit is in the modified form with a single supply.

$$I_E \approx 0.3 \text{ mA}, \qquad V_{CE} \approx 3 \text{ V}.$$

form of Fig. 6.4 rather than Fig. 6.3; that is, a base supply is used rather than an emitter supply. The emitter current is

$$I_E = (\beta_N + 1)\left(\frac{V_{BB} - V_{BE}}{R_B} + I_{CO}\right). \tag{6.9}$$

Fixed biasing circuits fall into two classes, depending on whether the base supply network approximates to a current or voltage source, that is, whether R_B is large or small compared with the input resistance of the transistor. In the current-feed circuit the base supply voltage V_{BB} is made large compared with V_{BE}—usually the collector supply V_{CC} is used—and R_B is necessarily made large to give I_E a typical value; the base is virtually fed from a current source. Figure 6.8 shows a circuit diagram. The emitter current depends critically on β_N and I_{CO}, but is almost independent of V_{BE} because V_{BB} is large in comparison. The circuit is so unsatisfactory that it should never be used; it is mentioned in this book only as an example of a circuit to be avoided.

The voltage-feed fixed biasing circuit is also unsatisfactory in its simple form, but it can be made satisfactory by using a nonlinear element as shown in Fig. 6.9. The circuit is made independent of the variation of V_{BE} with temperature, not by increasing V_{BB}, but by using a temperature-sensitive element to give V_{BB} a coefficient equal to that of V_{BE} (about -2.5 mV/°C). Although not obvious from Eq. 6.9, the emitter current can be made independent of both β_N and I_{CO} by reducing R_B and V_{BB}; ideally, R_B should be zero and V_{BB} should be equal to V_{BE}. This circuit is useful in large-signal amplifiers and is discussed more fully in Section 16.4.2.

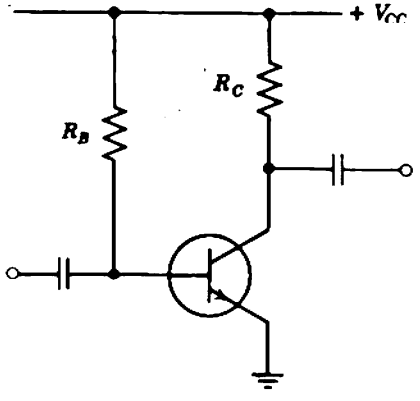


Fig. 6.8 Fixed biasing with current feed. This circuit should never be used.

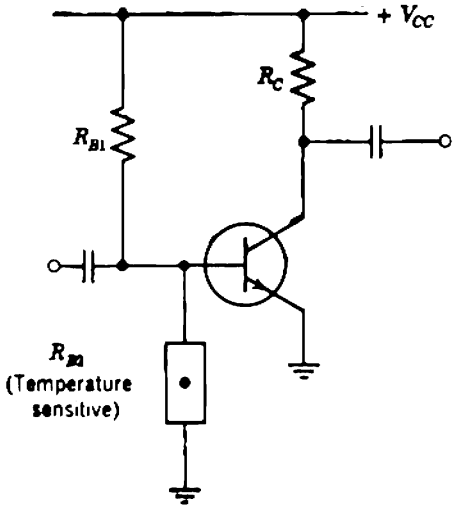


Fig. 6.9 Fixed biasing with voltage feed.

6.1.2 Cathode-Feedback Biasing

Vacuum tubes have no useful parameters corresponding to β_N and I_{CO} because recombination and thermal generation are negligible. However, residual gas in a tube produces a minute grid current of the order of 10^{-8} to 10^{-10} A except at grid voltages less negative than about -1 V. Special tubes such as electrometers have 10^{-14} A grid current or less. Figure 6.10 shows the circuit diagram for a *cathode-feedback biasing circuit* and by analogy with Eq. 6.4 the cathode current is

$$I_K = \frac{V_{KK} - V_{GK} - I_G R_G}{R_K}. \tag{6.10}$$

Notice that $-V_{GK}$ is a positive quantity because V_{GK} is the negative voltage of the grid relative to the cathode. Under all but the most unusual circumstances, $I_G R_G$ is negligible and Eq. 6.10 reduces to

$$I_K = \frac{V_{KK} - V_{GK}}{R_K}. \tag{6.11}$$

The condition for biasing stability is that V_{KK} should be very much greater than $|V_{GK}|$, when

$$I_K \approx \frac{V_{KK}}{R_K} \tag{6.12}$$

and I_K is independent of all the tube parameters. An example of cathode-feedback biasing appears in Section 6.2.1.2.

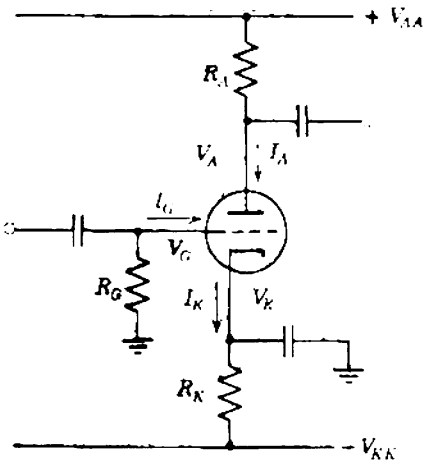


Fig. 6.10 Basic form of cathode-feedback biasing with two supplies of opposite polarity.

6.1.2.1 Fixed Biasing

Fixed biasing for a tube is the degenerate form of cathode-feedback biasing in which R_K is reduced to zero. Like fixed biasing for transistors, the circuit usually takes the practical form of Fig. 6.4 rather than that of Fig. 6.3; that is, the cathode supply is replaced by a grid supply. An important difference between the tube and transistor circuits is that V_{GK} and V_{GG} are negative for a tube, whereas V_{BE} and V_{BB} are positive for an *n-p-n* transistor.

Provided $I_G R_G$ is negligible, the grid voltage is fixed:

$$V_{GK} = V_{GG}.$$

The cathode current is therefore entirely determined by the characteristics of the tube, and the biasing stability is poor. However, fixed biasing of a tube is usable, whereas fixed biasing of a transistor is not (unless a nonlinear element is used), because tube parameters are virtually independent of ambient temperature. Further discussion of fixed biasing appears in Section 6.2.

6.1.2.2 Self Biasing

A special and important case of cathode-feedback biasing results from setting V_{KK} equal to zero. This arrangement is known as cathode biasing, or preferably as *self biasing* (to distinguish it from cathode-feedback biasing). Most vacuum-tube amplifiers use self biasing.

The operation of the circuit depends on the fact that V_{GK} is negative. Equation 6.10 therefore shows that I_K is positive even if V_{KK} is zero. No corresponding circuit exists for transistors because V_{BE} is positive (for an *n-p-n* transistor).

The stability of a self biasing circuit increases as $|V_{GK}|$ and the mutual conductance g_m increase. Typically, the tolerance on the current is $\pm 20\%$, which is roughly a factor of two improvement over fixed biasing. The design of self biasing circuits is discussed in Section 6.2.

6.2 GRAPHICAL METHODS

Section 6.1 discusses one biasing system that can define the quiescent point of a device with any degree of precision, provided the maximum spread in the device parameters is known. For a transistor, the uncertainty in quiescent point results principally from the spreads in β_N and I_{CO}; V_{BE} is relatively predictable. Further, all three parameters are nearly independent of V_{CE} and, therefore, of the design of the collector circuit. For a vacuum tube, no parameters correspond to β_N and I_{CO}, so that the uncertainty in quiescent point results entirely from the spread in V_{GK}. The value of V_{GK} necessary for a particular cathode current depends on the tube type, its age, and the anode (or screen) voltage.

A cathode-feedback biasing system can accommodate any spread in the value of V_{GK} if the cathode supply potential V_{KK} is sufficiently large. It is possible to keep the quiescent current inside specified limits even if the tube is replaced by another of totally different characteristics. However, a cathode-feedback biasing system requires a large value of V_{KK} if it is to accommodate even the variation of V_{GK} that occurs for a given tube over a reasonable range of quiescent current; considerable dc power is wasted in the cathode resistor R_K. If V_{KK} is reduced in an effort to reduce this power wastage, V_{GK} in Eq. 6.11 becomes a significant correction and the characteristics of the tube must be taken into account. In principle, the problem can be treated analytically, but because of the nonlinear nature of vacuum-tube characteristics a graphical approach is simpler and more rewarding.

There are two common graphical methods for presenting vacuum-tube data. The *anode characteristics* are plots of anode current versus anode voltage with grid voltage as parameter, while the *mutual characteristics* are plots of anode current versus grid voltage with anode and/or screen voltage as parameter. It must be emphasized that the characteristic curves supplied by manufacturers are averages, and that the characteristics of individual tubes may differ from these average curves by margins as large

as ±50%. Consequently, calculations for biasing circuits (like any other calculations based on average parameters) are subject to errors due to production tolerances and aging.

6.2.1 Anode Characteristics

Figure 6.11 compares the general shapes of the anode characteristics for triodes and pentodes. Triodes have a finite anode resistance of the order of thousands or tens of thousands of ohms; that is, the anode current depends appreciably on the anode voltage. Therefore, each anode curve (which is a plot of I_A versus V_A at constant V_G) has an appreciable gradient. In comparison, the anode resistance of a pentode is very large; I_A is almost independent of V_A, and the anode curves approximate to horizontal straight

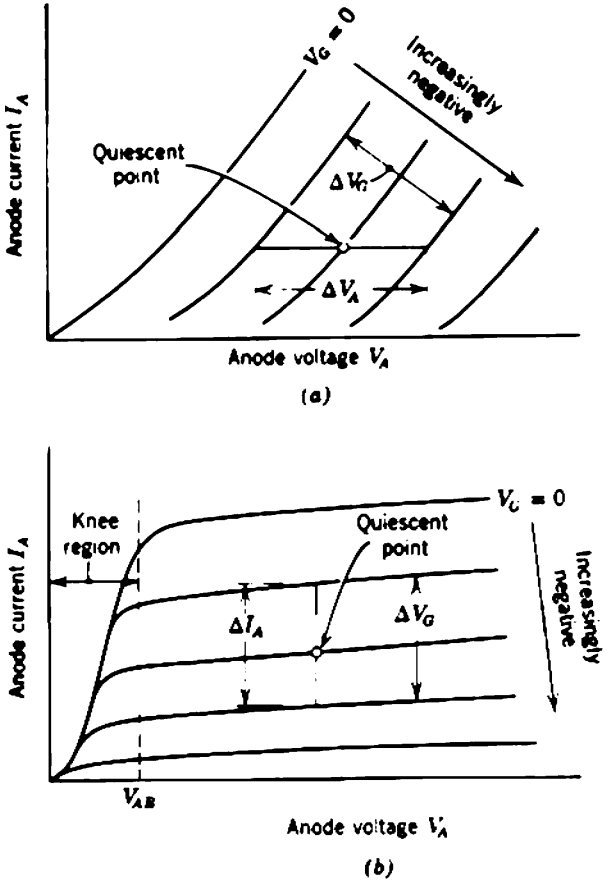


Fig. 6.11 Anode characteristics of typical tubes: (*a*) triode, showing the construction for finding μ; (*b*) pentode, showing the construction for finding g_m.

lines. The anode resistance of a pentode falls at low voltages and the curves converge toward the origin. Their shape in this region suggests the name *knee*, and the region of low anode resistance is said to be *below the knee*. This knee in the anode characteristics is explained in Section 3.1.3.

One set of anode characteristics characterizes a triode completely but does not characterize a pentode because the screen voltage can be used as a third independent variable. Tube manufacturers supply average anode characteristics for a pentode at one or perhaps two screen voltages. Curves for any other screen voltage can be constructed by the voltage conversion factor method of Section 3.4.2.1.

The values of the small-signal parameters μ, g_m, and r_A at any quiescent point can be deduced from the anode curves; for example, the anode resistance is defined as

$$r_A = \left(\frac{\partial V_A}{\partial I_A}\right)_{V_G} \approx \left(\frac{\Delta V_A}{\Delta I_A}\right)_{V_G}, \tag{6.13}$$

and is the reciprocal of the gradient of the curve at any point. The mutual conductance g_m can be calculated from the vertical spacing of the curves as in Fig. 6.11*b*:

$$g_m = \left(\frac{\partial I_A}{\partial V_G}\right)_{V_A} \approx \left(\frac{\Delta I_A}{\Delta V_G}\right)_{V_A}. \tag{6.14}$$

Similarly, μ can be calculated from the horizontal spacing as in Fig. 6.11*a*:

$$\mu = -\left(\frac{\partial V_A}{\partial V_G}\right)_{I_A} \approx -\left(\frac{\Delta V_A}{\Delta V_G}\right)_{I_A}. \tag{6.15}$$

The fact that the anode curves are not parallel straight lines with equal spacing indicates that the small-signal parameters change with cathode current and, to a small extent, with anode voltage. Consequently, the anode characteristics are a help in designing small-signal amplifiers; they provide a ready means for determining the average (or nominal) small-signal parameters at a chosen quiescent point or, conversely, the quiescent point necessary to give desired small-signal parameters.

The actual characteristics can be approximated by a set of parallel, equally-spaced straight lines (Fig. 6.12). These *piecewise-linear characteristics* are based on the assumption of a linear relation between the terminal currents and voltages, with the constant *large-signal parameters* $\bar{\mu}$, $\bar{g}_m$, and $\bar{r}_A$ as coefficients. For a triode

$$I_A = \bar{g}_m\left(V_G + \frac{V_A}{\bar{\mu}}\right) = \bar{g}_m V_G + \frac{V_A}{\bar{r}_A}. \tag{6.16}$$

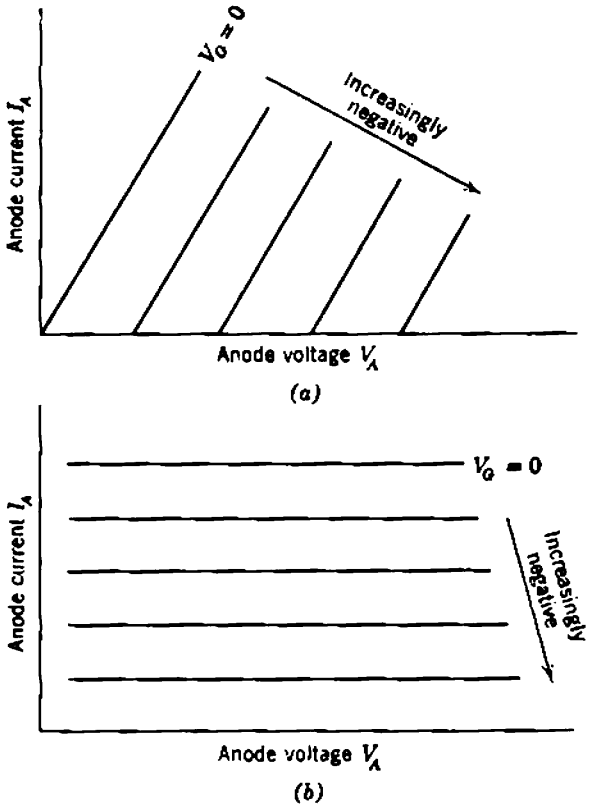


Fig. 6.12 Piecewise-linear representation of anode characteristics: (*a*) triode; (*b*) pentode.

whereas for a pentode

$$I_A = \overline{g_m}\left(V_G + \frac{V_S}{\bar{\mu}_S}\right). \tag{6.17}$$

Expressed another way, the piecewise-linear characteristics are constructed on the assumption that the small-signal parameters μ, g_m, and r_A are constant, independent of the quiescent point, and equal to the large-signal parameters $\bar{\mu}$, $\overline{g_m}$, and $\overline{r_A}$. Piecewise-linear characteristics are often of sufficient accuracy for the design of biasing systems; in addition, they are useful in the design of switching circuits, although these are beyond the scope of this book.

6.2.1.1 Dc Load Line

The anode characteristics of a triode permit a graphical calculation of the quiescent point in a given biasing circuit, but the anode characteristics of a pentode do not. One of the independent variables in the anode

characteristics is V_A, and V_A for a pentode has almost no effect on I_A at a normal quiescent point.

Consider first the fixed biasing circuit shown in Fig. 6.13*a*. The anode voltage is

$$V_A = V_{AA} - I_A R_A$$

which can be arranged to express the anode current as

$$I_A = V_A\left(-\frac{1}{R_A}\right) + \left(\frac{V_{AA}}{R_A}\right). \tag{6.18}$$

The anode current of a triode, however, is a function of V_A and V_G and is represented by

$$I_A = I_A(V_A, V_G). \tag{6.19}$$

In principle, Eqs. 6.18 and 6.19 can be solved to find the quiescent values of V_A and I_A (hence I_K) for given values of V_{AA}, V_G, and R_A. The analytical solution is tedious because of the nonlinear nature of Eq. 6.19, but a graphical solution can be obtained easily from the anode characteristics.

On axes of I_A versus V_A Eq. 6.19 represents the anode characteristics of a tube. On the same axes Eq. 6.18 represents the straight line shown in Fig. 6.13*b*; this straight line is called the *dc load line*. The load line joins the points

V_{AA} on the V_A axis,

$\frac{V_{AA}}{R_A}$ on the I_A axis,

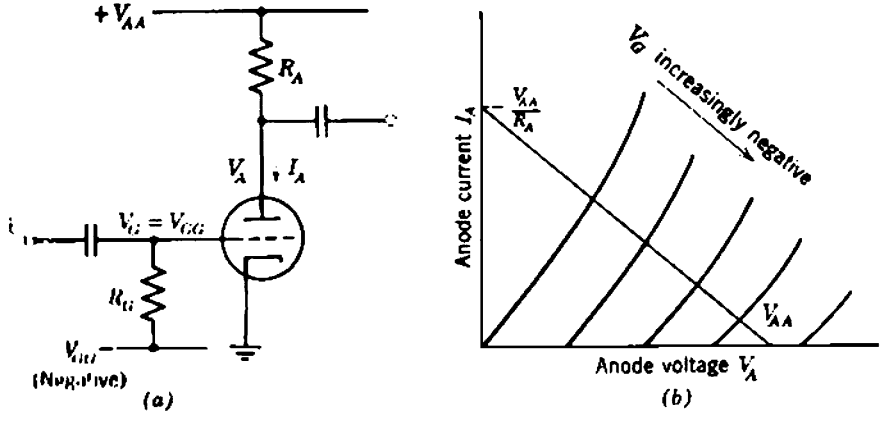


Fig. 6.13 Fixed biasing system and the dc load line.

and its gradient is

$$\frac{dI_A}{dV_A} = -\frac{1}{R_A}. \qquad (6.20)$$

The load line shows what the anode voltage would be if a certain anode current were flowing, but there is no implication that I_A and V_A can actually take up all sets of values corresponding to all points on the load line; clearly, the point

$$V_A = 0, \qquad I_A = \frac{V_{AA}}{R_A}$$

is absurd. The quiescent point is at the intersection of the load line with the anode characteristic corresponding to the given value of V_G, and the quiescent values of I_A and V_A can be read from the axes.

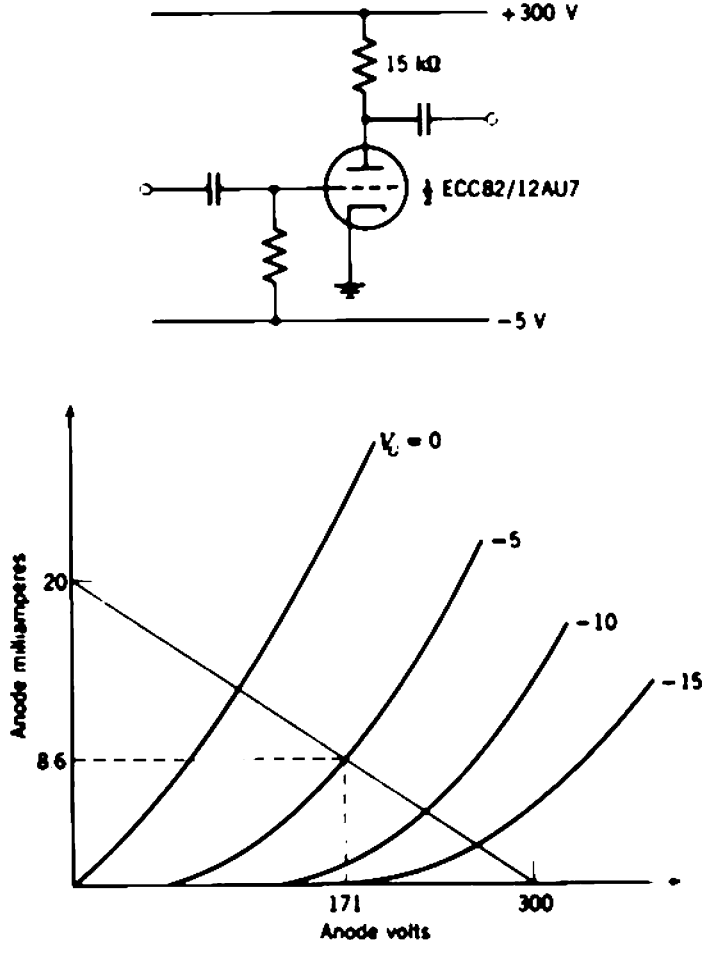


Fig. 6.14 Example of fixed biasing.

As an example, Fig. 6.14 shows in outline the anode characteristics of type ECC82/12AU7, with a load line corresponding to a 300-V supply and a 15-kΩ supply resistor. If the grid is maintained at −5 V with respect to the cathode, I_A and V_A are 8.6 mA and 171 V, respectively.

The load line provides a graphical means for determining the small-signal transfer functions G_T and A_V. The anode current and voltage corresponding to any grid voltage can be read from the load line and the ratios of changes can be calculated. In Fig. 6.14, if the grid is changed by −1 V from its quiescent potential −5 V, V_A changes by 13 V (from 171 to 184 V) so A_V is −13. For the same change in V_G, I_A changes from 8.6 to 7.7 mA so G_T is 0.9 mA/V.

Strictly speaking, the small-signal transfer functions are the ratios of infinitesimal changes, although the approximation is good if small finite changes are used:

$$G_T = \frac{dI_A}{dV_G} \approx \frac{\Delta I_A}{\Delta V_G}, \tag{6.21}$$

$$A_V = \frac{dV_A}{dV_G} \approx \frac{\Delta V_A}{\Delta V_G}. \tag{6.22}$$

A *transfer characteristic* is a plot of I_A or V_A versus V_G, and can be constructed from the load line. The *small-signal transfer function* at any quiescent point is the gradient of the transfer characteristic. Figure 6.15 shows the V_A/V_G transfer characteristic constructed from Fig. 6.14.

If a sine-wave signal is superimposed on the negative quiescent grid voltage, the anode voltage waveform can be determined from the transfer

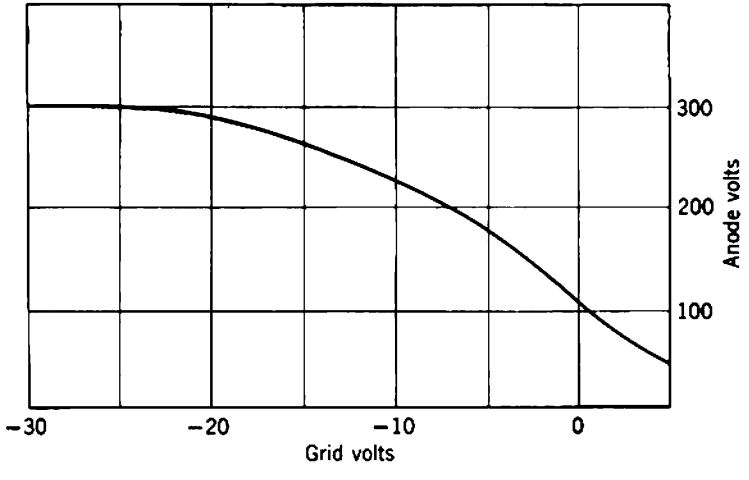


Fig. 6.15 V_A/V_G transfer characteristic constructed from Fig. 6.14.

characteristic, as in Fig. 6.16. As the amplitude of the input signal increases, so the output waveform departs further from a sine wave. Fundamentally, this *distortion* is due to the gain of a tube varying with its operating current. The changing gain is evidenced by the changing gradient of the transfer characteristic and is due to the variation of the small-signal parameters g_m and r_A with operating conditions. For example, g_m rises with the increasing cathode current on the positive half of the input waveform; the gain rises and the output waveform is "stretched" relative to the negative half of the input waveform. Chapter 9 contains quantitative information, but it is apparent that the anode characteristics and load line provide a qualitative measure of distortion. In addition, it is possible to choose a quiescent point at which distortion is minimum; such a point corresponds to a minimum of curvature in the transfer characteristic.

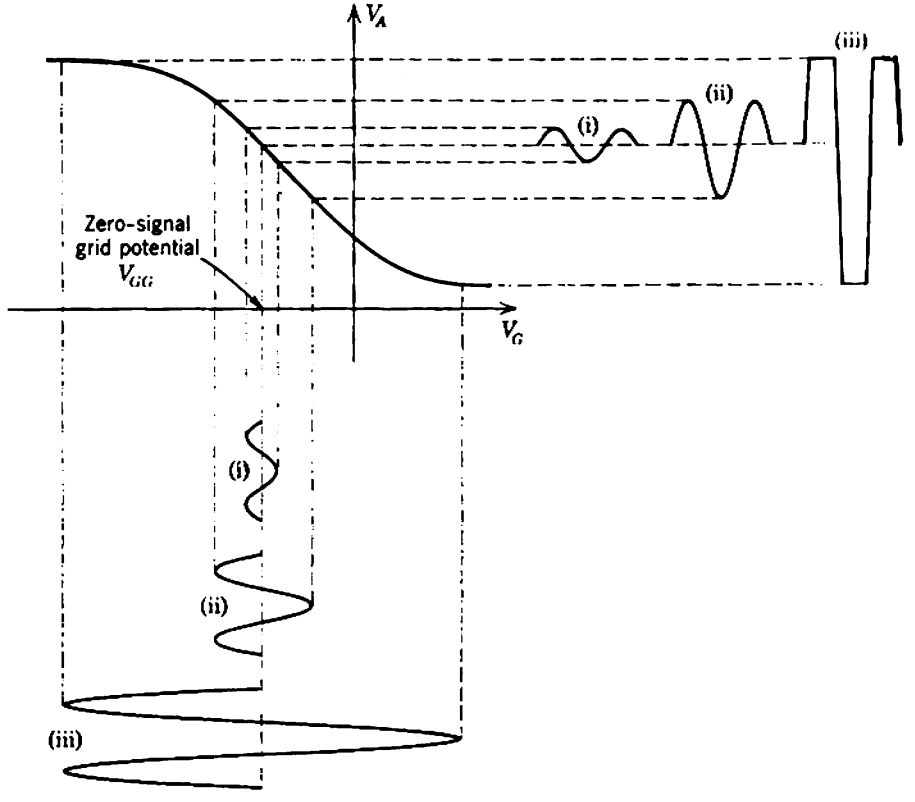


Fig. 6.16 V_A/V_G transfer characteristic, showing distortion

6.2.1.2 Ac Load Line

As discussed in Section 5.1.6, a vacuum tube is not usually operated with a simple anode supply resistor as its only load. Rather, some external load is coupled into the circuit and this external load makes use of the output signal. The two most common methods for coupling an external load to a tube are with a capacitor and a transformer (Fig. 6.17). The quiescent point can be found from the anode characteristics and the dc load line that corresponds to the dc resistance in the anode supply path—R_A in Fig. 6.17*a*, and the primary winding resistance R_1 of the transformer in Fig. 6.17*b*. Usually, the winding resistance of a transformer is negligibly small, and the dc load line becomes vertical. The load resistance at signal frequencies, however, is not R_A or R_1, so the dc load line cannot be used to calculate gain. With capacitance coupling, the load resistance is

$$R_L = \frac{R_A \times R_{ext}}{R_A + R_{ext}}, \tag{6.23}$$

provided that the signal frequency is high enough for the reactance of the coupling capacitor to be neglected. With transformer coupling

$$R_L = R_1 + N^2(R_2 + R_{ext}), \tag{6.24}$$

provided that the signal frequency is such that reactive effects in the transformer can be neglected. The *signal* or *ac load line* passes through

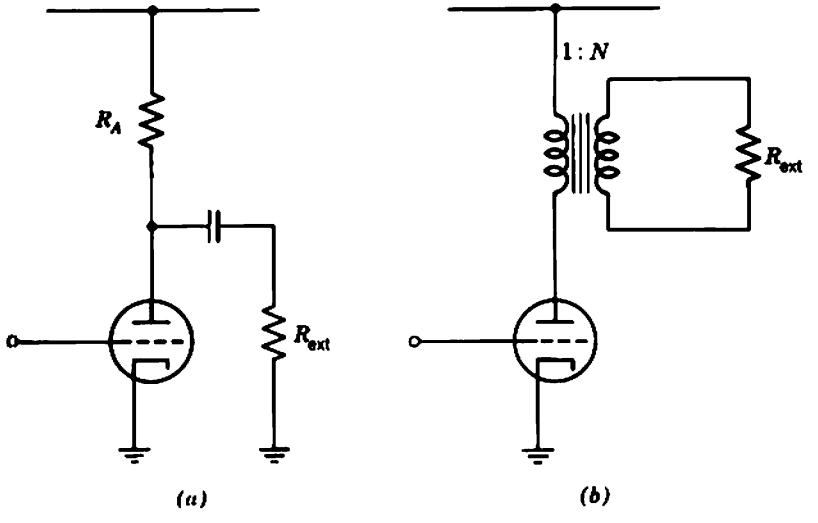


Fig. 6.17 Circuit arrangements for coupling an external load to an active device: (*a*) capacitance coupling; (*b*) transformer coupling.

the quiescent point found from the dc load line, and has a gradient corresponding to R_L:

$$\frac{dI_A}{dV_A} = -\frac{1}{R_L}. \tag{6.25}$$

For any signal-plus-quiescent voltage at the grid, I_A and V_A can be read from the intersection of the ac load line with the appropriate anode characteristic; gain and distortion can be investigated as before. Figure 6.18 compares the ac and dc load lines for capacitance and transformer coupling.

It is important to realize that Eqs. 6.23 and 6.24 apply only for signals of frequency such that all reactive elements in the coupling circuit can be neglected. At very low frequencies, the ac load line must be changed to take account of the reactance of the coupling capacitor or the primary

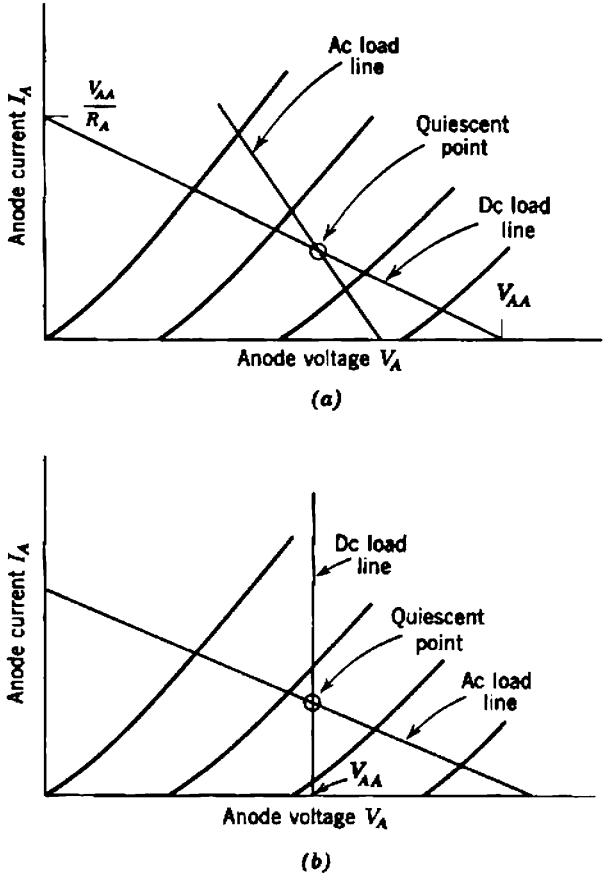


Fig. 6.18 Comparison of ac and dc load lines: (*a*) capacitance coupling; (*b*) transformer coupling.

inductance of the transformer. In the limit of zero frequency (i.e., dc), the ac load line degenerates to the dc load line. It is also important to realize that with capacitance coupling V_A cannot rise above the supply voltage V_{AA}. In contrast, with transformer coupling V_A can rise above V_{AA} at mid-band frequencies and is limited only by the intersection of the ac load line with V_A axis. The output voltage in excess of the supply voltage is produced by the back emf in the primary inductance of the transformer.

6.2.1.3 Cathode Bias Line

The anode characteristics of a triode can be used to find the quiescent point of a cathode-feedback biasing circuit such as Fig. 6.19*a*. Three simultaneous equations must be solved:

(i) *Anode characteristics*

$$I_A = I_A(V_{AK}, V_{GK}), \tag{6.26}$$

(ii) *Dc load line*

$$V_{AA} - V_{AK} = I_A R_A + I_K R_K, \tag{6.27}$$

(iii) *Cathode bias line*

$$V_{GK} = V_G - V_K = V_{GG} - I_K R_K. \tag{6.28}$$

Figure 6.19*b* shows the graphical solution. Notice that the dc load line corresponds to $(R_A + R_K)$, rather than R_A alone, because the dc voltage V_{AK} across the tube is V_{AA} less the drops across both R_A and R_K. Every point on the dc load line gives I_A at a particular value of V_{GK}. The *cathode bias line* gives V_{GK} at any I_K, and because $I_A = I_K$ for a triode the point of intersection with the dc load line is the quiescent point. The cathode bias line is constructed thus:

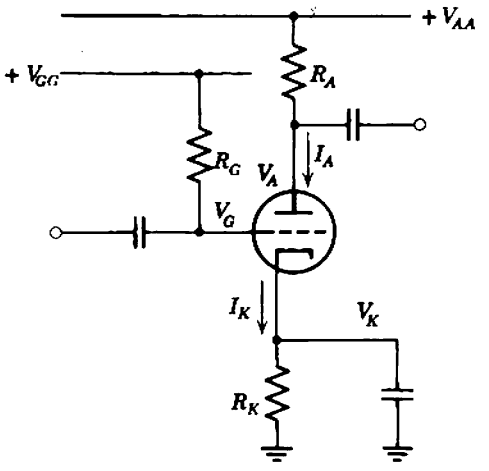
Suppose V_{GK} has some arbitrary value V_{GK0}; V_{GK} is a negative quantity. Equation 6.28 shows that the cathode current must be

$$I_{K0} = \frac{V_{GG} - V_{GK0}}{R_K}. \tag{6.29}$$

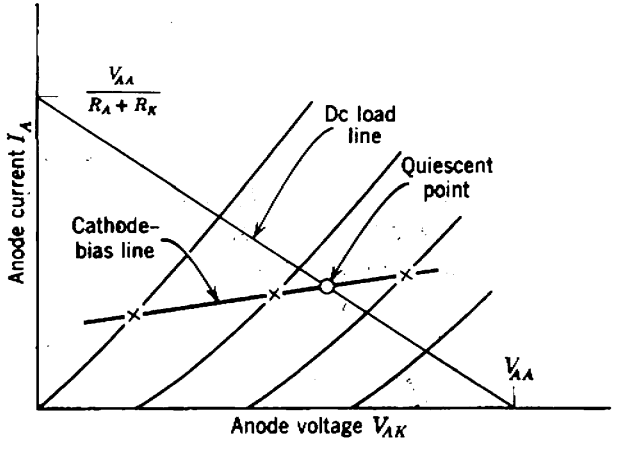
Therefore a point on the cathode bias line is at the level $I_K = I_{K0}$ on the $V_{GK} = V_{GK0}$ anode curve. Notice that the cathode bias line is slightly curved.

The special case of self biasing in which V_{GG} is set equal to zero follows directly. The only possible source of confusion is that V_{GK} is a negative quantity.

The ac load line, which can be used to investigate the signal performance, passes through the quiescent point found from the dc load line and cathode bias line. Because the cathode bypass capacitor constitutes a virtual short circuit at signal frequencies, the cathode resistor does not contribute to the ac load. Therefore the ac load is given by Eqs. 6.23 and 6.24 for



(a)



(b)

Fig. 6.19 Cathode-feedback biasing circuit and the cathode bias line. The circuit is in the modified form that requires two supplies of the same polarity.

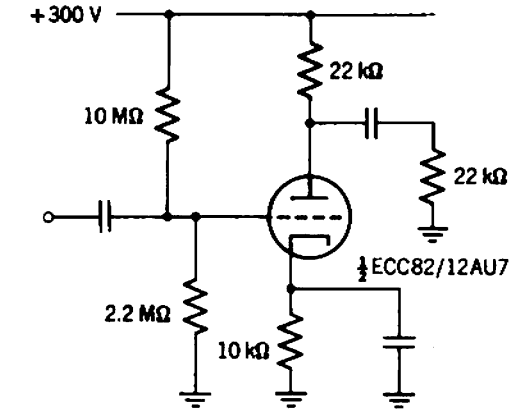
capacitance and transformer coupling respectively, and the gradient of the ac load line is

$$\frac{dI_A}{dV_A} = -\frac{1}{R_L}. \tag{6.30}$$

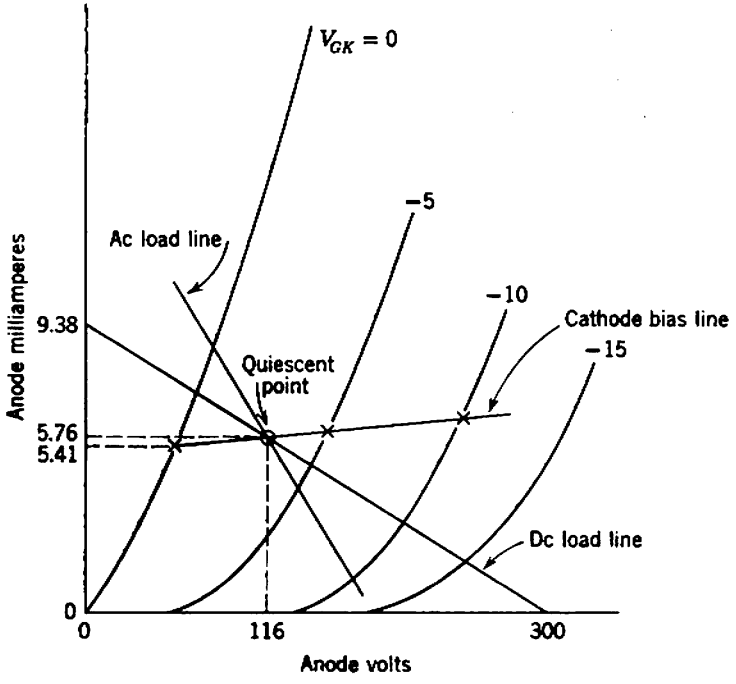
Figure 6.20 shows a worked-out example. The dc load line passes through the points

$$V_A = V_{AA} = 300 \text{ V at } I_A = 0,$$

$$V_A = 0 \text{ at } I_A = \frac{V_{AA}}{R_A + R_K} = 9.38 \text{ mA}.$$



(a)



(b)

Fig. 6.20 Example of cathode-feedback biasing, in the modified form that requires only one supply.

The effective grid supply potential V_{GG} is

$$V_{GG} = V_{AA}\left(\frac{R_{G2}}{R_{G1} + R_{G2}}\right) = 54.1 \text{ V}.$$

Therefore one point on the cathode bias line is

$$V_{GK} = \text{zero at } I_K = 5.41 \text{ mA}.$$

Other points can be calculated and the cathode bias line plotted. The quiescent point is at

$$V_{AK} = 116 \text{ V},$$

$$I_K = 5.76 \text{ mA},$$

$$V_{GK} = -3.5 \text{ V}.$$

The ac load line is drawn through this quiescent point and has a gradient corresponding to 11 kΩ.

6.2.1.4 Circuit Design

Sections 6.2.1.1 to 6.2.1.3 show how a cathode-feedback biasing system is analyzed when all component values and supply voltages are specified. The complementary problem of designing a circuit to bias a vacuum tube at any quiescent point is simpler. Moreover, the anode characteristics suffice for both triodes and pentodes.

Consider first the case of cathode-feedback biasing from a single supply (Fig. 6.21); V_{AA}, I_K, and R_A are specified, and V_{AK} (and V_{SK} for a pentode) is specified approximately. The cathode resistor R_K is calculated as follows:

$$V_A = V_{AA} - I_A R_A,$$

$$V_K = V_A - V_{AK},$$

$$R_K = \frac{V_K}{I_K}.$$

Therefore

$$R_K = \frac{V_{AA} - I_A R_A - V_{AK}}{I_K}. \tag{6.11}$$

[The anode and cathode currents are equal for a triode and related by the factor $(1 + k)$ for a pentode.] The equivalent grid supply V_{GG} becomes

$$V_{GG} = V_G = V_K + V_{GK};$$

that is,

$$V_{GG} = V_{AA} - I_A R_A - V_{AK} + V_{GK}, \tag{6.12}$$

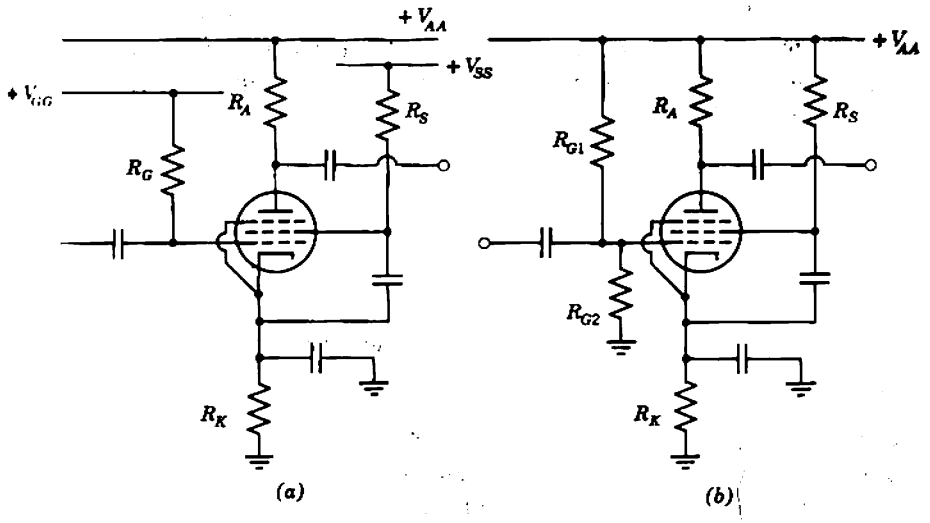


Fig. 6.21 Cathode-feedback biasing circuit using a pentode: (*a*) modified form, with three supplies of the same polarity; (*b*) usual practical arrangement, with one supply.

where V_{GK} at the specified I_K and V_{AK} (and V_{SK}) can be read directly from the anode characteristics. If the biasing is very stable, $|V_{GK}|$ in Eq. 6.32 is a small fraction of V_{GG} and can be neglected; the circuit component values becomes independent of the tube type. The screen supply resistor R_s of a pentode is given by

$$R_S = \frac{V_{SS} - V_K - V_{SK}}{I_S}$$

$$= \frac{V_{SS} - (V_{AA} - I_A R_A - V_{AK}) - V_{SK}}{I_S}.$$

Almost invariably the anode supply V_{AA} serves as the screen supply V_{SS}, so that

$$R_S = \frac{I_A R_A + V_{AK} - V_{SK}}{I_S}. \tag{6.33}$$

In the special cases of fixed and self biasing, V_{AK} is no longer an independent variable. A dc load line corresponding to R_A is drawn on the anode characteristics and the negative value of V_{GK} that gives the desired I_A can be read. The negative supply voltage V_{GG} for fixed bias, or the self-biasing resistor R_K $(= |V_{GK}|/I_K)$ follows. The error due to omitting R_K from the load line of a self biasing circuit is usually neglected; if desired, a more accurate design can be obtained by iteration. Another instance of V_{AK} not being independent is when the circuit takes the form of

Fig. 6.3*b*. The relations between Figs. 6.3 and 6.4 are such that V_{GG} in Fig. 6.21 is effectively specified. In these and other cases, V_{AK} can again be made independent by changing the effective anode supply voltage. Figure 6.22 shows a useful circuit trick:

$$V_{AA(\text{eff})} = V_{AA}\left(\frac{R_{A2}}{R_{A1} + R_{A2}}\right),$$

$$R_{A(\text{eff})} = R_{A1} \parallel R_{A2}.$$

6.2.2 Mutual Characteristics

The mutual characteristics of a vacuum tube are a set of plots of I_A versus V_G. For a triode V_A is the parameter for each curve, whereas for a pentode V_S is the significant parameter because I_A is almost independent of V_A. Usually each mutual curve for a pentode is plotted with $V_A = V_S$, although V_A is sometimes held constant throughout the entire set of curves. In addition, the mutual characteristics of a pentode include plots of I_g versus V_G. The general shape of the characteristics is the same for both triodes and pentodes; Figs. 6.23*a* and *b* show typical curves. Like the anode characteristics, manufacturers' published mutual characteristics for a given tube type are averages; because of differences in the averaging process, it is likely that points on the anode and mutual curves will not correspond exactly. Mutual characteristics are sometimes given a piecewise-linear representation.

The small-signal parameters at any quiescent point can be deduced from the mutual characteristics. For a triode g_m is the gradient of the curves,

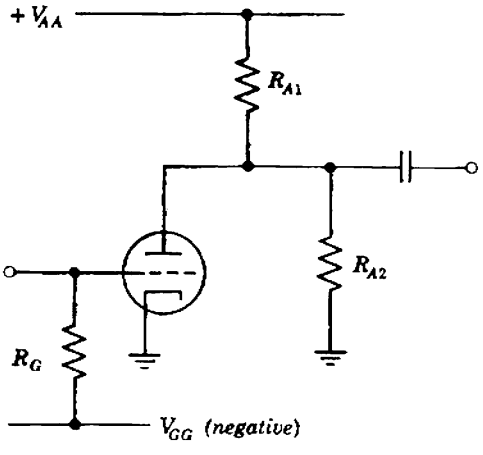


Fig. 6.22 Circuit trick for effectively reducing the anode supply voltage.

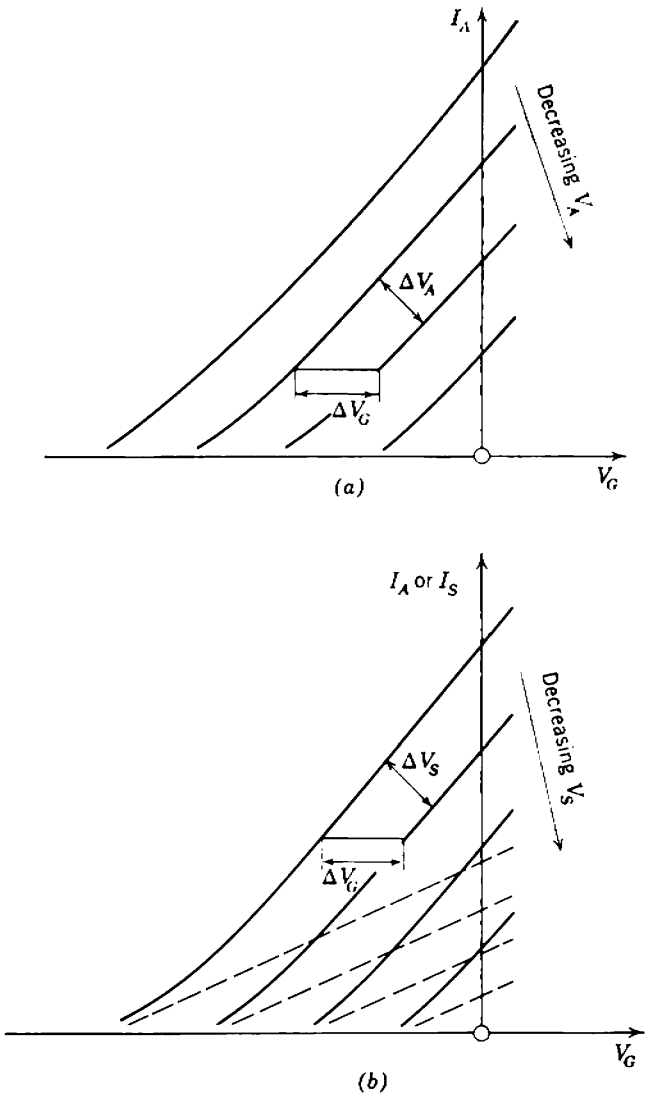


Fig. 6.23 Mutual characteristics of typical tubes: (*a*) triode, showing the construction for finding μ; (*b*) pentode, showing the construction for finding μ_S (—— = I_A, --- = I_S).

μ follows from the horizontal spacing as in Fig. 6.23*a*, and the anode resistance r_A follows from the vertical spacing (not shown in the diagram). For a pentode g_m is the gradient of the I_A curves, kg_m is the gradient of the I_S curves, and μ_S follows from the horizontal spacing of either set of curves (Fig. 6.23*b*); μ and r_A cannot be found because the variation of I_A with V_A is not shown on the curves.

In analyzing a circuit, a dc load line and cathode bias line can be drawn on the mutual characteristics of a tube to find its quiescent point; an ac load line and transfer characteristic can then be constructed to investigate the gain and distortion. These constructions on the mutual characteristics of a triode are less convenient than the corresponding constructions on the anode characteristics, but the opposite is true for a pentode. The reader is left to investigate methods of circuit design; the same principles apply for both anode and mutual characteristics. In general, biasing-circuit design is much simpler than biasing-circuit analysis.

6.2.2.1 Triode

In the triode circuit of Fig. 6.24*a*, the negative grid-to-cathode voltage is

$$V_{GK} = V_{GG} - I_K R_K,$$

and because the anode and cathode currents are equal

$$I_A = V_{GK}\left(-\frac{1}{R_K}\right) + \left(\frac{V_{GG}}{R_K}\right). \tag{6.34}$$

On axes of I_A versus V_{GK}, Eq. 6.34 represents a straight line of gradient $(-1/R_K)$ passing through the point V_{GG}/R_K on the I_A axis. The equation for the dc load line is

$$V_{AK} = V_{AA} - I_A R_A - I_K R_K,$$

and because I_A and I_K are equal

$$I_A = \frac{V_{AA} - V_{AK}}{R_A + R_K}. \tag{6.35}$$

Equation 6.35 represents a curved line, constructed as follows:

Suppose V_{AK} has some arbitrary value V_{AK0}. Equation 6.35 shows that the anode current must be

$$I_{A0} = \frac{V_{AA} - V_{AK0}}{R_A + R_K}. \tag{6.36}$$

Therefore a point on the load line is at the level $I_A = I_{A0}$ on the $V_{AK} = V_{AK0}$ mutual curve.

In the special case of self biasing for which V_{GG} is zero the cathode bias line passes through the origin. The term R_K in Eqs. 6.35 and 6.36 is usually ignored.

In principle, an ac load line can be constructed but, like the dc load line, it is curved and must be constructed on a point-by-point basis. It is therefore easier to deduce the transfer characteristic from the anode characteristics for which the load line is straight.

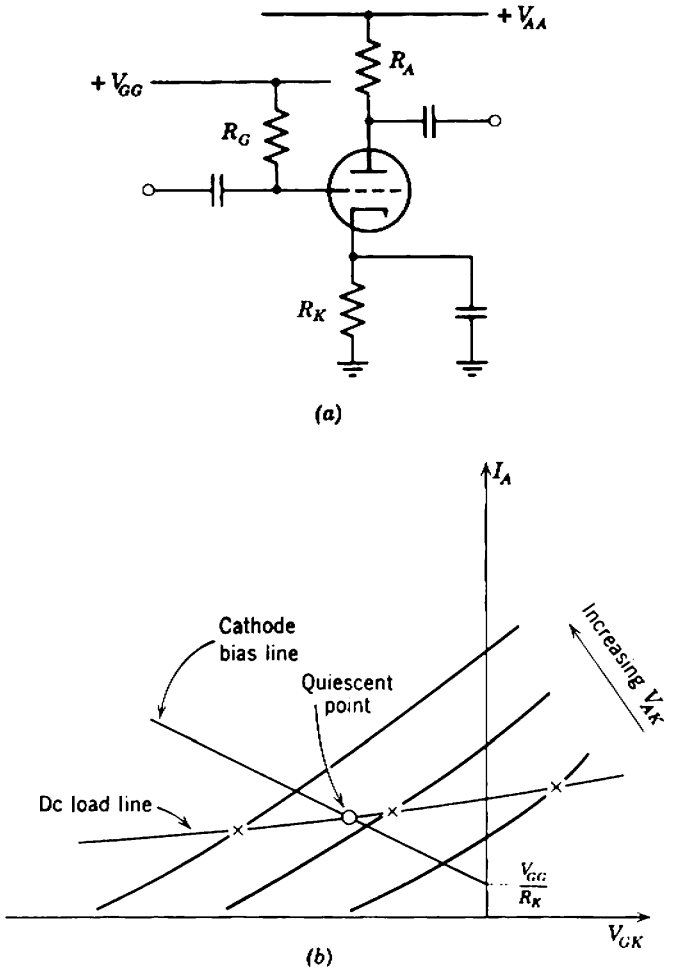


Fig. 6.24 Cathode-feedback biasing circuit shown in the modified form.

6.2.2.2 Pentode

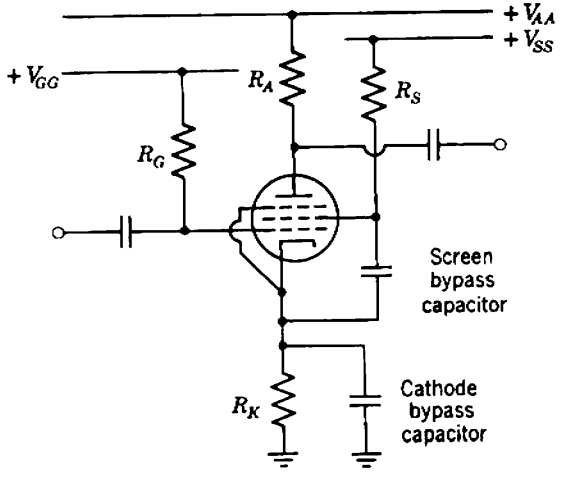
In using the mutual characteristic of a pentode, it is necessary to assume that I_A is independent of V_A, that is, to assume that both r_A and μ are infinite. This is a good approximation provided that V_A does not fall below the knee voltage; usually the error is less than the error due to the averaged nature of the characteristics published for a given tube type. Therefore the anode load resistance can be ignored in calculating the quiescent point and the gain of a pentode from its mutual characteristics, provided that the circuit is designed so that the tube operates above its knee.

In the pentode biasing circuit shown in Fig. 6.25*a* the cathode bias line is represented by

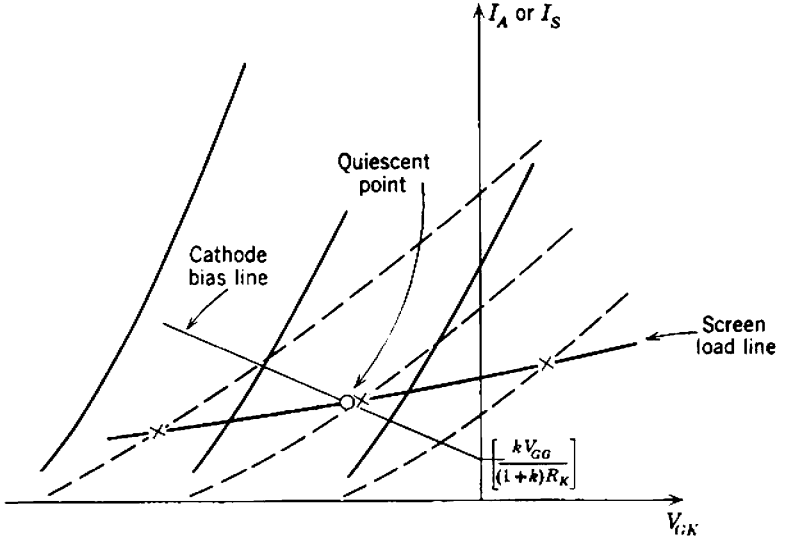
$$V_{GK} = V_{GG} - I_K R_K. \tag{6.37}$$

Because I_K is not given explicitly in the mutual curves, it is necessary to transform Eq. 6.37 to a form involving a known current:

$$V_{GK} = V_{GG} - I_S R_K \left(\frac{1 + k}{k}\right)$$



(*a*)



(*b*)

Fig. 6.25 Cathode-feedback biasing circuit shown in the modified form.

from which

$$I_S = V_{GK}\left[-\frac{k}{(1+k)R_K}\right] + \left[\frac{kV_{GG}}{(1+k)R_K}\right]. \tag{6.38}$$

The screen-to-cathode voltage is

$$V_{SK} = V_{SS} - I_S R_S - I_K R_K$$

from which

$$I_S = \frac{V_{SS} - V_{SK}}{R_S + R_K\left(\frac{1+k}{k}\right)}. \tag{6.39}$$

Equation 6.39 represents a curved *screen load line* on the mutual characteristics, and the quiescent point is at its intersection with the cathode bias line. The quiescent cathode current is

$$I_K = I_S\left(\frac{1+k}{k}\right). \tag{6.40}$$

This construction assumes that the ratio k ($= I_S/I_A$) is constant, whereas in fact k varies slightly. In the most precise work a better approximation can be obtained by iteration from the quiescent point found as above, but this is rarely necessary.

The screen of a pentode is normally bypassed to the cathode via a capacitor whose reactance at the lowest operating frequency is small. It can therefore be assumed that V_{SK} is constant for signal calculations. Because it is also assumed that I_A is independent of V_A, the mutual I_A curve corresponding to the quiescent V_{SK} is the I_A/V_G transfer characteristic. The gradient of this transfer characteristic is the transfer conductance G_T ($= g_m$ as expected for a pentode), and the curvature is indicative of the distortion that will be produced.

As an example, Fig. 6.26*a* shows the circuit diagram for an amplifier using a pentode type ECF80/6BL8. Figure 6.26*b* shows in outline the mutual characteristics. The ratio k of screen to anode current for this tube is about 0.33 so that

$$\frac{1+k}{k} \approx 4.0.$$

The screen and cathode bias lines are drawn on the characteristics and they intersect at

$$I_S = 2.0\text{ mA}, \qquad V_{GK} = -0.96\text{ V}.$$

Note that the contribution of R_K in Eq. 6.39 is negligible. The quiescent currents and voltages are

$$I_S = 2.0\text{ mA}, \qquad V_S = 100\text{ V},$$
$$I_A = 6.0\text{ mA}, \qquad V_A = 138\text{ V},$$
$$I_K = 8.0\text{ mA}, \qquad V_K = 0.96\text{ V}.$$

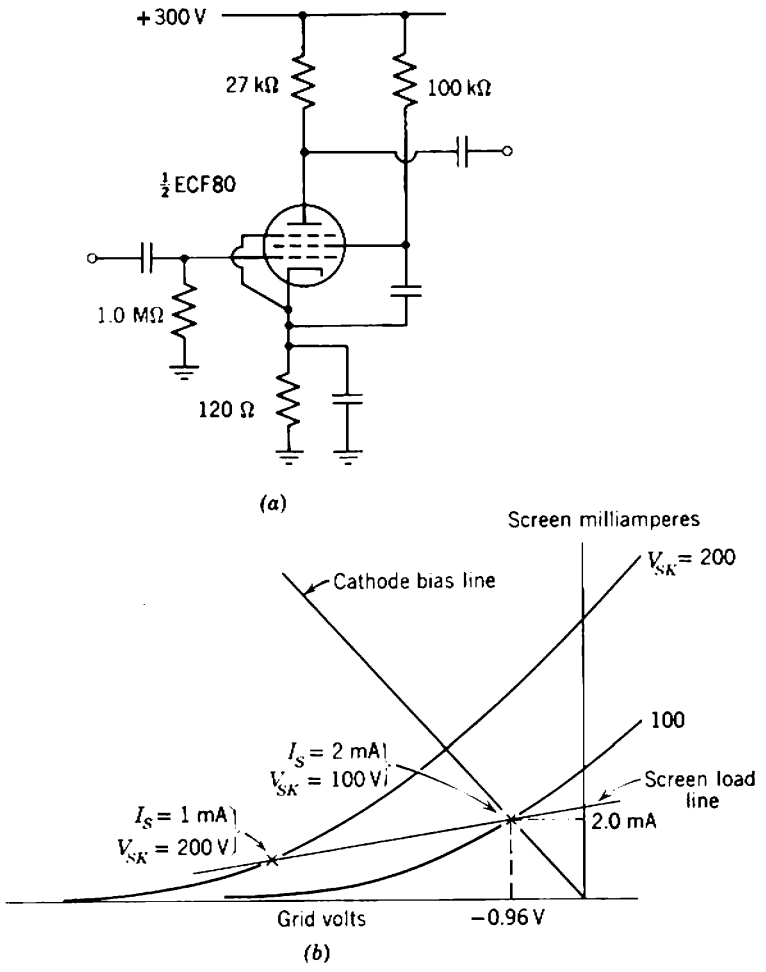


Fig. 6.26 Example of a self biasing circuit.

6.2.3 Graphical Methods for Transistors

All the foregoing graphical methods for vacuum tubes can be applied to transistors, but very little (if anything) is thereby accomplished.

The advantage of a graphical technique for designing a biasing circuit is its ease for incorporating the variation of V_{GK} or V_{BE} with quiescent point. The variation of V_{GK} for a tube is large enough to affect the design of a biasing system significantly. In contrast, V_{BE} for a transistor is an order of magnitude smaller and is known to within about 0.2 V for any operating conditions. Graphical methods are therefore unnecessary for designing transistor biasing circuits.

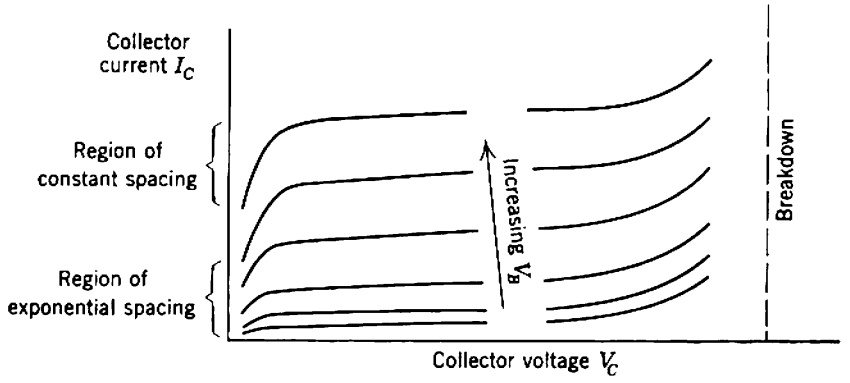


Fig. 6.27 Collector characteristics with V_{BE} as parameter.

In addition, graphs of transistor dc electrode voltages and currents provide no significant information about the variation of small-signal parameters. One set of *collector characteristics* is a plot of collector current versus collector voltage with base voltage as parameter. These characteristics correspond to the anode characteristics of a tube, but they are not often supplied by manufacturers. Because r_C for a transistor is large like r_A for a pentode, I_C is almost independent of V_C, and the characteristics approximate to horizontal straight lines (Fig. 6.27). Except at high currents, the spacing of the curves is exponential with V_B:

$$I_C \approx I_E \approx I_{E0} \exp\left(\frac{qV_B}{kT}\right). \tag{6.41}$$

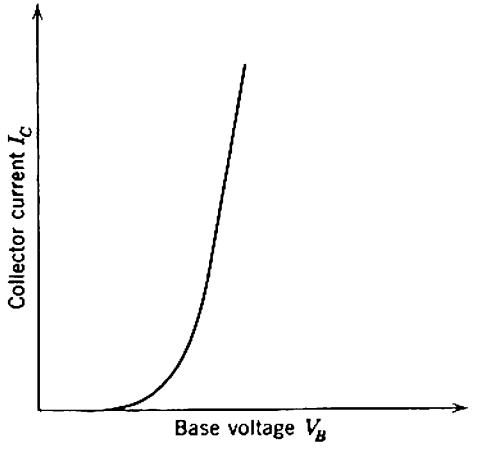


Fig. 6.28 Mutual characteristic of a transistor.

At high currents the spacing of the curves becomes approximately constant due to the presence of r_B. The base current causes a voltage drop across r_B, and the departure from the exponential spacing occurs when this drop becomes comparable with the drop across the internal junction. Thus the point of departure depends on base current, hence on β_N, and it is not reproducible. The region of the collector characteristics that does not follow Eq. 6.41 is therefore unreproducible, and average characteristics are of little use. The *mutual characteristics* of a transistor are plots of I_C versus V_B with V_C as parameter (Fig. 6.28). There is only one curve because I_C is almost independent of V_C; this curve follows the exponential law of Eq. 6.41 at low currents but becomes linear at high currents due to r_B.

Another set of *collector characteristics* that manufacturers often supply is an average plot of I_C versus V_C with I_B (rather than V_B) as parameter (Fig. 6.29). This set of curves shows how β_N varies with quiescent point for an "average" transistor but, since production transistors are likely to depart from the average β_N by a factor of two or more, the curves are of little use in either design or analysis.

In short, average graphs of dc terminal quantities provide no new, useful information. The measured characteristics of a particular transistor provide information on the variation of β_N, but this also is of little use because

(i) efficient circuit design does not require exact information on β_N;
(ii) β_N varies with the age, temperature, and environment of a transistor, even at a constant quiescent point.

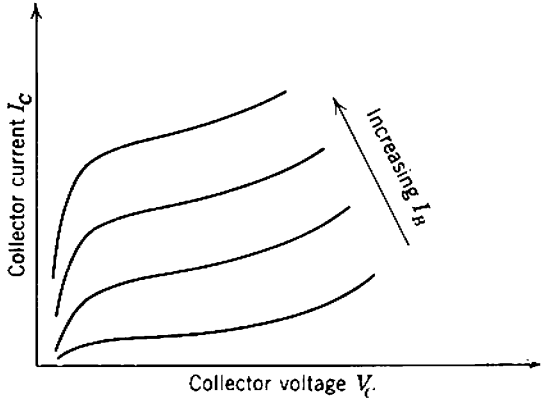


Fig. 6.29 Collector characteristics with I_B as parameter.

6.3 COLLECTING-ELECTRODE FEEDBACK*

A class of biasing systems quite different from those discussed in sections 6.1 and 6.2 is the family of collecting-electrode-feedback circuits. These biasing systems operate by holding the dc voltage of the collecting electrode constant and, thus, the emitting-electrode current is also constant. These biasing systems are based on Fig. 6.1; effectively, the collecting-electrode voltage is used to sense the emitting-electrode current.

6.3.1 Collector-Feedback Biasing

Figure 6.30 shows the circuit for a *collector-feedback biasing system* with the quiescent currents and voltages marked. Compared with the emitter-feedback circuit, collector-feedback biasing dispenses with two undesirable components:

(i) the emitter resistor R_E and its power wastage;
(ii) the emitter bypass capacitor C_E, simplifying low-frequency design (Sections 7.4.1 and 12.3).

Against this, a second power supply is required, and there is signal feedback via the biasing resistor R_F unless a bypass capacitor is connected from a tap on it to ground. Collector-feedback biasing is most commonly used

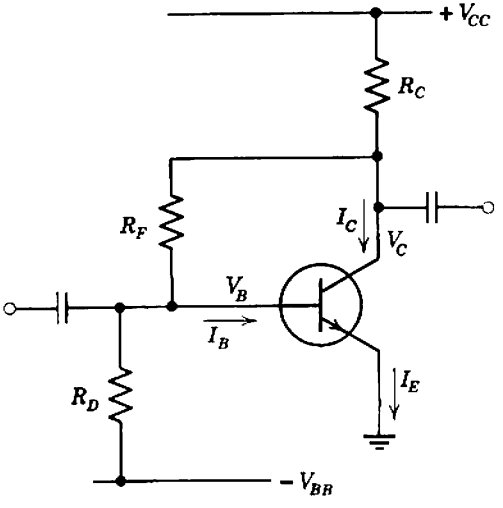


Fig. 6.30 Collector-feedback biasing circuit.

*Section 6.3 may be omitted on a first reading.

with the shunt-feedback stage discussed in Section 11.1.4. In outline, the operation of the circuit is as follows:

Suppose the circuit is disturbed in such a way that the emitter and collector currents increase; the collector voltage falls. This fall in voltage is transferred to the base via the divider chain R_F and R_D, where it tends to cut off the transistor and reduce the collector current. The disturbance is therefore counteracted and the circuit returns to its original state.

The collector voltage is given approximately by

$$V_C \approx V_{BB}\left(\frac{R_F}{R_D}\right) + \left[V_{CC} - V_{BB}\left(\frac{R_F}{R_D}\right)\right]\left(\frac{R_F}{R_F + \beta_N R_C}\right) + V_{BE}\left(\frac{R_D + R_F}{R_D}\right) - I_{CO}R_F. \quad (6.42)$$

The first term, $V_{BB}(R_F/R_D)$, makes the greatest contribution; the other three are the corrections due to finite values of β_N, V_{BE}, and I_{CO}, respectively. If the quiescent point is to be stable, these correcting terms must be small, and therefore the second-order cross-product corrections involving combinations of the variable parameters will be very small indeed. The approximation in Eq. 6.42 results from neglecting these cross-product terms entirely; in addition, it has been assumed that $\beta_N \approx (\beta_N + 1)$.

The philosophy underlying the design of a collector-feedback biasing system is basically the same as that for an emitter-feedback biasing system. The uncertain terms containing β_N, V_{BE}, and I_{CO} are made so small compared with the known term in Eq. 6.42 that the quiescent value of I_C always stays within specified limits. Thus, the following inequalities should be satisfied:

$$\beta_{NA} \gg \frac{R_F}{R_C}, \quad (6.43)$$

$$V_{BB} \gg V_{BEA}\left(\frac{R_D + R_F}{R_F}\right), \quad (6.44)$$

$$V_{BB} \gg I_{COA}R_D, \quad (6.45)$$

where β_{NA}, V_{BEA}, and I_{COA} are the average values of β_N, V_{BE}, and I_{CO}. A rule of thumb for about $\pm 10\%$ stability is to interpret the "much greater than" sign as "greater by a factor of 10." Equation 6.42 reduces to

$$V_C \approx V_{BB}\left(\frac{R_F}{R_D}\right). \quad (6.46)$$

The emitter current is

$$I_E = \frac{V_{CC} - V_C}{R_C} - \frac{V_{BB} + V_{EB}}{R_D} \quad (6.47)$$

to which a useful design approximation is

$$I_E \approx \frac{V_{CC} - V_C}{R_C} \approx \frac{V_{CC} - V_{BB}(R_F/R_D)}{R_C}. \tag{6.48}$$

Collector-feedback biasing should not be used when the external load is coupled via a transformer, because the primary winding resistance (corresponding to R_C) is small and the quiescent point is very unstable.

6.3.1.1 Self Biasing

The *self biasing circuit* shown in Fig. 6.31 is a degenerate form of collector-feedback biasing in which R_D and the V_{BB} supply vanish. (This transistor self biasing circuit bears no relation to the vacuum-tube self biasing circuit in Section 6.1.2.2.) The emitter current is

$$I_E = \left(\frac{V_{CC} - V_{BE} + I_{CO}R_F}{R_C}\right)\left(\frac{1}{1 + R_F/\beta_N R_C}\right). \tag{6.49}$$

Unlike all transistor biasing circuits considered so far, the correcting terms due to V_{BE}, I_{CO}, and particularly β_N are not made small, for then

$$I_E \approx \frac{V_{CC}}{R_C}$$

and the transistor is bottomed. Because these correcting terms are not small, the biasing stability is poor; self biasing should be used only in minimal-cost low-quality equipment.

It is good practice to design a self biasing circuit with

$$R_F \approx \beta_{NA} R_C \tag{6.50}$$

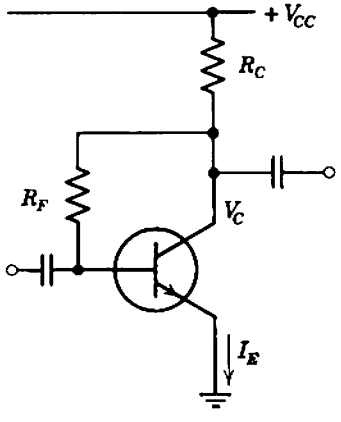


Fig. 6.31 Self biasing circuit.

so that

$$I_E \approx 0.5\left(\frac{V_{CC}}{R_C}\right), \qquad V_C \approx 0.5V_{CC}.$$

The conditions for stability against V_{BE} and I_{CO} are

$$V_{CC} \gg V_{BEA}, \tag{6.51}$$

$$V_{CC} \gg I_{COA}R_F. \tag{6.52}$$

Equation 6.51 is almost always satisfied, but Eq. 6.52 is difficult to satisfy at elevated temperatures. Section 6.3.2.1 describes a modified circuit that exchanges stability against V_{BE} for stability against β_N and I_{CO}.

6.3.2 Combined Collector- and Emitter-Feedback Biasing

Figure 6.32 is a circuit that combines collector and emitter feedback. It has a high order of biasing stability, and requires only a single supply voltage. Like collector-feedback biasing, combined-feedback biasing is particularly useful with shunt-feedback stages. The emitter current is

$$I_E = \left\{\frac{V_{CC} - V_{BE}[(R_C + R_F + R_D)/R_D] + I_{CO}R_F}{R_E}\right\} \times \left[\frac{R_D}{R_C(1 + R_D/R_E) + R_F + R_D}\right] \times \left\{1 + \left(\frac{R_F}{\beta_N R_E}\right)\left[\frac{R_D}{R_C(1 + R_D/R_E) + R_F + R_D}\right]\right\}^{-1}. \tag{6.53}$$

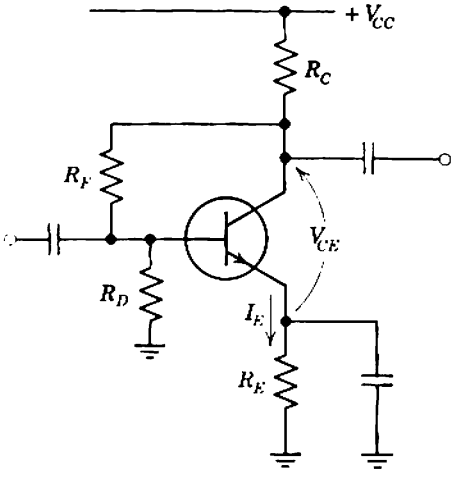


Fig. 6.32 Combined collector- and emitter-feedback biasing system.

The conditions necessary for stability are

$$\beta_{NA} \gg \frac{R_F}{R_E}\left[\frac{R_D}{R_C(1 + R_D/R_E) + R_F + R_D}\right], \tag{6.54}$$

$$V_{CC} \gg V_{BEA}\left(\frac{R_C + R_F + R_D}{R_D}\right), \tag{6.55}$$

$$V_{CC} \gg I_{COA}R_F. \tag{6.56}$$

If these conditions are satisfied, Eq. 6.53 reduces to

$$I_E \approx \frac{V_{CC}}{R_E}\left[\frac{R_D}{R_C(1 + R_D/R_E) + R_F + R_D}\right], \tag{6.57}$$

The collector-to-emitter voltage is

$$V_{CE} = V_C - V_E = \left[V_{CC} - R_C\left(I_E + \frac{V_B}{R_D}\right)\right] - I_E R_E$$

and since $V_B \approx V_E$ (because V_{BE} is small for a transistor), substitution from the approximate Eq. 6.57 gives

$$V_{CE} \approx V_{CC}\left[\frac{R_F}{R_C(1 + R_D/R_E) + R_F + R_D}\right]. \tag{6.58}$$

6.3.2.1 Modified Self Biasing

Figure 6.33 shows a degenerate form of the combined-feedback biasing system in which R_E becomes zero. Alternatively, the circuit can be

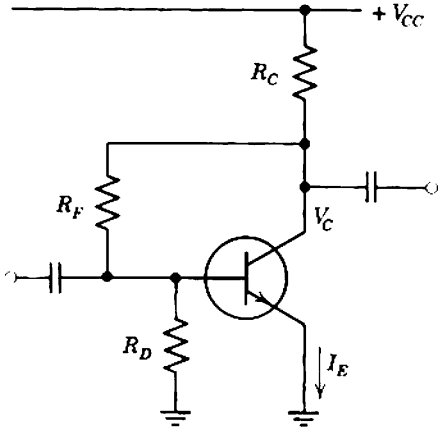


Fig. 6.33 Modified self biasing system (compare Fig. 6.31).

considered as a modified form of the self biasing system. The emitter current is

$$I_E = \left\{\frac{V_{CC} - V_{BE}[(R_C + R_F + R_D)/R_D] + I_{CO}R_F}{R_C}\right\}\left[\frac{1}{1 + R_F/\beta_N R_C}\right]. \quad (6.59)$$

Like the simple self biasing circuit, the modified circuit gives relatively poor stabilization of the quiescent point. Comparison of Eqs. 6.49 and 6.59 shows that R_F can be made much smaller in the modified circuit, with a consequent increase in stability against variations in β_N and I_{CO}. Some stability against variation in V_{BE} is lost, but this is no great disadvantage because V_{BE} is relatively predictable.

The conditions necessary for biasing stability are

$$\beta_{NA} \gg \frac{R_F}{R_C}, \quad (6.60)$$

$$V_{CC} \gg V_{BEA}\left(\frac{R_C + R_F + R_D}{R_D}\right), \quad (6.61)$$

$$V_{CC} \gg I_{COA}R_F. \quad (6.62)$$

These three inequalities cannot be simultaneously satisfied by large amounts at a reasonable quiescent point. The best compromise can be found only by trial and error; usually, it occurs when R_F is about one-third the value suggested by Eq. 6.50 for the simple circuit, with R_D adjusted to give I_E a quiescent value of about $0.5(V_{CC}/R_C)$. If properly designed, the modified circuit is a distinct improvement over the simple circuit; improperly designed, it can be much worse.

6.3.3 Vacuum-Tube Circuits

In this section, design equations are deduced for the vacuum-tube counterparts of the transistor circuits described in Sections 6.3.1 and 6.3.2. It should be noted that circuits corresponding to the two degenerate forms do not exist. The operation of the transistor circuits depends on the existence of a base current, so that there is a voltage drop across R_F. In comparison, the grid current of a vacuum tube at a normal quiescent point can be neglected.

6.3.3.1 Anode-Feedback Biasing

Figure 6.34 shows the *anode-feedback biasing circuit* for a triode. Since there are no vacuum-tube parameters corresponding to β_N and I_{CO}, the uncertainty in quiescent point arises entirely from the uncertainty in V_{GK}. By analogy with Eq. 6.42, the anode voltage is

$$V_A = V_{GG}\left(\frac{R_F}{R_D}\right) + V_{GK}\left(\frac{R_F + R_D}{R_D}\right). \quad (6.63)$$

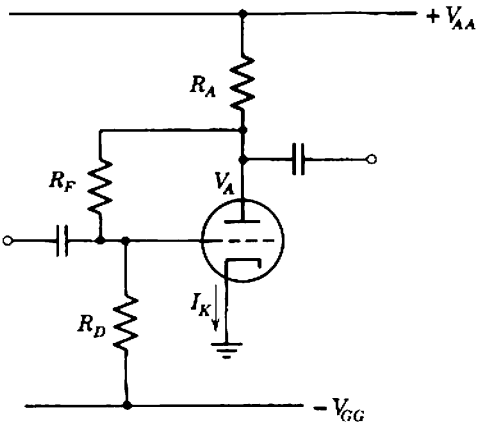


Fig. 6.34 Anode-feedback biasing system (suitable for both triodes and pentodes).

Note that V_{GK} is a negative quantity. If the uncertain correcting term due to V_{GK} is made small, Eq. 6.63 reduces to

$$V_A \approx V_{GG}\left(\frac{R_F}{R_D}\right), \tag{6.64}$$

The anode current is

$$I_A = \frac{V_{AA} - V_A}{R_A} - \frac{V_{GG} + V_{GK}}{R_D} \tag{6.65}$$

to which a useful design approximation is

$$I_A \approx \frac{V_{AA} - V_A}{R_A} \approx \frac{V_{AA} - V_{GG}(R_F/R_D)}{R_A}. \tag{6.66}$$

For a triode the anode and cathode currents are equal, whereas for a pentode, $I_K = I_A(1 + k)$.

The possible uncertainty in V_{GK} for a tube is much greater than the uncertainty in V_{BE} for a transistor. As with the emitting-electrode-feedback biasing system, it is always possible to design an anode-feedback circuit that will accommodate any value of V_{GK}, but to do so a very large value of V_{GG} is required. In circuit design it is therefore advisable to find the nominal value of V_{GK} at the required quiescent point from the anode or mutual characteristics, and include this in Eq. 6.63. In the analysis of a given circuit the approximate quiescent point can be found from Eqs. 6.64 and 6.66, the value of V_{GK} at this quiescent point can be found from the characteristic curves, and, finally, a more accurate solution is found by substituting this value of V_{GK} into Eqs. 6.63 and 6.65. If desired, a still more accurate solution can be found by repeated approximation.

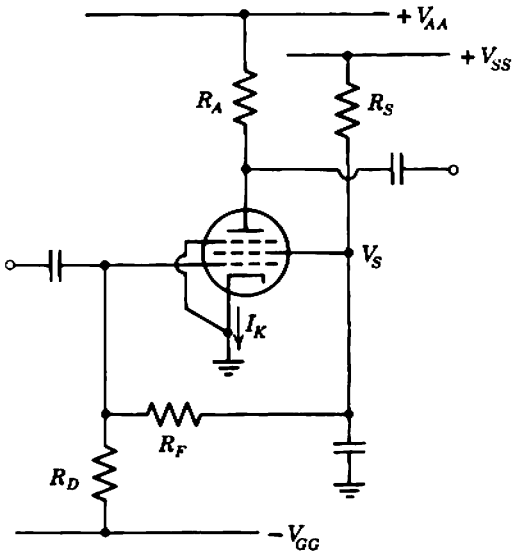


Fig. 6.35 Screen-feedback biasing system.

6.3.3.2 Screen-Feedback Biasing

Figure 6.35 shows a modified form of the anode-feedback circuit, suitable for use with pentodes. The feedback is taken from the screen and, because the screen is bypassed to ground at signal frequencies, there is no signal feedback via the biasing network. By analogy with Eqs. 6.66 and 6.64, the cathode current and screen voltage are approximately

$$I_K \approx \left(\frac{1+k}{k}\right)\left[\frac{V_{SS} - V_{GG}(R_F/R_D)}{R_S}\right], \tag{6.67}$$

$$V_S \approx V_{GG}\left(\frac{R_F}{R_D}\right). \tag{6.68}$$

It is left as an exercise for the reader to deduce the other design equations.

6.3.3.3 Combined Anode- and Cathode-Feedback Biasing

Figure 6.36 shows the circuit for combined anode and cathode feedback, suitable for both triodes and pentodes. The cathode current is

$$I_K = \left\{\frac{V_{AA} - V_{GK}[(R_A + R_F + R_D)/R_D]}{R_K}\right\} \times \left\{\frac{R_D}{R_A[1 + R_D/(1+k)R_K] + R_F + R_D}\right\}. \tag{6.69}$$

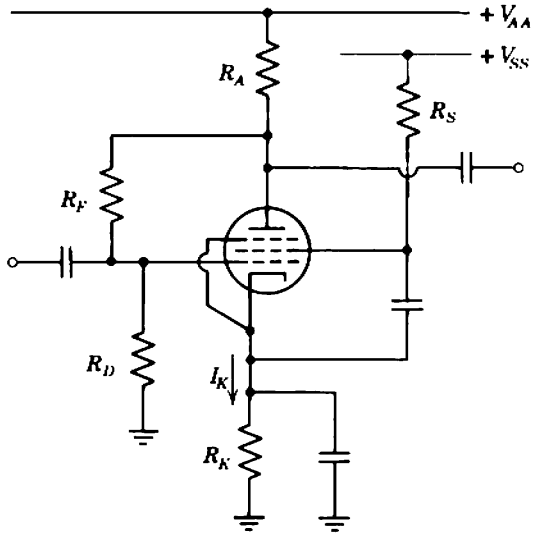


Fig. 6.36 Combined anode- and cathode-feedback biasing system (suitable for both triodes and pentodes).

If the correcting term due to V_{GK} is small, then

$$I_K \approx \frac{V_{AA}}{R_K}\left\{\frac{R_D}{R_A[1 + R_D/(1 + k)R_K] + R_F + R_D}\right\}, \tag{6.70}$$

$$V_{AK} \approx V_{AA}\left\{\frac{R_F}{R_A[1 + R_D/(1 + k)R_K] + R_F + R_D}\right\}. \tag{6.71}$$

For a triode, k is zero.

6.3.3.4 Combined Screen- and Cathode-Feedback Biasing

Figure 6.37 shows a circuit for pentodes that combines screen and cathode feedback. By analogy with Eqs. 6.70 and 6.71, the cathode current and screen-to-cathode voltage are

$$I_K \approx \frac{V_{SS}}{R_K}\left\{\frac{R_D}{R_S[1 + kR_D/(1 + k)R_K] + R_F + R_D}\right\}, \tag{6.72}$$

$$V_{SK} = V_{SS}\left\{\frac{R_F}{R_S[1 + kR_D/(1 + k)R_K] + R_F + R_D}\right\}. \tag{6.73}$$

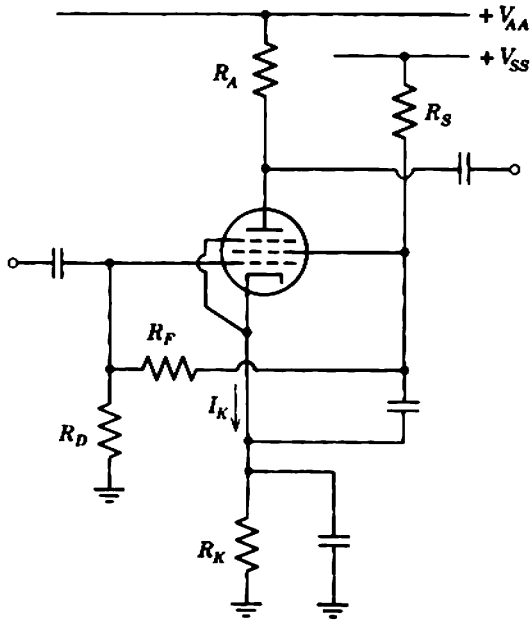


Fig. 6.37 Combined screen- and cathode-feedback biasing system.

6.4 BYPASS, COUPLING, AND DECOUPLING COMPONENTS

The purpose of a biasing system is to fix the relevant operating currents and voltages at values for which the small-signal parameters have known and desired values. Only if these parameters are known can the small-signal analysis of Chapter 5 be usefully applied to calculate the gain of an amplifier. Bypass, coupling, and decoupling components provide a means for applying time-varying signal currents and voltages to the devices without disturbing the quiescent currents and voltages.

6.4.1 Bypass Capacitors

With the exception of the special amplifier configurations discussed in Section 5.7, the whole mid-band analysis of Chapter 5 assumes that there is no signal voltage at the emitting electrode despite the presence of a signal current in it. The emitting electrode is taken as the datum point for signals. Therefore, when the emitting electrode is not connected directly to ground (as is the case for many biasing circuits), it is necessary to provide a low-impedance path from the electrode to ground; the cathode or emitter bypass capacitor introduced in Section 6.1 (Fig. 6.3) provides this path.

It is also assumed in Chapter 5 that no signal voltage exists between the screen and cathode of a pentode. The screen bypass capacitor introduced in Section 6.2.2.2 (Fig. 6.25) provides a low-impedance path at signal frequencies between these electrodes. Strictly speaking, the screen bypass capacitor should be connected between the screen and cathode, but if the cathode is bypassed to ground and is at zero signal voltage it is sufficient to connect the screen bypass capacitor between the screen and ground. This alternative arrangement usually leads to simpler construction of an amplifier.

The impedance of a bypass capacitor should be so small that it constitutes an effective short circuit and can, therefore, be neglected from small-signal calculations. The order of impedance that is negligible depends on the application; emitter bypass capacitors for most transistor stages are required to have an impedance as low as a few ohms, while screen bypass capacitors for most pentodes are allowed an impedance as high as several tens of kilohms. It is shown in Section 7.4.1 that, if the impedance of a bypass capacitor is not negligible for its particular application, the gain of the amplifier stages is reduced.

The reactance and impedance of a capacitor are inversely proportional to frequency. There is, therefore, some limiting frequency below which the impedance of any bypass capacitor cannot be neglected, and below which the gain must fall. Section 7.4.1 contains the relations between capacitance, 3-dB cutoff frequency, and the other circuit elements.

6.4.2 Coupling Capacitors and Transformers

Coupling components are used to transfer the signal (or ac) voltages and currents from one amplifier stage to the next, while isolating the quiescent (or dc) voltages and currents. The two basic circuits, capacitor and transformer coupling, are introduced in Section 6.2.1.2 (Fig. 6.17).

A coupling capacitor is chosen to have an impedance that constitutes a virtual short circuit at all the useful signal frequencies. Therefore, the signal voltage at the output of one stage is transferred without loss to the input of the following stage. Further, the load resistance for any device becomes the parallel combination of its own collecting-electrode supply resistor, the control-electrode supply resistor for the following device, and the input resistance of the following device. Like a bypass capacitor, the order of impedance that is negligible depends on the application. Acceptable values range from a few ohms to tens of megohms. A capacitor is a perfect coupling element at very high frequencies where its reactance tends to zero but not at low frequencies where its reactance rises. Section

7.4.2 contains the relations between capacitance, 3-dB cutoff frequency, and the other circuit elements.

Transformer coupling has both advantages and disadvantages compared with capacitor coupling. The greatest advantage is that a transformer can be used as an impedance changer, to change the effective source and load impedance seen by a device. It is shown in Section 8.5.2 that there is an optimum transformer turns ratio between a given source and device at which the signal-to-noise ratio is maximum. It is also shown in Section 16.2 that there is an optimum transformer turns ratio between a given device and load at which the available power output is maximum. A second advantage of a transformer is that the winding resistances are much less than the reflected impedance at signal frequencies; dc power wastage is minimized and, as discussed in Section 6.2.1.2, peak-to-peak output voltages in excess of the supply voltage can be obtained. Finally, transformer coupling is advantageous when a very high order of biasing stability is required. Figure 6.38 shows the circuit diagram for a transformer-coupled transistor amplifier with emitter-feedback biasing. The secondary winding resistance is small, so that the base supply resistance R_B in Eq. 6.4 is small and thus the biasing stability is good.

The disadvantage of a transformer is that its coupling efficiency falls at both low and high frequencies because of the reactive effects discussed in Section 7.6. The performance is reasonably predictable at low frequencies but not, in general, at high frequencies. Further, a coupling transformer costs much more than a coupling capacitor. Transformer

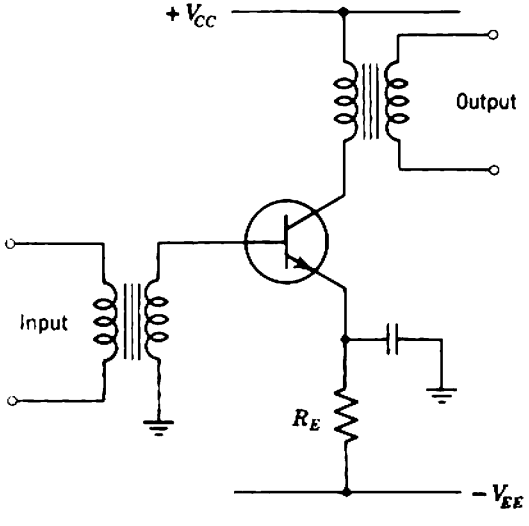


Fig. 6.38 Transformer-coupled transistor amplifier stage.

coupling is little used in modern low-pass amplifiers except when one or more of the advantages listed above is significant.

6.4.3 Bypass and Decoupling Capacitors in the Dc Supply System

The mid-band circuit theory of Chapter 5 assumes that there are no ac voltages on the dc supply rails. In practice, ac voltages can appear for either of two reasons.

The dc supply voltages in most electronic equipment are obtained by rectifying the ac power mains. The filtering that follows the rectification is not perfect, and invariably the dc voltages contain components at harmonics of the main frequency. Decoupling networks are connected in the supply rails to the early stages of most amplifiers to improve the filtering and reduce these *hum voltages* (Fig. 6.39). The decoupling network is a voltage divider whose attenuation ratio is higher for ac than for dc. Alternatively, the decoupling network is an RC low-pass filter. The ratio of the ac voltage v_M on the main supply rail to the voltage v_D at the decoupled supply point is

$$\frac{v_D}{v_M} \approx \frac{1}{1 + j\omega R_D C_D} \approx \frac{1}{j\omega R_D C_D}. \tag{6.74}$$

To be effective, a decoupling network must have a time constant long compared with the period of the hum. For example, if v_M is a 120 Hz hum originating from full-wave rectification of 60 Hz ac power mains, the attenuation is 70 times if $R_D C_D$ is 100 msec. The values shown in Fig. 6.39 (10 kΩ and 10 μF) are typical of vacuum-tube circuits. When a very

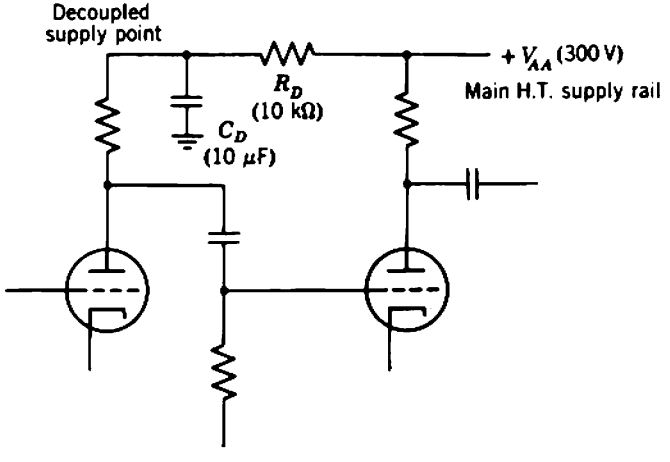


Fig. 6.39 Decoupling network.

high attenuation is required, two decoupling networks in cascade are more efficient than one network of very long time constant; the individual attenuations multiply rather than add. It is hardly necessary to point out that a decoupling resistor contributes to the dc load line.

In addition to hum voltages, signal-frequency voltages appear on the supply rails. Any power supply has a finite internal impedance, and therefore the signal currents drawn by the active devices produce signal-frequency fluctuations of the supply voltage. In a single-stage amplifier the power-supply impedance should be small compared with the appropriate supply resistor; if it is not, a bypass capacitor should be connected from the supply rail to ground to reduce the effective supply impedance at signal frequencies. In a multistage amplifier all stages contribute to the fluctuation of the supply voltage but the greatest part is due to the output stage where the signal currents are greatest. The fluctuation due to the output stage is fed back into the early stages, where it is amplified and finally reappears at the output. This is an example of a feedback system, and it is shown in Chapter 10 that a feedback system can break into oscillation. The remedy is to reduce the amplitude of the voltage fed back via the supply rail; this can be accomplished by using a low-impedance supply or, more simply, by decoupling the supply to the early stages.

Thus, decoupling networks serve the dual purposes of reducing hum and preventing signal feedback via the supply rails. It is shown in Section 7.4.3 that decoupling networks can also be used to compensate the low-frequency response.

6.5 DIRECT-COUPLED STAGES*

It is possible to design multistage amplifiers so that the quiescent collecting-electrode voltage of one device is the same as the control-electrode voltage of the following device. Under these circumstances the coupling capacitor may be omitted and the two stages are said to be *direct coupled.* Direct coupling has both advantages and disadvantages compared with resistance-capacitance coupling and transformer coupling. These are most conveniently discussed in relation to a specific circuit.

6.5.1 General Considerations

The left side of Fig. 6.40 is a transistor amplifier stage with emitter-feedback biasing; its emitter current I_{E1} and collector voltage V_{C1} can be

* Section 6.5 may be omitted on a first reading.

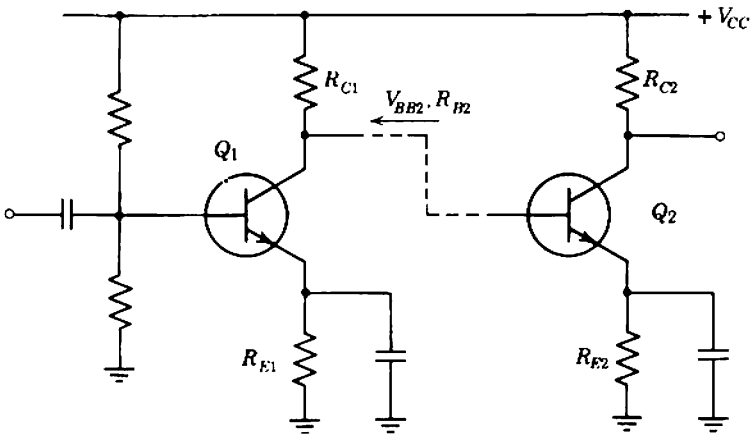


Fig. 6.40 Derivation of a direct-coupled group.

found from the equations of Section 6.1.1. According to Thevenin's theorem, the output circuit of this stage can be replaced by a voltage source in series with a resistance. As far as quiescent conditions are concerned, the equivalent voltage is V_{C1} and the equivalent resistance is R_{C1}. (The latter follows because the I_{C1} is almost independent of the collector voltage; complete independence is assumed in Section 6.1.1.) This equivalent voltage and resistance can be considered as the base supply V_{BB2} and R_{B2} of the emitter-feedback biasing system for the transistor on the right side of Fig. 6.40, and its quiescent point can be found. Notice that the collector voltage of Q_1 changes slightly when Q_2 is connected. A third stage can be direct coupled to Q_2, and if the process is repeated an amplifier of the type shown in Fig. 6.41 results. The quiescent point of all transistors in Fig. 6.41 is

$$I_E = 1 \text{ mA}, \qquad V_{CE} = 3 \text{ V}.$$

Nonpreferred resistor values are used to simplify the arithmetic.

Direct coupling offers two very great advantages in advanced amplifiers, though neither is significant in the simple amplifiers discussed in Chapter 5. First, direct coupling eliminates the coupling capacitor between two stages. It is shown in Section 7.4.2 that each coupling capacitor contributes a low-frequency pole and zero to an amplifier's singularity pattern, and it follows that eliminating a capacitor improves the low-frequency response. Direct coupling forms the basis of many of the low-frequency compensating schemes described in Section 12.3. Second, direct coupling eliminates the control-electrode supply resistor for one stage, thereby reducing the conductance in shunt with the signal path. It is shown in Section 11.2.2 that

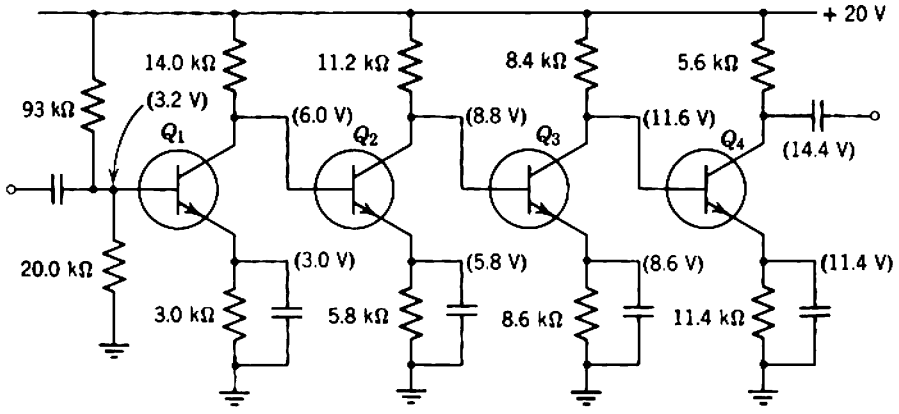


Fig. 6.41 Long direct-coupled group. The quiescent conditions for each transistor are

$$I_E = 1 \text{ mA}, \qquad V_{CE} = 3 \text{ V},$$

assuming $I_B = 20\,\mu\text{A}$. Nonpreferred resistor values are used to simplify the arithmetic.

the gain available from a feedback stage increases as the shunt conductance is reduced. Direct coupling is mandatory in integrated-circuit amplifiers because large coupling capacitors cannot be realized.

The principal disadvantage of direct coupling is that all resistor values and supply voltages must be known more accurately than in single-stage biasing systems, because errors in the biasing are transmitted through the amplifier and can accumulate. As an example, if the voltage at the base of Q_1 in Fig. 6.41 changes from 3.2 to 3.4 V (a change of only 7%), V_{CE} of Q_3 falls from 3 to 0 V, and Q_3 becomes inoperative. It is usually necessary to specify 5% or closer tolerance resistors rather than standard 10% tolerance types. It is also wise to restrict direct coupling to two, or at most three stages. A further disadvantage of direct coupling is that appreciable power is wasted in the supply resistors of any stage, because most of the supply voltage appears across them rather than across the active device. The remainder of this section is therefore confined to groups of two direct-coupled stages. The larger, less-designable groups are discussed briefly in Section 6.6.1.

The first of a group of direct-coupled stages need not have emitting-electrode-feedback biasing. Collecting-electrode-feedback systems are not only suitable but often more satisfactory. Section 10.7 shows that feedback changes the output impedance of an amplifier and, in the case of a collecting-electrode-feedback system, it reduces the output impedance to a low value. The second stage in a direct-coupled group is therefore

fed from a low-impedance control-electrode supply, and the biasing stability is increased.

Direct coupling is less suitable for vacuum tubes than for transistors. First, the power dissipated in the various R_A and R_K of a vacuum-tube circuit is much larger than the corresponding power in R_C and R_E, because of the much higher voltages in the vacuum-tube circuits. Secondly, there is a restriction on the maximum dc voltage allowed between the heater and cathode of a tube (Section 3.3); this rating will usually be exceeded in a multistage amplifier when all tubes have their heaters supplied from a common source. Finally, the voltage rating of the cathode bypass capacitors in vacuum-tube circuits must be inconveniently high. Miniature capacitors of about 12-V rating are adequate for direct-coupled transistor stages and most single vacuum-tube stages employing self bias. Capacitors of 100-V rating or more are necessary for direct-coupled vacuum-tube stages; such capacitors are expensive and bulky. Despite these limitations, direct-coupled vacuum-tube stages are worthy of consideration.

6.5.2 Two-Stage Transistor Circuits

Groups of two direct-coupled transistor stages are useful building-blocks for multistage amplifiers. They are perhaps more useful in feedback amplifiers than in simple amplifiers of the type discussed in Chapter 5. Three basic configurations exist with a number of modified forms.

The design equations given in the following sections are intended to show how a circuit works, rather than to give an accurate estimate of the quiescent conditions. The approximations may be quite poor, particularly for silicon transistors because of their large base-emitter voltage drop. The easiest way to analyze or design a circuit is not to derive an accurate set of equations, but to make common-sense allowances of a few tenths of a volt for the voltage drops due to base currents and for the base-emitter voltage drops.

6.5.2.1 Two-Stage Emitter-Feedback Biasing

Figure 6.42 shows the basic two-stage emitter-feedback biasing circuit. The design equations are:

$$I_{E1} \approx \frac{V_{EE}}{R_{E1}}, \tag{6.75}$$

$$I_{E2} = \frac{V_{E2}}{R_{E2}} \approx \frac{V_{C1}}{R_{E2}}, \tag{6.76}$$

where

$$V_{C1} \approx V_{CC} - I_{C1}R_{C1}. \tag{6.77}$$

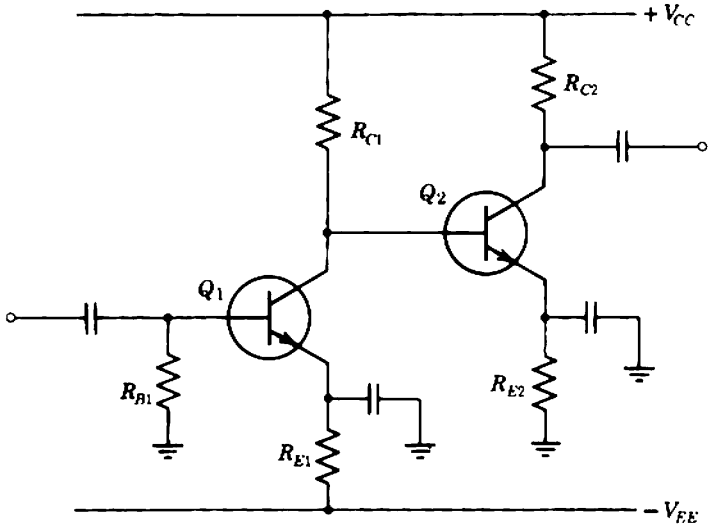


Fig. 6.42 Basic form of two-stage emitter-feedback biasing.

The corrections required for these equations are the following:

(i) All terms involving β_N, V_{BE}, and I_{CO} of Q_1 are omitted from Eq. 6.75.
(ii) V_{BE} of Q_2 is omitted from Eq. 6.76.
(iii) The base current of Q_2 is omitted from Eq. 6.77.

Two modified forms of this circuit exist. The first is the circuit shown in Fig. 6.40, for which only one supply voltage is required. In the second R_{E2} of the basic circuit is returned to the $-V_{EE}$ supply rather than to ground; this improves the biasing stability of the second stage at the cost of increased power wastage in R_{E2}. None of the three forms of two-stage emitter-feedback biasing is particularly useful.

6.5.2.2 Two-Stage Collector-Feedback Biasing

Many forms of the two-stage collector-feedback biasing circuit exist; Fig. 6.43*a* shows the basic circuit and Fig. 6.43*b* shows the most useful modification. As in the single-stage collector-feedback circuit (Fig. 6.30) the biasing resistor R_F in the basic two-stage circuit provides signal feedback around the first stage. The modified circuit is useful when this signal feedback must be eliminated; R_F is connected to the emitter of Q_2 rather than the collector of Q_1, and because these two points are at nearly the same dc voltage the quiescent conditions are similar to those in the basic circuit. The emitter of Q_2, however, is bypassed to ground, so there is no

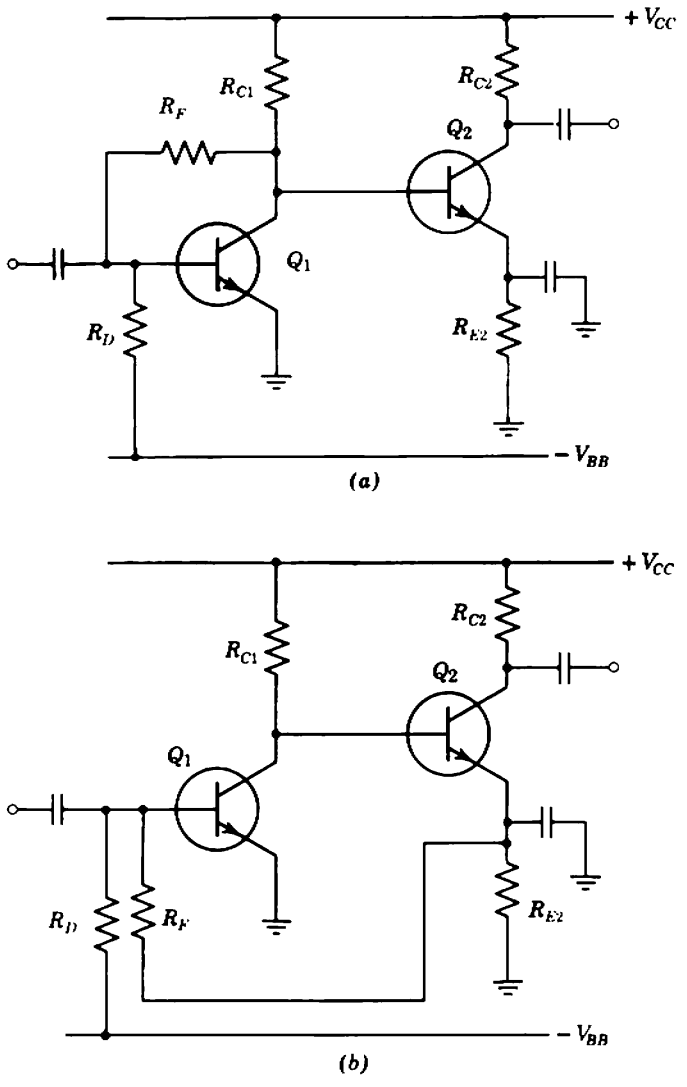


Fig. 6.43 Two-stage collector-feedback biasing: (*a*) the basic form; (*b*) a useful modification.

signal voltage at this point. There is therefore no signal feedback via R_F.

The design equations for the basic circuit are

$$I_{E1} \approx \frac{V_{CC} - V_{C1}}{R_{C1}}, \tag{6.78}$$

$$I_{E2} = \frac{V_{E2}}{R_{E2}} \approx \frac{V_{C1}}{R_{E2}}, \tag{6.79}$$

where

$$V_{C1} \approx V_{BB}\left(\frac{R_F}{R_D}\right). \tag{6.80}$$

The corrections required are the following:

(i) The current in R_D, and the base current of Q_2 are omitted from Eq. 6.78.
(ii) The base-emitter voltage drop is omitted from Eq. 6.79.
(iii) All terms involving β_N, V_{BE}, and I_{CO} of Q_1, and the base current of Q_2 are omitted from Eq. 6.80; the latter is usually an insignificant correction.

For the modified circuit

$$I_{E1} \approx \frac{V_{CC} - V_{C1}}{R_{C1}} \approx \frac{V_{CC} - V_{E2}}{R_{C1}}, \tag{6.81}$$

$$I_{E2} \approx \frac{V_{E2}}{R_{E2}}, \tag{6.82}$$

where

$$V_{E2} \approx V_{BB}\left(\frac{R_F}{R_D}\right). \tag{6.83}$$

There are many corrections for the transistor parameters.

Two other modified forms result from using self biasing (Section 6.3.1.1) and modified self biasing (Section 6.3.2.1) for the first stage, and two more forms are derived from these by connecting R_F to the emitter of Q_2 rather than the collector of Q_1. These circuits are not satisfactory, because the stability of self biasing circuits is poor. Such large variations occur in the collector voltage of the first stage that the emitter current of the second stage can increase to the point at which Q_2 "bottoms." Worthwhile modifications of Figs. 6.43*a* and *b* result from returning R_{E2} to the $-V_{BB}$ supply. The stability of the second stage is improved at the cost of increased power wastage in R_{E2}.

6.5.2.3 Two-Stage Combined-Feedback Biasing

Figure 6.44 shows the basic and modified circuits of two-stage combined-feedback biasing. The relation between these two circuits is the same as between Figs. 6.43*a* and *b*. The design equations for the basic circuit are

$$I_{E1} = \frac{V_{E1}}{R_{E1}} \approx \frac{V_{C1}}{R_{E1}}\left(\frac{R_D}{R_F + R_D}\right), \tag{6.84}$$

$$I_{E2} = \frac{V_{E2}}{R_{E2}} \approx \frac{V_{C1}}{R_{E2}}, \tag{6.85}$$

where

$$V_{C1} \approx V_{CC}\left[\frac{R_F + R_D}{R_{C1}(1 + R_D/R_{E1}) + R_F + R_D}\right]. \tag{6.86}$$

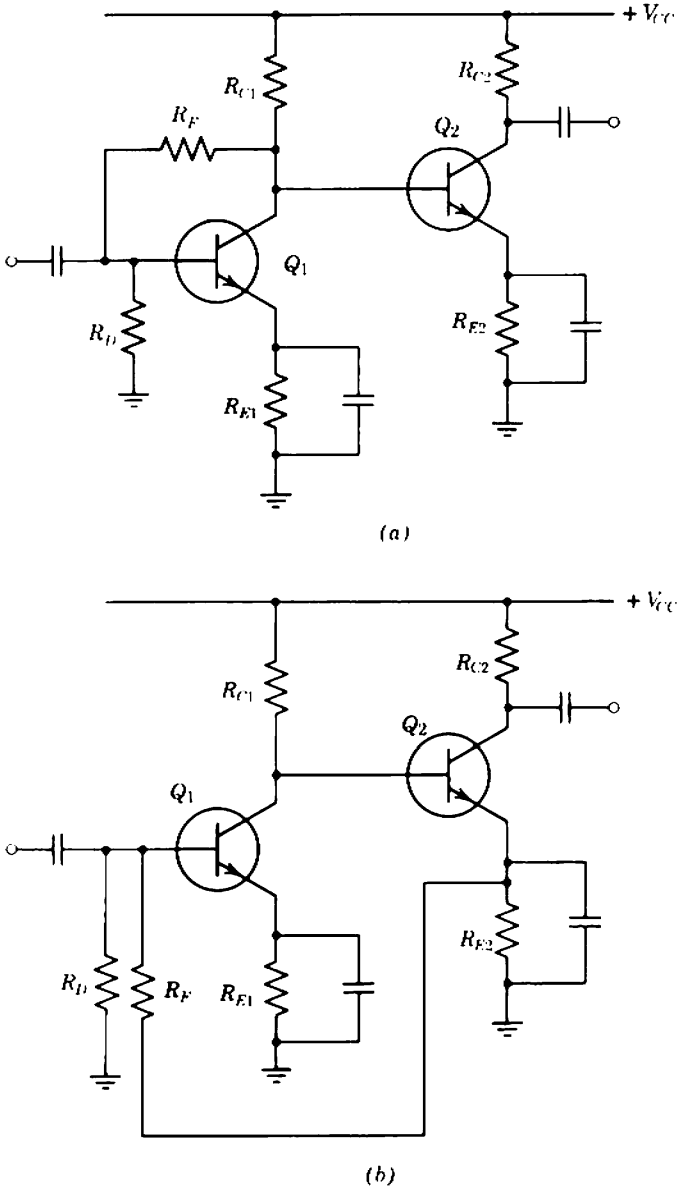


Fig. 6.44 Two-stage combined-feedback biasing: (*a*) the basic form; (*b*) a useful modification.

For the modified circuit,

$$I_{E1} = \frac{V_{E1}}{R_{E1}} \approx \frac{V_{E2}}{R_{E1}} \left(\frac{R_D}{R_F + R_D} \right), \tag{6.87}$$

$$I_{E2} \approx \frac{V_{E2}}{R_{E2}}, \tag{6.88}$$

where

$$V_{E2} \approx V_{CC}\left[\frac{R_F + R_D}{R_{C1}(R_D/R_{E1}) + R_F + R_D}\right]. \tag{6.89}$$

6.5.3 Two-Stage Vacuum-Tube Circuits

Vacuum-tube circuits exist corresponding to all the two-stage transistor biasing circuits except the self-biasing modifications of collector feedback; there is no vacuum-tube circuit corresponding to the single-stage transistor self biasing circuit (Section 6.3.1.1). There is, however, a vacuum-tube self biasing circuit (Section 6.1.2.2) and there are two screen-feedback circuits (Sections 6.3.3.2 and 6.3.3.4), from all of which two-stage direct-coupled groups can be derived.

A particularly useful vacuum-tube circuit, which has no transistor counterpart, is the cathode-to-screen feedback system shown in Fig. 6.45. Essentially, this is the modified two-stage combined-feedback biasing system, with the feedback applied to the screen rather than the grid. The circuit is most easily designed by considering the first stage as self biased and assuming that the screen is fixed at its nominal voltage. With the feedback to the screen, the quiescent point is stabilized by about one order of magnitude. The components R_S and C_S are sometimes omitted; R_S is always made so small that it has negligible effect on the quiescent conditions.

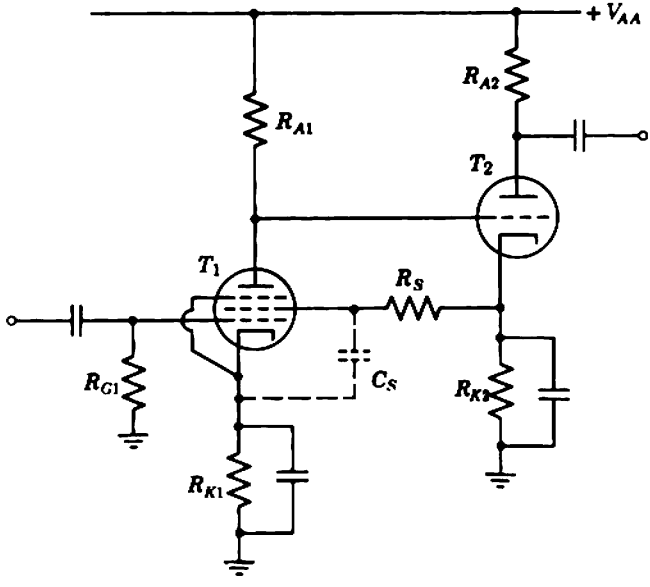


Fig. 6.45 Cathode-to-screen feedback biasing.

6.6 MISCELLANEOUS TOPICS*

Biasing-circuit design offers much scope for originality and circuit "cunning." A few useful tricks, and some other miscellaneous topics are discussed briefly in the following sections.

6.6.1 Long Direct-Coupled Groups

Occasionally, it is an advantage to direct couple a group of more than two amplifier stages, in order to improve the low-frequency response. As a general rule, long direct-coupled groups are not designable if the quiescent point of the first stage is determined only by its local biasing network; an accurate knowledge of device characteristics is required, and resistors of precise (and often unusual) values must be used. A long direct-coupled group can be made designable, however, by using a quiescent voltage in some later stage of the group as the biasing supply for the first stage. Figure 6.46 shows a circuit derived from Fig. 6.41 in Section 6.5.1; the 20.0 kΩ and 45.5 kΩ biasing resistors are chosen to give 3.2 V at

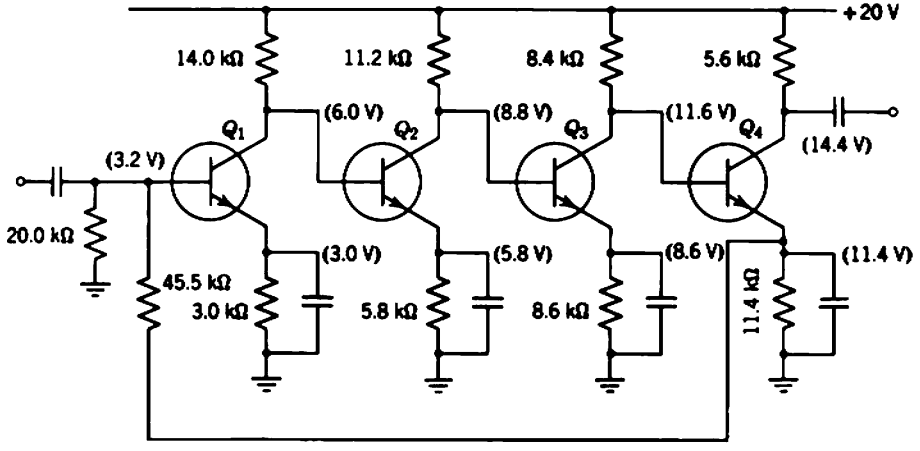


Fig. 6.46 Long direct-coupled group with over-all feedback biasing. The quiescent conditions for each transistor are

$$I_k = 1 \text{ mA}, \qquad V_{CE} = 3 \text{ V},$$

assuming $I_B = 20\,\mu\text{A}$. Nonpreferred resistor values are used to simplify the arithmetic.

* Section 6.6 may be omitted on a first reading.

the base of Q_1, assuming the base current is 20 μA. If the quiescent point of Q_1 shifts for any reason, the shift is transmitted through the amplifier, and the voltage at the emitter of Q_4 (which is the biasing supply for Q_1) moves in the direction to counteract the shift. Fundamentally, the biasing system is a feedback loop. The stability is so much improved compared with Fig. 6.41 that the nearest 10% preferred-value resistors could probably be used, although some adjustment might be necessary to secure optimum performance. It is esssential that these long groups be designed so that the change in biasing supply voltage opposes the shift in the first stage rather than aids it; that is, the feedback must be negative rather than positive.

Other disadvantages of long direct-coupled groups discussed in Section 6.5.1 are the power wastage in the supply resistors, and the high quiescent voltage to ground at the emitting electrodes of some devices. Figure 6.47 shows a transistor circuit that eliminates both disadvantages; the vacuum-tube counterpart exists also. Voltage dividers are used as coupling circuits between stages, so the emitter of each stage except the last can be grounded. The emitter of the last stage is near ground potential, and biasing feedback is taken from this point to the base of the first stage. One minor drawback is that the coupling dividers attenuate the signals and reduce the over-all gain. However, the attenuation is less than 2:1 with realistic values for $-V_{BB}$, and this is not too serious a loss. The capacitors in the upper arms of the dividers compensate for the input

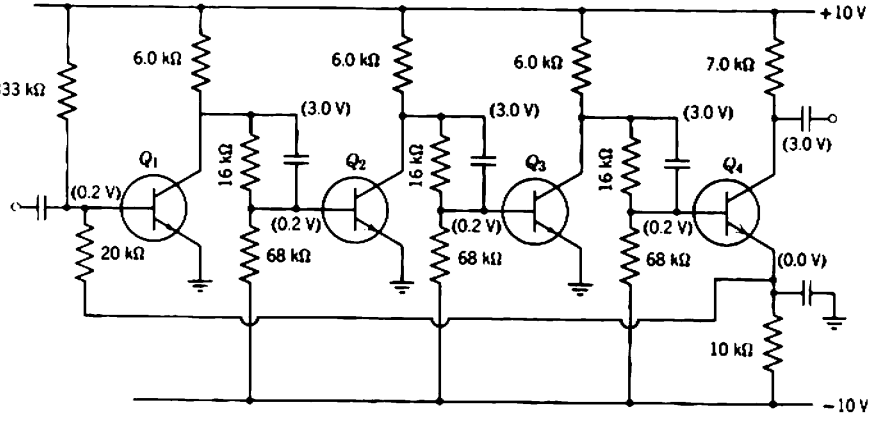


Fig. 6.47 Long direct-coupled group with interstage voltage dividers. The quiescent conditions for each transistor are

$$I_E = 1 \text{ mA}, \qquad V_{CE} = 3 \text{ V},$$

assuming $I_B = 20\,\mu\text{A}$. Nonpreferred resistor values are used to simplify the arithmetic.

capacitance of the following transistor, so that the attenuation ratio is constant for all signal frequencies.

6.6.2 Constant-Voltage Devices

A number of circuit tricks make use of the relatively constant voltage drop across a glow discharge tube or Zener diode, or even across a forward-biased silicon diode. Readily available glow discharge tubes operate with currents from 100 μA up to 50 mA, at drops between 50 and 150 V. Small Zener diodes operate over the same current range, at drops between 3 V and 20 V. Larger drops can be obtained by connecting devices in series. Manufacturers' data sheets give the slope resistance as a function of operating current; the minimum resistance (which occurs at an optimum current) may be as small as a few ohms. The temperature coefficient, and long- and short-term stability of the operating voltage are specified also.

The use of constant-voltage devices to derive auxiliary low-impedance dc voltage supplies from the main supply rails is obvious, and will not be discussed here. Less obvious is the direct application of these devices to biasing circuits. Perhaps the most useful arrangement is the transistor circuit shown in Fig. 6.48, or its vacuum-tube counterpart. This circuit is nothing more than a collector-feedback circuit (Fig. 6.30) rearranged so that only one supply voltage is required; Table 6.3 lists the relations

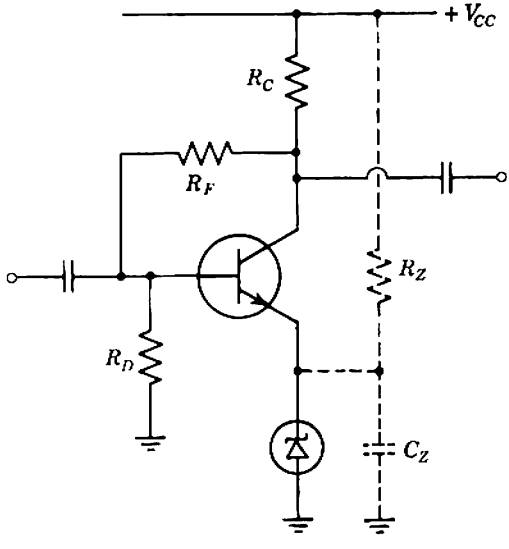


Fig. 6.48 Modified collector-feedback biasing circuit that operates from a single supply.

Table 6.3 Relations between Figs. 6.30 and 6.48

Parameter in Fig. 6.30	Parameter in Fig. 6.48
$V_{CC} + V_{BB}$	V_{CC}
V_{BB}	V_{Zener}

between Figs. 6.30 and 6.48. Resistor R_Z (shown broken in Fig. 6.48) may be required to increase the current through the Zener diode to a point at which its slope resistance can be neglected. Capacitor C_Z may be required to suppress the noise of the Zener diode.

Another use for constant-voltage devices is as coupling elements in long groups of direct-coupled stages. Figure 6.49 shows a transistor circuit; the vacuum-tube counterpart exists. One of the most convenient constant-voltage devices for the vacuum-tube circuit is a neon pilot lamp. These operate satisfactorily at currents as low as 100 μA, and the drop across them is about 60 V. Compared with the resistive dividers in Fig. 6.47, the constant-voltage divider has the advantage of no signal attenuation. Capacitors are usually connected across Zener diodes to suppress their noise, but should not be connected directly across any form of glow discharge tube as relaxation oscillation may occur; Fig. 6.50 shows a suitable arrangement of damping resistors.

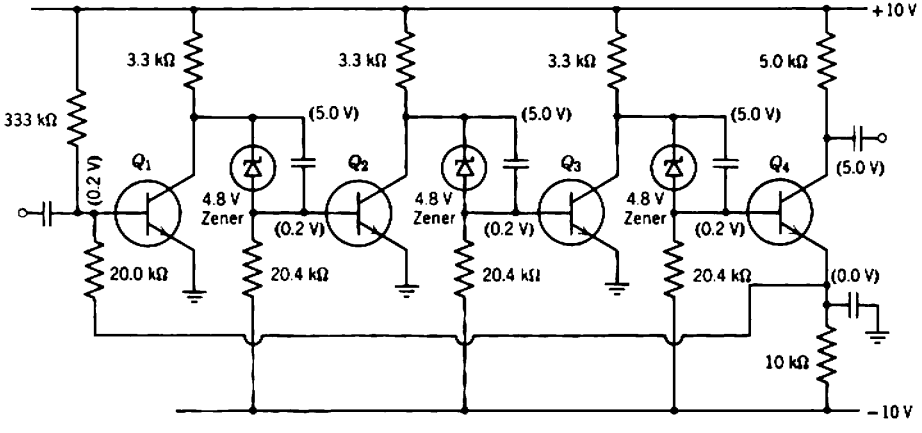


Fig. 6.49 Long direct-coupled group with constant-voltage coupling devices. The quiescent conditions for each transistor are

$$I_E = 1 \text{ mA}, \qquad V_{CE} = 5 \text{ V},$$

assuming $I_B = 20$ μA. The additional bleed current through the Zener diodes is 500 μA. Nonpreferred resistor values are used to simplify the arithmetic.

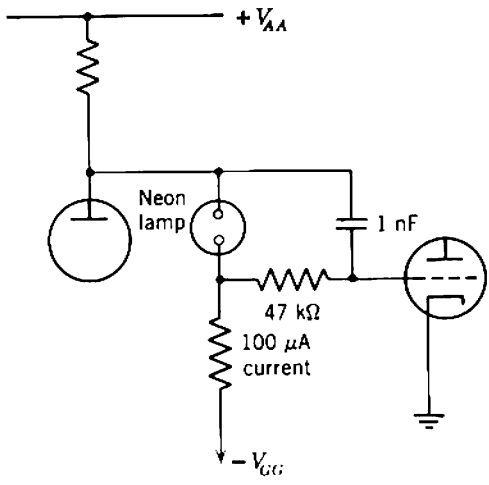


Fig. 6.50 Damping resistors necessary when a glow discharge tube is used as a coupling element.

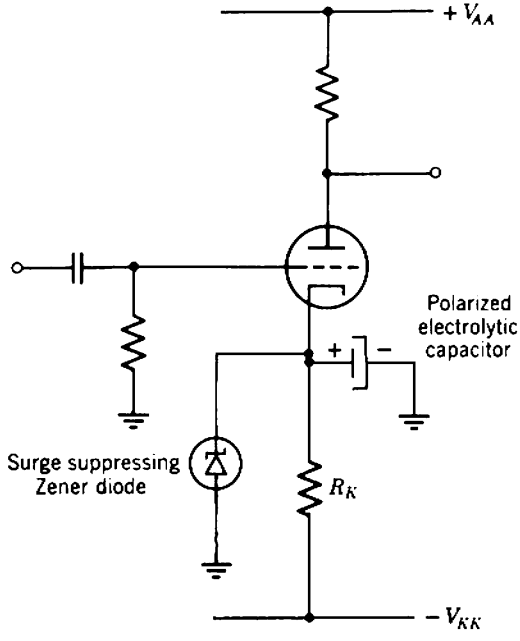


Fig. 6.51 Example of surge suppression with a Zener diode.

Zener diodes are useful for suppressing voltage surges. When an item of electronic equipment is first turned on, abnormally high voltages occur at some points until all capacitors attain their equilibrium charge, or until all vacuum-tube cathodes attain their operating temperature. Abnormal voltages can also occur when an amplifier is grossly overloaded by applying a very large input signal to it. The components most susceptible to damage by these voltage surges are the high-capacitance low-working-voltage bypass capacitors in the cathode and emitter circuits. These capacitors are usually polarized electrolytics, which are destroyed by small overvoltages or any reverse voltage. Figure 6.51 illustrates how a Zener diode can protect the bypass capacitor in a cathode-feedback biasing system. The diode is chosen to have a Zener breakdown voltage about 50% greater than the nominal quiescent voltage at the cathode. Under normal circumstances, therefore, the diode does not conduct. However, in the event of a positive-going surge the diode breaks down and limits the forward voltage across the capacitor. Before the cathode reaches its operating temperature, or in the event of a negative-going surge, the diode conducts in its forward direction, and protects the capacitor from a reverse voltage.

6.6.3 *n-p-n* and *p-n-p* Transistors in Combination

An advantage of transistors over tubes is that devices of both polarities exist, that is, both *n-p-n* and *p-n-p* transistors. The alternation of the

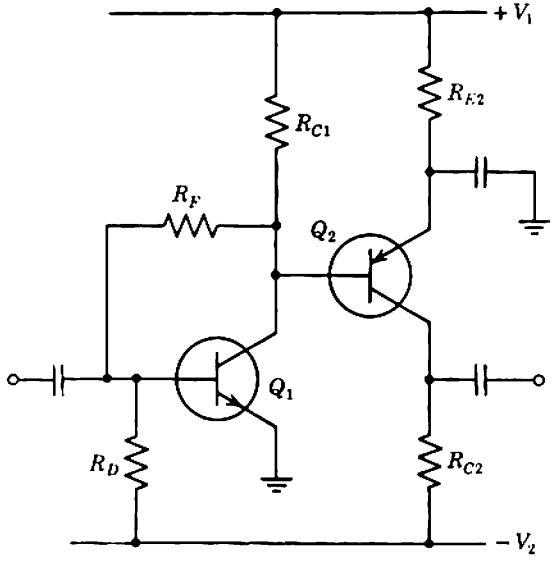


Fig. 6.52 *n-p-n* and *p-n-p* transistors in combination: a particularly useful two-stage group.

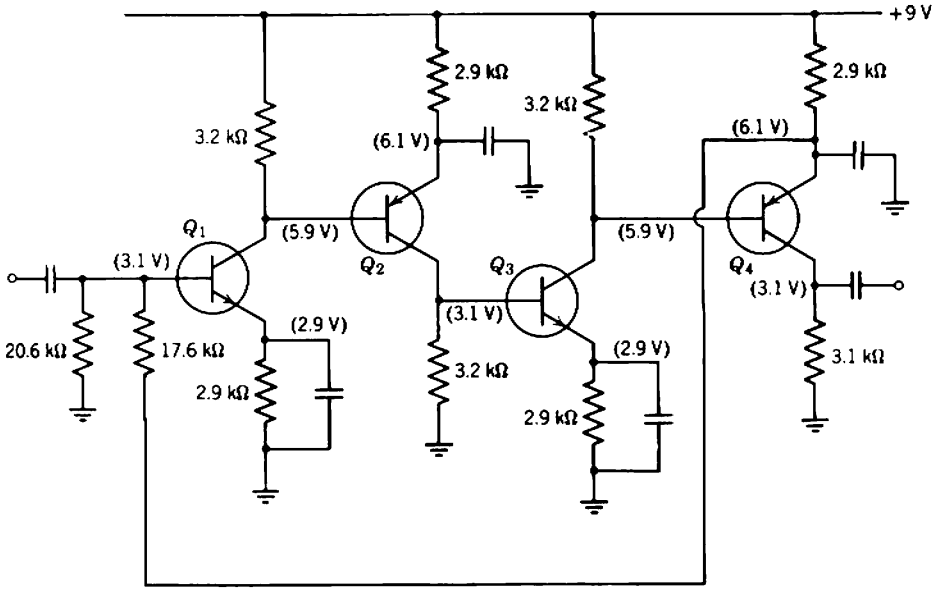


Fig. 6.53 *n-p-n* and *p-n-p* transistors in combination: a long direct-coupled group. The quiescent conditions for each transistor are

$$I_E = 1 \text{ mA}, \qquad V_{CE} = 3 \text{ V},$$

assuming $I_B = 20\,\mu\text{A}$. Nonpreferred resistor values are used to simplify the arithmetic.

two types in direct-coupled groups of stages prevents the dc voltage drops from accumulating and, thereby, reduces power wastage in supply resistors. Combinations of *n-p-n* and *p-n-p* transistors can be used with advantage in all the two-stage circuits of Section 6.5.2 and in the long groups of Section 6.6.1. Figure 6.52 shows a two-stage group that is particularly useful in feedback amplifiers, and Fig. 6.53 shows a group of four stages.

6.6.4 $R_{G(\text{max})}$ for Vacuum Tubes

One of the ratings for a vacuum tube is the maximum permissible supply resistor R_G between the grid and its bias source. This rating is not an absolute rating, and although not stated by many manufacturers the maximum permissible value of R_G depends on the type of biasing system.

As shown in Section 3.1.1.4, all vacuum tubes have a finite grid current. Usually the grid is biased more negative than the free-grid potential; as a result, the grid current is predominantly the positive ion component. This component increases with increasing tube temperature. The grid current flowing in R_G tends to make the grid less negative with respect to the

cathode and, therefore, tends to increase the cathode current and anode power dissipation. The larger the resistor, the greater is the change in grid voltage for a given change in grid current, and the greater is the increase in power dissipation. If R_G is large enough, the increase in power dissipation and tube temperature for a given increase in grid current can more than maintain the initial increase, and a regenerative condition is set up. One of the following two ends can result:

(i) In extreme cases, the power dissipation rises so far that the tube is destroyed.
(ii) The negative grid-to-cathode voltage falls to the point at which the grid attracts electrons from the cathode and the direction of the grid current reverses. The input admittance of the tube rises and loads the preceding stage, causing gross distortion of the signal.

The maximum allowed value of R_G is one for which the regenerative increase in dissipation is guaranteed not to occur.

Tube manufacturers normally specify the value of $R_{G(\max)}$ that applies for a tube operating at full rated dissipation in a self biasing circuit. This value can be exceeded, provided either:

(i) the tube is operating below maximum dissipation
(ii) the biasing system is more stable than the simple self biasing circuit.

A safe rule is that $R_{G(\max)}$ can be increased in inverse proportion to the total power dissipation (including the cathode heating power), and in direct proportion to the stability of the biasing system. The relative stability of two biasing systems is the inverse ratio of the changes in quiescent current resulting from the same change in device characteristics. For cathode-feedback and self biasing circuits, the relative stability is approximately the ratio of the voltage drops across R_K.

6.6.4.1 Grid-Leak Biasing

In a class of circuit known as grid-leak biasing $R_{G(\max)}$ has no significance. The grid is returned to the cathode through a resistor of the order of 10 MΩ, with no source of negative grid-to-cathode voltage. However, the grid takes up a voltage somewhat less negative than the free-grid potential, at which the current is due mainly to electrons striking the grid.

Grid-leak biasing is of very limited application. First, because it depends on the existence of a negative grid current (rather than a positive ion current), the temperature of the tube must be so low that no significant gas ion current is developed. Grid-leak biasing is therefore restricted to low-power applications. Secondly, grid-leak biasing is so imprecise that

it cannot be used except in the most noncritical applications. Grid-leak biasing is normally restricted to low-level audio-amplifier stages employing resistance-capacitance coupled triodes, because their gain is relatively independent of quiescent point and the power dissipation cannot rise to a destructive value.

6.6.5 Biasing Circuits for F.e.t.s

With two minor exceptions, biasing circuits for vacuum triodes and field-effect transistors follow the same principles. First, the sudden onset of input (grid) current for a tube occurs when the grid is slightly negative with respect to the cathode. In contrast, the onset of input (gate) current for an *n*-channel junction f.e.t. occurs when the gate is slightly positive with respect to the source, while there is no sudden onset of current in an insulated-gate f.e.t. Indeed, enhancement-mode i.g.f.e.t.s operate only when the gate is positive. More latitude is therefore available in f.e.t. circuits. Secondly, both *n*-channel and *p*-channel f.e.t.s exist, so there is a whole family of circuits corresponding to the *n-p-n* and *p-n-p* transistor combinations of Section 6.6.3.

6.6.6 Biasing Methods for Integrated Circuits

Biasing methods for integrated circuits tend to be dominated by the fact that large capacitances are difficult to integrate. Emitter-feedback biasing systems, which require large emitter bypass capacitors, are therefore unsatisfactory. Collector-feedback biasing, however, is extensively used.

A novel biasing method for integrated circuits is based on the voltage-feed fixed biasing circuit of Section 6.1.1.2, with a second transistor used as the temperature-dependent element (Fig. 6.54). The circuit depends for its operation on the fact that when two transistors are fabricated together in the same chip of silicon, their parameter values tend to be nearly equal and to track with variations in temperature. Moreover, the two transistors are thermally bonded together so that their temperatures are nearly equal. Thus the parameter values remain equal despite variations in ambient temperature and inequalities in power dissipation.

In Fig. 6.54 the bases of Q_1 and Q_2 are fed through equal resistors R_{B1} and R_{B2} from the same point, the collector of Q_1. Therefore the collector currents I_{C1} and I_{C2} are approximately equal. But I_{C1} is defined by the passive circuit elements as

$$I_{C1} \approx \frac{V_{CC}}{R_{C1}};$$

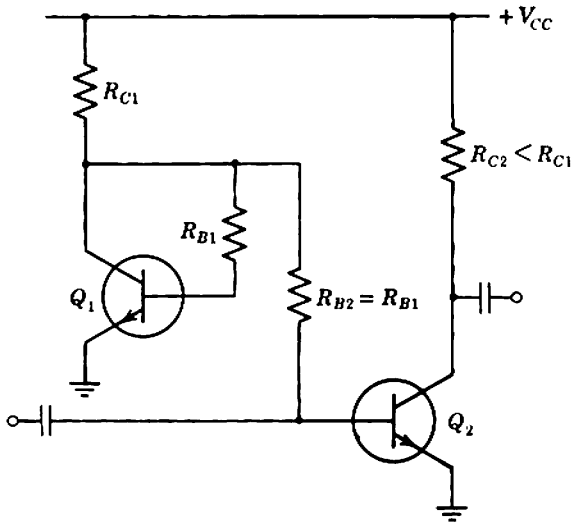


Fig. 6.54 Voltage-feed fixed biasing circuit suitable for integrated amplifiers.

Q_1 operates self biased and nearly (but not quite) bottomed. Therefore, if the transistors are identical, I_{C2} must be close to V_{CC}/R_{C1}, independent of ambient temperature. However, if the transistors have different emitter-junction areas, there is a predictable ratio between the emitter currents.

Chapter 7

Simple Practical Amplifiers I: Small-Signal Dynamic Response

Chapter 5 discusses the elementary circuit theory of amplifiers. Two simplifying assumptions are that

(i) the signal levels are small, so that the incremental elements in the equivalent circuits of the devices can be assumed constant;
(ii) the signal frequencies are in the mid-band range for which reactive elements in the equivalent circuit have no effect on the operation.

The purpose of this and the two succeeding chapters is to extend the elementary theory to the point at which it provides a basis for designing practical amplifiers. The major extension in this chapter is the consideration of the small-signal dynamic response when the signals have time rates of change for which reactive elements are important. In addition, there are brief discussions of the rather intangible approximation problem, and of the types of source and load that are commonly encountered.

7.1 THE APPROXIMATION PROBLEM

The primary aim of an amplifier is to increase the level of a signal without appreciable loss of signal information. Ideally, an amplifier should preserve a specified functional relation between the input and output signals. Often it is desired that the output should be a perfect replica of the input at a higher power level; in more specialized cases the output is required to differ from the input in some precisely defined way. The approximation problem results from the fact that no amplifier is perfect so that the ideal

response can never be achieved. Any practical amplifier corrupts the signal to some extent. The approximation problem is to decide what physically realizable transfer function is satisfactory for a given application.

The two main types of imperfection in an amplifier are introduced in Section 1.2. The first is the deformation of all time waveforms, that is, the limitation of restricted dynamic response, which is discussed in this chapter. The second is the appearance of signal components at the output that are not present in the input, that is, noise and distortion, which are discussed in Chapters 8 and 9. It is possible to reduce the degree of imperfection in an amplifier at the expense of increased complexity and cost. The problem facing the design engineer is to weigh the cost of improving an amplifier against the resultant improvement to the over-all system. In making a decision, it is necessary to know exactly what is expected of the complete system of which the amplifier is a part, and what the amplifier is expected to do in the system. In electronics as elsewhere no chain is stronger than its weakest link. There is therefore no advantage in designing an amplifier that is very much better than the rest of the system, but there is the disadvantage of the increased cost. Sometimes the amplifier itself is the weakest link; in such cases the problem is to decide what level of performance is tolerable and how much additional cost is worthwhile in order to achieve or better this minimum acceptable level. Experience, subjective tests, standards or specifications, measurements, and information theory may all influence the final decision.

As an illustration of the approximation problem, Table 7.1 lists the acceptable performance for various systems and the probable cause of the limitation in each case. Thus, in an excellent audio system there is no sense in extending the range of the amplifier beyond about 10 Hz to 40 kHz. Its response would be within a few percent at the over-all pass-band limits (20 Hz to 20 kHz), which is quite an insignificant variation compared with the imperfections in the microphone, tape, loudspeakers, and the room acoustics. The engineer who designs an amplifier with the object of extending the high-frequency response to 1 MHz is misguided in the extreme, and the engineer who proposes to build even a 50 kHz amplifier for use with a low-quality pickup and loudspeaker would be well advised to reconsider the distribution of his spending. This it not to imply that it is unwise to design an amplifier whose performance exceeds the system requirements; rather it is unwise to spend money with the sole purpose of exceeding requirements. Often a performance that is beyond requirements in one respect (say, frequency response) accrues as a by-product of some other aspect of the requirements (say, distortion).

Table 7.1 is not intended to be a complete specification for the various familiar systems. Although it is necessary for the systems to satisfy the

Table 7.1 Approximate Specifications for Familiar Systems

System	Frequency Range for Sinusoidal Signals	Distortion on Signal Peaks	Minimum Ratio Peak Signal to Noise
Telephone	250 Hz–3 kHz (microphone, earpiece, filters)	20% (microphone, earpiece)	40 dB (noise pick-up in plant)
Good AM radio receiver	150 Hz–5 kHz (loudspeaker principally; also the limited receiver pass-band)	10% (loudspeaker, circuit)	45 dB (static, mains-frequency hum)
Good phonograph	70 Hz–10 kHz (pick-up, loudspeaker)	5% (disk manufacture, pick-up, loudspeaker)	55 dB (disk manufacture, surface noise, hum and noise in amplifier)
Professional record/playback AF tape system	30 Hz–20 kHz (loudspeaker, room acoustics)	3% (tape nonlinearity, loudspeaker)	65 dB (tape noise)
Picture channel, average domestic TV receiver	20 Hz–3 MHz (limited receiver pass-band)	10% (video output stage)	30 dB (noise in first stage)
Picture channel, excellent domestic TV receiver	dc–5 MHz (channel width, C.C.I.R. standard)	5% (picture tube γ)	35 dB (noise in first stage)
Professional closed-circuit TV	dc–8 MHz (camera and picture tube resolution)	5% (camera and picture tube γ)	40 dB (noise in first stage of camera amplifier)

requirements listed, these requirements by themselves are not sufficient to ensure satisfactory performance. In video systems, for example, it is necessary to specify limits to the overshoot produced on transient input signals.

The most important single property of many amplifiers is their operating frequency range, and it is quite common to classify amplifiers according to the range of signal frequencies for which they are useful. Subaudio-frequency signals are encountered principally in control or servo systems. The component frequencies may extend as low as zero (i.e., dc), and as high as a few hundreds or thousands of hertz. Audio-frequency signals occur in sound systems and in many other applications such as certain measuring instruments and low-carrier-frequency communications circuits.

The frequency range extends from a few hertz to several tens of kilohertz; sometimes the low-frequency limit is dc. As the upper frequency limit is raised audio signals merge into wide-band or video signals. There is no accepted line of demarcation, but it would be unusual to describe a signal whose highest component frequency is less than 100 kHz as "wide band," or to describe as "audio" a signal with more than 1 MHz bandwidth. The upper frequency limit for wide-band signals extends almost without limit—certainly to 100 MHz. Wide-band amplifiers are used in most measuring instruments such as oscilloscopes and electronic voltmeters, in television (hence the term video), radar, and many nucleonic instruments. The lower frequency limit ranges from dc or a few hertz in measuring instruments and television, to 1 or more MHz in certain nucleonic applications.

7.2 SOURCES AND LOADS

Low-pass amplifiers are used in a multitude of applications, and a great diversity of sources and loads is encountered. Sources and loads can be classified first as single-ended or balanced and second according to their impedance and the relative driving-point impedance of the amplifier to which they should be connected. The following sections discuss a number of practical aspects of sources and loads. Satisfactory justifications are not given for some of the statements made; the reader may rectify these intentional omissions.

7.2.1 Single-Ended and Balanced Terminal Arrangements

Most sources, amplifiers, and loads have *single-ended* (or unbalanced) terminals. Each terminal pair consists of a *common terminal* (also known as ground or "cold") that is at the universally accepted zero of potential, ground voltage, and a *signal terminal* (also known as active or "hot") that is at the signal voltage relative to ground. The common terminal is usually connected to the frame or chassis. Items of equipment with single-ended terminals may be interconnected with separate signal and common wires or with a twin-wire pair. Alternatively, they may be interconnected with some form of concentric cable for which the inner conductor is the signal lead and the outer shield or covering is the common.

Some sources, amplifiers, and loads have *balanced* terminal arrangements. Each terminal group consists of two active terminals with the option of a third ground terminal. The signal is the difference between the voltages of the two active terminals which, ideally, are at equal but opposite signal

voltages relative to ground. A *differential amplifier* is insensitive to any inphase or *common-mode signal* that is present simultaneously on both active terminals, and responds only to the difference or *differential-mode signal* between them. The advantage of a balanced transmission path is that it is less susceptible to interference than a single-ended transmission path. Ideally, any stray electric or magnetic field induces the same emf in both signal wires and these interfering signals cancel. Perfect cancellation requires equal pickup in the two wires so that balance of the physical layout is important. Twin-leads are more satisfactory than separate wires, and a twin-lead with a grounded shielding cover is better still. Balanced systems are more expensive than single-ended systems as they involve a second active wire and often require a more complicated amplifier. They are therefore reserved for applications where a low susceptibility to interference is important; microphone cables carrying low-level signals are a common example.

7.2.2 Thevenin and Norton Representations

According to Thevenin's and Norton's theorems, any source can be regarded either as a voltage generator with a series impedance or as a current generator with a shunt impedance. The two representations are equivalent as long as the values of all the elements are known. In most transducers, however, the generator in one form is more closely allied to the transducer physics and is therefore more reproducible. For example, the equivalent voltage generator for an electromagnetic transducer is directly related to the time rate of change of flux; the equivalent current generator involves the transducer impedance as well as the flux change and is therefore less reproducible. An amplifier should be designed to make use of the more reproducible equivalent generator within a source.

Loads make direct use of the voltage applied to them or of the current forced through them. For example, electrostatic transducers make use of the voltage, whereas electromagnetic transducers make use of the current. (Loudspeakers are an exception to the latter statement.) The amplifier for use with a particular load should be designed to supply the preferred type of signal.

Inside a multistage amplifier the output of one stage is the source for the following stage. The design should be such that each stage makes use of the more reproducible output signal type from the preceding stage.

These conclusions can be summarized thus:

1. A source whose internal voltage is the signal to be amplified should work into an amplifier whose input impedance is much greater than the

internal impedance of the source, and which has a stable voltage-input transfer function (namely, G_T or A_V depending on the output circuit). Because the input resistance is large, the output voltage from the source is very nearly equal to the internal voltage generator, and the amplifier makes use of this known voltage.

2. Similarly, a source whose internal current is the signal to be amplified should work into an amplifier whose input impedance is much less than the internal impedance of the source, and which has a stable current-input transfer function (namely, A_I or R_T).

3. A load that makes use of the applied voltage should be fed from an amplifier whose output impedance is much less than the load impedance, and which has a stable voltage-output transfer function (namely, A_V or R_T depending on the input circuit). Because the output resistance is small, the voltage output from the amplifier is very nearly equal to the internal voltage generator, and the load makes use of this known voltage.

4. Similarly, a load that makes use of the applied current should be fed from an amplifier whose output impedance is much higher than the load impedance, and which has a stable current-output transfer function (namely, G_T or A_I).

Transducers of the types listed above are collectively described as *mismatched*, because there should be a gross impedance mismatch between them and the amplifier. *Matched* sources and loads are encountered in some rather specialized applications, and these (as their name implies) should be matched by the amplifier driving-point impedances. Common examples are open-wire transmission lines, strip lines, coaxial cables, and various communications networks such as attenuators and filters. With few exceptions, matched sources and loads are not transducers so neither the Thevenin nor the Norton representation has any more justifiable physical basis. In addition, the characteristic impedance is purely resistive and the question of conjugate matching does not arise. The accuracy required of the matching depends on the application; often the nearest 10% preferred-value resistor is adequate, but it may be necessary to match precise attenuators and filters to 1% or better.

There are two common impedance ranges for matched laboratory equipment—500 to 600 Ω for open-wire lines and associated networks, and 50 to 100 Ω for coaxial cables and associated networks—although other values may be encountered in different applications. Matched subaudio- and audio-frequency equipment is often designed for 600 Ω impedance and is usually balanced. However, provision is made for grounding one side of most 600-Ω amplifiers, filters, and attenuators, so that they can be used in a single-ended fashion if desired. Matched wide-band circuits are

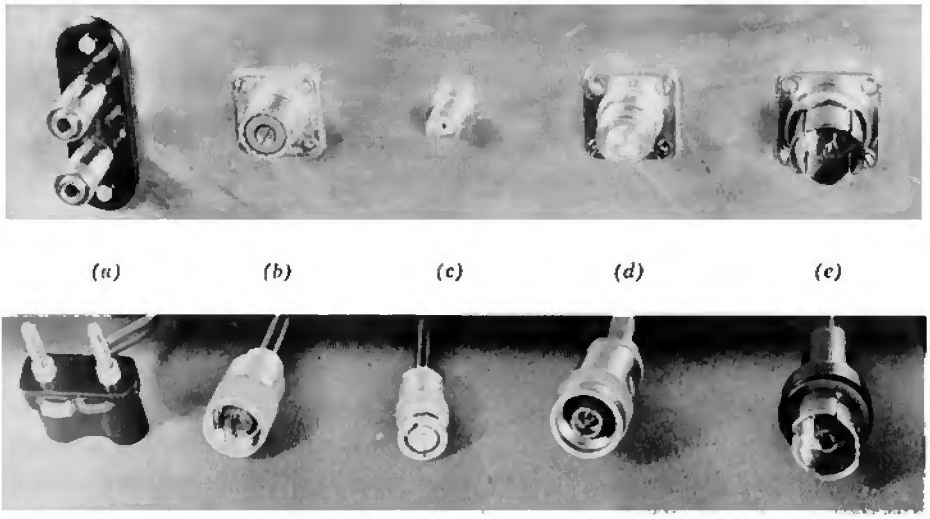


Fig. 7.1 Terminals and coaxial cable connectors: (*a*) G.R. terminals; (*b*) U.H.F. coaxial; (*c*) B.N.C. coaxial; (*d*) Type-N coaxial; (*e*) G.R. coaxial.

usually interconnected with coaxial cables; 50 Ω, 75 Ω, and 91 Ω are common impedance values. A vast number of plug and socket types are in common use and some are illustrated in Fig. 7.1.

7.2.2.1 Examples of Source and Load Transducers

There are a number of source transducers that have a low but imprecisely defined internal impedance, and for which the open-circuit output voltage is the signal to be amplified. Most transducers for servo systems, microphones, and phonograph pickups are of this class. These sources should work into an amplifier whose input impedance is very much higher than the internal impedance of the source, and whose stable and predictable transfer function is defined in terms of an input voltage. Practically any type of plug or terminal can be used with these approximate voltage sources at audio and subaudio frequencies. Coaxial plugs and sockets (without any terminating resistance) are often used at video frequencies.

For example, there are two common forms of phonograph pickup, the magnetic and the piezoelectric. Ideally, the open-circuit output voltage of either is directly related to the recorded signal; both have reactive output impedances. A magnetic pickup consists of a coil of finite winding resistance and self inductance; this coil moves in a magnetic field and the induced emf is proportional to the stylus velocity. Figure 7.2*a* shows the equivalent circuit. A piezoelectric pickup has a capacitive output impedance and an open-circuit voltage proportional to stylus displacement;

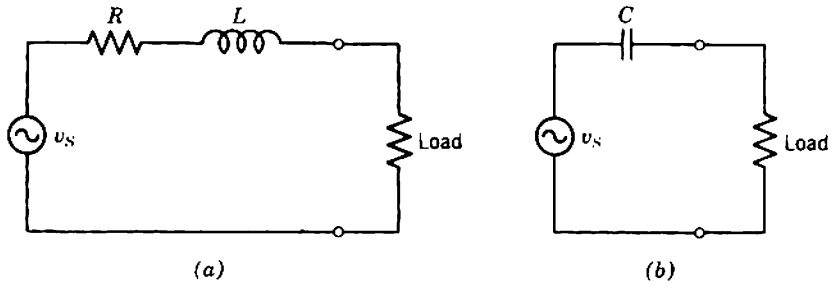


Fig. 7.2 Equivalent circuits for phonograph pickups: (*a*) magnetic; (*b*) piezoelectric.

the equivalent circuit is shown in Fig. 7.2*b*. If either pickup is loaded by a resistance that is high compared with the internal impedance over the audio-frequency range, the output voltage is independent of the recorded frequency and depends only on the appropriate velocity or displacement. However, if the load resistance is comparable to or less than the internal impedance, the output voltage is a function of frequency because of the reactive nature of the output impedance. Typical values for the winding resistance and self inductance of a magnetic pickup are 1000 Ω and 0.2 H; the load impedance should therefore be 25 kΩ or more if the output voltage is to be independent of frequency up to 20 kHz. The output capacitance of a crystal pickup is about 2 nF; the load resistance should therefore be 1.6 MΩ or more to maintain the output voltage down to 50 Hz.

Measuring instruments such as oscilloscopes and electronic voltmeters are usually designed to have a very high input impedance, so that they do not significantly load down the circuit under test. The input impedance of the instrument should be so high that, in comparison, the circuit under test behaves as a voltage source. Most test instruments have an input resistance of at least 1 MΩ and values as high as 10 MΩ are not uncommon. The shunt input capacitance ranges from about 100 to 10 pF or even less. Instruments for special purposes may have input impedances differing by orders of magnitude from these figures. Even at moderate frequencies the capacitive susceptance dominates the input admittance, and it follows that a test instrument disturbs the circuit more at high frequencies than at low frequencies. Audio-frequency instruments are usually fitted with G.R. terminals; video-frequency instruments or instruments that can be fitted with high-impedance probes usually have coaxial sockets. In either case the terminating resistor is omitted so that the input impedance is as high as possible. Occasionally a switch is provided so that the termination can be included if the instrument is to be connected to a matched source.

Some matched sources are occasionally treated as voltage sources and

are loaded with an impedance that is high compared with their characteristic impedance. Matched oscillators may be used this way, and their output voltage rises by a factor of two. Cables and transmission lines may be loaded by a high impedance only if there is a matching termination at the sending end and if the substantial wave reflected from the receiving end can be tolerated. It is never permissible to mismatch the output of a calibrated attenuator or filter as the performance will be radically changed and the calibration rendered meaningless.

Other sources have a high but imprecisely defined internal impedance, and their short-circuit current is the signal to be amplified. Such sources should work into an amplifier whose input impedance is much lower than the internal impedance of the source and which has a stable transfer function defined in terms of a current input. Sources of this type are usually associated with optical or particle transducers; examples are photoelectric cells, television cameras, and junction detectors and photomultipliers in nucleonic applications. The vidicon camera tube is one such source; it has an output resistance of several megohms in parallel with a few picofarads stray capacitance, and its short-circuit output current is directly related to the incident light intensity. Figure 7.3 shows the equivalent circuit. When a vidicon is connected to an amplifier whose input impedance is other than zero, the signal current into the amplifier falls at high frequencies. Typically, the output impedance of a vidicon is 10 MΩ shunted by 20 pF; if the input resistance of the amplifier is 10 kΩ, the input current to the amplifier is down 3 dB at 0.8 MHz. In addition it is shown in Section 8.5.3.2 that the load capacitance for these approximate current sources should be kept as small as possible in the interests of signal-to-noise ratio. Therefore, the input stage of the amplifier is often mounted as close to the source as possible in order to reduce the stray wiring capacitance. The output lead from the source is often soldered directly into the amplifier circuit to eliminate the capacitance of a plug and socket.

Many load transducers should be fed from an approximation to a voltage source, that is, an amplifier whose output impedance is much less than the

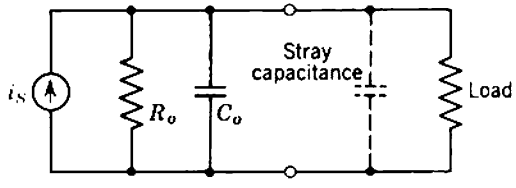


Fig. 7.3 Equivalent circuit for vidicon television camera tube.

load impedance and whose stable transfer function is defined in terms of an output voltage. Common examples are moving-coil loudspeakers, the intensity grid of a television picture tube, cathode-ray-tube deflection plates, piezoelectric transducers, and most nuclear counting circuits. A wide range of load impedances is encountered. Loudspeakers have an impedance of a few ohms; the deflecting plates in a cathode-ray tube or the grid of a television picture tube have an input capacitance of a few picofarads with almost infinite shunting resistance; piezoelectric transducers have a capacitive input impedance roughly of the order of 5 nF. A great variety of plugs and terminals is used for connecting such loads to an amplifier.

Some matched loads may be fed from voltage sources. Voltage feed is satisfactory for transmission lines and coaxial cables provided that they are matched at their receiving end and that reflections from the sending end can be tolerated. Attenuators retain their calibration when fed from a voltage source provided their load impedance is correct, but filter networks should not in general be fed from voltage sources.

Loads that should be fed from approximate current sources occur less frequently. Magnetic and electrochemical transducers are common examples. Amplifiers for use with such loads should have an output impedance that is high compared with the load resistance, and should have a stable transfer function defined in terms of an output current. Transmission lines and cables may be fed from current sources provided that they are matched at their receiving end and that single-ended termination is satisfactory. Attenuators, but not filters, may be fed from current sources.

Motors in servo systems may be fed from either current or voltage sources depending on the characteristics required. Field windings may be fed from either without change to the steady-state response, but current feed gives slightly faster transient response. Armatures fed from a voltage source have a constant velocity characteristic, whereas current feed gives constant acceleration.

7.2.3 Realization of Transfer Functions and Driving-Point Impedances

There are two fundamentally different approaches to designing an amplifier for use between a specified source and load. The direct approach illustrated in Fig. 7.4 is to design the amplifier so that its stable transfer function and driving-point impedances are as required. This is the more satisfactory approach for most applications. Its only fundamental disadvantage is that the driving-point impedances are not accurately defined; this applies particularly to transistor circuits, where the impedances are likely to differ by as much as a factor of two from the expected value

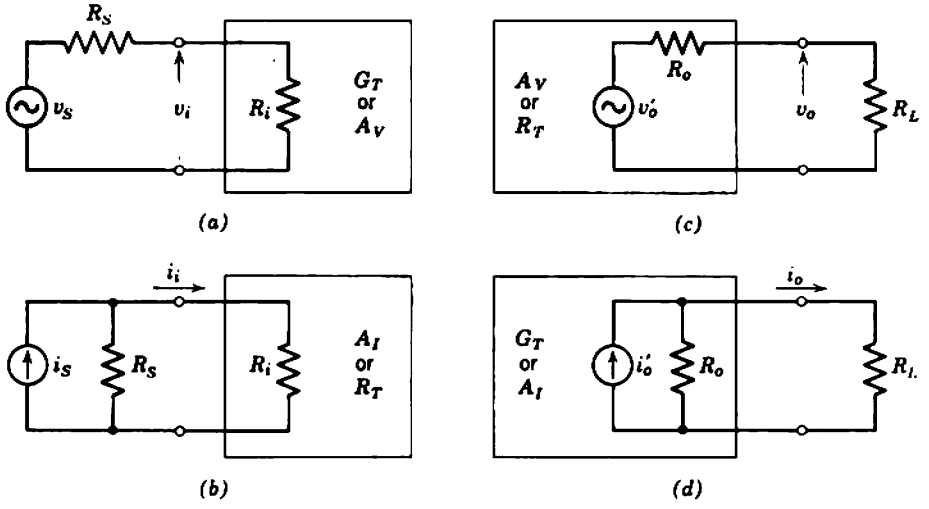


Fig. 7.4 Treatment of mismatched sources and loads—direct approach:

(*a*) Source voltage to be amplified.
If $R_S \ll R_i$, then $v_i \approx v_S$.

(*b*) Source current to be amplified.
If $R_S \gg R_i$, then $i_i \approx i_S$.

(*c*) Voltage applied to load.
If $R_o \ll R_L$, then $v_o \approx v_o'$.

(*d*) Current applied to load.
If $R_o \gg R_L$, then $i_o \approx i_o'$.

because of the large tolerance on β_N. This approach is therefore unsuitable for matched sources and loads, but is excellent for the more frequent case of mismatched sources and loads. For example, the voltage that a low-impedance source delivers into a high-input-impedance amplifier is almost independent of both impedances provided their ratio is large, and is almost equal to the open-circuit source voltage. Further, the feedback techniques described in Chapter 10 for raising the input impedance of an amplifier tend to stabilize the voltage-input transfer functions (namely, G_T or A_V) and make use of the known input voltage. Similar comments apply to the current that a high-impedance source delivers to a low-input-impedance amplifier, and to the various combinations of amplifier output impedance and load impedance.

The indirect approach illustrated in Fig. 7.5 is to design the amplifier so that its stable transfer function and driving-point impedances are opposite to those required, and then add passive elements at the input and

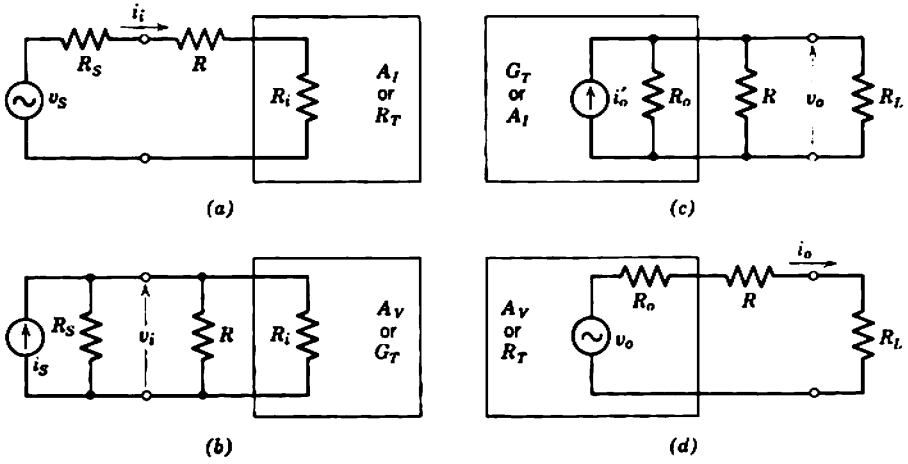


Fig. 7.5 Treatment of mismatched sources and loads – indirect approach:

(*a*) Source voltage to be amplified.
If $R \gg (R_S + R_i)$, then $i_i \approx v_S/R$.

(*b*) Source current to be amplified.
If $R \ll (R_S \parallel R_i)$, then $v_i \approx i_S R$.

(*c*) Voltage applied to load.
If $R \ll (R_o \parallel R_L)$, then $v_o \approx i_o R$.

(*d*) Current applied to load.
If $R \gg (R_o + R_L)$, then $i_o \approx v_o'/R$.

output to give the required functions. For example, if a high-input-resistance amplifier with stable voltage-input transfer function is required, a current-input amplifier with low input resistance is designed and a large resistance is connected in series with the input. This resistor converts the output voltage of the source into a current drive for the amplifier, and the total input resistance is the sum of the series resistor and the input resistance of the amplifier. Similarly, if a voltage-output amplifier with low output resistance is required, a current-output amplifier with high output resistance is designed and a small resistor is connected in shunt with its output terminals. The advantage of this approach is that it defines the driving-point impedances accurately, particularly if the ratio of the driving-point impedance of the amplifier proper to the desired value is very great. The relatively unknown impedance of the amplifier makes a small contribution to the total impedance compared with the accurately known contribution of the passive elements. Matched amplifiers are often designed in this way because cables, attenuators, and filters require an

accurate termination. The approach has quite serious disadvantages, however, and is not recommended for most applications. At the input the impedance-changing resistor dissipates some of the signal power available from the source and, therefore, reduces the ratio of the signal to the noise generated within the amplifier (see Section 8.5.2.1). An impedance-changing resistor at the output absorbs some of the output power, so that a larger and more expensive amplifier is required to give the same signal power in the load.

Although the above discussion is in terms of an amplifier and its external source and load, it applies also to the coupling networks between the stages in an amplifier. Any stage in an amplifier may be regarded as the source for the following stage and the load for the preceding stage.

No method is given in Chapter 5 for designing an amplifier that has a stable current-input transfer function. Rather, it is pointed out that A_I and R_T are almost meaningless concepts for a vacuum tube because the input current approaches zero, and that these transfer functions are unpredictable for a transistor amplifier because of their dependence on β_N. Later chapters discuss feedback methods by which stable amplification of an input current can be achieved. However, it is possible to produce an amplifier with stable A_I or R_T by the method illustrated in Fig. 7.5*b*. If the shunt resistor R is much smaller than the uncertain input resistance of the amplifier, the total input resistance is known and is approximately equal to R. An input current from the source develops a known input voltage across the known total input resistance and, because the voltage-input transfer function G_T or A_V is known, the output is known also. This method for producing a stable current-input transfer function is essentially the method used in Section 5.5.1 for stabilizing the gain of a transistor amplifier.

Transformers are often used in audio-frequency amplifiers for changing the effective source and load impedances, and the advantages discussed in Sections 8.5.2 and 16.2 accrue. A suitable input transformer can change the source impedance to the value that maximizes the signal-to-noise ratio. An output transformer can change the load impedance to the value for which the maximum output power is available from given vacuum tubes or transistors. Finally, a transformer can be used for converting a balanced circuit into a single-ended circuit, and vice versa.

7.3 FREQUENCY AND TIME RESPONSE

In general, the transfer function of any amplifier is different in magnitude and phase for sinusoidal signals of different frequencies. The output,

however, is a sinusoid of the same frequency as the input provided there is no noise or distortion in the amplifier, that is, provided the amplifier is *linear*. Any specification of the variation of the transfer function with frequency gives some measure of the *frequency response*. Also, no linear amplifier passes nonsinusoidal or transient signals without some deformation. A specification of the deformation suffered by a nonsinusoidal signal gives some measure of the *time* or *transient response* of the amplifier.

Both the frequency response and time response of any network are completely and uniquely determined by the positions of its poles and zeros in the complex frequency plane. Time and frequency response are therefore interdependent and any complete specification of one determines the other. At a common-sense level the two are related through Fourier analysis; the deformation of a nonsinusoidal waveform can be calculated from the different amplification of its various sinusoidal Fourier components. It is necessary to develop a flexibility of outlook, so that the performance of a circuit can be investigated in terms of either time or frequency response. The more pertinent approach depends on the problem; time response is generally more useful for signal applications, but frequency response permits a more direct approach to noise calculations.

7.3.1 Mid-Band Frequency Range

The transfer function of an amplifier is given as a function of complex frequency by the ratio of two polynomials in s. These polynomials can be factorized (complex factors are allowed), and the factors grouped such that

$$TF(s) = TF_m\left\{\prod_{k=1}^{n} [\psi_k(s)]\right\} = TF_m\,\psi(s). \tag{7.1}$$

TF_m is a constant multiplier called the *mid-band transfer function* and $\psi(s)$ is the *normalized frequency* (or *time*) *response*; $\psi(s)$ gives the poles and zeros of the transfer function. Each of the $\psi_k(s)$ factors is a complex, frequency-dependent term that is associated with one or more reactive elements in the equivalent circuit of the amplifier.

It is most important to realize that the poles and zeros throughout an amplifier are commutative. That is, the response of an amplifier is the same as long as the total singularity pattern is the same, quite independent of the distribution of the poles and zeros among the individual stages. This degree of freedom in designing for a specified response can be exploited to achieve a convenient practical realization of the amplifier.

The detailed forms of the $\psi_k(s)$ associated with particular reactive elements are discussed in Sections 7.4 and 7.5; there are two broad types:

LOW-FREQUENCY FACTORS

$$\psi_k(s) \to 1 \quad \text{as} \quad s \to \infty, \tag{7.2a}$$

HIGH-FREQUENCY FACTORS

$$\psi_k(s) \to 1 \quad \text{as} \quad s \to 0. \tag{7.2b}$$

The variation of the transfer function at low frequencies occurs when the product of the low-frequency factors departs appreciably from unity; similarly, the variation at high frequencies occurs when the product of the high-frequency factors departs from unity. In between is a *mid-band range of frequencies* for which $|\psi(s)| \approx 1$, and all reactive elements can be neglected. Evidently, TF_m is one of the real quantities G_T, A_V, A_I, or R_T given by the equations of Chapter 5; $TF(s)$ is the corresponding complex quantity Y_T, A_V, A_I, or Z_T.

Some amplifiers are not intended to have a transfer function that is independent of frequency over the operating frequency range. Examples are the equalizing stage in a phonograph (which compensates for the recording characteristic) and operational amplifiers (integrators and differentiators). Almost any desired response can be produced by a suitable array of poles and zeros, that is, a suitable quotient of polynomials in s. The idea of a mid-band range can be extended to accommodate these special cases by treating the ideal quotient of polynomials as the mid-band transfer function; the ψ function is what remains in the transfer function of a practical amplifier. This ψ function contains undesired singularities that cause departures from the ideal response at low and high frequencies. As before, the mid-band frequency range is the range over which the ψ function is approximately unity. The singularities found in following sections may be considered as belonging to TF_m or $\psi(s)$ as appropriate.

7.3.2 Bandwidth, Rise Time, and Tilt

The time response and frequency response of an amplifier can be calculated from its singularity pattern by exact analytical means or approximate graphical means; both methods are treated in specialized books. With experience, it is possible to approximate the response by inspection. Often this is easy, because the response of most amplifiers is dominated by a relatively few singularities. The complete singularity pattern contains much insignificant, or even useless and possibly confusing, information. One of two simple alternative specifications of the response is adequate for most purposes.

The first alternative specification of the response of an amplifier is its *bandwidth*, the frequency range over which the gain is approximately constant for sinusoidal signals. Usually the bandwidth is taken as the frequency range over which the power gain differs by less than a factor of two from its mid-band value, and the upper and lower frequency limits are called the *upper and lower half-power points* or *3-dB points*. Because power is proportional to the square of voltage or current, it follows that the voltage gain or current gain is constant to within a factor $\sqrt{2}$ over the 3-dB bandwidth. The adoption of half-power bandwidth as standard is convenient for analysis and realistic in practice. Analytically, the 3-dB point occurs at the frequency ω where

$$|\psi(j\omega)|^2 = \tfrac{1}{2}, \tag{7.3}$$

and in a single-pole system the 3-dB point is at the pole frequency. In practice, an amplifier is satisfactory for all but the most stringent requirements if its 3-dB bandwidth encloses the signal frequency range; the maximum change in gain over the passband is about 30%.

The second specification of the response of an amplifier is in terms of its deformation of a square wave of given frequency. A square wave is built up from a series of steps in opposite directions; because the step function is one of the basic signals in transform algebra, specification of the square-wave response allows ready determination of the response to most other simple waveforms. The two most significant parameters in the square-wave response of an amplifier are its *rise time* and *percentage tilt*. Unless the contrary is specified, the rise time is the time taken for the output to change from 10 to 90% of its final value when a square wave or step of very fast rise time is applied to the input. Associated with the rise time is the *percentage overshoot* and *ringing*; rise time, overshoot, and ringing are illustrated in Fig. 7.6.* The percentage tilt on the top of a square wave is a measure of the flatness of its top; this is illustrated in Fig. 7.7. The rise time of any network is related to its high-frequency response (though not uniquely), since the "edge" of a square wave is determined by the higher-frequency components in its Fourier series. Similarly, the tilt on a square wave of given frequency is related to the low-frequency response, since the "body" of a square wave is determined by its lower-frequency Fourier components.

The 3-dB bandwidth of an amplifier does not define its frequency response completely; there can be variations in phase or flatness of the

* The rise time, overshoot, and ringing on a square wave are often referred to as the *leading-edge* response. Sometimes, the tilt on a square wave is referred to as the trailing-edge response. The authors prefer *tilt* or *top response* and reserve *trailing-edge* for the downstroke of a square wave or pulse.

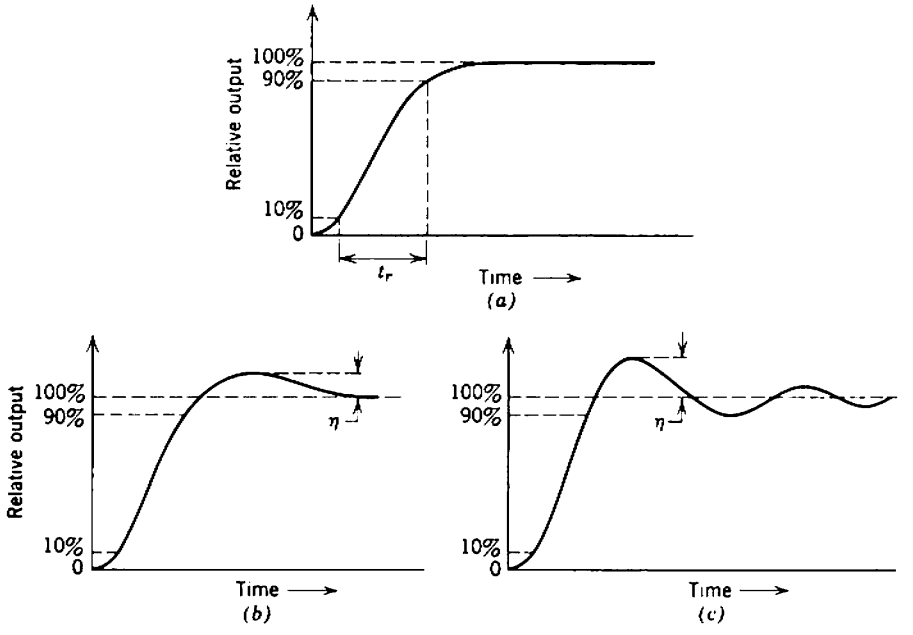


Fig. 7.6 Leading-edge response of an amplifier, that is, the response to a step-function input: (*a*) 10 to 90% rise time; (*b*) percentage overshoot; (*c*) percentage ring.

response inside the passband, and variations in the rate of attenuation outside the passband. The only complete specification is the singularity pattern. Similarly, the rise time and tilt do not define the square-wave response completely; there can be variations of the output waveform between the prescribed limits. Unfortunately, the realizable singularity patterns that give the flattest frequency response with specified 3-dB bandwidth do not give the best time response; there are considerable overshoot and tilt. It can be shown that the gain of an amplifier with optimum

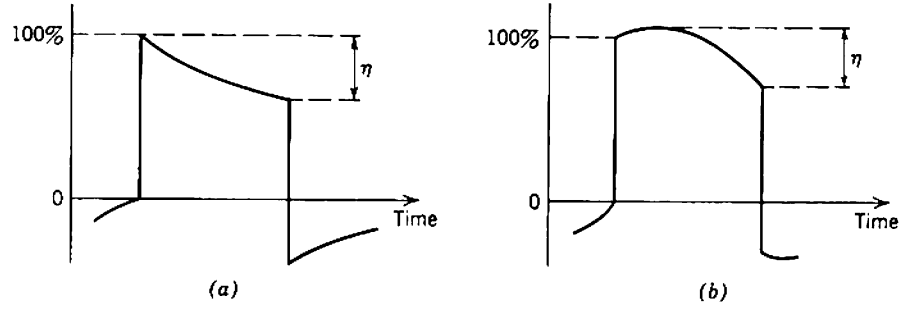


Fig. 7.7 Top response of an amplifier, that is, the response to a square-wave input: (*a*) simple tilt; (*b*) "bowed" tilt.

square-wave response is less than the gain of an amplifier with flat frequency response at high frequencies, but greater at low frequencies. There is, therefore, no one optimum design for an amplifier with specified bandwidth; the optimum singularity pattern depends on the waveform most commonly encountered. This is one example of the approximation problem. As a guiding rule, amplifiers for use in audio systems or with meters should have a flat frequency response, because the waveforms encountered are principally sinusoidal. Amplifiers for use in pulse or visual systems (television or oscilloscopes) should have optimum square-wave response.

7.3.3 Conventions

In general, the singularity pattern of an amplifier stage depends on which of the four transfer functions Y_T, A_V, A_I, and Z_T is being considered. As shown in Section 5.5.1, the stages within a multistage amplifier can be regarded as having voltage gain, current gain, and alternate transfer admittance and impedance. Although the individual singularity patterns depend on the breakup into stages, the total pattern for the stages inside an amplifier must always remain the same. However, the input and output circuit singularity patterns depend on the source and load impedance, and on the transfer function being considered. Given the singularity pattern for one combination of source impedance, load impedance, and transfer function, it requires only simple algebra to find the singularity pattern for any other combination. In the following sections the singularity pattern of a stage is derived in terms of the voltage applied to its input electrode. Voltage-input transfer functions are chosen for three reasons:

(i) the input voltage to a charge-controlled device gives the most direct control over the mobile charge;
(ii) the concepts of input current and current gain are almost meaningless for the vacuum-tube circuits thus far considered in this book;
(iii) the predictable transfer functions of simple transistor stages are those involving input voltage.

Other transfer functions are more suitable for some of the feedback amplifier configurations developed in later chapters of this book.

7.4 LOW-FREQUENCY RESPONSE OF *RC*-COUPLED AMPLIFIER STAGES

There are three general causes for the variation in gain of a resistance-capacitance-coupled (*RC*-coupled) amplifier stage at low frequencies. These are the interstage coupling capacitor, the bypass capacitors in the

biasing system, and the decoupling capacitor. Each of these capacitors contributes one pole and one zero to the amplifier's singularity pattern. Both singularities lie on the negative real axis in the complex frequency plane and are relatively close to the origin.

Coupling, bypass, and decoupling capacitors are introduced in Section 6.4. Coupling capacitors are used to transfer the signal voltages and currents to and from the input and output electrodes of the amplifying devices without affecting the quiescent dc conditions at these electrodes. Bypass capacitors hold the nonsignal electrodes such as the cathode, emitter, and screen at zero signal voltage when their quiescent voltages are obtained through biasing supply resistors. Decoupling capacitors are used to reduce unwanted ac signals on the dc supply rails.

It is possible to write down the network equations for the most general topology expected in an amplifier stage, and solve them to find the positions of all the low-frequency singularities. Simpler stages can be regarded as special cases of the general stage by letting certain component values become zero or infinite. While such an approach allows easy analysis of any multistage amplifier, it does little to facilitate amplifier design. It gives little insight into how each capacitor contributes to the over-all low-frequency response and, therefore, it does not suggest sound

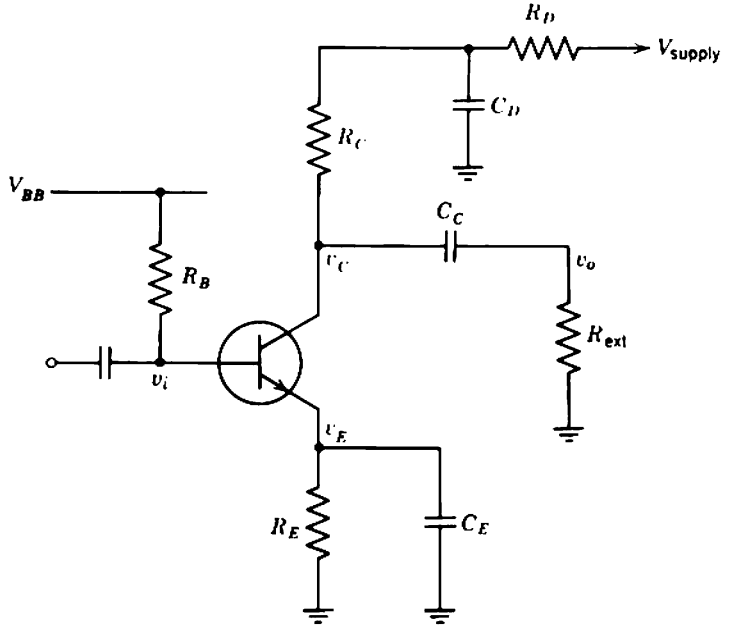


Fig. 7.8 Illustrative *RC*-coupled amplifier.

practical design procedures. In order to give some feeling for the operation of a circuit, the low-frequency behavior of the simple transistor amplifier shown in Fig. 7.8 is considered initially at a purely descriptive level. With this common-sense basis on which to build, the argument can proceed to an analytical approach from which the locations of the singularities can be found.

There are two coupling capacitors in Fig. 7.8, one at the input and one at the output. Only one of these need be considered as belonging to the stage; the other belongs to an adjacent stage. In accordance with the arbitrary standard proposed in Section 7.3.3, it is convenient to associate the output coupling capacitor with the stage. Suppose, therefore, that in Fig. 7.8 there is a signal voltage at the base of the transistor, whose frequency can be varied at will but whose amplitude is constant.

At frequencies that are not too low, the reactance of the emitter bypass capacitor C_E is so small that it constitutes a virtual short circuit to ground. The transfer admittance Y_T of the stage is given by G_T in the mid-band equations of Chapter 5, and is independent of frequency. At lower frequencies, the reactance of the emitter bypass capacitor rises and the signal current in the emitter develops a voltage across the finite impedance in the emitter circuit. This voltage subtracts from the input and reduces the voltage applied between base and emitter (Fig. 7.9). It follows that Y_T falls as the reactance of C_E increases with decreasing frequency. It also follows that the output impedance of the transistor rises, but this is not

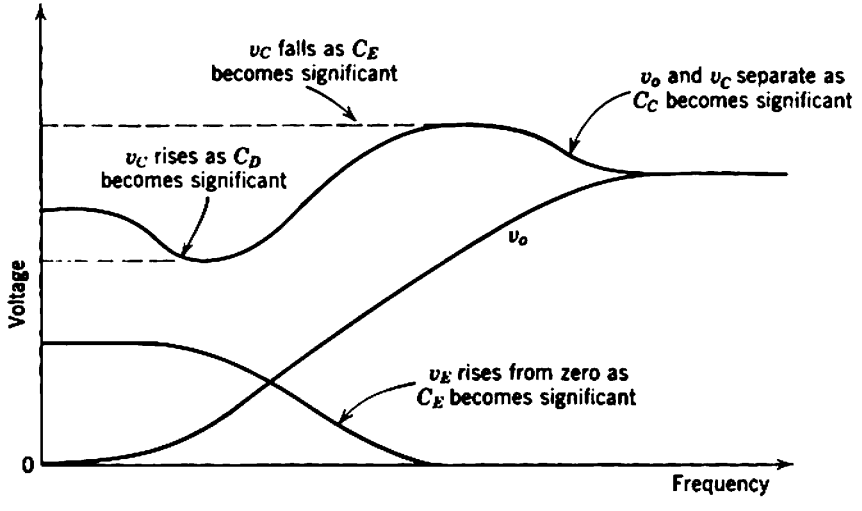


Fig. 7.9 Signal voltages versus frequency in the illustrative *RC*-coupled amplifier (Fig. 7.8).

apparent in Fig. 7.9. At very low frequencies the reactance of C_E is much greater than R_E; the impedance in the emitter circuit becomes constant and Y_T is again independent of frequency.

The voltage gain of the stage defined as v_C/v_i is determined by Y_T and the load in the collector circuit. At low frequencies, the reactance of C_C is very large and therefore the load is equal to R_C (neglecting R_D and C_D for the moment). The reactance of C_C falls at higher frequencies and therefore the load falls to R_C and R_{ext} in parallel. The voltage gain v_C/v_i falls to the mid-band gain calculated in Chapter 5, and v_C and v_o are equal. However, at the lower frequencies for which the reactance of C_C is appreciable, C_C forms a voltage divider with R_{ext} and v_o is less than v_C. At very low frequencies v_C remains finite, whereas v_o falls away toward zero.

Finally, there is the effect of the decoupling components R_D and C_D. At very low frequencies, the reactance of C_D increases and the load rises from R_C to $(R_C + R_D)$. The voltage gain v_C/v_i rises and there is a corresponding increase in v_o/v_i.

In a vacuum-tube circuit, the cathode bypass, coupling, and decoupling capacitors correspond to the emitter bypass, coupling, and decoupling capacitors. For a pentode, the screen bypass capacitor is similar in its final effect to the cathode bypass capacitor. With biasing circuits other than the emitting-electrode-feedback type, the effect of the various bypass capacitors is qualitatively similar to the emitter bypass considered in Fig. 7.8.

In general the various capacitors of a practical circuit do not influence the response over separate frequency ranges as is shown in Fig. 7.9 for simplicity. All the *break* or *corner frequencies* (the frequencies at which the gain begins to change) can lie quite close to each other. To an approximation, each capacitor in the circuit contributes an independent term to Eq. 7.1 of the form

$$\psi_k(s) = \frac{s - z}{s - p} = \frac{s + \omega_2}{s + \omega_1} = \frac{s + 1/\tau_2}{s + 1/\tau_1}, \tag{7.4}$$

where p and z = negative real numbers, and are the positions of the pole and zero on the negative real axis of the complex frequency plane,

ω_1 and ω_2 = the corner frequencies associated with the pole and zero, respectively,

τ_1 and τ_2 = the time constants associated with the pole and zero, respectively.

The corner frequencies are generally used in expressions for the frequency response, whereas the time constants are used in expressions for the transient response.

To assume that the various $\psi_k(s)$ terms in Eq. 7.1 are completely independent is an approximation, because the position of the pole associated with one capacitor very slightly influences the positions of the poles associated with the other capacitors. However, the pole positions also depend on the active device parameters such as g_m and β_N, and these have appreciable tolerances. Except for the case noted in Section 7.4.2.2, errors in pole positions that arise from neglecting cross-coupling are insignificant compared with the variations due to active device tolerances. It is therefore a sound engineering approximation (and extremely useful too) to neglect cross-coupling. In the following sections each of the possible $\psi_k(s)$ is evaluated on the assumption that there is no cross-coupling, although the direction in which a pole might shift is indicated. Additional approximate expressions are deduced which are useful for design purposes; these approximations are pessimistic in that the poles actually lie closer to the origin than the approximation suggests. Good practice is to overdesign rather than underdesign, so that small errors in the pole positions have little effect inside the passband.

7.4.1 Bypass Capacitors in the Biasing System

A multitude of biasing circuits is described in Chapter 6. Only the most common of these, the emitting-electrode-feedback circuit of Section 6.1, is discussed in this chapter because the general form of its results is typical of all biasing systems. The detailed analysis of other biasing circuits follows most easily from the feedback theory of Chapter 10; for the most part this is left as an exercise for the reader, but some further discussion is given in Chapter 12. Specifically, the cathode- and emitter-feedback biasing circuits are considered here, together with the effect of the screen bypass capacitor of a pentode. Vacuum-tube and transistor circuits are considered separately, as their analyses differ significantly.

Figure 7.10 shows the singularity pattern associated with any bypass

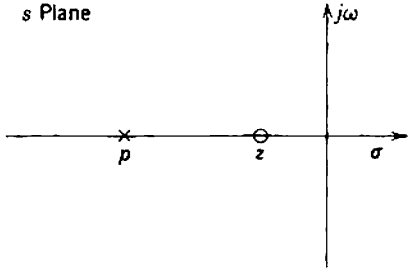


Fig. 7.10 Singularity pattern for $\psi_k(s)$ associated with a single bypass capacitor.

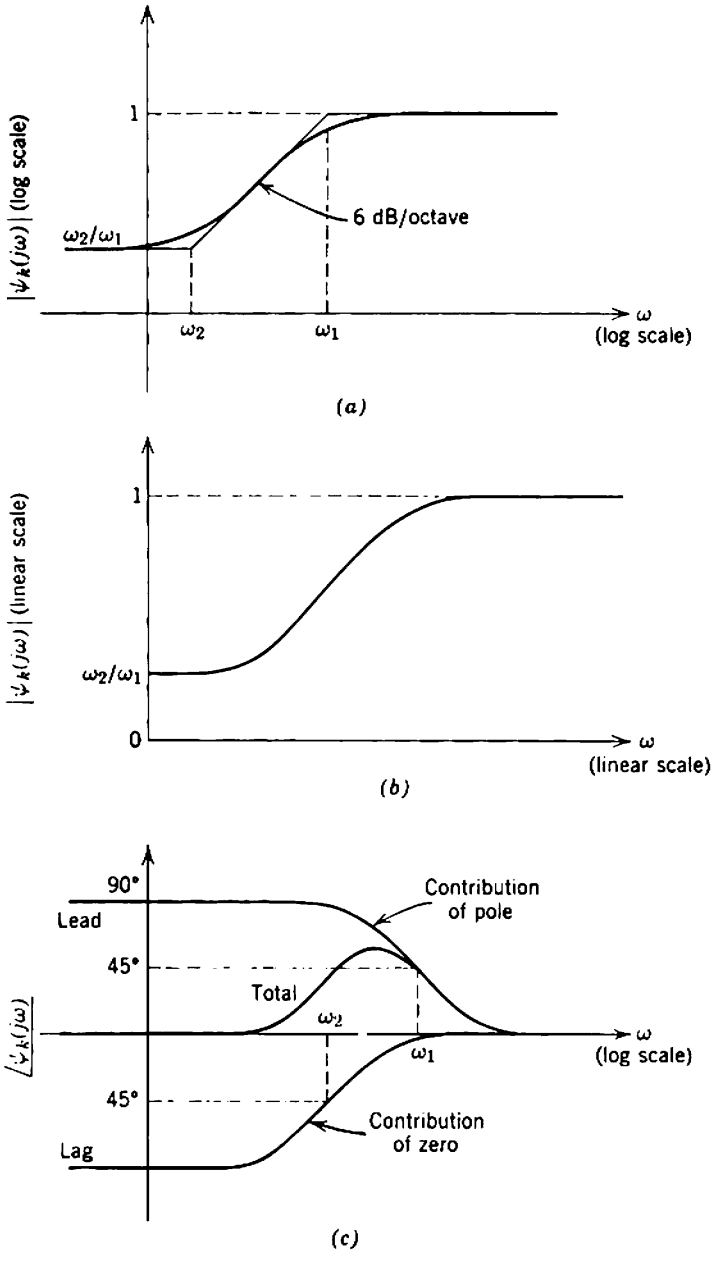


Fig. 7.11 Frequency response of $\psi_k(s)$ associated with a single bypass capacitor.

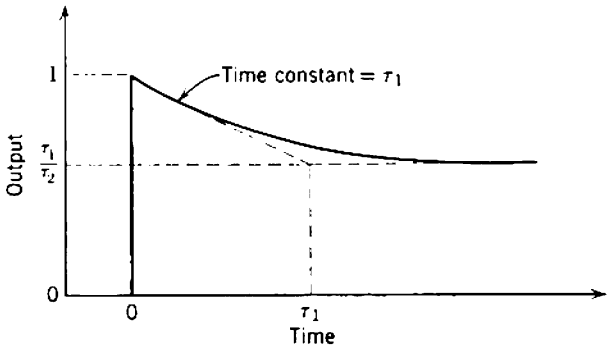


Fig. 7.12 Step response of $\psi_k(s)$ associated with a single bypass capacitor.

capacitor; $|z|$ is always less than $|p|$. Figure 7.11 shows the frequency response of the term $\psi_k(s)$, which may be considered as the normalized response of an amplifier whose only source of low-frequency variation is a single bypass capacitor. Figure 7.11*a* shows the gain versus frequency on logarithmic scales for which the asymptotes are apparent, whereas Fig. 7.11*b* is on linear scales to show the behavior at zero frequency. Finally, Fig. 7.11*c* shows the variation of phase with frequency. Similarly, $\psi_k(s)$ can be considered as the normalized transient response of an amplifier stage and Fig. 7.12 shows the response to a unit step function.

7.4.1.1 Cathode Bypass Capacitor

Figure 7.13*a* shows the circuit diagram for the general cathode-feedback biasing system. Although a triode is shown, the tube may equally well be a pentode. Figure 7.13*b* shows the elemental circuit used in calculating the effect of the cathode bypass capacitor; the components in the anode circuit are grouped as Z_L, and the cathode biasing network is grouped as Z_K, where

$$Z_K = \frac{R_K}{1 + sR_KC_K}. \tag{7.5}$$

Solution of the network equations gives the transfer admittance as

$$\frac{i_A}{v_G} = \frac{g_m r_A}{r_A + Z_L + (\mu + 1)Z_K}, \tag{7.6}$$

and the voltage gain is found by multiplying by Z_L:

$$\frac{v_A}{v_G} = -\frac{g_m r_A Z_L}{r_A + Z_L + (\mu + 1)Z_K}. \tag{7.7}$$

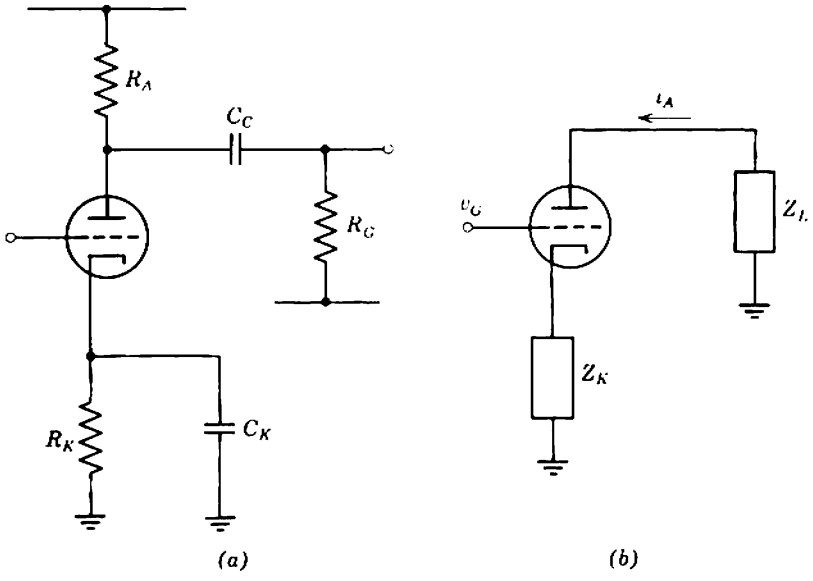


Fig. 7.13 Vacuum tube with cathode-feedback biasing: (*a*) complete circuit; (*b*) elemental circuit.

By comparison with Eq. 5.46 the inclusion of Z_K reduces the gain and increases the output impedance. Equation 7.6 can be manipulated into the form

$$\frac{i_A}{v_G} = \left(\frac{g_m r_A}{r_A + Z_L}\right)\left(\frac{s - z}{s - p}\right)$$

or, in terms of the mid-band transfer conductance G_T given by Eq. 5.45,

$$\frac{i_A}{v_G} = G_T\left(\frac{s - z}{s - p}\right) \tag{7.8}$$

so that the frequency-dependent factor $\psi_k(s)$ is explicit. The values of p and z are

$$-\frac{1}{p} = \frac{1}{\omega_1} = \tau_1 = R_K C_K\left[\frac{r_A + R_L}{r_A + R_L + (\mu + 1)R_K}\right] \tag{7.9}$$

and

$$-\frac{1}{z} = \frac{1}{\omega_2} = \tau_2 = R_K C_K. \tag{7.10}$$

where, for simplicity, it is assumed that Z_L is resistive and equal to R_L.

In the general case p and z can be evaluated. The zero is accurately defined by passive elements but, as it lies closer to the origin than the pole, the zero has less effect in the passband; for many practical purposes the

zero can be assumed to lie at the origin. The pole controls the essential features of the dynamic response, but as p depends on both μ and r_A the pole position is inaccurately defined. A pessimistic approximation is to assume that

$$r_A \gg Z_L$$

and

$$\mu + 1 \approx \mu \gg \frac{r_A}{R_K},$$

so that

$$p \approx -\frac{g_m}{C_K}. \tag{7.11}$$

This approximation is usually good for a pentode, but may be poor for a low-μ triode.

It is worth emphasizing that an amplifier which is to have a lower half-power frequency ω_L due to its cathode bypass capacitor does not have this capacitor chosen approximately such that it breaks with R_K at ω_L; that is,

$$\omega_L C_K \not\approx \frac{1}{R_K}.$$

Rather, the approximate value is

$$\omega_L C_K \approx g_m. \tag{7.12}$$

The value of C_K for specified ω_L is determined almost entirely by the tube type and not by the detail of the biasing circuit.

Very slight cross-coupling occurs between the poles due to the cathode bypass and coupling capacitors, because Z_L appears in Eq. 7.6 and thereafter, and Z_L involves the coupling capacitor in the anode circuit. However, Z_L is neglected in Eq. 7.11, and the design approximation for the cathode-bypass-capacitor pole does not involve the output coupling capacitor.

7.4.1.2 Screen Bypass Capacitor

In an amplifier stage using a pentode, the screen is usually biased at its desired quiescent voltage with an RC network of the form shown in Fig. 7.14*a*. The effect of the screen bypass capacitor C_S on the low-frequency response of the amplifier can be calculated from an elemental circuit such as 7.14*b*; all reactive elements except C_S are ignored and the components in the load and screen networks are grouped together as Z_L and Z_S, respectively:

$$Z_S = \frac{R_S}{1 + sR_SC_S}. \tag{7.13}$$

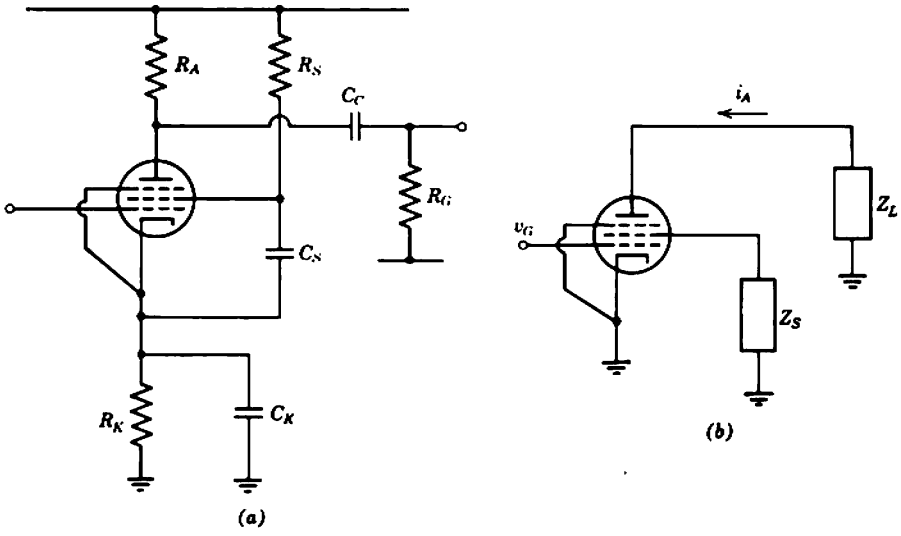


Fig. 7.14 Pentode with screen biasing network: (*a*) complete circuit; (*b*) elemental circuit.

Like all biasing networks, the screen *RC* supply network is essentially a feedback system which reduces the gain at low frequencies when the reactance of C_S becomes appreciable. Suppose that the grid voltage of a pentode is changing in the positive direction, so that the anode and screen currents increase. Due to the presence of Z_S, the screen voltage falls. This partially counteracts the rising grid voltage and reduces the amount by which the various currents increase; the effective transfer admittance and voltage gain are reduced. At moderately high frequencies where the reactance of C_S is small, Z_S is negligible and the mid-band equations of Chapter 5 apply. The gain falls with decreasing frequency as the reactance of C_S rises, but levels off again at very low frequencies when Z_S approaches R_S and becomes constant.

The transfer admittance of the stage is approximately

$$\frac{i_A}{v_G} = g_m\left(\frac{1}{1 + kg_mZ_S/\mu_S}\right), \tag{7.14}$$

where k is the ratio of screen to anode current and μ_S is the screen amplification factor. (The approximation assumes that r_A is large compared with Z_L, a valid assumption for any normal pentode.) Substitution from Eq. 7.13 gives

$$\frac{i_A}{v_G} = g_m\left(\frac{s - z}{s - p}\right).$$

Since the mid-band transfer conductance of a pentode is g_m,

$$\frac{i_A}{v_G} = G_T\left(\frac{s - z}{s - p}\right) \tag{7.15}$$

and the frequency-dependent term $\psi_k(s)$ is explicit. The values of p and z are

$$-\frac{1}{p} = \frac{1}{\omega_1} = \tau_1 = \frac{R_S C_S}{1 + kg_m R_S/\mu_S} \tag{7.16}$$

and

$$-\frac{1}{z} = \frac{1}{\omega_2} = \tau_2 = R_S C_S. \tag{7.17}$$

As with the cathode biasing network, the pole is more significant than the zero, even though the position of the pole is poorly defined because it depends on the tube parameters k, g_m, and μ_S. A pessimistic approximation for p is

$$p \approx -\frac{kg_m}{\mu_S C_S}. \tag{7.18}$$

The screen-circuit time constants react on both the RC coupling network and the cathode biasing network, and move their poles slightly toward the origin. The exact effect depends on whether the screen bypass capacitor is returned directly to the cathode or to ground; in either case the effect is quite small.

7.4.1.3 Emitter Bypass Capacitor

Qualitatively, the effect of the emitter bypass capacitor is the same as a cathode bypass capacitor. Quantitatively, the effect of the emitter bypass capacitor on the singularity pattern of a transistor amplifier is more complicated. The reactance of the capacitor affects the input impedance of the transistor, and plays an important part in determining the position of the pole due to the input coupling capacitor.

Figure 7.15 shows the elemental circuit design for a transistor amplifier stage that has an impedance Z_E in the emitter circuit. If the load impedance is not too large, the transfer admittance is

$$\frac{i_C}{v_B} = \frac{\alpha_N}{r_B/\beta_N + r_E + Z_E}. \tag{7.19}$$

If Z_E is the emitter biasing network such that

$$Z_E = \frac{R_E}{1 + sR_E C_E}, \tag{7.20}$$

then

$$\frac{i_C}{v_B} = \left(\frac{\alpha_N}{r_B/\beta_N + r_E}\right)\left(\frac{s - z}{s - p}\right)$$

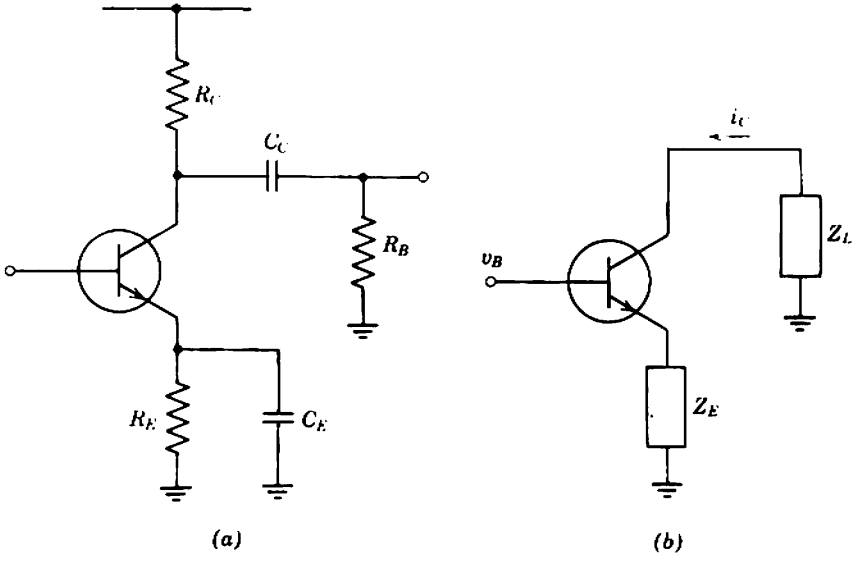


Fig. 7.15 Transistor with emitter-feedback biasing: (*a*) complete circuit; (*b*) elemental circuit.

or, in terms of the mid-band transfer conductance given by Eq. 5.76,

$$\frac{i_C}{v_B} = G_T\left(\frac{s - z}{s - p}\right), \tag{7.21}$$

where

$$-\frac{1}{p} = \frac{1}{\omega_1} = \tau_1 = R_E C_E\left(\frac{r_B/\beta_N + r_E}{r_B/\beta_N + r_E + R_E}\right) \tag{7.22}$$

and

$$-\frac{1}{z} = \frac{1}{\omega_2} = \tau_2 = R_E C_E. \tag{7.23}$$

If the amplifier stage has been designed for reasonable biasing stability such that its gain is predictable despite the production tolerance on β_N, then

$$R_E \gg r_E$$

and, as discussed in Section 5.3.1,

$$\frac{r_B}{\beta_N} \ll r_E.$$

The expression for p simplifies to

$$p \approx -\frac{1}{r_E C_E} = -\frac{qI_E}{kTC_E}. \tag{7.24}$$

The position of the pole is therefore almost independent of the detail of the biasing system and the transistor type, but is critically dependent on the quiescent dc emitter current I_E. On the other hand, if r_B/β_N is not negligible in comparison with r_E, the pole moves closer to the origin and the low-frequency performance is improved. Against this, both the pole position and the mid-band gain are unpredictable for they depend on β_N.

It is instructive to compare Eq. 7.24 with Eq. 7.11 of the vacuum-tube analysis. The mutual conductance of a transistor is

$$g_m = \frac{\alpha_N}{r_E} \approx \frac{1}{r_E}.$$

Substitution into Eq. 7.24 gives

$$p \approx -\frac{g_m}{C_E},$$

which is identical in form with Eq. 7.11.

The input impedance of a transistor is

$$Z_i = r_B + \beta_N r_E + \beta_N Z_E, \tag{7.25}$$

provided the load impedance is not too large. If Z_E is an emitter biasing network R_E and C_E, then Z_i has one pole and one zero at low frequencies. This gives rise to gross cross-coupling between the emitter bypass capacitor and the coupling capacitor preceding a stage. Further discussion appears in Section 7.4.2.2.

7.4.2 Coupling Capacitors

Any *RC* coupling network can be reduced to the form shown in Fig. 7.16. In the case of an amplifier stage that is coupled to a source:

i_{in} and v_{in} represent the source current or voltage, as appropriate;
R_1 represents the source resistance and its termination (if any);
R_2 represents the parallel combination of the input resistance of the device and its control-electrode supply resistor.

In the case of an interstage coupling network:

i_{in} represents the g_m generator of the first device;
R_1 represents the parallel combination of the output resistance of the first device and its collecting-electrode supply resistor;
R_2 represents the parallel combination of the input resistance of the second device and its control-electrode supply resistor.

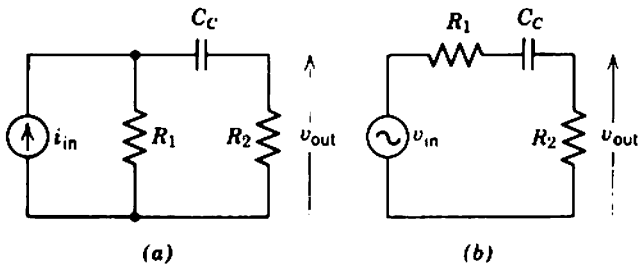


Fig. 7.16 *RC* coupling network in outline: (*a*) current-source representation; (*b*) voltage-source representation, $v_{in} = i_{in}R_1$.

Finally, in the case of a device that is coupled to an external load:

i_{in} represents the g_m generator of the device;
R_1 represents the parallel combination of the output resistance of the device and its collecting-electrode supply resistor;
R_2 represents the external load.

The coupling capacitor C_C forms a voltage divider with R_2, and the output voltage falls from the mid-band value calculated in Chapter 5 to zero at very low frequencies.

The network equations give

$$v_{out} = i_{in}\left(\frac{R_1R_2}{R_1 + R_2}\right)\left[\frac{s(R_1 + R_2)C_C}{1 + s(R_1 + R_2)C_C}\right]. \tag{7.26}$$

It follows that the frequency-dependent term $\psi_k(s)$ in Eq. 7.1 due to a coupling capacitor is

$$\psi_k(s) = \frac{s}{s - p} = \frac{s}{s + \omega_1} = \frac{s}{s + 1/\tau_1}, \tag{7.27}$$

where

$$-\frac{1}{p} = \frac{1}{\omega_1} = \tau_1 = (R_1 + R_2)C_C. \tag{7.28}$$

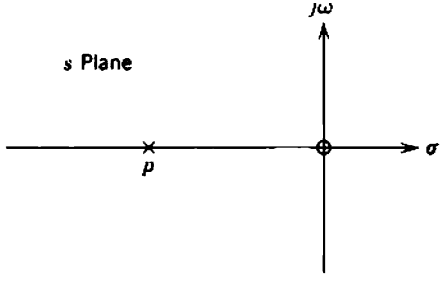


Fig. 7.17 Singularity pattern for $\psi_k(s)$ associated with an *RC* coupling network.

Figure 7.17 shows the singularity pattern for a coupling network; there is a zero at the origin and a pole whose position is determined by the circuit parameters. Figures 7.18 and 7.19 show the frequency response and time response. It is interesting to note that, when an *RC* coupling network with lower cutoff frequency f passes a square wave of repetition fre-

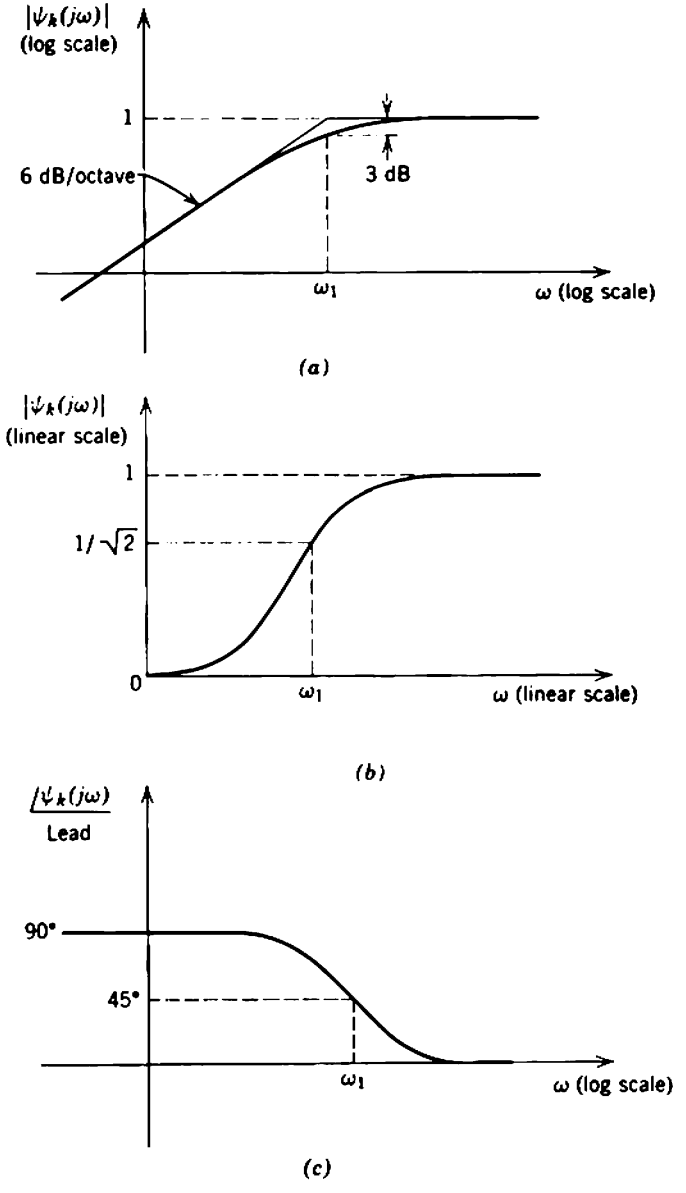


Fig. 7.18 Frequency response of $\psi_k(s)$ associated with an *RC* coupling network.

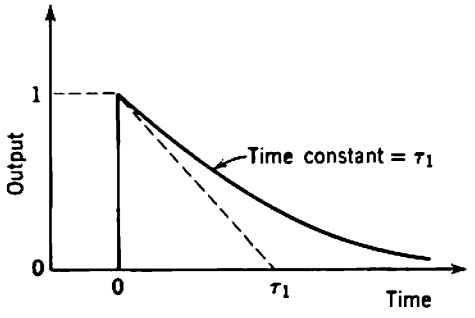


Fig. 7.19 Step response of $\psi_k(s)$ associated with an *RC* coupling network.

quency f, the output waveform is grossly deformed as shown in Fig. 7.20*a*. This gross deformation occurs despite the fact that the greatest change in amplitude of any Fourier component is only -30% (-3 dB), because of the phase shift in all the low-frequency components; the phase of the low-frequency components is advanced, and the "body" of the square wave is moved ahead in time. The lower cutoff frequency of the amplifier must

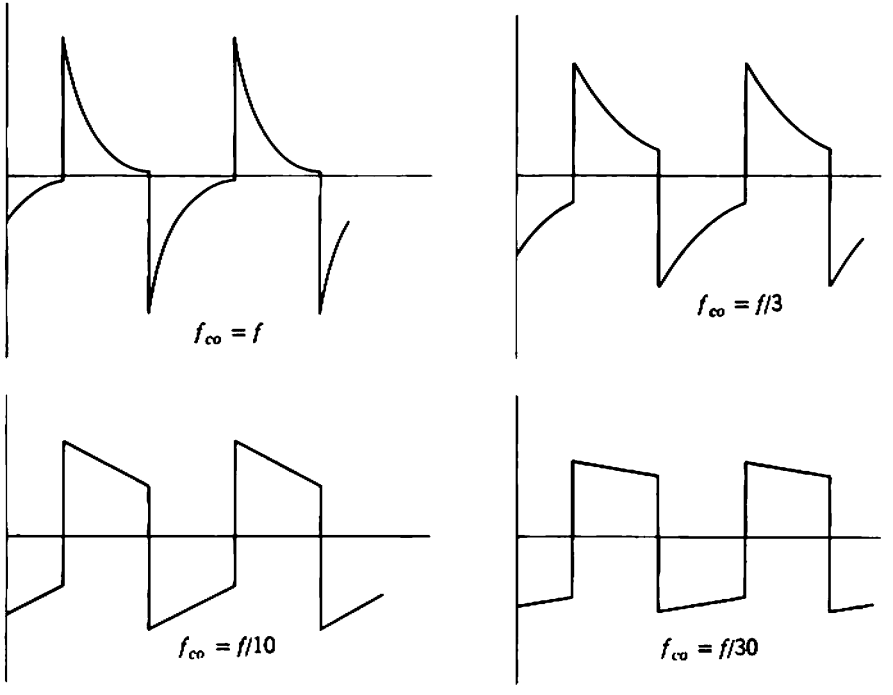


Fig. 7.20 Deformation of a square wave, frequency f, by *RC* coupling network, cutoff frequency f_{co}.

be less than the square-wave frequency by a factor 30 before the tilt is reduced to 5%.

7.4.2.1 Vacuum Tube

In the interstage coupling network between two vacuum tubes shown in Fig. 7.21, the output resistance of the first tube is

$$R_o \approx r_A \tag{7.29}$$

and the input resistance of the second tube is infinite. Therefore,

$$R_1 = \frac{r_A R_A}{r_A + R_A} \tag{7.30}$$

and

$$R_2 = R_G. \tag{7.31}$$

Therefore the pole is

$$-\frac{1}{p} = \frac{1}{\omega_1} = \tau_1 = \left(\frac{r_A R_A}{r_A + R_A} + R_G\right) C_C. \tag{7.32}$$

In almost all circuits, R_G is at least twice as large as R_A, and even larger ratios are often encountered. Therefore, a design approximation is

$$p \approx -\frac{1}{R_G C_C}. \tag{7.33}$$

This approximation is unlikely to be in error by more than 30%, and is often much better. Further, it is a pessimistic approximation. The interpretation of Eq. 7.33 is that the gain falls below the frequency at which R_G breaks with C_C, that is, the frequency at which the voltage dividing action of C_C with R_G becomes appreciable.

There is a small error in Eq. 7.29. The output impedance of a vacuum tube is r_A only if the cathode (and screen also if the tube is a pentode) is at

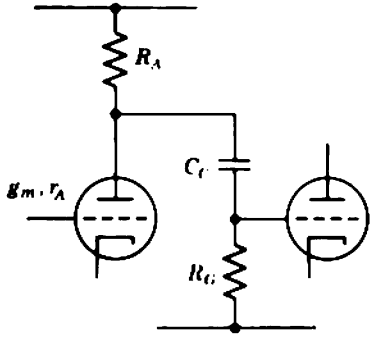


Fig. 7.21 *RC* coupling network between vacuum tubes.

zero signal voltage, that is, if these electrodes are perfectly bypassed to ground. Bypassing is not perfect at low frequencies and, therefore, the output impedance is greater than r_A. It follows that the coupling circuit pole is closer to the origin than the position given by Eq. 7.32; the error is small and is not taken into account in Eq. 7.33.

Expressions for the pole positions when a vacuum tube is capacitor coupled to an external source or load follow from trivial extensions of the above development.

7.4.2.2 Transistor

Figure 7.22 shows the interstage coupling network between two transistors. The output resistance of the first transistor is very large, and the input impedance of the second is Z_i given by Eq. 7.25. The values of R_1 and R_2 that must be substituted into Eq. 7.26 to find $\psi_k(s)$ are therefore

$$R_1 = R_C, \tag{7.34}$$

$$R_2 \to Z_2 = \frac{R_B Z_i}{R_B + Z_i}. \tag{7.35}$$

Notice that R_2 is not simply a resistance, because Z_i involves the impedance of R_E and C_E in the emitter circuit. It follows that $\psi_k(s)$ for the coupling capacitor has two poles and two zeros, not one of each as for vacuum tubes. Invariably, however, one of the zeros cancels the pole due to the emitter bypass capacitor (Eq. 7.22), so the complete system has a two-pole, two-zero transfer function.

The system zeros can be found easily. One is the emitter-bypass zero (Eq. 7.23); the other is due to the coupling capacitor and lies at the origin. General expressions for the coupling capacitor poles can be written down; however, the pole positions depend on both C_C and C_E (i.e., cross-coupling occurs), and the general expressions are so complicated that it is difficult

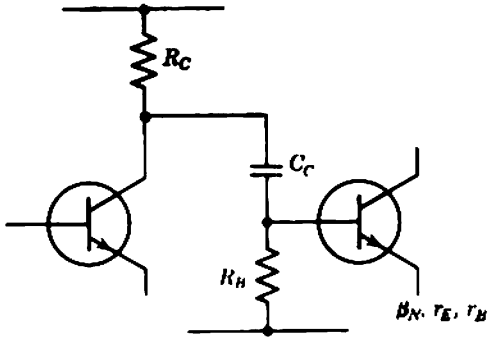


Fig. 7.22 *RC* coupling network between transistors.

to deduce anything useful from them. Three special cases account for the useful possibilities; Problem 12.5 contains further information.

CASE 1. If $R_B \ll (r_B + \beta_N r_E)$, then $Z_2 \approx R_B$ and cross-coupling between C_C and C_E is eliminated. The coupling capacitor contributes a zero at the origin, and a pole

$$-\frac{1}{p} = (R_C + R_B)C_C. \tag{7.36}$$

The emitter bypass capacitor contributes the pole and the zero is given in Section 7.4.1.3.

CASE 2. If C_E is very large so that $|Z_E|$ is small over the entire useful frequency range of the amplifier, Eq. 7.25 becomes

$$Z_i \approx r_B + \beta_N r_E.$$

The coupling capacitor contributes a dominant pole

$$-\frac{1}{p} = \left[R_C + \frac{R_B(r_B + \beta_N r_E)}{R_B + (r_B + \beta_N r_E)}\right]C_C, \tag{7.37}$$

and a second pole that lies much closer to the origin and close to the emitter-bypass zero. Overall, the system reduces approximately to the dominant pole plus a zero at the origin.

CASE 3. If $R_B \gg (r_B + \beta_N r_E)$ so that $Z_2 \approx Z_i$, the coupling capacitor contributes a dominant pole

$$-\frac{1}{p} = (R_C + r_B + \beta_N r_E)\left(\frac{C_C C_E}{\beta_N C_C + C_E}\right), \tag{7.38}$$

and a second pole that lies much closer to the origin and close to the emitter-bypass zero. Again, the system reduces approximately to the single dominant pole plus a zero at the origin. A useful practical simplification results from choosing

$$\beta_N C_C \gg C_E,$$

when Eq. 7.38 becomes

$$-\frac{1}{p} = \left(\frac{R_C + r_B}{\beta_N} + r_E\right)C_E. \tag{7.39}$$

This last case is the most generally useful.

Expressions for the pole positions when a transistor is capacitor coupled to an external source or load follow from trivial extensions of the above development.

7.4.3 Decoupling Networks

Decoupling networks contribute one real-axis pole and one real-axis zero to the singularity pattern of an *RC*-coupled amplifier stage. Like

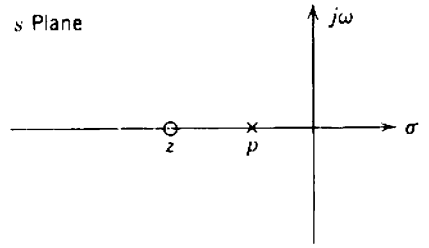


Fig. 7.23 Singularity pattern for $\psi_k(s)$ associated with a decoupling network.

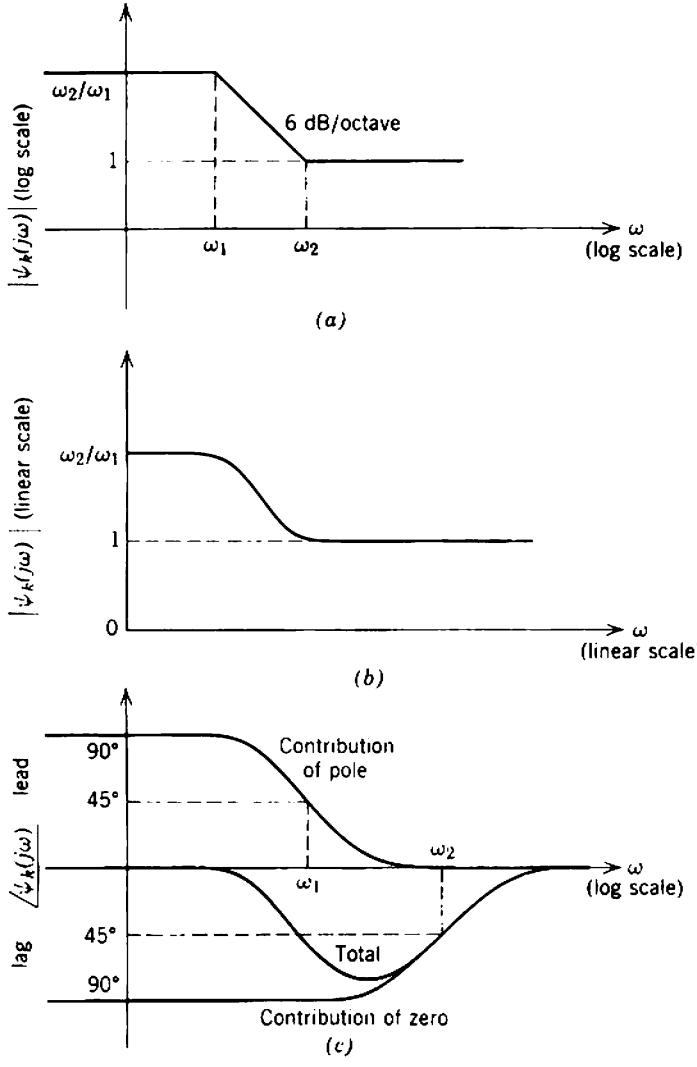


Fig. 7.24 Frequency response of $\psi_k(s)$ associated with a decoupling network.

the bypass capacitors, neither singularity is at the origin; in contrast to the bypass capacitors, the pole is closer to the origin than the zero. The form of the frequency-dependent term $\psi_k(s)$ in Eq. 7.1 due to a decoupling network is

$$\psi_k(s) = \frac{s - z}{s - p} = \frac{s + \omega_2}{s + \omega_1} = \frac{s + 1/\tau_2}{s + 1/\tau_1}, \tag{7.40}$$

where $|z|$ is greater than $|p|$. Figure 7.23 shows the singularity pattern, and Figs. 7.24 and 7.25 show the frequency and time response. Qualitatively, the operation of the circuit is as discussed in connection with Fig. 7.8.

The zero is more significant than the pole because it lies further from the origin, but often the circuit values in audio and subaudio amplifiers are such that both can be neglected. First, the pole and the zero lie so close to each other that, to an excellent approximation, they cancel. Second, both singularities are so close to the origin that they have negligible effect on the passband even if they do not cancel. However, a number of inequalities hold between the various resistances in the load network of a certain class of wide-band amplifier, and the decoupling network singularities move out from the origin; these amplifiers are discussed in Section 13.3. Because the positions of the decoupling network singularities are of interest only in this special case, these inequalities are assumed to be satisfied in the following analyses. Even if the inequalities are not satisfied, the position of the zero is still given exactly but the pole appears to lie closer to the origin than is actually the case. This error is insignificant because the pole has less effect on the passband than the zero. Decoupling networks have no effect on the singularity pattern of the bypass and coupling capacitors if the inequalities are satisfied, or if their own singularities lie well outside the passband.

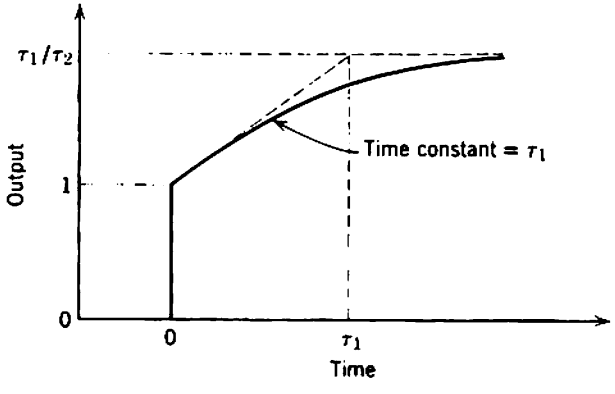


Fig. 7.25 Step response of $\psi_k(s)$ associated with a decoupling network.

7.4.3.1 Vacuum Tube

The following inequalities hold between the resistances in Fig. 7.26*a* if the amplifier is of the wide-band type discussed in Section 13.3.1.

$$R_A + R_D \ll r_A, \tag{7.41}$$

$$R_A + R_D \ll R_G. \tag{7.42}$$

The pole and the zero from the decoupling network are then

$$p = -\frac{1}{R_D C_D} \tag{7.43}$$

and

$$z = -\frac{R_A + R_D}{R_A R_D C_D}. \tag{7.44}$$

The pole and the zero lie very close together and cancel to a first approximation unless a further inequality is satisfied:

$$R_A \ll R_D. \tag{7.45}$$

The interpretation of Eq. 7.44 is that the gain increases below the frequency at which C_D breaks with $(R_A \parallel R_D)$, that is, the frequency at which the load impedance begins to increase. The gain levels off again at very low frequencies when the reactance of C_D becomes large compared with R_D (Eq. 7.43).

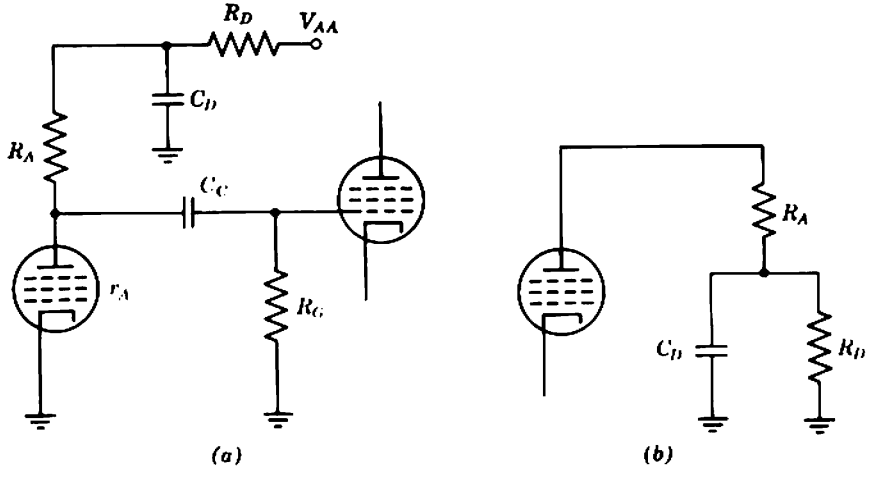


Fig. 7.26 Decoupling circuit for a vacuum tube: (*a*) complete circuit; (*b*) elemental circuit.

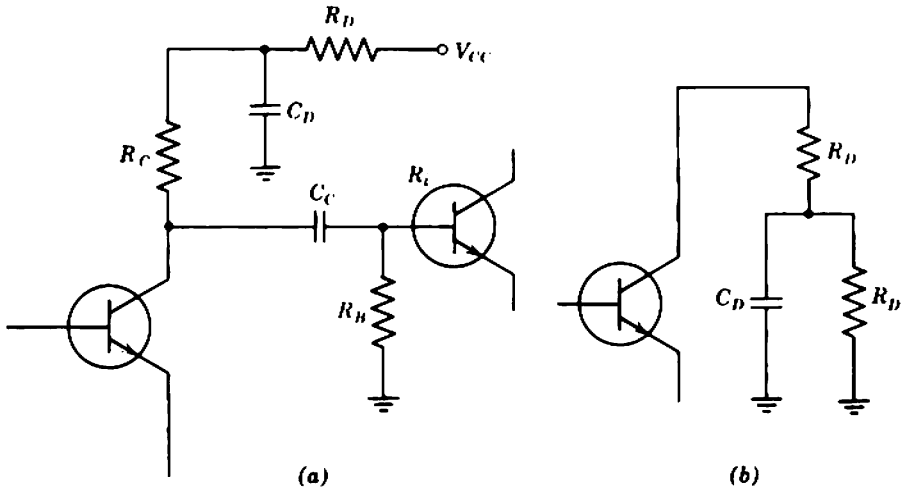


Fig. 7.27 Decoupling circuit for a transistor: (*a*) complete circuit; (*b*) elemental circuit.

7.4.3.2 Transistor

The inequality that must be satisfied in Fig. 7.27*a* before the decoupling components have a pronounced effect on the low-frequency response is

$$R_C + R_D \ll \frac{R_B R_t}{R_B + R_t}. \tag{7.46}$$

The pole and the zero are then

$$p = -\frac{1}{R_D C_D} \tag{7.47}$$

and

$$z = -\frac{R_C + R_D}{R_C R_D C_D}. \tag{7.48}$$

The pole and the zero lie very close to each other and cancel to a first approximation unless a further inequality is satisfied:

$$R_C \ll R_D. \tag{7.49}$$

These results have the same interpretation as for the vacuum-tube case.

7.5 HIGH-FREQUENCY RESPONSE OF RESISTANCE-COUPLED AMPLIFIER STAGES

The gain of an amplifier stage varies in both magnitude and phase for sinusoidal signals whose frequency lies above the mid-band range. In terms of an equivalent circuit, this variation is due to the capacitance that

exists from all points in the signal path to ground. There are three components of capacitance at any point:

(i) the capacitance of the device (or source) that precedes the point in question;
(ii) the stray capacitance of the interstage wiring and coupling components to ground;
(iii) the capacitance of the device (or load) that follows the point in question.

The device capacitances, (i) and (iii), can be further subdivided:

(a) the intrinsic capacitance associated with the mobile charge in the charge-control model;
(b) the extrinsic capacitances such as transition capacitances or direct capacitances between electrodes, lead wires, etc.

The discussion in this section is concerned only with the high-frequency response of resistance-coupled amplifier stages, that is, amplifier stages in which resistors are the only passive elements deliberately connected in shunt with the signal path. The *RC*-coupled amplifiers discussed in Section 7.4 are the most common examples; the direct-coupled amplifiers of Chapter 15 are another possibility. The shunting resistors in these amplifiers are the various supply resistors associated with the biasing networks. Coupling capacitors (which are in series with the signal path, rather than in shunt) can be neglected because the reactance of a coupling capacitor is negligible except at low frequencies. However, the stray capacitance to ground of a bulky coupling capacitor may well be the largest single component of stray capacitance associated with the interstage wiring. Further, all bypass capacitors can be neglected because their reactance is very small; the effective transfer admittance of any device is its mid-band transfer conductance G_T derived in Chapter 5.

7.5.1 General Theory

Any resistance-coupling network can be reduced to the form shown in Fig. 7.28*a* at high frequencies. In this diagram

g_m, R_o, and C_o represent the mutual conductance generator, output resistance, and output capacitance of the device preceding the coupling network, or the corresponding parameters of an external source;

R_P represents the parallel combination of all biasing resistors and is the same R_P as in the mid-band analysis of Chapter 5;

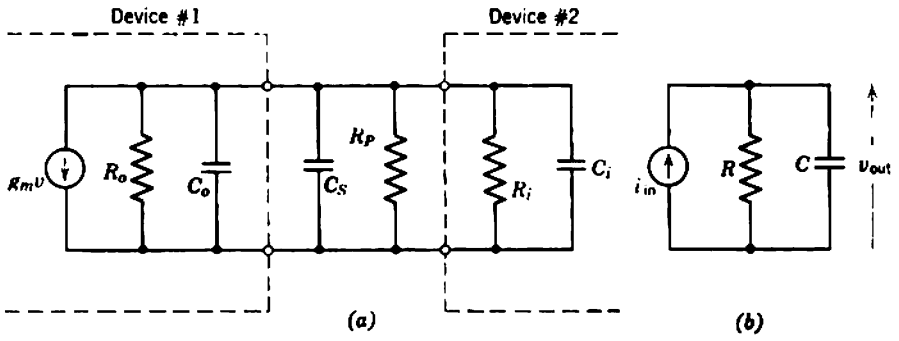


Fig. 7.28 High-frequency equivalent circuit for a resistance-coupled interstage network: (*a*) complete; (*b*) elements combined.

C_S is the stray wiring capacitance;
R_i and C_i represent the input resistance and input capacitance of the device following the coupling network, or the corresponding parameters of an external load.

The various elements can be combined to give the simple equivalent circuit of Fig. 7.28*b*. The output voltage is constant at medium frequencies for which the reactance of C is large, but falls away to zero at high frequencies as the modulus of the total load falls. The network equations give

$$v_{out} = i_{in} R \left[\frac{1}{1 + sRC} \right]. \tag{7.50}$$

It follows that the frequency-dependent term in Eq. 7.2 due to the stray capacitance that shunts a resistance-coupling network is

$$\psi_k(s) = \frac{1}{1 - s/p} = \frac{1}{1 + s/\omega_1} = \frac{1}{1 + s\tau_1}, \tag{7.51}$$

where

$$p = -\omega_1 = -\frac{1}{\tau_1} = -\frac{1}{RC}. \tag{7.52}$$

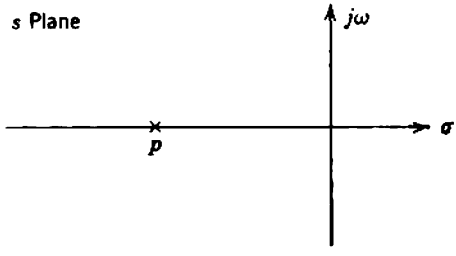


Fig. 7.29 Singularity pattern for $\psi_k(s)$ associated with a resistance-coupling network.

Figure 7.29 shows the high-frequency singularity pattern of a resistance-coupled amplifier stage; there is a single pole on the negative real axis. Figures 7.30 and 7.31 show the normalized frequency and time response $\psi_k(s)$, and the relations between the pole position, 3-dB cutoff frequency, and time constant.

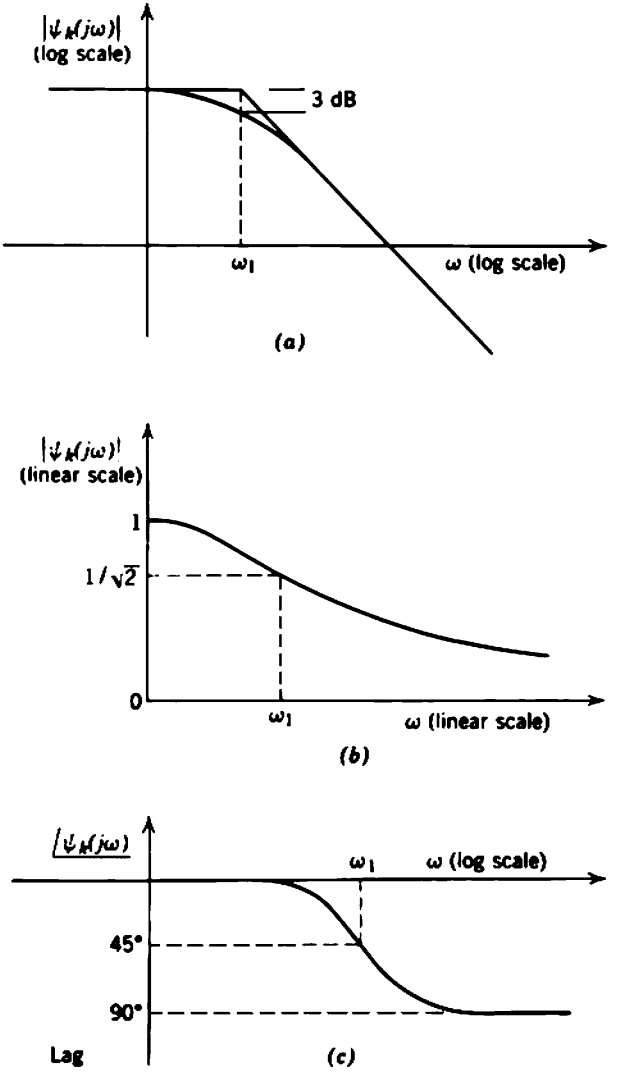


Fig. 7.30 Frequency response of $\psi_k(s)$ associated with a resistance-coupling network.

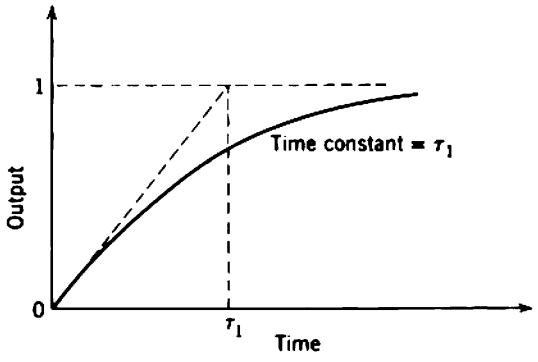


Fig. 7.31 Step response of $\psi_k(s)$ associated with a resistance-coupling network.

7.5.1.1 Vacuum Tube

The equivalent circuit for a vacuum-tube interstage network is shown in Fig. 7.32. The values of R and C in Fig. 7.28 are:

$$R = \text{(parallel combination of } r_A,\ R_A, \text{ and } R_G), \tag{7.53}$$

$$C = \text{(parallel combination of } C_o,\ C_S, \text{ and } C_i), \tag{7.54}$$

and the voltage gain from the grid of the first tube to the grid of the second is

$$\frac{v_{G2}}{v_{G1}} = -g_m R\left(\frac{1}{1 + sRC}\right). \tag{7.55}$$

A source of complication is in the values of C_o and C_i; these are not simply the anode-to-cathode and grid-to-cathode capacitances, respectively. Discussion of this follows in Section 7.5.2.

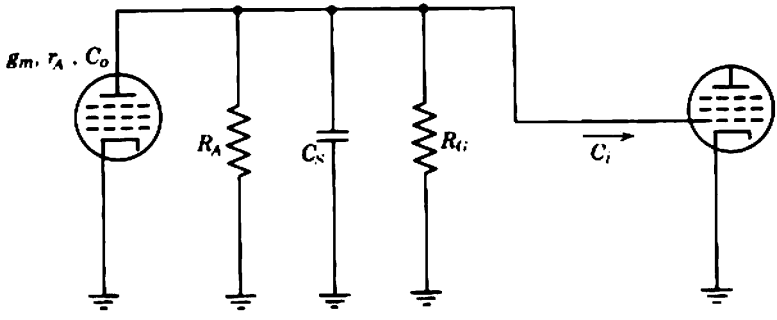


Fig. 7.32 Elemental circuit for a vacuum-tube interstage coupling network at high frequencies.

A figure of merit for an amplifier at high frequencies is the product of its mid-band gain and its 3-dB bandwidth, that is, its *gain-bandwidth product*:

$$\mathscr{GB} = |A_{Vm}| \times \omega_{co}, \tag{7.56}$$

where the cutoff frequency ω_{co} is the pole frequency. If a multistage amplifier consists of a number of identical stages, so that g_m, R, and C are the same for each, the gain-bandwidth product per stage is

$$\mathscr{GB} = (g_m R) \times \left(\frac{1}{RC}\right);$$

that is,

$$\mathscr{GB} = \frac{g_m}{C}. \tag{7.57}$$

Thus, the gain-bandwidth product of a given tube at constant quiescent conditions (i.e., constant g_m and r_A) and with constant stray wiring capacitance is a constant. The gain can be changed by varying R_A and/or R_G so as to vary R, but $\mathscr{GB}$ remains constant. The realizable $\mathscr{GB}$ of an amplifier stage can be increased somewhat by careful construction and wiring that reduce the stray capacitance. Realizable $\mathscr{GB}$ can also be increased by increasing the quiescent cathode current and hence g_m, but there remains a definite upper limit to $\mathscr{GB}$ that can be realized from a given tube:

$$\mathscr{GB}_{\max} = \frac{g_{m(\max)}}{C_i + C_o}, \tag{7.58}$$

where $g_{m(\max)}$ is the maximum g_m attainable without exceeding the tube's ratings. There is, therefore, an upper limit to the usefulness of a given vacuum-tube type as an amplifier of wide-band signals. The concept of a limiting value for $\mathscr{GB}$ is introduced in Section 2.5.1; further discussion of its significance in multistage amplifiers appears in Section 13.6.

7.5.1.2 Transistor

The exact analysis of a junction transistor interstage network at high frequencies is relatively complicated. As shown in Fig. 7.33*a*, r_B splits the shunting resistors and capacitors into two groups. Exact analysis shows that the voltage gain from the internal base of one transistor to the internal base of the next transistor has two poles, both of which lie on the negative real axis in the complex frequency plane. However, with any normal transistor and physical layout of the wiring, both the output capacitance C_o and the stray capacitance C_S are much smaller than the internal input capacitance c_C. (c_C and the characteristic frequency ω_C are defined in Section 4.7.2.1.) The two poles are therefore very far apart,

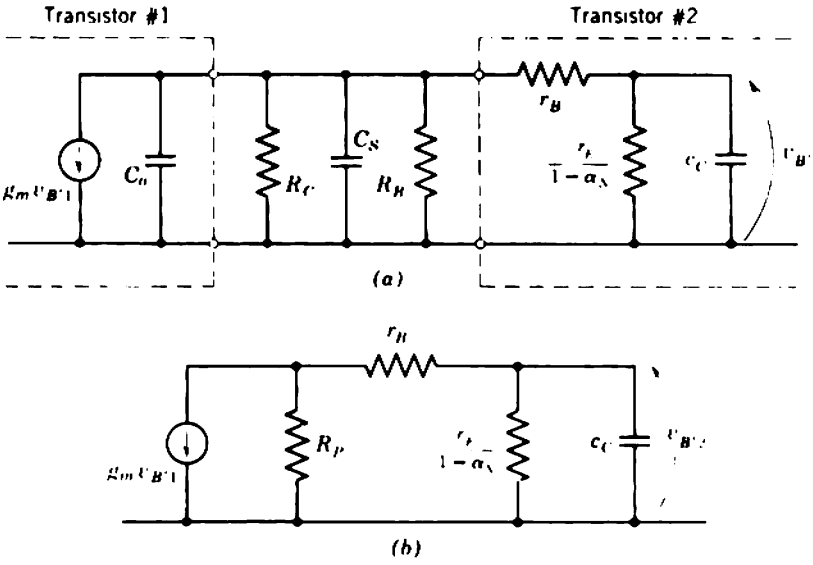


Fig. 7.33 Equivalent circuit for a transistor interstage coupling network at high frequencies: (*a*) complete; (*b*) simplified approximation.

and the high-frequency response inside the passband and for a long way outside is dominated by the innermost pole. Mathematically, it is a very good approximation to neglect C_o and C_S entirely, so as to reduce the network to the single-pole type shown in Fig. 7.33*b*. The voltage gain between the internal bases is

$$\frac{v_{B'2}}{v_{B'1}} = -g_m\left(\frac{R_P\beta_N r_E}{R_P + r_B + \beta_N r_E}\right)\left\{1 + s\left[\frac{(R_P + r_B)\beta_N r_E}{(R_P + r_B) + \beta_N r_E}\right]c_C\right\}^{-1}.$$

Since

$$g_m = \frac{\alpha_N}{r_E},$$

it follows that

$$\frac{v_{B'2}}{v_{B'1}} = -\left[\frac{\alpha_N R_P}{(R_P + r_B)/\beta_N + r_E}\right]\left\{1 + s\left[\frac{(R_P + r_B)\beta_N r_E}{(R_P + r_B) + \beta_N r_E}\right]c_C\right\}^{-1}. \quad (7.59)$$

The effective values of R and C in Fig. 7.28 are

$$R = \frac{(R_P + r_B)\beta_N r_E}{(R_P + r_B) + \beta_N r_E}, \quad (7.60)$$

$$C = c_C. \quad (7.61)$$

If a multistage transistor amplifier consists of a number of identical stages, the $\mathscr{GB}$ per stage is

$$\mathscr{GB} = |A_{Vm}| \times \omega_{co}$$
$$= \left[\frac{\alpha_N R_P}{(R_P + r_B)/\beta_N + r_E}\right]\left[\frac{(R_P + r_B) + \beta_N r_E}{(R_P + r_B)\beta_N r_E c_C}\right];$$

that is,

$$\mathscr{GB} = \frac{\alpha_N R_P}{(R_P + r_B) r_E c_C}. \tag{7.62}$$

But c_C is related to the characteristic frequency ω_C of a transistor by

$$\omega_C = \frac{\alpha_N}{r_E c_C}. \tag{7.63}$$

Thus,

$$\mathscr{GB} = \omega_C\left(\frac{R_P}{R_P + r_B}\right). \tag{7.64}$$

The realizable $\mathscr{GB}$ of a transistor is less than ω_C by an amount that depends on the ratio of R_P to r_B. Section 13.3.2 discusses the consequences in wide-band amplifiers, and derives an optimum emitter current at which $\mathscr{GB}$ is maximized.

7.5.2 Input and Output Capacitances: Miller Effect

The input capacitance to a device when it is used in an amplifier stage is not simply the sum of the capacitances from the input (control) electrode to all other electrodes. Similarly, the contribution of an active device to its own load capacitance is not simply the sum of capacitances from the output (collecting) electrode to all other electrodes. In both cases the capacitance is larger by an amount that depends on the voltage gain. This increase in capacitance is called the *Miller effect** after its discoverer and the excess capacitance is called the *Miller capacitance*.

Consider the charge-control model shown in Fig. 7.34. Three components of charge must be supplied to the control electrode in order to change the output current:

ΔQ_1 is the excess mobile charge required in the control region; this is represented by c_1;

ΔQ_{1e} is the extrinsic charge associated with changing the control-electrode voltage relative to the emitting electrode; this is represented by c_{1e}, which may be an electrostatic capacitance, a depletion layer (transition) capacitance, or a combination of the two;

* J. M. MILLER, "Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit," *Nat. Bur. Stand. Sci. Papers*, **15**, No. 351, 367, 1919–1920.

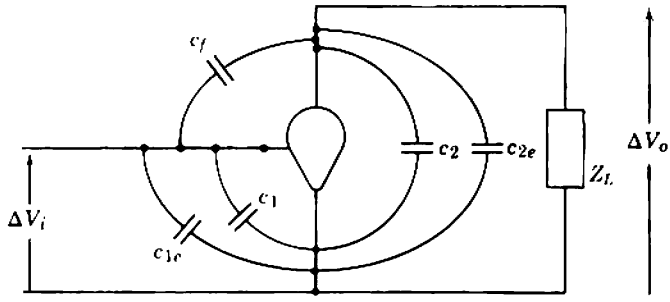


Fig. 7.34 Components of charge associated with a charge-controlled device which has voltage gain.

ΔQ_f is the charge associated with the change in voltage between the control and collecting electrodes; this is represented by c_f, which may be an electrostatic or transition capacitance, or a combination of the two.

Similarly, three components of charge must be stored in the device if the voltage of the collecting electrode is to change:

ΔQ_2 is the change in mobile charge in the conducting channel, represented by c_2;

ΔQ_{2e} is the charge associated with the stray capacitance c_{2e} between the collecting and emitting electrodes;

ΔQ_f has been considered above.

Note that the charge stored in the stray capacitance external to the device is not considered in this discussion.

The apparent input capacitance of the device is

$$C_i = \frac{\Delta Q_i}{\Delta V_i} \tag{7.65}$$

$$= \frac{\Delta Q_1 + \Delta Q_{1e} + \Delta Q_f}{\Delta V_i}$$

$$= \frac{c_1 \Delta V_i + c_{1e} \Delta V_i + c_f(\Delta V_i - \Delta V_o)}{\Delta V_i},$$

that is,

$$C_i = c_1 + c_{1e} + c_f(1 - A_V), \tag{7.66}$$

where A_V is the voltage gain:

$$A_V = \frac{\Delta V_o}{\Delta V_i}.$$

Note that A_V is negative (180° phase shift) for simple amplifiers, so that $(1 - A_V)$ is positive and greater than unity. By a similar argument, the contribution to the effective load capacitance is

$$C_o = c_2 + c_{2v} + c_f\left(1 - \frac{1}{A_V}\right). \tag{7.67}$$

7.5.2.1 General Miller Effect

Equations 7.66 and 7.67 are special cases of a general effect. In general, a bridging impedance Z_F connected between the input and output of an amplifier is equivalent to the shunt impedances shown in Fig. 7.35:

$$Z_{in} = \frac{Z_F}{1 - A_V}, \tag{7.68}$$

$$Z_{out} = \frac{Z_F}{1 - 1/A_V}. \tag{7.69}$$

Notice that Z_{out} is not the contribution of Z_F to the output impedance (the impedance looking back into the amplifier from the load). Rather, Z_{out} is the contribution to the load seen by the amplifier; the total effective load is the parallel combination of Z_{out} and any other load. Because $|A_V|$ is often much greater than unity, the term $1/A_V$ in Eq. 7.69 can usually be neglected and Z_{out} is approximately equal to Z_F. In contrast, Z_{in} may be very different from Z_F. Depending on the nature of Z_F and the magnitude and phase of A_V, Z_{in} can take on all possible impedance types—positive and negative resistance and positive and negative reactance. Two cases are of particular interest. In both, the amplifier is a single stage and Z_F is a capacitor C_F.

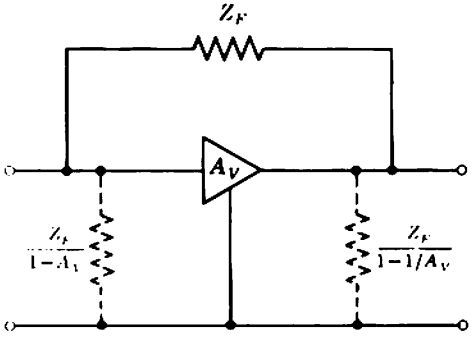


Fig. 7.35 Generalized Miller effect. The bridging impedance Z_F may be replaced by the shunt impedances at the input and output.

CASE 1. The gain increases with increasing frequency. This could be obtained by using a load impedance that increases with increasing frequency; an example is a load comprising inductance in series with resistance. Suppose the gain of the stage is of the form

$$A_V = -|A_{Vm}|(1 + s\tau). \tag{7.70}$$

The Miller input admittance follows from Eq. 7.68 as

$$Y_{in}(j\omega) = -\omega^2 C_F \tau |A_{Vm}| + j\omega C_F(1 + |A_{Vm}|), \tag{7.71}$$

where C_F is the total feedback capacitance. This admittance is the parallel combination of capacitance and frequency-dependent negative resistance plotted in Fig. 7.36.

The interpretation of a negative input resistance is that the input terminal of the amplifier is the source of energy. An LC tuned circuit is connected to the input terminal of the amplifying device in a band-pass tuned amplifier stage and, in combination with the negative input resistance, this tuned circuit may break into nondecaying oscillation. Historically, the tetrode and pentode were developed to have an anode-to-grid

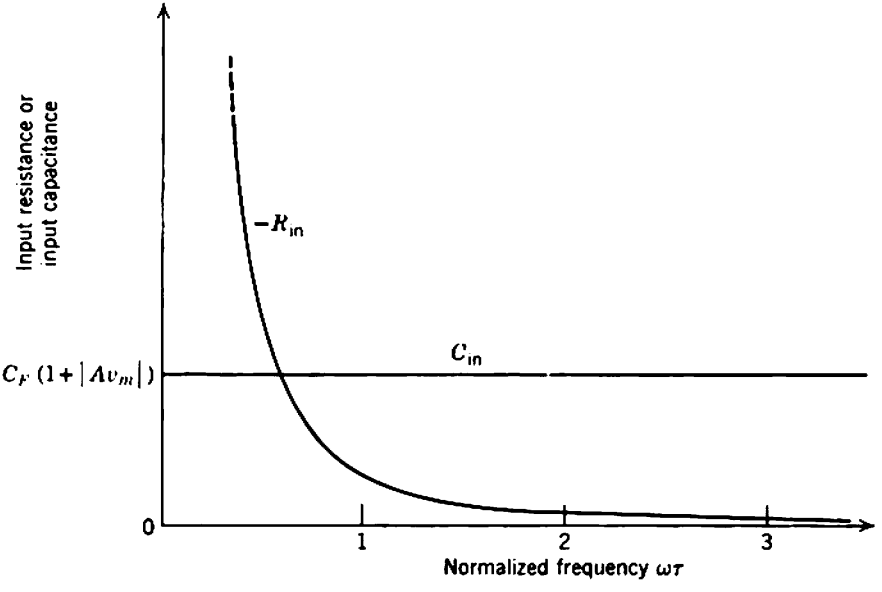


Fig. 7.36 Miller input admittance of an amplifier with rising high-frequency response:

$$A_V = A_{Vm}(1 + s\tau),$$

Z_F = output-to-input capacitance C_F.

Drawn for $A_{Vm} = -9$.

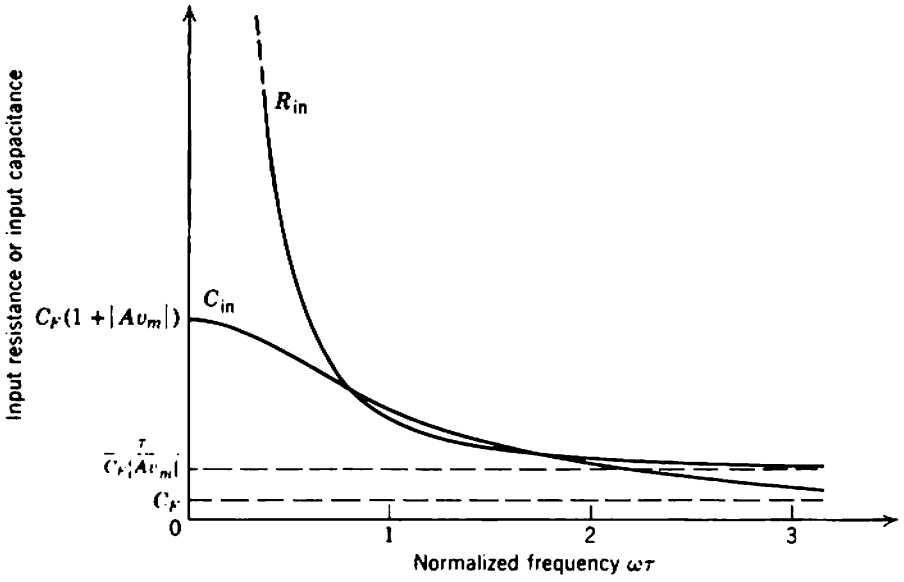


Fig. 7.37 Miller input admittance of an amplifier with falling high-frequency response:

$$A_V = A_{Vm}\left(\frac{1}{1 + s\tau}\right),$$

Z_F = output-to-input capacitance C_F.

Drawn for $A_{Vm} = -9$.

capacitance less than that of a triode, to reduce the tendency of tuned amplifiers to oscillate. Negative input resistance is unlikely to cause oscillation in a low-pass amplifier because the positive biasing resistance R_P is in shunt.

CASE 2. The voltage gain falls at high frequencies due, perhaps, to a capacitive load. Suppose that the voltage gain of the stage is of the form

$$A_V = -|A_{Vm}| \frac{1}{1 + s\tau} \tag{7.72}$$

as in the case of a resistance-coupled amplifier. The Miller input admittance follows as

$$Y_{in}(j\omega) = \left(\frac{\omega^2 C_F \tau |A_{Vm}|}{1 + \omega^2\tau^2}\right) + j\omega C_F\left[\frac{(1 + |A_{Vm}|) + \omega^2\tau^2}{1 + \omega^2\tau^2}\right]. \tag{7.73}$$

This admittance is the parallel combination of positive, frequency-dependent resistance and capacitance plotted in Fig. 7.37.

7.5.2.2 Practical Approximations

Figure 7.37 shows that the Miller input impedance of a resistance-coupled amplifier at high frequencies is a complicated function. It is perhaps worth pointing out that a vacuum tube, like a transistor, has a finite input resistance. The complete analysis of a multistage amplifier in which the input impedance of one device forms part of the load for the preceding device is a formidable undertaking. However, Fig. 7.37 shows that the Miller component of input capacitance is approximately constant at $C_F(1 + |A_{Vm}|)$ up to the cutoff frequency of the stage. Further, unless the component values are unusual in the extreme, the input resistance is very large indeed until well beyond the cutoff frequency. Therefore, it is a very good approximation to assume that the Miller input impedance is a constant capacitance over the passband. It is also a pessimistic approximation; the cutoff frequency of the preceding stage calculated using this simplified Miller impedance lies below the true value.

Vacuum-tube manufacturers do not quote separate values for c_1 and c_{1e}. Rather, these capacitances are lumped together as a so-called "input capacitance" c_{in}, which is the sum of the capacitances from the grid to all other electrodes except the anode. This is no disadvantage; as discussed in Section 3.4.2, both components are constant and independent of the quiescent operating conditions. For a triode, this "input" capacitance is the grid-to-cathode capacitance. The true input capacitance is

$$C_i = c_{\text{in}} + c_{AG}(1 + |A_{Vm}|). \tag{7.74}$$

Similarly, manufacturers quote an "output capacitance" c_{out}, rather than separate values for c_2 and c_{2e}. This "output" capacitance is the sum of the capacitances from the anode to all other electrodes except the grid, and for a triode c_{out} is the anode-to-cathode capacitance. The true contribution to the load capacitance is

$$C_o = c_{\text{out}} + c_{AG}\left(1 + \frac{1}{|A_{Vm}|}\right). \tag{7.75}$$

Usually, the term $1/|A_{Vm}|$ in Eq. 7.75 is negligible. Further, c_{AG} is so small in most pentodes that Miller effect can be neglected entirely.

The three components of input capacitance of a transistor should be considered separately. They have different laws of variation with quiescent point and, depending on circuit details, any one of them can dominate. The input capacitance follows from Eq. 7.66 as c_C first defined in Section 4.7.2.1:

$$C_i \equiv c_C = c_B + c_{tE} + c_{tC}(1 + |A_{Vm}|). \tag{7.76}$$

The components of apparent load capacitance corresponding to c_2 and c_{2e} in Eq. 7.67 are zero, so that

$$C_o = c_{tc}\left(1 + \frac{1}{|A_{vm}|}\right). \tag{7.77}$$

As in the case of a vacuum tube, the term $1/|A_{vm}|$ is often negligible.

7.6 COUPLING TRANSFORMERS

A transformer consists of two magnetically coupled coils of wire. Each coil has a finite winding resistance which is a sink for power, and the magnetic circuit is lossy also. Further, there is distributed capacitance between the turns of each coil and between the two coils. In the simplest case, the complete equivalent circuit for a transformer is as in Fig. 7.38*a*. The principal inaccuracies in Fig. 7.38*a* are

(i) representation of the distributed capacitance in each winding by a single capacitance C_1 or C_2 across its terminals,
(ii) representation of the interwinding capacitance by the single element C_3,
(iii) representation of the magnetic losses by a single resistor R_0.

The equivalent circuit is rather inconvenient in its present form, and a number of manipulations can be performed to give a more tractable representation:

1. The magnetically coupled coils are replaced by a T of inductances and an ideal transformer as shown in Fig. 7.38*b*:

$$N^2 = \frac{L_2}{L_1}, \tag{7.78}$$

$$k^2 = \frac{M^2}{L_1 L_2}. \tag{7.79}$$

The coefficient of coupling k is very close to unity for iron-cored transformers of the type considered in this chapter.

2. The secondary winding resistance and distributed capacitance are moved to the primary side of the ideal transformer (Fig. 7.38*c*).

3. The interwinding capacitance is replaced by a Π of capacitances (Fig. 7.38*d*).

Even this simplest equivalent circuit is so complex that a complete analysis is a formidable undertaking. The engineering problem with a coupling transformer is to establish the factors that determine its mid-band

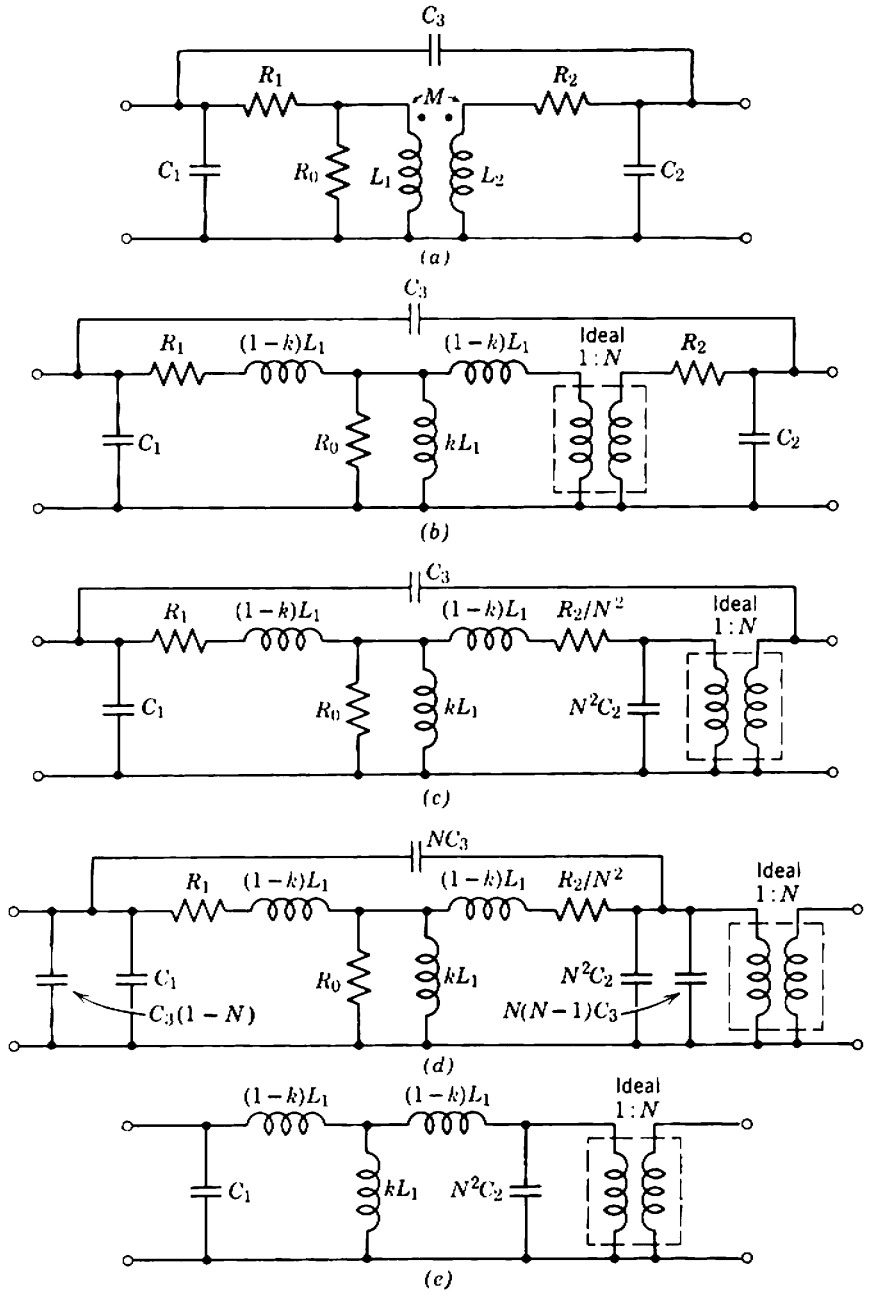


Fig. 7.38 Equivalent circuit of a coupling transformer:

L_1 and L_2 are primary and secondary inductances; M is mutual inductance; R_1 and R_2 are primary and secondary resistances; R_0 represents magnetic losses; C_1 and C_2 are primary and secondary capacitances; C_3 is interwinding capacitance;

$$k^2 = \frac{M^2}{L_1 L_2}; \qquad N^2 = \frac{L_2}{L_1}.$$

frequency range. The quantitative behavior outside this range is not of much importance, although a qualitative understanding is an aid to intelligent use of a transformer. Further simplifying approximations are therefore justifiable:

1. The mid-band range of a transformer lies between the resonant peaks due to the inductors and capacitors in the equivalent circuit. These resonances must be heavily damped if there is to be an appreciable mid-band range. The damping provided by R_0, R_1, and R_2 is quite small, so it follows that the damping provided by the external source and load resistances must be quite large. Therefore, R_0, R_1, and R_2 may be omitted in comparison.

2. The interwinding capacitance is usually much less than the distributed capacitance within the primary and secondary windings. The components of C_3 across the input and output terminals are therefore insignificant; the bridging component determines the response at the very highest frequencies, well beyond the mid-band range. All components representing C_3 are omitted from the final engineering equivalent circuit shown in Fig. 7.38*e*.

In any reasonable practical transformer the coefficient of coupling k is very close to unity. Therefore

$$kL_1 \approx L_1 \tag{7.80}$$

and

$$kL_1 \gg (1 - k)L_1. \tag{7.81}$$

Although it is strictly incorrect, it is common usage to call kL_1 the *primary inductance*; the true primary inductance is L_1. In accordance with common usage, no distinction will be made between kL_1 and L_1 in this book, and both will be called the primary inductance. The primary inductance is the inductance measured between the primary terminals when the secondary is open-circuited. It is necessary to use a low-frequency signal to make this or any other measurement of inductance, because the capacitances may introduce an appreciable error even at moderate frequencies; the apparent inductance looking into the primary terminals increases with increasing frequency until the resonant frequency is reached. The *secondary inductance* is the inductance looking into the secondary terminals with the primary open-circuited. Again, it is customary not to distinguish between L_2, N^2L_1 and kN^2L_1. The *primary leakage inductance* L_{L1} is the low-frequency inductance measured between the primary terminals when the secondary is short-circuited. Since k is close to unity,

$$L_{L1} \approx 2(1 - k)L_1. \tag{7.82}$$

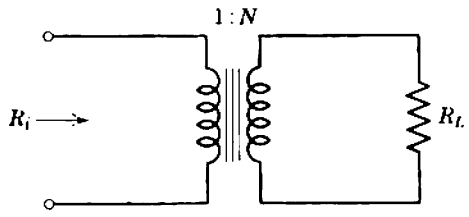


Fig. 7.39 Impedance ratio of a practical transformer.

The *secondary leakage inductance* L_{L2} is the inductance looking into the secondary with the primary short-circuited; by inspection,

$$L_{L2} = N^2 L_{L1} \approx 2N^2(1 - k)L_1. \tag{7.83}$$

Notice that the impedance ratio of a transformer is not given exactly by the square of the turns ratio. The resistance looking into the primary terminals of the transformer shown in Fig. 7.39 is not R_L/N^2, but

$$R_i = R_1 + \left(\frac{R_2 + R_L}{N^2}\right). \tag{7.84}$$

This correction is appreciable with almost all transformers. The true effective turns ratio is the square root of the secondary to primary inductance ratio (Eq. 7.78).

7.6.1 Frequency Response

Anticipating the result, a coupling transformer is unsatisfactory unless at least one out of the primary and secondary is loaded by a (relatively) small resistance. Under these circumstances, the lower 3-dB point can be found exactly and the high-frequency response can be evaluated semi-quantitatively. Sections 7.6.1.1 to 7.6.1.3 consider the response in the cases of either one or both windings being heavily loaded, and Section 7.6.1.4 indicates the behavior as both loading resistors become large.

7.6.1.1 Primary Loaded: Secondary Unloaded

Consider a transformer which is fed from a low-resistance source, and whose secondary load resistance is very high (Fig. 7.40*a*). At low frequencies the leakage inductance may be neglected because it is in series with the signal path and its reactance is small. Similarly, the stray capacitances can be neglected, because they are in shunt with the

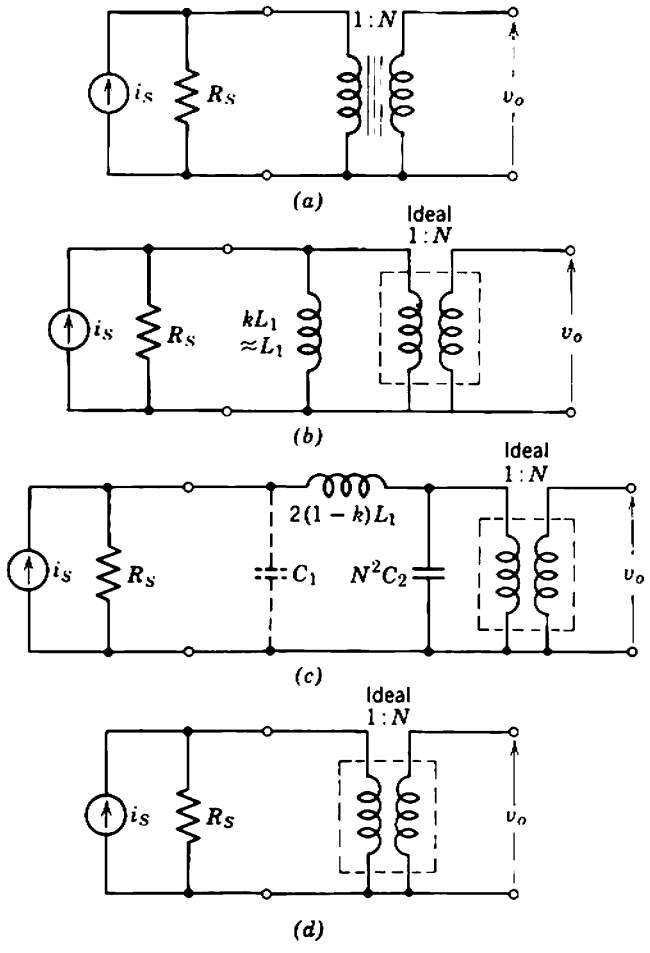


Fig. 7.40 Equivalent circuit for a transformer fed from a low-resistance source and with the secondary open-circuited: (*a*) circuit diagram; (*b*) low-frequency equivalent circuit; (*c*) high-frequency equivalent circuit; (*d*) mid-band equivalent circuit.

signal path and their reactances are large. The equivalent circuit reduces to Fig. 7.40*b* and the output voltage is given by

$$\frac{v_o}{i_S} = NR_S\left(\frac{s}{s + R_S/L_1}\right). \tag{7.85}$$

This is identical in form with the response of an *RC* coupling network. At high frequencies, the primary inductance may be neglected because it is in shunt with the signal path and its reactance is large. The primary capacitance is neglected also, because it is shunted by the small source resistance; however, C_1 does affect the response at very high frequencies

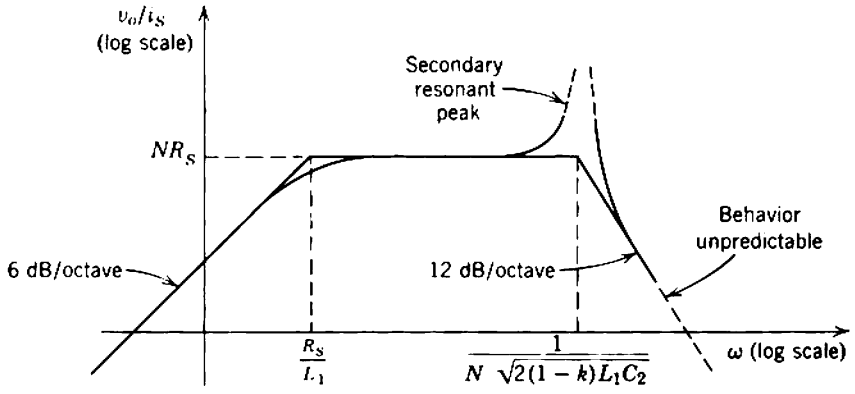


Fig. 7.41 Frequency response of a transformer fed from a low-resistance source and with the secondary open-circuited.

where its reactance becomes comparable with R_S. The equivalent circuit reduces to Fig. 7.40*c*, and the secondary resonant peak is at

$$\omega = \frac{1}{N\sqrt{2(1-k)L_1C_2}}. \tag{7.86}$$

At mid-band frequencies, all series reactances are small and all shunt reactances are large. The equivalent circuit reduces to Fig. 7.40*d*, and the output voltage is

$$\frac{v_o}{i_S} = NR_S. \tag{7.87}$$

Figure 7.41 shows the complete asymptotic frequency response and its smoothed shape. The height of the secondary resonant peak depends on R_S and the transformer losses. The behavior well beyond this peak depends on C_1 and possibly the interwinding capacitance.

7.6.1.2 Primary Unloaded: Secondary Loaded

Figure 7.42*a* shows a transformer which is fed from a high-impedance source, and which has a (relatively) small load resistor R_L connected across its secondary terminals. As in Section 7.6.1.1, the leakage inductance and winding capacitances may be neglected at low frequencies and the equivalent circuit reduces to Fig. 7.42*b*. The output voltage is given by

$$\frac{v_o}{i_S} = \frac{R_L}{N}\left(\frac{s}{s + R_L/N^2L_1}\right) \tag{7.88}$$

and, as above, this is identical in form with the response of an RC coupling network. At high frequencies, the primary inductance may be neglected, and the secondary capacitance is neglected also because it is shunted by

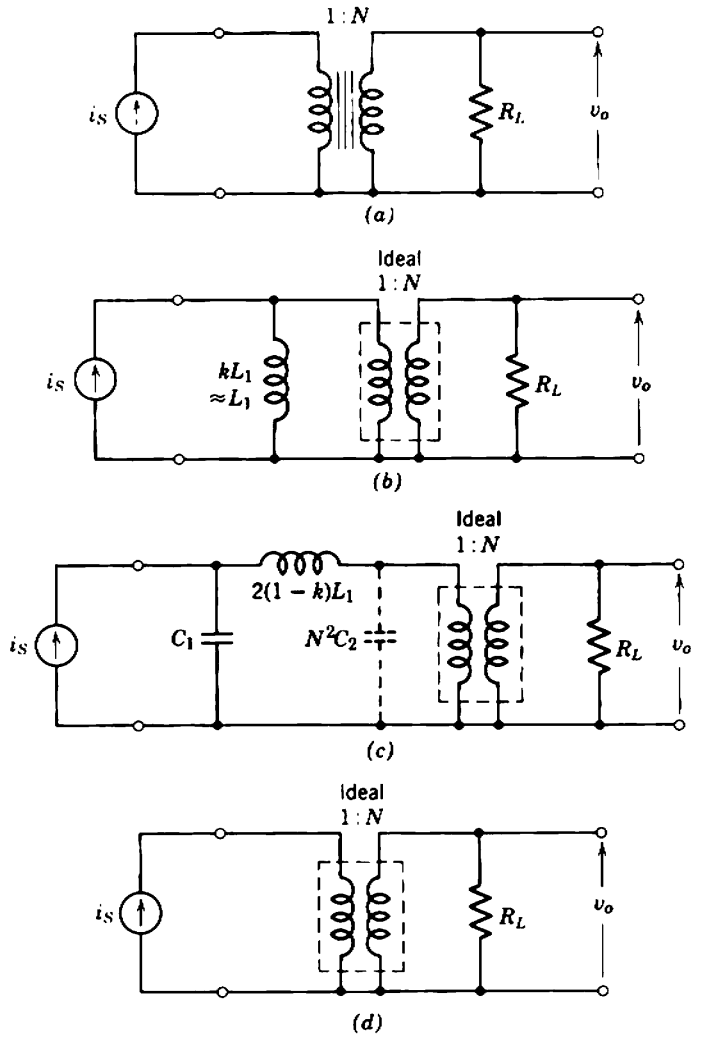


Fig. 7.42 Equivalent circuit for a transformer fed from a current source and with a low-resistance load: (*a*) circuit diagram; (*b*) low-frequency equivalent circuit; (*c*) high-frequency equivalent circuit; (*d*) mid-band equivalent circuit.

the small R_L. The equivalent circuit reduces to Fig. 7.42*c*, and the primary resonant peak is at

$$\omega = \frac{1}{\sqrt{2(1 - k)L_1C_1}}. \tag{7.89}$$

At mid-band frequencies all reactances may be neglected; the equivalent circuit reduces to Fig. 7.42*d*, and the output voltage is

$$\frac{v_o}{i_S} = \frac{R_L}{N}. \tag{7.90}$$

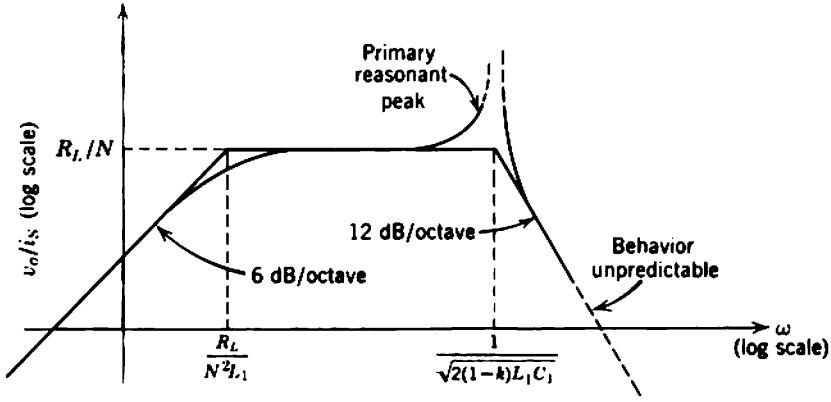


Fig. 7.43 Frequency response of a transformer fed from a current source and with a low-resistance load.

Figure 7.43 shows the complete asymptotic frequency response and its smoothed shape. In contrast to Section 7.6.1.1, the high-frequency response is dominated by the primary resonant peak; the height of this peak depends on R_L and the transformer losses. Well beyond this peak the behavior depends on C_2 and possibly the interwinding capacitance.

The winding of a transformer which has the more turns almost always has the lower resonant frequency. For example, consider a step-up transformer, that is, N is greater than unity. Comparison of Eqs. 7.86 and 7.89 shows that the secondary resonant frequency is lower than the primary resonant frequency by a factor N if the winding capacitances C_1 and C_2 are equal. However, the capacitance in a winding is approximately proportional to its number of turns, so that

$$C_1 \approx \frac{C_2}{N}.$$

It follows that the primary and secondary resonant frequencies are approximately in the ratio $1 : N^{-3/2}$.

7.6.1.3 Both Primary and Secondary Loaded

Figure 7.44*a* shows a transformer which is fed from a low-resistance source and which has a small load resistor connected across its secondary terminals. Again, the leakage inductance and winding capacitances can be neglected at low frequencies, and the equivalent circuit reduces to Fig. 7.44*b*. The output voltage is

$$\frac{v_o}{i_S} = NR\left(\frac{s}{s + R/L_1}\right), \tag{7.91}$$

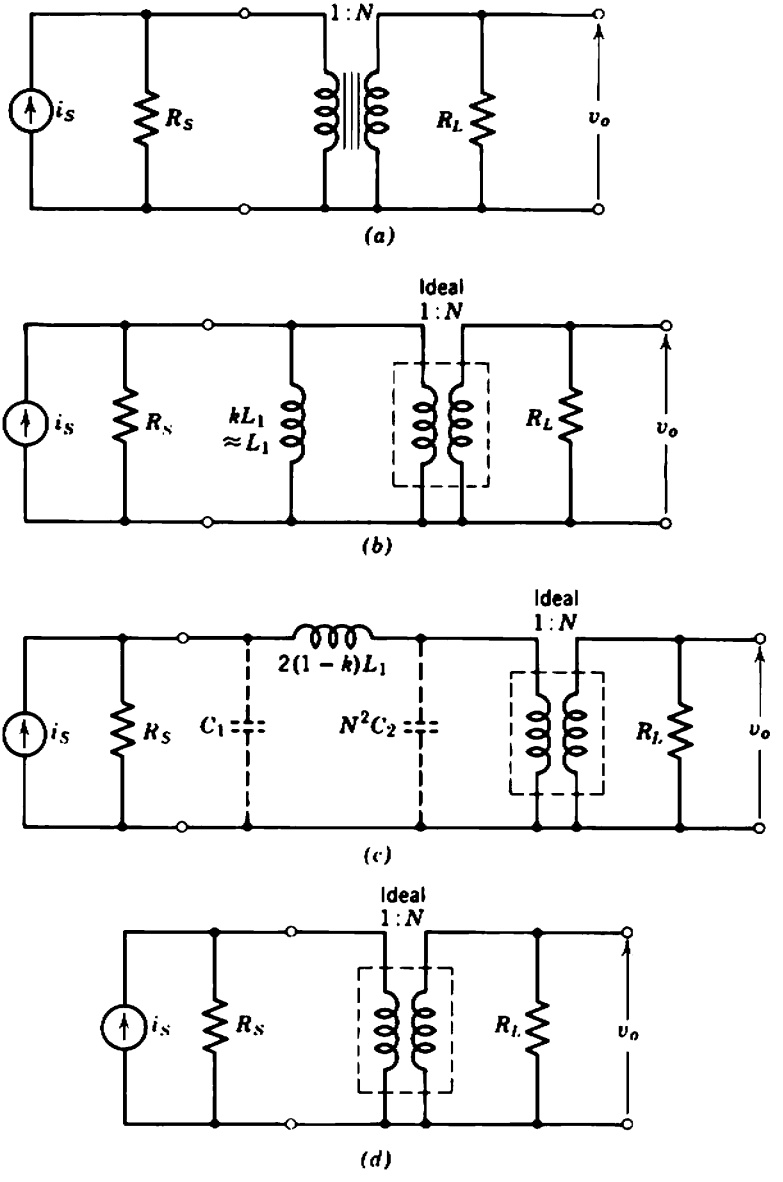


Fig. 7.44 Equivalent circuit for a transformer fed from a low-resistance source and with a low-resistance load: (*a*) circuit diagram; (*b*) low-frequency equivalent circuit; (*c*) high-frequency equivalent circuit; (*d*) mid-band equivalent circuit.

where

$$R = \frac{R_S(R_L/N^2)}{R_S + (R_L/N^2)}. \tag{7.92}$$

At high frequencies, the primary inductance may be neglected, and both winding capacitances are neglected also because they are shunted by the small source and load resistances. The equivalent circuit reduces to Fig. 7.44*c* and the output voltage is

$$\frac{v_o}{i_S} = NR\left(\frac{1}{1 + s\tau}\right), \tag{7.93}$$

where

$$\tau = \frac{2(1 - k)L_1}{R_S + R_L/N^2}. \tag{7.94}$$

There is no resonant peak; rather, the high-frequency response is a single-pole function of the same form as the response of a resistance-coupled stage. At very high frequencies C_1, C_2, and the interwinding capacitance become significant, and the response departs from the 6 dB/octave slope. At mid-band frequencies all reactances may be neglected; the equivalent circuit reduces to Fig. 7.44*d*, and the output voltage is

$$\frac{v_o}{i_S} = NR = N\left[\frac{R_S(R_L/N^2)}{R_S + (R_L/N^2)}\right]. \tag{7.95}$$

Figure 7.45 shows the complete frequency response.

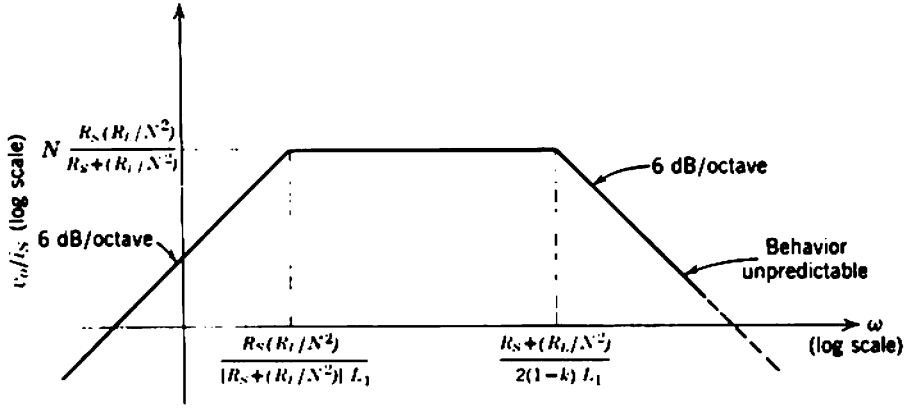


Fig. 7.45 Frequency response of a transformer fed from a low-resistance source and with a low-resistance load.

7.6.1.4 Neither Primary nor Secondary Loaded

As the source and load resistances for a transformer are increased, the mid-band frequency range shrinks and finally disappears. The limiting case in which both resistances become infinite corresponds to a transformer that is fed from a current source and whose secondary winding is open-circuited as in Fig. 7.46*a*. At low and medium frequencies the leakage inductance can be neglected in comparison with the primary inductance. The winding capacitances have no effect at low frequencies, but at medium frequencies, which would normally lie in the mid-band range, their very large reactance is comparable with the very large reactance of the primary inductance. The equivalent circuit reduces to the *LC*

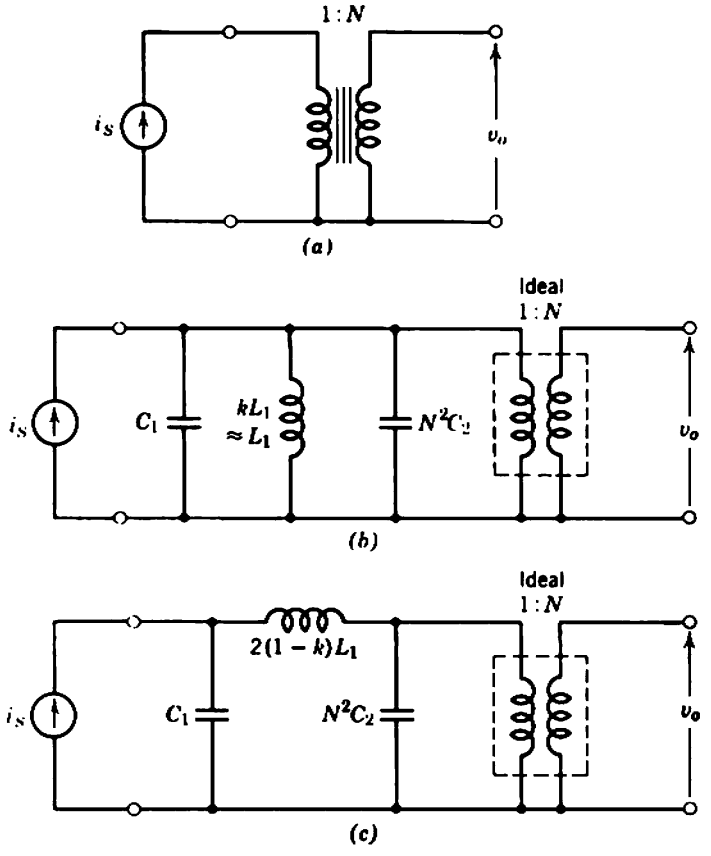


Fig. 7.46 Equivalent circuit for a transformer fed from a current source and with the secondary open-circuited: (*a*) circuit diagram; (*b*) low- and medium-frequency equivalent circuit; (*c*) high-frequency equivalent circuit.

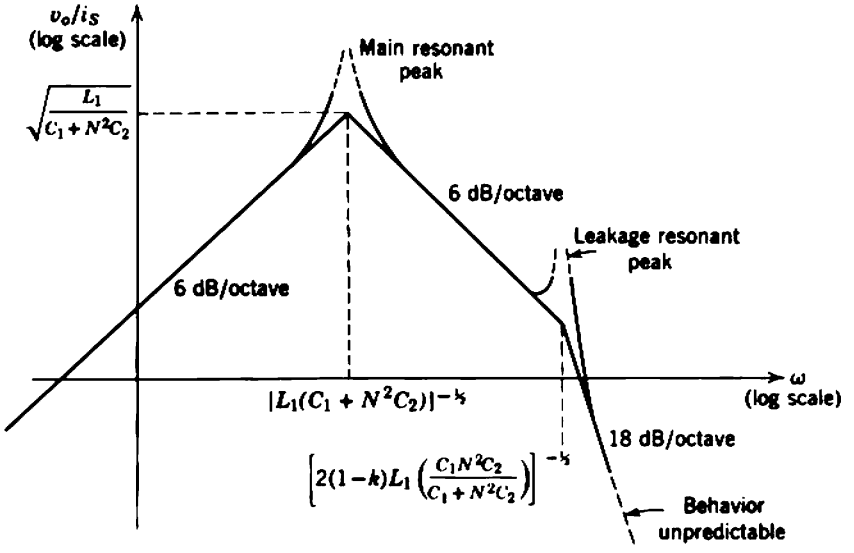


Fig. 7.47 Frequency response of a transformer fed from a current source and with the secondary open-circuited.

resonant circuit shown in Fig. 7.46*b* and the main transformer resonance is at

$$\omega = [L_1(C_1 + N^2C_2)]^{-\frac{1}{2}}. \tag{7.96}$$

At high frequencies, well beyond the main resonances, the reactance of the primary inductance becomes so large that it may be neglected, but the reactance of the leakage inductance becomes comparable with that of the winding capacitances. A second resonant circuit is formed as shown in Fig. 7.46*c*, and the leakage resonant frequency is

$$\omega = \left[2(1 - k)L_1\left(\frac{C_1N^2C_2}{C_1 + N^2C_2}\right)\right]^{-\frac{1}{2}}. \tag{7.97}$$

Figure 7.47 shows the frequency response.

7.6.2 Applications of Transformers

A sound guiding principle is to avoid the use of a coupling transformer in a low-pass amplifier unless this is inconvenient. The frequency response of a transformer is inferior to that of an *RC* coupling network, particularly at high frequencies. Further, it is well-nigh impossible to predict the response of a transformer accurately at high frequencies. Finally, transformers tend to be nonlinear because of nonlinearity in the *B–H*

curve of the core material; the elements in the equivalent circuit vary significantly with signal level. If a transformer must be used, one side must be loaded by a relatively small resistance to give satisfactory frequency response, and it is desirable that both sides should be loaded. (These comments apply to untuned transformers; unloaded tuned transformers are used extensively in band-pass amplifiers, but are not considered in this book.)

Transformers are used extensively for changing the effective source and load impedance for amplifiers. Proper choice of the effective source impedance maximizes the ratio of signal to noise in the first stage of an amplifier; this is discussed fully in Section 8.5.2. In this application, the transformer primary winding is loaded by the source resistance, and the secondary is loaded by the input resistance of the amplifier provided that this is (relatively) small. Proper choice of the effective load impedance for a vacuum tube or transistor maximizes the available signal power output, as discussed in Section 16.2.1. In this application, the load resistance damps the secondary winding, whereas the primary will be loaded by the output resistance of the amplifier if it is small.

Until about 1935 coupling transformers were used between low-μ triode amplifier stages while these triodes were the only readily available tube types. The relatively small anode resistance of these tubes loads the primary winding of the transformer and gives a satisfactory frequency response; the grid circuit of the following tube constitutes a virtual open circuit to the secondary. Substitution of the data

$$i_S = g_m v_G,$$

$$R_S = r_A,$$

into Eq. 7.87 gives the mid-band voltage gain per stage as

$$A_V = N g_m r_A = N\mu.$$

A typical transformer had N of the order of 3 so that the gain of a transformer-coupled stage was considerably greater than that of an RC-coupled stage. The step-up ratio N was limited to about 5 as a maximum, because the secondary resonant frequency became too low and entered the working frequency range of the amplifier. A typical high-quality transformer had

$$L_1 = 50 \text{ H},$$
$$N = 3,$$
$$\text{secondary resonant frequency} = 15 \text{ kHz},$$

and used in conjunction with triodes of 10-kΩ anode resistance it gave a 3-dB bandwidth of 30 Hz to 10 kHz. Transformer-coupled low-μ

triodes are rarely used now, because high-μ triodes and pentodes are readily available and can give as much, or more, gain per stage in *RC*-coupled circuits. Pentode stages should not be coupled by transformers because their anode resistance is very high and there is no loading on either winding of the transformer.

Transistor stages can be coupled together by step-down ($N < 1$) transformers to increase the available stable gain per stage. The input resistance of the transistor loads the secondary winding and gives a satisfactory frequency response. As an order of magnitude, two transformer-coupled stages are capable of as much stable gain as three *RC*-coupled stages. However, the cost of the transformers is usually greater than the cost of the extra transistors required to make up the gain of the *RC*-coupled stages. Further, the frequency response of the *RC*-coupled stages is superior.

Transformers are used extensively for driving the push-pull output stages of low-quality audio amplifiers (Section 16.5) such as those used in portable transistor radio receivers, and high-power public address systems.

7.7 PASSIVE ATTENUATORS

Often, it is necessary to provide some means for changing the gain of an amplifier. Broadly, there are three possible methods:

(i) by changing device parameters, particularly g_m,
(ii) by means of controlled negative feedback,
(iii) by means of passive attenuators.

In the first method the quiescent points of one or more of the devices in the amplifier are changed. This alters their small-signal parameters and, consequently, the gain is affected. This method of gain control is relatively imprecise and unpredictable, because of the variability of device parameters and the second-order perturbations in their variations with the quiescent point. Further, satisfactory design techniques for low-pass amplifiers aim at reducing the dependence of gain on device parameters and, therefore, frustrate this type of gain-control system. For these reasons, this method of gain control is not discussed in this book. However, it is used extensively in the *automatic gain-control systems* of band-pass amplifiers. Control of gain by means of negative feedback is discussed in Section 11.6.2.1.

The most important and generally useful method for changing the gain of a low-pass amplifier is by means of a passive attenuator. In this method the amplifier is designed to have the highest gain that is ever

required for its application and, when less gain is required, some of the available gain is "thrown away" by the attenuator. There are three classes of attenuator network:

(i) constant-impedance matched attenuators,
(ii) voltage dividers,
(iii) current dividers.

A constant-impedance attenuator is a T, Π, ladder or bridged-T network, whose design is considered in specialized books.* Almost always, a constant-impedance attenuator is switched and its loss can be varied only in discrete steps. These attenuators are precise devices and switching is necessary to preserve the correct relations between the elements; their accuracy is often better than 1%. Common impedance values are 600 Ω for audio frequencies, and 50 or 75 Ω for wide-band applications.

Voltage and current dividers can also be switched networks when precisely defined values of attenuation are required. When a quantitative measure of the attenuation is not required, however, a very simple attenuator can be made from a resistor with a variable tap—that is, a *potentiometer*. Figures 7.48*a* and *b* show circuit diagrams for a potentiometer used as a voltage and current divider, respectively. The attenuation is:

VOLTAGE DIVIDER

$$\frac{v_o}{v_S} = \frac{KRR_L}{(R_L + KR)(R_S + R) - K^2R^2}. \tag{7.99}$$

CURRENT DIVIDER

$$\frac{i_o}{i_S} = \frac{KRR_S}{(R_S + KR)(R_L + R) - K^2R^2}. \tag{7.100}$$

These equations are, of course, reciprocal in form.

Useful special cases occur when R_L and R_S respectively become very large. The output is directly proportional to K, the position of the tap on the potentiometer, so that there is a direct relation between the rotational position of the potentiometer shaft and the output:

VOLTAGE DIVIDER. If R_L becomes very large, then

$$\frac{v_o}{v_S} = \frac{KR}{R_S + R}. \tag{7.101}$$

* For example: W. C. JOHNSON, *Transmission Lines and Networks*, Chapter 12 (McGraw-Hill, New York, 1950).

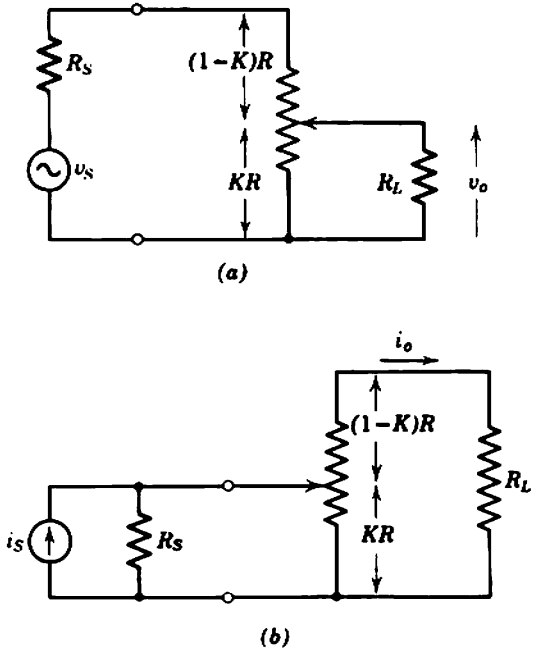


Fig. 7.48 Passive attenuators using a potentiometer: (*a*) voltage; (*b*) current.

CURRENT DIVIDER. If R_S becomes very large, then

$$\frac{i_o}{i_S} = \frac{KR}{R_L + R}. \tag{7.102}$$

These special cases can be achieved easily in practice. A voltage attenuating potentiometer is connected into the grid circuit of a vacuum tube where the input resistance is almost infinite. A current attenuating potentiometer is connected into the collector circuit of a transistor where the output resistance is almost infinite.

7.8 CONCLUDING COMMENTS ON DESIGN

It is now possible to formulate a general procedure for designing simple low-pass amplifiers. First, the small-signal circuit theory of Chapter 5 is used to find the number and types of stages required and their quiescent points. A biasing system is then designed using the principles outlined in Chapter 6. Next, the bypass and coupling capacitor values necessary to give the desired low-frequency response are found from Section 7.4. Finally, the high-frequency response is calculated from Section 7.5. If

noise or distortion is of interest, the material covered in Chapters 8 or 9 must be employed.

Notice that bypass and coupling capacitor values are *designed* to give a required low-frequency response whereas, at least for the simple amplifiers considered so far, the high-frequency response is *calculated*. The limitation to the dynamic response at low frequencies is artificial, because the coupling and bypass capacitors are introduced to simplify the design of the biasing system and reduce the number of fixed supply potentials which are required. In principle, the low-frequency response can be extended indefinitely by increasing the size of these capacitors. In contrast, the high-frequency limitation to the dynamic response is fundamental and arises from the limitation of gain-bandwidth product. Methods are discussed in Chapter 13 by which the high-frequency response can be extended about one octave. If this is still insufficient, there is no alternative to a complete redesign of the amplifier; the gain per stage must be reduced so that the individual stage bandwidths are increased, and the number of stages must be increased to restore the mid-band gain. Even this procedure may become inadequate. It is possible to show that there is a definite upper limit to the bandwidth which can be obtained with a given gain and a given device type, even if the number of stages is increased without limit. The whole matter is discussed in Section 13.6.

Practical amplifier design is illustrated by reference to the two following examples, one each of vacuum-tube and transistor circuits.

7.8.1 Vacuum-Tube Amplifier

The vacuum-tube amplifier is required to have a mid-band voltage gain about 1000 into an external load of 22 kΩ. The input resistance is not to be less than 500 kΩ, and the lower 3-dB point is to be below 50 Hz.

Figure 7.49 shows the proposed circuit diagram with the tube types and resistor values marked. It is left as an exercise for the reader to show that the mid-band gain is about 1000 and that the biasing system will achieve the following quiescent points:

FIRST STAGE

$$I_A = 1.35 \text{ mA},$$
$$I_S = 0.35 \text{ mA},$$
$$V_{SK} = 100 \text{ V};$$

therefore,

$$g_m = 1.1 \text{ mA/V};$$

therefore,

$$A_{Vm} = 100.$$

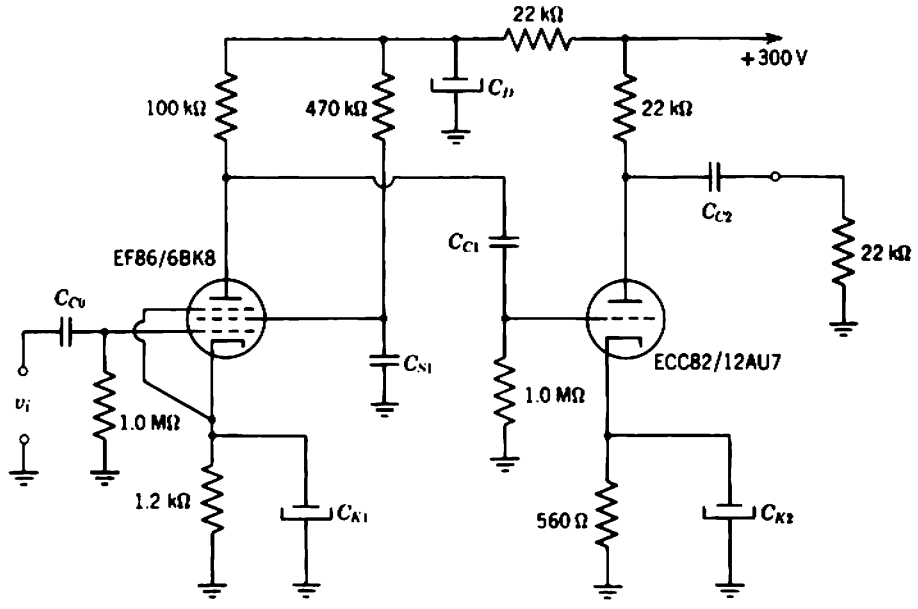


Fig. 7.49 Circuit diagram for × 1000 vacuum-tube amplifier.

SECOND STAGE

$$I_K = 7 \text{ mA},$$
$$V_{AK} = 140 \text{ V};$$

therefore,

$$g_m = 2 \text{ mA/V},$$

and

$$r_A = 8 \text{ k}\Omega;$$

therefore,

$$A_{Vm} = 10.$$

The values of the coupling and bypass capacitors are to be designed, and the high-frequency response is to be calculated.

7.8.1.1 Low-Frequency Response

The most pessimistic design approach assumes that all zeros are at the origin and uses the simplest approximations for the poles. This results in a design in which the capacitor values are larger than necessary; the true 3-dB point is somewhat below 50 Hz. However, no capacitor values are likely to be inconveniently large and the final design is entirely practical. There are six poles in the circuit, neglecting the decoupling network. These poles are set nominally coincident at

$$s = -100 \text{ radian/sec},$$

that is, about 16 Hz cutoff frequency for each capacitor. The total response is therefore 18 dB down at 16 Hz. It is shown in Section 12.1.2.2.2 that the response is 3 dB down at 50 Hz for this pole configuration. The capacitor values and reference equations are:

C_{C0}	(Eq. 7.33)	10 nF
C_{K1}	(Eq. 7.11)	10 μF
C_{S1}	(Eq. 7.18)	68 nF
C_{C2}	(Eq. 7.33)	10 nF
C_{K2}	(Eq. 7.11)	22 μF
C_{C3}	(modified Eq. 7.33)	470 nF

The decoupling components are chosen to give $\times 100$ attenuation of 100-Hz hum on the supply rail:

$$C_{D1} \quad \text{(Eq. 6.74)} \qquad 8.2\ \mu\text{F}.$$

The factor of 100 is much greater than the gain of the second stage so that any hum at the output is due to the anode supply voltage for this stage, rather than the first stage. Because Eq. 7.45 is not satisfied, the decoupling components have no effect on the frequency response. Notice that the screen is supplied from the decoupled supply rail; the decoupled supply is nominally 260 V.

7.8.1.2 High-Frequency Response

There are insufficient data to calculate the high-frequency response of the amplifier; the additional information required is:

(i) the source capacitance,
(ii) the source resistance,
(iii) the load capacitance.

However, the position of the interstage network pole, the only pole determined by the amplifier, can be found. The elements in Eqs. 7.53 and 7.54 that determine the high-frequency response are:

r_A is very large, from manufacturers' data,
$R_A = 100\ \text{k}\Omega$,
$R_G = 1.0\ \text{M}\Omega$;

therefore,

$R = 91\ \text{k}\Omega$.

$C_o = 6$ pF, from manufacturers' data and Eq. 7.75,
$C_S = 20$ pF, an intelligent guess,
$C_i = 17$ pF, from manufacturers' data and Eq. 7.74;

therefore,

$$C = 43 \text{ pF}.$$

Thus,

$$p = -\frac{1}{RC} = -2.5 \times 10^5 \text{ radian/sec}$$

and

$$f_{co} = 41 \text{ kHz}.$$

7.8.2 Transistor Amplifier

The transistor amplifier is designed for use between a 20-Ω ribbon (magnetic) microphone and a 22-kΩ external load. The voltage ratio

$$\frac{\text{output voltage}}{\text{open-circuit microphone voltage}}$$

is to be 1000 ± 20%, and the 3-dB passband is 30 Hz to 15 kHz.

Figure 7.50 shows the circuit diagram. The transistor is a rather high-gain audio-frequency type; its average parameters at $I_E = 1$ mA and $V_{CE} = 6$ V are:

$$r_B = 200\ \Omega, \qquad \tau_1 = 150 \text{ nsec},$$
$$r_C = 100 \text{ k}\Omega, \qquad c_{tE} = 200 \text{ pF},$$
$$\beta_N = 80, \qquad c_{tC} = 40 \text{ pF}.$$

It is left as an exercise for the reader to check that the specified gain and stability will be achieved if

$$r_E = 40\ \Omega,$$

so that

$$I_E = 625\ \mu\text{A},$$

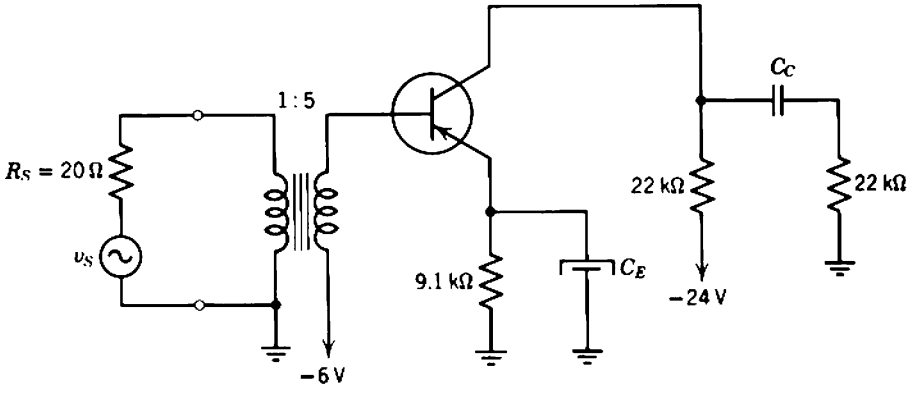


Fig. 7.50 Circuit diagram for × 1000 transistor amplifier.

and also that the biasing system achieves this quiescent point. Notice that the voltage gain of the transistor itself is about 200. This is quite large, and the simple expression for $G_T[= \alpha_N/(r_B/\beta_N + r_E)]$ is in error by about 10% due to the finite value of r_C. A correction has been made in the design by reducing r_E about 10% below the value calculated from the simple expression.

7.8.2.1 Low-Frequency Response

There are three low-frequency poles resulting from the presence of the emitter bypass capacitor, the output coupling capacitor, and the input transformer. These three poles are set coincident on the negative real axis at

$$s = -80 \text{ radian/sec}$$

to give a 3-dB point rather below the specified 30 Hz (see Section 12.1.2.2.2). The calculated values are:

C_E	(Eq. 7.24)	330 μF
C_C	(modified Eq. 7.36)	270 nF
L_1	(Eq. 7.85)	250 mH

The 3.5-kΩ load on the transformer secondary due to the input resistance of the transistor is neglected—compare Eqs. 7.85 and 7.91.

7.8.2.2 High-Frequency Response

The high-frequency response is likely to be dominated by the secondary resonance of the transformer, particularly since the effective secondary capacitance is increased by the large input capacitance of the transistor. An intelligent guess is to set the transformer resonant frequency one octave above the specified 3-dB frequency, and it is likely that the specification will be met.

The input capacitance of the transistor is given by Eq. 7.76 as

$$C_i \equiv c_C = \frac{\tau_1}{r_E} + c_{iE} + c_{iC}(1 - A_{Vm})$$
$$= 12.0 \text{ nF}.$$

This is so large that any likely secondary capacitance is negligible in comparison. Setting the desired secondary resonant frequency as

$$\omega = 2 \times 10^5 \text{ radian/sec}$$

in Eq. 7.86 gives

$$2(1 - k)L_1 = 83 \,\mu\text{H}.$$

But L_1 has already been calculated to be 250 mH. Therefore,

$$1 - k = 1.7 \times 10^{-4}.$$

Chapter 8

Simple Practical Amplifiers II: Noise

The discussion of small-signal amplifiers given in Chapters 5 and 7 assumes that there is a linear relationship between the input (excitation) and output (response) signals. This assumption may be invalidated in practical amplifiers by the presence of corrupting signals which arise from

(i) noise generated within the amplifier,
(ii) distortion produced by nonlinearity in the amplifying devices or other circuit components.

Noise is important when the signals are small in magnitude, whereas distortion becomes pronounced when the signal amplitude is large. Consequently, amplifier performance can be described adequately by simple linear circuit theory only within a range of signal amplitudes. Neither boundary of this range is fixed precisely; the tolerable level of the corrupting signal depends on the requirements of particular systems, and typical values for a number of familiar systems are given in Table 7.1. With few exceptions, the two sources of corrupting signals can be considered separately; usually the designer needs to consider only the noise performance of the input stage of an amplifier and the distortion performance of the output stage.

The purpose of this and the following chapter is to develop techniques for determining the magnitude of the corrupting signals present in the simple amplifier circuits discussed in preceding chapters. This involves, first, a general discussion of methods of calculating and specifying the magnitude of noise that originates within the amplifier, second, some comments on the possibility of additional noise-like signals due to hum,

microphony, and radio-frequency interference, and third, a discussion of methods for calculating and specifying distortion. These general discussions also provide a basis for determining the magnitude of corrupting signals in more advanced circuits discussed in subsequent chapters. At a design level rather than an analytical level, the material of these chapters provides a useful guide to practical methods for minimizing noise and distortion.

8.1 NETWORK THEORY OF NOISY AMPLIFIERS

It is shown in Sections 3.2, 4.4, and 4.8.5 that the noise performance of an amplifying device can be represented by superimposing noise current generators or noise voltage generators on the small-signal equivalent circuit. These generators represent the various significant physical sources of noise. Generators that represent thermal and shot noise have uniform power spectra, whereas generators representing flicker or induced grid noise have outputs that vary with frequency. Because of the multiplicity of noise generators in this physical noise model, it is inconvenient for direct use in calculations involving both signals and noise. A simple noise model can be obtained if the effect of the physical noise generators is represented by two equivalent noise generators situated at either the input or output of a noiseless equivalent circuit. The transformation from a physical noise model to the equivalent model with two external noise generators involves consideration of the frequency dependence of the physical noise generators and the equivalent circuit impedances. In general, the magnitudes of the external noise generators are functions of frequency, and are said to represent the *spot noise* at the frequency concerned. Because the spot noise generator magnitudes may be functions of frequency, calculations of total noise involve frequency-response methods even in cases where the signal performance is computed on a time-response basis. A further advantage of the model with equivalent external noise generators stems from the ease of determining their magnitudes by simple, direct measurements. The transformation from a physical noise model to the equivalent model is as follows:

Consider the spot noise model shown in Fig. 8.1*a*. This consists of a small-signal equivalent circuit for an intrinsic amplifying device (contained in the rectangle) and physical noise current generators $d(i_{N1}^2)$, $d(i_{N2}^2)$, and $d(i_{NB}^2)$. This model applies over a particular incremental bandwidth to the vacuum triode, the vacuum pentode (with screen bypassed to cathode), the bipolar transistor, and the field-effect transistor. In each case the physical significance of the generators is given in Table 8.1. The

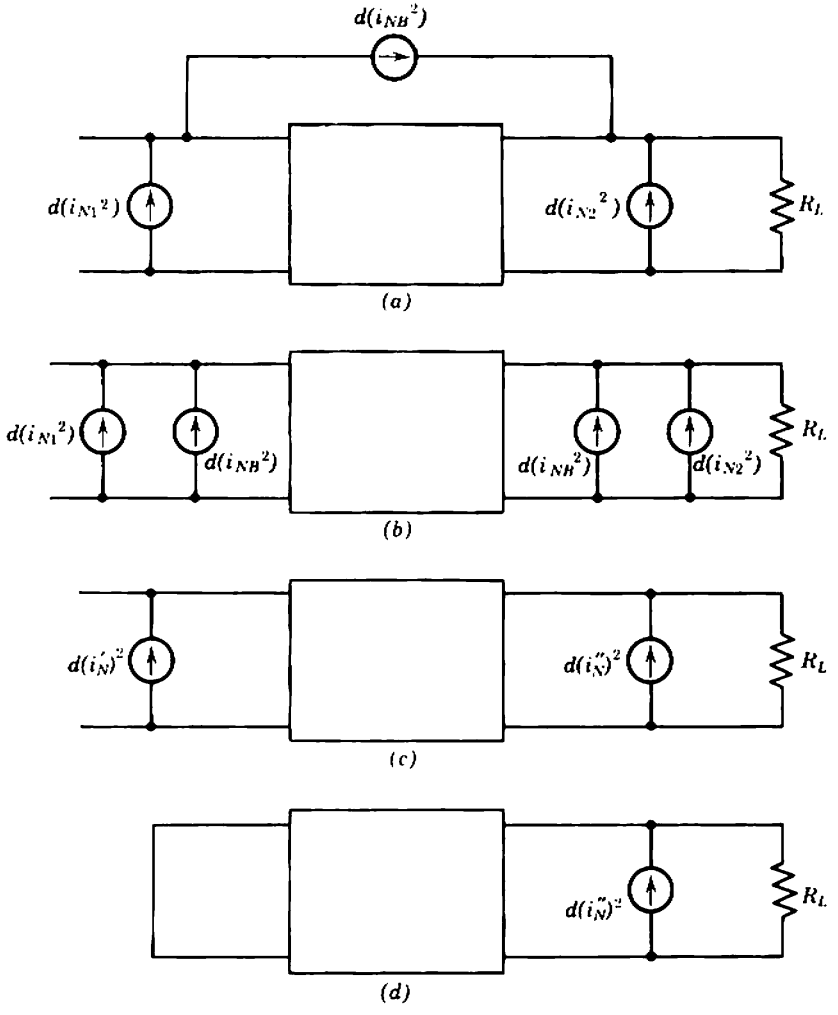


Fig. 8.1 Representation of physical noise sources by equivalent noise generators at the input.

Table 8.1 Noise Types in Active Devices

Device	$d(i_{N1}^2)$	$d(i_{N2}^2)$	$d(i_{NB}^2)$
Vacuum tube	Induced grid and grid current noise	Shot and flicker noise	Zero
Bipolar transistor	Shot and flicker noise	Shot noise	Shot and flicker noise
Field-effect transistor	Shot and flicker noise	Thermal and flicker noise	Zero

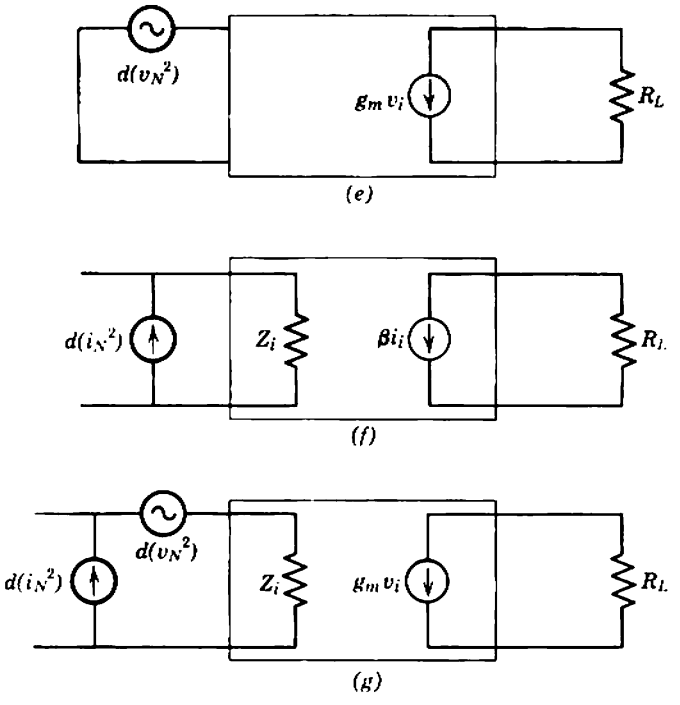


Fig. 8.1 *(continued)*

bridging generator $d(i_{NB}^2)$ can be replaced by two equal generators $d(i_{NB}^2)$ at the input and output as shown in Fig. 8.1*b*. Since $d(i_{NB}^2)$ is not correlated with $d(i_{N1}^2)$ or $d(i_{N2}^2)$, the generators at input and output can be combined to give $d(i_N')^2$ and $d(i_N'')^2$ (Fig. 8.1*c*) respectively where, at the particular frequency being considered,*

$$d(i_N')^2 = d(i_{N1}^2) + d(i_{NB}^2) \tag{8.1}$$

and

$$d(i_N'')^2 = d(i_{N2}^2) + d(i_{NB}^2). \tag{8.2}$$

If the input terminals are short-circuited (Fig. 8.1*d*) there can be no component of output noise current due to $d(i_N')^2$. Therefore, the noise current in the load resistance due to the device depends only on $d(i_N'')^2$. It is convenient to replace this noise current generator in the output circuit by a noise voltage generator in the input circuit. If the device has a mutual conductance g_m, the output noise current due to $d(i_N'')^2$ (with the

* The squares of $d(i_N')^2$ and $d(i_N'')^2$ are placed outside the brackets for easier reading.

input short-circuited) will be produced by a noise voltage generator

$$d(v_N^2) = \frac{d(i_N'')^2}{g_m^2}, \tag{8.3}$$

positioned as shown in Fig. 8.1*e*. Because $d(i_N'')^2$ is predominantly shot noise, and because g_m is independent of frequency over the range of validity of the charge-control model, $d(v_N^2)$ is approximately independent of frequency. It should be noted that the total noise current in R_L will include its own thermal noise contribution; this contribution is usually negligible in comparison with other components.

If the input terminals are open-circuited, the noise current in R_L due to the device depends on both $d(i_N')^2$ and $d(i_N'')^2$. The equivalent noise current generator at the input [$d(i_N)^2$ in Fig. 8.1*f*] is

$$d(i_N^2) = d(i_N')^2 + \frac{d(i_N'')^2}{|\beta(f)|^2}, \tag{8.4}$$

where $\beta(f)$ is the short-circuit current gain of the device. Formally, Eq. 8.4 is incorrect because the noise current $d(i_{NB}^2)$ occurs in both $d(i_N')^2$ and $d(i_N'')^2$. These two generators are therefore partially correlated and should not be summed as uncorrelated sources. However, the contribution of $d(i_{NB}^2)$ via $d(i_N')^2$ exceeds the contribution via $d(i_N'')^2$ by the current gain β^2. Because β is large over the useful frequency range, the correlation due to $d(i_{NB})^2$ can be neglected and Eq. 8.4 is valid for practical calculations. Notice that β is frequency-dependent for all charge-controlled devices, due to the input capacitance; in general

$$\beta(f) = g_m Z_i(f). \tag{8.5}$$

As frequency increases, the magnitude of β falls and $d(i_N^2)$ increases.

The spot noise model of Fig. 8.1*g*, consisting of a noiseless small-signal equivalent circuit with the two spot noise generators at the input, can be used for the limiting cases of zero and infinite source impedance. If the input terminals are short-circuited, $d(i_N^2)$ is ineffective and $d(v_N^2)$ determines the noise output. If the input terminals are open-circuited, $d(v_N^2)$ is ineffective and $d(i_N^2)$ determines the noise output. It is shown in the next section that, provided correlation between $d(v_N^2)$ and $d(i_N^2)$ can be neglected, the model represents the spot noise performance of the device for all source impedances. The total spot noise is found by summing the mean-square components due to $d(v_N^2)$ and $d(i_N^2)$. This simple model is of great practical utility and, with few exceptions, is used throughout the remainder of this book. The values of $d(v_N^2)$ and $d(i_N^2)$ can be found theoretically in many instances (Section 8.2) or can be deduced simply by measuring the noise output of the device with the input open-circuited and short-circuited, respectively.

8.1.1 Correlation

Throughout the above manipulations of the noise model of a device, it is assumed that noises are summed by adding squares of voltages or currents. This procedure is valid only if the noises originate from separate physical processes, that is, if the individual noise voltages and currents are not correlated. In the calculation of $d(i_N^2)$, this condition is not satisfied at all steps in the development. The bridging noise generator $d(i_{NB}^2)$ is split in two, added to $d(i_{N1}^2)$ and $d(i_{N2}^2)$, and the resultant generators $d(i_N')^2$ and $d(i_N'')^2$ are in turn rearranged and summed. During this final step, components of noise due to $d(i_{NB}^2)$ are added as if uncorrelated and an error is introduced. Expansion of Eq. 8.4 in terms of the physical noise generators gives the incorrect expression:

$$d(i_N^2) = d(i_{N1}^2) + \frac{d(i_{N2}^2)}{\beta^2} + d(i_{NB}^2)\left(1 + \frac{1}{\beta^2}\right). \qquad (8.6a)$$

The problem of correlation occurs when an attempt is made to manipulate the rather complicated physical noise model into a form more suited to network calculations. There are ways around the problem, but all are cumbersome. A basic (though brute-force) method for finding the total spot noise output is to sum the mean-square outputs due to all physical noise generators in the system. Alternatively, mathematical methods are available for summing partially correlated noise, and these are treated in specialized books. Fortunately, correlation is insignificant in many instances, and the simplifying manipulations of the physical noise model are therefore valid.

As an example, calculate the noise current into a short-circuit load of the device in Fig. 8.1*a* when its input terminals are open-circuited. Equation 8.6*a* is the formally incorrect result obtained by neglecting correlation. Using the accurate, brute-force method, the output currents due to the physical noise generators are:

$$\text{current due to } d(i_{N1}) = \beta d(i_{N1}),$$

$$\text{current due to } d(i_{N2}) = d(i_{N2}),$$

$$\text{current due to } d(i_{NB}) = (\beta + 1)d(i_{NB}).$$

The total noise current is therefore

$$d(i_{No}^2) = \beta^2 d(i_{N1}^2) + d(i_{N2}^2) + (\beta + 1)^2 d(i_{NB}^2).$$

The spot noise current generator at the input is

$$d(i_N^2) = \frac{d(i_{No}^2)}{\beta^2};$$

that is,

$$d(i_N^2) = d(i_{N1}^2) + \frac{d(i_{N2}^2)}{\beta^2} + d(i_{NB}^2)\left(1 + \frac{2}{\beta} + \frac{1}{\beta^2}\right). \qquad (8.6b)$$

By comparing Eqs. 8.6*a* and *b*, we find that the error from neglecting correlation is $2/\beta$ in the last term of Eq. 8.6*a*. This error is small provided $|\beta|$ is large.

The special conditions will now be derived under which correlation between $d(v_N^2)$ and $d(i_N^2)$ can introduce an appreciable error when the model of Fig. 8.1*g* is used to represent the noise performance of a device. Consider Fig. 8.2*a*, and suppose that the device spot noise is accurately represented by uncorrelated generators $d(i_N')^2$ and $d(i_N'')^2$. [This implies that correlation due to $d(i_{NB}^2)$ is negligible, as above.] The noise current in R_L due to the device is given correctly by

$$d(i_{No}^2) = \left[g_m^2 d(i_N')^2\left(\frac{R_S R_i}{R_S + R_i}\right)^2 + d(i_N'')^2\right]\left(\frac{R_o}{R_o + R_L}\right)^2. \qquad (8.7a)$$

Additional components exist due to the thermal noise of the source and of R_L itself, but will be ignored.

Using the representation of Fig. 8.2*b*, the noise current in R_L is

$$d(i_{No}^2) = \left[g_m^2 d(i_N^2)\left(\frac{R_S R_i}{R_S + R_i}\right)^2 + g_m^2 d(v_N^2)\left(\frac{R_i}{R_S + R_i}\right)^2\right]\left(\frac{R_o}{R_o + R_L}\right)^2. \qquad (8.8)$$

Any difference between Eqs. 8.7*a* and 8.8 appears because $d(v_N^2)$ and

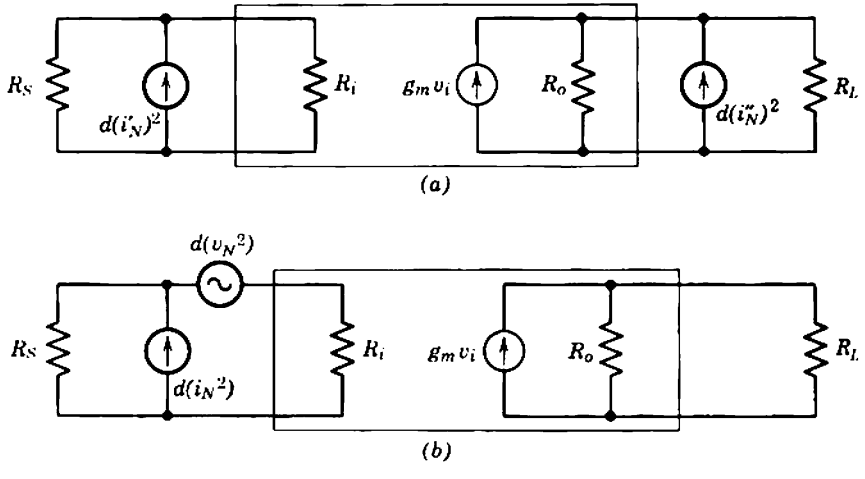


Fig. 8.2 Equivalence of noise representations with finite source resistance.

$d(i_N^2)$ are partly correlated due to the presence of $d(i_N'')^2$ in each. Substitution from Eqs. 8.3 and 8.4 into Eq. 8.8 gives

$$d(i_{No}^2) = \left[g_m^2 d(i_N')^2\left(\frac{R_S R_i}{R_S + R_i}\right)^2 + \frac{g_m^2}{\beta^2} d(i_N'')^2\left(\frac{R_S R_i}{R_S + R_i}\right)^2 + d(i_N'')^2\left(\frac{R_i}{R_S + R_i}\right)^2\right]\left(\frac{R_o}{R_o + R_L}\right)^2$$

and since

$$\beta = g_m R_i,$$

it follows that

$$d(i_{No}^2) = \left[g_m^2 d(i_N')^2\left(\frac{R_S R_i}{R_S + R_i}\right)^2 + d(i_N'')^2\left(\frac{R_S^2 + R_i^2}{R_S^2 + 2R_S R_i + R_i^2}\right)\right]\left(\frac{R_o}{R_o + R_L}\right)^2. \quad (8.7b)$$

Thus, the error in Eq. 8.7*b* is a term $2R_S R_i$ in the coefficient of $d(i_N'')^2$. This error is small unless two conditions are satisfied simultaneously: R_S and R_i must be of the same order of magnitude and β must be so small that $d(i_N')^2$ is not much greater than $d(i_N'')^2/\beta^2$.

At least one of these conditions is not satisfied in the majority of practical amplifiers. It is therefore valid to represent device spot noise by the model of Fig. 8.1*g* and neglect the correlation between $d(v_N^2)$ and $d(i_N^2)$. In amplifiers of the type discussed in Chapters 5 and 7, each device is fed from a low-impedance source; the first condition is not satisfied, and $d(v_N^2)$ alone provides a reasonably complete measure of device spot noise. In the feedback amplifiers which will be discussed in Chapter 11, there are impedance mismatches at the input of each stage such that R_S is either very small or very large in comparison with R_i; again the first condition is not satisfied, and either $d(v_N^2)$ or $d(i_N^2)$ provides a measure of device spot noise. In rare cases, R_S and R_i are comparable in magnitude, and both $d(v_N^2)$ and $d(i_N^2)$ contribute to the output noise; however, useful amplifying devices have large values of β at mid-band frequencies, and the second condition is not satisfied.

At high frequencies, $|\beta|$ falls and correlation between $d(v_N^2)$ and $d(i_N^2)$ is more likely to be significant. One example is given in Section 8.5.3.2 —the noise of a wide-band amplifier with a capacitive source. In this and other cases where correlation between $d(v_N^2)$ and $d(i_N^2)$ could cause appreciable errors, the recommended approach is to go back to the physical noise model and sum the squares of output noise due to the uncorrelated generators in this model.

Correlation is ignored throughout the remainder of this book unless the contrary is expressly stated. This is a sound, engineering approximation, except at high frequencies.

8.2 SPOT NOISE VOLTAGE AND NOISE CURRENT GENERATORS AT THE INPUT

It is shown in the preceding section that the spot noise performance of an amplifying device can be represented by two noise generators $d(v_N^2)$ and $d(i_N^2)$ at the input. This procedure simplifies the assessment of noise performance but has the disadvantage that $d(v_N^2)$ and $d(i_N^2)$ may not be related directly to the fundamental physical processes occurring in the device; in many cases, however, $d(v_N^2)$ and $d(i_N^2)$ are simply related to the dominant physical noise mechanism. It is the purpose of this section to evaluate $d(v_N^2)$ and $d(i_N^2)$ in terms of the basic device parameters and operating conditions.

8.2.1 Vacuum Triode

In a triode there is no bridging current $d(i_{NB}^2)$, and consequently there is no problem of correlation between $d(i_N')^2$ and $d(i_N'')^2$. These noise current generators are given by Eqs. 3.57, 3.61, 3.62, and 3.69 as

$$d(i_N')^2 = [2q \sum (I_G) + 4k(1.431T_K)g_i]\, df \tag{8.9}$$

and

$$d(i_N'')^2 = \left(2q\Gamma^2 I_A + \frac{KI_A^2}{f}\right) df. \tag{8.10}$$

The generator $d(i_N')^2$ represents components of noise due to dc and induced currents in the grid. All components of dc grid current contribute additive components of shot noise, even though the currents themselves may cancel; this is indicated by the summation sign. For a slightly negative grid bias there are components of $d(i_N')^2$ due to both electron and positive ion streams, and consequently the grid current noise is high. If the grid is taken more negative, the electron component vanishes and $\sum (I_G)$ may be identified with the net grid current. Thus, as the magnitude of negative grid voltage increases, the grid current noise is reduced. The generator $d(i_N'')^2$ represents shot and flicker noise on the cathode current. The space charge reduction factor Γ^2 can be replaced from Eqs. 3.58 and 3.59 to give

$$d(i_N'')^2 = \left[4k(0.644T_K)g_m + \frac{KI_A^2}{f}\right] df. \tag{8.11}$$

Over most of the useful frequency range, $d(i_N'')^2$ is usually the dominant component of tube noise. At very high frequencies (beyond 20 or 30 MHz) induced grid noise appears and $d(i_N')^2$ becomes significant. (There

are exceptions to this statement, notably applications involving very high impedance sources for which the dc grid current term makes an appreciable contribution to over-all noise.) Thus the input voltage generator $d(v_N^2)$ $[= d(i_N'')^2/g_m^2]$ usually accounts for noise performance except at high frequencies. It follows from Eq. 8.11 that

$$d(v_N^2) = \left[\frac{4k(0.644T_K)}{g_m} + \frac{KI_A^2}{fg_m^2}\right] df. \tag{8.12a}$$

For most triodes T_K is about 1000°K, so Eq. 8.12*a* may be written as

$$d(v_N^2) \approx \left[\frac{3.5 \times 10^{-20}}{g_m} + \frac{A}{f}\right] df, \tag{8.12b}$$

where the empirical constant A is a few times 10^{-13} for a low-noise triode at 1 mA anode current, but may be several orders of magnitude larger for triodes of poor noise performance.

The input current generator $d(i_N^2)$ becomes significant at high frequencies. The input impedance of a tube at (reasonably) high frequencies is

$$Z_i(f) = \frac{1}{j2\pi fc_{in}} \tag{8.13}$$

and the current gain is therefore

$$\beta(f) = \frac{g_m}{j2\pi fc_{in}}. \tag{8.14}$$

Substitution into Eqs. 8.4, 8.9, and 8.10 yields

$$d(i_N^2) = \left[2q\sum(I_G) + 4k(1.431T_K)g_i + \frac{2q\Gamma^2 I_A(2\pi fc_{in})^2}{g_m^2} + \frac{KI_A^2(2\pi fc_{in})^2}{g_m^2 f}\right] df. \tag{8.15}$$

The terms represent dc grid current noise, induced grid noise, shot noise, and flicker noise, respectively. At the high frequencies where $d(i_N^2)$ normally becomes significant, the first and last terms are negligible. Consequently, replacing Γ^2 gives

$$d(i_N^2) \approx 4k(0.644T_K)\left[\frac{20}{9}g_i + \frac{(2\pi fc_{in})^2}{g_m}\right] df. \tag{8.16}$$

It is instructive to compare the two high-frequency components of $d(i_N^2)$. Equation 3.41 gives the input conductance of a triode as

$$g_i \approx \frac{g_m\omega^2\tau_1^2}{20}.$$

Also, charge-control theory (Eq. 2.53) shows that

$$\tau_1 = \frac{c_1}{g_m}(1 + u_1),$$

and since $u_1 = 0.5$ for all vacuum tubes (Section 3.1.1.2),

$$\tau_1 = \frac{3c_1}{2g_m} = \frac{3c_{GK}}{2g_m}$$

where c_{GK} is the intrinsic input capacitance of a tube. Therefore,

$$g_i = \frac{9\omega^2 c_{GK}^2}{80g_m} = \frac{9(2\pi f)^2 c_{GK}^2}{80g_m},$$

and substitution into Eq. 8.16 gives

$$d(i_N^2) \approx 4k(0.644T_K)\frac{(2\pi f)^2}{g_m}\left(\frac{c_{GK}^2}{4} + c_{in}^2\right) df.$$

Recall from Section 3.1.2.2 that the input capacitance c_{in} of a tube is larger than the intrinsic capacitance c_{GK}. Therefore, the induced grid noise term in Eq. 8.16 is much smaller than the shot noise term, and can usually be neglected in low-pass amplifiers. However, c_{in} in a band-pass amplifier is tuned out by a resonating inductance, so the induced grid noise dominates. Detailed discussion of noise in tuned amplifiers is beyond the scope of this book; as a guiding rule a simplified physical noise model such as Fig. 8.1*c* is preferable to the artificial model (Fig. 8.1*g*).

8.2.2 Vacuum Pentode

A fairly complete discussion of noise sources in a pentode is given in Section 3.2.2. It is shown that shot and flicker components of noise are minimized by bypassing the screen to the cathode so that no noise voltage can be developed at the screen. For these conditions, partition components of shot and flicker noise exist, and it follows from Eqs. 3.66 and 3.68 that

$$d(i_N'')^2 = \left[2q\Gamma_A^2 I_A + K_A\left(\frac{I_A^2}{f}\right)\right] df. \tag{8.17}$$

Replacing Γ_A^2 from Eq. 3.67 gives

$$d(i_N'')^2 = \left[4kg_m(0.644\,T_K) + 2q\left(\frac{I_S I_A}{I_K}\right) + K_A\left(\frac{I_A^2}{f}\right)\right] df. \tag{8.18}$$

Finally, the expression for $d(v_N^2)$ which specifies the noise performance at low and mid-band frequencies follows from Eq. 8.3 as

$$d(v_N^2) = \left[\frac{4k(0.644T_K)}{g_m} + \frac{2q}{g_m^2}\left(\frac{I_S I_A}{I_K}\right) + \frac{K_A I_A^2}{f g_m^2}\right] df. \tag{8.19a}$$

Adopting the notation and numerical values of Eq. 8.12*b*,

$$d(v_N^2) \approx \left[\frac{3.5 \times 10^{-20}}{g_m} + \left(\frac{2.5 \times 10^{-19}}{g_m^2}\right) I_S + \frac{A}{f}\right] df, \qquad (8.19b)$$

where A is about 4×10^{-13} for a low-noise tube. The second term is the increase in noise of a pentode over a triode.

The high-frequency noise performance is specified by $d(i_N^2)$, given by Eq. 8.16 as in the triode case with Γ_A replacing Γ. Approximately,

$$d(i_N^2) \approx \left[4k(0.644T_K)\left(\frac{20}{9} g_i + \frac{(2\pi f c_{in})^2}{g_m}\right) + 2q\left(\frac{I_S I_A}{I_K}\right)\left(\frac{(2\pi f c_{in})^2}{g_m^2}\right)\right] df. \qquad (8.20)$$

8.2.3 Bipolar Transistor

It follows from the discussion of transistor noise in Section 4.4 that the two noise current generators $d(i_N')^2$ and $d(i_N'')^2$ for the intrinsic transistor in Fig. 8.3*a* are

$$d(i_N')^2 = \left\{2q[(1 - \alpha_N)I_E + I_{CO}] + \frac{KI_E^c}{f}\right\} df \qquad (8.21)$$

and

$$d(i_N'')^2 = 2q(\alpha_N I_E + I_{CO})\, df. \qquad (8.22)$$

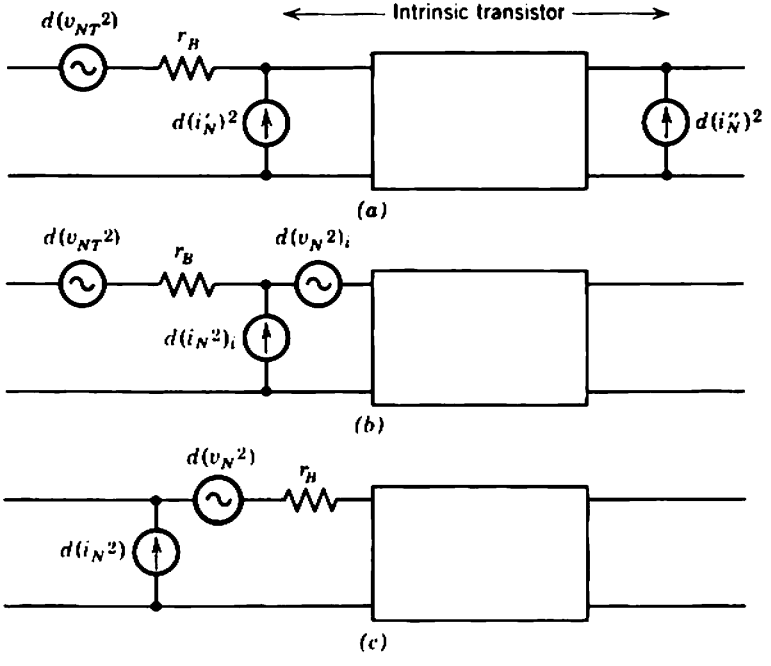


Fig. 8.3 Equivalent noise generators at the input of a transistor.

The constants K and c are associated with the flicker noise; the value of K varies over wide limits, but c lies between 1 and 2. Formally, the two noise current sources are partially correlated because of the presence of the I_{CO} term in each. However, I_{CO} is so small compared with I_E that it is usually valid to write

$$d(i_N'')^2 = 2q(\alpha_N I_E)\, df. \tag{8.23}$$

The two input spot noise generators for the intrinsic transistor, $d(v_N^2)_i$ and $d(i_N^2)_i$ in Fig. 8.3*b*, can be determined as follows. From Eq. 8.3

$$d(v_N^2)_i = 2q\left(\frac{\alpha_N I_E}{g_m^2}\right) df,$$

and since

$$g_m = \frac{\alpha_N}{r_E} \approx \frac{1}{r_E} = \frac{qI_E}{kT}$$

it follows that

$$d(v_N^2)_i \approx 4kT\left(\frac{r_E}{2}\right) df. \tag{8.24}$$

From Eq. 8.4

$$d(i_N^2)_i = \left\{2q\left[(1 - \alpha_N)I_E + I_{CO} + \frac{\alpha_N I_E}{|\beta(f)|^2}\right] + \frac{KI_E^c}{f}\right\} df, \tag{8.25}$$

where, from Eqs. 4.69 and 4.73, the short-circuit current gain $\beta(f)$ is

$$\beta(f) = \frac{i_C(f)}{i_B(f)} = \beta_N\left(\frac{1}{1 + jf\beta_N/f_T}\right). \tag{8.26}$$

For a complete transistor, the thermal noise voltage source $d(v_{NT}^2)$ associated with the extrinsic base resistance r_B is often significant and must be included in the input noise generators $d(v_N^2)$ and $d(i_N^2)$ shown in Fig. 8.3*c*. Provided r_B is small compared with the input impedance of the intrinsic transistor,

$$d(v_N^2) \approx d(v_{NT}^2) + d(v_N^2)_i = 4kT\left(\frac{r_E}{2} + r_B\right) df \tag{8.27}$$

and

$$d(i_N^2) \approx d(i_N^2)_i. \tag{8.28}$$

The power spectrum of $d(v_N^2)$ is uniform over the normal range of operating frequency, whereas $d(i_N^2)$ varies with frequency at both the low and the high ends of the spectrum. At low frequencies

$$d(i_N^2) \approx \left(\frac{KI_E^c}{f}\right) df. \tag{8.29}$$

At mid-band frequencies

$$d(i_N^2) \approx 2q[(1 - \alpha_N)I_E + I_{CO}]\,df. \tag{8.30}$$

At high frequencies

$$d(i_N^2) \approx 2q\left[\frac{\alpha_N I_E}{|\beta(f)|^2}\right] df \approx \left[2qI_E\left(\frac{f}{f_T}\right)^2\right] df. \tag{8.31}$$

8.2.4 Field-Effect Transistor

An introductory discussion of noise sources in field-effect transistors is given in Section 4.8.5, and it follows that the input and output noise current generators are approximately

$$d(i_N')^2 = 2q[\textstyle\sum(I_G)]\,df \tag{8.32}$$

and

$$d(i_N'')^2 = \left[4k(\nu T)g_m + \frac{KI_D}{f}\right] df. \tag{8.33}$$

Using Eqs. 8.3 and 8.4, we find that the noise generators at the input become

$$d(v_N^2) = \left[\frac{4k(\nu T)}{g_m} + \frac{KI_D}{g_m^2 f}\right] df = \left[\frac{4k(\nu T)}{g_m} + \frac{KV_P^2}{4I_{DO}f}\right] df \tag{8.34}$$

and

$$d(i_N^2) = \left[2q\textstyle\sum(I_G) + \dfrac{4k(\nu T)(2\pi f C_i)^2}{g_m} + \dfrac{KI_D(2\pi f C_i)^2}{g_m^2 f}\right] df, \tag{8.35}$$

where the second form of Eq. 8.34 follows from Eq. 4.141.

As with the vacuum triode, $d(v_N^2)$ alone accounts for f.e.t. noise unless the frequency is high or the circuit configuration unusual. Unlike the triode, however, the flicker component of $d(v_N^2)$ dominates over the white component throughout the useful frequency range. In view of this dominance of flicker noise in f.e.t.s, the safest procedure may be to rely on manufacturers' data or practical measurements for information on the magnitudes of $d(v_N^2)$ and $d(i_N^2)$ and their variation with operating conditions and frequency. Notice that Eq. 8.34 predicts that the dominant flicker component of $d(v_N^2)$ is independent of quiescent operating current, and this is confirmed in practice.

8.3 OTHER SOURCES OF NOISE

There are sources of noise in an amplifier other than the active devices. All resistors contribute their theoretically unavoidable thermal noise and, in addition, there is excess noise due to technological imperfections in some types of resistors and capacitors.

8.3.1 Noise Contributions of Biasing Resistors

A convenient analytical approach is to associate the noise from biasing resistors with a device in such a way that the resistor noise appears as a modification of the $d(v_N^2)$ and $d(i_N^2)$ generators of the device. The noise performance of any amplifying stage is thus specified by two generators in the same positions, and techniques can be developed for handling this general noise model. It is advantageous to associate a collecting-electrode supply resistor with the following device rather than with the device whose electrode current it supplies.

Figure 8.4*a* shows a device with a resistor R_1 shunting the input signal path to ground. In a practical example, R_1 represents the parallel combination of the control-electrode supply resistor for the device with the collecting-electrode supply resistor of the preceding device. Associated with R_1 is a noise current generator

$$d(i_N^2)_{R_1} = \frac{4kT}{R_1}\,df. \tag{8.36}$$

By inspection, the noise performance of the stage is specified by the two generators

$$d(v_N^2) = d(v_N^2)_D, \tag{8.37}$$

$$d(i_N^2) = d(i_N^2)_D + \frac{4kT}{R_1}\,df + \frac{d(v_N^2)_D}{R_1^2}, \tag{8.38}$$

where $d(v_N^2)_D$ and $d(i_N^2)_D$ are the spot noise generators of the device alone. The noise contribution of R_1 falls as R_1 is made larger. Notice that $d(v_N^2)$ and $d(i_N^2)$ are correlated to a small extent because $d(v_N^2)_D$ appears in each; usually, however, the last term in Eq. 8.38 is negligible, so that the total spot noise current generator at the input reduces to the sum of the device noise current generator with the resistor noise.

Figure 8.4*b* shows a device with a resistor R_2 in series with the signal input. Associated with R_2 is a noise voltage generator

$$d(v_N^2)_{R_2} = 4kTR_2\,df. \tag{8.39}$$

By inspection, the noise performance of the stage is specified by the two generators

$$d(v_N^2) = d(v_N^2)_D + d(i_N^2)_D R_2^2 + 4kTR_2\,df, \tag{8.40}$$

$$d(i_N^2) = d(i_N^2)_D. \tag{8.41}$$

Notice that $d(v_N^2)$ and $d(i_N^2)$ are correlated to a small extent due to the presence of $d(i_N^2)_D$ in each. Examples of a resistor deliberately added in series with a signal path are rare; a few possible situations are discussed in

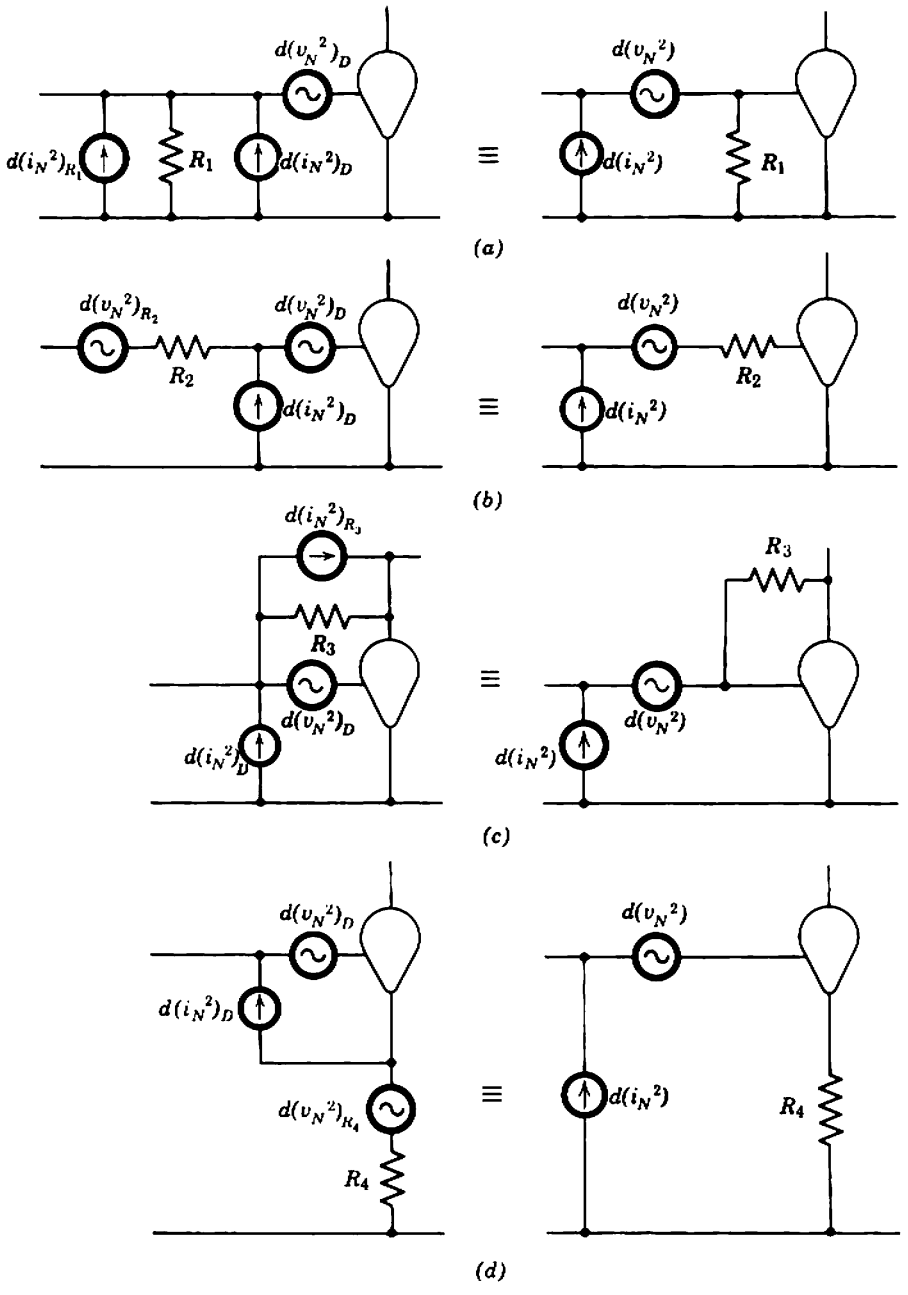


Fig. 8.4 Equivalent representations of noise due to biasing resistors.

Section 7.2.3. An example of an unavoidable resistor in series with the signal path is the base resistance r_B of a transistor. Strictly, the effect of r_B is specified by two equations of the form of Eqs. 8.40 and 8.41. However, r_B is so small that Eq. 8.27 is a good practical approximation for $d(v_N^2)$; the total spot noise voltage generator is the sum of the device noise voltage generator with the resistor noise. As R_2 is made much smaller, its noise contribution falls.

Figure 8.4*c* shows a device with a resistor R_3 connected between its input and output electrodes. This circuit configuration occurs in the collecting-electrode-feedback biasing systems of Section 6.3. It also occurs very commonly in shunt-feedback stages (Section 11.1). The noise current generator associated with R_3 is similar to the bridging generator $d(i_{NB}^2)$ in Fig. 8.1*a*, and contributes to both $d(v_N^2)$ and $d(i_N^2)$. By inspection,

$$d(v_N^2) = d(v_N^2)_D + \frac{4kT}{g_m^2 R_3}\,df, \tag{8.42}$$

$$d(i_N^2) = d(i_N^2)_D + \frac{4kT}{R_3}\left(1 + \frac{1}{\beta}\right)^2 df + \frac{d(v_N^2)_D}{R_3^2}\left(1 + \frac{1}{\beta}\right)^2, \tag{8.43}$$

where β is the current gain of the device. The noise contributions of R_3 to $d(v_N^2)$ and $d(i_N^2)$ are correlated, but the contribution to $d(v_N^2)$ is so small that it can usually be neglected. The noise falls as R_3 is made larger.

Figure 8.4*d* shows a device with a resistor R_4 connected into its common lead. This circuit configuration occurs in the series-feedback stage discussed in Section 11.1. It also occurs in a modified form in the emitting-electrode-feedback biasing circuits of Section 6.1; this modified form is discussed in Section 8.3.1.1. The noise performance of the stage is specified by the generators

$$d(v_N^2) = d(v_N^2)_D + d(i_N^2)_D R_4^2 + 4kTR_4\,df, \tag{8.44}$$

$$d(i_N^2) = d(i_N^2)_D. \tag{8.45}$$

These results are identical to Eqs. 8.40 and 8.41 for Fig. 8.4*b*. The noise is reduced by reducing R_4.

8.3.1.1 Cathode and Emitter Bias Resistors

A problem of some interest is the noise contribution of emitting-electrode supply resistors. As shown in Fig. 8.5, these resistors are in the common leg of an amplifier, similar to R_4 in Fig. 8.4*d*, but are bypassed by a capacitor. This capacitor short circuits the resistor's noise current generator at high frequencies. It remains then to consider the low-frequency noise voltage generator at the input due to this resistor.

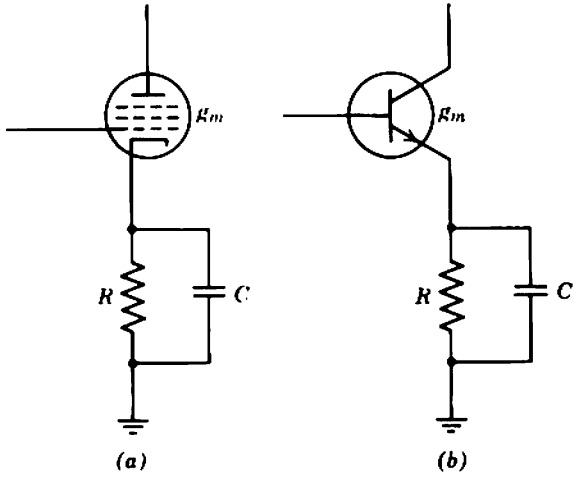


Fig. 8.5 Cathode and emitter biasing networks.

When the input terminals are short-circuited, the noise current in the output mesh of the stage due to its biasing resistor is

$$d(i_N'')^2_R = \frac{4kT}{R}\left[\frac{g_m{}^2R^2}{(1 + g_mR)^2 + \omega^2R^2C^2}\right] df.$$

The transfer admittance of the stage is

$$|Y_T|^2 = g_m{}^2\left[\frac{1 + \omega^2R^2C^2}{(1 + g_mR)^2 + \omega^2R^2C^2}\right].$$

Therefore, the noise voltage generator at the input due to the biasing resistor is

$$d(v_N{}^2)_R = \frac{d(i_N'')^2_R}{|Y_T|^2} = 4kTR\left(\frac{1}{1 + \omega^2R^2C^2}\right) df. \tag{8.46}$$

Strictly, $d(v_N{}^2)_R$ involves the device noise current generator $d(i_N{}^2)_D$ (see Eq. 8.44 of Section 8.3.1). This serves only to increase the noise attributable to the device and renders the thermal noise of the biasing network even less significant.

By inspection of Eq. 8.46, the noise contribution of the biasing network increases with decreasing frequency. Its contribution is therefore greatest at the lowest frequency of interest, the lower cutoff frequency ω_L of the amplifier. Now, ω_L cannot lie below the pole due to the biasing network; that is

$$\omega_L \geq \frac{1 + g_mR}{RC}.$$

Therefore

$$\omega_L \geq \frac{g_m}{C}. \tag{8.47}$$

The noise contribution of the biasing network at ω_L is

$$d(v_N^2)_{R,\omega_L} = 4kTR\left(\frac{1}{1 + \omega_L^2 R^2 C^2}\right) df$$

$$> 4kTR\left(\frac{1}{\omega_L^2 R^2 C^2}\right) df,$$

and substituting from Eq. 8.47

$$d(v_N^2)_{R,\omega_L} > \left(\frac{4kT}{g_m}\right)\left(\frac{1}{g_m R}\right) df. \tag{8.48}$$

For biasing stability in a vacuum-tube circuit, $g_m R$ is certainly not much less than unity, and may be much larger. Therefore, the noise contribution of the biasing network is of the order of, or else less than,

$$d(v_N^2)_{R,\omega_L} \approx \frac{4kT}{g_m} df. \tag{8.49}$$

The noise voltage generator of the tube is given by Eq. 8.12 or 8.19. Since the cathode temperature T_K is greater than the ambient temperature T by an order of magnitude, the biasing network noise is negligible in comparison with the white component of tube noise. If the flicker component of tube noise is considered, the biasing network noise is still less significant.

For biasing stability in a transistor circuit, $g_m R$ is much greater than unity. Therefore, the noise contribution of the biasing network is

$$d(v_N^2)_{R,\omega_L} \ll \frac{4kT}{g_m} df. \tag{8.50}$$

The noise contribution of the transistor is given by Eq. 8.27, and is very much larger.

Thus, the device noise is always much larger than the noise from a bypassed resistor in the common lead. It is left as an exercise for the reader to show that the noise from the screen biasing network of a pentode is also much less than the tube noise. Biasing-network noise can always be neglected.

8.3.2 Noise from Passive Elements

The above discussion assumes that the biasing resistors produce thermal noise only. This assumption may not be valid for certain common types

of resistor. Also, some other passive elements which should ideally be noiseless do in fact produce noise.

Resistors that employ carbon or similar granular materials have a significant component of flicker noise. This component is proportional to the dc current, and the corner frequency at which flicker and thermal components of noise are equal may be as high as 100 kHz. Wirewound resistors and some metal film resistors do not produce excess noise and may be required in the low-level section of an amplifier.

All polarized electrolytic capacitor types have a small but finite leakage current that appears to yield relatively high $1/f$ noise. The mechanism of noise generation does not seem to be well understood, but certainly the noise current increases with the leakage current. The magnitude of the leakage current depends mainly on the type of electrolyte used, the capacitance, temperature, and the time for which the polarizing voltage has been applied. As a rough rule, the leakage current may be written as

$$I_L = kCV, \tag{8.51}$$

where C is the nominal capacitance, V is the polarizing voltage, and k is a factor which decreases as the cost of the electrolyte increases, decreases for several seconds or minutes after the voltage is applied, and increases with temperature. Care must be exercised in using aluminum electrolytic capacitors in low-noise circuits, and whenever possible they should not be placed where their noise current can degrade amplifier performance. Failing this, tantalum or other low-leakage types should be used. The designer must accept a degradation of noise performance for a short time immediately after switching on.

Other components that may lead to unnecessary increases in the noise level are glow-discharge gas tubes and Zener diodes. Both types of device depend for their operation on a multiplicative ionization mechanism which is inherently noisy. The noise currents that arise from these devices can, in part, be bypassed from the amplifier by shunt capacitors. However, the shunt capacitance required is often large and if an electrolytic is used it contributes to the total noise. It is wise to avoid the use of constant-voltage devices in the early stages of amplifiers.

8.4 SIGNAL-TO-NOISE RATIO

It is shown in Section 8.1 that the spot noise performance of an amplifying device can be specified under given operating conditions (and, of course, at a given frequency) by two noise generators situated at the input

of a noiseless small-signal equivalent circuit. Section 8.3 discusses modifications to the values of these generators for taking into account the additional noise produced by the biasing resistors of a practical amplifying stage. Numerical values for the noise generators $d(v_N^2)$ and $d(i_N^2)$, and their laws of variation with operating conditions and frequency can be obtained in any of the following ways:

(i) from the relevant theoretical expressions given in Section 8.2,
(ii) from manufacturers' data,
(iii) from noise measurements.

Once the values of the generators are found, the small-signal equivalent circuit can be used to find the magnitude of the noise voltage or current at a given frequency at any point in the circuit. The small-signal equivalent circuit can also be used to calculate the signal voltage or current at the same point, and the ratio of signal to spot noise can be determined. This spot ratio is not a meaningful measure of noise performance, because the signal will not in general be corrupted by noise at the single spot frequency, but over an infinite bandwidth. Because $d(v_N^2)$ and $d(i_N^2)$ vary with frequency and because the amplifier gain is a function of frequency, the total noise voltage or current at any point must be determined by integration.

It is usual to define the *signal-to-noise ratio* S/N as the ratio of signal power to noise power; therefore, S/N is the ratio of mean-square signal voltage or current to the mean-square value of integrated noise voltage or current. The value of S/N is often expressed in decibels:

$$(S/N)_{\text{dB}} = 10 \log_{10} (S/N)_{\text{numeric}}. \tag{8.52}$$

The minimum acceptable values of peak signal to noise ratio for typical systems are given in Table 7.1. It is important to realize that S/N must be evaluated by integration over the frequency response characteristic of a complete system. In some instances human characteristics such as vision or hearing must be assessed in order to define a meaningful signal-to-noise ratio.

Because physical noise sources exist throughout an amplifier, the signal-to-noise ratio deteriorates progressively from stage to stage. Consequently, the signal-to-noise ratio S_o/N_o at the output is the meaningful quantity. However, the signal level increases from stage to stage, and therefore the noise from the second and later stages becomes less and less significant. Indeed, in the majority of cases, the signal-to-noise ratio at the output depends almost entirely on the noise performance of the first stage.

8.4.1 Signal-to-Noise Ratio of a Single-Stage Amplifier

Consider the situation shown in Fig. 8.6 where a single-stage amplifier is fed from a source i_S of admittance Y_S ($= G_S + jB_S$), and has a load resistor R_L. The noise generators $d(v_N^2)$ and $d(i_N^2)$ account for all of the physical sources of noise in the stage, and the conductive component G_S of the source gives rise to a noise current $4kT_SG_S\,df$. Notice that the effective source temperature T_S is not necessarily the ambient temperature; the product T_SG_S is so defined that the apparent thermal noise of the source is in fact equal to the total of all components of spot noise—thermal, shot, and flicker. In addition, thermal noise is generated within the load resistor R_L; however, it is assumed that this noise current is negligible compared with the output noise current from the device, and no noise generator for R_L is shown. In rare cases where the noise from R_L is significant, this noise may be considered as belonging to a second stage of the amplifier, and the results developed in Section 8.4.2 apply. For the disposition of noise generators shown in Fig. 8.6, the output ratio S_o/N_o of signal to noise in R_L can be obtained from a knowledge of the mean-square signal voltage v_{Si}^2 of frequency f_0 and the mean-square value of integrated noise voltage $d(v_{Ni}^2)$ which appears across Y_i.

The signal voltage of frequency f_0 is given by

$$v_{Si}^2 = \frac{i_S^2}{|Y_S + Y_i|^2}, \tag{8.53}$$

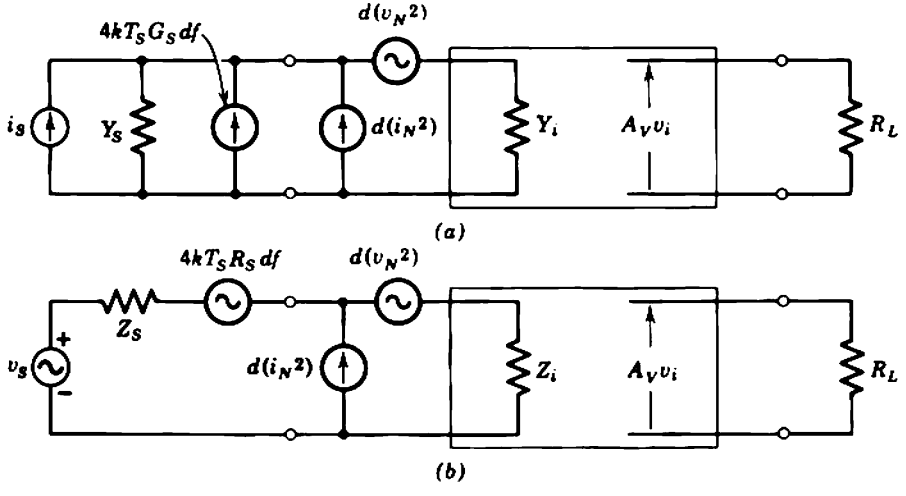


Fig. 8.6 Signal-to-noise ratio of a single-stage amplifier: (*a*) current-generator representation of source; (*b*) voltage-generator representation of source.

and the spot noise voltage of frequency f_0 is

$$d(v_{Ni}^2) = \frac{4kT_S G_S\, df + d(i_N^2) + d(v_N^2)|Y_S|^2}{|Y_S + Y_i|^2} \tag{8.54}$$

provided that all noise sources are uncorrelated. The signal power developed in R_L is

$$S_o = \frac{v_{Si}^2 |A_V(f_0)|^2}{R_L},$$

where $A_V(f_0)$ is the voltage gain of the amplifier at the signal frequency f_0. Similarly, the spot noise power of frequency f_0 in R_L is

$$dN_o = \frac{d(v_{Ni}^2)|A_V(f_0)|^2}{R_L}.$$

Substituting from Eqs. 8.53 and 8.54, the spot value of the signal-to-noise ratio is

$$\frac{S_o}{dN_o} = \frac{i_S^2}{4kT_S G_S\, df + d(i_N^2) + d(v_N^2)|Y_S|^2}, \tag{8.55}$$

which is completely independent of the transfer function $A_V/(Y_S + Y_i)$. This result can in fact be derived directly from Fig. 8.6 by applying Norton's theorem to the independent signal and noise generators to the left of Y_i. From a purely analytical viewpoint, it is unfortunate that this independence of transfer function is not preserved when the denominator is integrated to evaluate total noise power.

Each of the spot mean-square noise generators $d(i_N^2)$ and $d(v_N^2)$ is either constant or varies with frequency. The functional relationships can be written generally as

$$d(i_N^2) = \varphi_I(f)\, df, \tag{8.56a}$$

$$d(v_N^2) = \varphi_V(f)\, df, \tag{8.56b}$$

and in terms of this notation the mean-square spot noise voltage across Y_i is given by

$$d(v_{Ni}^2) = \left(\frac{4kT_S G_S + \varphi_I + \varphi_V|Y_S|^2}{|Y_S + Y_i|^2}\right) df. \tag{8.57a}$$

The total noise power output is

$$N_o = \int_0^\infty \frac{|A_V(f)|^2}{R_L}\, d(v_{Ni}^2)$$

and the output signal-to-noise ratio is therefore

$$\frac{S_o}{N_o} = \frac{i_S^2 \left| \dfrac{A_V}{Y_S + Y_i} \right|^2_{f=f_0}}{\displaystyle\int_0^\infty (4kT_S G_S + \varphi_I + \varphi_V |Y_S|^2) \left| \frac{A_V}{Y_S + Y_i} \right|^2 df} \cdot \qquad (8.58a)$$

The above calculation is not intended to be a model for all calculations of signal-to-noise ratio. Rather, it is intended to illustrate the principles; it is neither necessary nor desirable to adhere slavishly to the form of Eq. 8.58*a* in all cases. In some instances it is convenient to evaluate S_o/N_o as the ratio of the square of signal load current to the mean-square noise load current; A_V in Eq. 8.58*a* is replaced by Y_T. Again, it may be convenient to represent the source by a voltage generator and series impedance as in Fig. 8.6*b*; the spot noise voltage at the input becomes

$$d(v_{Ni}^2) = (4kT_S R_S + \varphi_I |Z_S|^2 + \varphi_V) \left| \frac{Z_i}{Z_S + Z_i} \right|^2 df \qquad (8.57b)$$

and the output signal-to-noise ratio is

$$\frac{S_o}{N_o} = \frac{v_S^2 \left| \dfrac{Z_i A_V}{Z_S + Z_i} \right|^2_{f=f_0}}{\displaystyle\int_0^\infty (4kT_S R_S + \varphi_I |Z_S|^2 + \varphi_V) \left| \frac{Z_i A_V}{Z_S + Z_i} \right|^2 df} \cdot \qquad (8.58b)$$

Finally, A_V (or Y_T) can be combined with Y_i (or Z_i) to express the signal-to-noise ratio in terms of Z_T or A_I.

8.4.2 Signal-to-Noise Ratio of a Multistage Amplifier

Consider the two-stage amplifier shown in Fig. 8.7*a*, in which the individual stages are characterized for spot noise performance by generators $d(v_{N1}^2)$, $d(i_{N1}^2)$, $d(v_{N2}^2)$, and $d(i_{N2}^2)$. It is required to find the relation between these generators and the two noise generators $d(v_N^2)$ and $d(i_N^2)$ situated at the input that represent the over-all noise performance of the complete amplifier. As shown in Fig. 8.7*b*, the uncorrelated generators $d(v_{N2}^2)$ and $d(i_{N2}^2)$ may be replaced by the single noise current generator

$$d(i_{N2}^2) + d(v_{N2}^2) |Y_{o1}|^2$$

situated at the output of the first stage. This noise current generator may in turn be represented by two noise generators at the input of the

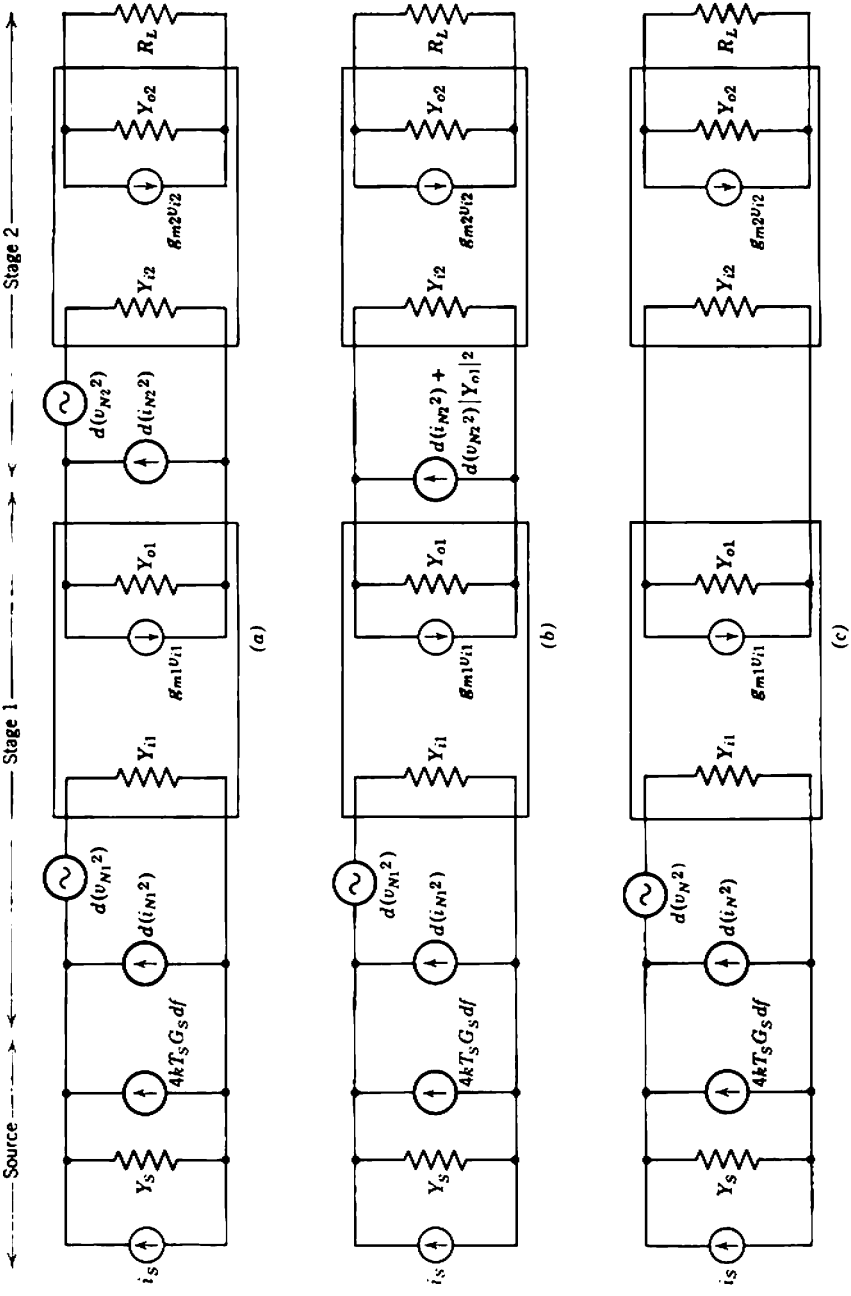


Fig. 8.7 Signal-to-noise ratio of a two-stage amplifier.

amplifier. By inspection, the total input noise generators $d(v_N^2)$ and $d(i_N^2)$ for the complete amplifier are

$$d(v_N^2) = d(v_{N1}^2) + \frac{d(i_{N2}^2) + d(v_{N2}^2)|Y_{o1}|^2}{g_{m1}^2}, \tag{8.59a}$$

$$d(i_N^2) = d(i_{N1}^2) + \frac{d(i_{N2}^2) + d(v_{N2}^2)|Y_{o1}|^2}{\beta_1^2}, \tag{8.59b}$$

where g_{m1} and β_1 are, respectively, the mutual conductance and short-circuit current gain of the device in the first stage. These noise generators may be expressed in the φ notation of Eqs. 8.56 as

$$d(v_N^2) \equiv \varphi_V\, df = \left(\varphi_{V1} + \frac{\varphi_{I2} + \varphi_{V2}|Y_{o1}|^2}{g_{m1}^2}\right) df, \tag{8.60a}$$

$$d(i_N^2) \equiv \varphi_I\, df = \left(\varphi_{I1} + \frac{\varphi_{I2} + \varphi_{V2}|Y_{o1}|^2}{\beta_1^2}\right) df. \tag{8.60b}$$

Finally, the output signal-to-noise ratio is given by substituting these values for the φ functions into Eqs. 8.58. Notice that A_V in Eqs. 8.58 must be interpreted as the voltage gain of the complete two-stage amplifier.

In most practical amplifiers, the values of g_{m1} and β_1 are so large that the noise contributions of the second stage to φ_V and φ_I are negligible. Therefore, from Eqs. 8.60

$$\varphi_V \approx \varphi_{V1}, \tag{8.61a}$$

$$\varphi_I \approx \varphi_{I1}. \tag{8.61b}$$

In amplifiers having more than two stages, the noise contributions of the third and succeeding stages to the over-all output noise will invariably be negligible if even a moderate signal gain is achieved in the first two stages. Thus, the proposition made at the outset, that the noise performance of a multistage amplifier is determined by the noise performance of its input stage, is true under all but the most unusual circumstances.

8.4.3 Practical Simplifications

Very often, Eqs. 8.58 for the output signal-to-noise ratio can be simplified when applied to a practical amplifier.

An amplifier for use with a reactive source is usually designed so that the transfer function from the source internal generator to the amplifier output has a flat mid-band frequency range. Consequently, the amplifier itself may not have a flat mid-band range. In this and all other situations, any transfer function of a system can be written as

$$TF(f) = TF_m\, \psi(f), \tag{8.62}$$

where TF_m is the system mid-band transfer function, and $\psi(f)$ is the system frequency response; $\psi(f)$ is the equivalent of $\psi(s)$ in real frequency notation. In Eq. 8.58*a*, the term $A_V/(Y_S + Y_i)$ is the system transfer impedance from source current generator to output voltage. Similarly, the term $Z_i A_V/(Z_S + Z_i)$ in Eq. 8.58*b* is the system voltage gain from source voltage generator to output voltage. If it is assumed that the signal frequency f_0 lies in the mid-band range of the system transfer function (as is usually the case), then

$$\psi(f_0) = 1$$

and Eqs. 8.58 can be simplified by means of Eq. 8.62 to

$$\frac{S_o}{N_o} = \frac{{i_S}^2}{\int_0^\infty (4kT_S G_S + \varphi_I + \varphi_V |Y_S|^2)|\psi(f)|^2 \, df}, \tag{8.63a}$$

$$\frac{S_o}{N_o} = \frac{{v_S}^2}{\int_0^\infty (4kT_S R_S + \varphi_I |Z_S|^2 + \varphi_V)|\psi(f)|^2 \, df}. \tag{8.63b}$$

Notice particularly that $\psi(f)$ *in these equations is the frequency response of the complete system*, the amplifier plus the source, not the amplifier alone. Only if the source impedance and amplifier input impedance are purely resistive does $\psi(f)$ reduce to the frequency response of the amplifier.

For many amplifier applications of the type considered in this book, the source impedance and amplifier input impedance are in fact substantially resistive over the frequency range of interest. Moreover, their ratio is such that one of the φ functions in Eqs. 8.58 dominates. For example, if $G_S \ll G_i$ (corresponding to an approximate current source), Eq. 8.58*a* becomes

$$\frac{S_o}{N_o} \approx \frac{{i_S}^2 |Z_T(f_0)|^2}{\int_0^\infty \varphi_I |Z_T(f)|^2 \, df} = \frac{{i_S}^2}{\int_0^\infty \varphi_I |\psi(f)|^2 \, df}. \tag{8.64a}$$

In this equation, $Z_T(f)$ and $\psi(f)$ are parameters of the amplifier alone. Similarly, if $R_S \ll R_i$ (corresponding to an approximate voltage source), Eq. 8.58*b* becomes

$$\frac{S_o}{N_o} \approx \frac{{v_S}^2 |A_V(f_0)|^2}{\int_0^\infty \varphi_V |A_V(f)|^2 \, df} = \frac{{v_S}^2}{\int_0^\infty \varphi_V |\psi(f)|^2 \, df}. \tag{8.64b}$$

Equations 8.63 can be used to define an *equivalent noise current* or *voltage referred to the input* of an amplifier in such a way that S_o/N_o is the

ratio of this quantity to the source current generator or source voltage generator:

$$\frac{S_o}{N_o} = \frac{i_S^2}{i_{Ni}^2{}_{eq}}, \tag{8.65a}$$

$$\frac{S_o}{N_o} = \frac{v_S^2}{v_{Ni}^2{}_{eq}}. \tag{8.65b}$$

In the general case, the mean-square noise current and voltage referred to the input are the denominators of Eqs. 8.63:

$$i_{Ni}^2{}_{eq} = \int_0^\infty (4kT_S G_S + \varphi_I + \varphi_V |Y_S|^2)|\psi(f)|^2\, df, \tag{8.66a}$$

$$v_{Ni}^2{}_{eq} = \int_0^\infty (4kT_S R_S + \varphi_I |Z_S|^2 + \varphi_V)|\psi(f)|^2\, df, \tag{8.66b}$$

where $\psi(f)$ is the frequency response of the system. In the special case of a dominant φ function with resistive source and input impedances,

$$i_{Ni}^2{}_{eq} \approx \int_0^\infty \varphi_I |\psi(f)|^2\, df, \tag{8.67a}$$

$$v_{Ni}^2{}_{eq} \approx \int_0^\infty \varphi_V |\psi(f)|^2\, df, \tag{8.67b}$$

where $\psi(f)$ is the frequency response of the amplifier alone.

Distinguish carefully between the various input noise quantities:

$d(i_N^2)$, $d(v_N^2)$ are the spot noise current and voltage generators at the amplifier input, representing the noise sources within the amplifier.

$d(i_{Ni}^2)$, $d(v_{Ni}^2)$ are the spot noise current that actually flows into the input impedance of the amplifier and the spot noise voltage that actually appears across the input impedance of the amplifier.

$i_{Ni}^2{}_{eq}$, $v_{Ni}^2{}_{eq}$ are the equivalent mean-square noise current and voltage referred to the amplifier input, immediately useful in calculating the signal-to-noise ratio from the source voltage or current generator.

Notice that, for a resistive source,

$$v_{Ni}^2{}_{eq} = R_S^2 i_{Ni}^2{}_{eq}. \tag{8.68}$$

Often, the frequency dependence of the terms in the above integrals is simple. For example, the ψ functions are likely to be dominated by one or two poles, whereas the noise functions may be independent of frequency (white noise), vary as $1/f$ (flicker noise), or vary as f^2 (high-frequency

noise). In cases where such simplifications do not apply, the integration may be carried out graphically. Some of the simpler integrals that occur in noise problems are listed in Table 8.2.* In these integrals, f_1 and f_2 are

Table 8.2 Useful Noise Integrals

	Integral	General Solution	Wide-band Approximation $f_2 \gg f_1$	Special Case $f_1 = f_2 = f_0$
1	$\int_0^\infty \left[\frac{1}{1+(f/f_2)^2}\right] df$	$\frac{\pi}{2} f_2$	$\frac{\pi}{2} f_2$	$\frac{\pi}{2} f_0$
2	$\int_0^\infty \left[\frac{(f/f_1)^2}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_2)^2}\right] df$	$\frac{\pi}{2}\left[\frac{f_2^2}{f_1+f_2}\right]$	$\frac{\pi}{2} f_2$	$\frac{\pi}{4} f_0$
3	$\int_0^\infty \frac{1}{f}\left[\frac{(f/f_1)^2}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_2)^2}\right] df$	$\left[\frac{1}{1-(f_1/f_2)^2}\right] \ln\left(\frac{f_2}{f_1}\right)$	$\ln\left(\frac{f_2}{f_1}\right)$	$\frac{1}{2}$
4	$\int_0^\infty \left[\frac{1}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_2)^2}\right] df$	$\frac{\pi}{2}\left[\frac{f_1 f_2}{f_1+f_2}\right]$	$\frac{\pi}{2} f_1$	$\frac{\pi}{4} f_0$
5	$\int_0^\infty \left[\frac{(f/f_1)^2}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_1)^2}\right] \times \left[\frac{1}{1+(f/f_2)^2}\right] df$	$\frac{\pi}{4}\left[\frac{f_1 f_2^2}{(f_1+f_2)^2}\right]$	$\frac{\pi}{4} f_1$	$\frac{\pi}{16} f_0$
6	$\int_0^\infty \left[\frac{1}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_2)^2}\right]^2 df$	$\frac{\pi}{4}\left[\frac{f_1 f_2(f_1+2f_2)}{(f_1+f_2)^2}\right]$	$\frac{\pi}{2} f_1$	$\frac{3\pi}{16} f_0$
7	$\int_0^\infty \left[\frac{(f/f_1)^2}{1+(f/f_1)^2}\right]\left[\frac{1}{1+(f/f_2)^2}\right]^2 df$	$\frac{\pi}{4}\left[\frac{f_2^3}{(f_1+f_2)^2}\right]$	$\frac{\pi}{4} f_2$	$\frac{\pi}{16} f_0$

break frequencies corresponding to real poles. Integrals (1) and (2) occur in computing the mean-square output noise from a simple amplifier with a white noise function. In (1) the amplifier has a single high-frequency break, and the signal bandwidth is f_2. In (2) the amplifier has one low-frequency and one high-frequency break, and the signal bandwidth is $(f_2 - f_1)$; for a wide-band amplifier, f_2 is much greater than f_1 and the solutions for integrals (1) and (2) converge. Integral (3) arises in computing the mean-square output noise from an amplifier of the type assumed in (2), when the noise has a $1/f$ spectrum. For a wide-band amplifier

* These and other integrals are listed in: A. B. GILLESPIE, *Signal, Resolution and Noise in Nuclear Counting Amplifiers* (McGraw-Hill, New York, 1953); A. H. SNELL (ed.), *Nuclear Instruments and Their Uses*, Vol. I (Wiley, New York, 1962).

the approximation implies that the noise output arises from the $1/f$ contribution in the signal bandwidth, because

$$\ln\left(\frac{f_2}{f_1}\right) = \int_{f_1}^{f_2} \frac{df}{f}.$$

The remaining integrals have several obvious interpretations, and are useful in not-too-specialized problems.

Sometimes the spot noise at the input of an amplifier is white. The most usual situation is when the source and input impedance are resistive, and $R_S \ll R_I$ so that the white noise function φ_V dominates. As φ_V is white (independent of frequency), it can be taken outside the integral in Eq. 8.67*b*. What remains is defined as the *noise bandwidth* of the amplifier:

$$\mathscr{B}_N = \int_0^\infty |\psi(f)|^2 \, df \tag{8.69}$$

and the equivalent noise voltage referred to the input is

$$v_{Ni}{}^2{}_{eq} = \varphi_V \int_0^\infty |\psi(f)|^2 \, df = \varphi_V \mathscr{B}_N.$$

Thus, noise bandwidth takes on the significance of the finite frequency range over which white noise at the input can be summed without regard to the frequency dependence of the amplifier's gain. Noise that is not white cannot be summed by means of the noise bandwidth. Usually the noise bandwidth of an amplifier is different from its 3-dB signal bandwidth; in integral (1) or the wide-band case of integral (2)

$$\mathscr{B}_N = 1.57\mathscr{B}_S.$$

However, the two are of the same order of magnitude. In view of the usual lack of precision with which the values of the noise generators are known, it is often a justifiable approximation to equate the signal and noise bandwidths. This is a particularly useful approximation in amplifiers for which the low-frequency and/or high-frequency rates of attenuation are complex, so that an exact evaluation of the noise bandwidth would be tedious.

8.4.4 Examples of Signal-to-Noise Calculations

The expressions developed in the preceding sections provide the formal basis for signal-to-noise calculations, but an engineering appreciation comes most readily from a study of some specific numerical examples.

It is informative to base the examples on the two designs described in Section 7.8. Although the examples differ in the working approximations which are permissible, the following general plan is adhered to in both calculations:

(i) establishing a functional or graphical representation for the transfer-function frequency response $\psi(s)$,
(ii) establishing a functional or graphical representation for the input mean-square noise generators, the φ functions of preceding sections,
(iii) an analytical or numerical integration to determine the total mean-square noise current or voltage at the output.

8.4.4.1 Vacuum-Tube Amplifier

The vacuum-tube amplifier discussed in Section 7.8.1 is represented diagrammatically in Fig. 8.8. The original design is such that $A_V(f)$ has a magnitude 1000 and five coincident low-frequency poles at $f_L = 16$ Hz. There is a sixth low-frequency pole due to the input coupling capacitor C_C. This capacitor is not included in $A_V(f)$; rather, it is included in the effective source admittance Y_S. In addition, $A_V(f)$ has a high-frequency pole due to the interstage network, at $f_H = 41$ kHz. The positions of other high-frequency poles depend on the source impedance and load capacitance. For the purpose of this calculation it is assumed that these poles are far

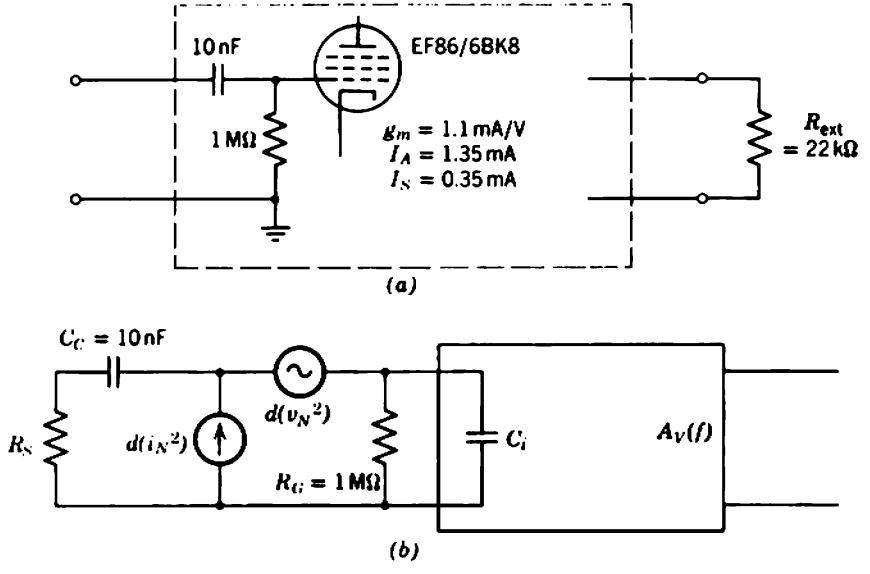


Fig. 8.8 Noise calculation on the vacuum-tube amplifier shown in Fig. 7.49: (*a*) outline circuit diagram; (*b*) circuit for noise calculations.

out, and that the response is dominated by the pole at 41 kHz. Thus, $A_V(f)$ may be written as

$$A_V(f) = 1000\left(\frac{jf/f_L}{1 + jf/f_L}\right)^5\left(\frac{1}{1 + jf/f_H}\right)$$

and the corresponding ψ function for use in noise calculations is

$$|\psi(f)|^2 = \left[\frac{(f/f_L)^2}{1 + (f/f_L)^2}\right]^5\left[\frac{1}{1 + (f/f_H)^2}\right]. \tag{8.70}$$

This function is plotted on logarithmic scales in Fig. 8.9*a*.

Because the voltage gain of the pentode is high (100), it is most unlikely that either its load resistor or the following triode will make significant contributions to over-all noise. The initial calculation will therefore assume that all noise arises from the pentode and its grid resistor. The validity of this assumption may be checked subsequently by comparing the various components of noise voltage developed at the pentode's anode.

As with most noise calculations, there is an initial difficulty in obtaining reliable information on the noise performance of the device to be used. For the EF86/6BK8 the only information provided by one manufacturer is that

"under normal conditions, . . . the equivalent noise voltage [referred to the input] is approximately 2 μV for the frequency range 25 to 10,000 Hz."

The incompleteness of this specification forces the designer to rely on basic calculations or measurements to obtain the necessary information on the input noise generators. However, since the noise at low frequencies is dominated by flicker components, the manufacturers' data may be used as follows to check on the value assumed for the constant A in Eq. 8.19*b*:

From Section 7.8.1, the operating conditions for the EF86/6BK8 are

$$I_S = 0.35 \text{ mA},$$
$$g_m = 1.1 \text{ mA/V}.$$

Substitution into Eq. 8.19*b* gives the spot noise voltage generator at the input as

$$d(v_N^2) = \left(10.5 \times 10^{-17} + \frac{A}{F}\right) df,$$

where the first term represents the shot noise on the anode current and the partition noise, and the second term represents the flicker noise. Thus,

$$\varphi_V = 10.5 \times 10^{-17} + \frac{A}{f}. \tag{8.71}$$

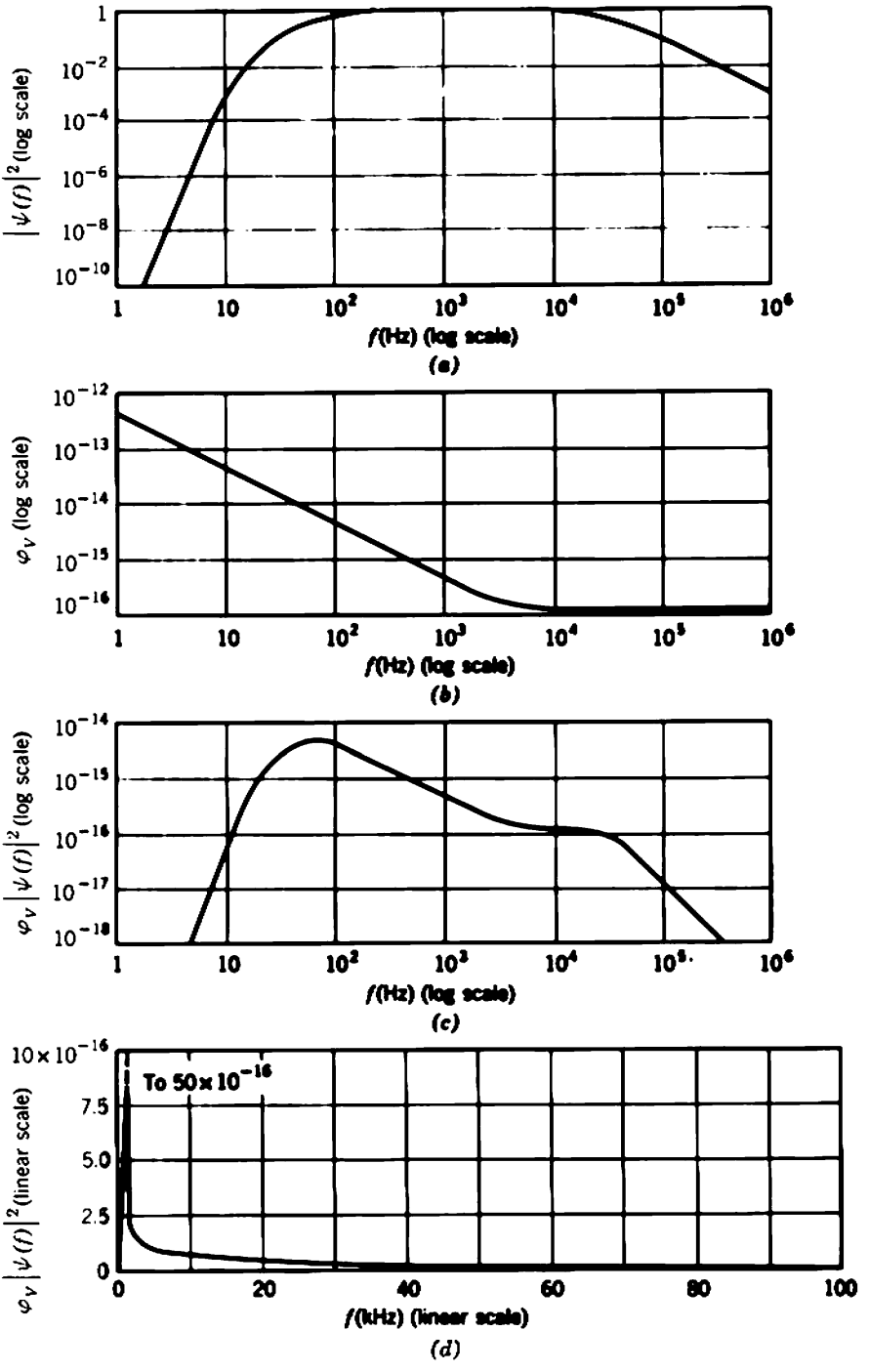


Fig. 8.9 Noise functions of the vacuum-tube amplifier shown in Fig. 7.49.

Assuming that the manufacturers' test conditions include a small value of impedance connected between grid and ground, the quoted value of equivalent noise voltage referred to the input is the contribution of $d(v_N^2)$ alone, and is given by Eq. 8.67*b*. Assuming further that the manufacturers' test conditions include an amplifier with simple 6 dB/octave fall-off in gain at low and high frequencies, then for the test circuit

$$|\psi(f)|^2 = \left[\frac{(f/f_1)^2}{1 + (f/f_1)^2}\right]\left[\frac{1}{1 + (f/f_2)^2}\right],$$

where f_1 and f_2 are respectively 25 Hz and 10 kHz. Finally, from the wide-band approximations of integrals (2) and (3) in Table 8.2, and Eq. 8.67*b*

$$\begin{aligned} v_{Ni}{}^2{}_{eq} &= (10.5 \times 10^{-17}) \times 1.57 f_2 + A \ln\left(\frac{f_2}{f_1}\right) \\ &= 1.65 \times 10^{-12} + 6.0A. \end{aligned}$$

But the manufacturers' data above gives

$$v_{Ni}{}^2{}_{eq} = 4.0 \times 10^{-12}.$$

Therefore,

$$A = 4.0 \times 10^{-13}.$$

This value is of the order of magnitude expected, and will be used throughout this example. The plot of mean-square spot noise generator voltage per cycle shown in Fig. 8.9*b* follows Eq. 8.71 with $A = 4.0 \times 10^{-13}$, and the corner frequency at which the shot and flicker components are equal is 3.8 kHz.

The mean-square spot noise current generator at the tube's input is given by Eq. 8.15. Over the audio-frequency range which is of interest in this calculation, the only component that could possibly be significant is that due to the dc grid current:

$$d(i_N^2)_D \approx 2q \sum (I_G)\, df.$$

In addition there is a component of noise current due to the 1 MΩ grid resistor R_G. The total spot noise current generator at the tube's grid follows from Eq. 8.38 as

$$d(i_N^2) = \varphi_I\, df = \left[2q \sum (I_G) + \frac{4kT}{R_G}\right] df.$$

Conservatively, $\sum (I_G)$ is taken as 10 nA, and therefore

$$\varphi_I = 3.2 \times 10^{-27} + 1.6 \times 10^{-26};$$

that is,

$$\varphi_I = 1.92 \times 10^{-26}. \tag{8.72}$$

The resistor noise is the dominant component.

Reference to the equivalent circuit (Fig. 8.8*b*), in which the noise generators have been included, indicates that the spot noise voltage at the input due to both the tube and R_G is given by Eq. 8.57*b* as

$$d(v_{Ni}^2) = \left[\varphi_V + \frac{\varphi_I}{(\omega C_C)^2}\right]\left[\frac{R_G^2}{R_G^2 + (1/\omega C_C)^2}\right] df,$$

provided that the source resistance R_S is small (a condition which is very likely to be true for this type of amplifier) and that the coupling capacitor C_C is much larger than the stray shunt input capacitance C_i (a condition which is certainly true). It is apparent that consideration of the contribution due to $d(i_N^2)$ is, in the context of this problem, quite naïve. Even at a frequency 1 Hz for which the capacitive reactance of C_C far exceeds the 1 MΩ resistance of R_G, the contribution to $d(v_{Ni}^2)$ from φ_I is only 1.92×10^{-14} which is masked by the φ_V contribution 4.0×10^{-13}. At higher frequencies the contribution due to φ_I is even less significant, so the total spot noise voltage at the input may be written as

$$d(v_{Ni}^2) \approx \varphi_V \, df.$$

The equivalent mean-square noise voltage referred to the input is given by Eq. 8.68*b*, which may be evaluated by multiplying ordinates of Fig. 8.9*a* and *b*, and measuring the area under the resulting curve. Figures 8.9*c*

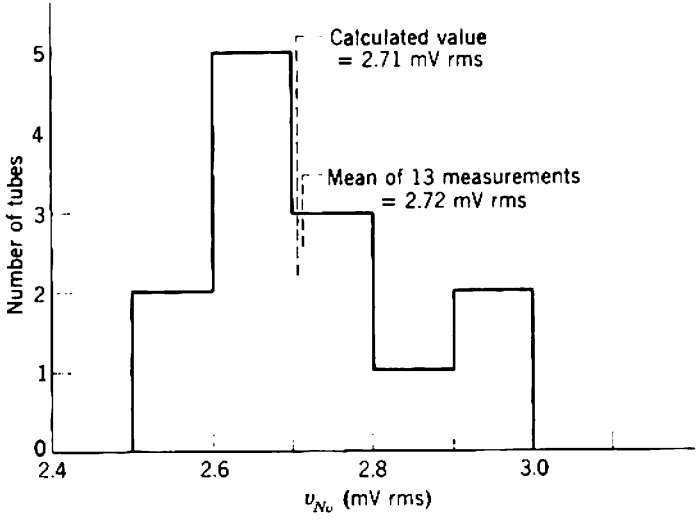


Fig. 8.10 Histogram of v_{No} for 13 different EF86/6BK8 pentodes in the amplifier circuit of Fig. 7.49.

and d show the curve on logarithmic and linear scales; the area is

$$v_{Ni}{}^2{}_{eq} = 7.36 \times 10^{-12}$$

and therefore

$$v_{Nieq} = 2.71\ \mu\text{V}. \tag{8.73}$$

Because the mid-band gain of the pentode is 100, the rms noise voltage referred to its anode is 271 μV. This far exceeds any contributions from its load resistor or the following triode, so Eq. 8.73 is an accurate calculated measure of the amplifier's noise performance.

The over-all gain of the amplifier is 1000, so the rms output noise v_{No} is 2.71 mV. The measured value of v_{No} for a number of tubes is indicated by the histogram of Fig. 8.10; the mean value 2.72 mV agrees well with the predicted figure. If a signal-to-noise ratio of 60 dB is to be achieved for mid-band signals, the input signal magnitude must exceed 2.7 mV.

8.4.4.2 Transistor Amplifier

The transistor amplifier discussed in Section 7.8.2 is represented diagrammatically in Fig. 8.11*a*. This amplifier has a mid-band voltage gain (including the 1 to 5 step-up transformer) of 1000, three low-frequency poles at 13.5 Hz, and a somewhat complex high-frequency behavior due to a resonance between the transformer secondary leakage inductance and the transistor input capacitance. In the interests of simplicity it is assumed that additional capacitance C_L shunts the load, so that the high-frequency behavior is dominated by a simple high-frequency break point at 15 kHz. All calculations can be based on the circuit shown in Fig. 8.11*b*, in which the source and transformer parameters are referred to the secondary winding. The voltage gain from the transistor base to the load is

$$A_V(f) = 200\left(\frac{jf/f_L}{1 + jf/f_L}\right)^2\left(\frac{1}{1 + jf/f_H}\right),$$

where $f_L = 13.5$ Hz and $f_H = 15$ kHz are the lower and upper break frequencies. The corresponding ψ function for use in the noise calculations is

$$|\psi(f)|^2 = \left[\frac{(f/f_L)^2}{1 + (f/f_L)^2}\right]^2\left[\frac{1}{1 + (f/f_H)^2}\right]. \tag{8.74}$$

There are three spot noise generators at the transistor base—the source noise current generator and the transistor noise voltage and current generators. Unlike the vacuum-tube amplifier considered above, the source resistance is specified in this example and its contribution to

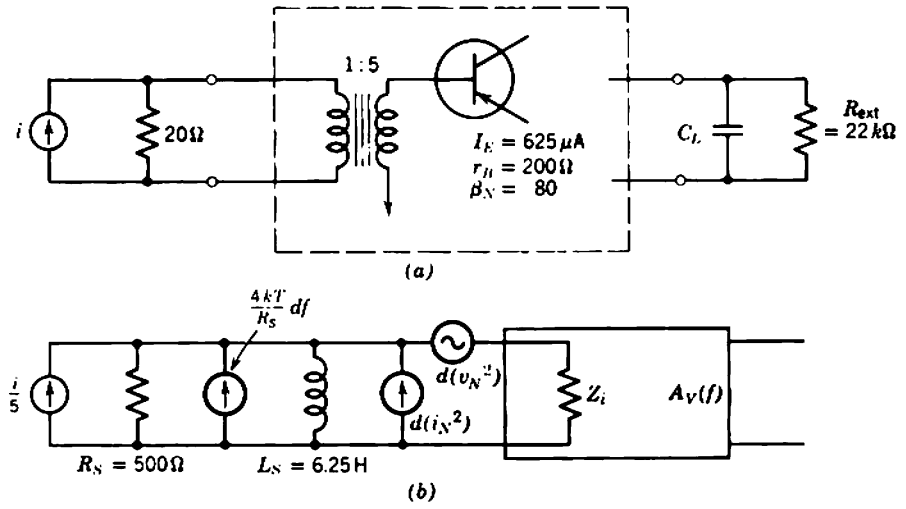


Fig. 8.11 Noise calculation on the transistor amplifier shown in Fig. 7.50: (*a*) outline circuit diagram; (*b*) circuit for noise calculations.

the noise is taken into account. The spot noise current generator of the resistive source is

$$\frac{4kT}{R_S}\,df = 3.2 \times 10^{-23}\,df. \tag{8.75}$$

The transistor spot noise generators are given by Eqs. 8.27 and 8.28. Substitution of the data

$$\begin{aligned} I_E &= 625\ \mu A, \\ r_E &= 40\ \Omega, \\ r_B &= 200\ \Omega, \\ \beta_N &= 80, \\ I_{CO} &= 2\ \mu A \end{aligned}$$

gives the corresponding φ functions as

$$\varphi_V = 3.52 \times 10^{-18}, \tag{8.76}$$

$$\varphi_I = 3.14 \times 10^{-24} + \frac{2.0 \times 10^{-22}}{|\beta(f)|^2} + \frac{A}{f}. \tag{8.77}$$

The spot noise voltage generator at the input is white, and the thermal noise of r_B is the dominant component. The spot noise current generator has three components which are significant at different parts of the frequency spectrum. In Eq. 8.77 the first term accounts for shot noise on

the dc base current (including I_{CO}), the second term accounts for the increase in noise current at high frequencies, and the third term represents the flicker noise. The upper corner frequency at which the first and second terms are equal occurs when

$$|\beta(f)| \approx \left(\frac{f_T}{f}\right) = 8.0.$$

It follows from Section 7.8.2 that f_T is 1.0 MHz, and therefore the upper corner frequency occurs at 125 kHz. The constant A is often a doubtful quantity; for the purpose of this calculation it is assumed that the lower corner frequency occurs at 200 Hz, in which case A has the value 6.28×10^{-22}.

Notice that φ_V is considerably smaller than in the vacuum-tube amplifier considered above, whereas φ_I is considerably larger. Even so the transistor spot noise current generator is masked by the source white noise current generator except at very high and very low frequencies. It is therefore a good approximation to represent the total spot noise current generator at the transistor base by a white noise generator

$$\left(\frac{4kT}{R_S} + \varphi_I\right) \approx 3.5 \times 10^{-23}. \tag{8.78}$$

This approximation simplifies the subsequent calculations.

The spot noise voltage at the input is given by Eq. 8.57*a*. Reference to Fig. 8.11 shows that

$$Z_S = \frac{1}{Y_S} = R_S\left[\frac{s(L_S/R_S)}{1 + s(L_S/R_S)}\right]$$

and

$$Z_i = \frac{1}{Y_i} = \left(\frac{s + \beta_N/R_iC_E}{s + 1/R_EC_E}\right)\left(\frac{1}{1 + sR_iC_i}\right)R_i,$$

where R_i $(= r_B + \beta_N r_E)$ is the mid-band input resistance. The first and second terms in Z_i account for its low-frequency and high-frequency variation, respectively. It follows from Section 7.8.2 that the impedances have the low- and high-frequency break points shown in Fig. 8.12. It is possible to adopt a similar approach to that of the preceding section, and evaluate $d(v_{Ni}^2)$ as a function of frequency on a point-by-point basis. However, it is preferable to use the experience gained in the vacuum-tube calculation as a guide for an approximate but more direct solution. A first approximation is that, since $|Y_S|$ is much larger than $|Y_i|$ over the passband and for a long way beyond it, and since all noise generators are

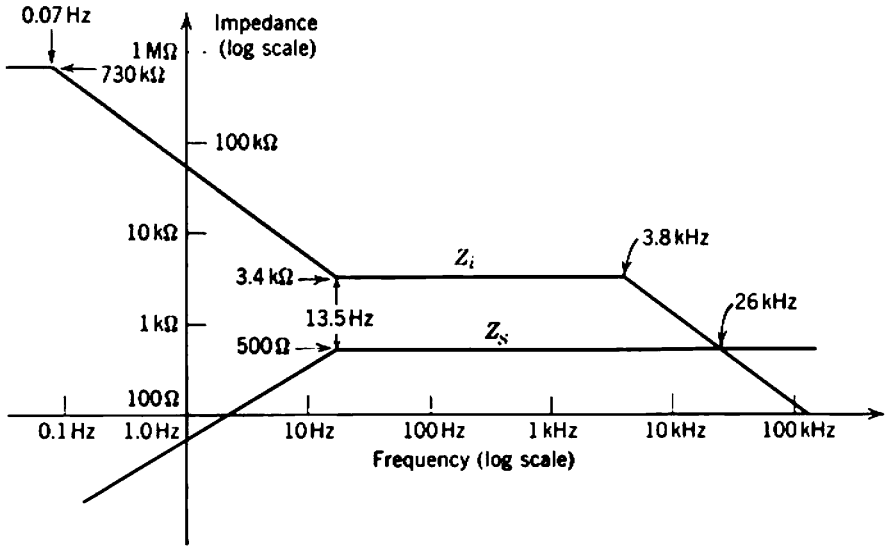


Fig. 8.12 Transistor input impedance and transformed source impedance for the transistor amplifier shown in Fig. 7.50.

white, it is reasonable to assume that the spot noise voltage $d(v_{Ni}^2)$ at the input is white:

$$d(v_{Ni}^2) = \left(\frac{4kT}{R_S} + \varphi_I + \frac{\varphi_V}{R_S^2}\right)(R_S \parallel R_I)^2\, df.$$

Therefore,

$$d(v_{Ni}^2) = 9.57 \times 10^{-18}\, df. \tag{8.79}$$

Because $d(v_{Ni}^2)$ is white, the equivalent mean-square noise voltage referred to the input is obtained by multiplying $d(v_{Ni}^2)$ by the noise bandwidth $\mathscr{B}_N$. The amplifier has a single high-frequency break at $f_H = 15$ kHz, and a second good approximation is to use the wide-band solution to integral (2) of Table 8.2 as the noise bandwidth. Thus

$$\mathscr{B}_N = 1.57 f_H = 23.5\ \text{kHz} \tag{8.80}$$

and therefore, from Eq. 8.79,

$$v_{Ni}^2{}_{eq} = 2.25 \times 10^{-13};$$

that is,

$$v_{Ni\,eq} = 0.47\ \mu\text{V rms}. \tag{8.81}$$

The transformer turns ratio and transistor voltage gain are respectively 1:5 and 200. Therefore, the equivalent noise voltage referred to the transformer primary and the noise voltage at the amplifier output are

respectively 94 nV and 94 μV. Experimental results for a number of different transistors are given in Fig. 8.13. If a signal-to-noise ratio of 60 dB or greater is to be achieved for mid-band signals, the input signal voltage to the transformer primary must be at least 94 μV.

It is interesting to compare this last result with the equivalent result for the vacuum-tube amplifier. An input signal magnitude of 2.7 mV which is required to give 60 dB signal-to-noise ratio for the vacuum-tube amplifier gives 90 dB for the transistor amplifier. The superiority of the transistor amplifier stems from a combination of three factors:

1. For comparable magnitudes of collecting electrode current, the g_m of a transistor far exceeds that of a vacuum tube, and in consequence the spot noise voltage generator at the input is smaller for a transistor than for a tube. Therefore, when the source impedance is low so that $d(v_N^2)$ dominates in the expression for input spot noise, transistors exhibit superior noise performance. Tubes, on the other hand, are superior when the source impedance is high. Typical values of $d(i_N^2)$ for a tube are much the smaller, as their dc grid current is much smaller than the base current of a transistor.

2. The use of a step-up transformer increases the signal level before the transistor noise is added. In effect the signal amplification exceeds the noise amplification. This technique can also be used with profit in vacuum-tube amplifiers. However, it is not an improvement to increase the transformer step-up ratio without limit for either tubes or transistors. Ultimately, the reflected source impedance seen by the amplifying device

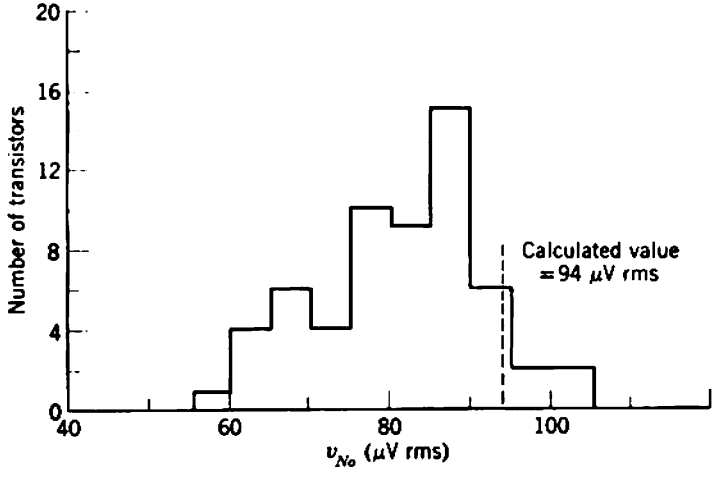


Fig. 8.13 Histogram of v_{No} for 59 germanium *p-n-p* transistors in the amplifier circuit of Fig. 7.50. These transistors are taken from five different type numbers.

becomes so large that the noise due to $d(i_N^2)$ becomes dominant. Any further increase in the turns ratio results in increased noise. A more detailed discussion is given in Section 8.5.2.

3. A proportion of the total improvement in the transistor circuit can be ascribed to the smaller bandwidth. Since the bandwidth ratio is

$$\frac{15 \text{ kHz}}{41 \text{ kHz}} = 0.36,$$

the ideal ratio of rms noise voltages for identical spot values of white noise is

$$\sqrt{0.36} = 0.60.$$

This certainly cannot account for the difference between the tube and transistor examples; in comparison with (1) and (2) above, the bandwidth factor is quite insignificant.

8.5 EFFECT OF SOURCE IMPEDANCE ON NOISE PERFORMANCE

It is shown in Section 8.4 that the output signal-to-noise ratio of an amplifier depends on the following factors:

(i) the input signal level,
(ii) the frequency response of the amplifier,
(iii) the magnitudes of the spot noise generators $d(v_N^2)$ and $d(i_N^2)$,
(iv) the source impedance.

The relative effects of these factors are summarized by Eqs. 8.58. Usually a circuit designer has no control over the input signal level, and the frequency response is fixed by the signal requirements. Thus, with few exceptions, the design problem of maximizing S_o/N_o reduces to attempting to optimize factors (iii) and (iv) over the useful frequency range of the system. In general, these factors are interdependent and it is difficult to achieve a real optimization throughout the frequency range; indeed for some types of reactive source no optimum combination of factors (iii) and (iv) exists even at one spot frequency. However, in many practical cases of amplifiers with moderately large bandwidths, it is possible to maximize the output signal-to-noise ratio by a suitable choice of magnitudes for $d(v_N^2)$, $d(i_N^2)$, and Z_S. For this reason it is appropriate to perform calculations on idealized combinations of source and device that indicate general trends, and can thus assist in the design of nonideal systems.

8.5.1 Resistive Sources

A useful measure of amplifier noise performance in a system is the *noise factor F*, defined as the ratio of source signal-to-noise to the output signal-to-noise. The more closely the noise factor approaches unity, the better is the noise performance. At any frequency, the spot ratio (S_o/dN_o) is given by Eqs. 8.55 and 8.56 as

$$\frac{S_o}{dN_o} = \frac{i_S^2}{(4kT_SG_S + \varphi_I + \varphi_V|G_S|^2)\,df}.$$

The ratio of source signal to spot noise is

$$\frac{S_S}{dN_S} = \frac{i_S^2}{4kT_SG_S\,df}$$

and therefore the *spot noise factor* is

$$F = \frac{S_S/dN_S}{S_o/dN_o} = 1 + \frac{\varphi_I + \varphi_V|G_S|^2}{4kT_SG_S}. \tag{8.82a}$$

The alternative form, obtained when the source is represented by a generator with a series impedance, is

$$F = 1 + \frac{\varphi_V + \varphi_I|R_S|^2}{4kT_SR_S}. \tag{8.82b}$$

Noise factor is discussed further in Section 8.6.1.

A related measure of noise performance is the relative increase in spot noise as the signal passes through an amplifier. It is convenient to introduce the notation of dN_{Ai} and dN_{Si} as the spot noise powers at the amplifier input due respectively to the amplifier noise generators and the source conductance. In terms of this notation, the *relative increase in spot noise* is

$$\frac{dN_{Ai}}{dN_{Si}} = \frac{\varphi_I + \varphi_V|G_S|^2}{4kT_SG_S} \tag{8.83a}$$

or the alternative form

$$\frac{dN_{Ai}}{dN_{Si}} = \frac{\varphi_V + \varphi_I|R_S|^2}{4kT_SR_S}. \tag{8.83b}$$

The increase in total noise can be found by integration. From Eqs. 8.82 and 8.83, the relation between the decrease in signal-to-noise ratio and the relative increase in noise is

$$F = 1 + \frac{dN_{Ai}}{dN_{Si}}. \tag{8.84}$$

Consider the combination of an amplifier and resistive source shown in Fig. 8.14. The resistor R_S is assumed to be the source of all noise introduced before the first amplifier stage, and the temperature T_S may lie above the actual temperature of R_S in order to represent components of source noise which cannot be ascribed to thermal noise in the source resistance. From Eqs. 8.83, the relative increase in spot noise due to the amplifier is

$$\frac{dN_{At}}{dN_{St}} = \frac{\varphi_I R_S + \varphi_V / R_S}{4kT_S}. \tag{8.85}$$

Now, φ_I and φ_V are parameters of the amplifying device and its operating conditions and, for a given R_S, it is often possible to adjust these parameters so that the relative increase in noise is a minimum. This is illustrated in the two simple examples which follow. Alternatively, for a given device under given operating conditions (φ_I and φ_V constant) there is a particular value of R_S for which the relative increase in noise is a minimum. This second condition is somewhat academic in that the designer is not always able to vary R_S; however, it is of considerable practical importance when an input transformer can be used, as in Section 8.5.2.

8.5.1.1 Transistor Amplifier

Take as a specific example the case of a transistor amplifier, where for simplicity and clarity it is assumed Eqs. 8.27 and 8.28 can be reduced to

$$\varphi_V = \frac{2(kT)^2}{qI_E} \tag{8.86}$$

and

$$\varphi_I = \frac{2qI_E}{\beta_N}. \tag{8.87}$$

Equations 8.86 and 8.87 are applicable only for a certain range of I_E; at low values of I_E the contribution of I_{CO} to φ_I becomes important, whereas at high values of I_E the contribution of r_B to φ_V becomes important. This

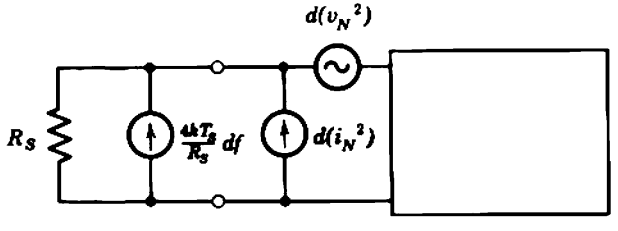


Fig. 8.14 Derivation of relative increase in noise due to amplifier.

latter restriction can be removed by including r_B with the source resistance so that the effective source resistance becomes $(R_S + r_B)$. These expressions are further restricted to the mid-band range of frequencies. Nevertheless the simple treatment which follows indicates a method which may be adopted when the φ functions have a more general character.

Substitution of Eqs. 8.86 and 8.87 into Eq. 8.85, and assuming that T_S and T are the same results in the equation

$$\frac{dN_{Ai}(I_E, R_S + r_B)}{dN_{Si}} = \frac{qI_E}{kT}\left[\frac{(R_S + r_B)}{2\beta_N}\right] + \frac{kT}{qI_E}\left[\frac{1}{2(R_S + r_B)}\right], \quad (8.88)$$

which (subject to the various assumptions involved) specifies the spot noise performance in terms of the source resistance and transistor parameters. The relative increase in noise above that generated by $(R_S + r_B)$ reaches a minimum value

$$\left(\frac{dN_{Ai}}{dN_{Si}}\right)_{\min} = \frac{1}{\sqrt{\beta_N}} \quad (8.89)$$

when I_E satisfies the condition

$$I_{E(\text{opt})} = \sqrt{\beta_N}\,\frac{kT/q}{R_S + r_B}. \quad (8.90)$$

However, it must not be inferred that slight divergences from this optimum condition result in a gross deterioration in noise performance. Indeed it can be demonstrated quite readily that the relative increase in noise is not a strong function of I_E. Perhaps the most convenient way of displaying the information contained in Eq. 8.88 is to plot the I_E versus $(R_S + r_B)$ loci for various constant values of the relative increase in noise. This is done in a fairly general way in Fig. 8.15 by normalizing both source resistance and relative increase in noise to β_N. The range of emitter current is restricted to avoid giving grossly misleading results for conditions beyond the range of validity of Eq. 8.87. As an example, if $(R_S + r_B)$ is $100\sqrt{\beta_N}\ \Omega$, the optimum emitter current is 0.25 mA and (dN_{Ai}/dN_{Si}) has a magnitude $1/\sqrt{\beta_N}$. If, however, (dN_{Ai}/dN_{Si}) is allowed a maximum increase of 20%, the emitter current may lie anywhere in the range from 0.14 to 0.46 mA. In other words, a change in emitter current from 56 to 182% of the optimum produces only a 20% change in the relative increase in noise. Notice that this does not correspond to a 20% decrease in signal-to-noise ratio. The percentage change in S_o/dN_o is equal to the percentage change in noise factor and, from Eq. 8.84, the percentage change in F is always less than the percentage change in (dN_{Ai}/dN_{Si}). Figure 8.15 is useful as a direct design aid for most transistors for the current range shown, since the influence of I_{CO} and the

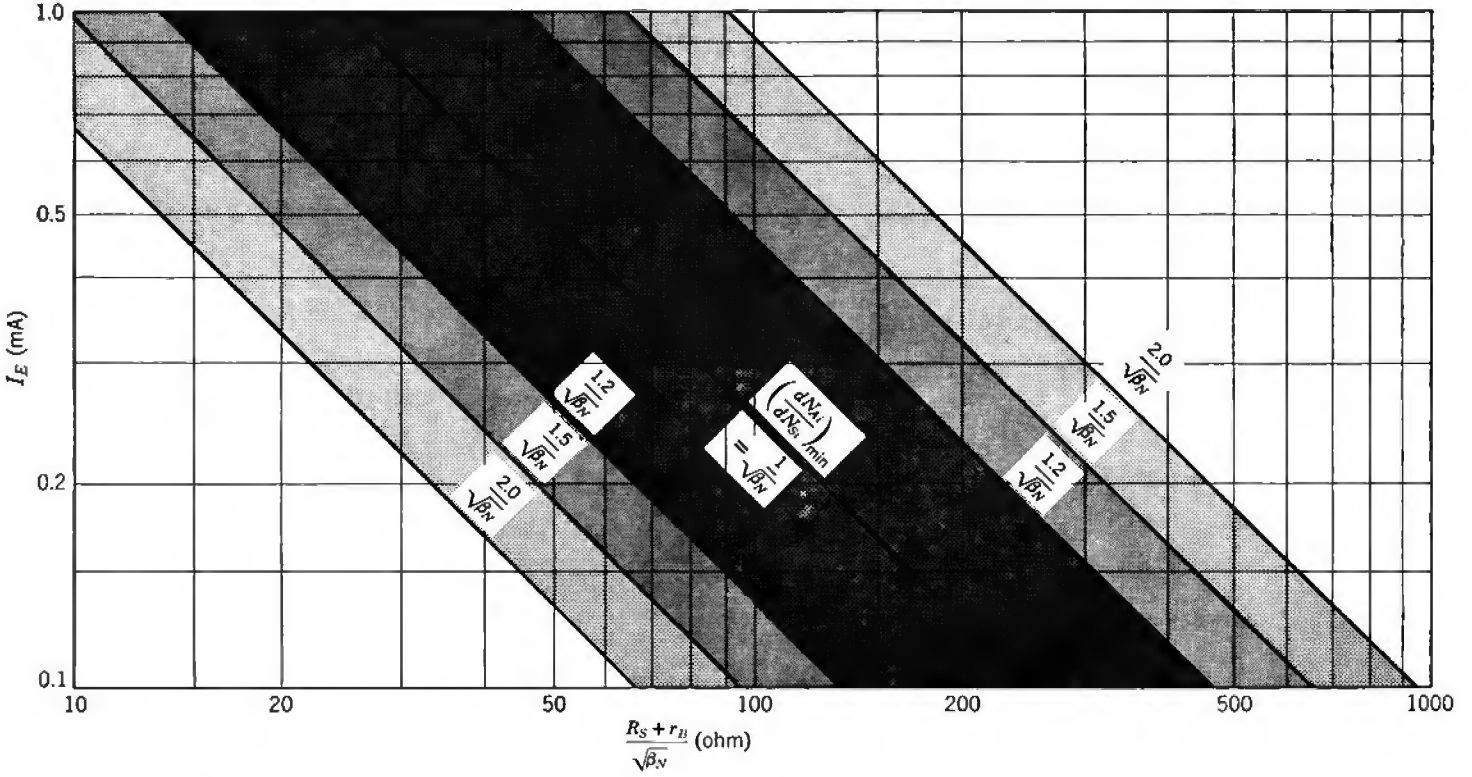


Fig. 8.15 Contours of constant transistor dN_{Ai}/dN_{Si} versus source resistance and emitter current.

current dependence of β_N are small. However, if the source impedance is high and the transistor must be operated at currents below 100 μA, these factors may have to be considered.

For amplifiers whose bandwidths are such that the frequency-dependent components of φ_V and φ_I are negligible in comparison with the constant components, the condition for minimum increase in spot noise (Eq. 8.90) is also the condition for minimum increase in wide-band noise (that is, for maximum signal-to-noise ratio). Roughly speaking, the noise theory presented here may be applied directly to transistor audio amplifiers having resistive sources, but probably should not be applied to subaudio amplifiers or video amplifiers. In these cases the frequency dependence of the spot noise generators requires that an integration be performed before the conditions for maximum signal-to-noise ratio can be determined. The principle of this more general (and more complicated) problem is considered in the discussion of reactive sources in Section 8.5.3.

8.5.1.2 Vacuum-Tube Amplifier

The vacuum-tube case is somewhat less tractable in that there is no simple general derivation of the optimum conditions of operation. Nevertheless, a semianalytical approach is possible, and it is shown that the noise performance that can be achieved is a function of the source resistance.

At mid-band frequencies the spot noise generators at the input of a triode are given by Eqs. 8.12*b* and 8.15:

$$d(v_N^2) = \frac{3.5 \times 10^{-20}}{g_m}\, df \tag{8.91}$$

and

$$d(i_N^2) = 2q\,\Sigma(I_G)\, df. \tag{8.92}$$

In order to proceed with a simple, moderately general form of analysis, it is necessary to express these noise generators as functions of anode current. Equation 3.36 shows that g_m is proportional to the cube root of I_A, and substituting values gives

$$d(v_N^2) = \frac{2.3 \times 10^{-20}}{G^{2/3} I_A^{1/3}}\, df = \frac{K_V}{I_A^{1/3}}\, df, \tag{8.93}$$

where G is the perveance and K_V is a constant for any given tube. There is no simple general relation between the anode and grid currents. However, it is usual to operate a tube with its grid somewhat more negative than the free grid potential. Under these conditions I_G depends largely on the ionization rate of the residual gas molecules, and this rate depends on the density and mean energy of the electrons in transit from cathode

to anode. There is some evidence to suggest that I_G for many tubes obeys the approximate law

$$\Sigma(I_G) = \nu V_A^{3/2} I_A \tag{8.94}$$

and the constant ν is of the order 10^{-10}. For normal values of V_A and moderate values of the amplification factor μ, V_A in Eq. 8.94 may be replaced from Eq. 3.35 by

$$V_A^{3/2} \approx \frac{\mu^{3/2}}{G} I_A$$

and it follows from Eq. 8.92 that

$$d(i_N^2) \approx \frac{2q\nu\mu^{3/2} I_A^2}{G} df = K_I I_A^2 \, df, \tag{8.95}$$

where K_I is a constant for any given tube.

Equations 8.93 and 8.95 express the input noise generators as simple functions of I_A. In practice, these precise functional relationships are not observed and it is convenient to generalize the equations by using exponents $-m$ and n for the powers of I_A. The equations become

$$d(v_N^2) = K_V I_A^{-m} \, df \tag{8.96}$$

and

$$d(i_N^2) = K_I I_A^n \, df. \tag{8.97}$$

The values of m and n can be determined experimentally over various limited ranges of operating conditions; Fairstein* quotes the results given in the first three lines of Table 8.3.

The relative increase in noise follows from Eq. 8.85 as

$$\frac{dN_{AI}}{dN_{SI}} = \frac{K_I I_A^n R_S + K_V I_A^{-m}/R_S}{4kT_S}. \tag{8.98}$$

Table 8.3 Noise Exponents for Typical Tubes

Tube Type	I_A	m	n
Triode, low anode voltage	1 mA	0.85	1.1
Triode, low anode voltage	10 mA	0.5	2.0
Pentode	1 mA	0.7	2.0
Approximate calculation (above)		0.33	2

* In the book: A. H. SNELL (ed.), *Nuclear Instruments and Their Uses*, Vol. 1 (Wiley, New York, 1962).

It is neither convenient nor useful to produce a general plot of Eq. 8.98, but in particular cases the optimum value of I_A for a given R_S is

$$(I_{A(\text{opt})})^{(m+n)/2} = \frac{\sqrt{K_V/K_I}}{R_S} \tag{8.99}$$

and the minimum relative increase in noise is

$$\left(\frac{dN_{At}}{dN_{St}}\right)_{\min} = \frac{\sqrt{K_V K_I}}{2kT_S} (I_{A(\text{opt})})^{(n-m)/2}. \tag{8.100}$$

Equation 8.100 gives the minimum relative increase in noise as a function of optimum anode current. It suggests that the absolute minimum occurs at zero anode current when the source resistance is infinite. This result is of course useless (and is questionable also because the exponents m and n in Eqs. 8.96 and 8.97 are not true constants) but it does indicate that there is no simple derivation of the conditions for absolute minimum noise. Studies of spot noise performance of vacuum tubes are best persued numerically for each particular case.

8.5.2 Noise Reduction by Input Transformer

In many cases the designer has no control over the magnitude of source resistance so that he can only choose a suitable amplifying device and then operate it so that its noise performance is acceptable. A very important exception occurs with audio amplifiers in which the turns ratio of an input transformer can be varied to change the effective source impedance seen by the first amplifying device.

For the reasons given in Section 7.6, it is very difficult to construct transformers which can be used over wide bandwidths. If attention is restricted to the mid-band frequency range of the transformer, the effect of noise performance is summarized by Fig. 8.16. The internal source current i_S, the noise associated with R_S, and the amplifier noise generators $d(v_N^2)$ and $d(i_N^2)$ are shown in Fig. 8.16*a*. All sources may be referred to the secondary as in Fig. 8.16*b*, and the relative increase in noise at the amplifier input is

$$\frac{dN_{At}}{dN_{St}} = \frac{d(v_N^2) + N^4 R_S^2 d(i_N^2)}{4kT_S N^2 R_S\, df} = \frac{\varphi_V/(N^2 R_S) + \varphi_I (N^2 R_S)}{4kT_S}. \tag{8.101}$$

This verifies the rather obvious fact that, as far as the amplifier is concerned, the apparent source resistance is $N^2 R_S$. The condition for optimum noise performance is

$$N_{\text{opt}}^2 = \frac{\sqrt{\varphi_V/\varphi_I}}{R_S} \tag{8.102}$$

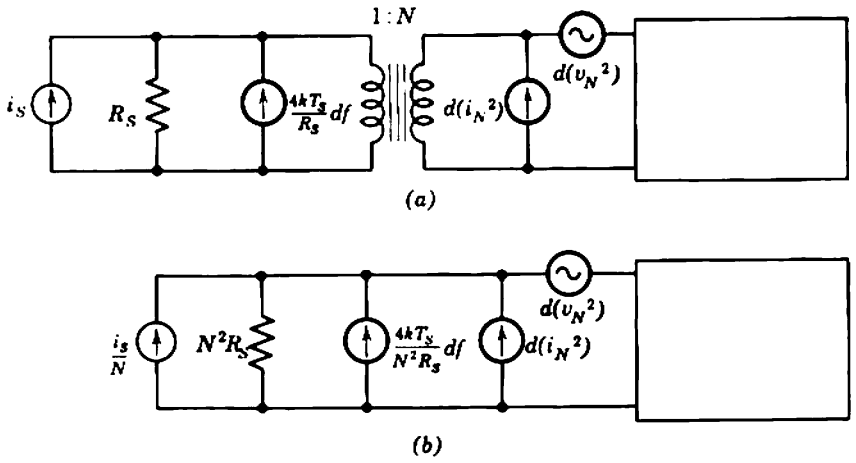


Fig. 8.16 Noise reduction with an input transformer: (*a*) circuit diagram; (*b*) equivalent representation.

and the minimum relative increase in noise is

$$\left(\frac{dN_{Ai}}{dN_{Si}}\right)_{\min} = \frac{\sqrt{\varphi_V \varphi_I}}{2kT_S}. \tag{8.103}$$

The optimum turns ratio can usually be achieved over the audio bandwidth for transistors, but may be more difficult to achieve with vacuum tubes. However, it must again be emphasized that moderate departures from the optimum condition do not produce a great deterioration in noise performance. This can be illustrated by reference to the transistor amplifier discussed in Section 8.4.4.2. For this amplifier, the mid-band spot noise generators are given by Eqs. 8.76 and 8.77 as

$$\varphi_V = 3.52 \times 10^{-18},$$

$$\varphi_I = 3.14 \times 10^{-24}.$$

Substituting these values together with $R_S = 20\ \Omega$ into Eqs. 8.102 and 8.103 gives the optimum turns ratio and the minimum relative increase in noise as 7.27 and 0.416, respectively. The actual turns ratio is only 5, that is, rather less than half the optimum impedance ratio, and the actual relative increase in noise follows as 0.538. Substituting these two values of (dN_{Ai}/dN_{Si}) into Eq. 8.84 shows that the increase in output signal-to-noise ratio (the increase in F) in going from the actual to the optimum turns ratio is only 8%.

8.5.2.1 Optimum Source Resistance

Equation 8.85 (or Eq. 8.101 for an amplifier with an input transformer) is often used to define an optimum source resistance. The relative increase in noise reaches a minimum value given by Eq. 8.103

$$\left(\frac{dN_{Ai}}{dN_{Si}}\right)_{\min} = \frac{\sqrt{\varphi_V \varphi_I}}{2kT_S} \tag{8.103}$$

when the source resistance has the value

$$R_{S(\text{opt})} = \sqrt{\frac{\varphi_V}{\varphi_I}}. \tag{8.104}$$

This optimum source resistance is of limited practical application. However, a number of mistaken ideas about it are quite prevalent, so it is worthy of a brief discussion.

The concept of optimum source resistance is significant only when an input transformer is used to change the effective source resistance to the value that maximizes the signal-to-noise ratio as discussed in Section 8.5.2. Even in this case, it is really the turns ratio which is optimized. The optimum source resistance is not the value that minimizes the noise output from an amplifier. Indeed, as R_S is increased, the output noise rises from a value determined by φ_V alone at zero R_S, reaches a maximum, and then falls to the value determined by φ_I alone at infinite R_S. The two noise minima thus occur at zero and infinite R_S. However, when an input transformer is used to change the effective source resistance, it changes the effective source generator also. As the turns ratio is increased from zero (that is, zero effective source resistance and voltage generator), the output signal-to-noise ratio rises initially even though the output noise is rising, because the signal output rises faster. S_o/N_o reaches a maximum at the optimum turns ratio, and falls again beyond this. At very high turns ratios when the effective source resistance approaches infinity, S_o/N_o falls even though the noise output is falling, because the signal output falls faster. In general, the optimum turns ratio which gives the peak of S_o/N_o does not give the peak in noise output.

If some means is available for changing the effective source resistance without changing the effective source generator, then

(i) for a constant source current generator, S_o/N_o is maximum for infinite R_S;
(ii) for a constant source voltage generator, S_o/N_o is maximum for zero R_S.

It follows that changing the effective source resistance by adding shunt or series resistors can only degrade the noise performance.

8.5.3 Reactive Sources

The simple considerations of Section 8.5.1 and 8.5.2 refer strictly only to spot noise performance, but in practice the conditions deduced for minimum spot noise are near optimum for total noise in many situations. They are usually adequate for amplifiers that have a flat frequency response and moderate bandwidth. However, in more general cases the derivation of the optimum conditions should take into account the variation of spot noise with frequency. This latter procedure is certainly necessary for amplifiers which have a very wide bandwidth, or whose gain is a function of frequency in the passband, or which are fed from a reactive source. The present section is concerned with reactive sources whose impedance varies appreciably over the signal frequency range. Two illustrative examples of interest are purely inductive and purely capacitive sources.

Under the most general circumstances, the output signal-to-noise ratio of a system is given by Eqs. 8.58 provided that there is no correlation between the noise generators at the input. Often these equations simplify. In the examples of Sections 8.5.1 and 8.5.2, both Y_S and Y_i are real. Another common situation is for the transfer function from source generator to amplifier output to be independent of frequency over the signal frequency range, even though Y_S, Y_i, and A_V may separately vary with frequency. Under these circumstances, Eqs. 8.63 give the signal-to-noise ratio and Eqs. 8.66 give the equivalent noise referred to the input *provided that $\psi(f)$ is interpreted as the frequency response of the system.* Equations 8.63 and 8.66 are a convenient starting-point for an investigation of amplifier noise when the source is reactive.

8.5.3.1 Inductive Sources

Inductive sources are quite commonly encountered in audio-frequency applications. Familiar examples are the magnetic transducers used for signal pickup from magnetic tapes and phonograph disks. In either transducer, a time-varying magnetic flux $\Phi(t)$ is established in the gap between pole pieces of a coil, and the resultant emf induced in the coil is proportional to the time rate of change of flux and the number of turns N. The precise manner in which the flux is varied in order to follow the signal information on the tape or disk is beyond the scope of this book. However, if the existence of a constant-flux recording characteristic is assumed, the open-circuit voltage of the transducer for a sinusoidal flux at any frequency f is given by

$$v_S = \text{const}_1 \, Nf.$$

The inductance of the source is proportional to the square of the number of turns in the coil:

$$L_S = \text{const}_2\, N^2.$$

These expressions may be combined to give

$$v_S = \kappa\sqrt{L_S}\, f, \tag{8.105}$$

where κ is a constant. The short-circuit output current from the transducer is therefore

$$i_S = \frac{v_S}{2\pi f L_S} = \frac{\kappa}{2\pi\sqrt{L_S}}, \tag{8.106}$$

independent of frequency. Thus, a constant current source representation is pertinent for a magnetic transducer with a constant flux input.

The circuit whose noise performance is to be investigated is shown in Fig. 8.17. The signal-to-noise ratio follows from Eq. 8.63*a* with G_S set at zero:

$$\frac{S_o}{N_o} = \frac{\kappa^2/4\pi^2 L_S}{\int_0^\infty [\varphi_I + \varphi_V/(2\pi f L_S)^2]|\psi(f)|^2\, df}. \tag{8.107}$$

This expression can be used in performing particular studies on optimum combinations of L_S, φ_V, and φ_I. It is perhaps most appropriate to restrict discussion in this section to the effect of L_S on noise performance. This is not a completely artificial restriction, because the designer may well have the value of L_S under his control directly, or he may be able to alter its effective value by using a transformer. The maximum signal-to-noise ratio for a given amplifier under given operating conditions occurs when

$$L_{S(\text{opt})}^2 = \frac{1}{4\pi^2}\frac{\int_0^\infty (\varphi_V/f^2)|\psi(f)|^2\, df}{\int_0^\infty \varphi_I|\psi(f)|^2\, df}. \tag{8.108}$$

in which case

$$\left(\frac{S_o}{N_o}\right)_{\max} = \frac{\kappa^2/8\pi^2 L_{S(\text{opt})}}{\int_0^\infty \varphi_I|\psi(f)|^2\, df}. \tag{8.109}$$

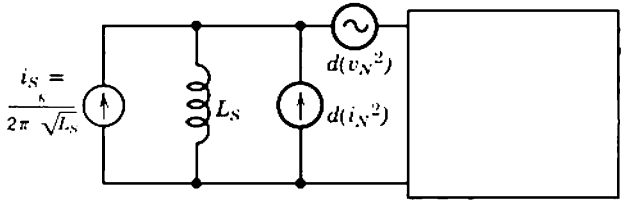


Fig. 8.17 Noise performance with an inductive source.

The integrals involved in Eq. 8.108 can often be evaluated simply. For example, suppose the system transfer function (v_o/i_S) has one low-frequency break at f_1 and one high-frequency break at f_2:

$$|\psi(f)|^2 = \left[\frac{(f/f_1)^2}{1 + (f/f_1)^2}\right]\left[\frac{1}{1 + (f/f_2)^2}\right].$$

If φ_V and φ_I are independent of frequency, integrals (4) and (2) of Table 8.2 may be used to show that

$$\int_0^\infty \frac{\varphi_V}{f^2} |\psi(f)|^2 \, df \approx \frac{\pi}{2}\frac{\varphi_V}{f_1}$$

and

$$\int_0^\infty \varphi_I |\psi(f)|^2 \, df \approx \frac{\pi}{2}\varphi_I f_2.$$

Therefore, the optimum source inductance is

$$L_{S(\text{opt})}^2 = \frac{\varphi_V/\varphi_I}{4\pi^2 f_1 f_2}.$$

For the transistor operating at 0.625 mA emitter current considered in Section 8.4.4.2, the mid-band noise functions are given by Eqs. 8.76 and 8.77 as

$$\varphi_V = 3.52 \times 10^{-18},$$

$$\varphi_I = 3.14 \times 10^{-24}.$$

If f_1 and f_2 are taken as 30 Hz and 15 kHz, respectively, the optimum source inductance is

$$L_{S(\text{opt})} = 250 \text{ mH},$$

which is a practically realizable value. However, as shown in other calculations, this optimum source impedance is not critical. A typical short-circuit current i_S from a 250-mH tape playback head on signal peaks is 3 μA rms; κ therefore has the value 10^{-5}. For the particular ψ function considered in this example, Eq. 8.109 reduces to

$$\frac{S_o}{N_o} = \frac{\kappa^2}{4\pi^3 \varphi_I L_{S(\text{opt})} f_2}$$

$$= 6.7 \times 10^7$$

$$= 78 \text{ dB}.$$

This is a very high signal-to-noise ratio for an audio system.

8.5.3.2 Capacitive Sources

Noise performance with a capacitance source is important for two reasons. First, a number of commonly encountered transducers do in fact have a capacitive output impedance; piezoelectric transducers, many nuclear particle detectors, and most photoelectric transducers such as television camera tubes are familiar examples. Second, and perhaps more important, all sources become capacitive at high frequencies when the reactance of the stray shunt capacitance becomes small. High-frequency noise in wide-band amplifiers is often calculated on the assumption that the source is capacitive.

Piezoelectric transducers and nuclear particle detectors have an open-circuit output voltage proportional to the charge present in the transducer at any instant. These transducers may be represented by the equivalent circuit shown in Fig. 8.18*a*, but it is convenient to transform the equivalent circuit from a series arrangement to the shunt arrangement shown in Fig. 8.18*b*. The stray capacitance can then be lumped with the

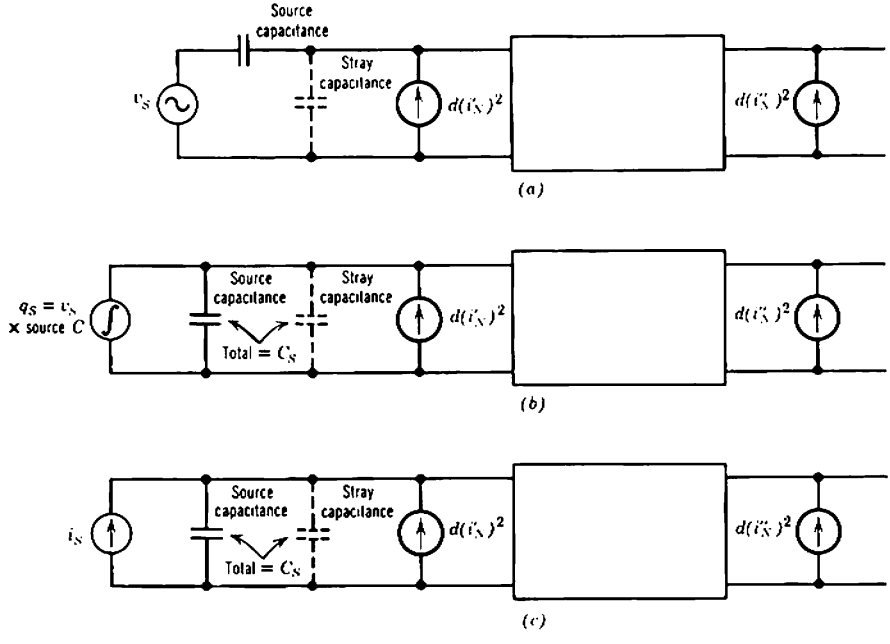


Fig. 8.18 Noise performance with a capacitive source: (*a*) charge-type source, voltage representation; (*b*) charge-type source, current representation; (*c*) current-type source.

transducer output capacitance to give a single effective source capacitance C_S. Notice that

$$q_S = \int i_S \, dt = \frac{i_S}{2\pi f} \tag{8.110}$$

for sinusoidal excitation. Photoelectric transducers such as vidicon camera tubes have a short-circuit output current that depends on the incident light intensity. The equivalent circuit is shown in Fig. 8.18*c*, in which the stray capacitance is lumped with the output capacitance of the transducer as an effective source capacitance C_S.

In general, the two spot noise generators $d(v_N^2)$ and $d(i_N^2)$ at the input of a device are correlated at high frequencies. Therefore, the noise model with two noise generators at the input should not be used when calculating the noise performance of a wide-band amplifier (unless, of course, correlation between the generators is taken into account). As discussed in Section 8.1.1, a model with uncorrelated generators should be used. The spot noise of a device can be represented by the generators $d(i_N')^2$ and $d(i_N'')^2$ in Fig. 8.1*c*, and it is shown in Section 8.1.1 that correlation between these generators is insignificant. The spot signal-to-noise ratios for Figs. 8.18*b* and *c* follow as

$$\frac{S_o}{dN_o} = \frac{(2\pi f)^2 q_S^2}{d(i_N')^2 + \dfrac{d(i_N'')^2}{|\beta|^2}}, \tag{8.111a}$$

$$\frac{S_o}{dN_o} = \frac{i_S^2}{d(i_N')^2 + \dfrac{d(i_N'')^2}{|\beta|^2}}, \tag{8.111b}$$

where β *is the short-circuit current gain from the source to the device output,* not the current gain of the device alone.

The relative influences of $d(i_N')^2$ and $d(i_N'')^2$ on spot noise depend on both the frequency and the device type. An integration must therefore be performed to find the conditions for minimum total noise. Subject to the assumption that the amplifier is designed so that the transfer function from source charge or current (as appropriate) to output is independent of frequency over the signal frequency range, the total signal-to-noise ratio follows from the discussion leading to Eqs. 8.63 and either Eq. 8.111*a* or *b* as

$$\frac{S_o}{N_o} = \frac{q_S^2}{\displaystyle\int_0^\infty \frac{|\psi(f)|^2}{(2\pi f)^2}\left[d(i_N')^2 + \frac{d(i_N'')^2}{|\beta|^2}\right]}, \tag{8.112a}$$

$$\frac{S_o}{N_o} = \frac{i_S^2}{\displaystyle\int_0^\infty |\psi(f)|^2\left[d(i_N')^2 + \frac{d(i_N'')^2}{|\beta|^2}\right]}. \tag{8.112b}$$

Note that $\psi(s)$ gives the poles and zeros of the system, not of the amplifier alone, and that β is the current gain from the source to the output. The denominators of Eqs. 8.112*a* and *b* may be interpreted as equivalent mean-square noise charge and noise current referred to the amplifier's input. The remainder of this section deals only with noise current, Eq. 8.112*b*; the topic of noise charge is taken up in Section 8.7.

For devices such as vacuum tubes and field-effect transistors which have a very small dc control-electrode current, $d(i_N')^2$ in the denominator of Eq. 8.112*b* can usually be neglected. The shot term in $d(i_N')^2$ is small because the dc control-electrode current of these devices is small, and the induced grid/gate noise term is small except at frequencies beyond several tens of megahertz. The value of $d(i_N'')^2$ is given in terms of the more familiar quantity, the noise voltage generator at the input, as

$$d(i_N'')^2 = g_m{}^2 d(v_N{}^2) = g_m{}^2 \varphi_V \, df$$

and the system current gain β is

$$|\beta| = \frac{g_m}{2\pi f(C_i + C_S)}.$$

Thus, Eq. 8.112*b* reduces to

$$\frac{S_o}{N_o} = \frac{i_S{}^2}{\int_0^\infty 4\pi^2 f^2 (C_i + C_S)^2 \varphi_V |\psi(f)|^2 \, df}.$$

But φ_V is independent of frequency, and may be taken outside the integral to give

$$\frac{S_o}{N_o} = \frac{i_S{}^2}{4\pi^2 (C_i + C_S)^2 \varphi_V \int_0^\infty f^2 |\psi(f)|^2 \, df}. \tag{8.113}$$

The noise increases monotonically with increasing source capacitance. There is therefore no finite optimum source capacitance, and the design problem of maximizing S_o/N_o reduces to minimizing φ_V. For triodes, Eq. 8.12 gives $d(v_N{}^2)$ $(\equiv \varphi_V \, df)$; its value decreases with increasing g_m, and therefore with increasing I_K. Consequently, there is no finite cathode current that minimizes the high-frequency noise. However, grid current tends to increase at high cathode currents, and the shot term in $d(i_N')^2$ may become significant. The practical conclusion is that a vacuum tube with a capacitive source should be operated at the highest cathode current (consistent with its ratings) that does not cause excessive grid current noise. It is probable that the same conclusion applies to f.e.t.s; increasing the drain current heats the device and increases the gate current noise.

The situation for a bipolar transistor is different in that a small optimum emitter current exists for any combination of transistor, source capacitance,

and bandwidth. Essentially, this is because of the appreciable dc base current which contributes significantly to $d(i_N')^2$. Although increasing the emitter current reduces the high-frequency noise as shown above for tubes and f.e.t.s, it also increases the dc base current and its noise contribution over the entire frequency range. The noise generators of an intrinsic transistor are given by Eqs. 8.21 and 8.23 as

$$d(i_N')^2 \approx 2q[(1 - \alpha_N)I_E + I_{CO}]\,df,$$

$$d(i_N'')^2 \approx 2q\alpha_N I_E\,df.$$

A general calculation of the optimum emitter current is tedious in the extreme; useful approximate results, however, can be obtained quite easily. Two approximations are made and both are good in practice.

1. The transistor is a high-gain high-frequency type for which β_N is large and τ_1 is small. It is also assumed that the optimum emitter current is small, so that the base-charging capacitance c_B is much smaller than the emitter transition capacitance c_{tE}. The current gain from the source to the transistor collector then follows as

$$|\beta(f)| \approx \frac{g_m}{2\pi f(C_i + C_S)} = \frac{(qI_E/kT)}{2\pi f(c_{tE} + C_S)}.$$

2. The factor $|\psi(f)|^2$ is the "brick wall" function discussed in Section 13.1, given by

$$|\psi(f)|^2 = 1 \text{ in the passband,}$$

$$|\psi(f)|^2 = 0 \text{ outside the passband.}$$

Therefore

$$\int_0^\infty [\text{function } (f)]|\psi(f)|^2\,df = \int_{f_L}^{f_H} [\text{function } (f)]\,df.$$

It is also assumed that the upper cutoff frequency f_H is much greater than the lower cutoff frequency f_L.

If these functions are substituted into Eq. 8.112*b*, the optimum emitter current follows as

$$I_{E(\text{opt})} = 2\pi \frac{kT}{q}(c_{tE} + C_S)\sqrt{\frac{\beta_N}{3}}\,f_H, \tag{8.114}$$

and the maximum signal-to-noise ratio is

$$\left(\frac{S_o}{N_o}\right)_{\max} = \frac{i_S^2}{2qI_{CO}f_H + 8\pi kT\dfrac{(c_{tE} + C_S)}{\sqrt{3\beta_N}}f_H^2}. \tag{8.115}$$

In addition to the small value of τ_1 assumed above, the desirable features of a low-noise transistor are:

(i) high β_N,
(ii) low I_{CO},
(iii) small c_{tE}.

The optimum emitter current for a typical low-noise transistor is of the order of 100 μA, dependent, of course, on the detail of the circuit.

If there were no recombination in transistors (that is, if the dc base current were negligible) the wide-band signal-to-noise ratio for a transistor with a capacitive source would be much higher than for a comparable vacuum tube with the same source. This is because φ_V is much smaller for a transistor than a tube, due to the former's higher ratio of g_m to I—see the discussion at the end of Section 8.4.4.2. However, recombination is an unfortunate reality, and the noise performance of a transistor suffers accordingly. Typical low-noise tubes and transistors have roughly comparable noise performance in applications such as television camera amplifiers—about 2 nA rms noise current in 5 MHz bandwidth when the source capacitance is 30 pF. Lower bandwidths and lower capacitances favor tubes whereas higher bandwidths and capacitances favor transistors. Field-effect transistors are undergoing rapid technological development, and already (1966) appear superior to either of the older device types; noise currents of 1 nA rms have been reported for room-temperature junction f.e.t.s,* and 0.5 nA for f.e.t.s cooled with liquid nitrogen to reduce their thermal noise.

8.6 OTHER METHODS OF SPECIFYING NOISE PERFORMANCE

Two main types of noise specification are in use; these may be classed as *fundamental* and *particular* specifications.

Fundamental specifications provide information that enables the noise performance of a particular device to be determined in any amplifier circuit. Examples are the fundamental noise models discussed in the appropriate sections of Chapters 3 and 4, or more convenient equivalent forms such as the input noise generators $d(v_N^2)$ and $d(i_N^2)$ introduced in Section 8.2. In such specifications the values of at least some of the noise generators and equivalent-circuit impedances are functions of frequency. Therefore, if the complete noise performance of any amplifier is to be

* A. J. SMITH, *Personal Communication*, U.K.A.E.A., Harwell, Berkshire.

determined, the magnitudes of the generators and/or impedances must be known at a number of spot frequencies.

Particular specifications provide direct information on the noise performance of a complete amplifier in its intended application. An example of this type of specification is the signal-to-noise ratio discussed in Section 8.3. Particular specifications therefore provide a direct basis for comparing the noise performance of competing devices or circuits which fulfill a given signal performance specification.

The most generally useful methods of these two classes for presenting noise specifications are probably the examples mentioned above, namely, a fundamental noise specification in terms of spot values of $d(v_N^2)$ and $d(i_N^2)$, or a particular specification in terms of output signal-to-noise ratio. However, other methods for specifying noise performance do exist and, despite being more cumbersome in amplifier noise calculations, they have gained acceptance under special circumstances. These other methods include spot noise factor, average noise factor, noise temperature, noise resistance, and number of ion pairs. Spot noise factor, noise temperature, and noise resistance are fundamental specifications, whereas average noise factor and the number of ion pairs are particular specifications. It is desirable to interrelate the information provided in these varying ways, and the basis of these additional noise specifications should be understood. However, it must be emphasized at the outset that, although it is always possible to determine a particular specification from a complete fundamental specification, the reverse procedure usually is impossible. For this reason, device manufacturers' noise data should be presented as a fundamental specification and not as a particular specification for some arbitrarily-chosen test circuit.*

8.6.1 Noise Factor

The spot noise factor F is introduced in Section 8.5.1. It can be defined in a variety of equivalent ways, of which the following is typical:

For a given amplifier, fed from a source of admittance Y_S at temperature T_S, the value of F at any specified frequency f_0 is

$$F = \frac{\text{total output noise power at } f_0}{\text{output noise power at } f_0 \text{ due to source}}. \tag{8.116}$$

Noise factor is unity for an ideal noiseless amplifier, but for any practical amplifier F exceeds unity. Noise factor is meaningless if the source is

* The remainder of Section 8.6 may be omitted on a first reading.

noiseless. Often F is quoted in decibels:

$$(F)_{\text{dB}} = 10 \log_{10} (F)_{\text{numeric}}. \tag{8.117}$$

The temperature T_S is often taken as a standard value 290°K.

Much of the theory of Section 8.5.1 can be applied to the general case in which Y_S has conductive and susceptive components G_S and B_S, respectively. Thus

$$F = 1 + \frac{\varphi_I + \varphi_V(G_S^2 + B_S^2)}{4kT_S G_S}. \tag{8.118}$$

The minimum value of F occurs when

$$G_{S(\text{opt})} = \sqrt{\frac{\varphi_I}{\varphi_V}} \tag{8.119a}$$

and

$$B_{S(\text{opt})} = 0. \tag{8.119b}$$

Under these circumstances

$$F_{\text{min}} = 1 + \frac{\sqrt{\varphi_I \varphi_V}}{2kT_S}. \tag{8.120}$$

For values of G_S above or below $G_{S(\text{opt})}$, the magnitude of F increases as shown in Fig. 8.19. If the source conductance differs markedly from $G_{S(\text{opt})}$, the noise factor depends only on one of the noise functions φ_V or

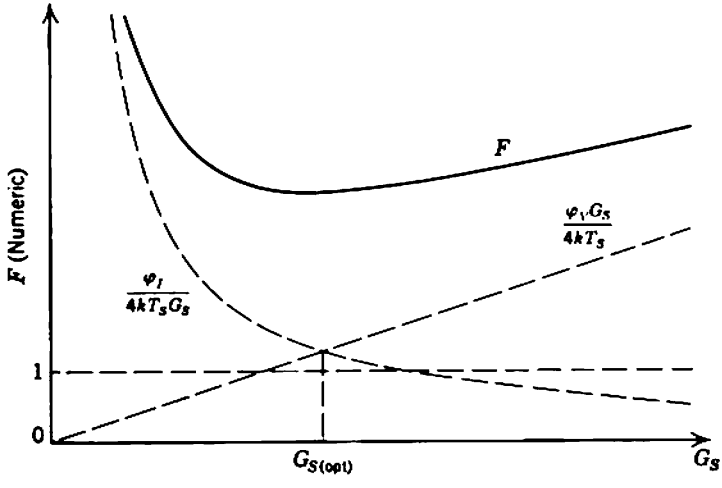


Fig. 8.19 Variation of noise factor with source conductance.

φ_I. Note that $G_{S(\text{opt})}$ depends on the ratio of φ_I to φ_V, whereas the magnitude of F_{min} depends on the product of φ_I and φ_V. Therefore, if information is provided on the values of both $G_{S(\text{opt})}$ and F_{min}, the values of φ_V and φ_I can be determined. In contrast, the mere statement of the magnitude of F at a given value of G_S is not complete specification of spot noise performance; it is unfortunate that manufacturers' data are often presented in this imprecise way. Finally, it should be emphasized that Eq. 8.118 is based on the assumption that the noise generators $d(i_N^2)$ and $d(v_N^2)$ are uncorrelated. If partial correlation exists, the minimum value of F occurs not when B_S is zero, but when B_S has some finite value. The reader is referred to specialized books for further information.

Average noise factor $\bar{F}$ is, as its name implies, the average value of F over some specified frequency range. Thus, the average noise factor over the frequency range f_L to f_H is

$$\bar{F} = \frac{1}{f_H - f_L} \int_{f_L}^{f_H} F \, df. \tag{8.121}$$

8.6.2 Noise Temperature

The noise temperature T_E of an amplifier may be defined as follows:

If the amplifier is fed from a source of admittance Y_S at absolute zero temperature, the spot noise output dN_{oo} originates entirely from sources within the amplifier. If the source temperature is raised, the source becomes noisy and the output noise increases. The temperature T_E at which the output noise doubles is the noise temperature of the amplifier.

In general, the ratio of spot noise output from an amplifier to the spot noise with the source at absolute zero is

$$\frac{dN_o}{dN_{oo}} = \frac{4kT_S G_S + \varphi_I + \varphi_V |Y_S|^2}{\varphi_I + \varphi_V |Y_S|^2}.$$

At the noise temperature T_E,

$$\frac{dN_o}{dN_{oo}} = 2$$

and therefore

$$4kT_E G_S = \varphi_I + \varphi_V |Y_S|^2.$$

But the noise factor of the amplifier follows from Eq. 8.118 as

$$F = 1 + \frac{\varphi_I + \varphi_V |Y_S|^2}{4kT_S G_S}.$$

Substitution gives

$$4kT_E G_S = (F - 1)4kT_S G_S$$

and therefore

$$T_E = (F - 1)T_S. \tag{8.122}$$

Often T_S is taken as the standard value 290°K.

The two spot noise parameters F and T_E are directly equivalent, and Table 8.4 gives the relation between a number of values. Noise temperature provides a convenient expanded measure of F near the point at which F is unity. It is unusual to find a conventional amplifier with a value of F less than about 1.1 (0.42 dB); the corresponding lower limit to noise temperature is about 30°K. However, the molecular amplifiers used at microwave frequencies can operate at rather lower noise temperatures. The noise temperature specification is mainly employed with these molecular amplifiers.

Table 8.4 Noise Factor and Noise Temperature

F(ratio)	F(dB)	T_E(°K)
1.00	0	0
1.01	0.04	2.9
1.05	0.21	14.5
1.10	0.42	29
1.50	1.76	145
2.0	3.01	290
2.5	3.98	435
3.0	4.77	580
4.0	6.02	870
8.0	9.03	2030

8.6.3 Noise Resistance

As the source resistance for an amplifier is increased from zero, the noise output first rises, reaches a maximum, and then falls again. The noise resistance is defined as the smaller value of the source resistance for which the noise power output is twice the value with zero source resistance.

Considering the mid-band case only, the spot noise voltage appearing across the input resistance of an amplifier is

$$d(v_{Ni}^2) = \left(\frac{R_i}{R_S + R_i}\right)^2 (4kTR_S + \varphi_V + \varphi_I R_S^2)\, df.$$

The noise resistance R_N satisfies the equation

$$d(v_{Nt}^2)_{R_S = R_N} = 2d(v_{Nt}^2)_{R_S = 0};$$

that is,

$$\left(\frac{R_t}{R_N + R_t}\right)^2 (4kTR_N + \varphi_V + \varphi_I R_N^2)\, df = 2\varphi_V\, df. \tag{8.123}$$

The smaller of the two solutions of Eq. 8.123 is

$$R_N \approx \frac{\varphi_V}{4(kT - \varphi_V/R_t)}. \tag{8.124}$$

For amplifiers of the type considered in this book, the second term in the denominator of Eq. 8.124 is negligible. Therefore,

$$4kTR_N \approx \varphi_V$$

and, multiplying by df,

$$4kTR_N\, df \approx \varphi_V\, df = d(v_N^2). \tag{8.125}$$

Thus, the noise resistance of an amplifier is the resistance whose thermal noise voltage is numerically equal to the noise voltage generator at the amplifier input.

8.7 NOISE IN NUCLEAR PARTICLE DETECTING SYSTEMS*

In nuclear particle detectors, there is an approximate proportionality between the energy of any particle which is stopped and the resulting peak charge produced in the detector. Until recently, ionization chambers were used as detectors and the charge could be expressed in terms of the number of ion pairs produced in the chamber. With the more modern *p-n* junction detectors, charge production arises from the formation of hole-electron pairs in the depletion layer. Nevertheless, the two processes are similar, and the term "ionization" may be applied to both. To a degree of accuracy sufficient for this book, the energy E of a particle is related to the number of ion pairs produced by

$$(\text{n.i.p.})_S = \frac{E}{q\Phi_i}, \tag{8.126}$$

where Φ_i is the ionization potential of the gas in the chamber, and the suffix S indicates a signal quantity. The charge collected is

$$q_S = \frac{E}{\Phi_i}. \tag{8.127}$$

* Section 8.7 may be omitted on a first reading.

To a good approximation the step-function signal charge in the detector generates a voltage step at the grid of the first tube in the amplifier shown in Fig. 8.20. In fact, the voltage decays slowly with time as the charge leaks away through the biasing resistors, but the peak amplitude of the input voltage step is very close to

$$\hat{v}_{Si} = \frac{q_S}{C_S + C_i}. \tag{8.128}$$

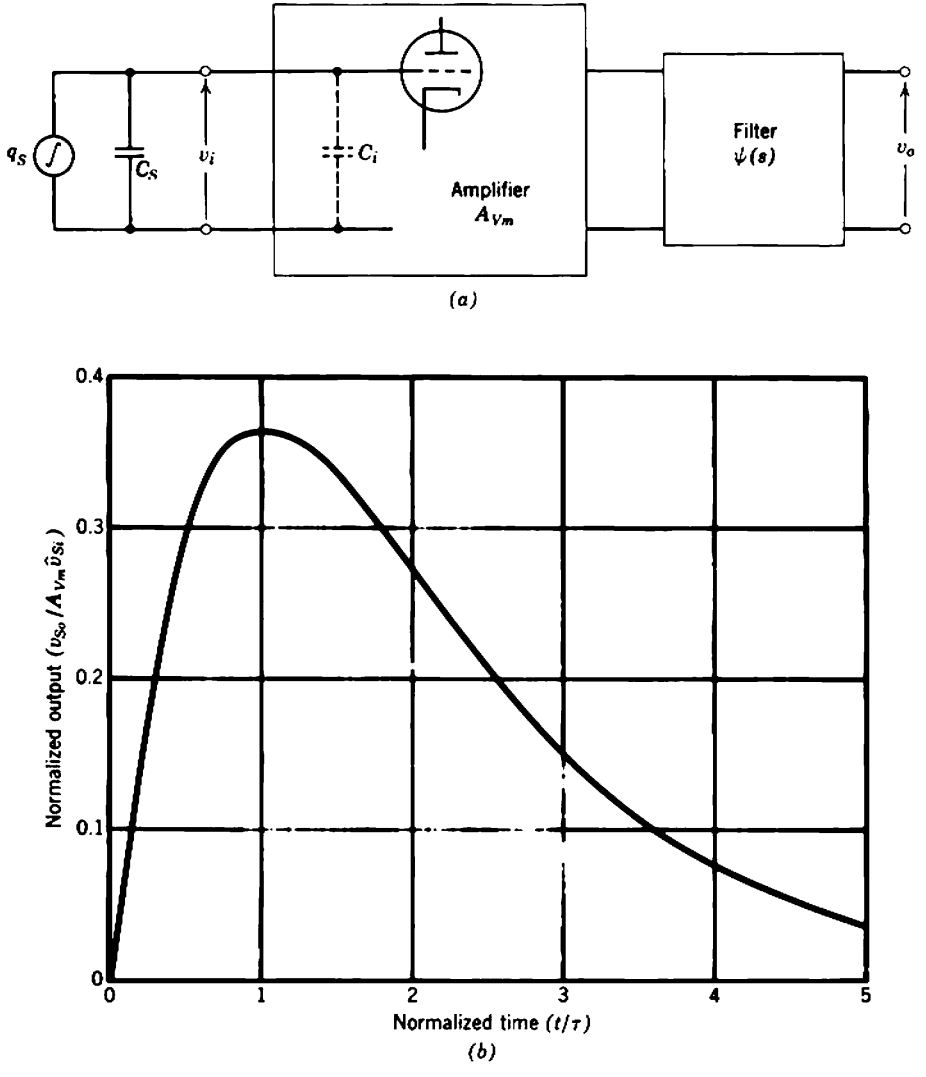


Fig. 8.20 Noise in a nuclear pulse amplifier: (*a*) outline circuit diagram; (*b*) form of the output pulse.

The resulting output voltage step from the amplifier is fed to some forming circuit which converts the step into a pulse suitable for counting and other purposes. One very commonly used pulse forming network is a combined *RC* integrator and differentiator, with the time constants equal at τ; the transfer function is

$$\psi(s) = \frac{s\tau}{(1 + s\tau)^2}. \tag{8.129}$$

Typical values of τ are of the order of a few microseconds. The output voltage from the system has the form shown in Fig. 8.20*b*:

$$v_{So}(t) = A_{Vm}\hat{v}_{Si}\left[\left(\frac{t}{\tau}\right)\exp\left(\frac{-t}{\tau}\right)\right]$$

and the peak amplitude, occurring at $t = \tau$, is

$$\hat{v}_{So} = \frac{A_{Vm}\hat{v}_{Si}}{e} \tag{8.130}$$

where $e = 2.71\ldots$.

8.7.1 Number of Ion Pairs

Noise in a particle detecting system is conveniently expressed as an equivalent number of ion pairs $(\text{n.i.p.})_N$ in the detector. The value of $(\text{n.i.p.})_N$ is defined as

$$(\text{n.i.p.})_N = \frac{\text{rms noise output}}{\text{peak signal output, when } (\text{n.i.p.})_S = 1}. \tag{8.131}$$

The peak signal output in response to one ion pair in the detector follows from Eqs. 8.128 and 8.130 as

$$\hat{v}_{So}|_{(\text{n.i.p.})_S = 1} = \frac{qA_{Vm}}{e(C_S + C_i)},$$

and the mean-square noise output is

$$v_{No}^2 = \left(\frac{A_{Vm}}{C_S + C_i}\right)^2 q_{Ni}^2{}_{eq}$$

where $q_{Ni}^2{}_{eq}$ is the equivalent mean-square noise charge referred to the input. Hence,

$$(\text{n.i.p.})_N = \frac{e}{q}\sqrt{q_{Ni}^2{}_{eq}}. \tag{8.132}$$

In general, the equivalent mean-square noise charge referred to the input of an amplifier is the denominator of Eq. 8.112*a*:

$$q_{Ni}{}^2{}_{eq} = \int_0^\infty \frac{|\psi(f)|^2}{(2\pi f)^2}\left[d(i_N')^2 + \frac{d(i_N'')^2}{|\beta|^2}\right], \tag{8.133}$$

where $\beta(f)$ is the current gain from the source to the output of the first device in the amplifier. For the special case of a particle detecting system whose transfer function $\psi(s)$ is given by Eq. 8.129,

$$\frac{|\psi(f)|^2}{(2\pi f)^2} = \frac{\tau^2}{[1 + (2\pi f\tau)^2]^2}. \tag{8.134}$$

The noise, expressed in ion pairs, can be found by substituting appropriate values of $d(i_N')^2$, $d(i_N'')^2$, and β into Eqs. 8.132 to 8.134.

8.7.1.1 Vacuum Tubes

The time constant τ in a typical nuclear particle detecting system is a few microseconds. Hence, the system passband is centred around a frequency of the order 100 kHz. At such frequencies, Section 8.2.1 shows that the dominant component of $d(i_N')^2$ for a triode is associated with the dc grid current, whereas $d(i_N'')^2$ is predominantly shot noise on the anode current. From Eqs. 8.9 and 8.10

$$d(i_N')^2 \approx 2q\,\Sigma(I_G)\,df,$$
$$d(i_N'')^2 \approx 2q\Gamma^2 I_A\,df.$$

Both noise generators are white. It is convenient to express $d(i_N'')^2$ in terms of an easily measurable quantity, the noise resistance:

$$d(i_N'')^2 = g_m{}^2 d(v_N{}^2) = g_m{}^2 4kTR_N\,df.$$

The system current gain is

$$|\beta(f)| = \frac{g_m}{2\pi f(C_S + C_i)}.$$

Substitution into Eqs. 8.132 to 8.134 gives the noise in ion pairs as

$$(\text{n.i.p.})_N = \frac{e}{q}\left(\int_0^\infty \left\{\frac{2q\,\Sigma(I_G)\,\tau^2}{[1 + (2\pi f\tau)^2]^2} + \frac{4kTR_N(C_S + C_i)^2(2\pi f\tau)^2}{[1 + (2\pi f\tau)^2]^2}\right\} df\right)^{1/2}.$$

Using integrals (4) and (2) of Table 8.2, and simplifying the result,

$$(\text{n.i.p.})_N = \frac{e}{q\sqrt{8}}\left[2q\,\Sigma(I_G)\,\tau + \frac{4kTR_N(C_S + C_i)^2}{\tau}\right]^{1/2}. \tag{8.135}$$

Notice that the grid current contribution to the noise is proportional to the shaping time constant τ, whereas the shot noise contribution is

inversely proportional. Minimum noise occurs at the optimum time constant

$$\tau_{opt} = \left[\frac{4kTR_N(C_S + C_i)^2}{2q\,\Sigma(I_G)}\right]^{1/2} \tag{8.136}$$

and the minimum noise is

$$(\text{n.i.p.})_{N(\min)} = \frac{e}{2q}\left\{[2q\,\Sigma(I_G)][4kTR_N(C_S + C_i)^2]\right\}^{1/4}. \tag{8.137}$$

8.8 HUM, MICROPHONY, AND RADIO-FREQUENCY INTERFERENCE

The discussion of noise up to this point is ideal, in that it considers only natural sources of noise. In addition to natural noise, the output of a practical amplifier contains noise components that are attributable to man-made phenomena and could in principle be reduced to zero. Unless care is taken with the electrical and mechanical design of equipment this man-made noise can cause a significant degradation in over-all noise performance. Unfortunately, man-made noise is not amenable to simple general analysis. The complication arises not from any inherent physical difficulty, but from the large numbers of variables involved; factors such as mechanical layout, environment, and interconnections between units are of vital importance. Nevertheless, each noise problem has a physical basis which points the way to its solution. In many instances satisfactory noise performance can be achieved by employing standard practice to minimize the effects of common sources of noise; this procedure may be complemented by recourse to intelligent experiment. The paragraphs which follow list the more important sources of man-made noise, and discuss some standard methods of noise reduction.

8.8.1 Hum

The term hum refers to undesired periodic signals introduced into the amplifier from the ac power mains. These signals are harmonically related to the mains frequency; in the vast majority of cases only the fundamental and/or second harmonic of the mains frequency are important. Sources of hum may be grouped into four classes.

8.8.1.1 Dc Supply Rails

The dc supply voltages in most items of electronic equipment are obtained by rectifying the ac supply mains. Any residual ac component

(*ripple*) on the dc supply voltages is introduced into all stages of an amplifier via the electrode supply resistors. Because the power supplies in most electronic equipment use full-wave single-phase rectifiers, the ripple voltages are at twice the mains frequency; exceptions occur with equipment using half-wave rectifiers or three-phase supplies. The hum voltages introduced into the early stages of an amplifier are amplified by the subsequent stages, and may build up to an appreciable level at the output. Obviously, higher ripple voltages can be tolerated on the supply rails for the later stages of an amplifier than for the early stages, and the most complete filtering is required for the supplies to the input stage. This type of hum can always be reduced to negligible proportions by proper design of the filter network in the rectifier and the decoupling networks in the amplifier. The amplifier design described in Section 7.8.1.1 employs a single decoupling network for the first stage.

8.8.1.2 Vacuum-Tube Heaters

It is standard practice in vacuum-tube equipment to use mains-frequency power to supply the heaters. The presence of an alternating voltage on the heaters can result in the introduction of hum into low-level stages. It might be expected that the hum would be due mainly to cyclic variations in cathode temperature; however, the large thermal capacity of the heater-cathode system ensures that the temperature variation over a cycle is negligible. The mechanisms which do account for hum are

(i) electrostatic coupling to the grid via the grid-heater capacitance,
(ii) conductive coupling to the cathode via the leakage between the heater and cathode,
(iii) magnetic coupling to the electron stream.

If a tube has a heater of simple geometry and if the heater-to-cathode insulation is of constant resistance, the simple *hum-bucking* scheme shown in Fig. 8.21*a* reduces heater hum appreciably. Ideally, equal but opposite hum voltages are induced into the signal path by the two sides of the heater. In practice, it is preferable to use a heater supply with some form of variable tap; adjustment of the tap can take up asymmetries in the heater structure, and give most nearly perfect cancellation of hum voltages. Figure 8.21*b* shows a common arrangement. Some further improvement often results from biasing the heater a few volts positive with respect to the cathode, as in Fig. 8.21*c*. This reduces thermionic emission of electrons from the heater to the cathode. In extreme cases, a rectifier and filter may be provided for powering the heater with direct current.

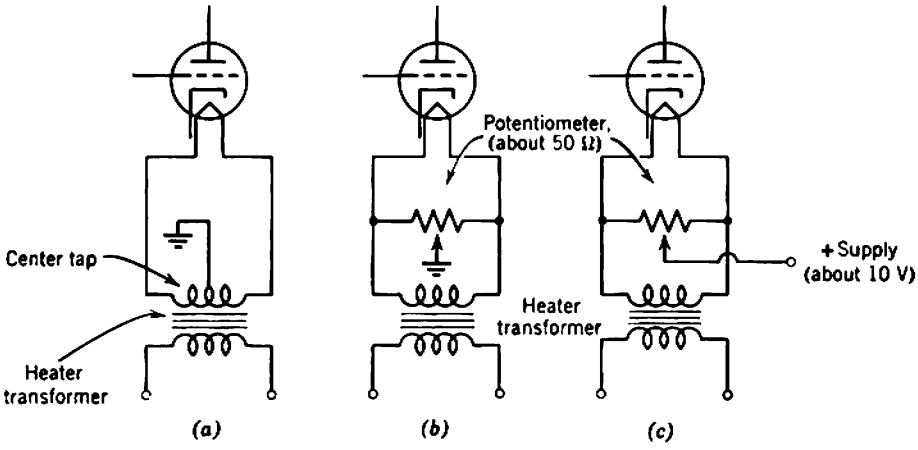


Fig. 8.21 Reduction of hum originating from ac voltage on a vacuum-tube heater.

8.8.1.3 Alternating Electric Fields

Alternating electric fields in an amplifier arise principally from mains-frequency voltages on ac power leads and heater wiring. Induction of hum voltages onto the signal leads by an electric field is easily visualized and, more important, it is easily reduced by the simple expedience of electrostatic shielding. However, even the need for shielding can be minimized if reasonable care is taken with an amplifier's physical layout. Leads carrying mains-frequency voltages should be kept well away from signal leads.

If the grid lead of a vacuum tube and an ac mains lead are run close to each other for a few inches, the stray capacitance C between them can easily amount to 1 pF. If the rms voltage on the mains lead is v_M and if the grid resistor of the tube is R_G, the voltage induced on the grid is

$$v_{G(\text{hum})} = 2\pi f_M R_G C v_M \tag{8.138}$$

where f_M is the mains frequency. Substituting the typical values $R_G = 1\ \text{M}\Omega$, $C = 1$ pF, $v_M = 110$ V, and $f_M = 60$ Hz gives

$$v_{G(\text{hum})} = 41.5\ \text{mV}.$$

If the tube concerned is the first stage of an audio amplifier, the resulting hum output is a veritable bellow. If the hum voltage at the grid is to be reduced to a value comparable to the equivalent noise voltage, the stray capacitance must be reduced by three or four orders of magnitude. A significant reduction can be achieved by rerouting the mains lead to a position remote from the grid lead. However, it is often necessary to

introduce additional electrostatic shielding between the grid and the mains leads. This can be accomplished by using shielded cable for the grid lead, by enclosing the tube envelope in a shielding can, and, if necessary, by placing the entire stage within a continuous metal shield. The heater leads should run directly away from the tube socket on the side remote from the grid lead. As an additional precaution, the heater leads may be shielded. The foregoing remarks (except those involving heater leads) are applicable to transistor amplifiers also. However, transistor circuits are different quantitatively because of the generally lower impedance levels. The base supply resistance R_B is smaller than R_G in Eq. 8.138 by two or three orders of magnitude and, consequently, the induced hum voltages are smaller also. Electrostatic hum pickup is less of a problem in transistor equipment than in vacuum-tube equipment.

8.8.1.4 Alternating Magnetic Fields

Alternating magnetic fields inside items of electronic equipment arise principally from the leakage fields of power transformers and from mains-frequency currents flowing in power leads and heater wiring. Alternating magnetic fluxes induce emfs in any conductor which they cut, and are a major problem in low-level amplifiers, particularly those using an input transformer. Magnetic shielding is not nearly so straightforward as electrostatic shielding for two reasons. First, massive shields of very high permeability alloys are required; it is standard practice to use single, double, or even triple shields on sensitive components such as input transformers. Second, the magnetic fields around components such as power transformers are not easy to visualize, and there is some uncertainty about the optimum placement of shields; experiments with small magnetometers can be of great assistance in locating areas of high field concentration. It is therefore most important to concentrate as far as possible on reducing magnetic fields at their source. Perhaps the most significant reductions can be achieved by

(i) twisting pairs of heater wires together tightly so that their equal and opposite magnetic fields cancel,
(ii) choosing power transformers which have a low leakage flux and orientating them so that high fluxes do not cut low-level sections of the amplifier,
(iii) avoiding *ground loops*, as described below.

A ground loop consists essentially of a closed path in the grounding system. The mechanism by which hum develops is most easily understood if the ground loop is formed when two items of equipment are connected together as in Fig. 8.22. The two units are connected together

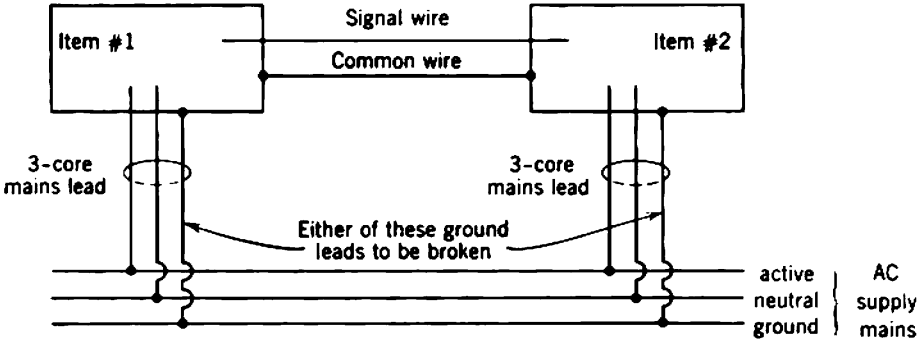


Fig. 8.22 Ground loop formed when two items of equipment are interconnected.

by signal and common wires, and are also connected to a common ground point via three-core power connectors. Any alternating magnetic field induces emfs in the various interconnecting wires. If the signal and common wires between the two units follow a similar path (particularly if they take the form of a concentric cable), the emfs induced in each are the same. There is therefore no differential voltage induced between them and no resultant hum is introduced into the system. However, the ground leads form a closed loop, and a current circulates in this ground loop due to the induced emfs. There are therefore IR voltage drops along these leads, whose sum is equal to the total induced emf. The drop along the common wire joining the two units appears as a differential voltage in the signal common path, and hum is introduced.

The remedy is to reduce the ground-loop current, which typically is several amperes. One approach is to open-circuit the ground lead in the power cable to one of the units, thereby interrupting the circulating current. This has the disadvantage that, when the two units are disconnected from each other, the unit with the open ground lead is a potential safety hazard. The ground lead should therefore be broken by means of one of the special adapter plugs available for just this purpose, which carries appropriate warnings and whose presence alerts the operator to the danger.

In gramophones and other audio systems, a preamplifier is often connected to a combination main amplifier and power supply by two cables: a shielded signal lead, and a multicore lead carrying heater and dc power to the preamplifier. The shield on the signal cable and the ground in the power cable constitute a ground loop, and hum can be introduced. Breaking the ground lead in the power cable is not a satisfactory remedy, because, if the signal cable should be disconnected, the chassis of the preamplifier goes off nearly to the supply voltage; anyone reconnecting

the signal cable will receive a shock. A satisfactory approach is to connect a resistor of about 50 Ω in the ground lead of the power cable. This is large enough to reduce the circulating ground current by 2 or 3 orders of magnitude, but small enough to keep the preamplifier chassis within a few volts of ground when the signal cable is disconnected.

Ground loops also occur inside items of equipment that are constructed on a metal chassis and use this chassis for ground connections. Eddy currents are induced in the chassis, and potentials of a few millivolts appear between different points on it. A remedy suitable for audio amplifiers is the *ground-bus technique*. Instead of using the chassis as a

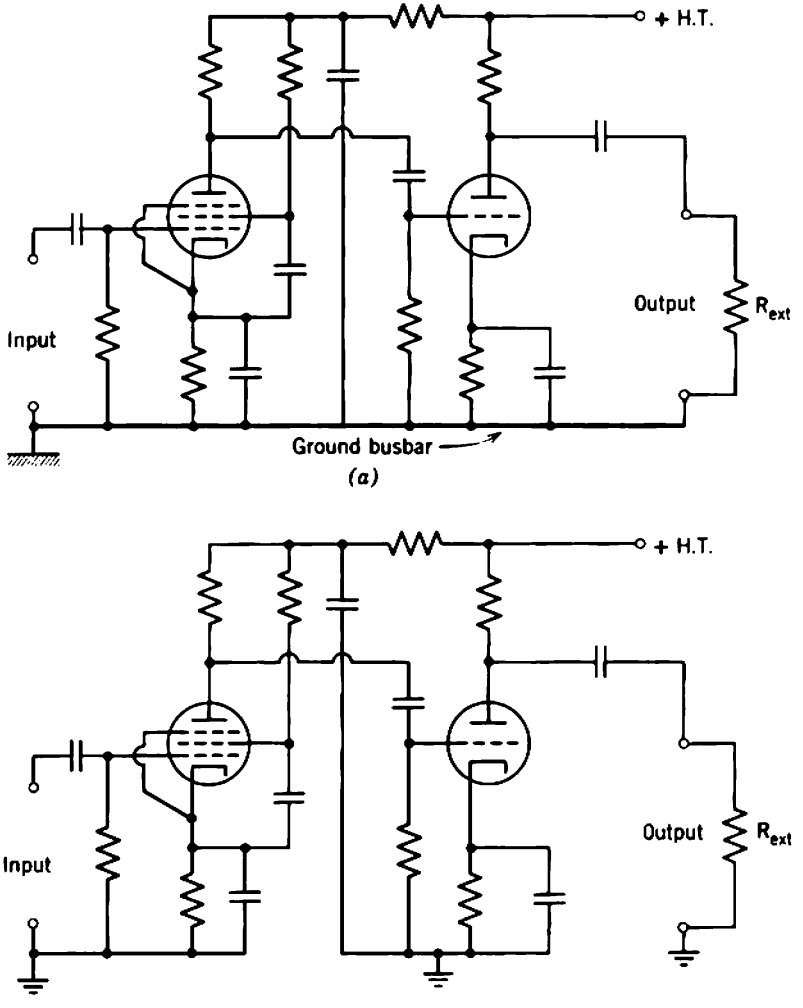


Fig. 8.23 Techniques for reducing the effects of ground loops in a metal chassis: (*a*) ground busbar; (*b*) grouped ground points.

ground, all ground points in the circuit are interconnected with a wire or *busbar*, and this is connected to the chassis at one point only, usually the input socket. In this way, the grounding system is reduced to a tree structure which has no closed paths, so no eddy currents can flow. This technique is unsuitable for wide-band amplifiers as coupling between stages via the inductive impedance of the ground bus may cause regeneration. Satisfactory hum performance can usually be obtained by grouping the ground points for each stage at one spot on the chassis in such a way that no hum ground voltage is introduced into signal loops. Figure 8.23 illustrates these techniques for the vacuum-tube amplifier described in Section 7.8.1.

8.8.2 Microphony

Microphony is a form of disturbance initiated by mechanical shock or vibration. The most common mechanism is that associated with a change of capacitance between electrodes due to shock-induced relative motion. Vacuum tubes are far more susceptible to microphony than transistors; microphony in tubes is caused mainly by vibration of the relatively flimsy grid wires. Microphonic noise can be reduced by mechanically isolating suspect components with antishock mounts. Further improvements can be achieved by using special ruggedized components.

8.8.3 Radio-Frequency Interference

Radio-frequency interference is an important fundamental problem for communications engineers, because with modern low-noise amplifiers the dominant components of noise may originate externally to the amplifier. Such sources of noise cover an extremely wide frequency spectrum, and include galactic noise from all the natural radio sources in the universe as well as an increasing quantity of man-made interference.

In the detailed design of conventional low-pass amplifiers, some types of man-made radio-frequency interference can be quite important. Electric motors, switches, welding machines, and leakage paths across very-high-voltage insulators are examples of *broad-band noise sources*. Such noise sources can induce interference into signal paths by electrostatic or magnetic coupling, that is, in much the same ways as hum is induced. The same noise reduction techniques are therefore satisfactory. An amplifier employing good construction practice in the low-level sections (screened input leads, screened transformers, twisted and screened heater wires, etc.) is usually not susceptible either to hum induction or radio-frequency interference.

A slightly different problem is that of interference from *narrow-band radio-frequency sources* such as nearby radio transmitters. This form of interference sometimes persists even when the effects of broad-band interference have been minimized, but it can usually be removed by keeping all signal leads very short and inserting appropriately tuned filter traps in any paths that join signal leads to supply rails. Thus, even if the relatively long supply leads act as efficient antennae, the rf voltages induced in them are blocked from the signal paths.

In conclusion, it should be noted that transistors and vacuum tubes are, roughly speaking, equally susceptible to radio-frequency interference. It is easy to become careless in the mechanical design of transistor circuits, purely because it is so easy to avoid hum pickup. This attitude has caused designers many problems when the equipment is used in an environment having a high level of rf interference.

8.9 CONCLUDING COMMENTS ON THE DESIGN OF LOW-NOISE AMPLIFIERS

At this point it is appropriate to summarize in general terms the contents of the preceding sections. Probably the most important points to emerge from the discussion are:

1. The designing of a low-noise amplifier falls into two parts—the purely electrical problem of minimizing circuit noise, and the predominantly mechanical problem of reducing the effects of man-made noise. Either type of noise can dominate. The practical conclusion is that a careful circuit design or the use of expensive, low-noise components may be completely negated unless the amplifier is well constructed.

2. If the mechanical design is such that the effects of man-made noise are reduced to negligible proportions, it is worth minimizing the circuit noise by careful design. There is no general procedure for maximizing the signal-to-noise ratio; each problem must be treated on its merits, but the trends indicated in Sections 8.4 and 8.5 can assist the designer in making an intelligent guess at methods for improving an unsatisfactory noise performance. Usually signal-to-noise ratio can be increased by exercising control over one or more of the following factors:

(a) the input device type,
(b) the input device quiescent point,
(c) the magnitude of the source impedance,
(d) the frequency response of the amplifier.

Invariably, at least one of these factors is fixed by other requirements.

Chapter 9

Simple Practical Amplifiers III: Nonlinearity and Distortion

All practical amplifying devices have nonlinear electrical characteristics in that their incremental (small-signal) parameters depend on the operating currents and voltages. Hence the output signal waveform from an amplifier does not follow the input waveform exactly; the small-signal parameters and instantaneous gain change from point to point on the input waveform. Both the input and output signals can be represented by Fourier series (or Fourier integrals if the waveforms are not repetitive) and any difference between the two spectra can be ascribed to the production of Fourier *distortion components* by the nonlinearity. Distortion components may be considered as a type of noise, and their magnitude relative to the signal is sometimes expressed as a signal-to-noise ratio. In general, the distortion components depend on the nature and level of the input signal and on the frequency response of the amplifier. The special case in which all significant distortion components lie in the mid-band frequency range is of considerable importance.

Once the sources of nonlinearity in amplifiers are understood, it is possible to minimize their effects by suitable design techniques. Significant improvements can be obtained by proper choice of device type, proper choice of operating conditions, and (as introduced in the next chapter) by employing negative feedback. Consequently, the linear device models used in preceding chapters may be used even when the signals are of quite large amplitudes. The commonly used terms *small-signal amplifier* and *large-signal amplifier* suggest a distinction which should not really exist. "Small-signal" conditions persist as long as the transfer characteristic of an amplifier may be assumed linear and the distortion

components are of "negligible" amplitude. The definition of just what is meant by "negligible" depends on the system requirements, where the system often must include a human observer. For example, in audio amplifiers the corrupting influence of distortion components can be assessed by carrying out subjective listening tests with varying amounts of nonlinearity, varying types of program material, and a variety of observers. The information from such subjective tests indicates the order of nonlinearity that is tolerable in particular applications. Under other circumstances, notably in amplifiers for use in measuring systems, the tolerable nonlinearity is specified directly by the constancy required in the amplifier's transfer function. Indeed, precise information about the nonlinearity is a fundamental type of specification, whereas information on the magnitudes of distortion components under some test condition is a particular type of specification. The relative advantages of the two are similar to those of the noise specifications discussed in Section 8.6. A fundamental specification is complete and can be used to determine the distortion under any signal conditions. Particular specifications allow simple, direct comparison of competing amplifiers under similar conditions of operation, but cannot ordinarily be used to determine the performance under any other conditions of operation.

9.1 BASIC IDEAS ON NONLINEARITY

In principle, the response of an amplifier to finite signals can be investigated by means of the nonlinear charge-control equations. As explained in Section 2.2.2, the stored charge Q_C can be eliminated to yield relations between the electrode currents and voltages of intrinsic devices. Equations to predict the large-signal performance of practical amplifiers should, of course, be modified to include the effects of extrinsic elements and passive circuit components. For some simple problems these nonlinear differential equations are capable of exact or approximate solution; this formal approach has been used extensively for predicting the response of simple pulse circuits.* However, for all but the simplest amplifiers, the approach becomes so tedious that an exact solution (or even a reasonable approximation thereto) is virtually unobtainable. Furthermore, the wisdom of attempting an exact solution is questionable. Precise knowledge of the character of the nonlinearity in a practical amplifier is seldom required; some reasonable estimate of the upper limit is sufficient. One convenient

* For example, R. BEAUFOY, "Transistor switching circuit design using the charge-control parameters," *Proc. Inst. Elec. Engrs., London*, **106B**, 1085, Suppl. 17, 1959.

measure of nonlinearity is the divergence of an amplifier from ideal linearity, known as the *differential error*; hence, a practical study can logically start from the known response to signals of infinitesimal amplitude.

Consider an amplifier containing one or more active devices. Assume arbitrarily that the amplifier is an attempt to realize an ideal transfer admittance, a linear relation between input voltage $V_i(t)$ and output current $I_o(t)$. Exactly parallel developments can be made for any other transfer function of the amplifier or of the complete system involving the source, amplifier, and load. In the notation of Chapters 5 and 7, small (infinitesimal) increments in V_i and I_o are represented by lowercase letters:

$$v_i = \Delta V_i,$$

$$i_o = \Delta I_o.$$

The relation between v_i and i_o can be expressed operationally as a transfer admittance $Y_T(s)$:

$$\frac{i_o(s)}{v_i(s)} = Y_T(s) = G_T\,\psi(s),$$

where G_T is the mid-band transfer conductance, and $\psi(s)$ is a function of complex frequency that gives the poles and zeros. In general, both G_T and $\psi(s)$ depend on the voltages and currents present at any instant on the active devices within the amplifier, because the small-signal parameters of a device depend on its electrode voltages and currents. Therefore, G_T and $\psi(s)$ vary from point to point on the waveforms of $V_i(t)$ and $I_o(t)$.

The variation of G_T is independent of complex frequency; it is a function of the amplifier alone, and is therefore known as the *static nonlinearity* in the transfer admittance. Static nonlinearity is easy to visualize and can be handled in many ways—among others graphically as in Section 6.2.1.1. For example, one *transfer characteristic* of a triode is a plot of anode voltage versus grid voltage for a particular load resistor (Figs. 6.14 and 6.15). The *transfer function* at any point—voltage gain in the case of Fig. 6.15—is the gradient of the transfer characteristic, and the output waveform in response to any input can be calculated as in Fig. 6.16. The variation of $\psi(s)$ is known as the *dynamic nonlinearity* in the transfer function, because $\psi(s)$ is a function of the complex frequency of the signal as well as of the instantaneous values of electrode voltages and currents. Dynamic nonlinearity is a function of the signal as much as it is of the amplifier, and is perhaps most easily visualized as a series of pole-zero plots for the particular instantaneous combinations of $V_i(t)$ and $I_o(t)$ that occur.

Because fewer variables are involved, static nonlinearity is much easier to treat analytically and to measure than dynamic nonlinearity. It would therefore be convenient if static nonlinearity alone were the criterion of an amplifier's acceptability. Fortunately, this is often the case. In the majority of low-pass amplifiers the useful signal information is largely contained in the mid-band range of frequencies, and static nonlinearity produces distortion components from the Fourier components lying inside this range. Additional distortion components arise from the effects of dynamic and static nonlinearities on the fringe Fourier components near the lower and upper cutoff frequencies, but since the amount of signal information involved is small the contribution to total distortion is small also. An amplifier with acceptable static nonlinearity therefore usually has an acceptable distortion performance under dynamic conditions, but there are undoubtedly exceptions to this statement.* This chapter considers only static nonlinearity.

9.1.1 Specification of Static Nonlinearity

Static nonlinearity in an amplifier can be specified in several ways. Among the most useful are:

1. A plot of the transfer characteristic enables the nonlinearity to be visualized at a glance. Moreover, the output waveform in response to any input can be calculated as in Section 6.2.1.1.

2. Equivalent to 1 above but more convenient analytically is a power series representation of the transfer characteristic. This formal representation of nonlinearity is used in much of the following algebraic development.

3. As an alternative to the transfer characteristic itself, the gradient is sometimes specified at a number of points. The most usual specification is the *differential error* γ (also known as *differential nonlinearity*) at any point, this being the fractional change in gradient from the value at the quiescent point. Since the small-signal transfer function at any point is the gradient of the transfer characteristic, γ can be written as

$$\gamma = \frac{TF - TF_Q}{TF_Q}, \tag{9.1}$$

where TF is the transfer function at the point for which γ is defined, and TF_Q is the value of TF at the quiescent point. Numerical values of γ are usually expressed as percentages.

* As an example, see J. B. POTTER, "On the measurement and interpretation of nonlinearity in a television system," *J. Brit. Inst. Radio Engrs.*, **26**, 141, August 1963.

Being defined in terms of the small-signal transfer function, differential error provides the formal link between the linear (small-signal) and nonlinear (large-signal) performance of an amplifier. In addition, differential error is by far the most useful measure of the nonlinearity in practical amplifiers. First, differential error is easy to calculate. The transfer function at any point on the signal waveform depends on the small-signal parameters of the active devices, and these vary in known ways with electrode voltages and currents. Second, the maximum differential error is in part the criterion of an amplifier's acceptability for many applications. As an example, oscilloscopes are often used to compare small-amplitude signals which are superimposed on a much larger signal; limits should be specified to the relative error between the incremental transfer function and its ideal constant value. Finally, there are simple relations between the differential error at the signal peaks and the two most common particular specifications of nonlinearity—low-order *harmonic distortion* and *intermodulation distortion.*

9.1.2 Clipping Points

Static nonlinearity in an amplifier can be divided into two types:

(i) The sharp curvature of the transfer characteristic near the points of maximum and minimum current. These are known as the *clipping points.*
(ii) The relatively gradual curvature well away from the clipping points.

The central portion of the transfer characteristic is the quasi-linear region within which supposedly linear circuits such as amplifiers must operate. Over this region the transfer characteristic can be approximated by a low-order power series. It is this low-order approximation that results in the simple relations between differential error at the signal peaks and low-order harmonic and intermodulation distortions. Naturally, the simple relations fail if an amplifier is driven into the regions of gross nonlinearity, but it is a truism to point out that linear amplifiers should not operate in these regions.

The clipping points are limits to be avoided in normal operation; only the region of moderate nonlinearity between these limits warrants detailed investigation. Determination of the clipping points is simple and almost self-evident for most amplifier stages, and ought not to involve more than a rudimentary graphical analysis (often piecewise-linear) with a load line (Section 6.2). The low-current clipping point of an amplifier usually corresponds to current cutoff in one or more amplifying devices, and is therefore defined fairly precisely. The high-current or saturation point

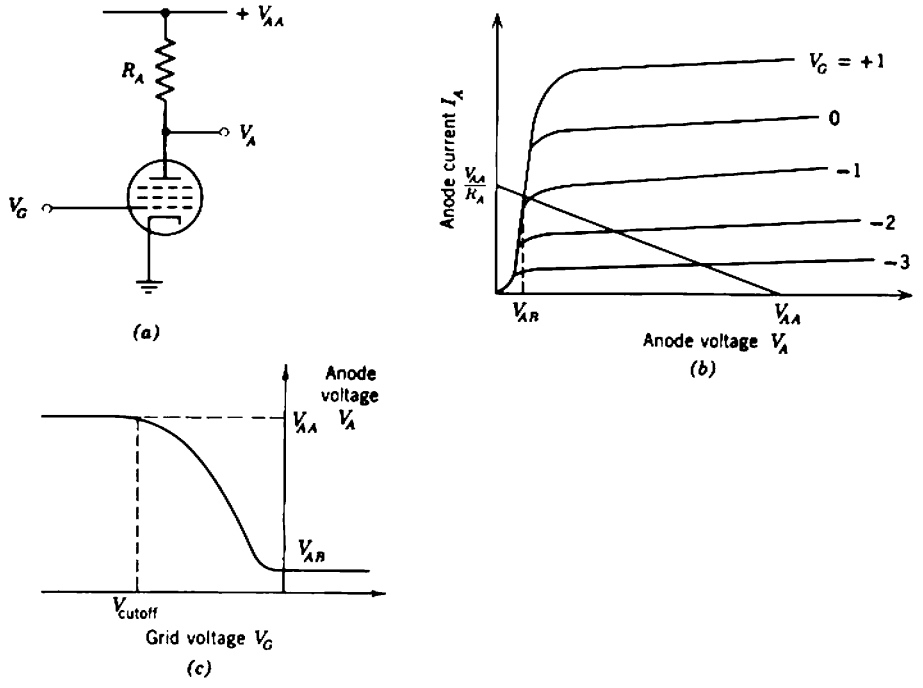


Fig. 9.1 Sharp saturation point for a pentode: (*a*) circuit diagram; (*b*) anode characteristics and load line; (*c*) V_A/V_G transfer characteristic.

can originate in any of several ways. For a triode, saturation may correspond to a very positive grid, for which the anode-to-cathode voltage falls to a low value and the tube is incapable of passing a larger anode current. More commonly, saturation corresponds to the point at which the grid becomes just positive and begins to draw heavy current, so that the preceding tube is incapable of driving the grid more positive. The same limits on the anode current may apply to a pentode. Another limiting mechanism is that the load line can intersect the anode characteristics below their knee (Fig. 9.1) so that a sharp current maximum occurs even though the grid is negative. For bipolar transistors, saturation usually corresponds to the collector current increasing to the point at which the collector-to-emitter voltage falls almost to zero, when the transistor is incapable of passing more current. For f.e.t.s, saturation corresponds either to the load line intersecting the drain characteristics in the knee region below pinch-off (analogous to pentode behavior), or in the case of junction types to forward bias of the gate (analogous to the onset of grid current in tubes). The quiescent point of a small-signal amplifier is normally chosen about mid-way between the clipping points so that signal

excursions of approximately the same magnitude are available in each direction.

9.2 HARMONIC DISTORTION

The nonlinearity in an amplifier is often specified by means of the distortion components produced for a standardized input signal. Two such specifications are in general use: *harmonic distortion* is specified when the standard input signal is a single sinusoid, and *intermodulation distortion* is specified when the input consists of two or more sinusoids of different frequencies.

In general, if the input signal to a nonlinear amplifier is a sinusoid of frequency ω, the output contains components at frequencies ω (the fundamental), 2ω (the second harmonic), 3ω (the third harmonic), and so on. Designating the amplitudes of these components as A_1, A_2, $A_3, \ldots$, the harmonic distortions for the various components are defined as

$$D_2 = \frac{A_2}{A_1}, \tag{9.2a}$$

$$D_3 = \frac{A_3}{A_1}, \tag{9.2b}$$

$$D_4 = \frac{A_4}{A_1}. \tag{9.2c}$$

Alternatively, the total harmonic distortion is defined as

$$D_T = \sqrt{D_2^2 + D_3^2 + D_4^2 + \cdots} = \left(\frac{A_2^2 + A_3^2 + A_4^2 + \cdots}{A_1^2}\right)^{1/2}. \tag{9.3}$$

Numerical values for D_2, $D_3, \ldots, D_T$ are usually expressed as percentages. Notice that the sum of the individual distortion components is found by adding distortion powers rather than amplitudes. This is appropriate because of the differing frequencies of the components, and is similar to noise summations. In many instances the power in harmonics beyond the third is so small as to be negligible.

9.2.1 Harmonic Distortion and Nonlinearity

Consider an amplifier whose significant transfer function is its transfer conductance, the relation between input voltage V_i and output current I_o. Exactly parallel developments can be made for other transfer functions. Suppose the transfer characteristic can be written in the functional form

$$I_o = I_o(V_i)$$

and that the quiescent values are V_{iQ} and I_{oQ}. Taking a Taylor expansion about the quiescent point yields

$$I_o = I_{oQ} + \left(\frac{dI_o}{dV_i}\right)_{V_i = V_{iQ}} (V_i - V_{iQ}) + \left(\frac{d^2I_o}{dV_i^2}\right)_{V_i = V_{iQ}} \frac{(V_i - V_{iQ})^2}{2!} + \cdots$$

which can be written as

$$I_o = I_{oQ} + a_1(V_i - V_{iQ}) + a_2(V_i - V_{iQ})^2 + \cdots. \tag{9.4}$$

Usually the transfer characteristic well away from the clipping points can be represented adequately by the first three or four terms in the series. In numerical problems, the constants $a_1, a_2, a_3, \ldots$ can be determined by any convenient curve-fitting technique; often a simple trial-and-error procedure is satisfactory.

Suppose now that a sinusoidal signal input voltage is applied to the amplifier:

$$V_i - V_{iQ} = \hat{V}_i \cos \omega t. \tag{9.5}$$

Substitution into Eq. 9.4 yields the following harmonic series for the signal component of output current:

$$\begin{aligned} I_o - I_{oQ} = {} & (\tfrac{1}{2}a_2\hat{V}_i^2 + \tfrac{3}{4}a_4\hat{V}_i^4 + \cdots) \\ & + (a_1\hat{V}_i + \tfrac{3}{4}a_3\hat{V}_i^3 + \cdots)\cos \omega t \\ & + (\tfrac{1}{2}a_2\hat{V}_i^2 + \tfrac{1}{2}a_4\hat{V}_i^4 + \cdots)\cos 2\omega t \\ & + (\tfrac{1}{4}a_3\hat{V}_i^3 + \tfrac{5}{16}a_5\hat{V}_i^5 + \cdots)\cos 3\omega t + \cdots. \end{aligned} \tag{9.6}$$

The series can be rewritten as

$$I_o - I_{oQ} = A_0 + A_1 \cos \omega t + A_2 \cos 2\omega t + \cdots, \tag{9.7}$$

where $A_0, A_1, A_2, \ldots$ are the amplitudes of the harmonic components of output current; their magnitudes depend on the input voltage V_i and the nonlinearity. Notice the term A_0, which represents a dc shift in the output current under signal conditions; effectively, the quiescent point shifts dynamically.

From the defining equations for harmonic distortion (Eqs. 9.2),

$$D_2 = \frac{A_2}{A_1} = \frac{\tfrac{1}{2}a_2\hat{V}_i^2 + \tfrac{1}{2}a_4\hat{V}_i^4 + \cdots}{a_1\hat{V}_i + \tfrac{3}{4}a_3\hat{V}_i^3 + \cdots} \approx \hat{V}_i\left(\frac{a_2}{2a_1}\right), \tag{9.8a}$$

$$D_3 = \frac{A_3}{A_1} = \frac{\tfrac{1}{4}a_3\hat{V}_i^3 + \tfrac{5}{16}a_5\hat{V}_i^5 + \cdots}{a_1\hat{V}_i + \tfrac{3}{4}a_3\hat{V}_i^3 + \cdots} \approx \hat{V}_i^2\left(\frac{a_3}{4a_1}\right), \tag{9.8b}$$

$$D_4 = \frac{A_4}{A_1} = \frac{\tfrac{1}{8}a_4\hat{V}_i^4 + \cdots}{a_1\hat{V}_i + \tfrac{3}{4}a_3\hat{V}_i^3 + \cdots} \approx \hat{V}_i^3\left(\frac{a_4}{8a_1}\right). \tag{9.8c}$$

The approximations in the last forms of these equations are based on the usually justifiable assumption that contributions from higher-order terms can be neglected. Notice that the general relation between a distortion component and the input signal amplitude is approximately

$$D_r \propto \hat{V}_i^{r-1}. \tag{9.9}$$

9.3 INTERMODULATION DISTORTION

Intermodulation distortion is specified when the input signal to an amplifier consists of two or more sinusoids at different frequencies. In general, if the input to a nonlinear network contains frequencies ω_a, ω_b, $\omega_c, \ldots$, the output contains an infinite number of frequencies of the form

$$\omega = x\omega_a + y\omega_b + z\omega_c + \cdots, \tag{9.10}$$

where the constants $x, y, z, \ldots$ take all integral values from $-\infty$ through zero to $+\infty$. Obviously, there could be a great many useful ways of specifying intermodulation, the generation of distortion components not present in the input. The simplest specifications are those in which the input contains only two frequencies ω_a and ω_b, but even for this elementary input signal two different specifications are in common use.*

In the first specification, ω_a is a large signal low in the mid-band frequency range, while ω_b is a small signal high in the mid-band range; 4:1 is a usual ratio for their magnitudes. The output signal can contain significant power at frequencies of the forms

$$\omega_a, 2\omega_a, 3\omega_a, \cdots$$

$$\omega_b, (\omega_b \pm \omega_a), (\omega_b \pm 2\omega_a), \cdots.$$

The harmonic frequencies ω_a, $2\omega_a$, $3\omega_a, \ldots$ in the output are of no immediate interest; they lie low in the mid-band frequency range and can be removed by a high-pass filter. The sum and difference frequencies ω_b, $(\omega_b \pm \omega_a)$, $(\omega_b \pm 2\omega_a), \ldots$ are involved in the specification of intermodulation. In general, the sum of three sinusoids

$$\cos \omega_b t + k \cos (\omega_b + \omega_a)t + k \cos (\omega_b - \omega_a)t$$

can be expressed as

$$(1 + 2k \cos \omega_a t) \cos \omega_b t.$$

Thus the three sinusoids together constitute an amplitude-modulated

* A very readable comparison of methods is: B. M. OLIVER, "Distortion and Intermodulation," *Hewlett-Packard Application Note* #15, December 1963.

wave. In the language of radio telephony, ω_b is the carrier frequency, ω_a is the modulating frequency, and the index of modulation is

$$m = 2k.$$

The index to which ω_b in the output of an amplifier is modulated by each of the frequencies ω_a, $2\omega_a$, $3\omega_a$, ... can be calculated from the magnitudes of the sum and difference frequency terms, and the intermodulation distortion is defined as

$$D_{I1} = \sqrt{m_1^2 + m_2^2 + m_3^2 + \cdots}. \tag{9.11}$$

In the second two-frequency specification of intermodulation distortion, ω_a and ω_b lie close together in the center of the mid-band range, and the two signals are of equal amplitude. The distortion component used in

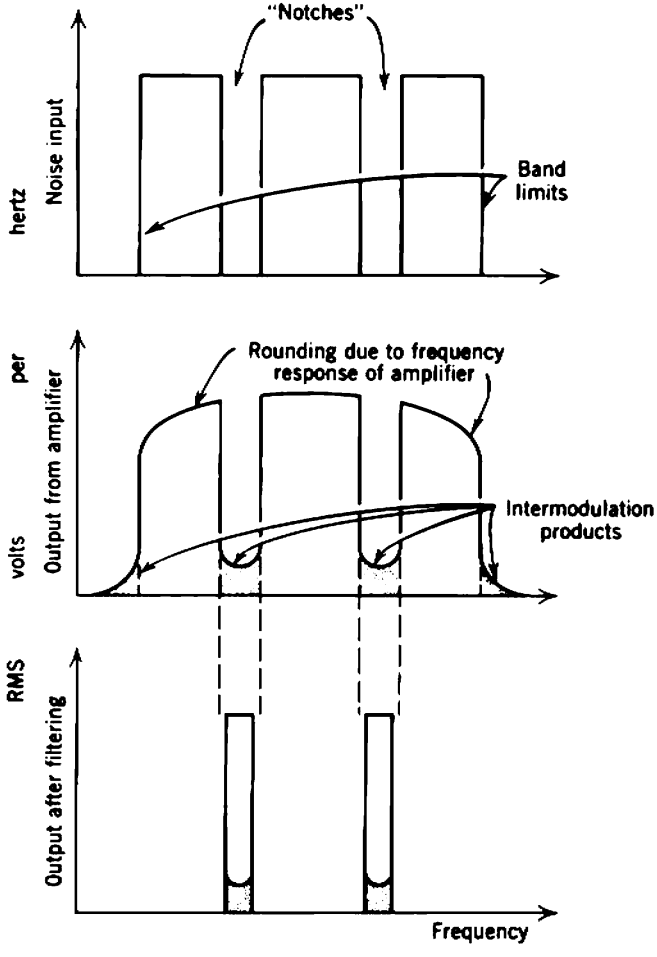


Fig. 9.2 Noise-loading test of intermodulation.

defining the intermodulation is the difference or beat frequency ($\omega_a - \omega_b$). The definition is

$$D_{I2} = \frac{A_{a-b}}{2A_a}, \tag{9.12}$$

where A_{a-b} is the amplitude of the beat frequency output,
$A_a = A_b$ are the amplitudes of the output components at ω_a and ω_b.

Single-frequency-input specifications of harmonic distortion, and two-frequency-input specifications of intermodulation distortion are rather unrealistic in that the test input signal is quite unlike a normal operating signal. Unless an amplifier has been grossly overdesigned for its particular application, the Fourier spectrum of its input signal occupies most of the mid-band range. A *noise-loading test* recognizes this fact, and gives a more realistic though less convenient measure of distortion. Basically, the input signal to the amplifier is white noise, which contains all frequencies simultaneously. However, the noise is band-limited by filters to lie within the mid-band range, and in addition has one or more "notches" cut into its spectrum by band-stop filters (Fig. 9.2). If the amplifier were linear, the spectrum of the output signal would contain only frequencies present in the input, but, with nonlinearity present, intermodulation products are formed and some of these lie in the vacant notches of the input spectrum. An rms meter is connected to the amplifier output via a band-pass filter which passes only frequencies within the notches, and the meter reading gives a measure of intermodulation. For standardized filters, meter readings can be defined in terms of percent intermodulation distortion, but the definition is arbitrary because it is intimately related to the filter characteristics.

9.3.1 Intermodulation Distortion and Harmonic Distortion

Subjective testing of audio systems suggests that total harmonic distortion and either one of the two-frequency measures of intermodulation distortion are roughly equivalent criteria of acceptability, provided the signal waveform does not approach the clipping points of the transfer characteristic. The specifications are related merely by numerical constants; approximately,

$$D_T = \tfrac{1}{3}D_{I1} = 2D_{I2}. \tag{9.13}$$

Under optimum listening conditions, distortion products in the center of the audio range can just be detected when

$$D_T \approx 3\%,$$
$$D_{I1} \approx 10\%,$$
$$D_{I2} \approx 1\tfrac{1}{2}\%.$$

These values may be taken as working limits for high-quality sound systems. They may be relaxed almost by an order of magnitude in less stringent applications (such as portable radio receivers) and at the ends of the audio-frequency range.

There is no simple relation between the nonlinearity of a transfer characteristic and the intermodulation distortion (defined in any of the above ways) that it produces. However, in view of the subjective equivalence of total harmonic distortion and intermodulation distortion in audio systems, it would be expected that some relation should exist. Brockbank and Wass* have performed a detailed analysis of the interactions between a number of sinusoids in a nonlinearity, and derived an expression for the total distortion power $P_{D(\text{tot})}$ in the intermodulation products as a function of the harmonic distortion of a single sinusoid:

$$P_{D(\text{tot})} = 4d_2\left(\frac{P}{P_0}\right)^2 + 24d_3\left(\frac{P}{P_0}\right)^3 + \cdots + 2^{r-1}r!\,d_r\left(\frac{P}{P_0}\right)^r, \qquad (9.14)$$

where P = the total power output (signal plus distortion)
d_r = the power output in the rth harmonic when the input is a sinusoid that produces arbitary power output P_0 at the fundamental (input) frequency.

Equation 9.14 represents an upper bound to the distortion power when the bandwidth and number of simultaneous input frequencies are both infinite. Brockbank and Wass indicate the order of correction required for typical practical systems; the reader is referred to their paper for details and for the derivation of Eq. 9.14. It is interesting (and perhaps significant) that the multiple-frequency input assumed by Brockbank and Wass approximates the noise input signal used in the noise-loading test of intermodulation. Their formula therefore gives a realistic measure of performance for an amplifier reproducing speech, music, and most other commonly encountered signals.

As an example in the use of Eq. 9.14, consider an audio amplifier whose distortion at full rated power output $P_{\max}$ is 5% third harmonic. This means that the *amplitude* of the third-harmonic component of output voltage or current is 5% of the fundamental (Eq. 9.2*b*). The third-harmonic *power* output is

$$P_{3(\max)} = P_{\max}\left(\frac{5}{100}\right)^2.$$

* R. A. Brockbank and C. A. A. Wass, "Nonlinear distortion in transmission systems," *J. Inst. Elec. Engrs., London*, **92**, part III, 45, March 1945. *Also* C. A. A. Wass, "A table of intermodulation products," *ibid.*, **95**, part III, 31, January 1948.

Equation 9.8*b* shows that the percentage third-harmonic distortion amplitude is proportional to the square of fundamental amplitude, and therefore proportional to fundamental power. Hence, the percentage third-harmonic distortion power is proportional to the square of fundamental power, and the actual third-harmonic distortion power is proportional to the cube of fundamental power:

$$P_3 \propto P_1^3.$$

Now, the total power output from any useful amplifier is dominated at all signal levels by the power in the fundamental. At maximum power output, therefore, the total power output is approximately equal to the fundamental power output. The third-harmonic distortion power at P_0 fundamental power follows as

$$d_3 \approx \left[P_{\max}\left(\frac{5}{100}\right)^2\right]\left(\frac{P_0}{P_{\max}}\right)^3.$$

Substitution into Eq. 9.14 gives the total distortion power output, at full-rated power output with a multiple-frequency input, as

$$P_{D(\text{tot})} = 24d_3\left(\frac{P_{\max}}{P_0}\right)^3 = 0.06P_{\max}.$$

The distortion power may be considered as a type of noise, and the output signal-to-noise ratio is

$$\frac{S_o}{N_o} = 10\log_{10}\left(\frac{P_{\max} - P_{D(\text{tot})}}{P_{D(\text{tot})}}\right) = 12.0\text{ dB}.$$

Brockbank and Wass perform the same calculation, but they include correcting factors for a finite bandwidth and obtain the less pessimistic signal-to-noise ratio of 14.0 dB.

9.4 RELATIONS BETWEEN DIFFERENTIAL ERROR AND HARMONIC DISTORTION

As stated in Section 9.1.1, simple relations exist between the low-order harmonic distortion in an amplifier and the differential error in its transfer characteristic at the signal peaks. Differential error can thence be related to intermodulation distortion via the equations of Brockbank and Wass, so that it provides a useful measure of amplifier performance on all the usual bases for comparison. As in previous sections, the following analysis is in terms of an amplifier's transfer conductance; a parallel development can be made for any other transfer function.

Consider an amplifier whose I_o/V_i transfer characteristic is represented by Eq. 9.4:

$$I_o = I_{oQ} + a_1(V_i - V_{iQ}) + a_2(V_i - V_{iQ})^2 + \cdots.$$

By differentiation, the transfer conductance at any input voltage is

$$G_T(V_i) = \frac{dI_o}{dV_i} = a_1 + 2a_2(V_i - V_{iQ}) + 3a_3(V_i - V_{iQ})^2 + \cdots. \quad (9.15)$$

If we suppose that the input signal voltage is the sinusoid given by Eq. 9.5,

$$V_i - V_{iQ} = \hat{V}_i \cos \omega t,$$

the values of G_T at the positive and negative signal peaks, and at the quiescent point, are respectively:

$$G_T' = a_1 + 2a_2\hat{V}_i + 3a_3\hat{V}_i^2 + \cdots, \quad (9.16a)$$

$$G_T'' = a_1 - 2a_2\hat{V}_i + 3a_3\hat{V}_i^2 - \cdots, \quad (9.16b)$$

$$G_{TQ} = a_1. \quad (9.16c)$$

Differential error γ is defined at any point on a transfer characteristic by Eq. 9.1. The values at the positive and negative peaks of the input signal are designated by γ' and γ'', and

$$\gamma' = \frac{TF' - TF_Q}{TF_Q}, \quad (9.17a)$$

$$\gamma'' = \frac{TF'' - TF_Q}{TF_Q}. \quad (9.17b)$$

Substituting from Eqs. 9.16 for the I_o/V_i transfer characteristic, and neglecting terms beyond the third,

$$\gamma' = \frac{G_T' - G_{TQ}}{G_{TQ}} \approx \frac{2a_2\hat{V}_i + 3a_3\hat{V}_i^2}{a_1}, \quad (9.18a)$$

$$\gamma'' = \frac{G_T'' - G_{TQ}}{G_{TQ}} \approx \frac{-2a_2\hat{V}_i + 3a_3\hat{V}_i^2}{a_1}. \quad (9.18b)$$

Subtracting and adding Eqs. 9.18, and comparing with the final forms of Eqs. 9.8 yields

$$\gamma' - \gamma'' = \frac{4a_2\hat{V}_i}{a_1} \approx 8D_2, \quad (9.19a)$$

$$\gamma' + \gamma'' = \frac{6a_3\hat{V}_i^2}{a_1} \approx 24D_3. \quad (9.19b)$$

Provided that the nonlinearity is moderate (and this is essential in useful amplifiers), the approximation of neglecting high-order terms in Eqs. 9.18

and 9.19 is justified. Equations 9.19 are therefore valid, simple relations between the differential error at the signal peaks and the second- and third-harmonic distortion.

Notice that D_2 vanishes if γ' and γ'' are equal, whereas D_3 vanishes if γ' and γ'' are equal in magnitude but have opposite signs. An amplifier whose only distortion is 5% of second harmonic has extreme values of γ equal to $\pm 20\%$. An amplifier whose only distortion is 5% of third harmonic has both extreme values of γ equal to 60%. Thus, a specification in terms of differential error expands the scale of measurement; quite large values of differential error correspond to quite small values of harmonic distortion.

In calculating the harmonic distortion of a practical amplifier, the values of γ' and γ'' to be substituted into Eqs. 9.19 can be found by several methods. Perhaps the most useful are the following:

1. If graphical data such as the anode characteristics of a tube are available, a load line can be drawn and the complete transfer characteristic constructed as in Section 6.2.1.1. The small-signal transfer function is the gradient of the transfer characteristic, and γ' and γ'' can be found from the gradient at the peaks of the signal waveform.
2. If the laws of variation of the small-signal parameters with operating conditions are known, the transfer function can be calculated at any point on the signal waveform. The values of γ' and γ'' follow from the transfer function at the signal peaks.
3. Occasionally, the transfer characteristic is known as an analytical function of the input. Differentiation yields the transfer function, and γ' and γ'' follow immediately.

As a simple example (of type 3), consider an ideal transistor amplifier whose transfer characteristic is

$$I_C = I_0 \exp\left(\frac{qV_B}{kT}\right). \tag{9.20}$$

The small-signal transfer conductance at any value of input (base) voltage is therefore

$$G_T(V_B) = \frac{dI_C}{dV_B} = \frac{qI_0}{kT} \exp\left(\frac{qV_B}{kT}\right).$$

Writing the base voltage in terms of its quiescent and signal components

$$V_B(t) = V_{BQ} + V_{BS}(t),$$

the transfer conductance becomes

$$G_T(V_{BS}) = \frac{qI_0}{kT} \exp\left[\frac{q(V_{BQ} + V_{BS})}{kT}\right] = \frac{qI_{CQ}}{kT} \exp\left(\frac{qV_{BS}}{kT}\right).$$

At the peaks of a sinusoidal input signal,

$$V_{BS} = \pm \hat{V}_B$$

so that the values of G_T at the signal peaks are

$$G_T' = \frac{qI_{CQ}}{kT} \exp\left(\frac{q\hat{V}_B}{kT}\right),$$

$$G_T'' = \frac{qI_{CQ}}{kT} \exp\left(-\frac{q\hat{V}_B}{kT}\right).$$

The extreme values of differential error follow as

$$\gamma' = \left[\exp\left(\frac{q\hat{V}_B}{kT}\right) - 1\right], \tag{9.21a}$$

$$\gamma'' = \left[\exp\left(-\frac{q\hat{V}_B}{kT}\right) - 1\right], \tag{9.21b}$$

and these are plotted in Fig. 9.3*a* as functions of the peak input $\hat{V}_B$. Notice that, for this ideal transistor, the nonlinearity is independent of the quiescent current I_{CQ}. Substitution into Eqs. 9.19 gives the second- and third-harmonic distortion as

$$D_2 = \frac{1}{4} \sinh\left(\frac{q\hat{V}_B}{kT}\right) \approx \frac{1}{4}\left(\frac{q\hat{V}_B}{kT}\right), \tag{9.22a}$$

$$D_3 = \frac{1}{12}\left[\cosh\left(\frac{q\hat{V}_B}{kT}\right) - 1\right] \approx \frac{1}{24}\left(\frac{q\hat{V}_L}{kT}\right)^2. \tag{9.22b}$$

There are plotted in Fig. 9.3*b*. Notice that second-harmonic distortion dominates, and that the distortion is large for quite small input signals; a number that is easily remembered is 10% second harmonic at 10 mV ($= 0.4\ kT/q$) peak input.

9.4.1 A Useful Approximation

Equations 9.19 express the harmonic distortion in an amplifier as a function of the differential error at the signal peaks when the *input* is a sine wave. This is unfortunate from a practical viewpoint, because the small-signal parameters of an active device are usually known in terms of the *output* current (or what is proportional to it, the emitting-electrode current). If distortion is present, the output current waveform is not a sine wave and is not symmetrical; the current and transfer function at the signal peaks are therefore anything but convenient to calculate if graphical data are not available, and the distortion cannot be found without a good

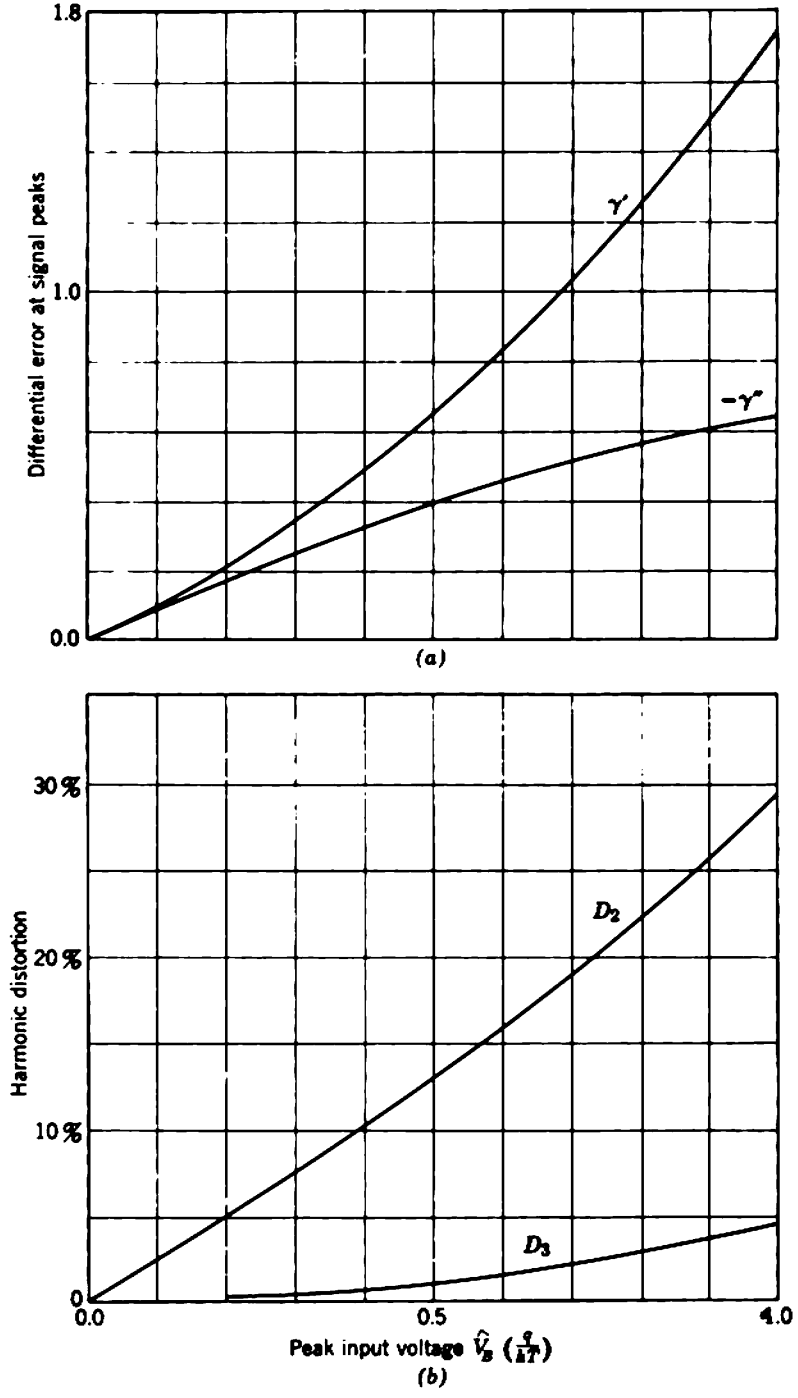


Fig. 9.3 Nonlinearity and distortion in an ideal transistor amplifier.

deal of labor. Section 9.5 considers such calculations. (The ideal transistor amplifier chosen for the example above is one of the few cases in which the calculation is straightforward.)

Fortunately, there is a simple approximation that obviates the need for a tedious calculation. This approximation is valid for many practical amplifiers, though not all. It can be shown that γ' and γ'' defined by Eqs. 9.17 may be approximated by the values of γ at the peaks of a symmetrical output waveform of the same amplitude as the actual output waveform, provided either of two conditions is satisfied. If the actual peak-to-peak output current is I_{pp} (not ordinarily symmetrical about the quiescent current I_{oQ}), then

$$\gamma' \approx \frac{TF(I_{oQ} + \frac{1}{2}I_{pp}) - TF_Q}{TF_Q}, \tag{9.23a}$$

$$\gamma'' \approx \frac{TF(I_{oQ} - \frac{1}{2}I_{pp}) - TF_Q}{TF_Q}. \tag{9.23b}$$

The conditions that must be satisfied before Eqs. 9.23 become valid approximations are either

(i) the nonlinearity must be small, the meaning of "small" being discussed in connection with Eq. 9.30 following;
(ii) the input signal amplitude must be so small that third-harmonic distortion is negligible compared with second-harmonic.

As before, the analysis is given in terms of the I_o/V_i transfer characteristic.

Over the quasi-linear region the power series representation of the transfer characteristic can be approximated by

$$I_o = kV_i^n. \tag{9.24}$$

The exponent n gives a measure of the nonlinearity, and is unity for a linear amplifier. By differentiation, the transfer conductance is

$$G_T = \frac{dI_o}{dV_i} = nkV_i^{n-1} \tag{9.25}$$

or alternatively, by substitution from Eq. 9.24,

$$G_T = nk\left(\frac{I_o}{k}\right)^{(n-1)/n} \tag{9.26}$$

The method is to find the differential error in the transfer characteristic at the peaks of a symmetrical input waveform from Eq. 9.25, and at the peaks of a symmetrical output waveform from Eq. 9.26. The conditions under which the two are the same are also the conditions under which Eqs. 9.23 are valid approximations.

As before, the signal input voltage is the sinusoid given by Eq. 9.5:

$$V_i - V_{iQ} = \hat{V}_i \cos \omega t.$$

The values of output current at the signal peaks follow from Eq. 9.24 as

$$I'_o = k(V_{iQ} + \hat{V}_i)^n,$$

$$I''_o = k(V_{iQ} - \hat{V}_i)^n.$$

Taking the first three terms in the binomial expansion, and then subtracting, the peak-to-peak signal output current is

$$\frac{I_{pp}}{I_{oQ}} = 2n\left(\frac{\hat{V}_i}{V_{iQ}}\right). \tag{9.27}$$

The values of G_T at the peaks of the symmetrical input waveform follow immediately from Eq. 9.25 as

$$G'_T = nk(V_{iQ} + \hat{V}_i)^{n-1},$$

$$G''_T = nk(V_{iQ} - \hat{V}_i)^{n-1}.$$

Taking the first three terms in the binomial expansions and then subtracting, the differential errors at the peaks are

$$\gamma' = (n-1)\left(\frac{\hat{V}_i}{V_{iQ}}\right) + \frac{(n-1)(n-2)}{1 \cdot 2}\left(\frac{\hat{V}_i}{V_{iQ}}\right)^2, \tag{9.28a}$$

$$\gamma'' = -(n-1)\left(\frac{\hat{V}_i}{V_{iQ}}\right) + \frac{(n-1)(n-2)}{1 \cdot 2}\left(\frac{\hat{V}_i}{V_{iQ}}\right)^2. \tag{9.28b}$$

The values of G_T at the peaks of a symmetrical output waveform of the same peak-to-peak amplitude as the actual output waveform follow immediately from Eq. 9.26 as

$$G_T(I_{oQ} + \tfrac{1}{2}I_{pp}) = nk\left(\frac{I_{oQ} + \frac{1}{2}I_{pp}}{k}\right)^{(n-1)/n},$$

$$G_T(I_{oQ} - \tfrac{1}{2}I_{pp}) = nk\left(\frac{I_{oQ} - \frac{1}{2}I_{pp}}{k}\right)^{(n-1)/n}.$$

Replacing I_{pp} from Eq. 9.27 and then taking the first three terms in the binomial expansions and subtracting, the differential errors at the peaks are

$$\gamma(I_{oQ} + \tfrac{1}{2}I_{pp}) = (n-1)\left(\frac{\hat{V}_i}{V_{iQ}}\right) + \frac{(n-1)(-1)}{1 \cdot 2}\left(\frac{\hat{V}_i}{V_{iQ}}\right)^2, \tag{9.29a}$$

$$\gamma(I_{oQ} - \tfrac{1}{2}I_{pp}) = -(n-1)\left(\frac{\hat{V}_i}{V_{iQ}}\right) + \frac{(n-1)(-1)}{1 \cdot 2}\left(\frac{\hat{V}_i}{V_{iQ}}\right)^2. \tag{9.29b}$$

The first terms on the right-hand sides of Eqs. 9.28 and 9.29 are the same, but the second terms are not, in general. Thus, Eqs. 9.23 are not universally valid approximations. Notice, however, that the first terms in Eqs. 9.28 and 9.29 alone appear in the expression for second-harmonic distortion. The conditions that must be satisfied before Eqs. 9.29 reduce to Eqs. 9.28 are either:

(i) The exponent n must be close to unity. If

$$n = 1 + \delta, \tag{9.30}$$

the second terms are nearly equal if $|\delta|$ is less than about 0.2.

(ii) The input voltage $\hat{V}_i$ must be so small that the second terms in Eqs. 9.28 and 9.29 can be neglected. This implies that third-harmonic distortion is neglible.

The former condition, n close to unity, normally applies only to amplifiers in which effective steps have been taken to reduce the nonlinearity. Feedback amplifiers (discussed in later chapters) are the most notable examples, and for these amplifiers Eqs. 9.23 are good approximations. For simple amplifiers of the types considered thus far in this book, Eqs. 9.23 can be used to predict the second-harmonic distortion, but they are quite seriously in error for the third-harmonic.

9.5 ESTIMATES OF DIFFERENTIAL ERROR

This section considers a method for estimating differential error at any point on a transfer characteristic, graphical data not being available. Such information may be required in order to define an operating range over which the small-signal transfer function remains within specified limits, or in a calculation of harmonic or intermodulation distortion. In essentials, the method is to express the transfer function in terms of device small-signal parameters whose laws of variation with operating conditions are known, and then substitute particular values into Eq. 9.1. Use is made of the ideal laws derived in Chapters 3 and 4; for a vacuum tube

$$g_m \propto I_K^{1/2}, \tag{9.31}$$

whereas for a transistor

$$r_E \propto I_E^{-1}. \tag{9.32}$$

As practical devices may not follow these laws exactly, the results obtained may not be exact either. They are, however, approximately correct and they have the great advantage of being general, whereas results derived from the actual law of a particular device would not.

The problem is to find the differential error at any point on the transfer characteristic. This is straightforward when the point is specified in terms of the output signal, because the small-signal parameters and transfer function are known as functions of the output current (or what is proportional to it, the emitting-electrode current). When the point is specified in terms of the input signal, the calculation might proceed thus:

Consider, for example, the transfer conductance, which is known as a function of output current:

$$G_T = \frac{dI_o}{dV_i} = G_T(I_o). \tag{9.33}$$

Integration yields

$$\int dV_i = \int \frac{dI_o}{G_T(I_o)}.$$

This can be written in functional notation as

$$V_i = H(I_o), \tag{9.34}$$

where H depends, in general, on the quiescent current I_{oQ}. In principle, Eq. 9.34 can be solved for the output current:

$$I_o = H^{-1}(V_i). \tag{9.35}$$

Equation 9.35 is the I_o/V_i transfer characteristic. By differentiation, the transfer conductance as a function of input voltage is

$$G_T = \frac{dI_o}{dV_i} = \frac{d}{dV_i}[H^{-1}(V_i)] = G_T(V_i). \tag{9.36}$$

The value of γ at any input voltage can now be found by direct substitution into Eq. 9.1.

In practice, Eq. 9.34 cannot often be solved in a convenient form. The method adopted in the following sections is to plot G_T (given by Eq. 9.33) against V_i (given by Eq. 9.34); the resulting curve is Eq. 9.36. Suitable normalization is incorporated so that values of γ can be read directly from the graphs. The analyses are, of course, valid only in the region of the transfer characteristic well away from the clipping points.

9.5.1 Vacuum Tubes

The small-signal parameters of an ideal vacuum tube vary according to the laws

$$g_m \propto I_K^{1/3} \propto I_A^{1/3},$$
$$r_A \propto I_K^{-1/3} \propto I_A^{-1/3},$$
$$\mu = \text{constant}.$$

Equation 5.45 gives the transfer conductance of a tube as

$$G_T = \frac{i_A}{v_G} = \frac{dI_A}{dV_G} = \frac{\mu}{r_A + R_L}$$

and nonlinearity stems from the variation of r_A with anode current. At any instantaneous current I_A and for a quiescent current I_{AQ},

$$G_T(I_A) = \frac{dI_A}{dV_G} = \frac{\mu}{r_{AQ}\left(\frac{I_A}{I_{AQ}}\right)^{-1/3} + R_L} = \frac{g_{mQ}}{\left(\frac{I_A}{I_{AQ}}\right)^{-1/3} + \left(\frac{R_L}{r_{AQ}}\right)}. \quad (9.37)$$

Equation 9.37 corresponds to Eq. 9.33 of the general analysis.

Writing V_G and I_A in terms of their quiescent and signal components

$$V_G = V_{GQ} + V_{GS},$$

$$I_A = I_{AQ} + I_{AS},$$

and integrating Eq. 9.37 about the quiescent point yields

$$\int_{V_{GQ}}^{V_{GQ}+V_{GS}} g_{mQ}\, dV_G = \int_{I_{AQ}}^{I_{AQ}+I_{AS}} \left[\left(\frac{I_A}{I_{AQ}}\right)^{-1/3} + \left(\frac{R_L}{r_{AQ}}\right)\right] dI_A.$$

This can be rearranged into the form of Eq. 9.34:

$$V_G = \left(\frac{I_{AQ}}{g_{mQ}}\right)\left[\frac{3}{2}\left(1 + \frac{I_{AS}}{I_{AQ}}\right)^{2/3} + \frac{R_L}{r_{AQ}}\left(1 + \frac{I_{AS}}{I_{AQ}}\right) - \frac{3}{2} - \frac{R_L}{r_{AQ}}\right]. \quad (9.38)$$

Figure 9.4 is a normalized plot of Eq. 9.37 versus Eq. 9.38. The ordinate is the differential error γ, and the abscissa is the signal excursion (normalized to I_{AQ}/g_{mQ}) about the quiescent point. The variable parameter in the family of curves is the ratio of load resistance to quiescent anode resistance, R_L/r_{AQ}. This ratio approaches zero for pentodes, because their anode resistance is very large. Notice that Fig. 9.4 applies for all transfer functions of a vacuum tube, not merely the transfer conductance; Table 5.3 shows that other transfer functions differ from the transfer conductance only by a constant multiplier.

As an example in the use of Fig. 9.4, consider the amplifier stage discussed in Section 5.2.2. A triode, type ECC83/12AX7, is operated at the quiescent point

$$V_{AQ} = 150 \text{ V},$$

$$I_{KQ} = 0.5 \text{ mA}.$$

and with a load resistance

$$R_L = 100 \text{ k}\Omega.$$

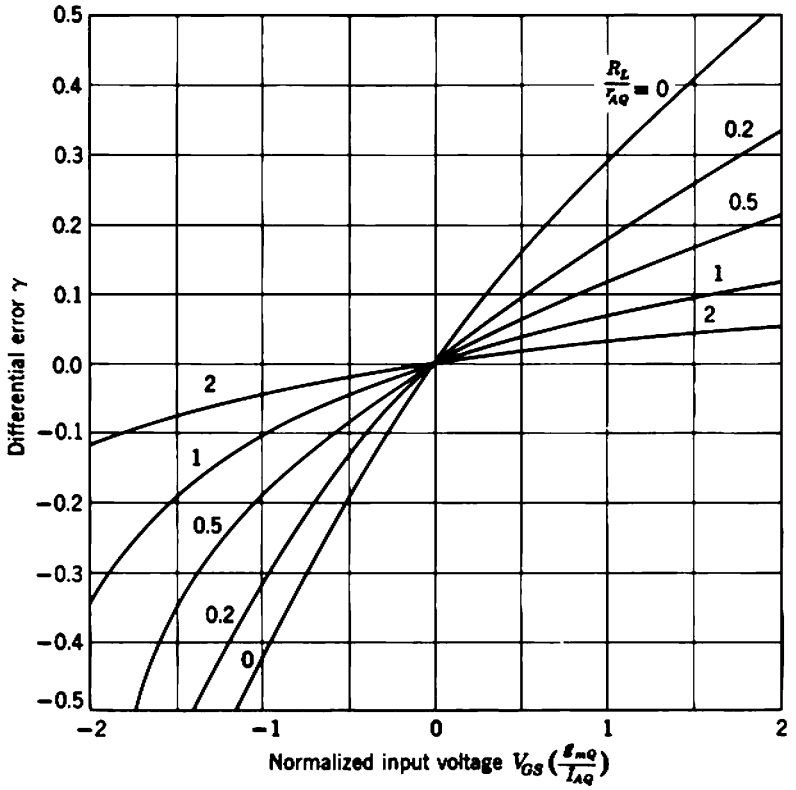


Fig. 9.4 Universal curve for estimating the differential error in a vacuum-tube amplifier.

The small-signal parameters at the chosen quiescent point are

$$g_{mQ} = 1 \text{ mA/V},$$
$$r_{AQ} = 100 \text{ k}\Omega,$$
$$\mu = 100.$$

It is required to find the values of γ at peak signal inputs of ± 1 V. The normalized signal input and parameter for the family of curves are

$$V_{GS}\left(\frac{g_{mQ}}{I_{AQ}}\right) = \pm 2,$$
$$\frac{R_L}{r_{AQ}} = 1,$$

and the values of differential error are read from Fig. 9.4 as

$$\gamma' = +0.12,$$
$$\gamma'' = -0.33.$$

The corresponding values computed from manufacturers' published anode characteristics are

$$\gamma' = +0.14,$$

$$\gamma'' = -0.36.$$

For this particular tube, there is excellent agreement between practice and the idealized theory.

As a general rule, Fig. 9.4 gives quite accurate estimates of differential error provided that

$$\frac{R_L}{r_{AQ}} < 1.$$

At larger values, the predicted nonlinearity due to variations of g_m and r_A is small and the second-order variation of μ (neglected in the derivation) becomes the dominant source of nonlinearity. Figure 9.4 tends to be optimistic by a factor of two or more for medium-μ triodes with large loads. For other cases, the error is unlikely to exceed 20%.

9.5.2 Transistors

The small-signal parameters of an ideal transistor vary according to the laws

$$r_E \propto I_E^{-1} \propto I_C^{-1},$$

$$\alpha_N, \beta_N, \text{ and } r_B \text{ are constants.}$$

Any transfer function can be manipulated into the form

$$TF = \frac{A}{R + r_E/\alpha_N},$$

where A and R are constants, and nonlinearity arises from the variation of r_E with emitter current. Two examples from Section 5.3.2 are Eqs. 5.76 and 5.81:

$$G_T = \frac{i_C}{v_B} = \frac{\alpha_N}{(1 - \alpha_N)r_B + r_E} = \frac{1}{r_B/\beta_N + r_E/\alpha_N},$$

$$\frac{i_C}{i_S} = \frac{\alpha_N R_S}{(1 - \alpha_N)(R_S + r_B) + r_E} = \frac{R_S}{(R_S + r_B)/\beta_N + r_E/\alpha_N}.$$

Nonlinearity stems from the variation of r_E with collector current:

$$\frac{r_E}{\alpha_N} = \frac{kT}{\alpha_N q I_E} \approx \frac{kT}{qI_C}.$$

Consider the transfer conductance:

$$G_T(I_C) = \frac{dI_C}{dV_B} = \frac{1}{r_B/\beta_N + kT/qI_C}, \tag{9.39}$$

which corresponds to Eq. 9.33 of the general analysis. Writing V_B and I_C in terms of their quiescent and signal components,

$$V_B = V_{BQ} + V_{BS},$$

$$I_C = I_{CQ} + I_{CS},$$

and integrating Eq. 9.39 about the quiescent point yields

$$\int_{V_{BQ}}^{V_{BQ}+V_{BS}} dV_B = \int_{I_{CQ}}^{I_{CQ}+I_{CS}} \left(\frac{r_B}{\beta_N} + \frac{kT}{qI_C}\right) dI_C.$$

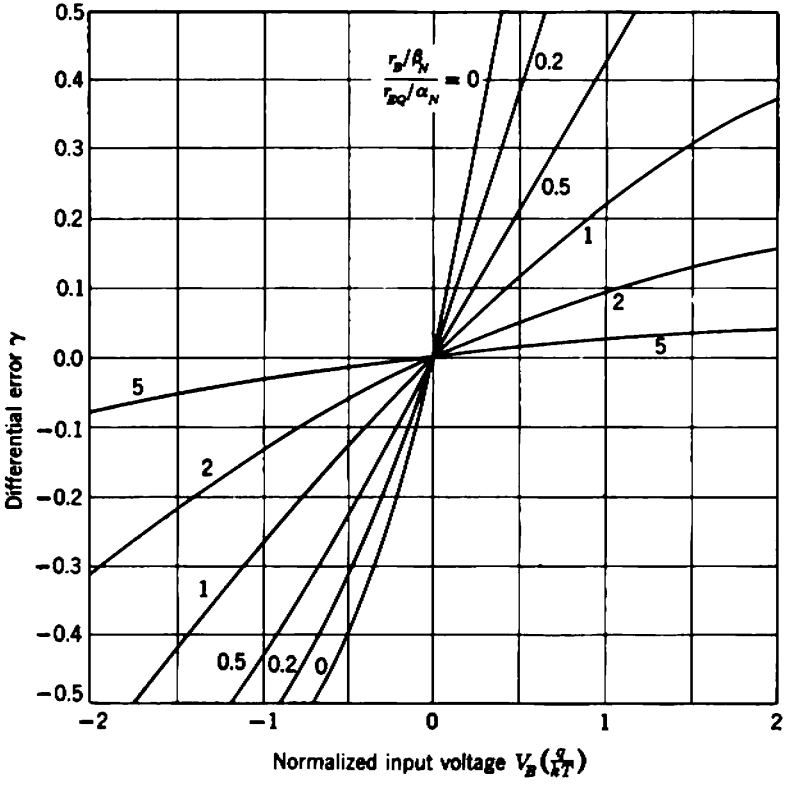


Fig. 9.5 Universal curve for estimating the differential error in a transistor amplifier. These curves can be applied to a transfer function defined in terms of the source internal voltage generator, provided the following substitutions are made:

$$V_B \rightarrow V_S$$

$$r_B \rightarrow R_S + r_B$$

This can be manipulated into the form of Eq. 9.34:

$$V_{BS} = \frac{kT}{q}\left[\frac{r_B/\beta_N}{r_{EQ}/\alpha_N}\left(\frac{I_{CS}}{I_{CQ}}\right) + \ln\left(1 + \frac{I_{CS}}{I_{CQ}}\right)\right]. \tag{9.40}$$

Figure 9.5 is a normalized plot of Eq. 9.39 versus Eq. 9.40. The ordinate is the differential error γ, and the abscissa is the signal excursion (normalized to kT/q) about the quiescent point. The variable parameter in the family of curves is

$$\frac{r_B/\beta_N}{r_{EQ}/\alpha_N} \quad \text{or} \quad \frac{(R_S + r_B)/\beta_N}{r_{EQ}/\alpha_N},$$

as appropriate for the transfer function under consideration. Figure 9.5 applies reasonably accurately to the nonlinearity in any transistor amplifier, provided that

$$\text{parameter} < 2.$$

At larger values, the predicted nonlinearity due to variation of r_E is small, and the second-order variations of other parameters (notably β_N) become the dominant sources of nonlinearity.*

9.6 CONCLUDING COMMENTS ON NONLINEARITY

Section 9.5 conclusively demonstrates a very important fact: the differential error at any point on the transfer characteristic of a simple amplifier increases as the point in question moves further from the quiescent point. In other words, the effect of nonlinearity in an amplifier becomes more significant as the signal amplitude is made larger. This fact is hinted at in a number of other places; for example, Eq. 9.9 shows that harmonic distortion is proportional to signal amplitude raised to some power.

In typical amplifiers the gain per stage is such that the signal amplitude increases by an order of magnitude or more for each stage. By far the largest signal occurs in the output stage. Therefore the nonlinearity of the output stage dominates in the generation of distortion components, and the distortion performance of an amplifier can be estimated quite accurately from consideration of the output stage alone. In this respect distortion is the antithesis of noise; the noise performance of an amplifier is largely determined by its input stage.

* A good discussion of nonlinearity in transistor amplifiers is: G. M. RIVA, P. J. BENETEAU, and E. DALLA VOLTA, "Amplitude distortion in transistor amplifiers," *Proc. Inst. Elec. Engrs., London*, 111, 481, March 1964.

Numerical values for γ are typically some tens of percent at the peaks of a signal waveform that drives an amplifier through half the quasi-linear region of its transfer characteristic between the clipping points. Corresponding values of harmonic and intermodulation distortion are a few percent. This order of nonlinearity is quite unacceptable in precision amplifiers such as those used in measuring instruments and, indeed, in many less stringent applications. Distortion can be reduced by using amplifiers that are capable of much more output than is actually required so that signal excursions occupy a smaller proportion of the transfer characteristic, but this is very uneconomical. There is a pressing demand for amplifier circuits of greater sophistication than those discussed so far in this book. Negative feedback, which forms the subject matter of later chapters, is the standard technique for reducing nonlinearity.

9.7 OVERLOAD AND OVERLOAD RECOVERY

This final section describes briefly the effects of overloading an amplifier, that is, of applying an input signal large enough to drive the amplifier out of the quasi-linear region of its transfer characteristics. An obvious effect is that the output is grossly distorted during the time interval for which the overload persists. Less obvious is the fact that the output may remain distorted for some time after the overloading signal is removed.

Equation 9.7 shows that the average output-electrode current of a device is a function of the applied signal amplitude; the term A_0 represents a dynamic shift in operating current. This shift is small for signal amplitudes in the quasi-linear range, but tends to be large under overload conditions because the coefficients of high-order terms in $\hat{V}_i$ are large. Usually A_0 is positive, corresponding to an increase in average current under signal conditions.

The biasing circuit of an amplifier stage is designed expressly to hold the average device current constant. When a signal is first applied to an amplifier the biasing voltages have their quiescent values but, as the biasing circuit senses the change in average current and the charges on the bypass capacitors change, the biasing voltages change in the directions that tend to restore the average current to its original value. When the signal is removed the excess charge leaks away through the biasing resistors and the amplifier slowly reverts to its original quiescent point. Such transient changes in biasing voltages with signal result in an output from the amplifier, a form of low-frequency noise. Under overload conditions, the changes in biasing voltages may be so great that, at the conclusion of the overload, the amplifier is biased out of its quasi-linear operating region.

As a result, small signals, which would ordinarily be within the amplifier's capability, are in fact grossly distorted until the excess charge on the bypass capacitors leaks away. Distortion or complete suppression of a small signal immediately following an overload is called *blocking* or *paralysis*. The *dead time* elapses between removal of the overload and the amplifier's reverting to normal operation.

A similar mechanism for blocking an amplifier is that the average control electrode current changes under overload conditions, again resulting in a disturbance of a biasing voltage. Grid current in a vacuum tube is negligible for normal operating conditions but rises sharply if the grid is driven positive with respect to the cathode. In a transistor, β_N falls to a low value at the saturation clipping point for which V_{CE} approaches zero, and base current increases. As well as being a potential source of blocking, the change in control electrode current constitutes a nonlinearity in the input impedance and causes nonlinearity in the transfer characteristic of the preceding stage.

As an example of blocking, Fig. 9.6 shows the effect of overloading a self-biased triode amplifier stage. The behavior of more complicated circuits is qualitatively similar, but may be complicated in detail by the presence of additional bypass capacitors or other energy storage elements.

In principle, blocking can be eliminated entirely by using amplifiers of dynamic range large enough to cope with any signal that might be encountered. This, however, is quite unrealistic, and the practical conclusion is that all amplifiers may be subjected to occasional overloads, while some amplifiers may be required to operate in the presence of almost continuous overload signals. At the one extreme are amplifiers such as those in radio receivers; these have adequate dynamic range for normal signals, but may be overloaded under fault conditions, for example, by a burst of static. At the other extreme are certain nuclear pulse amplifiers which are required to amplify small pulses despite the presence of interfering pulses some 100 times larger occurring at a higher mean rate. In the former case, no particular precautions need be taken because blocking occurs only under fault conditions. Overload performance in the latter case is of extreme importance, and is discussed in specialized books. A few general principles apply to all amplifiers:

The extent to which an amplifier is blocked by an overload depends on the duration of the overload relative to the time constants of the biasing system, and on the magnitude of the overload. The dead time following an overload depends on the extent of the blocking, and on the time constants of the biasing system. An amplifier with long biasing time constants may not block at all in the presence of a short duration overload, because

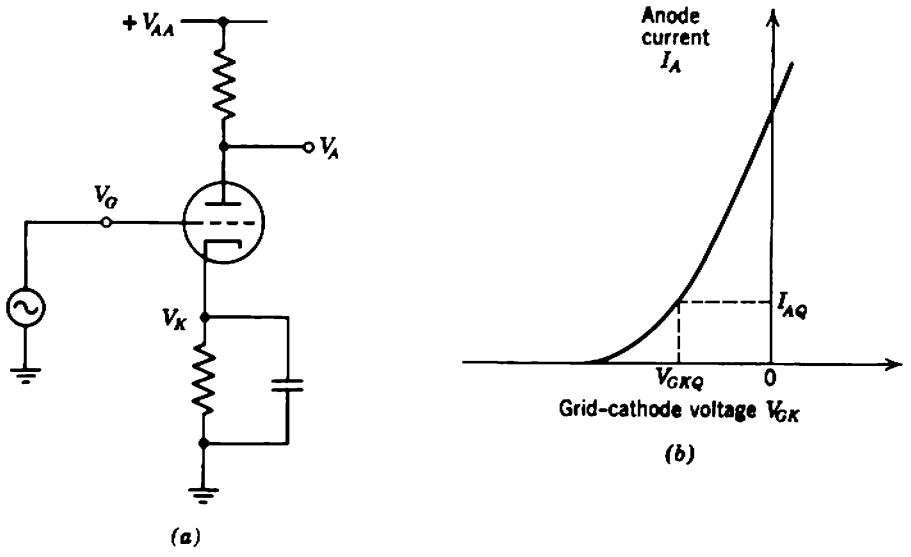


Fig. 9.6 Overload in a simple amplifier: (*a*) circuit diagram; (*b*) transfer characteristic; (*c*) to (*f*) waveforms.

insufficient excess charge accumulates on the large capacitors to bias the amplifier out of its linear operating region. The larger an overload, the greater the change in mean electrode currents, and the faster an amplifier approaches the blocked condition. Overloads beyond a certain amplitude and duration cause complete blocking in which the biasing voltages reach quasi-equilibrium values and further overloading causes no change.

For given signal conditions, there are circuit techniques for increasing the average time for which an amplifier is operative relative to the time for which it is blocked. As an example, reducing the biasing time constants speeds recovery, but also reduces (in the same ratio) the overload duration required to cause blocking. However, reducing the biasing time constant has the additional effect of increasing the probability of a complete blockage; the recovery time after a complete blockage is independent of the duration of the overload, and overall there is a fractional increase in the operating time. The price paid for reducing the biasing time constants is a degradation of the low-frequency response. Another circuit technique is the use of catching diodes to reduce the biasing voltage excursions at complete blockage, thereby speeding the recovery. Blocking can be prevented entirely by eliminating all capacitors and other energy stores from the biasing system, that is, by using a dc amplifier, but this introduces new problems described in Section 15.3.

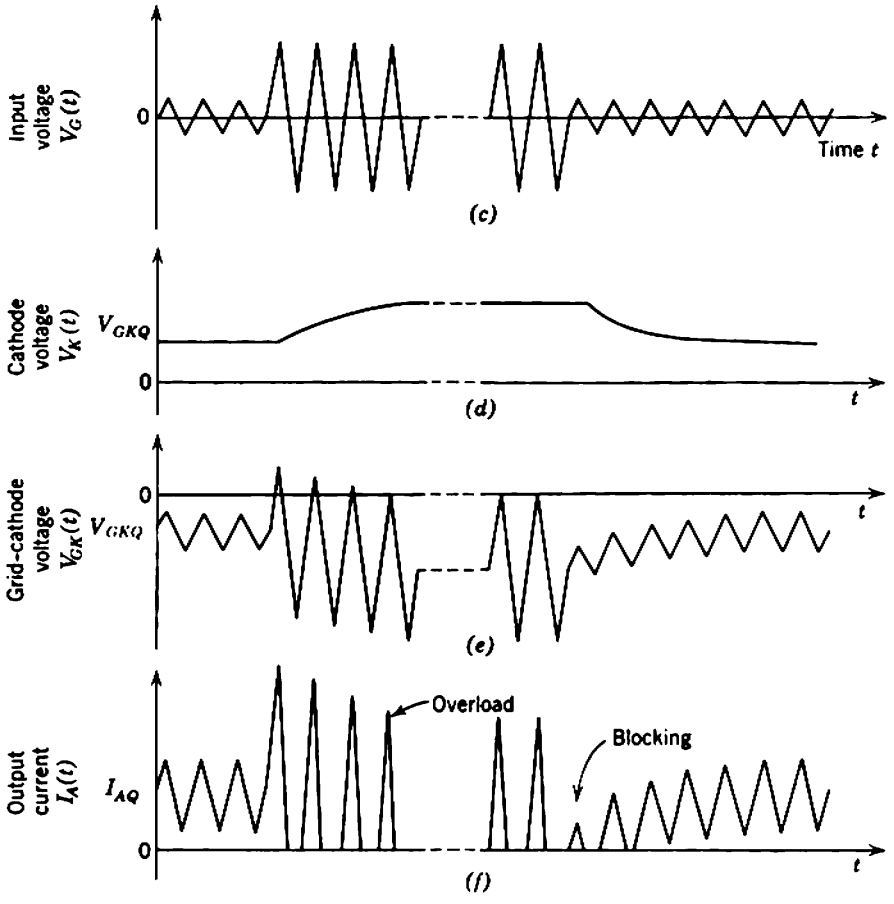


Fig. 9.6 (*continued*)

Chapter 10

Introduction to Negative Feedback Theory

It is shown in Chapter 5 that amplifier stages using unselected active devices can be designed to have gains of moderate magnitude and stability simply by fixing the quiescent voltages and currents within specified limits and by shunting the signal path with appropriate resistors. In this way stage gains ranging from a few tens to a few hundreds can be achieved with a tolerance of perhaps $\pm 20\%$. In general, there can be an exchange between gain magnitude and gain stability; vacuum pentodes yield the highest gain per stage but transistors are capable of the most stable gain. Multistage amplifiers can be formed by cascading stages. The gain can be increased without limit if enough stages are used, but the over-all stability falls as the number of stages is increased. A practical conclusion is that it is virtually impossible to achieve large gains with stability better than $\pm 10\%$ when the amplifier uses unselected components. This order of stability is inadequate for many purposes.

Because selection of components is not permissible in economic, designable circuits, some different design technique is required for high-stability amplifiers. The established technique is to use *negative feedback* around an amplifier as illustrated in Fig. 10.1. The input signal to the system is compared continuously with a *feedback signal* which is a constant fraction β of the output; β is known as the *feedback factor*, and may be dimensional and/or frequency-dependent. The difference between the input and feedback signals (the *error signal*) is fed into the *forward path* μ or amplifying part of the system, and the phase relations are made such that the circuit adjusts itself for minimum error signal. If the forward gain μ is very large, the error signal must be very small; the input and feedback

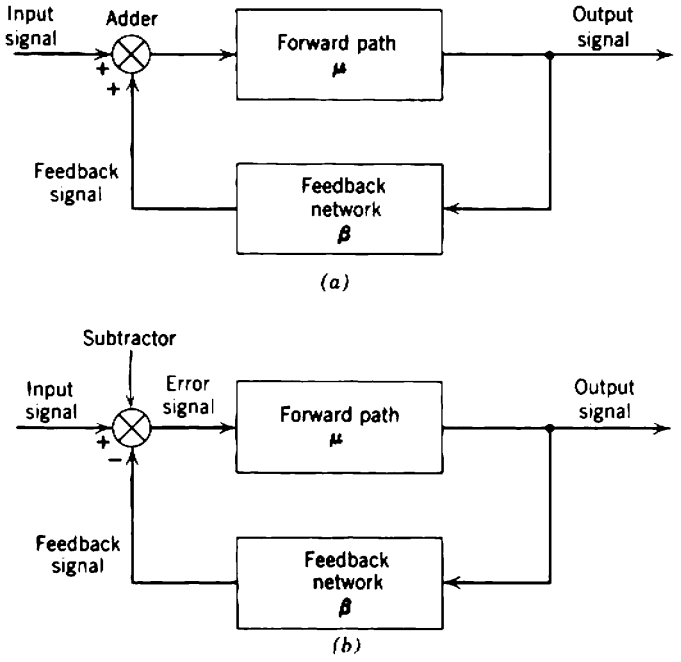


Fig. 10.1 Block diagram of a feedback system: (*a*) electronics sign convention; (*b*) control sign convention.

signals are equal in magnitude but of opposite signs, and the over-all gain of the system is $1/\beta$. Because β is a constant fraction, which can be fixed to a high degree of precision by passive circuit elements, the over-all gain of the amplifier with feedback is (ideally) independent of the amplifier itself.

In practical feedback circuits, the input and feedback signals are not exactly equal in magnitude and, therefore, the gain is not determined entirely by the passive feedback network. Nevertheless, gain stabilities of almost any finite degree can be achieved.

Because the input-output relation of a feedback amplifier is largely independent of its amplifying section, feedback amplifiers have at least two further advantages over nonfeedback amplifiers:

(i) distortion due to nonlinearities is reduced;
(ii) the close tolerances on bias currents and voltages can be relaxed somewhat.

A further, less obvious use for negative feedback is to change the input and output impedances. Unfortunately, the advantages and general

flexibility of feedback circuits are achieved at the expense of new design complications; in particular, the small-signal dynamic response of an amplifier with feedback is radically different from that of the same amplifier without feedback.

The various aspects of feedback circuits introduced above are considered in the sections that follow. The treatment in this chapter is mainly theoretical; specific applications of the theoretical principles are given in Chapters 11, 12, 13, and 14.

10.1 ELEMENTARY EXAMPLE OF NEGATIVE FEEDBACK

Consider, as an introductory example of the use of negative feedback, the pentode amplifier circuits shown in Fig. 10.2. The circuit of Fig. 10.2*a* is obviously not a feedback amplifier; the signal between grid and cathode is due only to the input voltage. This voltage produces the anode signal current which develops the output signal voltage across the load resistor, and the mid-band voltage gain is

$$A_V = \frac{v_o}{v_i} = -g_m R_L\left(\frac{1}{1 + R_L/r_A}\right). \tag{10.1}$$

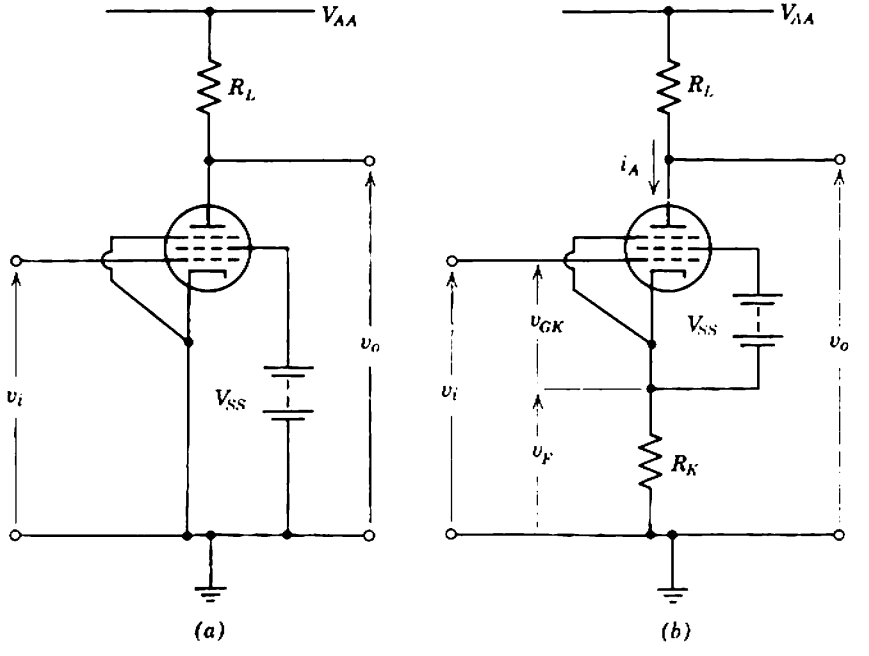


Fig. 10.2 Elementary example of negative feedback: (*a*) amplifier without feedback; (*b*) amplifier with feedback.

For normal values of r_A and R_L this reduces to

$$A_V \approx -g_m R_L, \tag{10.2}$$

and the tolerance on g_m (adopted throughout this book as lying between 60 and 150% of the nominal value) becomes the tolerance on A_V.

The amplifier can be modified as shown in Fig. 10.2*b*, by adding the resistor R_K in the cathode lead. Under these circumstances the anode current is controlled directly by the grid-to-cathode voltage v_{GK}, which is the difference between the input voltage and the feedback voltage v_F developed across R_K:

$$v_{GK} = v_i - v_F.$$

This feedback voltage is derived from the output current by means of the passive element R_K:

$$v_F = R_K i_A.$$

If the mutual conductance of the pentode approaches infinity, v_{GK} must approach zero and v_F approaches v_i. Hence

$$G_T = \frac{i_o}{v_i} = \frac{i_A}{v_i} \to \frac{1}{R_K} \tag{10.3}$$

and

$$A_V = -G_T R_L \to -\frac{R_L}{R_K}. \tag{10.4}$$

For finite values of g_m network analysis shows that

$$A_V = -g_m R_L \left[\frac{r_A}{r_A + R_L + (\mu + 1)R_K}\right]. \tag{10.5}$$

However, μ is very large for a pentode, so that

$$\mu + 1 \approx \mu = g_m r_A.$$

Therefore A_V reduces to

$$A_V \approx -g_m R_L \left(\frac{1}{1 + R_L/r_A + g_m R_K}\right). \tag{10.6}$$

As $g_m R_K$ is increased from zero, it ultimately becomes the dominating term in the denominator of Eq. 10.6; G_T and A_V approach their asymptotic values $(1/R_K)$ and $(-R_L/R_K)$.

The effect of feedback on the magnitude and stability of voltage gain is shown more clearly by a numerical example. Suppose a pentode is to be operated with a 10 kΩ load at a quiescent point for which the nominal mutual conductance is

$$g_{mA} = 5 \text{ mA/V}.$$

Subject to the assumption that (R_L/r_A) is negligibly small, the effect of varying R_K from zero to 1.4 kΩ is to be investigated.

The variation of A_V with R_K for tubes having nominal, upper-limit and lower-limit values of g_m can be determined from Eq. 10.6 and is plotted in Fig. 10.3. Also shown is the asymptotic value of voltage gain, R_L/R_K. A logarithmic scale for A_V enables percentage spreads in gain to be determined directly. Notice the following points:

1. Increasing R_K decreases the voltage gain. Ultimately, the magnitude of A_V becomes too small to be useful.
2. Increasing R_K decreases the relative spread in gain between upper-limit and lower-limit values of g_m. When R_K is zero, A_V varies from -40 to $+50\%$ about the nominal value, whereas, when $R_K = 1.4$ kΩ, A_V varies from -7.68 to $+4.32\%$ about the nominal value.
3. Increasing R_K decreases the error between the actual gain and the asymptotic value given by R_L/R_K. For the useful range of gain magnitude

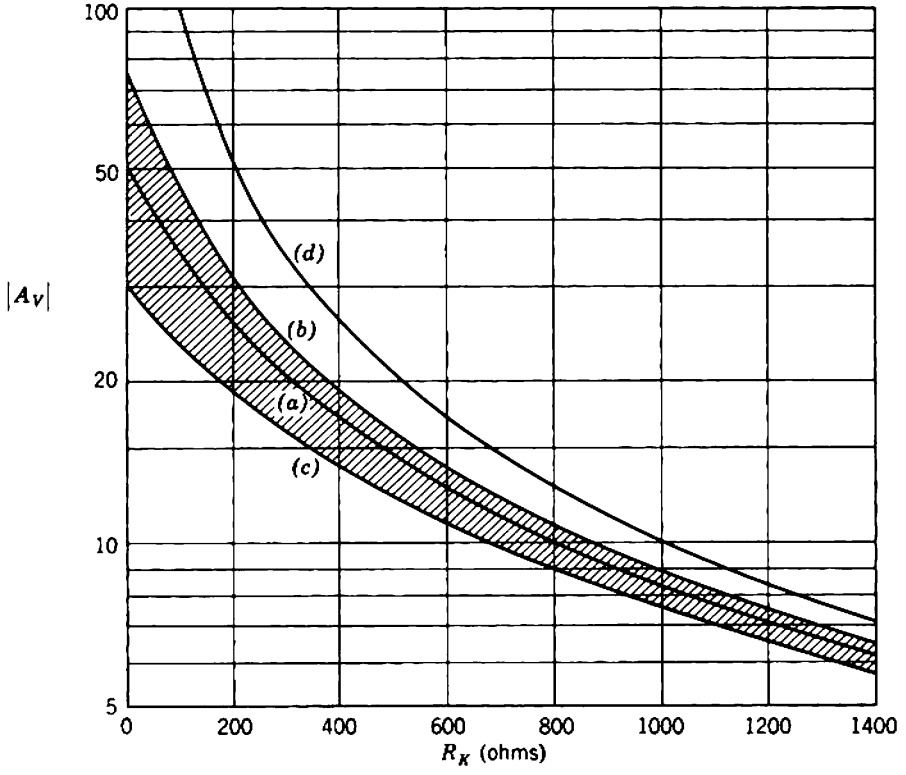


Fig. 10.3 Voltage gain versus R_K for Fig. 10.2*b*: curve (*a*)—nominal g_m; curve (*b*)—upper-limit $g_m = 1.5\,g_{mA}$; curve (*c*)—lower-limit $g_m = 0.6\,g_{mA}$; curve (*d*)—ideal case $A_V = R_L/R_K$.

shown, however, there is still a substantial error between the actual value and the ratio R_L/R_K. For example, when $R_K = 1.4\ \text{k}\Omega$, R_L/R_K is 14.4% greater than the nominal gain; this error is greater than the total percentage spread in gain. In many other practical cases similar discrepancies exist between accurate results and approximate results based on the assumption of infinite amplifier gain. It is therefore prudent to use these approximate results only as order of magnitude checks.

10.2 COMPUTATIONAL DIFFICULTIES

Linear passive network theory is one of the really basic tools of electrical engineering. This theory can be extended readily to combinations of linear networks and active devices, but while straightforward in principle, such extensions are not always useful in design work. Network theoreticians are concerned with the more general aspects of formal active network theory and, as a result of many basic investigations, substantial theoretical contributions have been made to the totality of knowledge. Formal approaches based on matrices and determinants can be used to great effect in analyzing general feedback circuits,* but in specific numerical examples they have the major deficiency of not showing clearly which engineering approximations are valid until after an elaborate analysis has been completed.

To illustrate this point consider the three networks shown in Fig. 10.4. These are a passive network, a network containing a controlled source (though without feedback), and a feedback network. All can be analyzed by a formal node-equation determinant approach.

The network of Fig. 10.4*a* is driven from a current source i_i, and it is required to find the transfer impedance v_o/i_i. The node equations are

$$\begin{pmatrix} i_i \\ 0 \end{pmatrix} = \begin{pmatrix} Y_1 + Y_F & -Y_F \\ -Y_F & Y_2 + Y_F \end{pmatrix} \begin{pmatrix} v_i \\ v_o \end{pmatrix}. \tag{10.7}$$

Thence

$$\frac{v_o}{i_i} = \frac{\Delta_{12}}{\Delta},$$

where Δ is the network determinant

$$\Delta = \begin{vmatrix} Y_1 + Y_F & -Y_F \\ -Y_F & Y_2 + Y_F \end{vmatrix}$$

* A classic book of this type which will repay study is: H. W. Bode, *Network Analysis and Feedback Amplifier Design* (van Nostrand, Princeton, N.J., 1945).

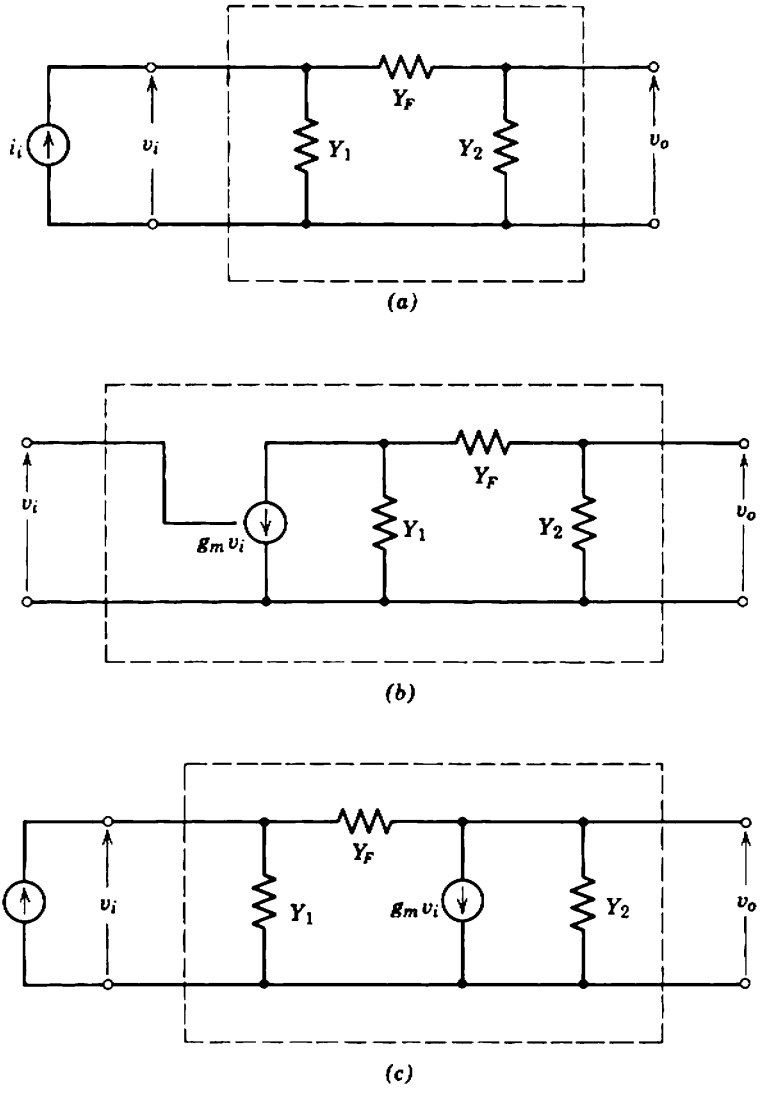


Fig. 10.4 Examples of networks: (*a*) passive network; (*b*) network with controlled source; (*c*) feedback network.

which is symmetrical about its main diagonal, and

$$\Delta_{12} = Y_F$$

is a cofactor. Hence

$$\frac{v_o}{i_i} = \frac{Y_F}{Y_1(Y_2 + Y_F) + Y_2 Y_F}. \tag{10.8}$$

Next, consider the example of Fig. 10.4*b* in which a controlled current source $g_m v_i$ is included in the network. The controlling voltage v_i is an independent variable, and from Eq. 10.8

$$\frac{v_o}{v_i} = -g_m\left[\frac{Y_F}{Y_1(Y_2 + Y_F) + Y_2 Y_F}\right]. \tag{10.9}$$

In this example the presence of the controlled source introduces no significant complication because the passive network and the controlled source may be considered as separate entities.

Consider finally the example shown in Fig. 10.4*c* in which the controlled source $g_m v_i$ is controlled by a voltage that depends in part on the controlled source itself. This type of behavior is exhibited by all feedback circuits. If the input current i_i is provided by an independent source, a formal evaluation of v_o/i_i can be based on the following node equations:

$$\begin{pmatrix} i_i \\ 0 \end{pmatrix} = \begin{pmatrix} Y_1 + Y_F & -Y_F \\ g_m - Y_F & Y_2 + Y_F \end{pmatrix}\begin{pmatrix} v_i \\ v_o \end{pmatrix}. \tag{10.10}$$

Thence

$$\frac{v_o}{i_i} = \frac{\Delta'_{12}}{\Delta'},$$

where

$$\Delta' = \begin{vmatrix} Y_1 + Y_F & -Y_F \\ g_m - Y_F & Y_2 + Y_F \end{vmatrix}$$

is the determinant for the passive network with a controlled source imbedded in it. This determinant is not symmetrical about its main diagonal, the presence of the feedback being indicated by the off-diagonal term g_m. Thus

$$\frac{v_o}{i_i} = \frac{-(g_m - Y_F)}{Y_1(Y_2 + Y_F) + Y_2 Y_F + g_m Y_F}. \tag{10.11}$$

Aside from the slightly more complicated determinant, there is no great conceptual difficulty in the application of formal network theory to feedback networks. Difficulties do arise, however, when attempts are made to replace the algebraic determinant solution with the direct calculations that are more suitable for numerical problems. For example, the transfer impedance of the passive network in Fig. 10.4*a* would normally be written down directly from the division of input current between Y_1 and $(Y_2 + Y_F)$: the current flowing in Y_F and Y_2 is

$$\left[\frac{Y_2 Y_F/(Y_2 + Y_F)}{Y_1 + Y_2 Y_F/(Y_2 + Y_F)}\right] i_i,$$

from which

$$\frac{v_o}{i_i} = \frac{Y_F}{Y_1(Y_2 + Y_F) + Y_2 Y_F}$$

as before. In general, a direct approach is more convenient in practice because it enables order-of-magnitude simplifications to be made as the calculation proceeds. For example, if Y_1 were much greater than Y_F, the calculation might run

$$\frac{v_i}{i_i} \approx \frac{1}{Y_1},$$

and therefore

$$\frac{v_o}{i_i} = \frac{v_o}{v_i} \times \frac{v_i}{i_i} = \left(\frac{Y_F}{Y_2 + Y_F}\right) \frac{1}{Y_1}.$$

The same direct approach is possible for the nonfeedback active network of Fig. 10.4*b* and, of course, for all of the nonfeedback amplifier circuits considered in preceding chapters, but it is not obvious how such a direct calculation can be performed in the feedback example of Fig. 10.4*c*. A useful approach to feedback should

(i) provide a consistent approach for the analysis of any feedback circuit,
(ii) enable order-of-magnitude approximations to be made early in numerical calculations,
(iii) highlight the factors that are of basic importance, so as to facilitate design of circuits in which the over-all performance depends largely on a few fixed parameters.

At least two commonly used approaches satisfy these requirements:

(i) an extension of the *principle of superposition* to circuits containing controlled sources,
(ii) the method of *signal flow graphs*, which is effectively a topological equivalent of (i).

The choice between methods is largely a matter of personal preference. Although the discussion of feedback circuits in this book is based on the superposition approach, individual readers may profit from the use of signal flow graphs.*

* A particularly good discussion of signal flow graphs is given in: J. G. TRUXAL, *Automatic Feedback Control System Synthesis* (McGraw-Hill, New York, 1955).

10.3 SUPERPOSITION ANALYSIS OF FEEDBACK AMPLIFIERS

Application of the principle of superposition to linear feedback amplifier circuits enables complex problems to be broken up into a number of simpler problems. This is achieved by writing a consistent set of equations that relates the terminal quantities of interest to each other and to a particular interpal reference variable. This reference variable may be anything convenient—a voltage, a current, a controlled source, or a passive element. Proper choice of the reference variable usually results in considerable simplification. Factors that may influence the choice are considered in Section 10.3.4.

Consider, by way of introduction, the linear two-port shown in Fig. 10.5. This two-port contains n node pairs of which only the ports are accessible. The signals at the ports are voltages v_1 and v_2, and currents i_1 and i_2. For definiteness, assume that port 1 is the "input" and as such is connected to a source; port 2 is the "output" and is connected to a load. Imbedded in the network is a controlled source and, again for definiteness, assume that the output current i_k of the source is controlled by the voltage v_j. As reference variable take the mutual conductance g_m of the source:

$$i_k = g_m v_j. \tag{10.12}$$

Notice the reference direction for i_k. The network equations can be written as

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1j} & \cdots & y_{1n} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2j} & \cdots & y_{2n} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3j} & \cdots & y_{3n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ & & & & & & \\ y_{k1} & y_{k2} & y_{k3} & \cdots & y_{kj} & \cdots & y_{kn} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ & & & & & & \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nj} & \cdots & y_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ \\ v_j \\ \vdots \\ \\ v_n \end{pmatrix} = \begin{pmatrix} i_1 \\ i_2 \\ 0 \\ \vdots \\ 0 \\ -i_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} i_S - Y_S v_1 \\ -Y_L v_2 \\ 0 \\ \vdots \\ 0 \\ -g_m v_j \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{10.13}$$

The determinant of the left-hand side of Eqs. 10.13 depends on the two-port alone, and is independent of Y_S, Y_L, and the reference variable g_m. Equations 10.13 can be solved to find the ratio of any two terminal quantities.

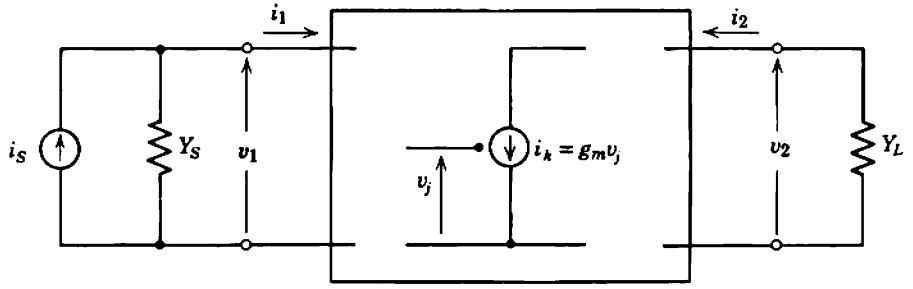


Fig. 10.5 General linear active network.

10.3.1 Voltage Gain

If for example the voltage gain v_2/v_1 is to be found, it is convenient to transfer the load admittance to the left-hand side of the equations:

$$\begin{pmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1j} & \cdots & y_{1n} \\ y_{21} & y_{22}+Y_L & y_{23} & \cdots & y_{2j} & \cdots & y_{2n} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3j} & \cdots & y_{3n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{k1} & y_{k2} & y_{k3} & \cdots & y_{kj} & \cdots & y_{kn} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nj} & \cdots & y_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} i_1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ -i_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (10.14)$$

The solutions for v_1, v_2, and v_j are

$$v_1 = \frac{\Delta'_{11}}{\Delta'} i_1 - \frac{\Delta'_{k1}}{\Delta'} i_k, \quad (10.15a)$$

$$v_2 = \frac{\Delta'_{12}}{\Delta'} i_1 - \frac{\Delta'_{k2}}{\Delta'} i_k, \quad (10.15b)$$

$$v_j = \frac{\Delta'_{1j}}{\Delta'} i_1 - \frac{\Delta'_{kj}}{\Delta'} i_k, \quad (10.15c)$$

where Δ' is the network determinant and Δ'_{pq} are the relevant cofactors. The primes indicate that the determinants are not of the two-port alone, but of the network comprising the two-port plus Y_L.

Elimination of i_1 between Eqs. 10.15 yields

$$v_2 = \left(\frac{\Delta'_{12}}{\Delta'_{11}}\right)v_1 + \left(\frac{\Delta'_{12}\,\Delta'_{k1}}{\Delta'\,\Delta'_{11}} - \frac{\Delta'_{k2}}{\Delta'}\right)i_k, \tag{10.16a}$$

$$v_j = \left(\frac{\Delta'_{1j}}{\Delta'_{11}}\right)v_1 + \left(\frac{\Delta'_{1j}\,\Delta'_{k1}}{\Delta'\,\Delta'_{11}} - \frac{\Delta'_{kj}}{\Delta'}\right)i_k. \tag{10.16b}$$

In conjunction with Eq. 10.12 these equations are of the form

$$v_2 = a_{12}v_1 + z_{k2}i_k, \tag{10.17a}$$

$$v_j = a_{1j}v_1 + z_{kj}i_k, \tag{10.17b}$$

$$i_k = g_m v_j, \tag{10.17c}$$

and it follows that

$$\frac{v_2}{v_1} = a_{12} + \frac{g_m z_{k2} a_{1j}}{1 - g_m z_{kj}}. \tag{10.18}$$

From the principle of superposition, each of the parameters a_{12}, a_{1j}, z_{k2}, and z_{kj} has a simple physical interpretation:

$$a_{12} = \left.\frac{v_2}{v_1}\right|_{i_k = 0}, \tag{10.19a}$$

$$z_{k2} = \left.\frac{v_2}{i_k}\right|_{v_1 = 0}, \tag{10.19b}$$

$$a_{1j} = \left.\frac{v_j}{v_1}\right|_{i_k = 0}, \tag{10.19c}$$

$$z_{kj} = \left.\frac{v_j}{i_k}\right|_{v_1 = 0}. \tag{10.19d}$$

The four subsidiary calculations involved in determining these factors are usually simple because they are carried out on reduced forms of the network. There is no need to evaluate the determinants in Eqs. 10.16.

10.3.1.1 Example of Superposition Calculation

As an example of the method, consider the calculation of the mid-band voltage gain for the pentode feedback amplifier discussed in Section 10.1. The small-signal equivalent circuit is shown in Fig. 10.6*a*; the voltage that controls the source current (v_j in the general development above) is the grid-cathode voltage v_{GK}.

Figure 10.6*b* shows the reduced form of the circuit for the source current i_k set to zero. The first two subsidiary calculations are so trivial that a_{12}

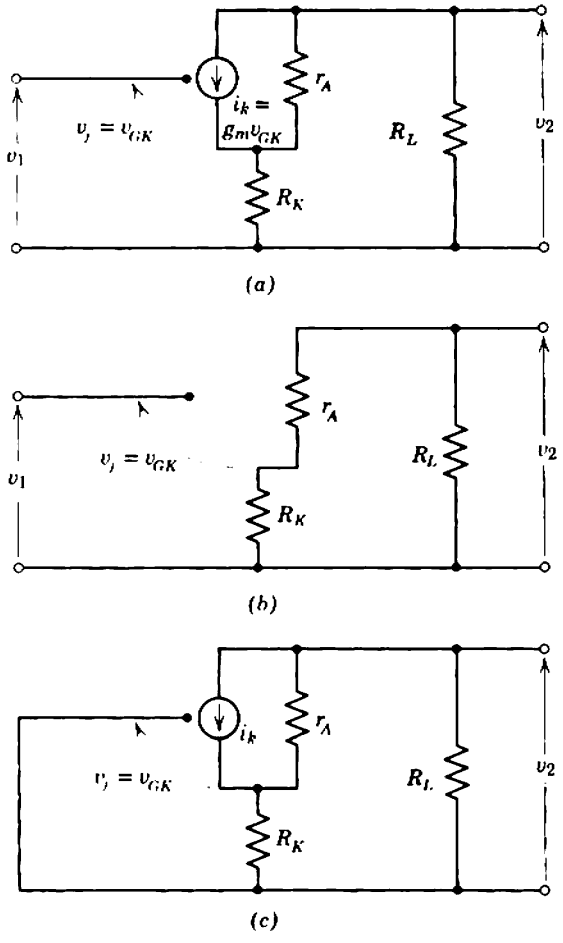


Fig. 10.6 Superposition calculation on Fig. 10.2*b*: (*a*) complete equivalent circuit; (*b*) reduced form, $i_k = 0$; (*c*) reduced form, $v_1 = 0$.

and a_{1j} can be written down by inspection as

$$a_{12} = \left.\frac{v_2}{v_1}\right|_{i_k = 0} = 0$$

and

$$a_{1j} = \left.\frac{v_{GK}}{v_1}\right|_{i_k = 0} = 1.$$

Figure 10.6*c* shows the reduced form of the circuit for the input voltage v_1 set to zero. Here, the calculations are somewhat less trivial, but the

results can still be written down directly as

$$z_{k2} = \left.\frac{v_2}{i_k}\right|_{v_1=0} = \frac{-r_A R_L}{R_K + R_L + r_A}$$

and

$$z_{kj} = \left.\frac{v_{GK}}{i_k}\right|_{v_1=0} = \frac{-r_A R_K}{R_K + R_L + r_A}.$$

The over-all voltage gain follows from Eq. 10.18 as

$$\frac{v_2}{v_1} = -\frac{g_m\left(\dfrac{r_A R_L}{R_K + R_L + r_A}\right)}{1 + g_m\left(\dfrac{r_A R_K}{R_K + R_L + r_A}\right)}. \tag{10.20}$$

Simple algebraic reduction yields

$$\frac{v_2}{v_1} = -g_m R_L\left[\frac{r_A}{r_A + R_L + (\mu + 1)R_K}\right], \tag{10.21}$$

which is the result assumed in Section 10.1 (Eq. 10.5). This result can, of course, be obtained from a conventional node or loop analysis. Indeed, the formal result (Eq. 10.20) is not particularly compact and further algebra is required for the reduction to Eq. 10.21. This is generally true when exact literal expressions are obtained, but in practice analytical simplifications result from order-of-magnitude approximations or the use of numerical values in the subsidiary calculations.

10.3.2 Other Transfer and Driving-Point Functions

In the formal derivation of other transfer and driving-point functions, the development runs parallel as far as Eq. 10.14. The equations corresponding to Eq. 10.17 and their solution corresponding to Eq. 10.18 are set out below for current gain, transfer admittance, and transfer impedance. It is left as an exercise for the reader to obtain the equations for the driving-point functions.

1. CURRENT GAIN. Evaluate coefficients of

$$i_2 = a_{12} i_1 + a_{k2} i_k, \tag{10.22a}$$

$$v_j = z_{1j} i_1 + z_{kj} i_k, \tag{10.22b}$$

$$i_k = g_m v_j, \tag{10.22c}$$

so that

$$\frac{i_2}{i_1} = a_{12} + \frac{g_m a_{k2} z_{1j}}{1 - g_m z_{kj}}. \tag{10.23}$$

2. TRANSFER ADMITTANCE. Evaluate coefficients of

$$i_2 = y_{12}v_1 + a_{k2}i_k, \tag{10.24a}$$

$$v_j = a_{1j}v_1 + z_{kj}i_k, \tag{10.24b}$$

$$i_k = g_m v_j, \tag{10.24c}$$

so that

$$\frac{i_2}{v_1} = y_{12} + \frac{g_m a_{k2} a_{1j}}{1 - g_m z_{kj}}. \tag{10.25}$$

3. TRANSFER IMPEDANCE. Evaluate coefficients of

$$v_2 = z_{12}i_1 + z_{k2}i_k, \tag{10.26a}$$

$$v_j = z_{1j}i_1 + z_{kj}i_k, \tag{10.26b}$$

$$i_k = g_m v_j, \tag{10.26c}$$

so that

$$\frac{v_2}{i_1} = z_{12} + \frac{g_m z_{k2} z_{1j}}{1 - g_m z_{kj}}. \tag{10.27}$$

The superposition principle is not restricted to functions relating voltage and current; for example, input charge is commonly encountered in nucleonic applications—see Section 8.7.

A definite pattern exists for all transfer and driving-point functions, and the foregoing results can be generalized by writing the closed-loop transfer function as TF (or the driving-point function as DF) and the parameter chosen for the reference variable as g. Then

$$TF = TF_D + \frac{gac}{1 - gb}, \tag{10.28}$$

where TF_D, a, b, and c are the superposition factors which, in general, are functions of complex frequency, and have the significance implied in:

(system output) = TF_D (system input) + c (reference variable output),
(reference variable input) = a (system input) + b (reference variable output),
(reference variable output) = g (reference variable input).

Four definitions can be formulated from Eq. 10.28:

1. TF_D is the value of TF when the reference variable g vanishes, and is called the *direct component* of TF.

2. gac is the value of TF when b vanishes, and is called the *open-loop transfer function* TF_0. Physically, TF_0 is the value of $(TF - TF_D)$ when

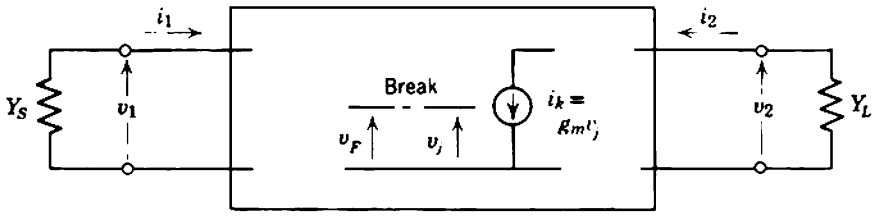


Fig. 10.7 Interpretation of the feedback voltage v_F for the reference variable g_m.

the reference variable is controlled only by an input quantity, that is, when no signal is fed back from the output of the reference variable.

3. b is called the *feedback factor for the reference variable* g. (The symbol β introduced earlier is a special case of b.) Physically, b is the ratio of the signal fed back from the output of the reference variable to its controlling terminals, to the output of the reference variable. Obviously, this interpretation is meaningful only if the circuit is broken at some point. Figure 10.7 illustrates the interpretation of b for the case of g_m as reference variable:

$$b = \frac{v_F}{i_k}.$$

4. gb is a dimensionless quantity called the *loop gain for the reference variable* g. It follows from the defining equations for b that gb is the ratio of the signal fed back to the controlling terminals for a given output of the reference variable, to the controlling signal actually required to produce the same output. As in the case of b, this interpretation of gb is meaningful only when the controlling signal and feedback signal are separated by a break in the equivalent circuit. In Fig. 10.7, the loop gain referred to g_m is

$$gb = \frac{v_F}{v_j}.$$

10.3.3 Formal Definition of Loop Gain

The loop gain for any reference variable g is a most important quantity in feedback amplifier design. It is shown in Section 10.4 that loop gain determines the stability of the closed-loop transfer function against changes in the reference variable, and in Section 10.5 that there is a relation between loop gain and the closed-loop dynamic response. For these reasons it is desirable to formulate a general definition of loop gain in terms of the circuit determinants. Practical definitions of loop gain for use in amplifier design are discussed in Sections 10.4 and 10.6.

Equations 10.13 can be rewritten in two different forms. First, all elements are transferred to the left-hand side:

$$\begin{pmatrix} y_{11}+Y_S & y_{12} & y_{13} & \cdots & y_{1j} & \cdots & y_{1n} \\ y_{21} & y_{22}+Y_L & y_{23} & \cdots & y_{2j} & \cdots & y_{2n} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3j} & \cdots & y_{3n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{k1} & y_{k2} & y_{k3} & \cdots & y_{kj}+g_m & \cdots & y_{kn} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nj} & \cdots & y_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} i_S \\ 0 \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{10.29}$$

The determinant of the left-hand side is designated by Δ, and is the determinant of the complete system—the two-port, source, load, and reference variable g_m. The relation between Δ and Δ' is

$$\Delta' = \Delta - Y_S\Delta_{11} - g_m\Delta_{kj}.$$

Second, all elements except the reference variable g_m are transferred to the left-hand side of Eqs. 10.13:

$$\begin{pmatrix} y_{11}+Y_S & y_{12} & y_{13} & \cdots & y_{1j} & \cdots & y_{1n} \\ y_{21} & y_{22}+Y_L & y_{23} & \cdots & y_{2j} & \cdots & y_{2n} \\ y_{31} & y_{32} & y_{33} & \cdots & y_{3j} & \cdots & y_{3n} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{k1} & y_{k2} & y_{k3} & \cdots & y_{kj} & \cdots & y_{kn} \\ \vdots & \vdots & \vdots & & \vdots & & \vdots \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nj} & \cdots & y_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} i_S \\ 0 \\ 0 \\ \vdots \\ 0 \\ -i_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{10.30}$$

The determinant of the left-hand side is designated by Δ^0, and is the determinant of the complete system with the reference variable set to zero. The value of Δ^0 is

$$\Delta^0 = \Delta - g_m\Delta_{kj}.$$

However, $\Delta_{kj} = \Delta^0_{kj}$, and therefore

$$\Delta = \Delta^0 + g_m\Delta^0_{kj}$$

or, in general notation with g as reference variable,

$$\Delta = \Delta^0 + g\Delta^0_{kj}. \tag{10.31}$$

The feedback factor b is the ratio of the signal fed back from the output of the reference variable to its controlling terminals, to the output of the reference variable:

$$b = \left.\frac{v_j}{i_k}\right|_{i_S = 0},$$

and from Eq. 10.30

$$b = -\frac{\Delta^0_{kj}}{\Delta^0}. \tag{10.32}$$

Multiplying by the reference variable yields the loop gain

$$gb = -g\,\frac{\Delta^0_{kj}}{\Delta^0},$$

and using Eq. 10.31

$$gb = 1 - \frac{\Delta}{\Delta^0}. \tag{10.33a}$$

This result is perfectly general, and is the formal definition of loop gain:

$$\begin{pmatrix}\text{loop gain for any}\\ \text{reference variable}\end{pmatrix} = 1 - \frac{\begin{pmatrix}\text{complete circuit determinant,}\\ \text{including source and load}\end{pmatrix}}{\begin{pmatrix}\text{circuit determinant with reference}\\ \text{variable set at zero}\end{pmatrix}}. \tag{10.33b}$$

Notice that loop gain depends on the entire system, including the source and load.

10.3.4 Choice of Reference Variable

The aim in choosing a reference variable is to simplify the writing down of the superposition constants. For single-stage feedback amplifiers that use charge-controlled devices, the most appropriate choice is usually the mutual conductance g_m (as in the preceding development), since the removal of the only controlled source from the equivalent circuit results in considerable simplification. Other active circuit parameters, however, may be used as the reference variable. It is often appropriate to choose β_N as the reference variable in transistor circuits. In multistage feedback amplifiers, convenient simplifications result if one of the open-loop transfer functions of the amplifier is chosen; this procedure is discussed in Section 10.6.

Passive elements too may be chosen as the reference variable; for example, if a resistance R is taken as the reference variable, it is possible to take the voltage as the output quantity and the current as the controlling quantity and write

$$v_2 = a_{12}v_1 + a_{R2}v_R, \tag{10.34a}$$

$$i_R = y_{1R}v_1 + y_{RR}v_R, \tag{10.34b}$$

$$v_R = Ri_R. \tag{10.34c}$$

In some circuits a substantial simplification is achieved when v_R vanishes; perhaps the most notable example is when R is a feedback resistor in a cathode or emitter circuit.

In some complicated networks no single choice of reference variable results in the necessary simplification. For such cases two approaches are possible. First, the reduced networks can themselves be tackled by further use of superposition. Second, it may be convenient to use two reference variables such as g_{m1} and g_{m2} in the following sample set:

$$v_2 = a_{12}v_1 + z_{k_1 2}i_{k_1} + z_{k_2 2}i_{k_2}, \tag{10.35a}$$

$$v_{j_1} = a_{1j_1}v_1 + z_{k_1 j_1}i_{k_1} + z_{k_2 j_1}i_{k_2}, \tag{10.35b}$$

$$v_{j_2} = a_{1j_2}v_1 + z_{k_1 j_2}i_{k_1} + z_{k_2 j_2}i_{k_2}, \tag{10.35c}$$

$$i_{k_1} = g_{m1}v_{j_1}, \tag{10.35d}$$

$$i_{k_2} = g_{m2}v_{j_2}. \tag{10.35e}$$

Such a set can be solved for the required transfer function.

There are no hard and fast rules for the optimum choice of reference variable and there is no real substitute for experience. However, the possibility of analytical difficulties is remote even if the choice is not optimum, because designable circuits must have simple topology and the superposition calculations cannot become very complicated.

10.4 LOOP GAIN

The primary aim in using negative feedback is to stabilize the mid-band transfer function of an amplifier so that it depends substantially on a passive feedback network alone. It is not obvious from the form of Eq. 10.28 that this is accomplished. Invariably, however, a reference variable exists that satisfies all the following conditions:

(i) The reference variable includes all active devices and other uncertain parameters; it is customary to represent this special reference variable by μ.

(ii) The corresponding feedback factor is the transfer function of the passive feedback network in isolation; it is customary to represent this special feedback factor by β.
(iii) $ac = 1$.
(iv) $TF_D = 0$.

For this particular reference variable, the general Eq. 10.28 reduces to

$$TF = \frac{g}{1 - gb};$$

that is,

$$TF = \frac{\mu}{1 - \mu\beta} = -\frac{1}{\beta}\left(\frac{1}{1 - 1/A_l}\right), \tag{10.36a}$$

where

$$A_l = \mu\beta. \tag{10.37}$$

is the loop gain for μ. As A_l is made increasingly negative, TF approaches $(-1/\beta)$ and the uncertainty in its value due to the uncertainty in the reference variable μ becomes small. In the limit of infinite loop-gain magnitude,

$$\lim_{|A_l| \to \infty} \{TF\} = -\frac{1}{\beta}, \tag{10.38}$$

which is quite independent of the value of μ. Thus, the feedback factor β takes on the significance of the closed-loop transfer function demanded under ideal circumstances. The same behavior is expected for large positive values of A_l, but these are excluded for reasons given in Section 10.5.2.1. For positive values of A_l in the vicinity of $+1$, TF becomes very large and critically dependent on the value of μ.

The special reference variable μ can be evaluated by rearranging Eq. 10.28. Suppose a circuit has been analyzed using any convenient reference variable g. Extracting the known feedback factor $(-1/\beta)$ from the right-hand side of Eq. 10.28 and rearranging the remaining terms gives

$$TF = \begin{pmatrix}\text{demanded transfer} \\ \text{function, } -1/\beta\end{pmatrix}\left(\frac{1}{1 - 1/(\text{loop gain, } \mu\beta)}\right), \tag{10.36b}$$

where the quantity defined as the loop gain can be evaluated. Since β is known, μ can be found. Further rearrangement gives

$$\frac{1}{\text{loop gain}} = 1 - \frac{\text{demanded transfer function}}{\text{actual closed-loop transfer function}}. \tag{10.39}$$

From an analytical viewpoint, loop gain can be referred to a number of arbitrarily chosen reference variables, and there has been much speculation

in technical literature as to which is the most basic. Many and varied arguments have been put forward but, although such contributions may be of philosophical interest, they are of little use in circuit design. A feedback network whose ideal transmission β can be evaluated exists in any feedback amplifier, and from a design viewpoint loop gain takes on the significance given in Eqs. 10.36. Equations 10.36 are among the most basic in feedback theory. There is, of course, no compulsion to use μ as reference variable when analyzing a circuit; any convenient parameter may be used, and the resulting equation (of the form of Eq. 10.28) can be manipulated into the form of Eq. 10.36.

The quantity $(1 - A_l)$ occurs in many expressions, and is called the *return difference* F:

$$F = 1 - A_l. \tag{10.40}$$

The reader should be warned that the two conventions illustrated in Fig. 10.1 exist for the sign of loop gain. For both conventions the feedback is considered as negative if $|TF|$ is less than $|\mu|$, that is, if F is greater than unity. In this and most other books on electronics, the sign convention follows Fig. 10.1*a* for which negative feedback corresponds to a negative value of loop gain; this emphasizes the fact that there must be a net phase reversal around the feedback loop. Control engineers, on the other hand, are most conscious of the error between the input and feedback signals, and therefore assume a subtractor at the input. This subtractor contributes the loop phase reversal, and as a result the loop gain (which does not include the subtractor) must be positive for negative feedback. Equations 10.36 and 10.40 become

$$TF = \frac{\mu}{1 + \mu\beta} = \frac{1}{\beta}\left(\frac{1}{1 + 1/A_l}\right),$$

$$F = 1 + \mu\beta = 1 + A_l.$$

10.4.1 Example of Change in Reference Variable

As an example in the use of one reference variable to analyze a circuit followed by an algebraic manipulation into the useful form of Eq. 10.36, consider the transistor shunt-feedback stage shown in Fig. 10.8. This circuit is discussed extensively in Section 11.1.4. For simplicity r_C and $\beta_N r_C$ are omitted from the transistor equivalent circuit. The mutual conductance

$$g_m = \frac{\alpha_N}{r_E}$$

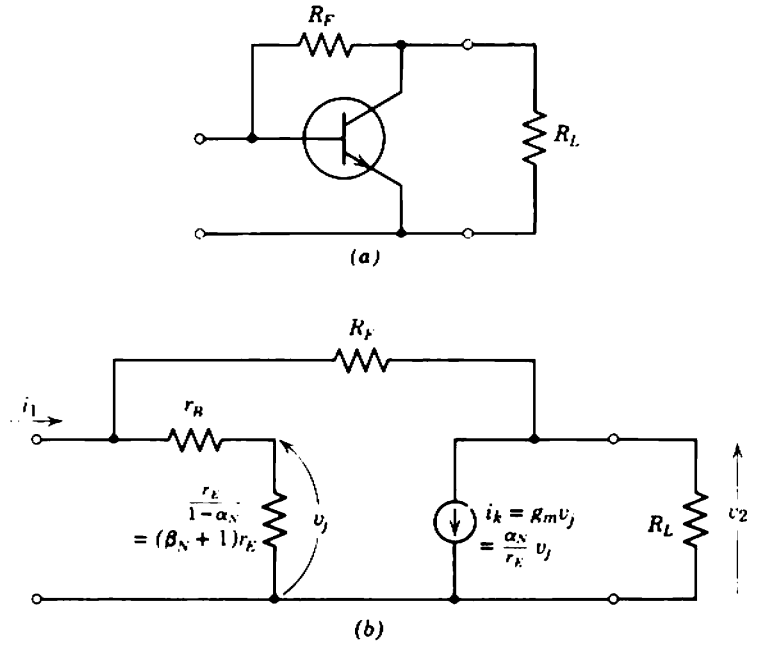


Fig. 10.8 Transistor shunt-feedback stage.

is taken as the reference variable, and the superposition equations for transfer resistance R_T become

$$v_2 = r_{12}i_1 + r_{k2}i_k, \tag{10.41a}$$

$$v_j = r_{1j}i_1 + r_{kj}i_k, \tag{10.41b}$$

$$i_k = g_m v_j = \frac{\alpha_N}{r_E} v_j, \tag{10.41c}$$

for which

$$R_T = \frac{v_2}{i_1} = r_{12} + \frac{g_m r_{k2} r_{1j}}{1 - g_m r_{kj}}. \tag{10.42}$$

The superposition parameters can be written down as

$$r_{12} = \left.\frac{v_2}{i_1}\right|_{i_k = 0} = \frac{[r_B + (\beta_N + 1)r_E]R_L}{r_B + (\beta_N + 1)r_E + R_F + R_L},$$

$$r_{k2} = \left.\frac{v_2}{i_k}\right|_{i_1 = 0} = -\frac{[r_B + (\beta_N + 1)r_E + R_F]R_L}{r_B + (\beta_N + 1)r_E + R_F + R_L},$$

$$r_{1j} = \left.\frac{v_j}{i_1}\right|_{i_k = 0} = \frac{(\beta_N + 1)r_E(R_F + R_L)}{r_B + (\beta_N + 1)r_E + R_F + R_L},$$

$$r_{kj} = \left.\frac{v_j}{i_k}\right|_{i_1 = 0} = -\frac{(\beta_N + 1)r_E R_L}{r_B + (\beta_N + 1)r_E + R_F + R_L}.$$

Direct substitution gives an unwieldy expression for the transfer resistance in the form of Eq. 10.28. Manipulation gives the preferred form

$$R_T = -R_F\left(\frac{1}{1 - 1/A_l}\right) \tag{10.43}$$

where $(-R_F)$ is the transfer resistance expected under ideal circumstances, and the loop gain is

$$A_l = -\left\{\frac{\beta_N R_F R_L}{(R_F + R_L)[R_F + r_B + (\beta_N + 1)r_E]}\right\}\left[1 - \frac{r_B(1 - \alpha_N) + r_E}{\alpha_N R_F}\right]. \tag{10.44}$$

The last term in Eq. 10.44 is the ratio of the direct transfer conductance through the feedback resistor $(= -1/R_F)$, to G_T of the transistor $\{= \alpha_N/[r_B(1 - \alpha_N) + r_E]\}$. Invariably the latter dominates; if this were not so, there would be no point in using an active device. Therefore Eq. 10.43 reduces to

$$R_T = -\frac{R_F}{1 + \dfrac{(R_F + R_L)[R_F + r_B + (\beta + 1)r_E]}{\beta_N R_F R_L}}. \tag{10.45}$$

10.4.2 Generalizations

Three important generalizations follow from the preceding discussion.

First, it is seen from Eq. 10.38 that the feedback factor β has dimensions inverse to those of the closed-loop transfer function *TF*. For example, if the transfer impedance of an amplifier is to be stabilized by means of feedback, β must have the dimensions of admittance. In the above example β is the ideal ratio of feedback current to output voltage, namely, $1/R_F$.

Second, the reference variable μ is recognized as an open-loop transfer function of the complete forward path. Moreover, it is seen from Eqs. 10.37 and 10.38 that μ must have the same dimensions as *TF*; only under these circumstances is the loop gain a dimensionless quantity which can be added to unity. Thus, if transfer impedance is to be stabilized, μ must be an open-loop transfer impedance.

Third, if μ and β can be evaluated simply, a feedback amplifier can be analyzed without recourse to superposition or any other formal technique. Although β can usually be written down at sight, μ, unfortunately, cannot. *The reference variable μ for a feedback amplifier is not simply the transfer function of the amplifier with the feedback network removed.* Rather, it is the forward transmission of the complete system, and must include all

loading and feedthrough effects associated with the feedback network β. In general, the feedback network cannot be separated from the forward path to determine the reference variable, and this introduces considerable complication.

As an example, the feedback resistor in the transistor shunt-feedback stage constitutes a load on both the input and the output circuits, and provides a path for direct signal feedthrough. The open-loop transfer resistance μ is therefore not R_T derived in Chapter 5, namely, $-\beta_N R_L$. As a matter of interest, the value of μ in this example can be deduced from the known values of A_l and β as

$$\mu = \frac{A_l}{\beta} = -\left\{\frac{\beta_N R_F R_L}{(R_F + R_L)[R_F + r_B + (\beta_N + 1)r_E]}\right\}\left[R_F - \frac{r_B(1 - \alpha_N) + r_E}{\alpha_N}\right].$$

Such an expression can hardly be described as an obvious choice for reference variable!

The practical conclusion is that, if it is desired to build a designable feedback amplifier for which the closed-loop transfer function can be written down easily, the circuit must be designed so that the feedback network β is separable from the forward path μ. Section 10.6.1 lists the necessary conditions. It is assumed throughout the remainder of this chapter that μ is known, either by direct calculation or from A_l and β as above.

10.4.3 Gain Stability

Feedback is usually employed to reduce the effect of fluctuations in the reference variable μ on the closed-loop transfer function TF; μ is defined in such a way that it includes all unpredictable parameters. If a fluctuation $\Delta\mu$ produces a fluctuation ΔTF, the *sensitivity* of TF with respect to μ is defined as

$$S_\mu = \frac{\Delta TF/TF}{\Delta\mu/\mu}. \tag{10.46}$$

If the changes in μ and TF are small, differentiation of Eq. 10.36 yields

$$S_\mu = \frac{1}{1 - A_l} = \frac{1}{F} \tag{10.47}$$

Thus, if there is a change of 5% in the open-loop transfer function and F is 10, the change in closed-loop transfer function is about 0.5%. Sensitivity falls as $|A_l|$ is increased.

It must be emphasized that Eq. 10.47 is valid only for small changes. However, a typical tolerance on the open-loop transfer function of a single-stage amplifier is $\pm 20\%$; much larger ranges occur with multistage

amplifiers. Therefore, some other relation between changes in μ and TF must be obtained. In general, it is not meaningful to measure the gain stability of a feedback amplifier by the sensitivity defined in Eq. 10.46, because the ratio between a finite change in μ and the resulting change in TF depends on the magnitude of the change in μ. Equations 10.48 (in the next paragraph) and 10.78 (in Section 10.8) give useful measures of the stabilizing effect of feedback.

Suppose that the nominal (average) values of μ and TF are μ_A and TF_A, and that the actual values are

$$\mu = m\,\mu_A,$$
$$TF = M\,TF_A.$$

The multiplying factors m and M are often expressed as percentages. Then from Eq. 10.36

$$M\,TF_A = -\frac{1}{\beta}\left[\frac{1}{1 - 1/(\beta m \mu_A)}\right],$$

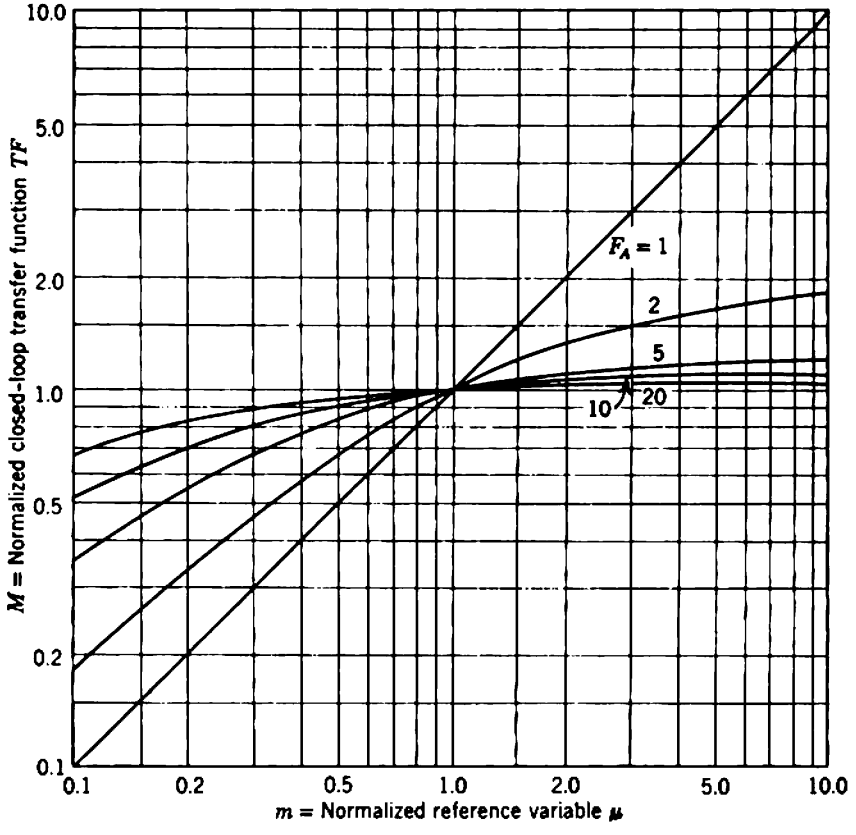


Fig. 10.9 Gain stability of feedback amplifiers.

which leads to

$$M = \frac{F_A}{F_A + \left(\frac{1-m}{m}\right)}, \tag{10.48}$$

where F_A is the nominal return difference:

$$F_A = 1 - \mu_A\beta.$$

Equation 10.48 is plotted in Fig. 10.9 for various values of F_A. Notice that, if $F_A > 1$, M remains finite even if m approaches infinity, although M falls to zero when m falls to zero. As an example, if F_A is 10 and the open-loop transfer function can vary from half its nominal value to infinitely large, then

$$50\% \leq m \leq \infty$$

and it follows that

$$91\% \leq M \leq 111\%.$$

Thus the closed-loop transfer function can vary roughly $\pm 10\%$.

10.5 DYNAMIC RESPONSE

The preceding discussion is tied somewhat to considerations of the midband performance, and to the great advantage of reduced sensitivity that feedback offers. However, the addition of a feedback path to an amplifier has a profound effect upon its dynamic response. Unless the proper steps are taken in designing the amplifier, the decrease in sensitivity may be completely negated by a deterioration in the dynamic response; all too often a feedback amplifier exhibits gross peaks in its frequency response and ringing in its transient response, or even breaks into self-sustained oscillation. Bode* makes the following comment:

"The engineer who embarks upon the design of a feedback amplifier must be a creature of mixed emotions. On the one hand, he can rejoice in the improvements in the characteristics of the structure which feedback promises to secure him. On the other hand, he knows that unless he can finally adjust the phase and attenuation characteristics around the feedback loop so the amplifier will not spontaneously burst into uncontrollable [oscillation], none of these advantages can actually be realized. The emotional situation is much like that of an impecunious young man who

* H. W. Bode, "Relations between attenuation and phase in feedback amplifier design," *Bell System Tech. J.*, **19**, 421, July 1940. (Reproduced by courtesy of American Telephone and Telegraph Company.)

has impetuously invited the lady of his heart to see a play, unmindful, for the moment, of the limitations of the $2.65 in his pockets. The rapturous comments of the girl on the way to the theatre would be very pleasant if they were not shadowed by his private speculation about the cost of the tickets."

Regeneration* (or oscillation) is an extreme case of undesired dynamic response which must obviously be avoided; useful regeneration criteria are given in Sections 10.5.1 and 10.5.2. The mere avoidance of regeneration, however, does not guarantee a satisfactory dynamic response; Sections 10.5.2.4 and 10.5.3.4 suggest design methods for achieving desired singularity patterns. The treatment is necessarily brief and lacks rigor, but readers may overcome these deficiencies by studying any of the excellent books available on feedback theory.

10.5.1 Analytical Methods

Equation 10.36 is the starting point for determining the dynamic response of a feedback amplifier. It is pointed out in Section 10.3.2 that the general superposition constants TF_D, g, a, b, and c may be functions of complex frequency, and correspondingly the special functions μ and β may also be functions of s. Thus

$$TF(s) = \frac{\mu(s)}{1 - \mu(s)\beta(s)} = -\frac{1}{\beta(s)}\left[\frac{1}{1 - 1/A_l(s)}\right]. \tag{10.49}$$

The functions μ and β can be expressed as quotients of polynomials in s, with a constant multiplying factor. In the notation of Section 7.3.1

$$\mu(s) = \mu_m \psi_\mu(s), \tag{10.50}$$

$$\beta(s) = \beta_m \psi_\beta(s), \tag{10.51}$$

where μ_m and β_m are the mid-band values, and the ψ functions give the poles and the zeros. The loop gain becomes

$$A_l(s) = \mu_m\beta_m \psi_\mu(s) \psi_\beta(s) = A_{lm} \psi_\mu(s) \psi_\beta(s). \tag{10.52}$$

At mid-band frequencies the ψ functions are unity and Eq. 10.49 reduces to

$$TF_m = -\frac{1}{\beta_m}\left(\frac{1}{1 - 1/A_{lm}}\right). \tag{10.53}$$

* Many writers use the term *instability* for regeneration. The authors of this book reserve the concept of stability for the measure of an amplifier's sensitivity to changes in its active devices; a stable amplifier is one whose gain remains substantially constant despite changes in the circuit parameters (other than the feedback network).

Thus A_{lm} determines the stability of TF over the mid-band range of A_l, and A_{lm} should be among the specifications for a feedback amplifier. Outside this frequency range $|A_l|$ and the stability fall. Therefore the bandwidth of A_l (the frequency range over which a given degree of stability must be maintained) may also be specified. Notice that $|TF|$ does not fall by 3 dB until $|A_l|$ has fallen nearly to unity; the bandwidth of TF is much larger than the bandwidth of A_l.

Rearrangement of Eq. 10.49 shows that $TF(s)$ has

zeros at the zeros of $\psi_\mu(s)$,

zeros at the poles of $\psi_\beta(s)$,

poles at the zeros of $[1 - A_{lm}\,\psi_\mu(s)\,\psi_\beta(s)]$.

Thus the zeros of the closed-loop transfer function can be found by inspection; they are readily under the control of the circuit designer and are independent of A_{lm}. In contrast, the poles of the closed-loop transfer function are the solutions of

$$F(s) = 1 - A_{lm}\,\psi_\mu(s)\,\psi_\beta(s) = 0 \qquad (10.54)$$

and they do depend on the value of A_{lm}.

In principle, Eq. 10.54 can be solved to complete the singularity pattern. Alternatively, if a certain closed-loop singularity pattern is desired, it is possible to work back from Eq. 10.54 to the necessary relation between A_{lm}, ψ_μ, and ψ_β. Invariably, however, constraints which reduce the number of independent variables are imposed on this latter procedure; A_{lm} is constrained as above, and ψ_β is made a simple function of s, often unity. Thus the principal variable at the designer's disposal is the frequency response ψ_μ of the forward path. Quantitative discussion of the desirable closed-loop singularity patterns for an amplifier must be deferred until Chapters 12 and 13, but qualitative trends (with which the reader should be familiar) are listed in Section 10.5.2.3. One inviolable rule is that all poles must lie in the left half of the complex frequency plane; poles in the right half correspond to a regenerative (oscillatory) response.

In practice, an analytical approach to the dynamic response based on the solution of Eq. 10.54 is unsatisfactory for two reasons. First, the actual solving of Eq. 10.54 becomes tedious if there are more than two poles. Second, except in trivial cases the approach gives no indication of the effects of changes in A_{lm}, ψ_μ, and ψ_β. Although analysis may show that the dynamic response of a feedback amplifier is intolerable, it does not show how to remedy the situation. A satisfactory approach should be based on some simple relation between the open-loop transfer function

and the closed-loop poles. One such relation is provided by the paths along which the closed-loop poles move as $|A_{lm}|$ varies from zero to infinity, $\psi_\mu(s)$ and $\psi_\beta(s)$ being fixed. These paths are known as the *root loci*, and Section 10.5.3.2 lists rules for constructing them. The root-locus approach has the minor disadvantage that it requires a knowledge of the open-loop transfer function in analytical form, and this is not always available (for example, when there is a transformer in the circuit). *Nyquist's criterion* provides a relation between the empirical, measured open-loop frequency response and the closed-loop response and is useful in such cases. The root-locus construction requires more input data, but gives more information in return and is certainly the more general approach. Nevertheless, Nyquist's criterion is considered first as it establishes clearly the important result quoted in Section 10.4, that A_{lm} must under normal circumstances be negative.

10.5.2 Nyquist's Criterion

In 1932 Nyquist* developed a test for regeneration in terms of the frequency response of the return difference F. This function is either known as an analytic expression or can, in principle, be measured. The test shows if regeneration will occur, and gives the closed-loop frequency response. It does not give the positions of the closed-loop poles or the transient response, although both can be estimated.

Consider the function $F(s)$ whose poles and zeros can be plotted in the complex frequency plane. Suppose some arbitrary contour is drawn into the s plane, enclosing n_p poles and n_z zeros of $F(s)$ in a clockwise direction as shown in Fig. 10.10*a*. Cauchy's integral† shows that if this same contour is mapped into the F plane (Fig. 10.10*b*) it encircles the origin clockwise $(n_z - n_p)$ times. If the $j\omega$ axis of the s plane is chosen for the contour (Fig. 10.10*c*) the number of clockwise encirclements is the difference between the number of zeros and poles of F in the right half of the s plane. But from Eq. 10.54, the poles of F are the poles of either ψ_μ or ψ_β, and these cannot lie in the right half of the s plane or the amplifier is regenerative without feedback. Therefore the number of clockwise encirclements is the number of zeros of F in the right half-plane. From Eq. 10.49 the zeros of F are the poles of the closed-loop transfer function TF. Therefore, if the contour encircles the origin, TF has poles in the right half of the s plane and the amplifier will regenerate.

---

* H. NYQUIST, "Regeneration theory," *Bell System Tech. J.*, **11**, 126, January 1932.

† See for example: E. T. WHITAKER and G. N. WATSON, *Modern Analysis* (4th ed.), 119 (Cambridge University Press, London, 1958).

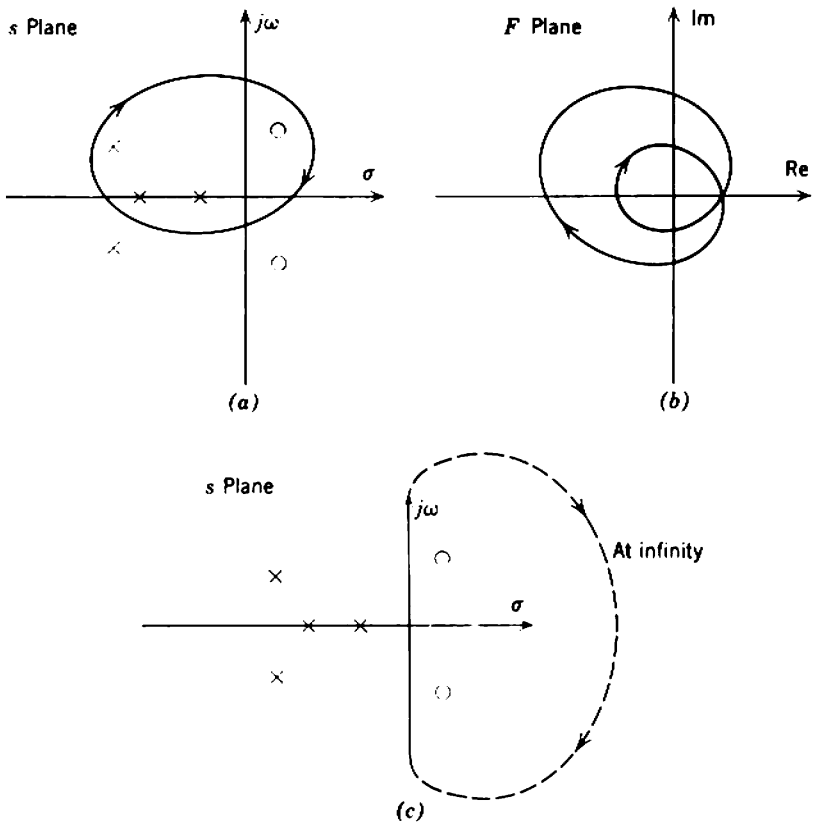


Fig. 10.10 Derivation of Nyquist's stability criterion.

In the F plane this special contour is a polar plot of F as ω runs from $-\infty$ to $+\infty$. Because the singularities of F must lie on the real axis or occur as complex conjugate pairs, the portion of the contour for negative values of ω is the mirror image of the portion for positive ω. Therefore the Nyquist criterion for regeneration can be stated:

> If the polar plot of $F(j\omega)$ plus its mirror image encircles the origin in a clockwise direction as ω runs from zero to infinity, the amplifier will regenerate.

From an experimental viewpoint, return difference is an unfortunate variable for regeneration criteria, as it cannot be measured easily. Loop gain and return difference, however, are related by

$$F(s) = 1 - A_l(s)$$

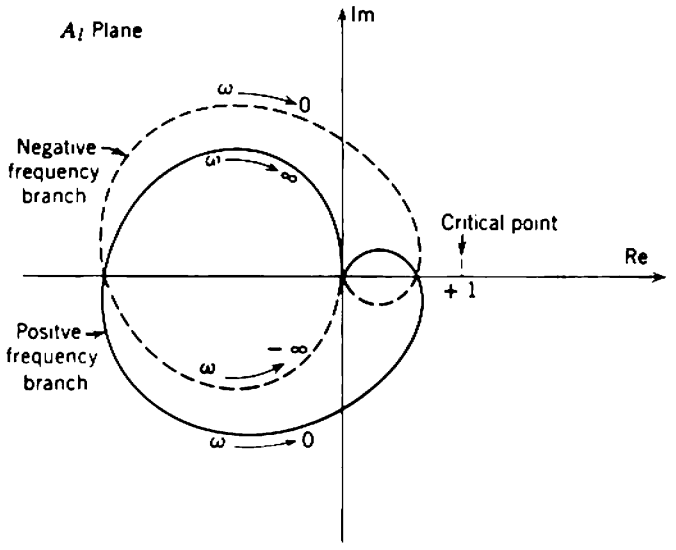


Fig. 10.11 Typical Nyquist diagram. The portion for negative ω is shown broken.

and loop gain can be measured by breaking the feedback loop.* It is therefore convenient to reframe Nyquist's criterion in terms of loop gain:

> If the polar plot of loop gain $A_l(j\omega)$ plus its mirror image encircles the point $(+1, j0)$ in a clockwise direction as ω runs from zero to infinity, the amplifier will regenerate when the feedback loop is closed.

A polar plot of loop gain with ω as parameter is called a *Nyquist diagram*, and $(+1, j0)$ is the *critical* or *Nyquist point*. Figure 10.11 is a typical example.

10.5.2.1 Nature of the Nyquist Diagram

At mid-band frequencies, $\psi_\mu(s)$ and $\psi_\beta(s)$ are unity, and A_l reduces to A_{lm}. The corresponding point on a Nyquist diagram lies on the real axis, as shown in Figs. 10.12 and 10.13 for positive and negative values of A_{lm}, respectively. With decreasing frequency, $|\psi_\mu|$ (and possibly $|\psi_\beta|$) fall to zero and the phase develops a lead; the loop-gain locus spirals counterclockwise toward the origin. Similarly, with increasing frequency the locus spirals clockwise toward the origin. Over-all there is a clockwise rotation as frequency runs from zero to infinity. Often it is monotonically clockwise, although exceptions exist as discussed in connection with Fig. 10.15.

* Problems 10.3, 10.4, and 10.6 discuss ways of measuring A_l.

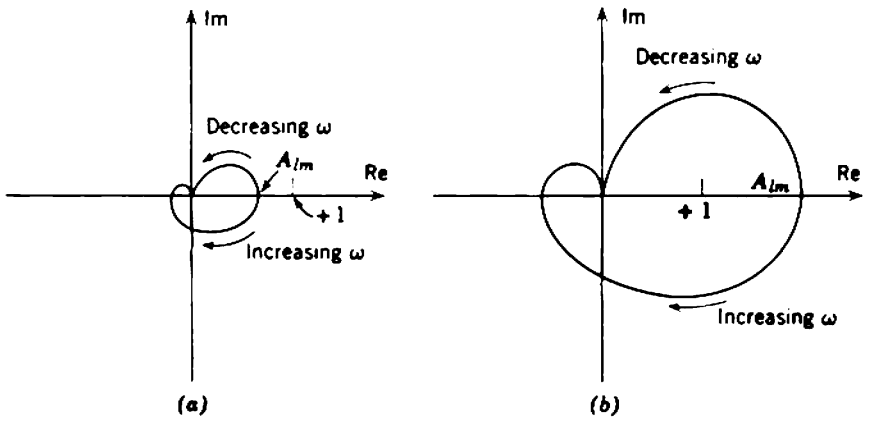


Fig. 10.12 Nyquist diagram for positive A_{lm}: (*a*) $|A_{lm}| < 1$, degenerative; (*b*) $|A_{lm}| > 1$, regenerative. The negative-frequency portion of the curve is omitted.

If A_{lm} is positive but less than unity, the locus will not ordinarily enclose the critical point (Fig. 10.12*a*). If A_{lm} is one or more, however, the critical point is enclosed as in Fig. 10.12*b* and the feedback is regenerative. Thus large positive values of A_{lm} are inadmissible, even though Section 10.4 shows that they would give a stable closed-loop transfer function.

If A_{lm} is negative and $|A_{lm}|$ is greater than unity but not too large, the critical point is not enclosed (Fig. 10.13*a*). However, if $|A_{lm}|$ is made large enough, the locus moves outside the critical point and the feedback

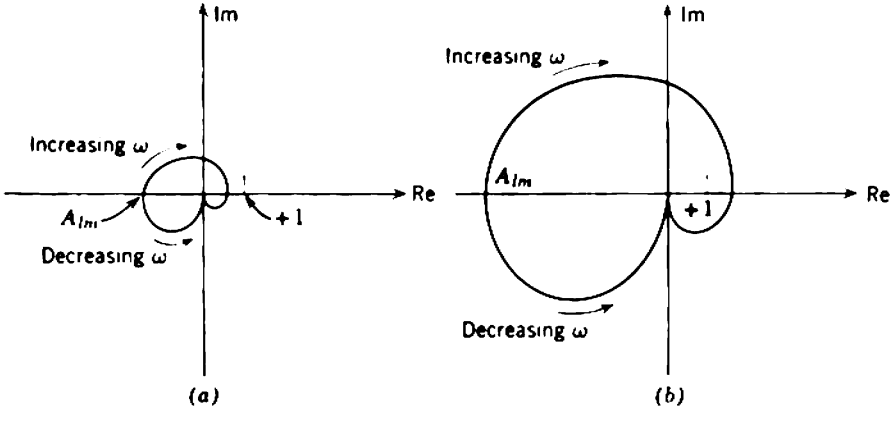


Fig. 10.13 Nyquist diagram for negative A_{lm}: (*a*) $|A_{lm}| > 1$, degenerative; (*b*) $|A_{lm}| \gg 1$, regenerative. The negative-frequency portion of the curve is omitted.

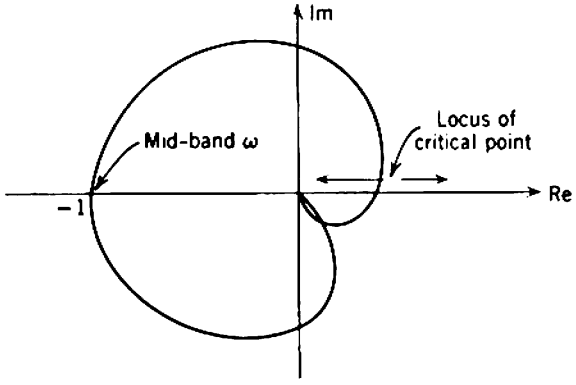


Fig. 10.14 Normalized Nyquist diagram (for $A_{lm} < 0$). The critical point is $1/|A_{lm}|$.

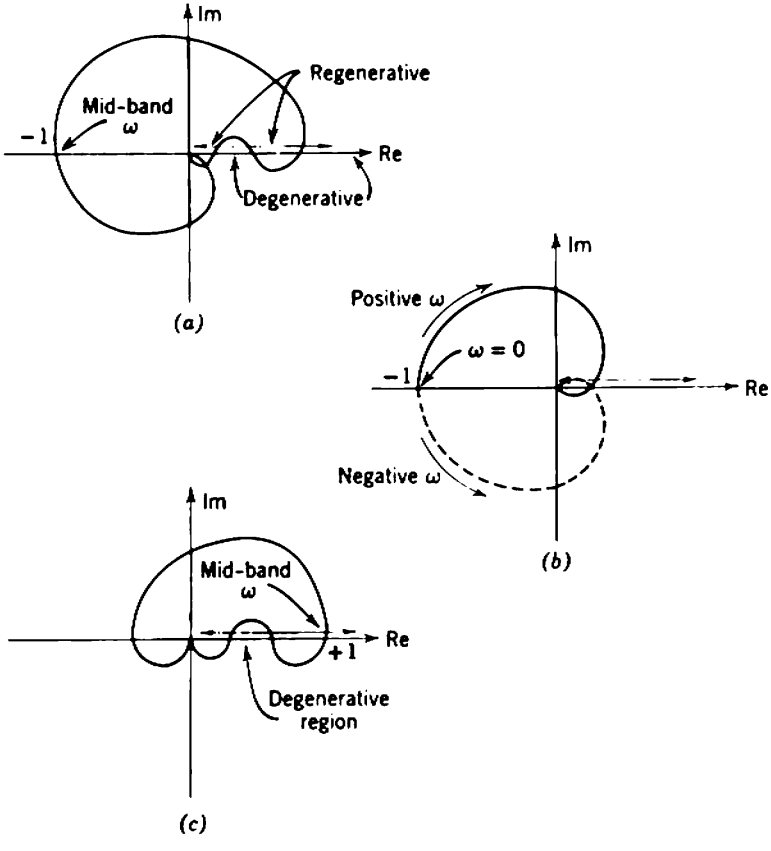


Fig. 10.15 Nyquist diagram oddities (normalized).

changes from degenerative to regenerative. The design problem in feedback amplifiers is to ensure that $|A_{lm}|$ can be made large enough to ensure stability of TF before the onset of regeneration is reached.

Perhaps more useful than the Nyquist diagram is a normalized form in which $|A_{lm}|$ is unity. The critical point becomes $1/A_{lm}$, and its locus lies along the positive real axis as shown in Fig. 10.14. The value of $|A_{lm}|$ for which regeneration occurs can be read directly as the reciprocal of the point at which the loop-gain locus cuts the positive real axis.

Oddities are encountered in Nyquist diagrams if the ψ functions are such that the direction of spiraling goes through a temporary reversal. Figure 10.15*a* shows the normalized Nyquist diagram for an amplifier that goes through the four stages of degenerative, regenerative, degenerative, and finally regenerative as $|A_{lm}|$ is increased. The special dc amplifiers considered in Chapter 15 have no low-frequency poles or zeros, so that the low-frequency portion of the locus coincides with the mid-band point; Fig. 10.15*b* is a typical normalized Nyquist diagram. Finally, some amplifiers are degenerative over a range of A_{lm} positive and greater than unity (Fig. 10.15*c*).

10.5.2.2 Bode Diagrams

For amplifiers in which the mid-band feedback is negative Nyquist's criterion can be restated in a simpler way:

> If the excess phase shift around the feedback loop (that is, the phase angle of $\psi_\mu\psi_\beta$) is less than 180° at the frequency for which $|A_l|$ is unity, the amplifier will not regenerate when the feedback loop is closed.

If this condition is satisfied, the loop-gain locus does not enclose the critical point. Two useful conclusions are the following:

(i) It is not necessary to construct the entire Nyquist diagram; only the region in the vicinity of the critical point need be considered.
(ii) Rather than construct a Nyquist diagram at all, it is sufficient to plot the magnitude and phase of the loop gain; such plots are commonly called *Bode diagrams.*

Normalized Bode diagrams are the most generally useful; often the accuracy of asymptotic plots is adequate.

As an example, consider an amplifier having a number of coincident low-frequency poles and an equal number of zeros at the origin. This corresponds to an amplifier with a number of identical RC coupling networks (Section 7.4.2). For simplicity, the variation of the ψ functions

at high frequencies is ignored; only the low-frequency sections of the Nyquist and Bode diagrams are shown. (The case of coincident high-frequency poles is considered in Section 10.5.3.3.)

If the amplifier has a single pole, the normalized asymptotic Bode diagrams are as shown in Figs. 10.16*a* and *b*. The excess phase shift never exceeds 90°, so regenerations cannot occur no matter how much feedback is applied. The Nyquist diagram is a semicircle. If the number of poles is increased to two, the excess phase shift reaches 180° at zero frequency and, in principle, regeneration can occur if $|A_{lm}|$ is infinite. For three coincident poles, the excess phase shift reaches 270° and regeneration occurs with a finite value of $|A_{lm}|$. Asymptotic Bode diagrams show that the excess phase shift reaches 180° at a frequency about one octave below the break, and that the attenuation is then rather less than 10. Accurate construction of the Bode or Nyquist diagrams shows that regeneration occurs when $|A_{lm}|$ is 8 and that the frequency of 180° excess phase shift is $\omega_{co}/\sqrt{3}$.

The possible excess phase increases by a further 90° for each additional pole, and there is a decrease in the amount of mid-band feedback that can be applied before the onset of regeneration. An important practical conclusion is that it becomes increasingly difficult to apply large amounts of feedback to an amplifier as the number of stages is increased, because each stage contributes one high-frequency pole and at least one low-frequency pole.

Bode* has shown that for minimum-phase networks there is a relationship between the magnitude and phase of the frequency response. (A minimum-phase network has all its zeros as well as all its poles in the left half of the s plane.) Many amplifiers are minimum-phase structures, and in such cases regeneration can be predicted from a plot of the loop-gain magnitude alone. The exact expression for the phase at any frequency is complicated; it involves the magnitude at all frequencies from zero to infinity, but the dominating factor is the rate of change of magnitude at the particular frequency. A rule of thumb is the following:

> If the rate of attenuation of the loop gain does not exceed 33 dB/decade at the frequency of unity loop gain, the amplifier is unlikely to regenerate when the feedback loop is closed.

Rates of attenuation other than multiples of 20 dB/decade can be approximated over a large bandwidth by interlacing poles and zeros on the negative

* H. W. Bode, *Network Analysis and Feedback Amplifier Design*, Chapter 14. See also Section 14.2.1 of this book.

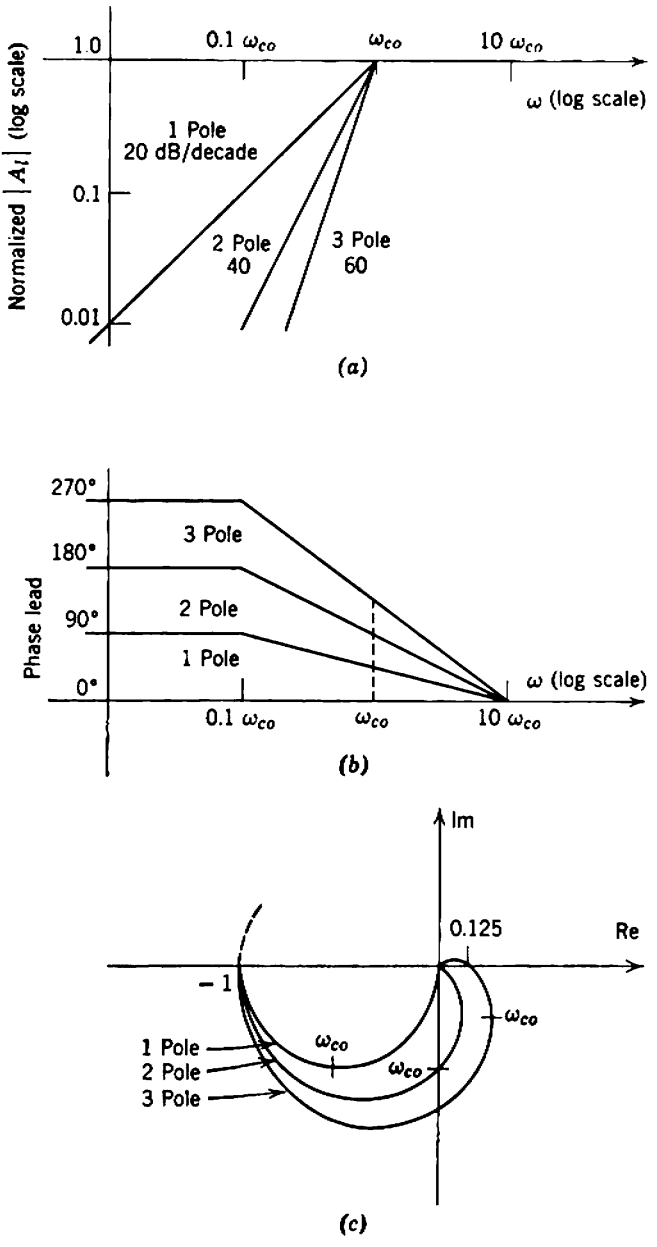


Fig. 10.16 Normalized asymptotic Bode diagrams, and normalized Nyquist diagram for a feedback amplifier with a number of coincident low-frequency breaks at ω_{co}: (*a*) Bode plot, magnitude (log scales); (*b*) Bode plot, phase (log frequency scale); (*c*) Nyquist diagram (linear polar).

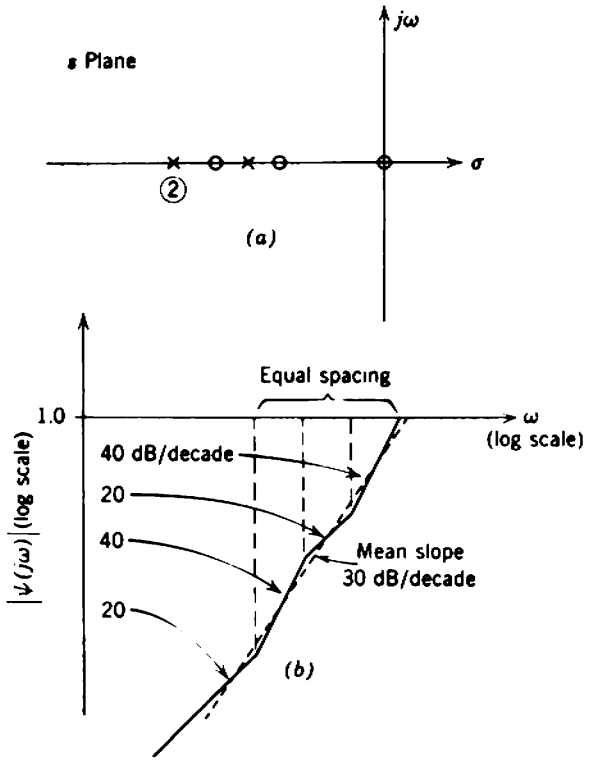


Fig. 10.17 Build-up of 30 dB/decade attenuation rate from interlaced poles and zeros: (*a*) singularity pattern of $\psi(s)$; (*b*) asymptotic plot of $|\psi(j\omega)|$.

real axis. As an example, the spacing shown in Fig. 10.17 gives 30 dB/decade average attenuation rate and 135° phase shift. Section 14.2.1.3 contains more details. Low-frequency zeros on the axis can be produced by bypass or decoupling capacitors (Section 7.4); high-frequency zeros can be produced by the *RC* ladder networks described in Section 14.2.1.3.

10.5.2.3 Closed-Loop Frequency Response

In a Nyquist diagram the vector drawn from the origin to the loop-gain locus is $A_l(j\omega)$ and the vector from the locus to the critical point is $F(j\omega)$. Equation 10.49 can be rewritten as

$$TF(j\omega) = \frac{1}{\beta(j\omega)} \left[\frac{A_l(j\omega)}{F(j\omega)}\right]. \tag{10.55}$$

Thus the closed-loop frequency response can be calculated graphically from the behavior of these vectors and the known value of $\beta(j\omega)$. The

calculation is simplified if contours of constant $|A_l/F|$ are drawn on the Nyquist diagram; Fig. 10.18 shows the contours for

$$\left|\frac{A_l}{F}\right| = \frac{1}{\sqrt{2}} = -3\text{ dB},$$

$$\left|\frac{A_l}{F}\right| = 1 = 0\text{ dB},$$

$$\left|\frac{A_l}{F}\right| = \sqrt{2} = +3\text{ dB},$$

whereas the critical point and origin correspond to infinite and zero closed-loop transfer function, respectively. In the special, but not unusual, case of an amplifier that has β independent of frequency, Eq. 10.55 reduces to

$$TF(j\omega) = \frac{1}{\beta_m}\left[\frac{A_l(j\omega)}{F(j\omega)}\right] \tag{10.56}$$

and the closed-loop frequency response can be read directly from the Nyquist diagram.

The singularity pattern, frequency response, and time response of any network are uniquely interrelated, so that if any one is known the other two can be found. Therefore it is possible to work back from the Nyquist diagram of a feedback amplifier through the closed-loop frequency response to the singularity pattern and time response. Quantitative

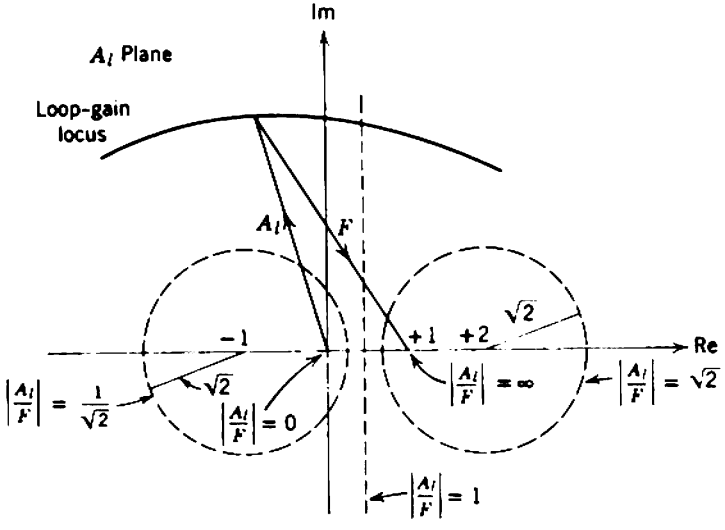


Fig. 10.18 Contours of constant $|A_l(j\omega)/F(j\omega)|$ on a Nyquist diagram.

interrelations are given in Sections 12.1 and 13.2, but the qualitative trends are as follows:

If a pole in the transfer function of an amplifier lies in the left half of the s plane but close to the $j\omega$ axis, there is a peak in the frequency response. The transient response has a "bow" on the top of a square wave (for a low-frequency pole) or a ring on the leading edge (for a high-frequency pole). For the limiting case in which a pole lies on the $j\omega$ axis, the amplifier just oscillates at the frequency of the pole; expressed another way, there is an infinite peak in the frequency response. If poles of the transfer function lie in the right half of the s plane, the amplifier oscillates with increasing amplitude. The amplitude builds up until the amplifier overloads and the s plane representation is invalidated.

From another viewpoint a Nyquist diagram is a map of the $j\omega$ axis of the s plane, and the critical point represents a pole (or poles) of the closed-loop transfer function. Therefore, the proximity of the loop-gain locus to the critical point gives an indication of the positions of the closed-loop poles. As $|A_{lm}|$ is increased from zero, the locus approaches the critical point and the closed-loop transfer function develops a peak at the frequency of closest approach. At the value of A_{lm} for which the locus touches the critical point, the pole(s) lie on the $j\omega$ axis and the frequency of oscillation is the frequency of 180° excess loop phase shift. For larger values of $|A_{lm}|$ the locus crosses the critical point, that is, the pole(s) move into the right half of the s plane.

10.5.2.4 Design Principles

The essential aim in designing a feedback amplifier is to keep the loop-gain locus away from the critical point. Moreover, the design must be such that the locus stays away from the critical point even with an unfavorable combination of component tolerances. In short, some safety margin must be allowed in the design.

If prevention of regeneration is the sole criterion of satisfactory dynamic response, standard practice is to allow safety margins of 30° in phase and a factor of two in magnitude. The loop-gain locus must not enter the region marked in Fig. 10.19. These margins are based on typical tolerances for components in the forward path of the amplifier; the reader may verify that they are adequate except under unusual circumstances.

Quite a different problem is to design an amplifier whose response is to be substantially flat and free from peaks. Figure 10.18 shows contours of constant gain, from which can be deduced the additional regions forbidden to the loop-gain locus for various degrees of flatness in the closed-loop response.

There are two basic techniques for steering the loop-gain locus away

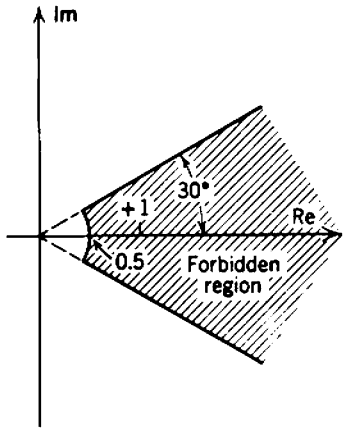


Fig. 10.19 Forbidden region on the Nyquist diagram, for 2:1 magnitude and 30° phase margins.

from the critical point—compensation and the use of dominant poles. Compensation is achieved by adding zeros to the loop gain in order to reduce the loop phase shift. The frequency (or frequencies) at which the zero(s) will be most beneficial can be estimated from the Nyquist diagram, remembering that a real-axis zero contributes 45° phase shift at its own frequency. The approximate Bode criterion of Section 10.5.2.2 is a good example of compensation; the interlaced poles and zeros in Fig. 10.17 keep the rate of loop-gain attenuation below 30 dB/decade and the phase shift below 135°, thereby steering the locus away from the critical point. This design technique is used extensively in maximum-feedback amplifiers (Section 14.2).

In the dominant pole technique the loop-gain cutoff is arranged to be due essentially to one pole; it is shown in Section 10.5.2.2 that a single-poled feedback loop cannot regenerate because the excess phase is limited to 90°. In practice, one pole dominates the cutoff to such an extent that the loop-gain magnitude falls below unity before any other poles contribute undesirably to the phase shift. This technique (or a modification of it) is used extensively in feedback peaking of amplifiers (Sections 12.2 and 13.5). In this application the primary purpose of the feedback is to extend the bandwidth—the attendant increase in stability is merely a bonus. Therefore the fact that the loop gain and stability fall toward the passband edges is of no consequence; (the closed-loop bandwidth is much greater than the bandwidth of the loop gain, Section 10.5.1). However, the technique is inefficient if a specific loop-gain magnitude and gain stability must be maintained over some bandwidth.

The dominant low- and high-frequency poles are placed at the edges of this useful passband, and because the rate of loop-gain cutoff is 20 dB/decade all other poles must lie outside the dominant poles by a factor at least as large as $|A_{lm}|$. Thus the bandwidths of all stages in the amplifier except one, as well as the closed-loop bandwidth, must be much greater than the useful bandwidth. The most efficient design results from using the maximum permissible rate of loop-gain cutoff, namely, 33 dB/decade, corresponding to 150° excess phase shift and, therefore, to 30° safety margin as above.

10.5.3 The Root-Locus Method

Both the strength and weakness of Nyquist's regeneration criterion as a tool in circuit design are that it is concerned only with frequency response. Loop-gain magnitude can be calculated or measured easily, and in the case of minimum-phase amplifiers it provides all the information necessary to determine the closed-loop frequency response; in the general case, the variation of phase with frequency is required also. Because of these characteristics, Nyquist's methods are ideally suited to amplifiers for which the input signal is band-limited to lie within a certain frequency range, and the specifications call for a loop-gain magnitude to be maintained over this range. Telephone repeater amplifiers are the classic example. In this application many (perhaps hundreds) amplifiers are connected in tandem, spread over the length of a cable to make up its losses, and individual gain stabilities of perhaps 0.01 dB are required to keep the total gain constant to within a few decibels. Loop gains of hundreds are maintained over the required frequency range, and the telephone signals are band-limited by filters to lie within it. The behavior outside this range is unimportant. Peaks of a few decibels are permitted near the edges of the closed-loop passband, and design criteria such as the safety factor illustrated in Fig. 10.19 are adequate. Black, Nyquist, Bode, and many other well-known feedback theoreticians were telephone engineers.

Since the mid-1940's feedback has commonly been applied to systems for which the input signal is not limited to lie well within the closed-loop passband. Often, the frequency response of a mechanical control system or electronic amplifier is inadequate, due mainly to the limitations of one component. If feedback is applied to the system, this limiting component sets the loop-gain bandwidth but the closed-loop system bandwidth is increased. Thus, the feedback potentially enables the system output to follow the input waveform more closely but, because the input signal is not band-limited, any resonances (peaks) in the frequency response can be excited, giving an unacceptable transient response.

Although the singularity pattern and transient response can be deduced approximately from the closed-loop frequency response, this extension of Nyquist's criterion is too tedious for many amplifier design studies. Evans's root-locus method* introduced in 1948 has rapidly gained acceptance as the most powerful design aid for control systems. With few exceptions, books on the root-locus method have been written by control engineers; the reader is hereby cautioned that the positive sign convention for loop gain is by far the more common.

It is shown in Section 10.5.1 that the closed-loop singularity pattern is made up of

$$\text{zeros at the zeros of } \psi_\mu(s),$$

$$\text{zeros at the poles of } \psi_\beta(s),$$

$$\text{poles at the zeros of } [1 - A_{lm}\,\psi_\mu(s)\,\psi_\beta(s)].$$

The closed-loop zeros can be written down at sight, and are independent of A_{lm}. For given ψ functions, the closed-loop poles move about the complex frequency plane as $|A_{lm}|$ is increased from zero. The paths along which the poles move are termed the *root loci*, and the root-locus method is essentially a graphical solution of the characteristic equation (Eq. 10.54) by means of these loci. Because of the special form of this equation, the loci obey a number of geometrical rules and can be constructed easily.

10.5.3.1 The Root-Locus Equations

The characteristic equation can be rewritten as

$$A_{lm}\,\psi_\mu(s)\,\psi_\beta(s) = 1. \tag{10.57}$$

The ψ functions give the poles and zeros of the loop gain, so Eq. 10.57 can be expanded as

$$\nu A_{lm}\left[\frac{(s - z_1)(s - z_2)\cdots}{(s - p_1)(s - p_2)\cdots}\right] = 1, \tag{10.58}$$

where $p_1\,p_2\cdots$ = the poles of $A_l(s)$,
$z_1\,z_2\cdots$ = the zeros of $A_l(s)$,
ν = a constant, shown in the next paragraph to be real.

Notice that, while the ψ functions considered in Chapter 7 contain only real poles and zeros, complex singularities are permissible provided they occur in conjugate pairs. Often, all singularities lie in the left half of the s plane. Zeros occur in the right half-plane if either the forward path μ

* W. R. Evans, "Graphical analysis of control systems," *Trans. Am. Inst. Elect. Engrs.*, 67, Part 1, 547, 1948.

or feedback network β is not a minimum-phase structure, whereas poles occur in the right half-plane if the forward path is regenerative (an unlikely, though not forbidden situation). The constant νA_{lm} in Eq. 10.58 is called the *loop level.*

The ψ functions are the products of a number of low- and high-frequency factors. By definition, the low-frequency ψ factors approach unity as $|s|$ becomes very large, whereas the high-frequency ψ factors approach unity as $|s|$ becomes very small. In general, the low-frequency factor associated with any one capacitor is of the form

$$\psi_{\text{low}}(s) = \frac{s - z}{s - p},$$

where z may vanish, and either p or z may be one of a complex conjugate pair. The high-frequency factors are of the form

$$\psi_{\text{high}}(s) = \frac{1 - s/z}{1 - s/p} = \frac{p(s - z)}{z(s - p)},$$

where z may be infinite, and either p or z may be one of a complex conjugate pair. Summing over all finite singularities, the constant ν becomes

$$\nu = \left[(-1)^{(n_p - n_z)}\left(\frac{p_1 p_2 \cdots}{z_1 z_2 \cdots}\right)\right]_{\text{h.f. singularities only}} \tag{10.59}$$

Because complex singularities occur only as conjugate pairs, ν must be real. Moreover, ν is positive provided all singularities lie in the left half of the s plane. In order to avoid confusion over a sign ambiguity, it is hereinafter assumed that this condition is satisfied and ν is positive. The extension to non-minimum-phase or open-loop-regenerative amplifiers is essentially trivial, and is left as an exercise for the reader.

If a feedback amplifier has a mid-band frequency range over which the ψ functions are approximately unity, the low- and high-frequency roots of the characteristic equation can be considered separately. When the low-frequency roots are considered alone, Eq. 10.58 reduces to

$$A_{lm}\left[\frac{(s - z_1)(s - z_2)\cdots}{(s - p_1)(s - p_2)\cdots}\right]_{\text{l.f. singularities}} = 1, \tag{10.60}$$

whereas at high frequencies

$$A_{lm}\left[\frac{p_1 p_2 \cdots (s - z_1)(s - z_2)\cdots}{(-1)^{(n_p - n_z)} z_1 z_2 \cdots (s - p_1)(s - p_2)\cdots}\right]_{\text{h.f. singularities}} = 1. \tag{10.61}$$

In the complex frequency plane, the term $(s - p_k)$ represents the vector from the point p_k to the general point s (Fig. 10.20). Similarly, $(s - z_k)$ represents the vector from z_k to s. Therefore, after transposing the loop

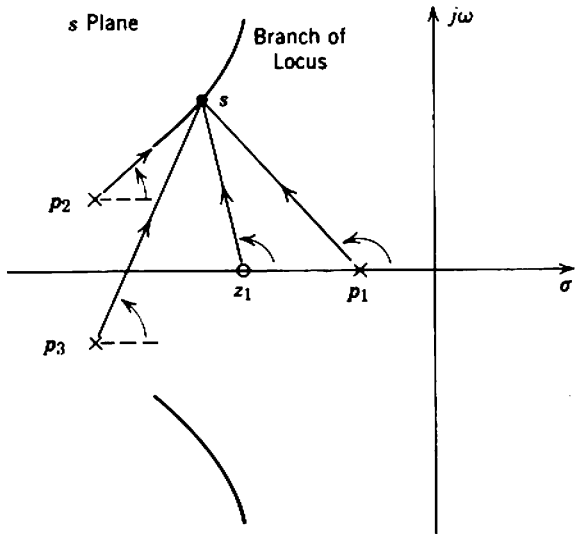


Fig. 10.20 Vectors $(s - p_k)$ and $(s - z_k)$ in the complex frequency plane.

level νA_{lm} from the left to the right side of Eq. 10.58, it follows that for every point on the root loci

$$\frac{\text{product of vectors from zeros}}{\text{product of vectors from poles}} = \frac{1}{\nu A_{lm}}. \tag{10.62}$$

Equation 10.62 provides the geometrical basis for constructing the loci.

10.5.3.2 Rules for Constructing the Loci*

Three rules for the number of branches in the root-locus plot follow from the general form of Eq. 10.58. For very small values of $|A_{lm}|$ the left-hand side is finite only at the poles of loop gain. Therefore:

RULE 1. At very small values of $|A_{lm}|$, the root loci start out from the poles of $A_l(s)$.

As a direct corollary:

RULE 2. The number of branches in the root locus plot is the number of poles in $A_l(s)$.

* For a root-locus plotting machine, see Y. J. KINGMA, "Root-locus plotting machine for classroom demonstration," *Trans. Inst. Elec. Electronics Engrs.*, E-7, 86, June–September 1964. Digital computer programs for plotting root loci are available from many sources.

By reasoning similar to Rule 1:

RULE 3. At very large values of $|A_{lm}|$, the root loci finish on the zeros of $A_l(s)$.

Invariably, $A_l(s)$ has more poles than finite zeros; the behavior of the branches that cannot terminate on finite zeros is discussed in connection with Rule 5.

Equation 10.62 leads to a number of rules that apply for finite values of $|A_{lm}|$. Remembering that A_{lm} is negative for negative feedback and that ν is positive,

$$\frac{\text{product of vectors from zeros}}{\text{product of vectors from poles}} = \text{(a negative real number)}.$$

Therefore:

RULE 4. The sum of the angles of the vectors from all zeros of $A_l(s)$ to any point on the loci, less the sum of the angles from all poles, must be an odd multiple of 180°:

$$\sum \left(\underline{/\text{vectors from all zeros}} \right) - \sum \left(\underline{/\text{vectors from all poles}} \right) = (2n + 1)\pi. \quad (10.63)$$

Rule 4 is in itself sufficient to construct the entire root-locus plot on a trial-and-error basis. However, a few further rules enable the loci to be sketched roughly from Eq. 10.63 so, only if required, need the precise detail be filled in. In any practical construction of the loci it is more convenient to measure the angles indicated in Fig. 10.21 rather than the angles in Fig. 10.20 because the center of the protractor remains stationary. A special protractor called a *Spirule** is designed to add as well as measure angles, and assists greatly in plotting root loci. The following are corollaries of Rule 4:

RULE 5. If there are n_p poles and n_z zeros, the $(n_p - n_z)$ branches that cannot finish on finite zeros radiate out toward infinity as $|A_{lm}|$ becomes very large, at the angles

$$\theta = \frac{(2n + 1)\pi}{n_p - n_z}. \quad (10.64)$$

Their asymptotes intersect on the negative real axis at

$$\sigma = \frac{\Sigma(p) - \Sigma(z)}{n_p - n_z}. \quad (10.65)$$

* Available from The Spirule Co., 9728 El Venado Drive, Whittier, California.

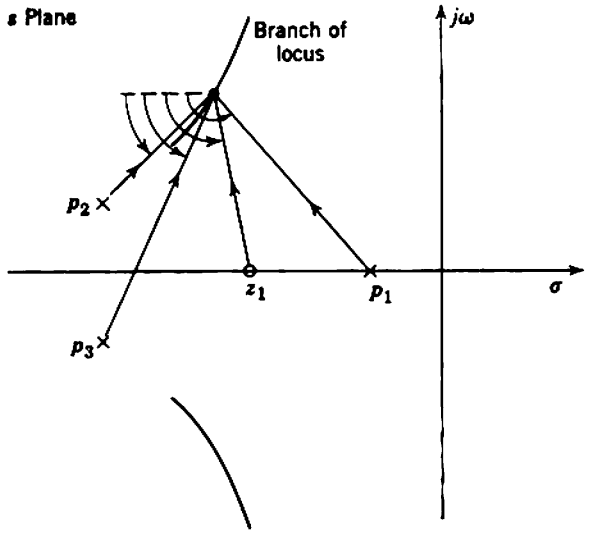


Fig. 10.21 Angles usually measured in root-locus plotting.

RULE 6. One branch of the locus lies along the negative real axis when the number of singularities to the right is odd.

RULE 7. When two branches of the locus meet on the negative real axis, they break away at right angles. If all open-loop singularities are real, the point σ at which the break occurs satisfies the relation

$$\sum_{\substack{\text{all poles}\\ \text{to left}}} \left(\frac{1}{\sigma - p}\right) - \sum_{\substack{\text{all zeros}\\ \text{to left}}} \left(\frac{1}{\sigma - z}\right) = \sum_{\substack{\text{all poles}\\ \text{to right}}} \left(\frac{1}{p - \sigma}\right) - \sum_{\substack{\text{all zeros}\\ \text{to right}}} \left(\frac{1}{z - \sigma}\right). \tag{10.66}$$

RULE 8. The locus leaves a pole at the angle

$$\theta = (2n + 1)\pi - \sum \left(\underline{/\text{vectors from all other poles}}\right) + \sum \left(\underline{/\text{vectors from all zeros}}\right). \tag{10.67}$$

RULE 9. The locus approaches a zero from the angle

$$\theta = (2n + 1)\pi + \sum \left(\underline{/\text{vectors from all poles}}\right) - \sum \left(\underline{/\text{vectors from all other zeros}}\right). \tag{10.68}$$

As an example in the use of these rules, Fig. 10.22 shows a 5-pole 2-zero pattern and two steps in the construction of the loci. Rules 1 to 3 give

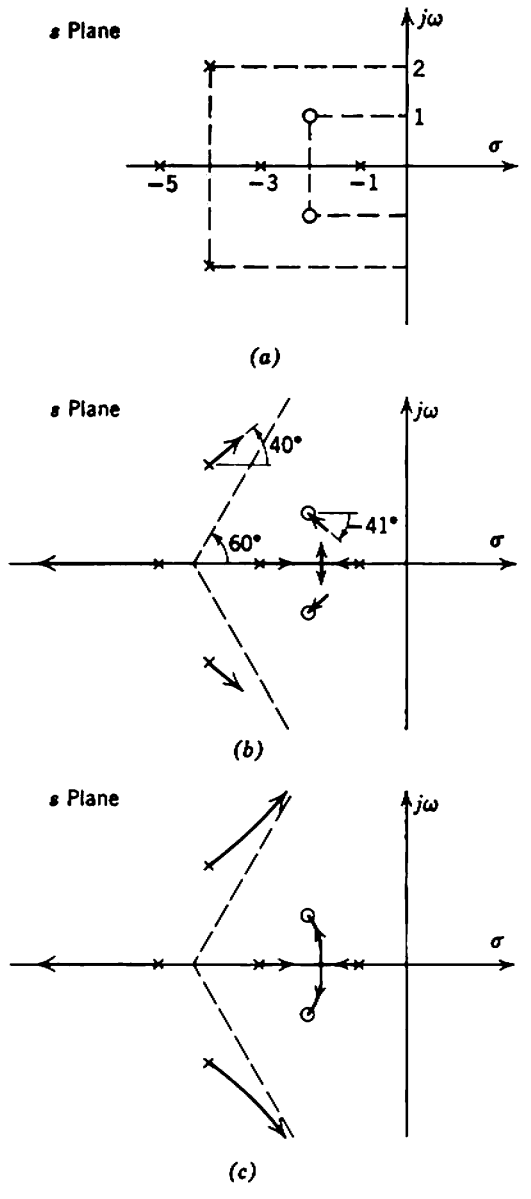


Fig. 10.22 Example of root-locus plotting; (*a*) singularity pattern of $A_i(s)$; (*b*) quantitative information from the rules; (*c*) guess at filling in the loci.

general information while Rules 5 to 8 give the specific information shown in Fig. 10.22*b*:

Rule 5:

$$\begin{aligned} n_p &= 5, \\ n_z &= 2, \\ \Sigma(p) &= -17, \\ \Sigma(z) &= -4. \end{aligned}$$

Therefore the asymptotes are at angles $\pi/3$, π, and $5\pi/3$, and intersect at -4.33.

Rule 6 shows that the locus lies along the negative real axis between -1 and -3, and to the left of -5.

Rule 7 in conjunction with Rule 1 shows that the loci break away at right angles to the negative real axis somewhere between -1 and -3. The exact point cannot be found from Eq. 10.66 because some singularities do not lie on the negative real axis.

Rule 8 for the pole at $(-4, +2)$

$$\sum \left(\underline{/\text{vectors from all other poles}} \right) = +417°,$$

$$\sum \left(\underline{/\text{vectors from all zeros}} \right) = +277°.$$

Therefore $\theta = +40°$. Similarly, for the pole at $(-4, -2)$, the angle is $-40°$.

Rule 9 for the zero at $(-2, +1)$

$$\sum \left(\underline{/\text{vectors from all poles}} \right) = +229°,$$

$$\sum \left(\underline{/\text{vectors from all other zeros}} \right) = +90°.$$

Therefore $\theta = -41°$. Similarly, for the zero at $(-2, -1)$, the angle is $+41°$.

In view of these facts, it is almost certain that the loci behave qualitatively as in Fig. 10.22*c*.

10.5.3.3 Determination of $|A_{lm}|$

The root loci are the trajectories of the closed-loop poles as $|A_{lm}|$ runs from zero to infinity. Once the loci have been constructed, the mid-band

loop gain corresponding to any point on them can be found from Eq. 10.62. Rearranging and taking moduli to eliminate the minus sign associated with A_{lm} yields

$$|A_{lm}| = \frac{1}{\nu}\left(\frac{\text{product of distances from poles}}{\text{product of distances from zeros}}\right). \tag{10.69}$$

Notice that "vectors" have been replaced by "distances." The Spirule is designed to multiply lengths as well as add angles, and is again convenient.

As an example, consider an amplifier having n_p coincident high-frequency poles. This corresponds to an amplifier having a number of identical stages; the example is the inverse of that in Section 10.5.2.2 for which Nyquist methods were used. Suppose the poles lie on the negative real axis at $-\omega_{co}$ as in Fig. 10.23 so that ν is $(\omega_{co})^{n_p}$.

For the case of one pole the single root-locus branch lies entirely along the negative real axis, and the closed-loop pole moves from $-\omega_{co}$ out toward $-\infty$ as $|A_{lm}|$ is increased from zero (Fig. 10.23*a*). When the pole

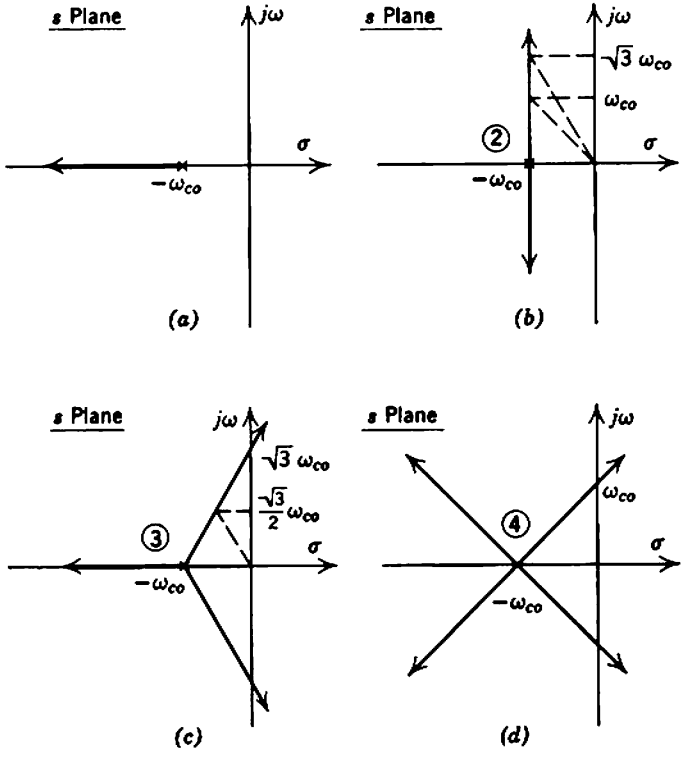


Fig. 10.23 Root-locus diagrams for multi-coincident-pole systems: (*a*) 1 pole; (*b*) 2 pole; (*c*) 3 pole; (*d*) 4 pole.

is at $-3\omega_{co}$, for example, $|A_{lm}|$ is 2. The pole cannot enter the right half of the s plane, so the system cannot become regenerative. Indeed, since the pole remains real no matter how large $|A_{lm}|$ becomes, the system cannot even develop a peak in its closed-loop frequency response or an overshoot in its transient response.

For a double-pole system the two branches of the root locus lie parallel to the imaginary axis (Fig. 10.23*b*) and again regeneration cannot occur no matter how large $|A_{lm}|$ becomes. The closed-loop poles, however, are complex for all finite values of $|A_{lm}|$, and the impulse response is a decaying oscillation. From the geometry of the diagram the poles lie at

$$45° \text{ for } |A_{lm}| = 1,$$

$$60° \text{ for } |A_{lm}| = 3.$$

Poles at 60° correspond to half-critical damping. The resulting overshoot in the step response is 16.4%; this is about the maximum tolerable in any system. Thus 3 is a limit to the useful loop gain that can be applied to a two-coincident-pole system.

For a triple-pole system the three branches of the root locus radiate out at 60°, 180°, and 300° as in Fig. 10.23*c*, and one pair of poles is complex for all finite values of $|A_{lm}|$. Their loci cross the imaginary axis at the points $\pm j\sqrt{3}\,\omega_{co}$ when $|A_{lm}| = 8$, and enter the right half of the s plane. Thus, 8 is an absolute limit to the feedback that can be applied to a three-coincident-pole system, and for this value the system just oscillates at the frequency $\sqrt{3}\,\omega_{co}$. A more useful limit, corresponding to 60° poles, is $|A_{lm}| = 1$. Notice how the maximum useful feedback falls as the number of poles is increased. The trend continues for four poles, when the loci radiate out at 45°, 135°, 225°, and 315°, Fig. 10.23*d*, and all poles are complex for all finite values of $|A_{lm}|$. Two branches of the locus cross the imaginary axis into the right half of the s plane at the points $\pm j\omega_{co}$ corresponding to $|A_{lm}| = 4$. This example illustrates how little feedback can be applied to some systems, and therefore what little increase in gain stability can be achieved through the use of feedback if the open-loop singularity pattern is unsuitable.

10.5.3.4 Design Principles

The aim in designing a feedback system from root-locus considerations is to find an open-loop singularity pattern for which the root loci pass through the locations desired for the closed-loop poles. Further, the positions of the poles on the loci must correspond to the desired mid-band loop gain. In elementary cases it may be possible to track backward along the loci from the desired closed-loop pole positions at the nominal

loop gain to the open-loop positions at zero loop gain, and thereby obtain the open-loop transfer function directly.* However, this is not possible in general. Design is essentially a trial-and-error process, and is very time consuming unless some liberties are taken with the final closed-loop singularity pattern.

The greatest practical simplification is that only two or three root-locus branches are significant and need be controlled with precision, because the response of any useful amplifier is dominated by only a few singularities. Chapters 12 and 13 contain the relevant information. Moreover, there is a simple rule for distorting the loci, so that the branches for a given open-loop pattern can be changed if they pass near but not through the desired dominant closed-loop pole positions; open-loop poles repel the loci, whereas zeros attract them. If two branches occur as in the incomplete root-locus plots of Fig. 10.24, moving the pole in Fig. 10.24*a* or the zero in Fig. 10.24*b* will distort the loci as shown.

If the loop gain $A_l(s)$ has a dominant high-frequency pole (much closer to the origin than all other high-frequency singularities), changes in this pole have little effect on the root-locus plot. However, changing the pole alters the constant ν in Eq. 10.62, and can therefore be used to control the value of $|A_{lm}|$ at which the root locus passes through a given point. The same conclusion applies to a dominant low-frequency pole (much further from the origin than all other low-frequency singularities), although the mechanism is different.

The weakness of these and the Nyquist design methods described in Section 10.5.2.4 is that the closed-loop singularity pattern depends on

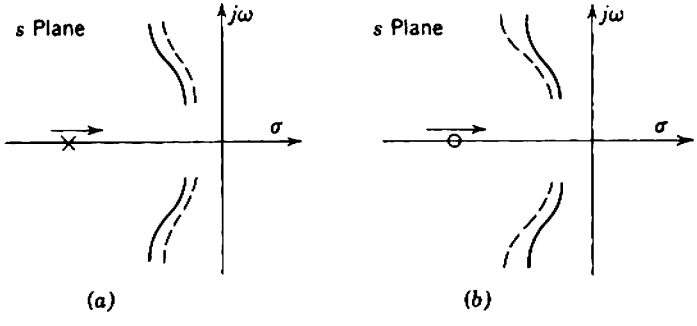


Fig. 10.24 The design technique of "locus pushing."

* R. J. D. REEVES, "Feedback amplifier design," *Proc. Inst. Elec. Engrs.* (*London*), **100**, Part III, 43, January 1953; *ibid.*, **99**, Part IV, 383, December 1952.

$A_l(s)$, and therefore on the forward path $\mu(s)$. While the feedback stabilizes the closed-loop transfer function magnitude over the mid-band range of $A_l(s)$, the closed-loop singularity pattern is not stabilized. An exceedingly powerful design technique* makes use of the fact that the root loci terminate on the zeros of $A_l(s)$ as $|A_{lm}|$ approaches infinity. Therefore, if $|A_{lm}|$ is made large and the feedback factor $\beta(s)$ is designed to have zeros at the points where dominant poles of $TF(s)$ are required, these poles will lie close to their desired positions almost independent of $\mu(s)$. (It is no use setting zeros of $\mu(s)$ at the desired closed-loop poles because these zeros appear in $TF(s)$ and would almost cancel the poles.) Small corrections can be included in the zeros of $\beta(s)$ so that the closed-loop poles lie exactly at the desired locations with nominal A_{lm}. In addition, the feedback network can be used to set the dominant closed-loop zeros because poles of $\beta(s)$ appear as zeros in $TF(s)$. Thus, the approximate form of Eq. 10.36

$$TF(s) \approx -\frac{1}{\beta(s)}$$

is true for the singularity pattern as well as the mid-band magnitude. All dependence on the forward path is minimized. Zeros of $\beta(s)$ which control the dominant poles of $TF(s)$ are often called *phantom zeros.*

10.6 TRANSFER FUNCTIONS

Most practical feedback amplifiers approximate to one of the ideal structures shown in Fig. 10.25.

For two voltages to add, they must be in series; it is meaningless to speak of two unequal voltage sources being connected in shunt. In Figs. 10.25*a* and *b*, a voltage is fed back into the input circuit and added in series with the true input voltage to give the input voltage applied to the forward path. This arrangement is known as *series* or *voltage feedback at the input.* Similarly, for two currents to add they must be in shunt; it is meaningless to speak of two unequal current sources being connected in series. In Figs. 10.25*c* and *d*, a current is fed back into the input circuit and added in shunt with the true input current to give the input current applied to the forward path. This arrangement is known as *shunt* or *current feedback at the input.*

A circuit connected in shunt with a load can sample the voltage across it, but cannot detect the current through the load. In Figs. 10.25*b* and *d*,

* Perhaps the first explicit statement of this technique is: M. S. GHAUSI and D. O. PEDERSON, "A new design approach for feedback amplifiers," *Trans. Inst. Radio Engrs.*, CT-8, 274, September 1961.

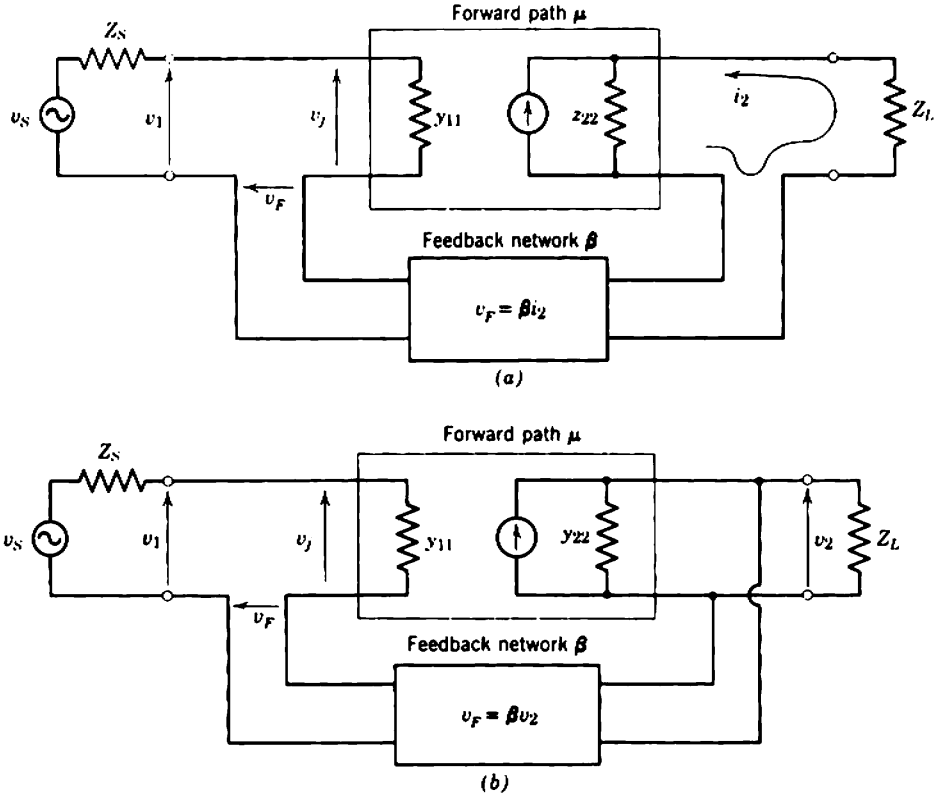


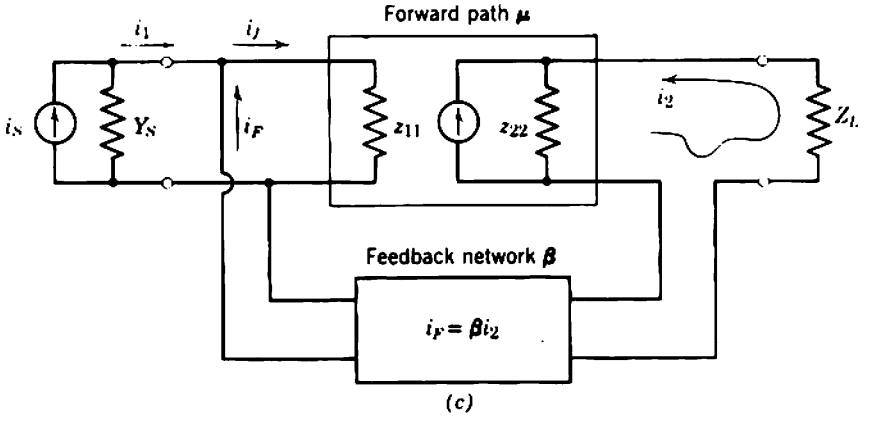
Fig. 10.25 Block diagrams for ideal feedback structures.

the feedback is derived from the load voltage and the arrangement is known as *shunt* or *voltage feedback at the output*. Similarly, a circuit connected in series with a load can sample the current through it, but cannot detect the voltage across the load. In Figs. 10.25*a* and *c*, the feedback is derived from the load current and the arrangement is known as *series* or *current feedback at the output*.

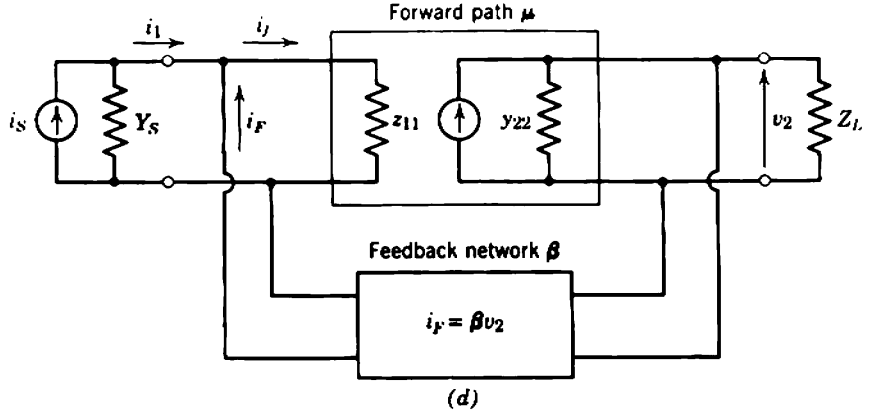
Table 10.1 lists the stable closed-loop transfer function types for the four structures in Fig. 10.25, together with the μ and β functions; these follow from the discussion in Section 10.4.2. The closed-loop transfer function magnitude is, of course, given by Eq. 10.36, namely,

$$TF = -\frac{1}{\beta}\left(\frac{1}{1 - 1/\mu\beta}\right) \approx -\frac{1}{\beta}.$$

It is worth emphasizing that only one transfer function is stable for each configuration.



(c)



(d)

Fig. 10.25 (*continued*)

Table 10.1 Transfer Functions of Ideal Feedback Structures

Stable Closed-Loop Transfer Function *TF*	Reference Variable μ (Open-Loop Transfer Function)	Feedback Factor β
$Y_T = \dfrac{i_2}{v_1}$	$Y_{T0} = \dfrac{i_2}{v_j}$	$v_F = \beta i_2$
$A_V = \dfrac{v_2}{v_1}$	$A_{V0} = \dfrac{v_2}{v_j}$	$v_F = \beta v_2$
$A_I = \dfrac{i_2}{i_1}$	$A_{I0} = \dfrac{i_2}{i_j}$	$i_F = \beta i_2$
$Z_T = \dfrac{v_2}{i_1}$	$Z_{T0} = \dfrac{v_2}{i_j}$	$i_F = \beta v_2$

Figure 10.26 shows practical approximations to the four basic configurations. The ideal closed-loop transfer functions for parts *a* to *d* of the diagram (corresponding to $-1/\beta$ in Table 10.1) are

$$Y_T = \frac{1}{Z_{F1}}, \tag{10.69a}$$

$$A_V = \frac{Z_{F1} + Z_{F2}}{Z_{F1}}, \tag{10.69b}$$

$$A_I = -\frac{Z_{F1} + Z_{F2}}{Z_{F1}}, \tag{10.69c}$$

$$Z_T = -Z_{F2}. \tag{10.69d}$$

In order to obtain negative feedback, the amplifier without feedback must have the correct phase shift at mid-band frequencies. Restrictions must

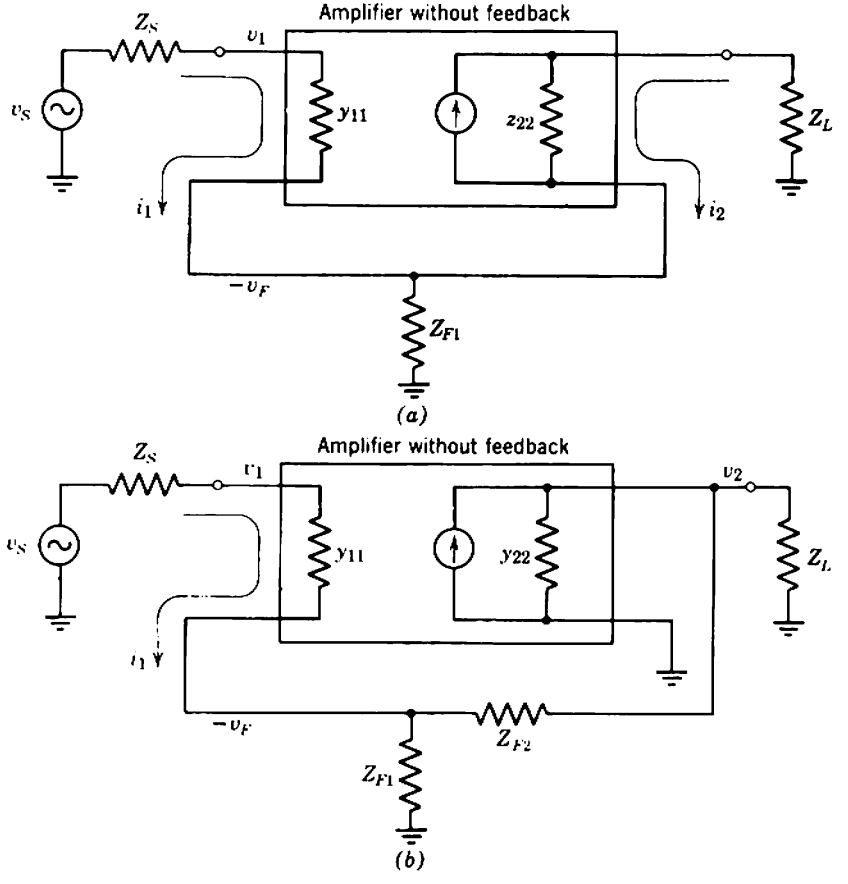


Fig. 10.26 Block diagrams for practical feedback structures.

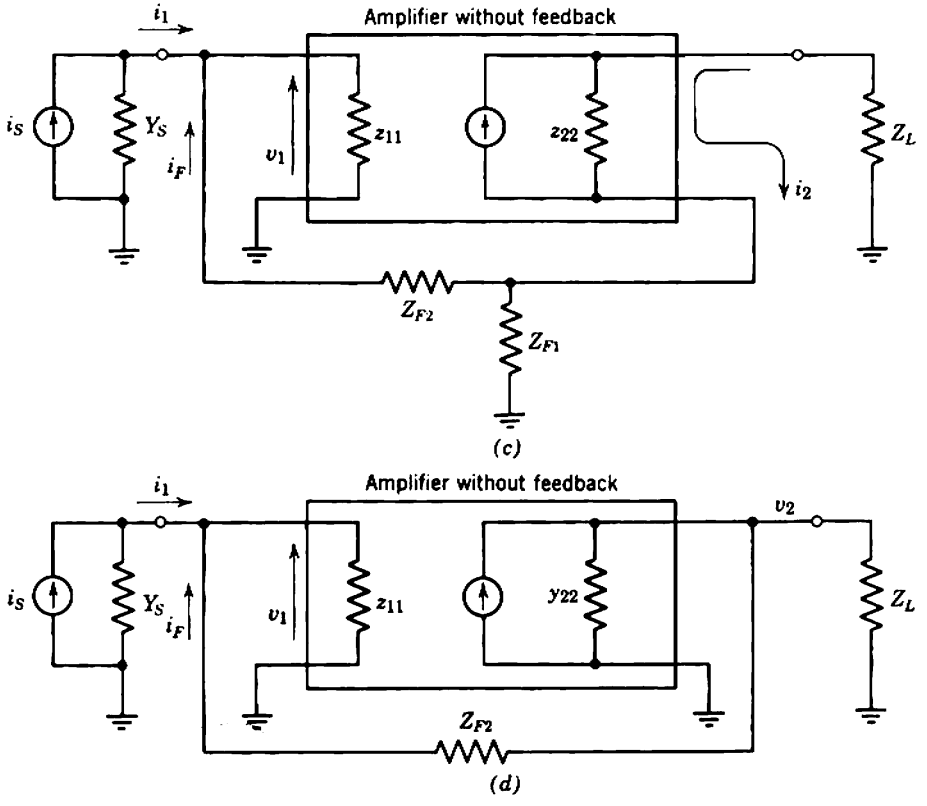


Fig. 10.26 *(continued)*

therefore be placed on the number of its stages. If common-emitting-electrode stages are used throughout, the amplifiers in *a* and *d* must have an odd number of stages, whereas those in *b* and *c* must have an even number.

10.6.1 Approximations for Loop Gain

Loop gain determines the stability and dynamic response of a feedback amplifier, and is the most basic design parameter. Although loop gain can always be found from β and TF as in Section 10.4.1, practical amplifier design is much simplified if μ can be written down at sight. Formally, the forward path and feedback network in the practical realizations shown in Fig. 10.26 are not separable; that is, *the reference variable μ is not the transfer function of the amplifier with the feedback network removed.* In Figs. 10.26*a* and *b* this complication stems from the fact that the "feedback" voltage across Z_{F1} contains a component due to v_1 unless the input admittance y_{11} of the amplifier without feedback vanishes, forcing i_1 to

zero. Similarly, the "feedback" current through Z_{F2} in Figs. 10.26*c* and *d* contains a component due to i_1 unless the input impedance z_{11} of the amplifier without feedback vanishes, forcing v_1 to zero.

In practice, μ need not be known exactly, because a small error in its value resulting from an approximation in an analytical expression is no more significant than a component being off tolerance in an exact expression. Therefore μ can be approximated by the transfer function of the amplifier without feedback, provided:

(i) The input resistance of the amplifier without feedback is large or small as appropriate.

(ii) Loading of the amplifier output by the feedback network is taken into account. In parts *a* and *c* of Fig. 10.26 the effective load is

$$Z_{L(\text{eff})} \approx Z_L + Z_{F1}, \tag{10.70a}$$

whereas in parts *b* and *d*

$$Z_{L(\text{eff})} \approx Z_L \parallel Z_{F2}. \tag{10.70b}$$

This approximate separation of μ and β is an exceedingly useful design technique.

As an illustration, the loop gain for the transistor shunt-feedback stage in Fig. 10.8 can be written down, provided the input resistance of the transistor is small:

(i) transfer resistance without feedback

$$\mu \approx R_{T0} = -\beta_N R_{L(\text{eff})} = -\beta_N (R_L \parallel R_F),$$

(ii) feedback factor

$$\beta = \frac{1}{R_F}.$$

Therefore the loop gain is

$$A_l = \mu\beta \approx -\beta_N \left(\frac{R_L}{R_F + R_L} \right). \tag{10.71}$$

Equation 10.44 reduces to Eq. 10.71 if $[r_B + (\beta_N + 1)r_E]$ is small, as required for separability.

10.7 DRIVING-POINT FUNCTIONS

In the general case the driving-point functions of a feedback amplifier follow from Eq. 10.28 as

$$DF = DF_D + \frac{gac}{1 - gb}, \tag{10.72}$$

where DF_D, a, b, and c are superposition constants, and g is the reference variable used in analyzing the circuit. In contrast to the transfer functions there is rarely some feedback network deliberately added to stabilize the driving-point functions and, correspondingly, there is no special reference variable for which the form of Eq. 10.72 simplifies. Equation 10.72 as it stands is the useful, general expression for the driving-point functions of a feedback amplifier.

Under special circumstances there may be a simple relation between the input immittances with and without feedback and the loop gain $\mu\beta$; for example, in either the ideal structures of Fig. 10.25 or the practical approximations of Fig. 10.26, the input immittances in parts *a* to *d* respectively are

$$Y_i = \frac{i_1}{v_1} = \frac{y_{11}}{1 - \mu\beta}, \tag{10.73a}$$

$$Y_i = \frac{i_1}{v_1} = \frac{y_{11}}{1 - \mu\beta}, \tag{10.73b}$$

$$Z_i = \frac{v_1}{i_1} = \frac{z_{11}}{1 - \mu\beta}, \tag{10.73c}$$

$$Z_i = \frac{v_1}{i_1} = \frac{z_{11}}{1 - \mu\beta}. \tag{10.73d}$$

Rather more complicated relations exist between the output immittances with and without feedback and the loop gain $\mu\beta$. For the ideal structure in Fig. 10.25*a*

$$Z_o = z_{22}\left(1 - \frac{\mu_{sc}\beta}{1 + Z_S y_{11}}\right), \tag{10.74a}$$

where z_{22} = the output impedance of the forward path,
μ_{sc} = the transfer admittance of the forward path into a short-circuit load,
Z_S = the source impedance.

For parts *b* to *d* of Fig. 10.25

$$Y_o = y_{22}\left(1 - \frac{\mu_{oc}\beta}{1 + Z_S y_{11}}\right), \tag{10.74b}$$

$$Z_o = z_{22}\left(1 - \frac{\mu_{sc}\beta}{1 + Y_S z_{11}}\right), \tag{10.74c}$$

$$Y_o = y_{22}\left(1 - \frac{\mu_{oc}\beta}{1 + Y_S z_{11}}\right). \tag{10.74d}$$

Although Eqs. 10.73 and 10.74 apply only to special amplifier configurations, the same qualitative trends are observed for all configurations.

Series feedback increases driving-point impedances, whereas shunt feedback reduces driving-point impedances.

At the input of an amplifier series feedback is voltage feedback and stabilizes a voltage-input transfer function (Y_T or A_V depending on the output circuit). Because the feedback raises the input impedance, any source behaves more like a voltage (zero impedance) source in comparison and, as stated in Section 7.2.3, the input voltage to the amplifier approaches the open-circuit source voltage, that is, the signal to be amplified. Shunt feedback at the input is current feedback; a current-input transfer function (A_I or Z_T) is stabilized, the input impedance falls, and the current into the amplifier approaches the short-circuit source current. At the output of an amplifier, series feedback is current feedback and stabilizes a current-output transfer function (Y_T or A_I depending on the input circuit). Because the feedback raises the output impedance, any load behaves more like a short circuit in comparison and the load current approaches the short-circuit output current of the amplifier. Shunt feedback at the output is voltage feedback; a voltage-output transfer function (A_V or Z_T) is stabilized, the output impedance falls, and the load voltage approaches the open-circuit output voltage of the amplifier.

10.8 NOISE

There are many requirements for precise amplification of very small signals; preamplifiers for oscilloscopes and electronic voltmeters, strain gauge amplifiers, and head amplifiers for television cameras and nuclear particle detectors are familiar examples. Negative feedback should be applied to these amplifiers to stabilize their gain, and it is necessary to consider the influence of this feedback on their noise. As discussed in Chapter 8, there is no simple, general approach for optimizing noise performance; each problem must be treated on its own merits. However, it is pertinent to discuss here one aspect of noise in feedback amplifiers, namely, the effect of combining part of the noisy output signal with the relatively noise-free input signal.

Consider the ideal amplifier shown in Fig. 10.27, having series feedback at both input and output. Suppose the spot noise of the forward path is represented by generators $d(v_N^2)_\mu$ and $d(i_N^2)_\mu$ at its input, and that the feedback network is noiseless; this latter assumption is rarely justified in practice. The spot noise generators at the input to the complete system can be found by short- and open-circuiting the input terminals. With the input terminals short-circuited, the noise output is determined by $d(v_N^2)_\mu$

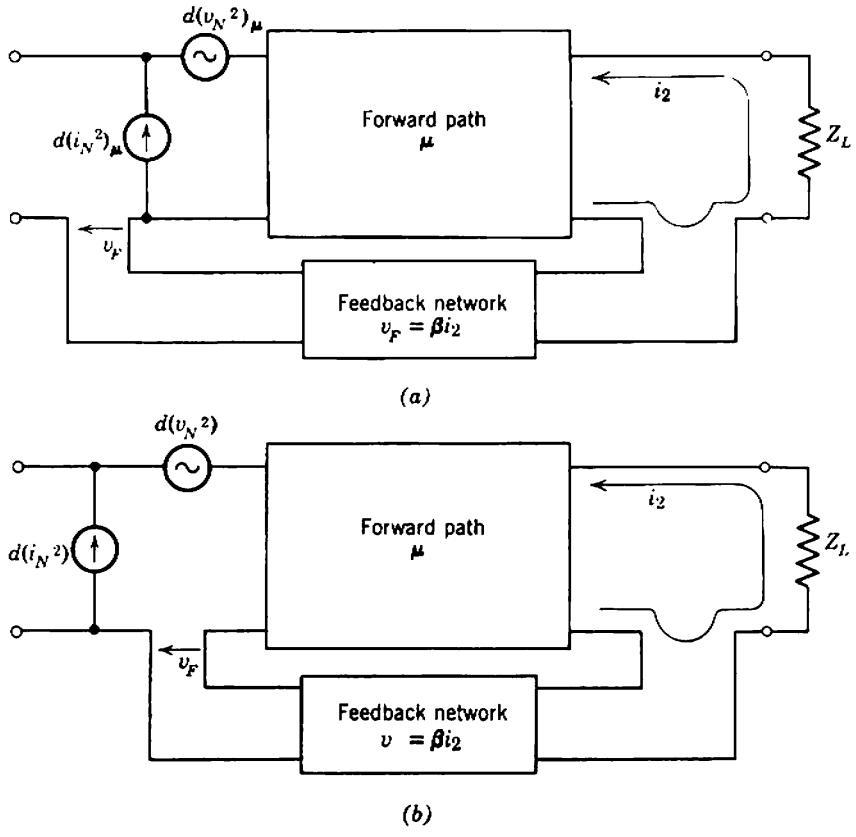


Fig. 10.27 Noise in an ideal series-input/series-output feedback amplifier.

alone, because $d(i_N^2)_\mu$ cannot develop any voltage across the ideal feedback network:

$$v_F = \beta i_2.$$

The noise current output is less with feedback than without because the transfer admittance is reduced by the factor $(1 - \mu\beta)$, but by inspection

$$d(v_N^2) = d(v_N^2)_\mu. \tag{10.75}$$

With the input terminals open-circuited, the noise output is determined by $d(i_N^2)_\mu$ alone, and it is almost trivial to note that

$$d(i_N^2) = d(i_N^2)_\mu. \tag{10.76}$$

The noise current output is the same both with and without feedback because open-circuiting the input terminals open-circuits the feedback loop.

Thus for series-input/series-output feedback the noise generators at the input are the noise generators of the forward path. The same conclusion applies to all other configurations, and ideally feedback has no effect on the noise performance of any amplifier. In practice, this conclusion can be invalidated in either of two ways:

1. If $|A_{fm}|$ is large so that the closed-loop transfer function is small, noise from subsequent stages in the amplifier may be significant. In other words, the principle that feedback does not change the noise performance is true only when the feedback loop encloses all significant sources of noise.
2. In almost all practical amplifiers the feedback network contains resistors that contribute thermal noise.

Practical applications of these general principles are given in Sections 11.4 and 14.1.2.

10.9 NONLINEARITY

The small-signal parameters of an amplifying device depend on the currents and voltages impressed on the device. Accordingly, the small-signal parameters change from point to point on a signal waveform, and the gain changes also. Different portions of the signal waveform are amplified by different amounts and practical amplifiers are said to be nonlinear.

It is shown in Chapter 9 that a useful measure of the nonlinearity in an amplifier is the differential error γ, defined at any point on the signal waveform as

$$\gamma = \frac{TF - TF_Q}{TF_Q}, \tag{10.77}$$

where TF_Q is the transfer function under quiescent conditions. Simple relations exist between the harmonic distortion, low-order intermodulation distortion, and the values γ' and γ'' at the peaks of a signal waveform. The most general method for calculating distortion is the same for both feedback and nonfeedback amplifiers: the transfer functions TF_Q, TF', and TF'' are found from any convenient formulas, substitution into Eq. 10.77 gives the parameters γ' and γ'', and these in turn are substituted into the relevant equations of Chapter 9. An example of this type of calculation is given in Section 11.5.

An alternative method for calculating the differential error in a feedback amplifier is based on Section 10.4.3. Differential error γ is the

fractional change in transfer function at any point on the signal waveform. Equations 10.46 and 10.47 can be rearranged to show that if the differential error in the forward path μ of a feedback amplifier is γ_0 then the differential error in the closed-loop transfer function is

$$\gamma = \frac{\gamma_0}{1 - A_{IQ}(1 + \gamma_0)} \approx \frac{\gamma_0}{1 - A_{IQ}} = \frac{\gamma_0}{F_Q} \tag{10.78}$$

where A_{IQ} and F_Q are, respectively, the loop gain and return difference at the quiescent point. The approximate form applies when γ_0 is small. Thus, the differential error and distortion in a feedback amplifier are approximately $1/F_Q$ of the differential error and distortion in the forward path, provided these are fairly small.

Differential error at frequencies outside the mid-band range depends on changes in both the mid-band transfer function and the singularity pattern. Negative feedback reduces the former component but may reduce or increase the latter depending on the precise form of the closed-loop singularity pattern and the signal frequency. The design technique suggested at the end of Section 10.5.3.4 is particularly helpful in reducing the differential error caused by a shifting singularity pattern, as the dominant poles and zeros of the closed-loop transfer function are determined almost entirely by the passive feedback network, and are therefore fixed.

Chapter 11

Amplifiers with Single-Stage Feedback

The outline of negative feedback theory in Chapter 10 shows that feedback is a circuit technique of great utility. Its most valuable application is in stabilizing the mid-band transfer function of an amplifier to be relatively independent of device parameters, bias conditions, load impedance, and signal level. The price that must be paid for these advantages is, of course, the complication incurred in the dynamic response; regeneration (self-oscillation) occurs in extreme cases.

The simplest feedback amplifiers are those in which feedback is applied to single stages only and a study of practical feedback circuits is most conveniently begun at this level. Single-stage feedback amplifiers usually have the advantage of a simple and easily controlled dynamic response; the vast majority of useful single-stage feedback building blocks have only one or two poles and regeneration cannot possibly occur. Exceptions to this statement do exist, notably some narrow-band tuned amplifiers (which are beyond the scope of this book) and wide-band emitter followers under certain conditions of loading (which can scarcely be considered as simple, designable circuits and are certainly beyond the scope of this chapter). Therefore it is convenient and permissible to consider separately the mid-band performance and small-signal dynamic response of the two single-stage feedback building-blocks introduced in this chapter. The present chapter is concerned with the mid-band performance; Chapters 12 and 13 consider the dynamic response.

The use of single-stage feedback is not merely a low-level introduction to the more sophisticated technique of multistage feedback, but is in itself useful in the design of two types of amplifier:

1. For amplifiers requiring moderately stable transfer function magnitudes (say $\pm 5\%$ per stage), appropriate single-stage feedback amplifiers can be cascaded. There is no regeneration problem and the dynamic response can be predicted to the same order of accuracy as the transfer function magnitude.

2. For amplifiers requiring highly stable transfer function magnitudes (say $\pm 1\%$), over-all feedback can be applied around a number of feedback stages in cascade. The use of feedback stages in the forward path provides effective control over the reference variable μ, and thereby simplifies the design of the over-all feedback loop. Section 14.3 contains details of this technique.

There has been much speculation in the technical literature as to which amplifier configuration (common-emitting-electrode, common-control-electrode, or common-collecting-electrode) is the most basic, and which are merely feedback derivatives. As shown in Chapter 5, the common-emitting-electrode amplifier is unique in that it alone yields useful values for all four transfer functions, and in this practical sense it may be considered as "basic." Because of the phase change between input and output, there are only two possible methods for applying negative feedback that satisfy the following conditions:

(i) the circuitry does not include phase-reversing transformers,
(ii) the magnitude of the closed-loop transfer function can be controlled by the feedback network.

One method is the use of series feedback to stabilize the transfer conductance and the other is the use of shunt feedback to stabilize the transfer resistance. The series- and shunt-feedback stages shown in Figs. 11.1*a* and *b* are much more flexible building-blocks than the common-control-electrode or the common-collecting-electrode stages shown in Figs. 11.1*c* and *d*. These latter configurations can be considered as feedback derivatives of the common-emitter stage, in which the feedback stabilizes current gain and voltage gain, respectively. However, as shown in Chapter 5, their gains have fixed magnitudes of slightly less than unity so there can be no exchange of gain magnitude for gain stability; condition (ii) above does not hold.

The material of this chapter is moderately self-sufficient and does not depend critically on the formal feedback theory of Chapter 10. However, some facility in analyzing feedback circuits is assumed; the transfer and driving-point functions of useful amplifiers are quoted without derivation. These expressions can be verified by normal circuit analysis, but the reader may profit from the use of the superposition technique discussed in Section 10.3.

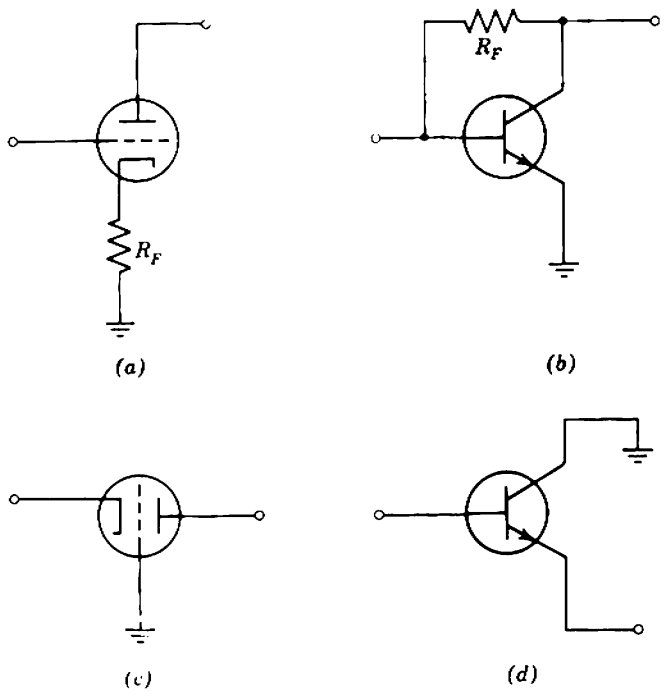


Fig. 11.1 Possible feedback stages: (*a*) series feedback (triode); (*b*) shunt feedback (transistor); (*c*) current feedback (grounded grid); (*d*) voltage feedback (grounded collector).

11.1 ISOLATED SERIES- AND SHUNT-FEEDBACK STAGES

This section begins with a brief but general description of the series- and shunt-feedback stages shown in Fig. 11.2, and goes on to a detailed discussion of the effects of active device parameters.

In the series-feedback stage a signal voltage is developed across R_F by the signal component of current flowing in the emitting electrode. This voltage appears in the input mesh of the circuit, where it (numerically) subtracts from the input voltage to give the voltage applied between the control and emitting electrodes:

$$v_F = -i_3 R_F, \tag{11.1}$$

$$v_1 = v_i + v_F. \tag{11.2}$$

Now, the emitting-electrode current is directly proportional to the collecting-electrode or output current i_o:

$$i_3 = (\text{const}) i_o. \tag{11.3}$$

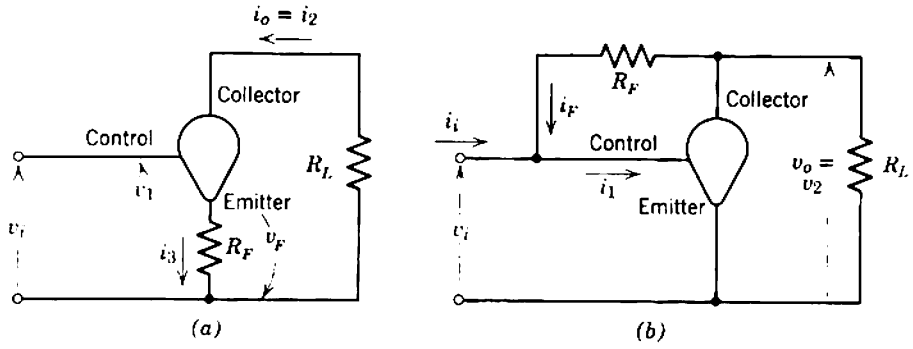


Fig. 11.2 Generalized feedback stages: (*a*) series feedback; (*b*) shunt feedback.

Therefore the feedback voltage is directly proportional to the output current. It follows that if the loop gain is very large (that is, if v_1 tends to zero) the transfer conductance of the stage is

$$G_T = \frac{i_o}{v_i} = \frac{1}{(\text{const})R_F}, \tag{11.4}$$

independent of the mutual conductance g_m of the device. The value of the constant is

- 1 for a triode, a pentode which has its screen bypassed to the cathode, or a field-effect transistor;
- $1 + k$ for a pentode which has its screen bypassed to ground;
- $1/\alpha_N$ for a bipolar transistor.

In the shunt-feedback stage a current proportional to the difference between the input and output voltages is fed back to the input node:

$$i_F = \frac{v_o - v_i}{R_F}, \tag{11.5}$$

and if the voltage gain is large

$$i_F \approx \frac{v_o}{R_F}. \tag{11.6}$$

This feedback current (numerically) subtracts from the input current to give the signal current input to the device:

$$i_1 = i_i + i_F. \tag{11.7}$$

If the loop gain is very large (that is, if i_1 tends to zero), the transfer resistance of the stage is

$$R_T = \frac{v_o}{i_i} = -R_F. \tag{11.8}$$

11.1.1 Vacuum-Tube Series-Feedback Stage

Figure 11.3 shows the equivalent circuit for a vacuum tube with series feedback. The transfer conductance and voltage gain are shown in Section 10.1 to be

$$G_T = \frac{i_o}{v_i} = \frac{g_m}{1 + (R_L/r_A) + (\mu + 1)R_F/r_A}, \tag{11.9}$$

$$A_V = \frac{v_o}{v_i} = -G_T \times R_L, \tag{11.10}$$

and both transfer functions depend on the tube's parameters. The dependence on μ and r_A, however, is incidental and, in principle, can be eliminated without reference to the feedback:

(i) If μ is much greater than unity, then

$$\frac{\mu + 1}{r_A} \approx \frac{\mu}{r_A} = g_m. \tag{11.11}$$

(ii) If R_L is chosen such that

$$R_L \ll r_A, \tag{11.12}$$

the second term in the denominator of Eq. 11.9 vanishes.

Thus Eq. 11.9 can be reduced to

$$G_T = \frac{g_m}{1 + g_m R_F} = \frac{1}{R_F}\left(\frac{1}{1 + 1/g_m R_F}\right) \approx \frac{1}{R_F}. \tag{11.13}$$

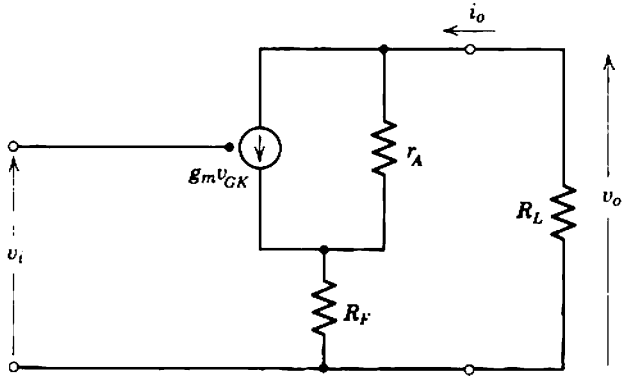


Fig. 11.3 Equivalent circuit of a vacuum-tube series-feedback stage.

As shown in Eq. 11.4, the demanded transfer conductance of a series-feedback vacuum-tube stage is $1/R_F$. By comparison of Eq. 11.13 with Eq. 10.36 the loop gain is identified as

$$A_l = -g_m R_F, \tag{11.14}$$

which is, of course, the voltage gain between the input electrodes of the tube (the grid and cathode) and the feedback resistor as a load. The mutual conductance g_m is identified as the reference variable μ.

The loop gain required for given stability in G_T against variation of g_m can be found from Section 10.4.3. For example, suppose the tolerance on g_m is

$$0.6\, g_{mA} \leq g_m \leq 1.5\, g_{mA}$$

and that the required tolerance on G_T is $\pm 5\%$. It follows from Eq. 10.48 that the nominal loop gain must be 11.66 to give 5% fall in G_T when g_m falls to $0.6\, g_{mA}$. A nominal loop gain of only 6 is sufficient to give 5% rise in G_T when g_m rises to $1.5\, g_{mA}$, but the larger value 11.66 must be used to hold G_T within the specified $\pm 5\%$. Table 11.1 lists the nominal values of $|A_l|$ $(= g_{mA} R_F)$ required for various degrees of stability.

Table 11.1 Nominal Loop Gain Required for Specified Tolerance on Transfer Function, with Tolerance on g_m as Parameter

Specified Tolerance on Transfer Function	Tolerance on g_m	Required Nominal Loop Gain $\lvert A_l \rvert$	Actual Tolerance on Transfer Function
$\pm 5\%$	60 to 150%	11.66	95.0 to 102.5%
	80 to 125%	3.72	95.0 to 104.5%
$\pm 10\%$	60 to 150%	5.00	90.0 to 108.3%
	80 to 125%	1.25	90.0 to 109.7%
$\pm 25\%$	60 to 150%	1.00	75.0 to 120.0%
	80 to 125%	0.00	80.0 to 125.0%

The uncertainty in G_T due to the finite value of μ is negligible, provided that the nominal value of μ is 10 or more, because μ is a relatively stable and reproducible parameter. However, R_L must not be so large that appreciable uncertainty is introduced through the dependence on r_A.

If the tube is a pentode with its screen bypassed to ground rather than

to its cathode, the screen component of signal current flows in R_F. Equation 11.13 becomes

$$G_T = \left[\frac{1}{(1 + k)R_F}\right]\left[\frac{1}{1 + 1/g_m(1 + k)R_F}\right] \approx \frac{1}{(1 + k)R_F}. \quad (11.15)$$

11.1.2 Vacuum-Tube Shunt-Feedback Stage

Figure 11.4*a* shows the equivalent circuit for a vacuum tube with shunt feedback. It is shown in Section 11.2.2.2 that a pentode is much more satisfactory than a triode as a shunt-feedback amplifier. The anode resistance of a pentode is much larger than any likely load and the equivalent circuit therefore reduces to Fig. 11.4*b*. The transfer and driving-point functions are:

$$R_T = \frac{v_o}{i_i} = -\frac{R_F}{1 + \dfrac{R_F + R_L}{g_m R_F R_L}} \approx -R_F, \quad (11.16)$$

$$A_I = \frac{i_o}{i_i} = -\frac{R_T}{R_L}, \quad (11.17)$$

$$A_V = \frac{v_o}{v_i} = -g_m\left(\frac{R_F \times R_L}{R_F + R_L}\right), \quad (11.18)$$

$$G_T = \frac{i_o}{v_i} = -\frac{A_V}{R_L}, \quad (11.19)$$

$$R_i = \frac{v_i}{i_i} \approx \frac{R_F + R_L}{g_m \times R_L}. \quad (11.20)$$

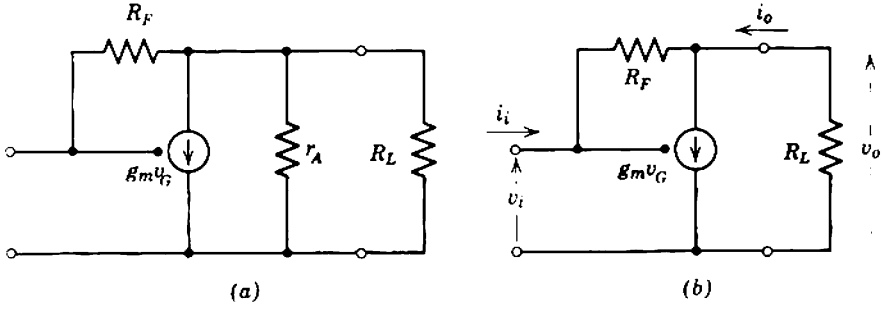


Fig. 11.4 Equivalent circuit of a vacuum-tube shunt-feedback stage: (*a*) complete; (*b*) simplified for a pentode.

As expected, the inclusion of R_F stabilizes the transfer resistance to approximately $-R_F$ despite variations in the tube parameters and the load resistance. In contrast, R_F has no effect on the transfer functions defined in terms of an input voltage (that is, G_T and A_V) save that, as evidenced by Eq. 11.18, the effective load resistance is the parallel combination of R_L and R_F.

Equation 11.16 can be compared with Eq. 10.36, and the loop gain is identified as

$$A_l = -g_m\left(\frac{R_F \times R_L}{R_F + R_L}\right), \tag{11.21}$$

which is, rather surprisingly, the voltage gain of the stage. The explanation is that, if the mutual conductance g_m is reduced to zero, all the output voltage is applied to the tube grid via R_F; the feedback factor is therefore unity and the loop gain is equal to the gain of the forward path provided that the source resistance is infinite so that no voltage dividing action occurs between R_F and the source. There is, then, an apparent anomaly; Eq. 11.16 applies for all source resistances, whereas the loop gain appears to depend on the source resistance. This and several related topics are discussed in Section 11.3; they are sources of complication that appear only if a circuit is improperly used.

In the present case the anomaly can be avoided by proper choice of source resistance. Because the stable transfer function of a shunt-feedback stage is

$$R_T = \frac{\text{output voltage}}{\text{input current}},$$

the stage should be fed from an approximate current source. Such a source has a high output resistance (ideally approaching infinity) and there is no voltage dividing action between R_F and the source. It is only when a shunt-feedback stage is fed from a low-impedance (approximate voltage) source that the complication appears. Although this may seem obvious and somewhat trivial, the importance of proper use of feedback stages cannot be too strongly emphasized.

As in the series-feedback stage, a table may be prepared to give the nominal value of $|A_l|$ $[= g_{mA}R_FR_L/(R_F \times R_L)]$ required for a given degree of stability in R_T against variation of g_m; this table is identical with Table 11.1. For a given value of closed-loop transfer resistance ($\approx -R_F$), the loop gain increases as R_L is increased. The loop gain is usually very large indeed (because the voltage gain of a pentode is high) and the uncertainty in R_T is therefore minute.

11.1.3 Transistor Series-Feedback Stage

Figure 11.5 shows the equivalent circuit for a transistor with series feedback. The transfer and driving-point functions are

$$G_T = \frac{i_o}{v_i} = \frac{\alpha_N}{(1 - \alpha_N)r_B + r_E + R_F} \approx \frac{\alpha_N}{r_B/\beta_N + r_E + R_F} \approx \frac{1}{R_F}, \quad (11.22)$$

$$A_V = \frac{v_o}{v_i} = -G_T \times R_L, \quad (11.23)$$

$$A_I = \frac{i_o}{i_i} = \beta_N, \quad (11.24)$$

$$R_T = \frac{v_o}{i_i} = -A_I \times R_L, \quad (11.25)$$

$$R_i = \frac{v_i}{i_i} = r_B + \frac{r_E + R_F}{1 - \alpha_N} \approx r_B + \beta_N(r_E + R_F). \quad (11.26)$$

As expected, the inclusion of R_F stabilizes the transfer conductance at approximately $1/R_F$ but has no effect on the transfer functions defined in terms of an input current. Equation 11.22 can be manipulated into the form of Eq. 10.36, namely,

$$G_T = \frac{\alpha_N}{R_F}\left[\frac{1}{1 + (r_B/\beta_N + r_E)/R_F}\right] \quad (11.27)$$

when (α_N/R_F) is identified as the closed-loop transfer conductance expected for infinite loop gain, and the loop gain is

$$A_l = -\frac{R_F}{r_B/\beta_N + r_E}. \quad (11.28)$$

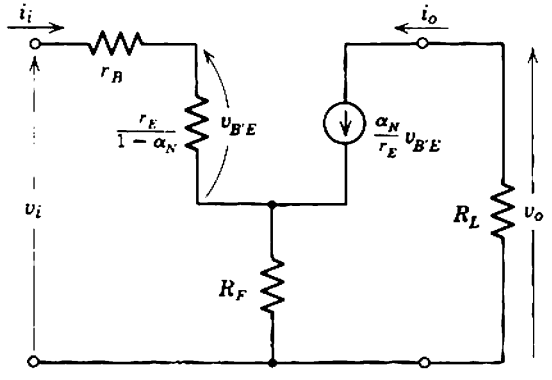


Fig. 11.5 Equivalent circuit of a transistor series-feedback stage.

Two classes of uncertainty arise in Eq. 11.22 from uncertainty in the transistor parameters. The first is due to α_N; this uncertainty exists even when the loop gain is large, but it is a small uncertainty because α_N is predictable and very close to unity. For many purposes α_N in Eq.11.22 can be taken as unity; a working rule for the most precise calculations is

$$\alpha_N = 0.99 \text{ for high-gain transistors,}$$

$$\alpha_N = 0.98 \text{ for medium-gain transistors,}$$

$$\alpha_N = 0.96 \text{ for low-gain transistors.}$$

The uncertainty in loop gain results principally from the uncertainty in r_E which varies with both emitter current and temperature; however, r_E follows the known law

$$r_E = \frac{kT}{qI_E},$$

so that its variation is predictable. The fractional uncertainty in r_B/β_N is relatively great, but because r_E can always be made much greater than r_B/β_N this need not be a significant source of error. It is not really practicable to produce a table giving the nominal value of loop gain required for specified stability in G_T, because there are two sources of uncertainty. Rather, common sense should be used in choosing the values of I_E and R_F when designing a series-feedback stage of specified transfer conductance.

Finally, there is a source of error in G_T not apparent in Eq. 11.22. Because an approximate equivalent circuit is used, there is the implication that R_L must not be too large. The error due to nonzero R_L is

$$\frac{G_{T(\text{true})}}{G_{T(\text{apparent})}} = \frac{1}{1 + R_L/r_C\,(1 - A_l)}. \tag{11.29}$$

This error is small until R_L becomes comparable with r_C, and is negligible if a series-feedback stage has a small load resistance.

11.1.4 Transistor Shunt-Feedback Stage

Figure 11.6*a* shows the simplified equivalent circuit for a transistor with shunt feedback. Under unusual circumstances, R_L and R_F may be so large that there is some doubt as to the validity of this equivalent circuit. Should such doubt arise, R_L and R_F in the following equations may be replaced by R'_L and R'_F which represent the parallel combinations of R_L with r_C, and R_F with $\beta_N r_C$ in the complete equivalent circuit of Fig. 11.6*b*.

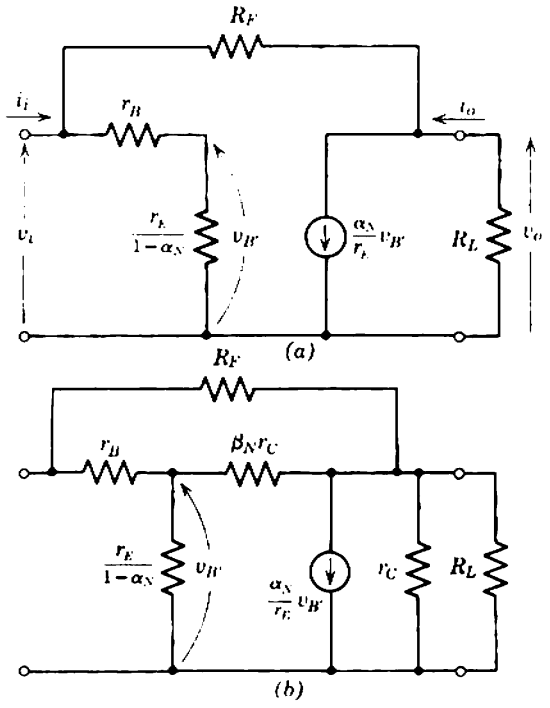


Fig. 11.6 Equivalent circuit of a transistor shunt-feedback stage: (*a*) simplified; (*b*) complete.

This is an excellent approximation, because r_B is small compared with R_F and $\beta_N r_C$. The transfer and driving-point functions are:

$$R_T = \frac{v_o}{i_i} = -\frac{R_F}{1 + \dfrac{R_F + R_L}{\beta_N R_L}} \approx -R_F, \tag{11.30}$$

$$A_I = \frac{i_o}{i_i} = -\frac{R_T}{R_L}, \tag{11.31}$$

$$A_V = \frac{v_o}{v_i} = -\left(\frac{\alpha_N}{r_B/\beta_N + r_E}\right)\left(\frac{R_F \times R_L}{R_F + R_L}\right), \tag{11.32}$$

$$G_T = \frac{i_o}{v_i} = -\frac{A_V}{R_L}, \tag{11.33}$$

$$R_i = \frac{v_i}{i_i} \approx \left(\frac{r_B}{\beta_N} + r_E\right)\left(1 + \frac{R_F}{R_L}\right). \tag{11.34}$$

Strictly, R_i is the parallel combination of $R_F/(1 - A_V)$ and the input resistance $(r_B + \beta_N r_E)$ of the transistor itself; this reduces approximately to the value given in Eq. 11.34 (see Section 11.2.2.2). As in the vacuum-

tube case, the inclusion of R_F stabilizes the transfer resistance, but has no stabilizing effect on the transfer functions defined in terms of an input voltage. The loop gain is identified as

$$A_l = -\beta_N\left(\frac{R_L}{R_F + R_L}\right). \tag{11.35}$$

This might have been written down at sight; β_N is the current gain of the transistor, and $R_L/(R_F + R_L)$ is the fraction of the collector current that flows back to the input via R_F.

Two apparent anomalies exist in connection with the loop gain. The first involves the source resistance and is similar to that for a vacuum-tube shunt-feedback stage; it is discussed in Section 11.3. The second is that the careful derivation of Eq. 10.45 in Section 10.4.1 yields

$$R_T = -\frac{R_F}{1 + \dfrac{(R_F + R_L)[R_F + r_B + (\beta_N + 1)r_E]}{\beta_N R_F R_L}} \tag{11.36}$$

and the extra term $[r_B + (\beta_N + 1)r_E]$ is the input resistance of the transistor itself. Equation 11.36 reduces to Eq. 11.30 only if

$$R_F \gg r_B + (\beta_N + 1)r_E. \tag{11.37}$$

With reasonable values of loop gain the error in Eq. 11.30 is quite small even if Eq. 11.37 is satisfied by only a factor of two or three. Nevertheless, it is prudent to design a shunt-feedback stage so that Eq. 11.37 is satisfied by a larger amount, five at least, as the high-frequency analysis of Section 13.5.2.2 is very much simplified.

The nominal value of $|A_l|$ $[= \beta_{NA}R_L/(R_F + R_L)]$ necessary to give a specified tolerance on R_T can be calculated from the theory of Section 10.4.3 when the spread in the value of β_N is known. Table 11.2 gives some useful values.

Table 11.2 Nominal Loop Gain Required for Specified Tolerance on Transfer Resistance, with Tolerance on β_N as Parameter

Specified Tolerance on R_T	Tolerance on β_N	Required Nominal Loop Gain $\lvert A_l \rvert$	Actual Tolerance on R_T
±5%	50% to ∞	20	95.5 to 105.0%
	67 to 200%	9.5	95.5 to 105.0%
±10%	50% to ∞	10	91.7 to 110.0%
	67 to 200%	4.5	91.7 to 110.0%
±25%	50% to ∞	4	83.3 to 125.0%
	67 to 200%	1.5	83.3 to 125.0%

11.1.5 Field-Effect Transistor Stages

Field-effect transistors are undergoing rapid technological development, but the trend is to smaller gate current, larger pinched-off drain resistance, and larger mutual conductance. Any of the resistances r_{GS}, r_{DG}, and r_{DS} in the equivalent circuits of Fig. 11.7 may be very large indeed.

The transfer conductance of a series-feedback f.e.t. stage is very closely

$$G_T = \frac{i_o}{v_i} = \frac{g_m}{1 + g_m R_F} \approx \frac{1}{R_F}, \tag{11.38}$$

and the loop gain is identified as

$$A_l = -g_m R_F. \tag{11.39}$$

Table 11.1 may be used to find the nominal value of $|A_l|$ $(= g_{mA} R_F)$ necessary for a required degree of stability. The voltage gain is

$$A_V = \frac{v_o}{v_i} = -G_T \times R_L. \tag{11.40}$$

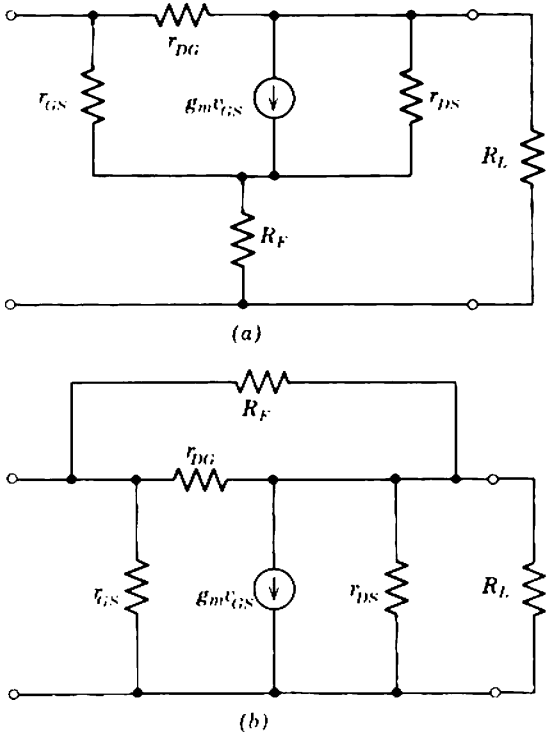


Fig. 11.7 Equivalent circuits for f.e.t. feedback stages: (*a*) series feedback; (*b*) shunt feedback.

The input resistance is the parallel combination of two components R_{i1} and R_{i2} which are due, respectively, to r_{GS} and r_{DG}. From the feedback theory of Section 10.7, the former component is

$$R_{i1} = r_{GS}(1 + |A_l|), \tag{11.41}$$

whereas the latter follows from the general Miller effect theory of Section 7.5.2.1 as

$$R_{i2} = \frac{r_{DG}}{1 + |A_V|}. \tag{11.42}$$

Either or both of R_{i1} and R_{i2} may be significant; both are likely to be several tens of megohms or higher.

In an f.e.t. shunt-feedback stage, r_{DS} and r_{DG} are likely to be much larger than R_L and R_F, respectively, and can be neglected; r_{GS} is very large also. The performance is identical with that of a pentode shunt-feedback stage, the relevant transfer and driving-point functions being:

$$R_T = \frac{v_o}{i_i} = -\frac{R_F}{1 + \dfrac{R_F + R_L}{g_m R_F R_L}} \approx -R_F, \tag{11.43}$$

$$A_l = -g_m\left(\frac{R_F \times R_L}{R_F + R_L}\right), \tag{11.44}$$

$$A_V = \frac{v_o}{v_i} = -g_m\left(\frac{R_F \times R_L}{R_F + R_L}\right), \tag{11.45}$$

$$R_i = \frac{v_i}{i_i} = \frac{R_F}{1 - A_V} \approx \frac{R_F + R_L}{g_m R_L}. \tag{11.46}$$

Table 11.1 gives the nominal value of $|A_l|$ $[= g_{mA} R_F R_L/(R_F + R_L)]$ required for various degrees of stability.

11.2 CASCADED FEEDBACK STAGES

The effects of feedback on the transfer and driving-point functions of an amplifier are discussed fairly generally in Sections 10.6 and 10.7, and some specific results for series- and shunt-feedback stages are given in Section 11.1. In summary, the important results are:

SERIES-FEEDBACK STAGE

(i) is a voltage-in-to-current-out amplifier;
(ii) has a high input impedance, yet should be fed from a low-impedance (ideally zero) source;
(iii) has a high output impedance, yet should work into a low-impedance (ideally zero) load.

SHUNT-FEEDBACK STAGE

(i) is a current-in-to-voltage-out amplifier;
(ii) has a low input impedance, yet should be fed from a high-impedance (ideally infinite) source;
(iii) has a low output impedance, yet should work into a high-impedance (ideally infinite) load.

Intuitively, therefore, it is expected that the most satisfactory method for cascading series- and shunt-feedback stages is to alternate them. The low input impedance of a shunt stage is a suitable load for a series stage, and the high input impedance of a series stage is a suitable load for a shunt stage. Similarly, the low output impedance of a shunt stage is a suitable source for a series stage, and the high output impedance of a series stage is a suitable source for a shunt stage. The alternate cascade is discussed extensively in Sections 11.2.1 and 11.2.2.

Although the arrangement of series- and shunt-feedback stages in alternation is in fact the most satisfactory, it is not nearly so well known as the cascades of similar stages. It is therefore worth discussing these cascades in some detail, and pointing out their shortcomings. The conclusion reached is that the alternate cascade is so far superior to a cascade of either series-feedback stages or shunt-feedback stages that it is the only arrangement of the three worth considering for practical amplifier designs.

11.2.1 General Theory of the Alternate Cascade

The most important feature of the alternate cascade is that there is a gross difference in magnitude between the output resistance of one stage and the input resistance of the following stage. The input resistance of a shunt-feedback stage is so low compared with the output resistance of a series-feedback stage that the transfer conductance of the series stage is very close to the short-circuited transfer conductance. This short-circuited transfer conductance can be calculated readily, and is

$$G_T \approx \frac{1}{R_F}. \tag{11.47}$$

Similarly, the transfer resistance of a shunt-feedback stage within an alternate cascade is very nearly the open-circuited transfer resistance. This too can be calculated readily, and is

$$R_T \approx -R_F. \tag{11.48}$$

Expressed another way, there is a gross impedance mismatch between stages, and this minimizes interaction. *The transfer function of a stage within an alternate cascade is very nearly the transfer function in isolation.*

The over-all transfer function of an alternate cascade is the product of the individual stage transfer conductances and transfer resistances. It is possible to construct an amplifier which has any one of its over-all transfer functions stable by suitably choosing the input and output stage types, and whether the total number of stages is even or odd. Thus, in Figs. 11.8*a* to *d*, respectively,

(a) odd number of stages

$$G_T = G_{T1} \times R_{T2} \times \cdots \times R_{T(n-1)} \times G_{Tn}; \tag{11.49}$$

(b) even number of stages

$$A_V = G_{T1} \times R_{T2} \times \cdots \times G_{T(n-1)} \times R_{Tn}; \tag{11.50}$$

(c) even number of stages

$$A_I = R_{T1} \times G_{T2} \times \cdots \times R_{T(n-1)} \times G_{Tn}; \tag{11.51}$$

(d) odd number of stages

$$R_T = R_{T1} \times G_{T2} \times \cdots \times G_{T(n-1)} \times R_{Tn}. \tag{11.52}$$

It is worth emphasizing that the significant signal that flows from a series stage into a shunt stage is a current, not a voltage. The signal voltage between the two stages is of no consequence; its value varies between wide limits because it depends on the unstable input resistance of the shunt-feedback stage. Correspondingly, the significant signal that flows from

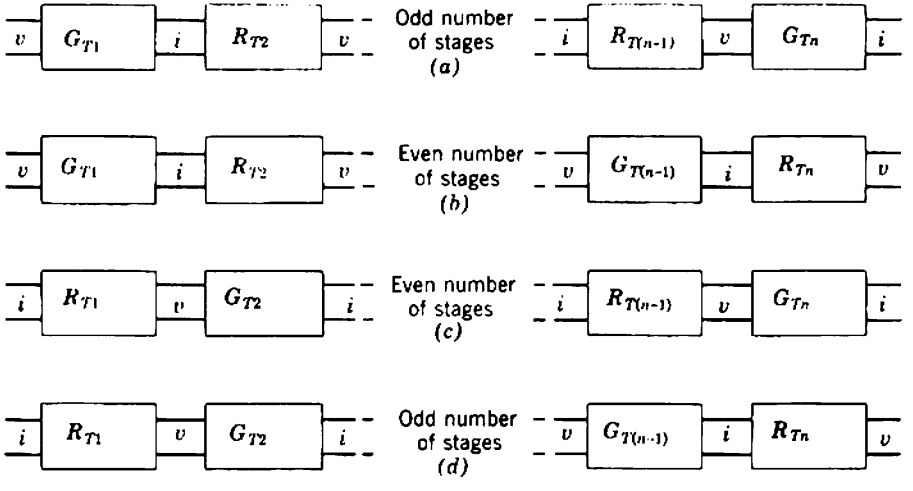


Fig. 11.8 The alternate cascade: (*a*) stable G_T; (*b*) stable A_V; (*c*) stable A_I; (*d*) stable R_T. Notice the pertinent signal between stages indicated by *v* or *i*.

a shunt stage into a series stage is a voltage; the current is unstable because it depends on the unstable input resistance of the series-feedback stage. Thus the alternate cascade is unique in that only one transfer function is stable between any two points. In any other type of amplifier, all transfer functions of any stage are equally stable (or unstable)—see Section 5.5.1 for stages without feedback, and Sections 11.2.3 and 11.2.4 for cascaded series- or shunt-feedback stages. It is also worth pointing out that, within an alternate cascade, the voltage gain of a series-feedback stage and the current gain of a shunt-feedback stage are small. For a series-feedback stage

$$A_V = -G_T \times R_L = -G_T \times R_{i(\text{shunt})},$$

which is small because $R_{i(\text{shunt})}$ is small. For a shunt-feedback stage

$$A_I = \frac{-R_T}{R_L} = \frac{-R_T}{R_{i(\text{series})}},$$

which is small because $R_{i(\text{series})}$ is large.

11.2.2 Gain Limits of the Alternate Cascade

When feedback amplifier stages are cascaded, there are definite upper limits to the gain per stage that can be obtained with a given stability for specified tolerances on the active device parameters. These limits involve not only the limits to the values of G_T and R_T of the individual stages, but also the slight interaction that does occur between stages. The situation is somewhat similar to that of the transistor amplifiers discussed in Sections 5.3 and 5.5; for given stability, more gain can be obtained from an isolated stage than from a stage within a cascade.

Two fundamental limits which can never be exceeded in any circuit whatsoever are that the voltage and current gains of any stage must be less than the voltage and current amplification factors of the device. These limits are more of academic interest than practical importance, but are implicit in the equations. Two practical limits are discussed in the following sections—loading of a shunt stage by the input resistance of the following series stage, and shunting of the signal path by the biasing resistors.

11.2.2.1 Limit Due to Recombination

With devices that have appreciable recombination (notably the bipolar transistor), an upper limit to the gain results because of interaction between stages. Such devices have a finite input resistance which is inversely

proportional to their transfer conductance; for a transistor it follows from Eqs. 11.22 and 11.26 that

$$R_i = \frac{\beta_N}{G_T}. \tag{11.53}$$

The input resistance of a series-feedback stage in a long cascade constitutes a load on the preceding shunt-feedback stage, and therefore sets an upper limit to the transfer resistance that can be realized with a given degree of stability. The current gain of the two stages, which is the product of the transfer resistance and transfer conductance of the first and second stages respectively, is therefore limited also. Specifically, for the two transistor stages shown in Fig. 11.9,

$$R_{L(\text{shunt})} = R_{i(\text{series})} \approx \beta_{N2} R_{F2} \tag{11.54}$$

if all biasing resistors in shunt with the signal path are neglected. Now, the condition for R_T of the shunt-feedback stage to be stable follows from Eq. 11.35 as

$$|A_l| = \beta_{NA1}\left[\frac{R_{L(\text{shunt})}}{R_{F1} + R_{L(\text{shunt})}}\right] \gg 1$$

and substituting from Eq. 11.54 gives

$$\frac{\beta_{NA1}\beta_{NA2}R_{F2}}{R_{F1} + \beta_{NA2}R_{F2}} \gg 1;$$

that is,

$$\frac{\beta_{NA1}\beta_{NA2}}{R_{F1}/R_{F2} + \beta_{NA2}} \gg 1$$

or

$$\frac{\beta_{NA1}\beta_{NA2}}{A_I + \beta_{NA2}} \gg 1, \tag{11.55}$$

where

$$A_I = R_{T1} \times G_{T2} \approx \frac{R_{F1}}{R_{F2}}$$

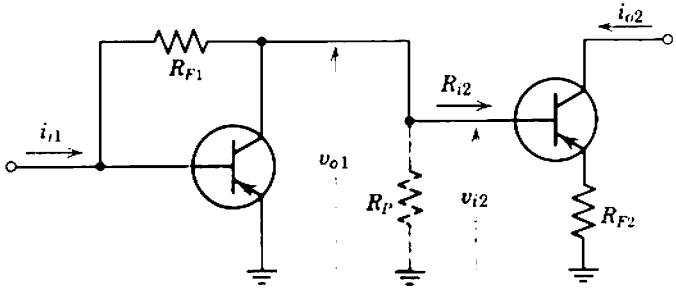


Fig. 11.9 Limit to the stable current gain of two transistor stages.

is the current gain of the two stages. Typically, A_I of the two stages is much greater than β_{NA2}, so that Eq. 11.55 reduces to

$$\beta_{NA1}\beta_{NA2} \gg A_I. \tag{11.56}$$

Table 11.3 lists the maximum current gains that can be obtained with a given degree of stability for various production tolerances. Often the same type of transistor is used in both stages, so that β_{NA1} and β_{NA2} are the same. Notice that the limits on over-all A_I given in Table 11.3 make no allowance for the tolerance on G_T of the series-feedback stage; in addition to choosing the ratio of R_{F1} to R_{F2} that gives the specified maximum current gain, it is necessary to make the absolute value of R_{F2} large enough to stabilize G_{T2} significantly more accurately than the specified tolerance on A_I.

Notice too that in a practical circuit which includes biasing resistors, the load resistance for the shunt-feedback stage is less than the input resistance of the series-feedback stage:

$$R_{L(\text{shunt})} = \frac{R_{i(\text{series})} \times R_P}{R_{i(\text{series})} + R_P}, \tag{11.57}$$

Table 11.3 Maximum Nominal Current Gain Available with Specified Tolerance from a Transistor Shunt-Feedback Stage Followed by a Transistor Series-Feedback Stage

Specified Tolerance on $A_{I(\max)}$	Tolerance on β_N	Required Nominal $\frac{R_{F1} \times G_{T2}}{\beta_{NA1}\beta_{NA2}}$	Resultant Nominal $\frac{A_{I(\max)}}{\beta_{NA1}\beta_{NA2}}$	Actual Tolerance on $A_{I(\max)}$
±5%	50% to ∞	0.018	0.018	95 to 102%
	67 to 200%	0.043	0.041	95 to 103%
±10%	50% to ∞	0.038	0.037	90 to 106%
	67 to 200%	0.096	0.088	90 to 107%
±25%	50% to ∞	0.125	0.111	75 to 113%
	67 to 200%	0.355	0.262	75 to 124%

The difference between the third and fourth columns is due to the difference in magnitude between the transfer resistance R_T of a shunt-feedback stage and its feedback resistor R_F:

$$A_I = R_{T1} \times G_{T2} \approx R_{F1} \times G_{T2}.$$

This distinction is ignored in the text.

where R_P represents the parallel combination of all biasing resistors. This further reduces the transfer resistance that can be realized from the shunt stage with a given degree of stability and, correspondingly, reduces the current gain that can be realized from the two stages. The biasing resistors are essentially parasitic in nature and reduce the gain below the ideal limit. This is in direct contrast to the nonfeedback stages considered in Section 5.5.1 in which the biasing resistors form an essential part of the load, and are used to stabilize the gain. It is not practicable to give a general expression for the reduction of gain due to loading by the biasing resistors; each case is best considered on its own merits. As a guiding rule, half the ideal gain can be realized fairly readily, and more than 90% of the ideal gain can be obtained with sophisticated circuit techniques such as *bootstrapping* (discussed in Section 11.6.1.1). It is most important that loading by the input resistance of the following series-feedback stage and the biasing resistances be taken into account in the design of a transistor shunt-feedback stage. The gain of a multistage transistor feedback amplifier is almost always limited by the loading on the shunt-feedback stages.

There is no corresponding limit (due to loading) on the voltage gain that can be realized from a transistor series-feedback stage followed by a shunt stage. In all but the most unusual cases, the input resistance of the shunt-feedback stage is so low that it constitutes a virtual short-circuit load for the series-feedback stage, as required for ideal operation. The upper limit to G_T of a series-feedback stage that can be realized with a given stability depends principally on the quiescent emitter current (Eq. 11.28), although R_i may become undesirably small if G_T is raised too far (Eq. 11.53). The upper limit to R_T of the shunt-feedback stage is set by the external load resistance (Eq. 11.35). Under favorable circumstances the product of G_T and R_T (i.e., the voltage gain of the two stages) can be much larger than the current-gain limits given in Table 11.3.

Recombination is negligible in vacuum tubes and field-effect transistors and therefore the input resistance of vacuum-tube and f.e.t. series-feedback stages is virtually infinite. Consequently, there is no theoretical limit (due to loading) on the current gain that can be realized from a shunt-feedback stage followed by a series-feedback stage. Also, there is no limit (due to loading) on the voltage gain that can be realized from a series stage followed by a shunt stage, because the small input resistance of a shunt stage is a desirable load for a series stage. A shunt-feedback stage, however, is loaded by biasing resistors, so there is a practical limit to the transfer resistance that can be realized with a given degree of stability. Loading by biasing resistors must be taken into account in the design of vacuum-tube or f.e.t. shunt-feedback stages.

11.2.2.2 Limit Due to Coupling Efficiency

Although there is no theoretical limit other than device amplification factor to the voltage gain that can be realized from a series-feedback stage followed by a shunt-feedback stage, a practical limit can result from imperfect coupling between the stages. In any practical amplifier, biasing resistors represented by R_P in Fig. 11.10 are in shunt with the signal path, and the input current to the shunt stage is less than the output current of the series stage. The coupling efficiency η is given by

$$\eta = \frac{i_{i(\text{shunt})}}{i_{o(\text{series})}} = \frac{R_P}{R_P + R_{i(\text{shunt})}}, \tag{11.58}$$

and the voltage gain of the two stages is

$$A_V = G_{T1} \times \eta \times R_{T2}. \tag{11.59}$$

The coupling efficiency and gain are raised by increasing the ratio of R_P to $R_{i(\text{shunt})}$. As when a series stage follows a shunt stage, the biasing resistors are essentially parasitic and reduce the gain.

The input resistance of a transistor shunt-feedback stage is given by Eq. 11.34 and is very small (typically, a few hundred ohms). This low input resistance results principally from the very high mutual conductances of a transistor, α_N/r_E. Often η is as large as 0.95 and can be neglected in calculating the gain of a transistor amplifier. Also, since η is close to unity, any error in the value of $R_{i(\text{shunt})}$ results only in a very small error in the value of η. It is therefore permissible to use an approximate expression such as Eq. 11.34 even in a precise calculation of gain.

Vacuum tubes and field-effect transistors have much smaller mutual conductances than transistors, and their shunt-feedback stages have much higher input resistances. Coupling efficiency can therefore become an

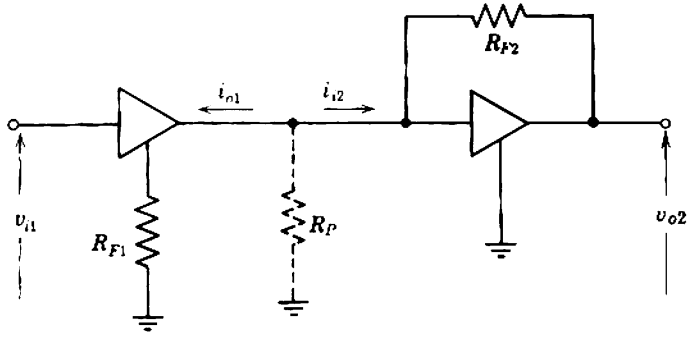


Fig. 11.10 Coupling efficiency between series- and shunt-feedback stages.

appreciable correction. Further, $R_{i(\text{shunt})}$ has a large production tolerance because it depends on g_m (Eq. 11.20) and, consequently, there is some uncertainty in the value of η. Table 11.4 lists the tolerance on various values of η for given tolerances on g_m. In general, the tolerance on η can be reduced only by reducing the value of $R_{i(\text{shunt})}$. Other things being constant, this involves reducing R_F of the shunt stage and, consequently, reducing the voltage gain of the two stages. Stability can be achieved only at the price of a reduction in gain; coupling efficiency therefore sets a practical limit to the stable gain of cascaded vacuum-tube or f.e.t. feedback stages.

Pentodes are much more satisfactory than triodes as shunt-feedback stages because, for the same transfer resistance, their input resistance is much smaller. From the discussion of generalized Miller effect in Section 7.5.2.1, the input resistance of a shunt-feedback stage is the parallel combination of $R_F/(1 - A_V)$ and the input resistance of the device itself. Invariably, the former component dominates for all devices and the input resistance is

$$R_i \approx \frac{R_F}{1 - A_V} \approx -\frac{R_F}{A_V}. \tag{11.60}$$

Equation 11.60 reduces to Eqs. 11.20, 11.34, and 11.66 in the special cases of vacuum tubes, bipolar and field-effect transistors. The voltage gain of a pentode is much higher than that of a triode, and the input resistance of a pentode stage is therefore much the less. Triode shunt-feedback stages

Table 11.4 Nominal Coupling Efficiency Required to Achieve a Specified Tolerance, with the Tolerance on R_i (= Tolerance on $1/g_m$) as Parameter

Specified Tolerance on η	Tolerance on R_i (or $1/g_m$)	Required Nominal Ratio R_i/R_P	Resultant Nominal Value of η	Actual Tolerance on η
±5%	67 to 167%	0.086	0.920	95 to 103%
	80 to 125%	0.267	0.790	95 to 105%
±10%	67 to 167%	0.200	0.833	90 to 106%
	80 to 125%	0.800	0.555	90 to 110%
±25%	67 to 167%	1.000	0.500	75 to 120%

In a cascade of shunt-feedback stages (Section 11.2.4), the third column of this table is ratio of R_i to R_S required to define the input current with specified tolerance.

are unsatisfactory and should not be used. In low-noise applications, in which the partition noise of a pentode would be intolerable, two triodes connected as a *cascode* (see Problem 8.1) can be used in place of a pentode. In contrast, there is little to choose between triodes and pentodes as series-feedback stages. If anything, the triode has a slight advantage because the g_m of a triode at a specified cathode current tends to be higher than the g_m of a pentode due to the division of current between the anode and screen.

It is worth pointing out that the input resistance of a vacuum-tube shunt-feedback stage may be so high that it does not constitute a virtual short-circuit load for the preceding series-feedback stage. It may be necessary to take R_L into account in the calculation of G_T if the series-feedback stage is a triode—Eq. 11.9 rather than Eq. 11.13.

11.2.3 Cascaded Series-Feedback Stages*

Series feedback stabilizes the transfer conductance i_o/v_i of an amplifier stage. In the cascade of series-feedback stages shown in Fig. 11.11, the output current from the first device produced by a given input voltage can be stabilized to any required degree by the application of sufficient series feedback. Similarly, the output current from the second device can be stabilized if its input voltage is known. This input voltage to the second device is the output voltage from the first device, and is given by

$$\frac{v_{i2}}{v_{i1}} = \frac{v_{o1}}{v_{i1}} = A_{V1} = -G_{T1} \times R_{L1}. \tag{11.61}$$

Thus, the input voltage to the second stage depends on the load resistance R_{L1}. In general, R_{L1} has three components:

(i) the collecting-electrode supply resistor for the first device,
(ii) the control-electrode supply resistor for the second device,
(iii) the input resistance R_{i2} of the second device.

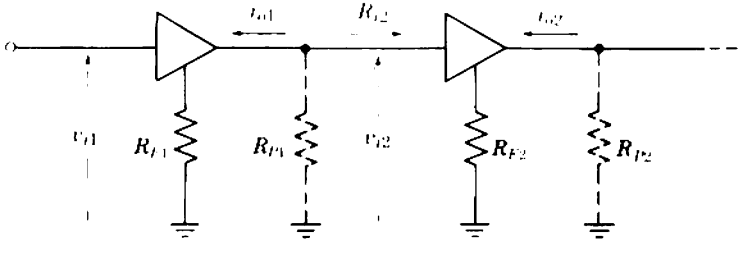


Fig. 11.11 Cascaded series-feedback stages.

* Section 11.2.3 may be omitted on a first reading.

The two supply resistors may be combined as R_{P1}, and

$$R_{L1} = \frac{R_{P1} \times R_{i2}}{R_{P1} + R_{i2}}. \tag{11.62}$$

Devices in which there is appreciable recombination (notably the bipolar transistor) have a finite input resistance. For a transistor series-feedback stage the input resistance is given by Eq. 11.53 and, therefore, the load resistance in a cascade of series-feedback stages is

$$R_{L1} \approx \frac{R_{P1} \times \beta_{N2} R_{F2}}{R_{P1} + \beta_{N2} R_{F2}}. \tag{11.63}$$

Thus, there is interaction between stages; the load resistance of one stage, and therefore its voltage gain, depend on β_N of the following transistor. Because β_N has a production tolerance, *application of series feedback to the individual stages does not stabilize the over-all gain*; it is necessary also to stabilize the load resistance for each stage. This can be accomplished by shunting the input resistance of each device by a small, known resistance, that is, by giving R_P a small value. The whole situation is similar to the cascade of nonfeedback stages considered in Section 5.5.1. As R_P is made smaller compared with the nominal value of R_i, the gain becomes more stable but also becomes smaller. Table 11.5 lists the gain per stage

Table 11.5 Maximum Nominal Gain Available with Specified Tolerance for Each Transistor in a Cascade of Series-Feedback Stages

Specified Tolerance on A_{max}	Tolerance on β_N	Allowed Nominal $\frac{A_{max}}{\beta_{NA}}$	Allowed Nominal $\left(\frac{A_{max}}{\beta_{NA}}\right)^2$	Actual Tolerance on A_{max}
±5%	50% to ∞	0.048	0.0023	95.5 to 105.0%
	67 to 200%	0.095	0.0090	95.5 to 105.0%
±10%	50% to ∞	0.091	0.0083	91.7 to 110.0%
	67 to 200%	0.182	0.0332	91.7 to 110.0%
±25%	50% to ∞	0.200	0.04	83.3 to 125.0%
	67 to 200%	0.400	0.16	83.3 to 125.0%

The fourth column is the maximum nominal gain of two stages (with twice the tolerance) for comparison with the alternate cascade. For example, with ±10% total tolerance on the gain (equivalent to ±5% per stage) and with β_N lying between 67 and 200% of its nominal value, the current gain of the alternate cascade is read from Table 11.3 as $0.088\,\beta_{NA}^2$, compared with $0.009\,\beta_{NA}^2$ for cascaded series-feedback stages in the above Table.

available with various degrees of uncertainty for specified production tolerances on β_V. There is virtually no improvement over stages without feedback, save that the requirement of stable emitter current is relaxed; r_E in the equations of Chapter 5 is replaced by $(r_E + R_F)$. Notice that Table 11.5 is constructed on the assumption that the series feedback applied to each device is sufficient to stabilize its transfer conductance much more accurately than the voltage gain is to be stabilized. Comparison of Tables 11.3 and 11.5 shows that two cascaded series-feedback stages are inferior to two alternate stages by about a factor of ten. The practical conclusion is that cascaded transistor series-feedback stages should never be used.

Devices in which recombination can be neglected (vacuum tubes and field-effect transistors) have almost infinite input resistance. There is therefore no uncertainty in the load resistance for a device within a cascade of series-feedback stages. The application of series feedback to the individual stages stabilizes the over-all gain provided that the individual stages are designed for stability; the voltage gain of a stage can under no circumstances exceed the amplification factor of the device, and the load (and gain) must not be made too large or the second-order corrections in the expressions for G_T become appreciable. There is no theoretical advantage of the alternate cascade over a cascade of series-feedback stages, but in practice the alternate cascade is usually superior because, with reasonable supply voltages, it is not possible to make the supply resistors large enough to realize the gain inherently available in cascaded series-feedback stages.

Finally, it is worth pointing out that, if a cascade of series-feedback stages is designed so that the voltage gain of each stage is stable, all transfer

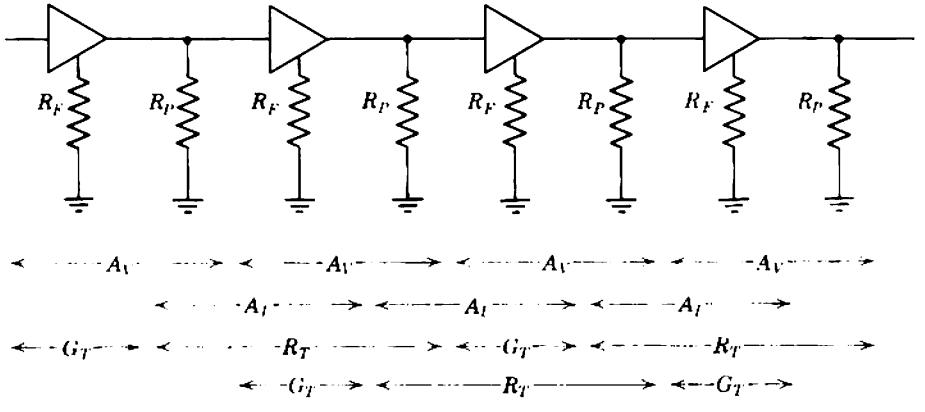


Fig. 11.12 Cascaded series-feedback stages, showing that all transfer functions are stable.

functions of every device are stable. Figure 11.12 illustrates the point. Table 11.5 therefore applies for both voltage gain and current gain.

11.2.4 Cascaded Shunt-Feedback Stages*

Shunt feedback stabilizes the transfer resistance v_o/i_i of an amplifier stage. In the cascade of shunt-feedback stages shown in Fig. 11.13, the output voltage from the first stage produced by a given input current can be stabilized to any required degree by the application of sufficient feedback. Similarly, the output voltage from the second stage can be stabilized if its input current is known. It is necessary to interpose a series resistor R_S between the stages to convert the voltage output of one stage into a current drive for the following stage; this resistor may be reduced to zero under special circumstances. The input current to the second stage is

$$i_{i2} = \frac{v_{o1}}{R_{S1} + R_{i2}}$$

and therefore the current gain of the first stage is

$$\frac{i_{i2}}{i_{i1}} = A_{I1} = \frac{v_{o1}}{i_{i1}}\left(\frac{1}{R_{S1} + R_{i2}}\right);$$

that is,

$$\frac{i_{i2}}{i_{i1}} = A_{I1} = \frac{R_{T1}}{R_{S1} + R_{i2}}. \tag{11.64}$$

The load resistance of the first stage is

$$R_{L1} = \frac{R_{P1} \times (R_{S1} + R_{i2})}{R_{P1} + (R_{S1} + R_{i2})} \tag{11.65}$$

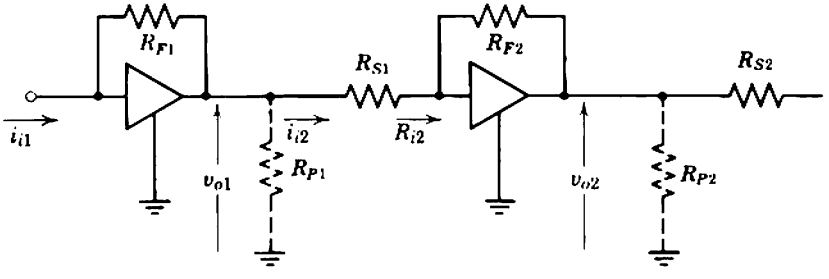


Fig. 11.13 Cascaded shunt-feedback stages.

* Section 11.2.4 may be omitted on a first reading.

and, in the limiting case in which the biasing resistors R_P are infinitely large,

$$R_{L1} = R_{S1} + R_{i2}. \tag{11.66}$$

The value of R_S must be a compromise between high gain per stage and stable gain. If R_S is reduced in value, the current gain per stage is increased provided that R_{T1} remains constant. The load resistance of the first stage, however, is reduced and this reduces the loop gain and the stability of the transfer resistance. Thus a high value for R_S favors a stable gain whereas a low value favors a high gain. Notice that the biasing resistors R_P are essentially parasitic; they reduce the load resistance for each stage and therefore reduce the transfer resistance that can be realized with a given degree of stability.

With bipolar transistors a definite upper limit exists to the gain per stage that can be realized with a given degree of stability. Consider a cascade of shunt-feedback stages in which the biasing resistors R_P are so large that Eq. 11.66 applies. Table 11.2 lists the nominal values of loop gain necessary for various tolerances on R_T, with given tolerances on β_N. In general, Table 11.2 is of the form

$$|A_{tA}| = \beta_{NA}\left(\frac{R_L}{R_F + R_L}\right) \geq \kappa;$$

that is,

$$\frac{\beta_{NA}}{(R_F/R_L) + 1} \geq \kappa$$

or

$$\frac{R_F}{R_L} \leq \frac{\beta_{NA}}{\kappa} - 1. \tag{11.67}$$

But, from Eq. 11.48,

$$|R_T| \approx R_F,$$

and R_L is given by Eq. 11.66. Therefore, from Eq. 11.64, the current gain per stage that can be realized with a given degree of stability is

$$A_{I1} = \frac{R_{T1}}{R_{S1} + R_{i2}} \approx \frac{R_{F1}}{R_{L1}} \leq \frac{\beta_{NA}}{\kappa} - 1;$$

that is,

$$A_{I(\max)} = \frac{\beta_{NA}}{\kappa} - 1, \tag{11.68}$$

where κ is the nominal value of loop gain required for the specified degree of stability of an isolated stage. Table 11.6 lists typical values of the maximum current gain available per stage. Comparison of Tables 11.3 and 11.6 shows that a cascade of shunt-feedback stages is inferior to the alternate cascade by about a factor of 10 for every two stages. Cascaded shunt-feedback stages are also slightly inferior to cascaded series-feedback

stages. In practice the gain per stage available from cascaded shunt-feedback stages is even less than the limit imposed by Eq. 11.68, because of the additional loading by the biasing resistors.

No theoretical limit exists to the gain per stage of cascaded vacuum tubes or field-effect transistors with shunt feedback, other than the limit imposed by device amplification factor μ. However, as with series-feedback stages considered in Section 11.2.3, loading by the biasing resistors imposes a practical limit which is less than for the alternate cascade.

In the design of a cascade of shunt-feedback stages it is necessary to make R_S large enough to mask the uncertainty in R_i due to the tolerance on g_m or β_N, thereby defining the input current produced by the output voltage of the preceding stage. The problem is in some ways akin to coupling efficiency in the alternate cascade; Table 11.4 can be interpreted to find the minimum value of R_S required in vacuum-tube or f.e.t. circuits. However, as R_S is made larger, loading by the biasing resistors becomes more significant.

Finally, Fig. 11.14 shows that all transfer functions of a shunt-feedback stage are stable, so that Table 11.6 applies both for current gain and voltage gain.

Table 11.6 Maximum Nominal Gain Available with Specified Tolerance for Each Transistor in a Cascade of Shunt-Feedback Stages

Specified Tolerance on A_{max}	Tolerance on β_N	Required Nominal $\frac{1 + R_F/R_L}{\beta_{NA}}$	Resultant Nominal $\frac{1 + A_{max}}{\beta_{NA}}$	Actual Tolerance on A_{max}
±5%	50% to ∞	0.050	0.047	95.5 to 105.0%
	67 to 200%	0.105	0.095	95.5 to 105.0%
±10%	50% to ∞	0.100	0.090	91.7 to 110.0%
	67 to 200%	0.222	0.182	91.7 to 110.0%
±25%	50% to ∞	0.250	0.200	83.3 to 125.0%
	67 to 200%	0.667	0.400	83.3 to 125.0%

The difference between the third and fourth columns is due to the difference in magnitude between the transfer resistance R_T of a shunt-feedback stage and its feedback resistor R_F:

$$A = \frac{R_T}{R_L} \approx \frac{R_F}{R_L}.$$

This distinction is ignored in the text.

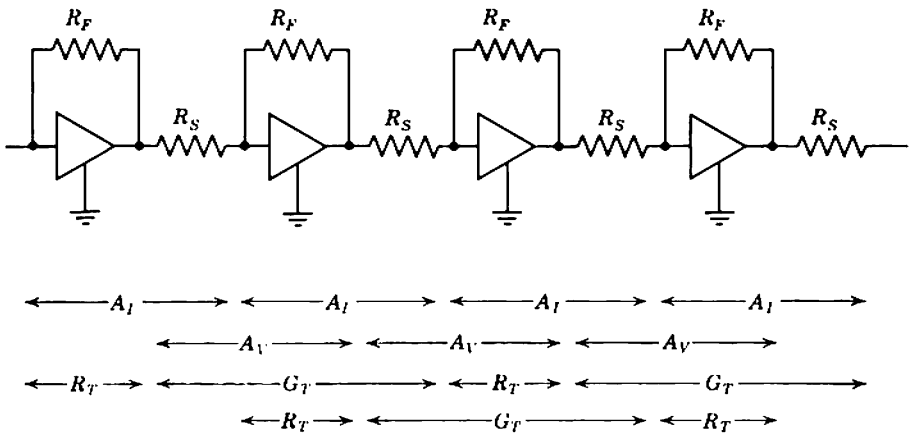


Fig. 11.14 Cascaded shunt-feedback stages, showing that all transfer functions are stable.

A number of circuits have been published in which shunt-feedback stages are cascaded directly with R_S reduced to zero. This appears thoroughly undesirable. At the outset it can be stated that the gain available per stage is certainly not greater than the limit imposed by Eq. 11.68, because R_S does not appear in this equation. There is therefore no possible advantage in reducing R_S to zero. There are, however, at least two disadvantages. First, the treatment of the isolated transistor shunt-feedback stage is based on the assumption (Eq. 11.37) that R_F is much larger than the input resistance of the transistor itself. This condition is unlikely to be satisfied if R_S is either small or omitted, because R_F must then be small to maintain a high loop gain. The result is a reduction in the available gain. Second, there is the theoretical problem that it is not at all clear just what the feedback accomplishes. Shunt feedback stabilizes the transfer resistance, v_o/i_i; unless R_S is included, the anomalous situation exists of a voltage-output stage feeding into a current-input stage!

11.3 LOOP GAIN*

The expressions for loop gain given in Section 11.1 are based on the assumption that the stages are isolated; that is, the source resistance for a series-feedback stage is zero whereas the source resistance for a shunt-feedback stage is infinite. These conditions are approached quite closely

* Section 11.3 may be omitted on a first reading.

in the alternate cascade because of the impedance mismatch between stages, but they may not be satisfied in cascades of similar stages. This section provides a mathematical discussion of the effects of source resistance on the performance of feedback stages. The theory is of limited practical application; circuit design can always be based on the first-order theory of Section 11.2.1 with the correcting factors due to coupling efficiency and loading of a shunt stage by a series stage.

11.3.1 Series-Feedback Stage

Figure 11.15*a* is the equivalent circuit for a charge-controlled device with series feedback being fed from a source of finite resistance R_S. For simplicity it is assumed that r_2 and g_f are large enough to be neglected. This approximation is valid in almost every practical case. Even if the approximation is not justified, the inclusion of r_2 and g_f merely adds algebraic complexity to the analysis; no new concepts are introduced.

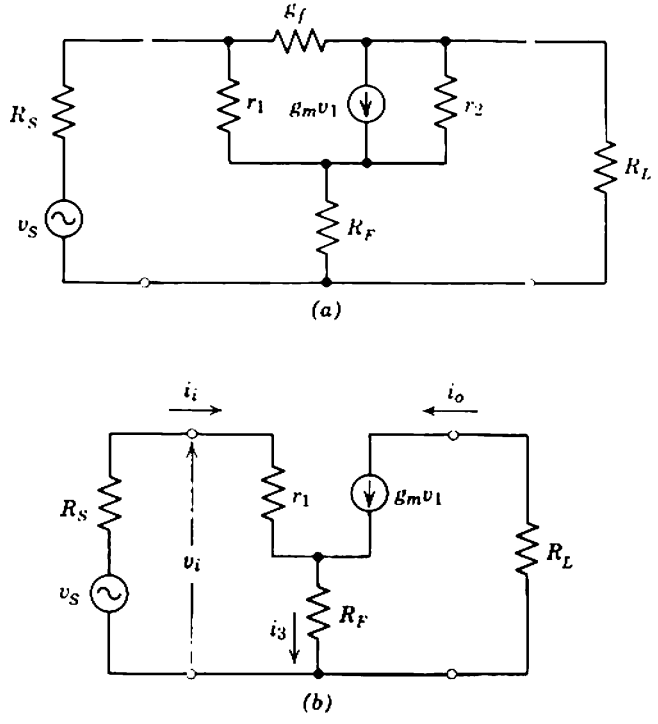


Fig. 11.15 The series-feedback stage in a system: (*a*) complete charge-control model; (*b*) simplified charge-control model.

The equivalent circuit may therefore be simplified to Fig. 11.15*b*. The notation can be simplified by introducing two symbols α and β:

(i) the current gain of the device is

$$\beta = g_m r_1 \tag{11.69}$$

and

$$g_m v_1 = \beta i_i; \tag{11.70}$$

(ii) the ratio of collecting- to emitting-electrode currents is

$$\frac{i_o}{i_3} = \alpha = \frac{\beta}{\beta + 1}. \tag{11.71}$$

From the definition of loop gain (Eq. 10.36), the ratio of output current to source voltage can be written as

$$\frac{i_o}{v_S} = \frac{\alpha}{R_F}\left(\frac{1}{1 - 1/A_l'}\right), \tag{11.72}$$

where A_l' is the loop gain of the complete system. (The prime is to distinguish the system loop gain A_l' from the stage loop gain A_l derived in Section 11.1.) Network analysis shows that

$$\frac{i_o}{v_S} = \frac{\alpha}{R_F}\left[\frac{1}{1 + (R_S + r_1)/(\beta + 1)R_F}\right] \tag{11.73}$$

and therefore the system loop gain is identified as

$$A_l' = -\frac{(\beta + 1)R_F}{R_S + r_1}. \tag{11.74}$$

Thus, the loop gain of a series-feedback stage in a system depends on its source resistance (as expected from Section 10.3.2, Eq. 10.33).

For a bipolar transistor (the only charge-controlled device with a significant input resistance), r_B can be included with R_S and Eq. 11.74 becomes

$$A_l' = -\frac{R_F}{(R_S + r_B)(1 - \alpha_N) + r_E} \tag{11.75}$$

compared with the expression

$$A_l = -\frac{R_F}{r_B(1 - \alpha_N) + r_E} \approx -\frac{R_F}{r_B/\beta_N + r_E}$$

derived in Section 11.1.3. For a vacuum tube or f.e.t., β and r_1 are very large indeed; it therefore follows from Eqs. 11.69 and 11.74 that A_l and A_l' are numerically equal, and are both equal to $(-g_m R_F)$.

Undoubtedly, A_l' is the true loop gain of a series-feedback stage in a system, and determines the stability of the output current when the stage is fed from a given source. In contrast, A_l is the loop gain of an isolated

state. The distinction is that A_l assumes a constant input voltage v_i to the stage whereas A_l' assumes a constant internal source voltage v_S. In the latter case the actual input voltage is the source voltage divided down by the source and input resistances:

$$\frac{v_i}{v_S} = \frac{R_i}{R_S + R_i}, \tag{11.76}$$

and A_l' takes account of any uncertainty in this division that arises from uncertainty in R_i. Notice that A_l' reduces to A_l if R_S is zero, that is, if the stage is fed from a pure voltage source so that its input voltage is constant and equal to the source voltage; A_l' is very nearly equal to A_l in the alternate cascade because the output resistance of the preceding shunt-feedback stage is small.

The relation between the internal source voltage and output current of a series-feedback stage can of course be derived with reference to the stage loop gain A_l and the division of voltage between source and input resistances. For the general stage shown in Fig. 11.15*b*,

$$R_i = \frac{v_i}{i_i} = r_1 + R_F(\beta + 1),$$

$$G_T = \frac{\beta}{R_i} = \alpha\left(\frac{\beta + 1}{R_i}\right),$$

from which

$$\frac{i_o}{v_S} = \frac{v_i}{v_S} \times \frac{i_o}{v_i} = \frac{\alpha}{R_F}\left[\frac{1}{1 + (R_S + r_1)/(\beta + 1)R_F}\right]$$

as expected from Eq. 11.73. For a bipolar transistor Eq. 11.73 becomes

$$\frac{i_o}{v_S} = \frac{\alpha_N}{(R_S + r_B)(1 - \alpha_N) + r_E + R_F} \approx \frac{\alpha_N}{(R_S + r_B)/\beta_N + r_E + R_F}. \tag{11.77}$$

Stage loop gain A_l is more useful in practical amplifier design than system loop gain A_l'. The significant parameter of a series-feedback stage is the stability of the output current when an input voltage is actually applied to the stage, and this stability is determined by the stage loop gain. The uncertainty in the input resistance (which results in an uncertainty in v_i when the stage is fed from another stage of finite output resistance) is most conveniently considered as an uncertainty in the load of this preceding stage, and should be taken into account in calculating its output voltage. The alternative approach using system loop gain is to

(i) take v_S as the open-circuit output voltage of the preceding stage, that is the voltage that would exist at the input of the series-feedback stage if its input resistance were infinite, which it most certainly is not in the general case;

(ii) take R_S as the output resistance of the preceding stage (including all biasing resistors);
(iii) determine the output current from the series stage by means of Eq. 11.73.

This approach seems far less convenient except possibly for the input stage of an amplifier designed for use only with a particular source. In all other cases, the significant transfer function is (output current)/(actual input voltage), and stage loop gain is the more useful quantity.

11.3.2 Shunt-Feedback Stage

Figure 11.16*a* is the equivalent circuit for a charge-controlled device with shunt feedback, being fed from a source of finite resistance R_S. As in the case of the series-feedback stage considered above, r_2 and g_f can be neglected or, if not neglected, included in the values of R_L and R_F. The equivalent circuit reduces to Fig. 11.16*b*.

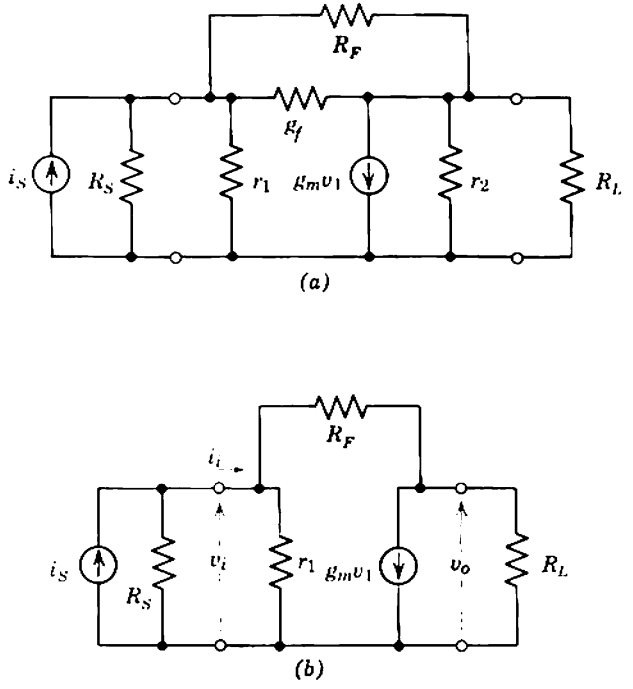


Fig. 11.16 The shunt-feedback stage in a system: (*a*) complete charge-control model; (*b*) simplified charge-control model.

From the definition of loop gain, the ratio of output voltage to source current can be written as

$$\frac{v_o}{i_S} = -\frac{R_F}{1 - 1/A_l'}, \tag{11.78}$$

where A_l' is the system loop gain. Network analysis shows that

$$A_l' = -\left(g_m - \frac{1}{R_F}\right)\left(\frac{R_F \times R_L}{R_F + R_L}\right)\left(\frac{r_1 R_S}{r_1 R_F + r_1 R_S + R_F R_S}\right).$$

In the first bracket, g_m is the forward transmission via the mutual conductance of the active device, whereas $(-1/R_F)$ is the direct transmission through the feedback resistor (compare with Eq. 10.44 of Section 10.4.2). Invariably the former component is much the larger in any practical amplifier, and the system loop gain reduces to

$$A_l' = -g_m\left(\frac{R_F R_L}{R_F + R_L}\right)\left(\frac{r_1 R_S}{r_1 R_F + r_1 R_S + R_F R_S}\right). \tag{11.79}$$

For a vacuum tube or f.e.t. r_1 is very large, and therefore Eq. 11.79 becomes

$$A_l' = -g_m\left(\frac{R_F R_L}{R_F + R_L}\right)\left(\frac{R_S}{R_F + R_S}\right). \tag{11.80}$$

It is interesting that this value follows directly from Section 11.1.2; the first two terms are the stage loop gain derived in Eq. 11.21, and the last term is the voltage dividing factor between R_F and R_S.

For a bipolar transistor, r_B can be included with $\beta_N r_E$ in the expression for r_1, and therefore Eq. 11.79 can be rearranged as

$$A_l' = \beta_N\left(\frac{R_L}{R_F + R_L}\right)\left[\frac{R_F R_S}{(r_B + \beta_N r_E)(R_F + R_S) + R_F R_S}\right]. \tag{11.81}$$

The first two terms are the stage loop gain derived in Eq. 11.35; there is no simple interpretation of the third term. Usually, however, r_1 is much less than either R_F or R_S, so that the third term is nearly unity and A_l' reduces approximately to A_l.

The distinction between A_l' and A_l for a shunt-feedback stage is dual to that for a series-feedback stage. The system loop gain A_l' determines the stability of the output voltage when the stage is fed from a given source, whereas the stage loop gain A_l determines the stability of the output when a given current actually flows into the stage. The input current to the stage is

$$\frac{i_i}{i_S} = \frac{R_S}{R_S + R_i} \tag{11.82}$$

and, for constant i_S, i_i depends on R_i. In the limiting case of infinite source resistance, the input current is constant and equal to the source current; A_I' reduces to A_I (Eqs. 11.80 and 11.21, or 11.81 and 11.35).

There are three methods for calculating the output voltage from a shunt-feedback stage. The formal method (dual to that for the series-feedback stage in Section 11.3.1) is to

(i) take the effective source current generator as the g_m generator of the preceding stage, that is, the current that would be delivered into a short-circuit load;

(ii) the effective source resistance is the parallel combination of R_o of the preceding stage with all biasing resistors.

The system loop gain A_I' determines the stability of the output voltage.

Section 11.2.2.2 describes the practical method for calculating the output, using coupling efficiency. The output current from the preceding

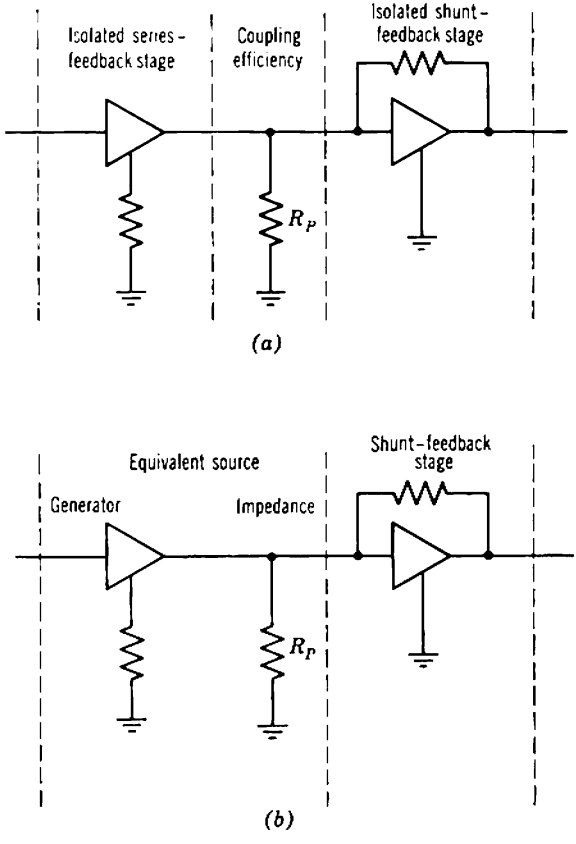


Fig. 11.17 Alternative ways of regarding a shunt-feedback stage: (*a*) isolated stage with coupling efficiency at its input; (*b*) as part of a system, as in Fig. 11.16.

series-feedback stage divides between the biasing resistor R_P and the shunt-feedback stage as in Fig. 11.17*a*, and output voltage from the shunt-feedback stage is

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = -\eta \times R_T. \tag{11.83}$$

A second practical method that is useful in calculating the frequency response is illustrated in Fig. 11.17*b*. The actual output current from the preceding series-feedback stage (determined by G_T of that stage) is regarded as the source current generator, and the biasing resistor R_P is regarded as the source resistance. From Eq. 11.78

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = \frac{R_F}{1 - 1/A_l'}, \tag{11.84}$$

where the loop gain for this particular system is given by Eq. 11.79 with R_P replacing R_S.

There are two classes of uncertainty in Eq. 11.83: the uncertainty in η due to the tolerance on R_i (Section 11.2.2.2), and the tolerance on R_T determined by the stage loop gain (Section 11.1.2 or 11.1.4). The total uncertainty in v_o is the combination of these two effects, and must of course reduce to the value obtained from a system loop gain calculation. Expansion of Eq. 11.83 gives

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = \left(\frac{1}{1 + R_i/R_P}\right)\left(\frac{R_F}{1 - 1/A_l}\right).$$

Substitution of Eqs. 11.20 and 11.21 for a vacuum tube gives

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = \frac{R_F}{1 + \left(\dfrac{R_F + R_L}{g_m R_F R_L}\right)\left(\dfrac{R_F + R_P}{R_P}\right) + \begin{pmatrix}\text{very small}\\ \text{terms}\end{pmatrix}} \tag{11.85}$$

as expected from Eqs. 11.80 and 11.84; the very small error term arises from the approximate expression for R_i (Eq. 11.20). Similarly, for a bipolar transistor substitution of Eqs. 11.34 and 11.35 gives

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = \frac{R_F}{1 + \left(\dfrac{R_F + R_L}{\beta_N R_L}\right)\left[\dfrac{(r_B + \beta_N r_E)(R_F + R_P) + R_F R_P}{R_F R_P}\right] + \begin{pmatrix}\text{very small}\\ \text{terms}\end{pmatrix}} \tag{11.86}$$

as expected from Eq. 11.81.

11.4 NOISE

The noise in a feedback amplifier stage can be divided into two broad classes—the noise of the active device and the thermal noise associated with the feedback resistors. The formal theory of noise in single-stage feedback amplifiers follows directly from the general discussions in Chapters 8 and 10, so it remains to consider the practical application. Section 10.8 shows that the two spot noise generators at the input of a device are not changed by feedback, and Section 8.3.1 lists the noise contributions of the resistors in an amplifier stage.

11.4.1 Series-Feedback Stage

Figure 11.18*a* shows the circuit diagram and noise generators for a series-feedback stage. The circuit is identical to Fig. 8.4*d* in Section

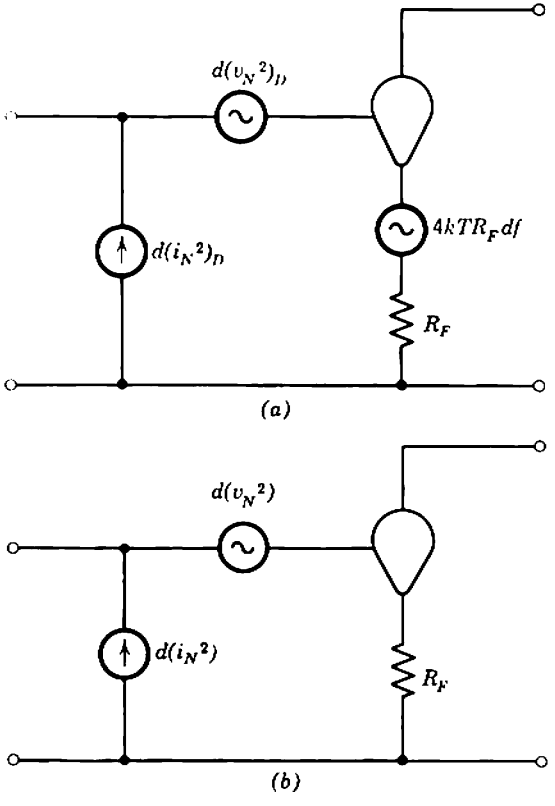


Fig. 11.18 Noise in a series-feedback stage: (*a*) noise generators; (*b*) equivalent representation.

8.3.1, and from Eqs. 8.44 and 8.45 the two spot noise generators at the input of a series-feedback stage are:

$$d(v_N^2) = d(v_N^2)_D + d(i_N^2)_D R_F^2 + 4kTR_F\,df, \tag{11.87}$$

$$d(i_N^2) = d(i_N^2)_D. \tag{11.88}$$

The spot noise voltage at the input is

$$d(v_{Ni}^2) = [4kTR_S\,df + d(v_N^2) + d(i_N^2)R_S^2]\left(\frac{R_i}{R_S + R_i}\right)^2. \tag{11.89}$$

Equation 11.89 is formally incorrect because of correlation between the noise generators. The spot noise generators of the stage $d(v_N^2)$ and $d(i_N^2)$ are correlated due to the presence of $d(i_N^2)_D$ in each, and the device spot noise generators $d(v_N^2)_D$ and $d(i_N^2)_D$ are themselves slightly correlated. Section 8.1.1, however, shows that correlation can be neglected unless two conditions are satisfied simultaneously: R_S and R_i must be of the same order of magnitude, and the short-circuit current gain must be small. For correct operation a series-feedback stage must be fed from a low-resistance source, so the former condition is not satisfied. Correlation can therefore be neglected, and Eq. 11.89 is a realistic measure of the mean-square spot noise voltage at the input of a series-feedback stage.

Qualitatively, the noise voltage at the input of a series-feedback stage increases monotonically with increasing loop gain, that is, with increasing feedback resistor. The noise voltage generator $d(v_N^2)$ increases with R_F, and no other term in Eq. 11.89 falls to offset this. A poor noise performance is the greatest disadvantage of the series-feedback stage. If there is sufficient feedback to ensure gain stability, the total noise is far more than that of the active device alone.

In normal operation the source resistance for a series-feedback stage is so much smaller than its input resistance that $d(i_N^2)$ does not contribute appreciably to $d(v_{Ni}^2)$. The noise voltage generator $d(v_N^2)$ therefore provides an almost complete measure of stage noise, and the spot signal-to-noise ratio is

$$\frac{S_o}{dN_o} \approx \frac{v_S^2}{d(v_N^2)}. \tag{11.90}$$

Further, the device noise-current generator $d(i_N^2)_D$ is usually so small that Eq. 11.87 reduces to

$$d(v_N^2) \approx d(v_N^2)_D + 4kTR_F\,df. \tag{11.91}$$

Thus, the mean-square noise voltage generator at the stage input is the sum of the device mean-square noise voltage generator and the thermal

noise of the feedback resistor. Equation 11.91 is the most generally useful expression for the noise of a series-feedback stage.

Equation 11.91 allows a general, order-of-magnitude estimate of the increase in noise with feedback. Inspection of Section 8.2 reveals that the white component of the spot noise voltage generator at the input of any device is

$$d(v_N^2)_D \approx 4kT\left(\frac{\nu}{g_m}\right) df. \tag{11.92}$$

The constant ν is about one half in all cases—0.644 for a triode, $\frac{1}{2}$ for a bipolar transistor, and ranges from $\frac{1}{2}$ to $\frac{2}{3}$ for an f.e.t. Inspection of Section 11.1 reveals that the loop gain for any series-feedback stage is of the order of

$$|A_l| \sim g_m R_F. \tag{11.93}$$

Substitution into Eq. 11.91 and rearrangement yields the ratio of the noise with feedback to the noise of the device alone:

$$\frac{d(v_N^2)}{d(v_N^2)_D} \sim 1 + 2|A_l|. \tag{11.94}$$

It is worth pointing out that the left-hand side of Eq. 11.94 is a ratio of mean-square voltages, not rms voltages; the increase in rms voltage is only threefold when $|A_l|$ is 4.

11.4.2 Shunt-Feedback Stage

Figure 11.19*a* shows the circuit diagram and noise generators for a shunt-feedback stage. The circuit is identical to Fig. 8.4*c* in Section 8.3.1, and from Eqs. 8.42 and 8.43 the spot noise generators at the input of a shunt-feedback stage are:

$$d(v_N^2) = d(v_N^2)_D + \frac{4kT}{{g_m}^2 R_F} df, \tag{11.95}$$

$$d(i_N^2) = d(i_N^2)_D + \frac{4kT}{R_F}\left(1 + \frac{1}{\beta}\right)^2 df + \frac{d(v_N^2)_D}{{R_F}^2}\left(1 + \frac{1}{\beta}\right)^2, \tag{11.96}$$

where β is the short-circuit current gain of the device. The spot noise current at the input is

$$d(i_{Ni}^2) = \left[\frac{4kT}{R_S} df + \frac{d(v_N^2)}{{R_S}^2} + d(i_N^2)\right]\left(\frac{R_S}{R_S + R_i}\right)^2, \tag{11.97}$$

where there is a formal error in Eq. 11.97 due to correlation between the noise generators.

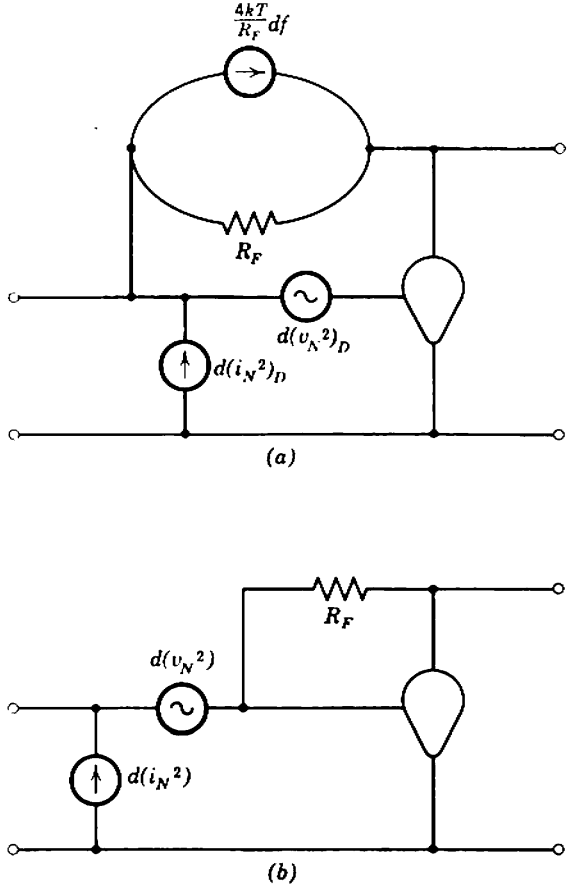


Fig. 11.19 Noise in a shunt-feedback stage: (*a*) noise generators; (*b*) equivalent representation.

As with a series-feedback stage, the noise at the input increases monotonically with increasing loop gain, that is, with decreasing R_F. If there is sufficient feedback to ensure gain stability, the total noise is far more than that of the active device alone. In normal operation the source resistance for a shunt-feedback stage is so much larger than its input resistance that $d(v_N^2)$ does not contribute appreciably to $d(i_{Ni}^2)$. Therefore, the noise current generator $d(i_N^2)$ provides an almost complete measure of stage noise, and the spot signal-to-noise ratio is

$$\frac{S_o}{dN_o} \approx \frac{i_S^2}{d(i_N^2)}. \tag{11.98}$$

Further, the current gain β of useful amplifying devices is so large that Eq. 11.96 reduces to

$$d(i_N^2) \approx d(i_N^2)_D + \frac{4kT}{R_F}\,df. \tag{11.99}$$

The spot noise current generator at the stage input is thus the sum of the device noise current generator and the thermal noise of the feedback resistor. No general expression (corresponding to Eq. 11.94 of the series stage) exists for the increase in noise with shunt feedback, because there is no general expression for the device noise current generator $d(i_N^2)_D$.

11.5 NONLINEARITY

Nonlinearity in an amplifier can be expressed in terms of the differential error γ at any point in its transfer function:

$$\gamma = \frac{TF - TF_Q}{TF_Q},$$

where TF_Q is the transfer function at the quiescent point. Relations discussed in Chapter 9 exist between the differential errors γ' and γ'' at the peaks of a signal waveform, and the low-order harmonic and intermodulation distortion. There are two alternative methods for calculating the nonlinearity in a feedback amplifier. The formal approach based on Eq. 10.78 of Section 10.9 is to determine the differential errors γ_0' and γ_0'' for the amplifier without feedback, and then divide by the return difference to obtain γ' and γ'' with the feedback applied. For cascades of series- and shunt-feedback stages (and in many other cases) a simpler alternative is to calculate γ' and γ'' directly from the known transfer function with feedback. Notice that, because the nonlinearity in a feedback amplifier is small, the approximate theory of Section 9.4.1 applies; Eqs. 9.23 are adequate approximations for γ' and γ'', and should be used in preference to Eqs. 9.17.

11.5.1 Sources of Nonlinearity

In a cascade of alternate series- and shunt-feedback stages, nonlinearity can arise in three ways—variation of G_T of series-feedback stages, variation of R_T of shunt-feedback stages, and variation of the coupling efficiency η. Although all stages contribute to the total nonlinearity, the bulk is due to the last two stages in which the signals are largest and they alone need to be considered.

Equations 11.13, 11.22, and 11.38 express the transfer conductance of vacuum-tube, transistor, and f.e.t. series-feedback stages respectively in terms of device parameters. Nonlinearity results because these parameters change with operating current. For an ideal vacuum tube, g_m is proportional to $I_K^{1/2}$ so γ' and γ'' can be calculated at any peak-to-peak current output. As a better alternative for practical tubes, g_m at the peak currents can be determined from graphical data. For field-effect transistors, g_m varies quite closely as $I_S^{1/2}$ so there is no need for graphical data. For bipolar transistors, the dominant source of nonlinearity is the variation of r_E with I_E; second-order contributions result from β_N and possibly r_B or α_N, and these can be included if graphical data are available.

Equations 11.16, 11.30 (or 11.36), and 11.43 give the transfer resistance of shunt-feedback stages, and again nonlinearity results because the device parameters change with operating current. Notice that a shunt-feedback stage develops its output voltage across R_F as well as R_L, so that the peak-to-peak device current is

$$i_{pp} = \frac{v_{o(pp)}}{(R_F \parallel R_L)}. \tag{11.100}$$

For vacuum-tube and f.e.t. shunt-feedback stages, g_m alone contributes to the nonlinearity which can be calculated similarly to that for series-feedback stages above. The loop gain of these stages, however, is equal to their voltage gain and is ordinarily so large that nonlinearity can be neglected in comparison with the other stages in a multistage amplifier. For transistor shunt-feedback stages β_N is the principal contributor to nonlinearity, although r_E (in Eq. 11.36) or R_L can be significant; R_L involves the input resistance of the following stage and this depends on I_E. Again nonlinearity is small because β_N does not depend strongly on operating current. Shunt-feedback stages of any type rarely contribute to the nonlinearity of a multistage amplifier.

Equation 11.58 expresses the coupling efficiency between a series- and shunt-feedback stage as a function of the input resistance of the shunt-feedback stage. This input resistance depends on the device parameters (Eqs. 11.20, 11.34, and 11.46), and therefore varies with operating current. For tubes and f.e.t.s, g_m alone contributes to this nonlinearity, whereas r_E is the principal contributor in transistor circuits.

In summary, when an n-stage amplifier has a series-feedback output stage, this stage usually dominates the nonlinearity and

$$\gamma \approx \frac{(G_{Tn}) - (G_{TnQ})}{(G_{TnQ})}. \tag{11.101a}$$

However, when the output stage has shunt feedback, the series-feedback penultimate stage and the coupling efficiency between it and the output stage are the main contributors:

$$\gamma \approx \frac{[G_{T(n-1)} \times \eta_n] - [G_{T(n-1)Q} \times \eta_{nQ}]}{[G_{T(n-1)Q} \times \eta_{nQ}]}. \qquad (11.101b)$$

Except when *p-n-p* and *n-p-n* transistors are used in adjacent stages the nonlinearities in G_T and η partially compensate each other.

11.6 MISCELLANEOUS TOPICS

This section discusses some unrelated but important practical topics. Included are suggestions for biasing feedback stages and means for designing amplifiers with controlled (but not constant) gain.

11.6.1 Loading by Biasing Resistors

The principal loss of gain in a cascade of alternate series- and shunt-feedback stages is due to loading by the biasing supply resistors. Some circuit arrangements for biasing amplifier stages give less loading than others and are, therefore, preferable.

Figure 11.20 shows the preferred methods for biasing a series-feedback stage. In all cases, some form of the emitting-electrode-feedback circuit is recommended. Collecting-electrode feedback is undesirable because it increases the loading on the preceding stage; the biasing resistor provides a shunt-feedback path which reduces the input resistance of the stage. If only one supply voltage is available, a derived form of emitting-electrode-feedback biasing is best. In some cases (usually for vacuum tubes) the signal feedback resistor is large enough to ensure biasing stability, and the circuit of Fig. 11.20*a* is used. In other cases (usually for transistors) the signal feedback resistor is not large enough, and additional biasing resistance is included as in Fig. 11.20*b*. There is, of course, no restriction of the former arrangement to tubes or the latter to transistors. If both positive and negative supplies are available, the basic form of emitting-electrode-feedback biasing is useful; Fig. 11.20*c* shows the circuit arrangement for an *n-p-n* transistor.

Figure 11.21 shows the preferred methods for biasing a shunt-feedback stage. In all cases some form of collecting-electrode feedback is used and the signal feedback resistor doubles as a biasing resistor. If only one supply is available, combined-feedback biasing is an obvious choice and is illustrated for a pentode in Fig. 11.21*a*. An excellent alternative is to

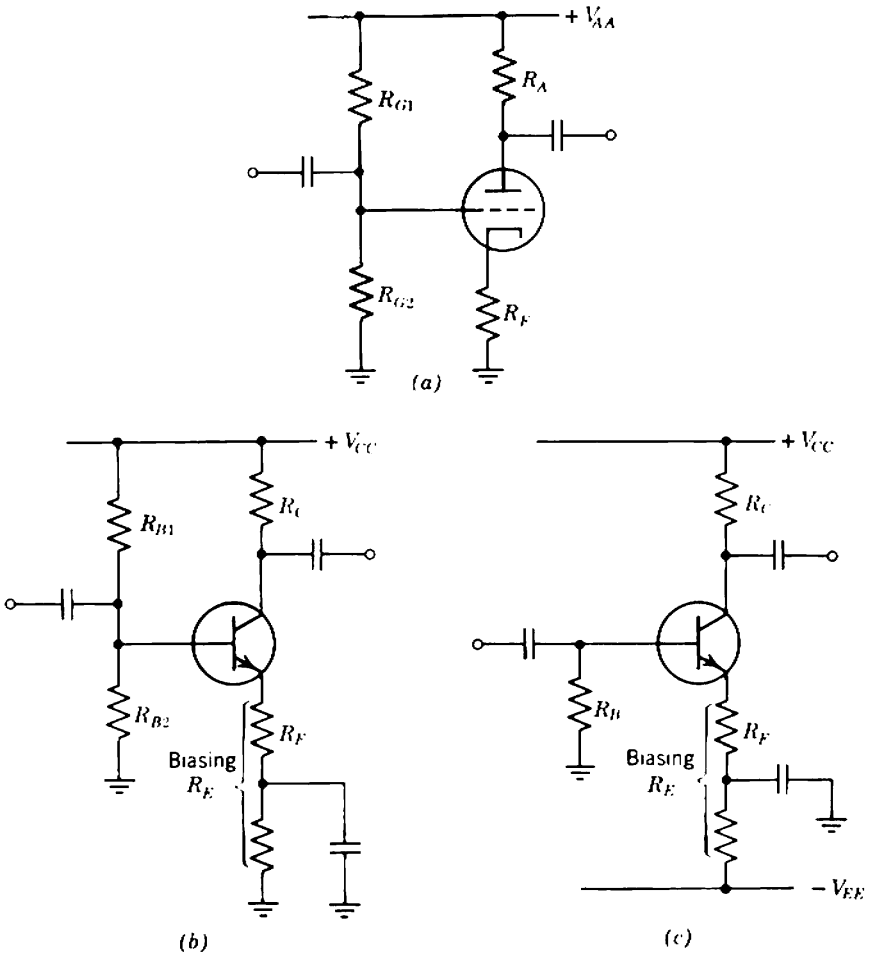


Fig. 11.20 Preferred biasing methods for a series-feedback stage: (*a*) single supply voltage; (*b*) arrangement when signal R_F is too small for biasing; (*c*) two supply voltages.

replace the cathode/emitter *RC* network by a Zener diode (Fig. 6.48 of Section 6.6.2). If both positive and negative supplies are available, simple collecting-electrode-feedback biasing is recommended (Fig. 11.21*b*). Sometimes the signal feedback resistor is rather too small for convenience in the biasing circuit as an excessive bleed current is required. In such cases the circuit arrangement in Fig. 11.21*c* is useful. (The reverse situation, the signal feedback resistor being too large for biasing stability, never occurs.) Although Fig. 11.21*c* shows the circuit modification applied to a collecting-electrode-feedback biasing system, it can equally well be applied to a combined-feedback biasing system.

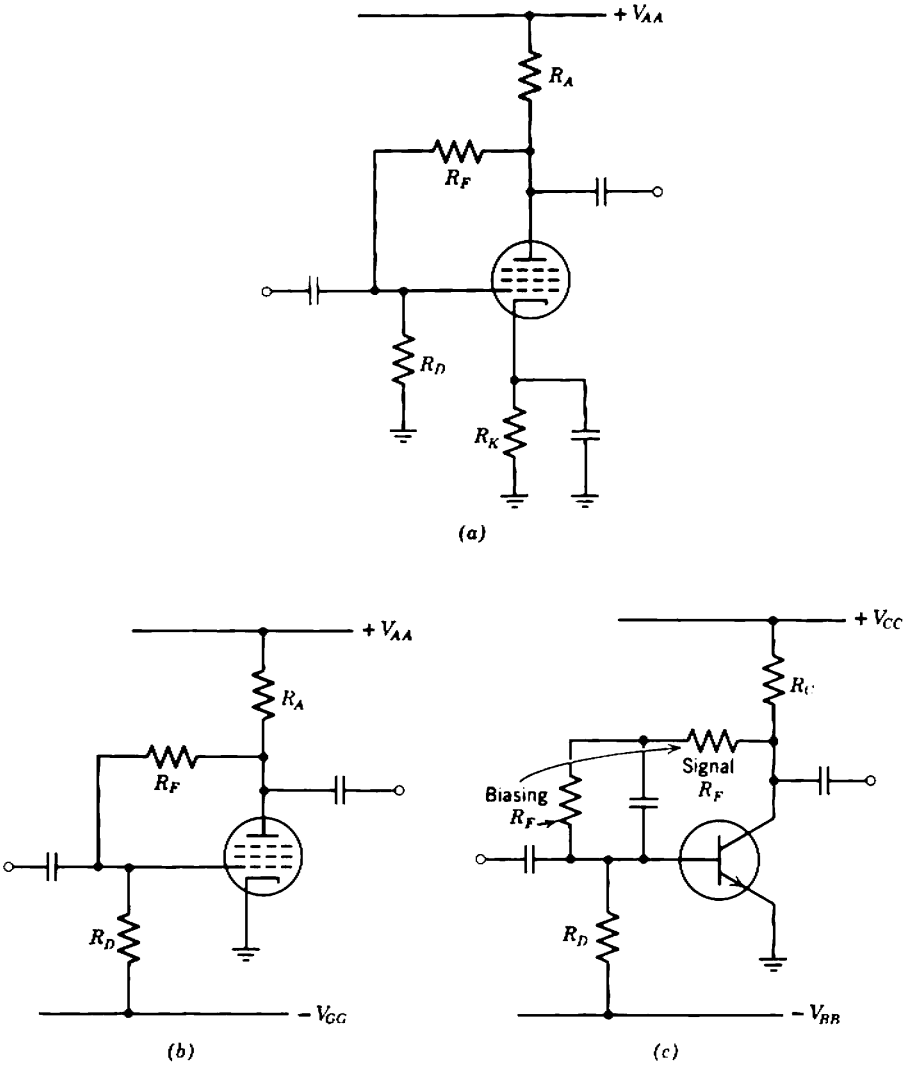


Fig. 11.21 Preferred biasing methods for a shunt-feedback stage: (*a*) single supply voltage; (*b*) two supply voltages; (*c*) arrangement when signal R_F is too small for biasing.

Often, loading by biasing resistors can be reduced by direct coupling stages, as the control-electrode supply resistor for the second stage is eliminated. Section 6.5 discusses the principles of direct coupling and suggests a number of circuits. Perhaps most useful are the three arrangements shown in Fig. 11.22 in which a series stage is direct coupled to the preceding shunt stage. Although transistor realizations are shown,

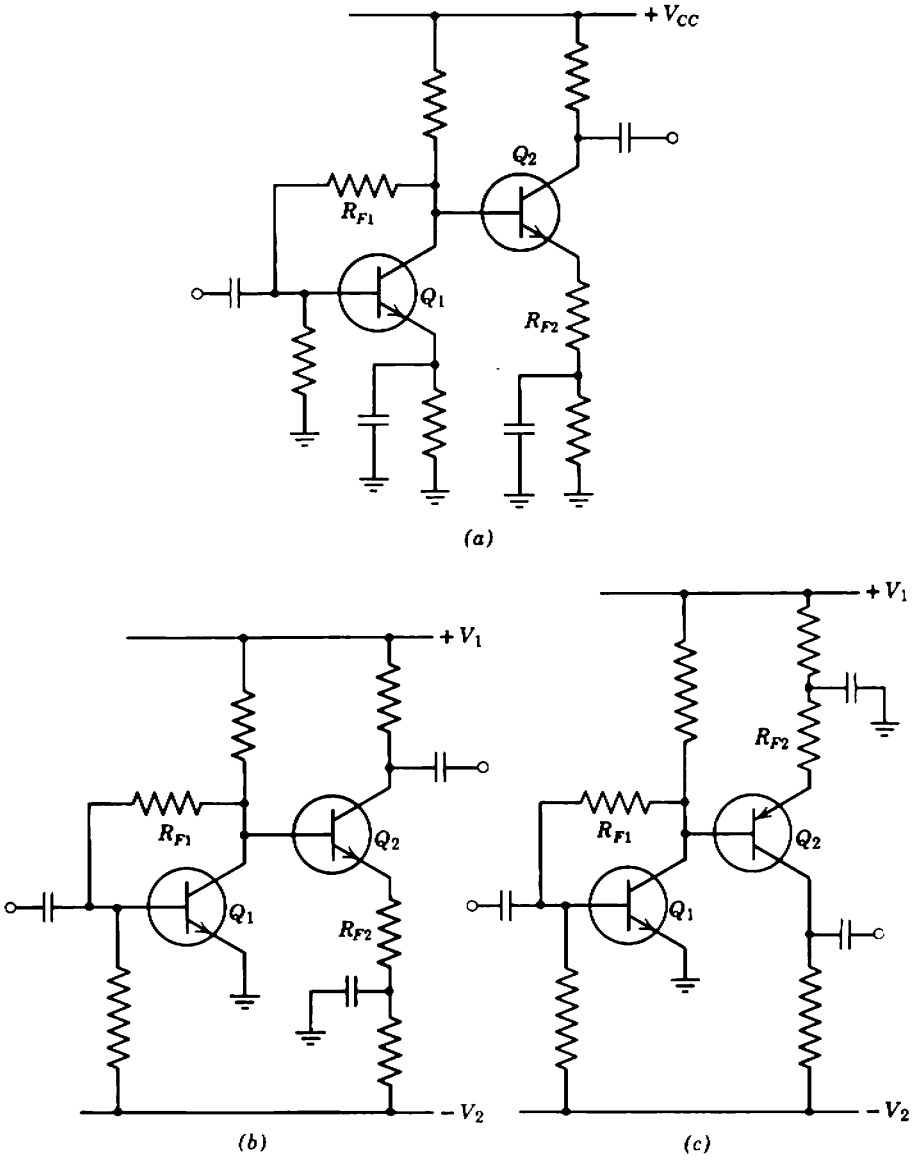


Fig. 11.22 Examples of direct-coupled stages.

tubes are satisfactory. Figures 11.22*a* and *b* are straightforward applications of the biasing circuits for an isolated shunt-feedback stage using one and two supply voltage respectively; the emitter *RC* network for Q_1 in Fig. 11.22*a* can be replaced by a Zener diode. Combinations of *p-n-p* and *n-p-n* transistors lead to elegant circuits such as Fig. 11.22*c*; the input

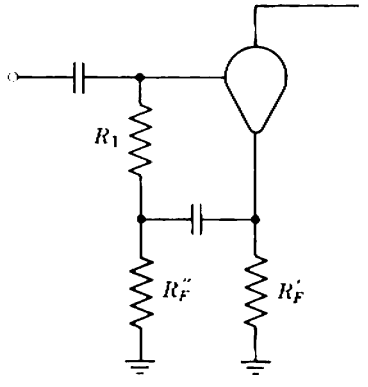
base and output collector are near the same dc voltage, and the supply voltage requirements are no greater than for a single stage.

11.6.1.1 Bootstrapping of Supply Resistors

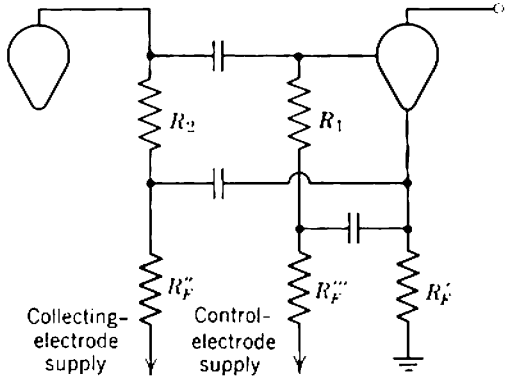
A powerful method for reducing the loading by supply resistors on the output of a shunt-feedback stage is *bootstrapping*. Essentially bootstrapping is the application of series feedback to raise the total input impedance (including all biasing resistors) of a stage.

Figure 11.23*a* illustrates a simple case of bootstrapping. If the series feedback is removed by short circuiting R_F'', the input resistance is

$$R_{i0} = (R_1 \parallel r_1),$$



(*a*)



(*b*)

Fig. 11.23 Theory of bootstrapping.

where r_1 is the input resistance of the device. With the short circuit removed, the feedback resistor is $(R'_F \parallel R''_F)$ so that

$$|A_l| = g_m(R'_F \parallel R''_F),$$

and the input resistance rises to

$$R_i = (R_1 \parallel r_1)(1 + |A_l|).$$

Figure 11.23*b* illustrates a more general case in which

$$R_{i0} = (R_1 \parallel R_2 \parallel r_1), \tag{11.102}$$

$$R_F = (R'_F \parallel R''_F \parallel R'''_F), \tag{11.103}$$

and therefore

$$|A_l| = g_m R_F \tag{11.104}$$

and

$$R_i = R_{i0}(1 + |A_l|). \tag{11.105}$$

Figure 11.24 illustrates the practical application of bootstrapping to transistor stages. The load resistance for the shunt-feedback stage is

$$R_L = [R_C \parallel R_B \parallel (r_B + \beta_N r_E)](1 + |A_l|).$$

The feedback resistance for the series-feedback stage is

$$R_F = (R'_{F2} \parallel R''_{F2} \parallel R'''_{F2}),$$

and

$$|A_l| = \frac{R_F}{r_B/\beta_N + r_E}.$$

If R_C and R_B are large compared with $(r_B + \beta_N r_E)$, the effective load becomes

$$R_L \approx r_B + \beta_N(r_E + R_F)$$

and the loading on the shunt stage is reduced to the theoretical minimum (Section 11.2.2.1). Figure 11.24*a* shows bootstrapping applied to *RC*-coupled stages, Fig. 11.24*b* shows direct-coupled stages, and Fig. 11.24*c* is an elegant arrangement using *n-p-n* and *p-n-p* transistors in combination.

Notice that the supply resistors at the input of a shunt-feedback stage cannot be bootstrapped simply, because no signal voltage appears at the emitting electrode.

11.6.2 The Use of Feedback Networks for Gain Variation

The gain of a well-designed feedback amplifier is determined almost entirely by the feedback resistors, independent of the active device

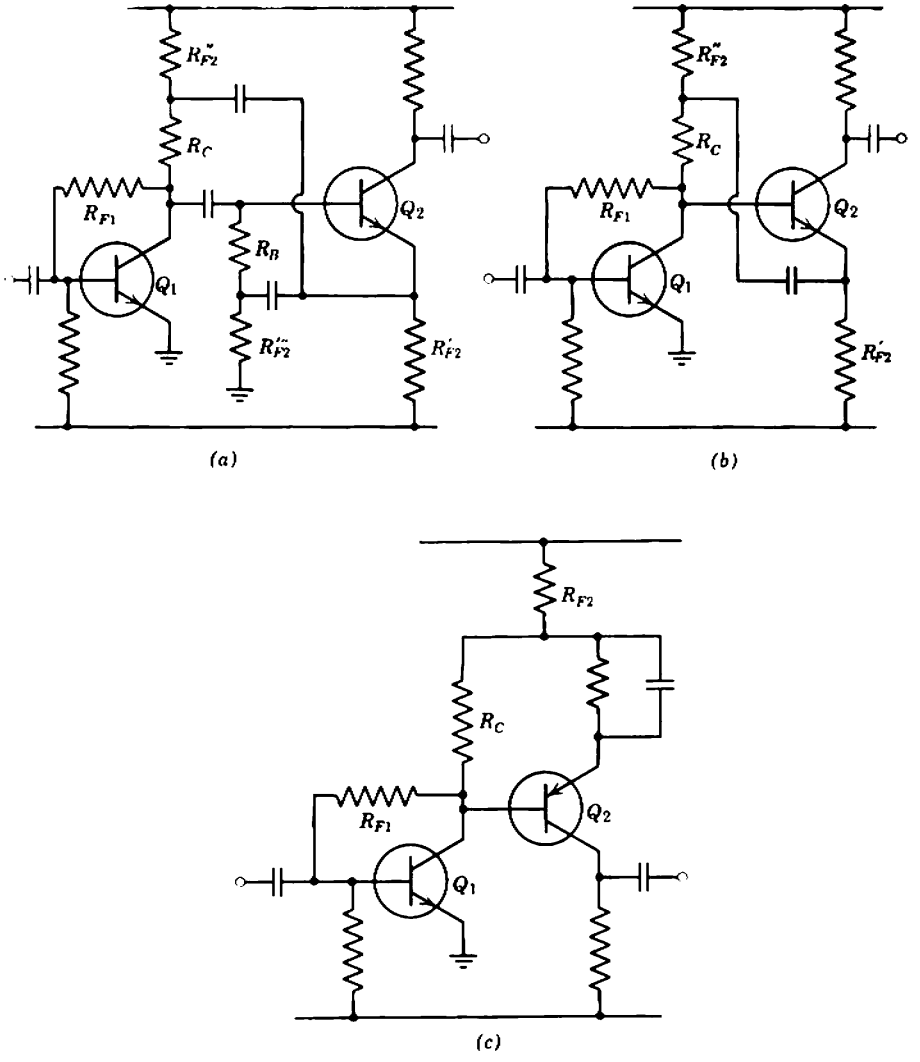


Fig. 11.24 Bootstrapping as applied to transistor circuits.

parameters and the biasing or load resistances. This suggests that the feedback resistors might be used to vary the gain in an accurately controlled way. The two major applications are amplifiers using feedback attenuators (in contrast to the passive attenuators of Section 7.7) and the special amplifiers whose mid-band gain varies with frequency (introduced in Section 7.3.1).

The important design rule for these amplifiers is that the loop gain should be high for all the values of the feedback impedance, not just the

value that gives minimum transmission. Provided this condition is satisfied, the transfer admittance of a series-feedback stage and the transfer impedance of a shunt-feedback stage are given respectively by

$$Y_T \approx \frac{1}{Z_F}, \tag{11.106}$$

$$Z_T \approx -Z_F. \tag{11.107}$$

11.6.2.1 Feedback Attenuators

In amplifiers using feedback gain controls, one or more feedback resistors is made variable. If the control is used simply to preset the gain at some specified value (and then not changed) the design is straightforward, but care must be taken with the biasing circuit if the gain is to be varied during normal operation. The variable feedback resistor must not be a part of the biasing system, because changing the gain will cause a variation in the quiescent point and this can have two undesirable consequences:

(i) In extreme cases, the change in the quiescent point may be so large that the amplifying device moves out of its linear range of operation.
(ii) The variation of quiescent point is a low-speed transient which is amplified by the subsequent stages. These stages may overload and produce gross distortion during the transient.

Figure 11.25 shows the preferred methods for connecting gain-control potentiometers into transistor series- and shunt-feedback stages; vacuum-tube stages follow identical principles. In either case the effective feedback resistance is the parallel combination of the biasing resistor and the

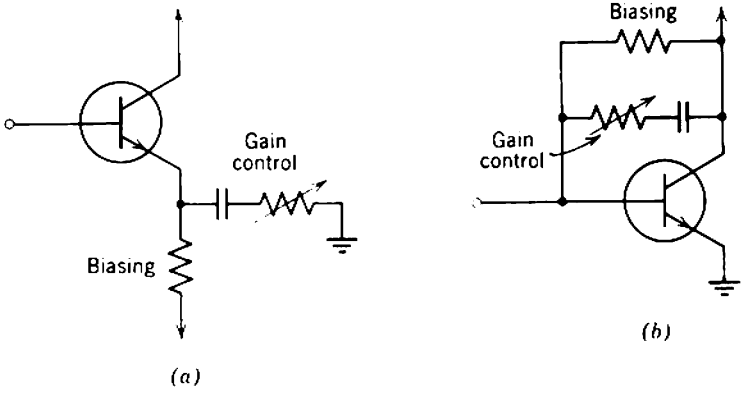


Fig. 11.25 Preferred methods for connecting feedback attenuators into series- and shunt-feedback stages.

potentiometer. The maximum feedback resistance (corresponding to infinite resistance of the potentiometer) is the biasing resistor.

Passive attenuators can have an infinite ratio of maximum to minimum gain because their transmission can be reduced to zero. In contrast, feedback gain controls have a relatively small gain ratio because they cannot reduce the over-all transmission to zero. A practical limit is about ten-to-one; larger ratios imply very high loop gain at minimum transmission, and regeneration may occur.

11.6.2.2 Varying Gain-Versus-Frequency Amplifiers

Frequency-dependent feedback impedances built up from resistors, capacitors, and possibly inductors can be used to give a gain that varies in a controlled fashion with signal frequency. For a shunt-feedback stage the poles and zeros of the feedback impedance are the poles and zeros of the transfer impedance provided the loop gain is large. For a series-feedback stage the poles of the feedback impedance become zeros of the transfer admittance and vice versa.

Inductors are not only more expensive than resistors and capacitors, they are less perfect; therefore, inductors should be avoided as far as possible. If inductors are excluded from the feedback impedances, a series-feedback stage can give only a gain characteristic that rises with increasing frequency whereas a shunt stage can give only a falling gain/frequency characteristic. An example of an amplifier with controlled frequency response appears in Section 11.7.2.

11.7 CONCLUDING COMMENTS ON DESIGN

This chapter is rather unbalanced, in that more space is given to complications and refinements than to the main principles. For everyday circuit design, the important material is

(i) the expressions for G_T and R_T of the series- and shunt-feedback stages, respectively, listed in Section 11.1,
(ii) the principle of the alternate cascade as set out in Section 11.2.1.

The main complication is interaction between stages, which gives rise to a theoretical limit on the stable gain. The advantage of the alternate cascade over cascades of similar stages is that interaction is minimized through the use of impedance mismatches, and is not significant unless the gain approaches the theoretical limit. For example, compare the approximate and accurately calculated gains of the two amplifiers described in Section 11.7.1. However, the circuit designer should understand how interaction

occurs even in the alternate cascade, as he can then avoid complications in the majority of circuit designs. Further, in cases where maximum gain is required, it is obvious which steps should be taken in order to approach the theoretical limit for the vacuum tubes or transistors employed. Such a design is illustrated in Section 11.7.3.

11.7.1 The Convenience of the Alternate Cascade

The greatest convenience of the alternate cascade is that, unless the total gain approaches a limiting value, the gains of the individual stages can be calculated quite independently and then multiplied to give the over-all gain. Figures 11.26 and 11.27 are circuit diagrams for $\times 100$ voltage amplifiers using transistors and vacuum tubes, respectively. These illustrate the convenience and simplicity of the alternate cascade.

11.7.1.1 Transistor Amplifier

A first estimate of the voltage gain of the amplifier follows from Eq. 11.50 as

$$A_V = G_{T1} \times R_{T2},$$

and using the approximate forms of Eqs. 11.22 and 11.30

$$A_V \approx \left(\frac{1}{R_{F1}}\right) \times R_{F2} = \frac{22{,}000}{150} = 147.$$

This estimate is of course high, because it assumes infinite loop gain in both stages. An intelligent guess is that the gain will be 20% less per stage, that is, about 100 over-all.

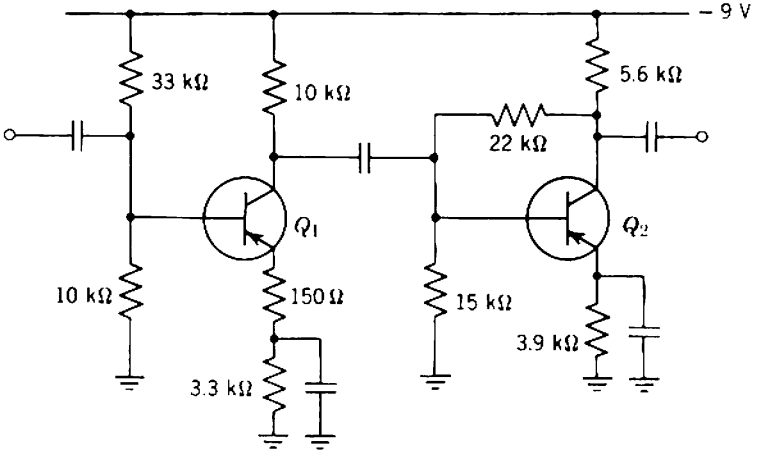


Fig. 11.26 $\times 100$ voltage amplifier. Any medium-gain transistors may be used.

The biasing circuits are such that both transistors operate at about 0.5 mA emitter current, a convenient value for a small-signal amplifier. Both transistors therefore have

$$r_E = 50\ \Omega$$

and substitution of data into Eq. 11.22 gives the transfer conductance of the first stage as

$$G_{T1} = 4.75\ \text{mA/V}.$$

Notice that r_B/β_N for any transistor likely to be used makes only a minute correction; the gain is virtually independent of Q_1. The transfer resistance of the second stage follows from Eq. 11.30, and depends slightly on both β_N and R_L. Assuming that the transistor is a medium-gain type (average β_N about 80) and that the 5.6 kΩ collector supply resistor is the only load,

$$R_{T2} = 20.6\ \text{V/mA}.$$

Hence the over-all voltage gain is

$$A_V = G_{T1} \times R_{T2} = 98.0.$$

To find the gain of the amplifier exactly, it is necessary to make a correction for the coupling efficiency between stages. From Eq. 11.34, the input resistance of the shunt-feedback stage is

$$R_{i2} = 260\ \Omega.$$

The effective biasing resistor R_P is the parallel combination of the 10 kΩ collector supply resistor with the 15 kΩ base biasing resistor, that is, 6 kΩ. Therefore, from Eq. 11.58, the coupling efficiency is

$$\eta = 0.96$$

and the true voltage gain is

$$A_V = 94.$$

Thus the correction due to interaction between stages is in fact small, one of the desirable features of the alternate cascade.

At most the gain could vary by ± 10% from this last calculated value if both transistors were at the extreme high or low tolerance limits of β_N. Unless the 150 Ω and 22 kΩ feedback resistors are 5% tolerance types, the gain tolerance due to β_N is completely masked by resistor tolerances. Even with 5% resistors the possible tolerances due to resistors and β_N are only equal.

The input resistance of the first transistor follows from Eq. 11.26 as 16 kΩ. In combination with the 33- and 10-kΩ biasing resistors, the total input resistance of the amplifier is 5 kΩ. If two of the amplifiers are

cascaded, the input resistance of the second amplifier loads the shunt-feedback output stage of the first amplifier and reduces its gain. From Eq. 11.30 the transfer resistance of the second stage in the first amplifier falls by 4.5%, from 20.6 to 19.7 V/mA. This is a small change; to a good approximation, the amplifiers can be cascaded without interaction.

11.7.1.2 Vacuum-Tube Amplifier

The first estimate of the voltage gain of the amplifier shown in Fig. 11.27 is

$$A_V = G_{T1} \times R_{T2} \approx \left(\frac{1}{R_{F1}}\right) \times R_{F2} = \frac{68{,}000}{5600} = 121.$$

The corrections necessary to find the gain exactly are similar to those for the transistor amplifier considered above.

FIRST STAGE

$$I_K = 2 \text{ mA},$$
$$V_{AK} = 100 \text{ V}.$$

Therefore

$$g_m = 1.5 \text{ mA/V}$$

and from Eq. 11.13

$$G_{T1} = 159 \ \mu\text{A/V}.$$

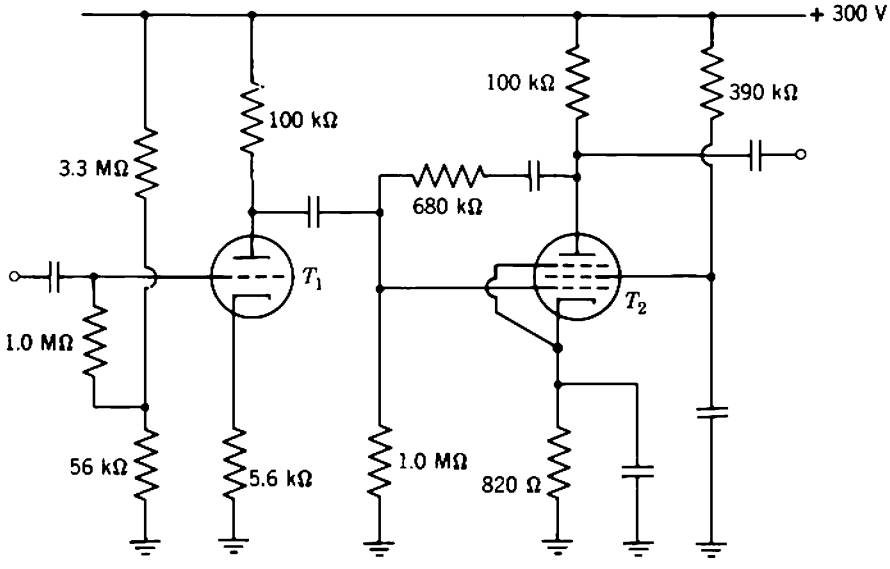


Fig. 11.27 × 100 voltage amplifier, using type ECF80/6BL8 triode-pentode.

SECOND STAGE

$$I_K = 2.5 \text{ mA},$$
$$V_{SK} = 80 \text{ V}.$$

Therefore

$$g_m = 2.5 \text{ mA/V}$$

and from Eqs. 11.16 and 11.20

$$R_{T2} = 652 \text{ V/mA},$$
$$R_{i2} = 3.12 \text{ k}\Omega.$$

COUPLING EFFICIENCY. From Eq. 11.58

$$\eta = 0.97.$$

VOLTAGE GAIN. From the above information,

$$A_V = G_{T1} \times \eta \times R_{T1} = 105.$$

It is left as an exercise for the reader to investigate the effect of cascading these amplifiers.

11.7.2 The Versatility of the Alternate Cascade

Figure 11.28 is the simplified circuit diagram* for a phonograph preamplifier. This circuit illustrates the application of frequency-selective feedback to yield a controlled gain/frequency characteristic. A phonograph preamplifier should have the following characteristics:

1. Sufficient mid-band gain to amplify the very low output of the pickup cartridge to a value suitable for the main amplifier. In the example, the maximum mid-band gain is about 200.
2. A high peak-signal-to-noise ratio, at least 60 dB. In the example, the output of the pickup cartridge is 3 mV rms at 10 cm/sec recorded velocity, and S_o/N_o is 80 dB.
3. A low distortion level at peak output. In the example, the preamplifier is designed to deliver 250 mV rms into a high-impedance load, with about 0.008% distortion.
4. A gain/frequency characteristic that compensates for a frequency weighting introduced in the recording process. In the example, the law followed is the R.I.A.A. characteristic whose asymptotes are shown in Fig. 11.29*a*.
5. Tone controls which can boost or cut the bass (frequencies below about 500 Hz) and treble (frequencies above about 2 kHz).

---

* The complete circuit is given in E. M. CHERRY, "An engineering approach to the design of transistor feedback amplifiers," *J. Brit. Instn. Radio Engrs.*, **27**, 127, February 1963.

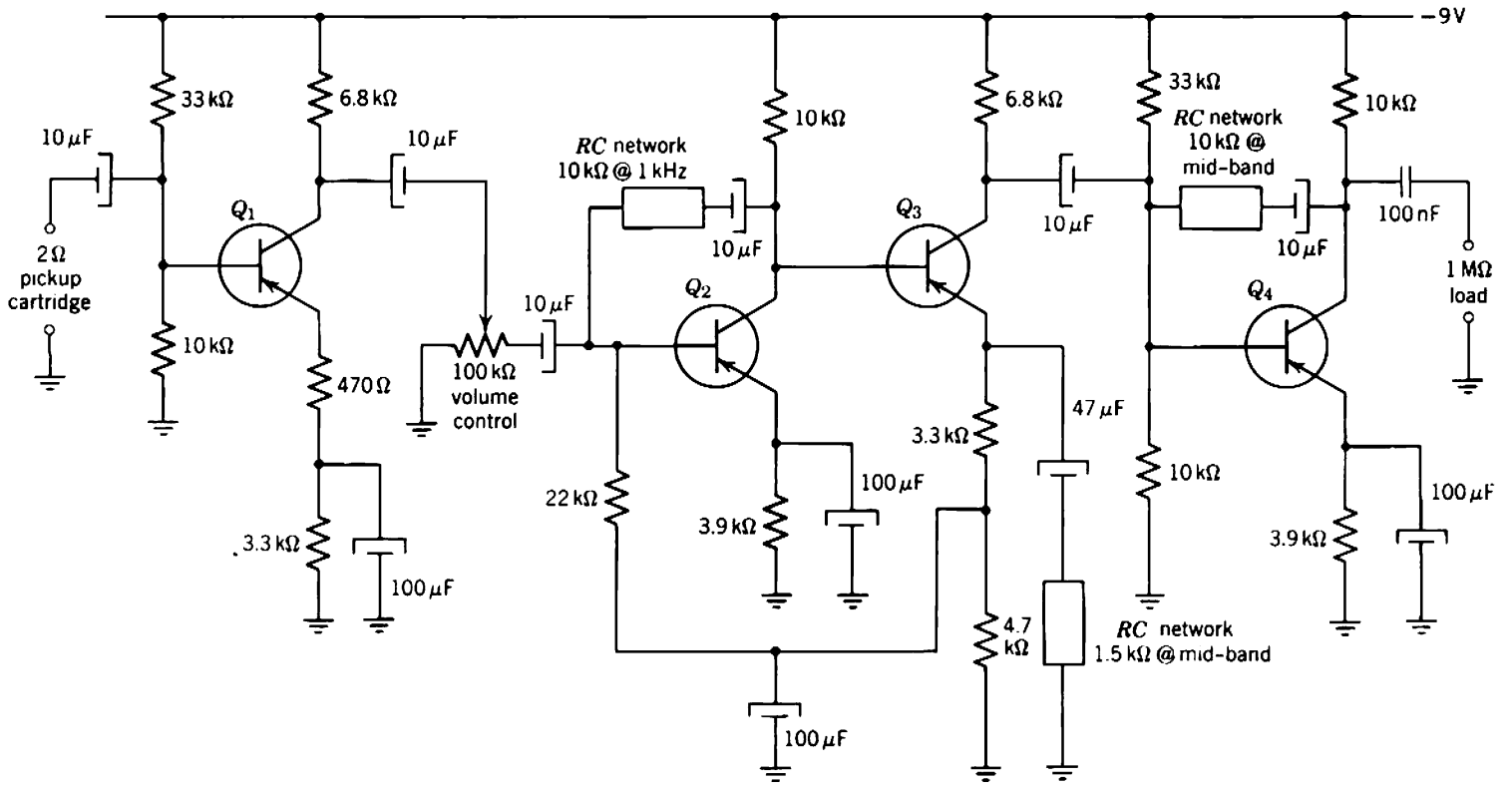


Fig. 11.28 Simplified circuit diagram for a phonograph preamplifier. Details of the *RC* networks appear in Figs. 11.29 and 11.30. All transistors are type OC44.

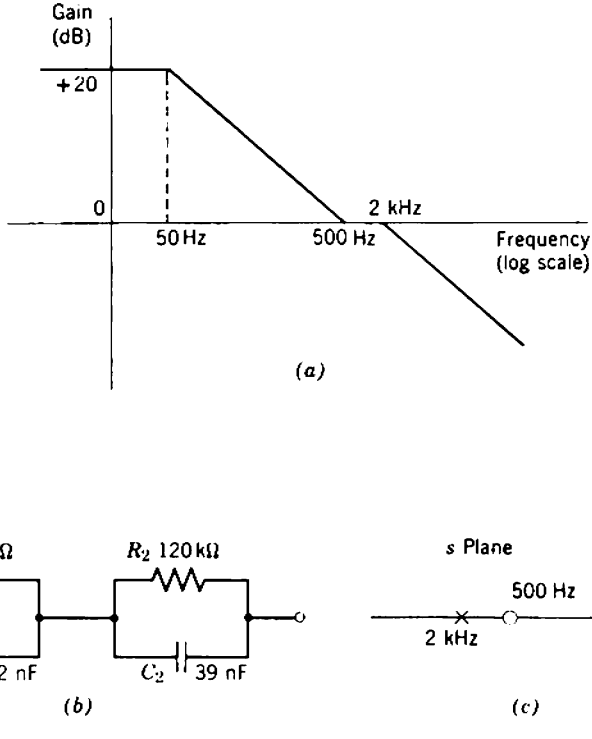


Fig. 11.29 The feedback impedance for Q_2: (*a*) desired R.I.A.A. equalizing characteristic; (*b*) the *RC* feedback network Z_F; (*c*) the poles and zeros of Z_F.

In the circuit diagram, Q_1 is a low-noise series-feedback stage whose transfer conductance is about 2 mA/V. Q_2 is a shunt-feedback stage whose 1 kHz transfer impedance is about 9 V/mA, but the feedback impedance is frequency-dependent; this stage provides equalization for the R.I.A.A. recording characteristic. Q_3 is a series-feedback stage whose feedback impedance can be made frequency-dependent. This stage can provide a response that rises with increasing frequency (that is, bass cut and treble boost) and its mid-band transfer conductance is about 1 mA/V. Finally, Q_4 is a shunt-feedback stage that can provide a falling gain/frequency characteristic (bass boost and treble cut) and its mid-band transfer resistance is about 10 V/mA. A passive current attenuator (the volume control) is provided between Q_1 and Q_2 to vary the gain of the preamplifier. Q_2 is direct coupled to Q_3 to reduce loading by the biasing resistors; the biasing feedback is taken from a point in the emitter circuit of Q_3 that is bypassed to ground, to reduce signal feedback via the biasing network. All transistors operate at about 0.5 mA emitter current, with 3 V between emitter and collector.

The input series-feedback stage is conventional, having a flat gain versus frequency characteristic. However, the noise performance of this stage warrants attention as the noise in the input stage of an amplifier is almost always the dominant component. There are two principal sources of noise in this stage—the transistor itself and the 470-Ω feedback resistor. Equation 11.91 shows that the noise voltage at the input is the sum of the transistor and feedback resistor spot noise voltage generators, given respectively by Eqs. 8.27 and 2.61 as

$$d(v_N^2)_D = 1.6 \times 10^{-18}\ \text{V}^2/\text{Hz},$$

$$4kTR_F\, df = 7.5 \times 10^{-18}\ \text{V}^2/\text{Hz}.$$

Both generators are substantially white and the total spot noise voltage is

$$d(v_N^2) = 9.1 \times 10^{-18}\ \text{V}^2/\text{Hz}.$$

Over a 10 kHz bandwidth, the noise is 320 nV rms. With a peak signal input of 3 mV rms, the signal-to-noise ratio is 79.4 dB which is an excellent performance.

Q_2 is the equalizing stage whose gain versus frequency asymptotes are shown in Fig. 11.29*a*. The load resistance for this stage is dominated by its 10-kΩ collector-supply resistor, and its feedback impedance is shown in Figs. 11.29*b* and *c*. At a frequency of 1 kHz, the feedback impedance is 9.2 kΩ; the loop gain is large and the transfer impedance is about 9 V/mA. Moreover, the zero and the pole in the feedback impedance at 500 Hz and 2 kHz appear in the transfer impedance. However, at frequencies around 50 Hz, the feedback impedance is about 120 kΩ; the loop gain is quite small, so the pole at 34 Hz in the feedback impedance is shifted appreciably in the transfer impedance. The 120-kΩ feedback resistor is calculated to given 90 V/mA transfer resistance at low frequencies (20 dB boost relative to 1 kHz), taking account of the finite loop gain. If the loop gain were large, the resistor would be 82 kΩ.

Q_3 and Q_4 provide a mid-band gain of 10, and provide tone control facilities through the frequency-dependent feedback impedances shown in Fig. 11.30. Notice that the total feedback impedance for Q_3 is the parallel combination of its 3.3-kΩ emitter biasing resistor with the 1.5-kΩ network.

The nonlinearity of the preamplifier occurs principally in Q_3 and in the coupling efficiency into Q_4. For 250 mV rms output (= 700 mV peak-to-peak) at mid-band frequencies, the signal current in Q_4 follows from Eq. 11.100 as 140 μA peak-to-peak and, since the transfer resistance of this

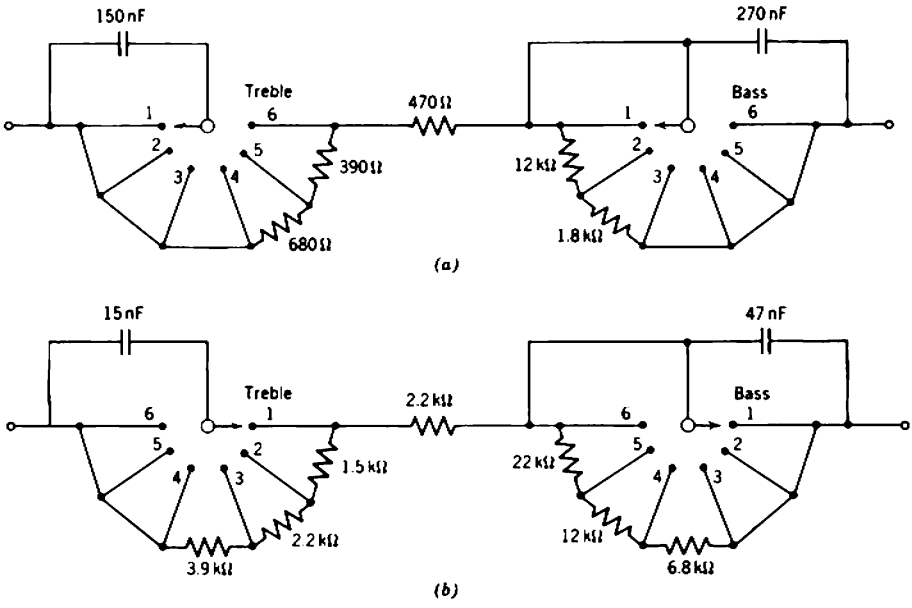


Fig. 11.30 The feedback impedances for Q_3 and Q_4: (*a*) 1.5 kΩ in the emitter of Q_3; (*b*) 10 kΩ in the collector of Q_4. The switch positions are:

	Treble	Bass
1	−12 dB	−8 dB
2	−8 dB	−4 dB
3	−4 dB	flat
4	flat	+4 dB
5	+4 dB	+8 dB
6	+8 dB	+12 dB

stage is 10 V/mA, the signal current in Q_3 is 70 μA peak-to-peak. The differential error at the signal peaks follows from Eq. 11.101*b* as

$$\gamma' = -0.0203\%,$$
$$\gamma'' = -0.0790\%.$$

(Equations 11.22, 11.34, and 11.58 are used in the calculation.) Therefore, from Eq. 9.22, the second- and third-harmonic distortions are 0.007% and 0.004%, respectively; the rms total harmonic distortion follows as 0.008%.

11.7.3 The Potential Capabilities of the Alternate Cascade

Figure 11.31 is the circuit diagram for a transistor current amplifier designed to give a high, stable gain, close to the ideal limit of the transistors.

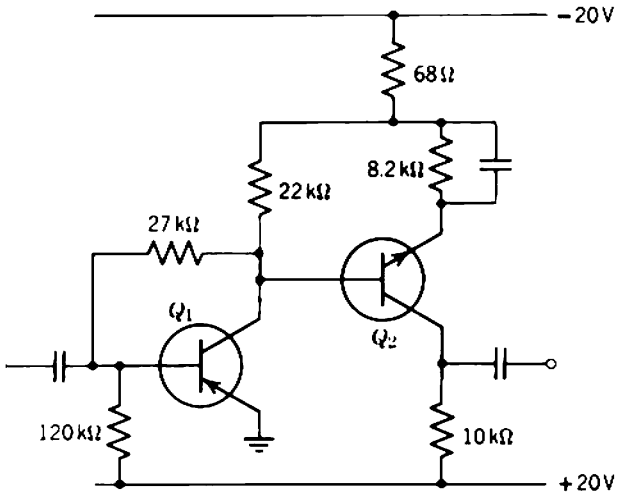


Fig. 11.31 High-current-gain transistor amplifier.

The gain is about 280, with $\pm 10\%$ tolerance in worst cases if the transistors have average β_N about 60.

The two stages are direct-coupled, with the collector supply resistor of the first stage bootstrapped to the emitter of the second; *p-n-p* and *n-p-n* transistors are alternated to give convenient dc conditions. The second transistor operates at a rather high current (about 2 mA) to reduce its input resistance and hence reduce loading by the biasing resistors in comparison. In addition, the first transistor operates at a low current (about 0.5 mA) so its collector supply resistor can be large.

The transfer conductance of the second stage follows from Eq. 11.22 as

$$G_{T2} \doteq 11.5 \text{ mA/V},$$

and the loop gain follows from Eq. 11.28 as

$$|A_l| = 4.$$

Equation 11.105 gives the total input resistance (including the 22-kΩ collector-supply resistor) as

$$R_{i2} = 5.0 \text{ k}\Omega.$$

Equations 11.30 and 11.34 give the transfer resistance and input resistance of the first stage as

$$R_{T1} = 24.4 \text{ V/mA},$$

$$R_{i1} = 352 \ \Omega.$$

Finally, the current gain of the two stages is given by Eq. 11.51 as

$$A_I = 280.$$

If two of these amplifiers are cascaded, the coupling efficiency between them must be taken into account in calculating the over-all current gain. The resistance shunting the signal path to ground is the 10-kΩ collector supply resistor of the second stage in the first amplifier, in shunt with the 120-kΩ base biasing resistor of the first stage in the second amplifier. These amount to 9.25 kΩ, and in combination with the 352-Ω input resistance proper, the coupling efficiency follows from Eq. 11.58 as

$$\eta = 0.963.$$

Chapter 12

Low-Frequency Compensation of Amplifiers

Chapter 7 contains a fairly complete discussion of the singularity patterns of elementary nonfeedback amplifier stages, and Section 7.8 gives two examples of amplifier design. The design process in these examples is somewhat rudimentary:

(i) capacitances and inductances are calculated to set all low-frequency poles coincident at a position obtained from a quoted formula;
(ii) no consideration is given to the fact that some component values might be inconveniently large;
(iii) the positions of the high-frequency singularities are merely calculated from the various circuit parameters.

This and the following chapter are concerned with *designing* amplifiers to have a specified dynamic response. In each case, the treatment begins with a discussion of the dynamic response of physically realizable singularity patterns and ends with circuit details that locate singularities accurately. Series and shunt feedback form the basis of many such circuits, so analysis of the two feedback stages is an integral part of the chapters.

The term *compensation* is applied to modifying an amplifier to improve its dynamic response. It might be argued that an amplifier should be "designed" from the outset to have the optimum response, so that no "compensation" is required. However, design is essentially a trial-and-error process of weighing one improvement against another, and compensation (in the sense of choosing the best compromise) is thus an integral part. Compensation (in the sense of experimental fiddling with an

amplifier after it has been constructed) is a thoroughly reprehensible practice.

Inspection of Sections 7.4 and 7.5 shows that, apart from the inconvenience associated with very large capacitors, the low-frequency singularities can be placed at any finite point on the negative real axis. There is therefore no fundamental limitation to the dynamic response at low frequencies. Indeed, all low-frequency singularities can be eliminated through the use of a dc amplifier as in Chapter 15. In contrast, the product of the distances of the high-frequency poles from the origin is set by the stage gain and realizable device gain-bandwidth product. The realizable $\mathscr{G}\mathscr{B}$ is limited by the intrinsic $\mathscr{G}\mathscr{B}$, and this in turn is limited by the mean transit time of carriers through the control region (Eq. 2.60). High-frequency limitations to the dynamic response are therefore fundamental.

This difference in nature between the low- and high-frequency limitations to dynamic response results in different approaches to compensation. Formally, high-frequency compensation involves complicated mathematical methods for finding the optimum locations for the singularities; often some of the poles should occur as complex conjugate pairs. There are, however, only two basic methods for producing complex poles (inductance and feedback peaking described in Sections 13.4 and 13.5), so the mechanics of high-frequency compensation are quite simple. Fortunately, much of the mathematics can be bypassed also. Low-frequency compensation is more in the nature of ingenious circuit design which seeks an acceptable response with convenient capacitor sizes. There is no pressing need to achieve the ultimate performance from given capacitors because larger capacitors can be used. The same situation does not exist at high frequencies; the limitation of gain-bandwidth product is fundamental.

12.1 SINGULARITY PATTERNS

The following sections list some simple rules for calculating the low-frequency dynamic response of an amplifier from its singularity pattern, and for selecting a singularity pattern that yields a desirable response. It is assumed that all singularities lie on the negative real axis.

12.1.1 Rules for Combining the Effects of Low-Frequency Singularities

The general form of the low-frequency factor $\psi_k(s)$ associated with any bypass, coupling or decoupling capacitor, or associated with a transformer

primary inductance is

$$\psi_k(s) = \frac{s - z}{s - p} = \frac{s + 1/\tau_2}{s + 1/\tau_1} = \frac{s + \omega_2}{s + \omega_1}. \tag{12.1}$$

The time constants τ_1 and τ_2 associated with the kth ψ factor are more often used in expressions for transient response, whereas the corner frequencies are used in the frequency response:

$$\omega_1 = \frac{1}{\tau_1},$$

$$\omega_2 = \frac{1}{\tau_2}.$$

Each of the $\psi_k(s)$ factors has one pole and one zero, and for bypass and decoupling capacitors neither singularity is at the origin. In these two cases it is convenient to split Eq. 12.1 into two terms each of which does have one singularity at the origin:

$$\psi_k(s) = \psi_{k1}(s) \times \psi_{k2}(s), \tag{12.2}$$

where

$$\psi_{k1}(s) = \frac{s}{s - p} = \frac{s}{s + 1/\tau_1} = \frac{s}{s + \omega_1} \tag{12.3}$$

and

$$\psi_{k2}(s) = \frac{s - z}{s} = \frac{s + 1/\tau_2}{s} = \frac{s + \omega_2}{s}. \tag{12.4}$$

These terms with a singularity at the origin are easier to handle than the more general term. On a frequency basis Eq. 12.3 corresponds to a response which falls with a simple 6 dB/octave asymptote for decreasing frequency, whereas Eq. 12.4 corresponds to a rising asymptote. On a time basis, Eq. 12.3 corresponds to a falling exponential, whereas Eq. 12.4 corresponds to a rising ramp. In calculating the response of a multistage amplifier, the positions of all low-frequency singularities are found first. Next, the number of canceling pole-zero pairs required to split the $\psi_k(s)$ terms into two are added at the origin, and finally the simple responses of all $\psi_{k1}(s)$ and $\psi_{k2}(s)$ are combined.

12.1.1.1 Frequency Response

The modulus and phase of $\psi_{k1}(s)$ and $\psi_{k2}(s)$ for a sinusoidal signal ω are:

$$|\psi_{k1}(j\omega)| = \left(\frac{\omega^2}{\omega^2 + \omega_1^2}\right)^{1/2}, \tag{12.5}$$

$$\underline{/\psi_{k1}(j\omega)} = \arctan\left(\frac{\omega_1}{\omega}\right), \tag{12.6}$$

$$|\psi_{k2}(j\omega)| = \left(\frac{\omega^2 + \omega_2^2}{\omega^2}\right)^{1/2}, \tag{12.7}$$

$$\underline{/\psi_{k2}(j\omega)} = \arctan\left(\frac{-\omega_2}{\omega}\right). \tag{12.8}$$

Figures 12.1 and 12.2 show logarithmic plots for one decade in frequency either side of the corner frequencies ω_1 and ω_2. The complete response of an amplifier with a number of different corner frequencies is found in magnitude and phase by adding the ordinates of these curves with the abscissas shifted horizontally by an amount appropriate for the difference in corner frequencies.

Some approximations can be made to simplify calculating the contributions of corner frequencies that lie well below the passband. For these

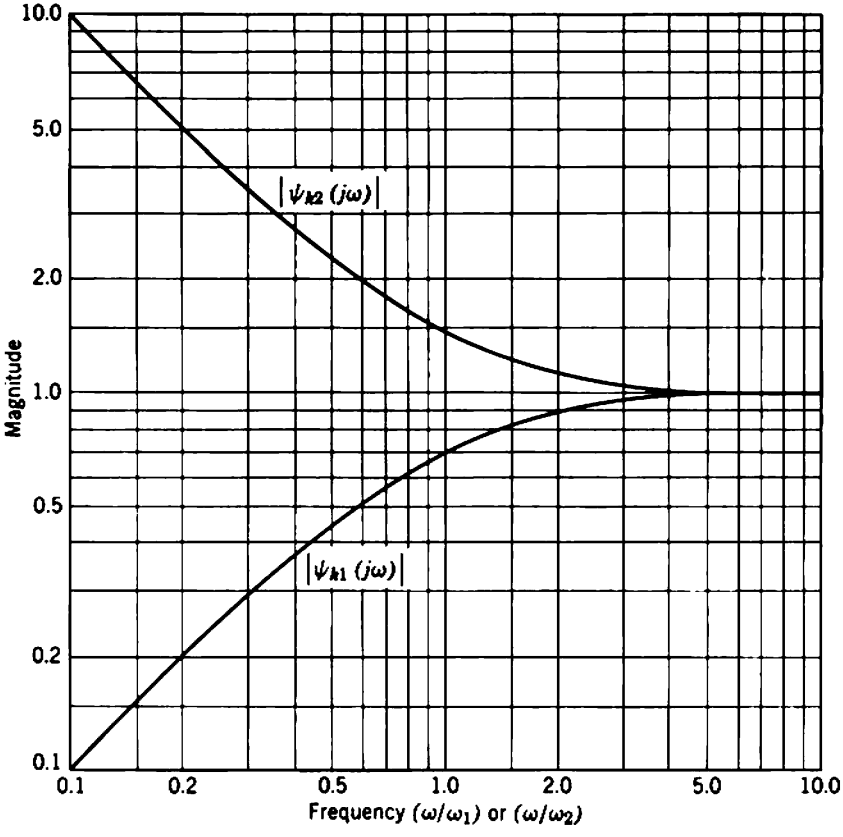


Fig. 12.1 Normalized frequency response of $\psi_{k1}(s)$ and $\psi_{k2}(s)$: magnitude.

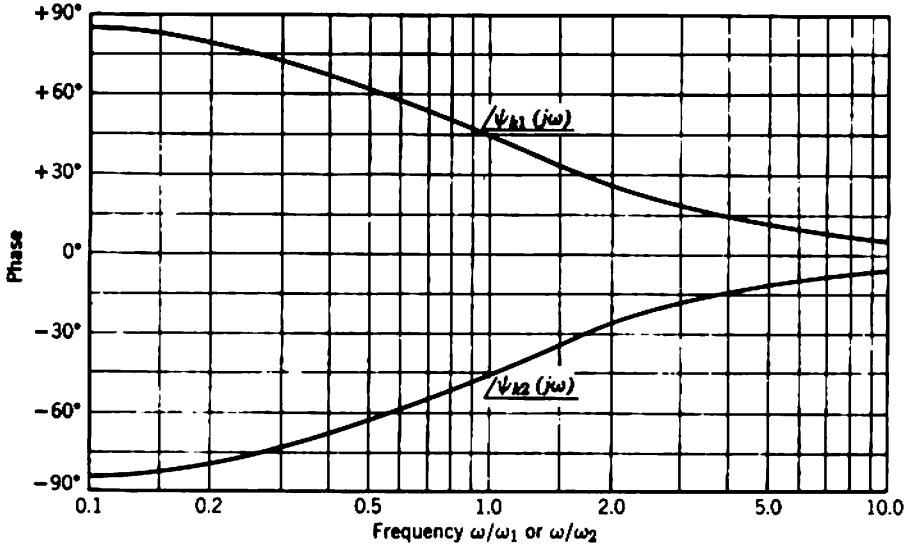


Fig. 12.2 Normalized frequency response of $\psi_{k1}(s)$ and $\psi_{k2}(s)$: phase.

corner frequencies, (ω_1/ω) and (ω_2/ω) are very small in the passband, and

$$|\psi_{k1}(j\omega)| \approx 1 - \frac{1}{2}\left(\frac{\omega_1}{\omega}\right)^2, \tag{12.9}$$

$$\underline{/\psi_{k1}(j\omega)} \approx \frac{\omega_1}{\omega}, \tag{12.10}$$

$$|\psi_{k2}(j\omega)| \approx 1 + \frac{1}{2}\left(\frac{\omega_2}{\omega}\right)^2, \tag{12.11}$$

$$\underline{/\psi_{k2}(j\omega)} \approx -\frac{\omega_2}{\omega}. \tag{12.12}$$

The total magnitude response for n_p poles and n_z zeros that lie well below the passband is

$$\left|\prod_{k=1}^{n_p}[\psi_{k1}(j\omega)]\right| \times \left|\prod_{k=1}^{n_z}[\psi_{k2}(j\omega)]\right| = 1 - \frac{1}{2\omega^2}\left[\sum_{k=1}^{n_p}(\omega_1{}^2) - \sum_{k=1}^{n_z}(\omega_2{}^2)\right] \tag{12.13}$$

and similarly the phase response is

$$\underline{/\prod_{k=1}^{n_p}[\psi_{k1}(j\omega)]} - \underline{/\prod_{k=1}^{n_z}[\psi_{k2}(j\omega)]} = \frac{1}{\omega}\left[\sum_{k=1}^{n_p}(\omega_1) - \sum_{k=1}^{n_z}(\omega_2)\right]. \tag{12.14}$$

In most practical design work the lower 3-dB frequency can be evaluated to within about 1 octave by inspection. If an accurate value is required, the curves of Figs. 12.1 and 12.2 are used to find the contributions of the singularities within 2 or 3 octaves of the first approximation, and Eqs. 12.13 and 12.14 are used for the singularities outside this range. Notice that, because of the squares in Eq. 12.13, singularities more than 1 decade outside the passband have less than 1% effect on the magnitude response inside the passband. However, the phase contribution is not less than 1% of 1 radian until the singularities lie 2 decades outside the passband.

12.1.1.2 Time Response

The response to a unit step of a network that has a transfer function $\psi_{k1}(s)$ is

$$\rho_1'(t) = \exp\left(\frac{-t}{\tau_1}\right), \tag{12.15}$$

whereas the response to the step-plus-ramp $(1 + t/T)$ is

$$\rho_1''(t) = \frac{\tau_1}{T} + \left(1 - \frac{\tau_1}{T}\right)\exp\left(\frac{-t}{\tau_1}\right). \tag{12.16}$$

These exponentials can be approximated by the first two terms in their expansions for values of t that are short compared with the time constant τ_1:

$$\rho_1'(t) \approx 1 - \left(\frac{t}{\tau_1}\right), \tag{12.17}$$

$$\rho_1''(t) \approx 1 - t\left(\frac{1}{\tau_1} - \frac{1}{T}\right). \tag{12.18}$$

A square wave consists of a series of steps in opposite directions, equally spaced in time. If a square wave passes through a network with transfer function $\psi_{k1}(s)$, it follows from Eq. 12.15 that the output consists of a series of exponentials. However, if the period t_0 of the square wave is short compared with the time constant τ_1, the output can be approximated by a set of sloping straight lines (Eq. 12.17). Figure 12.3 shows that the fractional tilt η on a square of repetition frequency f_0 $(= 1/t_0)$ is

$$\eta \approx \frac{1}{4f_0\tau_1} = \frac{\pi}{2}\left(\frac{f_1}{f_0}\right), \tag{12.19}$$

where f_1 is the corner or pole frequency in hertz. Because a square wave with tilt can be approximated by a series of step-plus-ramp functions equally spaced in time, it follows from Eq. 12.18 that the tilt on the output waveform is

$$\eta_o \approx \eta_i + \frac{1}{4f_0\tau_1}, \tag{12.20}$$

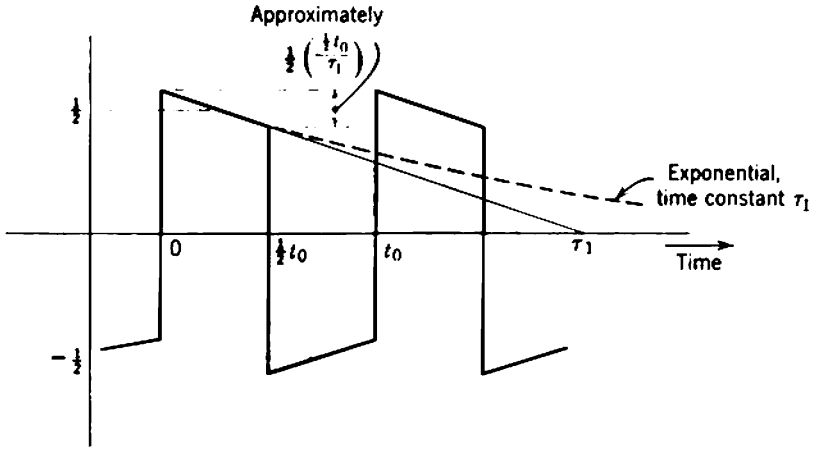


Fig. 12.3 Normalized square-wave response of $\psi_{k1}(s)$.

where η_i is the input tilt. Thus, the output tilt is the sum of the input tilt plus the tilt that would be produced on a true square wave. Notice that these equations apply only for small tilts for which the exponential can be approximated by a straight line—10%, or at the most 20%; the corner frequency (in hertz) must not be more than about one tenth of the repetition frequency.

A parallel development can be made for a network whose transfer function is $\psi_{k2}(s)$. The response to a unit step is exactly

$$\rho_2'(t) = 1 + \frac{t}{\tau_2}, \tag{12.21}$$

so that the fractional tilt on a square wave of repetition frequency f_0 is

$$\eta = -\frac{1}{4f_0\tau_2} = -\frac{\pi}{2}\left(\frac{f_2}{f_0}\right). \tag{12.22}$$

As shown in Fig. 12.4, the negative tilt corresponds to a rising slope. The response to a step-plus-ramp is

$$\rho_2''(t) = 1 + t\left(\frac{1}{\tau_2} + \frac{1}{T}\right) + \frac{t^2}{2\tau_2 T} \tag{12.23}$$

and for small values of t

$$\rho_2'' \approx 1 + t\left(\frac{1}{\tau_2} + \frac{1}{T}\right). \tag{12.24}$$

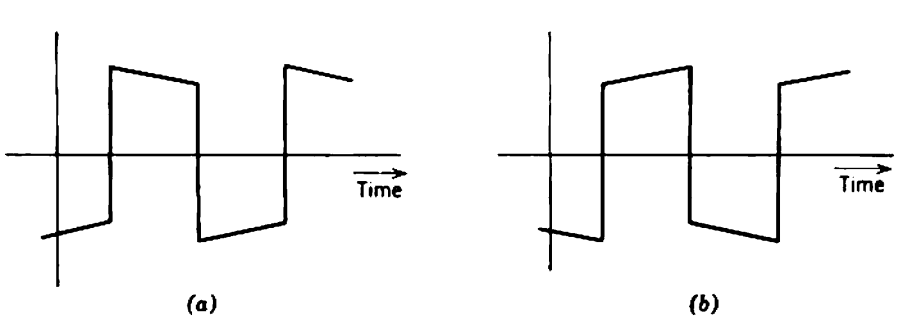


Fig. 12.4 Tilt on a square wave: (*a*) positive tilt; (*b*) negative tilt.

The tilt at output when the input is a square wave with tilt η_i is

$$\eta_o \approx \eta_i - \frac{1}{4f_0\tau_2} \tag{12.25}$$

and again the tilts add algebraically.

From the above discussion it is apparent that the time response of an amplifier can be calculated by simply adding the tilts (either fractional or percentage) due to the various $\psi_{k1}(s)$ and $\psi_{k2}(s)$, provided these tilts are small. Notice that the tilts due to the $\psi_{k1}(s)$ are positive, whereas those due to the $\psi_{k2}(s)$ are negative. Notice too that all ψ factors are commutative, so that the order in which they occur in the stages of an amplifier does not matter. Consider, for example, an amplifier that has a number of $\psi_{k1}(s)$ in its early stages and a number of $\psi_{k2}(s)$ in the later stages, but has been designed so that the sum of the tilts is zero. There will be a very large tilt on a square wave near the middle of the amplifier, but this will be corrected by the later stages even though it is large.

12.1.2 Compensated Patterns

The principle of all low-frequency compensating circuits is to introduce a number of $\psi_{k2}(s)$ into the transfer function, so that their rising frequency response and negative tilt improve the over-all response. Expressed another way, a number of real-axis zeros are introduced into the transfer function, and these partially offset the effect of the poles. All capacitors introduce a zero as well as a pole. The zero from a bypass or coupling capacitor, however, lies closer to the origin than the associated pole (that is, $|z| < |p|$), so that the net contribution of these capacitors can only be a falling frequency response or positive tilt, and no over-all compensation is achieved. Decoupling capacitors, on the other hand, contribute a zero that lies further from the origin than the associated pole, and

therefore yield an over-all compensation. Shunt-feedback stages (analyzed in Section 12.2.3) also contribute zeros that can be used for compensation.

12.1.2.1 Time Response

Compensating the time response of an amplifier is inherently easier than compensating the frequency response, and is therefore treated first. There is little more to be said than to reiterate the rule derived in Section 12.1.1.2:

Tilts (percentage or fractional) add algebraically, provided that they are small.

The total tilt produced by an amplifier on a square wave of repetition frequency f_0 follows from Eqs. 12.19 and 12.22 as

$$\eta = \frac{1}{4f_0}\left[\sum_{k=1}^{n}\left(\frac{1}{\tau_1}\right) - \sum_{k=1}^{n}\left(\frac{1}{\tau_2}\right)\right]. \tag{12.26}$$

Notice that the numbers of poles and zeros are equal to each other and to the total number n of capacitors in the circuit. If the total tilt is to be zero,

$$\sum_{k=1}^{n}\left(\frac{1}{\tau_1}\right) = \sum_{k=1}^{n}\left(\frac{1}{\tau_2}\right) \tag{12.27a}$$

or, in terms of the corner frequencies,

$$\sum_{k=1}^{n}(\omega_1) = \sum_{k=1}^{n}(\omega_2). \tag{12.27b}$$

The sum of the distances of the poles from the origin must be equal to the sum of the distances of the zeros.

As discussed in Section 7.4.2 in connection with an *RC* coupling network, the 3-dB cutoff frequency of an amplifier must be very much lower than the repetition frequency of a square wave that is to be passed with small tilt. Thus, all the singularities must be well below the frequency of the square wave, and Eqs. 12.13 and 12.14 can be used to calculate the total magnitude and phase distortion of the individual Fourier harmonics. As shown in the discussion following the derivation of these equations, singularities a decade below the frequency of a sinusoid have negligible effect on its magnitude; therefore, an amplifier with small tilt in its square-wave response introduces no magnitude distortion of the various Fourier harmonics. Comparison of Eqs. 12.14 and 12.27*b* shows that the condition for zero tilt is also the condition for zero phase shift; the equations are in fact the same, with

$$n_p = n_z = n, \quad \text{the total number of capacitors.}$$

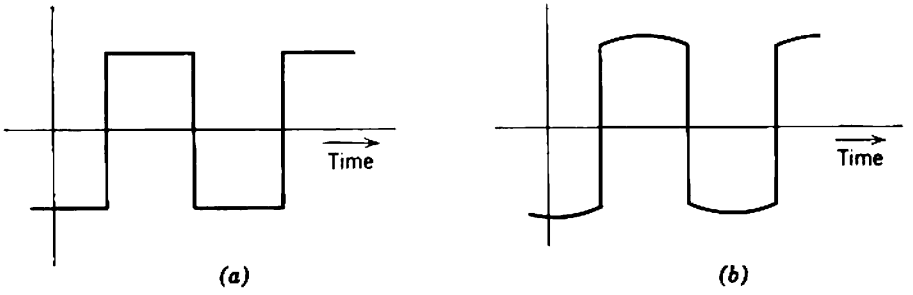


Fig. 12.5 Compensated square-wave response: (*a*) ideal; (*b*) bowed.

The two results, zero magnitude distortion and zero phase shift, are of course a special case of the well-known general condition for distortionless transmission of a waveform, and are hardly unexpected.

The magnitude response to sinusoidal signals of an amplifier that has been compensated for optimum square-wave response is not flat. There is a peak at very low frequencies (well below the frequency of a square wave which can be passed without tilt) where Eqs. 12.13 and 12.14 cease to apply.

A situation sometimes arises in which some of the tilt contributions cannot be reduced below the desirable maximum of 10% with convenient capacitor sizes. It is possible to compensate these larger tilts approximately but, in general, the top of a square wave becomes "bowed" as shown in Fig. 12.5. Optimum compensation corresponds to minimum peak-to-peak variation of the top and, although not strictly correct, it is common usage to describe the waveform distortion as a percentage tilt. Equation 12.27 is not satisfied exactly at optimum compensation; rather

$$\sum_{k=1}^{n} (\omega_1) < \sum_{k=1}^{n} (\omega_2). \tag{12.28}$$

Equation 12.27 serves as a useful starting point for the design of a compensating system, but there is no simple method for calculating the optimum singularity pattern exactly. This is one of the rare occasions when experimental adjustment may be preferable to a theoretical study. As a guiding rule, less than 5% "bowing" can be achieved if no corner frequency (in hertz) is more than one third of the square-wave repetition frequency.

12.1.2.2 Frequency Response

In principle, there is no problem in designing a singularity pattern for an amplifier to compensate the magnitude and phase of the low-frequency sinusoidal response. The compensating zeros are simply positioned to

give the desired total response. A considerable practical difficulty, however, is that there is no simple analytical method for summing the responses of singularities that lie near the cutoff frequency; Eqs. 12.13 and 12.14 apply only for the singularities close to the origin. In general, there is no alternative to a trial-and-error design based on graphical methods. This section lists a number of special cases for which the labor is reduced.

12.1.2.2.1 *Pole-Zero Cancellation*

The technique of designing an amplifier so that some of the low-frequency poles and zeros cancel is most useful, as it reduces the number of effective singularities and therefore simplifies the calculation. Still greater simplification results if the total pattern is reduced to one of the special types described in Sections 12.1.2.2.2 to 12.1.2.2.4, and its response can be calculated easily. Notice that, although exact cancellation of the various poles and zeros is desirable, even an error as large as an octave has quite a small effect on the magnitude response. Also, errors in pole positions due to tolerances on device parameters tend to average out, as it is unlikely that all parameters will have a bias in the same direction. Nevertheless, the more exact formulas for the positions of the poles should be used in preference to the simplest design approximations. Notice too that any *RC*-coupled amplifier stage can be designed so that its low-frequency singularity pattern reduces to a single pole plus a zero at the origin. This follows because each capacitor contributes one pole and one zero; the zero at the origin is due to the coupling capacitor.

12.1.2.2.2 *Multiple Coincident Poles*

There is a simple rule for calculating the 3-dB cutoff frequency of an amplifier that has an n_p-order multiple pole on the negative real axis and, of course, an n_p-order zero at the origin. This type of pattern can be obtained by cascading a number of stages whose individual patterns have been reduced to a single pole plus zero as in Section 12.1.2.2.1. Alternatively, because poles and zeros are commutative, the singularities throughout the whole amplifier may be arranged to give the multiple-coincident-pole pattern; cancellation of poles and zeros occurs between the stages rather than within each stage. Consider Fig. 12.6; the response is 3 dB down at a frequency ω_{co} such that

$$|\psi(j\omega_{co})|^2 = \frac{(\text{product of distances from zeros})^2}{(\text{product of distances from poles})^2} = \frac{1}{2}; \qquad (12.29)$$

that is,

$$\left(\frac{\omega_{co}^2}{\omega_{co}^2 + \omega_1^2}\right)^{n_p} = \frac{1}{2}$$

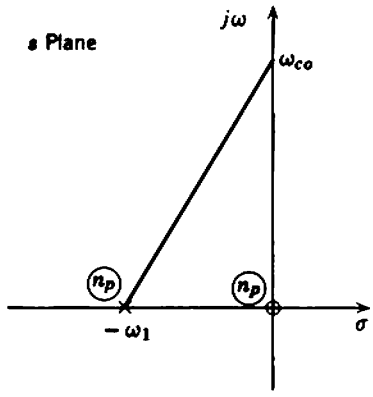


Fig. 12.6 Lower half-power point for multiple coincident poles.

which reduces to

$$\omega_{co} = \frac{\omega_1}{\sqrt{2^{1/n_p} - 1}}. \tag{12.30}$$

As the order n_p of the pole is increased, the cutoff frequency rises. Table 12.1 lists the normalized cutoff frequency $1/\sqrt{2^{1/n_p} - 1}$ for n_p up to 10. The results in this Table are used in the examples of Section 7.8 to find the positions of the multiple poles necessary for the specified cutoff frequencies.

Table 12.1 Normalized Cutoff Frequency for Multiple Low-Frequency Poles

Order of Pole n_p	$\frac{1}{\sqrt{2^{1/n_p} - 1}}$
1	1.000
2	1.552
3	1.961
4	2.299
5	2.590
6	2.857
7	3.096
8	3.317
9	3.535
10	3.732

12.1.2.2.3 *In-Band Zeros*

The use of compensating zeros to cancel some of the circuit poles is an easily calculated but inefficient method for compensating the low-frequency response of an amplifier. A better method is to position the zeros further from the origin than the poles, so that the zeros lie inside the passband. Over a limited frequency range the rising response of a few zeros can offset the falling response of a larger number of poles and lower the 3-dB cutoff frequency. The general case is not amenable to analysis, and must be treated graphically on a trial-and-error basis. There is, however, a simple design rule for the special case in which all poles are coincident at one point and all compensating zeros are coincident at another.

Consider the singularity pattern of Fig. 12.7, in which there is an n_p-order pole at $-\omega_1$ and an n_z-order zero at $-\omega_2$ and, of course, an $(n_p - n_z)$-order zero at the origin. From Eqs. 12.5 and 12.7, the magnitude of the ψ function at any frequency is

$$|\psi(j\omega)|^2 = \left(\frac{\omega^2}{\omega^2 + \omega_1^2}\right)^{n_p} \times \left(\frac{\omega^2 + \omega_2^2}{\omega^2}\right)^{n_z}. \tag{12.31}$$

This response can be made maximally flat (i.e., the only turning point with respect to ω^2 is at infinity) by choosing

$$\frac{\omega_2}{\omega_1} = \left(\frac{n_p}{n_z}\right)^{1/2}. \tag{12.32}$$

By definition the 3-dB cutoff frequency occurs when

$$|\psi(j\omega_{co})|^2 = \tfrac{1}{2}.$$

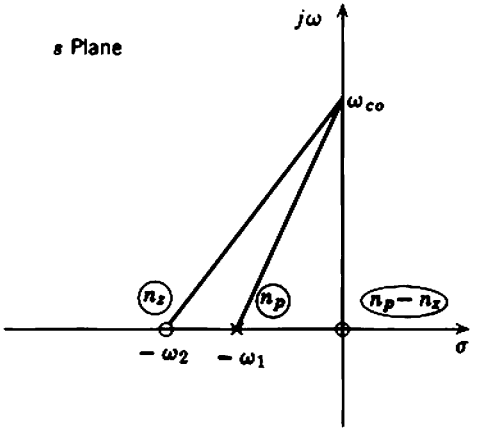


Fig. 12.7 Lower half-power point with in-band zeros.

Substituting Eq. 12.32 into Eq. 12.31 therefore gives the cutoff frequency as the solution of

$$\left(\frac{\omega_{co}^2}{\omega_{co}^2 + \omega_1^2}\right)^{n_p} \times \left[\frac{\omega_{co}^2 + \omega_1^2(n_p/n_z)}{\omega_{co}^2}\right]^{n_z} = \frac{1}{2}. \tag{12.33}$$

Table 12.2 lists some useful results. The 5th column of this table is the cutoff frequency that results from setting $\omega_2 = \omega_1$ so as to give partial cancellation of the multiple pole by the compensating zeros. Comparison with the 4th column shows that the bandwidth advantage of in-band zeros is rather more than a factor 1.5.

Table 12.2 Low-Frequency Compensation by In-Band Zeros (Normalized so that $\omega_1 = 1$)

Order of Pole n_p	Order of Compensating Zero n_z	Position of Zero ω_2/ω_1	Cutoff Frequency ω_{co}/ω_1	Cutoff Frequency if $\omega_2 = \omega_1$ (Cancellation)
2	1	1.414	0.643	1.000
3	1	1.732	1.000	1.552
4	1	2.000	1.260	1.961
3	2	1.224	0.605	1.000
4	2	1.414	0.921	1.552

12.1.2.2.4 *Dominant Singularities*

Sometimes it is not convenient to set all low-frequency poles coincident, as required in Sections 12.1.2.2.2 and 12.1.2.2.3. The necessary values of some capacitors become very large, whereas others are still quite small. A more convenient design results from increasing these small capacitors, thereby moving the associated poles toward the origin so that they have little effect on the passband. The remaining outer group of singularities is called the *dominant group* because it largely controls the response. For a specified cutoff frequency, the dominant poles can be moved further from the origin as their number is reduced. Reducing the number of dominant poles by increasing some capacitors therefore allows other capacitors to be reduced.

The use of a dominant singularity group has the further advantage of simplifying the design work. Because the dominant group is of low order, the labor in finding its response graphically from Figs. 12.1 and 12.2 is quite small. The extreme case is when the dominant group is a single dominant pole, but any other grouping such as a multiple pole or multiple

pole with in-band zeros may be used. The response due to nondominant singularities can be found analytically from Eqs. 12.13 and 12.14; any grouping may be used but a multiple coincident pole is the easiest to handle.

12.2 LOW-FREQUENCY RESPONSE OF FEEDBACK STAGES

The gain of feedback amplifier stages varies at low frequencies due to the presence of coupling, bypass, and decoupling capacitors. Each capacitor contributes one pole and one zero to the open-loop singularity pattern, and all singularities lie on the negative real axis (except that the zero due to a coupling capacitor lies at the origin). However, the poles migrate to their closed-loop positions when the feedback loop is closed, and these can, in general, be complex. Feedback stages should be designed so that all their dominant closed-loop poles lie on the negative real axis, as the same types of compensating singularity patterns are then applicable to both feedback and nonfeedback stages.

12.2.1 Coupling Capacitors

For the most part, coupling capacitors are external to the feedback loops around series- and shunt-feedback stages. These capacitors can therefore be treated in much the same way as the coupling capacitors between nonfeedback stages.

When a series-feedback stage is coupled to the input circuit of a shunt-feedback stage as in Fig. 12.8, the signal output current from the first device divides between the various biasing resistors and the input resistance

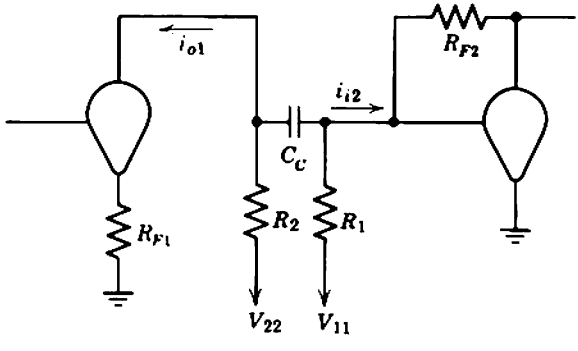


Fig. 12.8 Coupling capacitor between series- and shunt-feedback stages.

of the shunt stage proper. In a well-designed circuit the coupling efficiency is nearly 100% at mid-band frequencies, and most of the current flows into the shunt stage. At low frequencies, however, the reactance of the coupling capacitor becomes large, and the distribution of the current is changed; progressively more of the signal current output from the first device flows into its own supply resistor and less flows on into the shunt stage. The frequency-dependent factor in the gain is

$$\psi_k(s) = \frac{s}{s - p}, \tag{12.34}$$

where

$$p = -\frac{1}{[(R_o \parallel R_2) + (R_i \parallel R_1)]C_C}. \tag{12.35}$$

The condition for near-unity mid-band coupling efficiency is

$$R_i \ll (R_2 \parallel R_1)$$

and a good approximation for p is therefore

$$p \approx -\frac{1}{(R_o \parallel R_2)C_C}. \tag{12.36}$$

Further, the output resistance of a series-feedback stage is very large and a second approximation is

$$p \approx -\frac{1}{R_2 C_C}. \tag{12.37}$$

Equation 12.37 is by far the most useful expression for p. Notice that p is accurately reproducible because it (substantially) depends only on the passive elements R_2 and C_C. This pole can be found exactly from the complete analysis of a shunt-feedback stage (Section 12.2.3.3).

When a shunt-feedback stage is coupled to the input circuit of a series-feedback stage as in Fig. 12.9, the coupling capacitor forms a voltage divider with the input resistance of the series stage. At low frequencies the input voltage to the series stage falls away to zero as the reactance of C_C rises. The frequency-dependent factor in the gain is again $\psi_k(s)$ given by Eqs. 12.34 and 12.35, but a totally different set of approximations apply. The output resistance of a shunt-feedback stage is small and therefore

$$p \approx -\frac{1}{(R_i \parallel R_1)C_C}. \tag{12.38}$$

For a vacuum tube R_i is infinite and Eq. 12.38 reduces to

$$p \approx -\frac{1}{R_G C_C}, \tag{12.39}$$

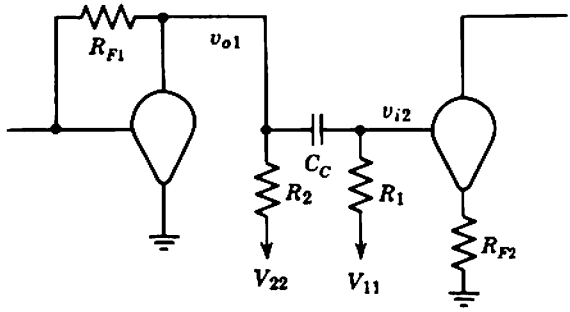


Fig. 12.9 Coupling capacitor between shunt- and series-feedback stages.

where R_G is the grid supply resistor. The position of the pole is accurately defined by the passive elements R_G and C_C. For a well-designed transistor circuit, R_i of a series stage at mid-band frequencies is made much smaller than the supply resistors in order that R_i should be the dominating component of the load for the preceding shunt stage. The input impedance at low frequencies, however, depends on any biasing network in the emitter circuit; the situation is similar to that of the nonfeedback stages discussed in Section 7.4.2.2, and gross cross-coupling occurs between the circuit poles. Section 12.2.2.1 considers the most useful special case in which the poles can be found easily. Another special case is when the biasing time constants in the emitter circuit are so long that R_i can be approximated by Eq. 11.26 over the entire useful frequency range. In this case, Eq. 12.38 reduces to

$$p \approx -\frac{1}{R_i C_C} = -\frac{1}{[r_B + \beta_N(r_E + R_F)]C_C} \approx -\frac{1}{\beta_N R_F C_C}. \quad (12.40)$$

The position of the pole depends on β_N, and is therefore not defined accurately. Problem 12.5 contains further information.

12.2.2 Bypass Capacitors in a Series-Feedback Stage

A biasing network in the emitting-electrode circuit of a series-feedback stage can be considered as part of the feedback impedance. The reactance of the bypass capacitor in Fig. 12.10 is small at medium and high frequencies, and the effective feedback impedance is the feedback resistor R_F. As the frequency falls, the reactance of C_E increases so that the magnitude of the feedback impedance increases and the transfer admittance of the stage falls. At very low frequencies, where the reactance of C_E

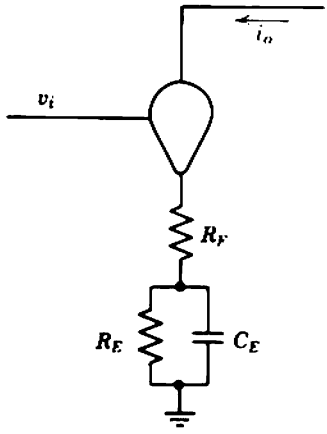


Fig. 12.10 Emitting-electrode bypass capacitor in a series-feedback stage.

is very large, the feedback resistance becomes $(R_F + R_E)$. The frequency-dependent factor in the transfer admittance is

$$\psi_k(s) = \frac{s - z}{s - p} \tag{12.41}$$

and the values of p and z are

$$p = -\frac{1 + G_T R_E}{R_E C_E}, \tag{12.42}$$

$$z = -\frac{1}{R_E C_E}, \tag{12.43}$$

where G_T is the mid-band transfer conductance of the stage. If the mid-band loop gain is large, G_T is known accurately and the position of the pole is defined accurately. In addition, the transfer conductance of a series stage with large loop gain is

$$G_T \approx \frac{1}{R_F},$$

so the value of p reduces to

$$p \approx -\frac{1}{(R_F \mid R_E)C_E}. \tag{12.44}$$

12.2.2.1 Special Considerations for Transistors

The input impedance of a transistor series-feedback stage depends on the feedback impedance. At low frequencies the input impedance of a stage with an emitter biasing network is

$$Z_i = r_B + \beta_N\left(r_E + R_F + \frac{R_E}{1 + sR_E C_E}\right) = R_i\left(\frac{s - p}{s - z}\right), \tag{12.45}$$

where R_i is the mid-band input resistance, and p and z are given by Eqs. 12.42 and 12.43. (Notice that p and z are interchanged between the numerators and denominators of Eqs. 12.41 and 12.45.) This frequency-dependent input impedance reacts on the RC coupling network that precedes the stage, and gross cross-coupling occurs between the poles. As in the case of nonfeedback stages (Section 7.4.2.2), the system has two poles and two zeros. One zero is due to the coupling capacitor and lies at the origin; the other is due to the emitter bypass capacitor and is given by Eq. 12.43. General expressions for the poles are very complicated, but if the circuit has been well designed so that

$$R_i \gg R_B$$

then one pole cancels the emitter bypass zero and the other is at

$$p \approx -\frac{\beta_N C_C + C_E}{[r_B + \beta_N(r_E + R_F)]C_C C_E}. \tag{12.46}$$

A useful practical simplification results from choosing

$$\beta_N C_C \gg C_E,$$

when Eq. 12.46 becomes

$$p \approx -\frac{1}{(r_B/\beta_N + r_E + R_F)C_E} = \frac{G_T}{C_E}. \tag{12.47}$$

Perhaps the best approach of all is to direct couple a transistor series-feedback stage to the preceding shunt stage (Fig. 11.22 of Section 11.6.1, for example) and thereby eliminate the problem of cross-coupling.

12.2.2.2 Screen Bypass Capacitor for a Pentode

Section 7.4.1.2 describes the effect of an RC screen supply network at low frequencies, and derives expressions for the pole and zero when these are the only low-frequency singularities. Any impedance in the screen circuit applies a form of series feedback to a pentode. The screen current (which is proportional to the cathode current) develops a signal voltage across the screen impedance and, after division by the screen amplification factor μ_S, this signal voltage effectively subtracts from the input applied to the grid. Thus an impedance in the screen circuit can be moved into the cathode circuit after division by the scale factor $[\mu_S(1 + k)/k]$. The effective feedback impedance in Fig. 12.11 becomes

$$Z_F(s) = R_F + \frac{R_K}{1 + sR_K C_K} + \frac{k}{\mu_S(1 + k)}\left(\frac{R_S}{1 + sR_S C_S}\right) \tag{12.48}$$

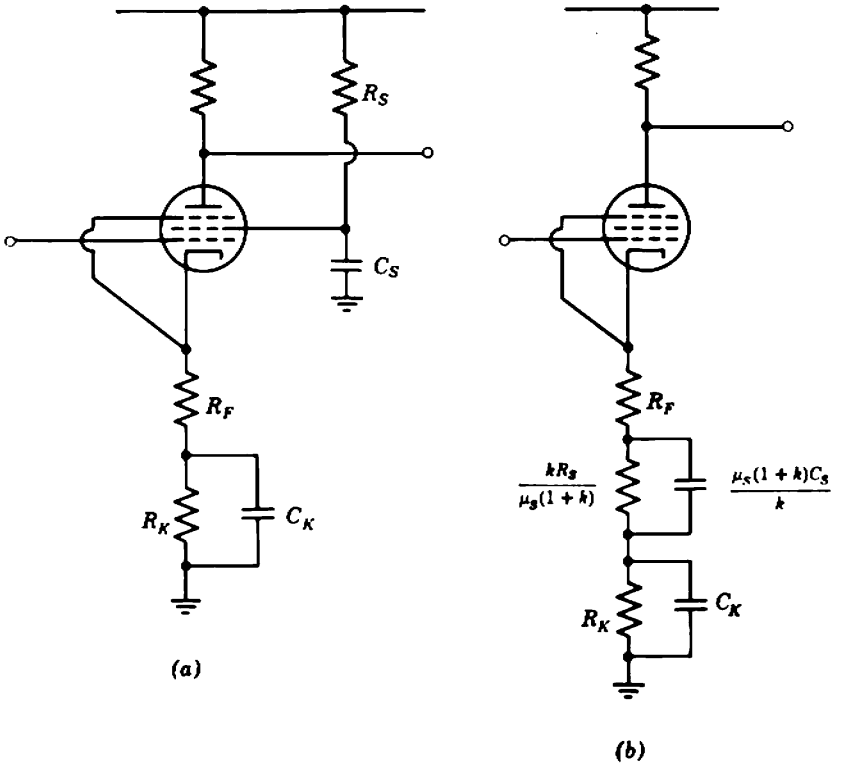


Fig. 12.11 Screen bypass capacitor in a series-feedback stage: (*a*) circuit diagram; (*b*) equivalent circuit arrangement.

and the transfer admittance of the stage is

$$Y_T(s) = \frac{g_m}{1 + g_m(1 + k)Z_F(s)}. \tag{12.49}$$

A useful simplification results from designing the circuit with

$$R_K C_K = R_S C_S.$$

The mid-band transfer conductance is G_T given by Eq. 11.15, and the ψ factor is given by Eq. 12.41 with

$$p = -\frac{1 + G_T[(1 + k)R_K + kR_S/\mu_S]}{R_K C_K}, \tag{12.50}$$

$$z = -\frac{1}{R_K C_K} = -\frac{1}{R_S C_S}. \tag{12.51}$$

12.2.3 The Shunt-Feedback Stage

The analysis of a shunt-feedback stage at low frequencies is quite complicated, because there can be more than two poles and because these poles can move off the negative real axis. A root-locus approach is the most generally satisfactory. In the following sections the effects of bypass capacitors and a feedback blocking capacitor are first considered separately; these simple cases give insight into the operation of the circuit. Finally, a general expression is derived for the open-loop singularity pattern, and this can be used in a complete root-locus analysis.

12.2.3.1 Bypass Capacitors

Connecting an impedance in the emitting-electrode circuit of a device applies series feedback to it and reduces its transfer admittance and voltage gain. There is no change, however, in its current gain and transfer resistance—compare Eqs. 5.77 and 11.24, for example. Now, the transfer resistance of a shunt-feedback stage is

$$R_T = -\frac{R_F}{1 - 1/A_l} \approx -\frac{R_F}{1 - R_F/R_{T0}},$$

where R_{T0} is the transfer resistance without feedback. Because R_{T0} is independent of any impedance in the emitting-electrode circuit, the transfer resistance of a shunt-feedback stage is, to a first order, independent of any emitting-electrode biasing network. The emitting-electrode bypass capacitor shown in Fig. 12.12 introduces no poles or zeros into the transfer resistance of a shunt-feedback stage.

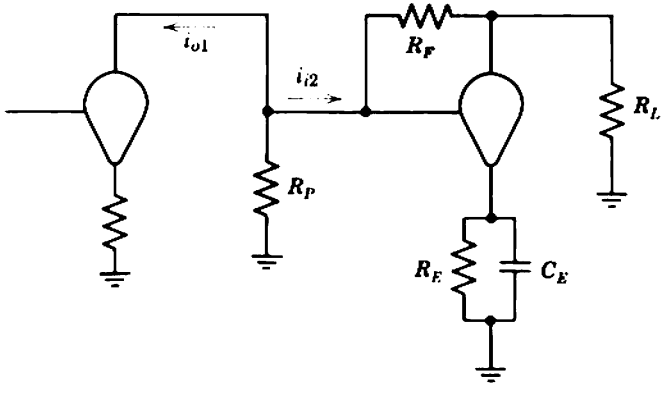


Fig. 12.12 Emitting-electrode bypass capacitor for a shunt-feedback stage.

Equation 11.60 gives the input impedance of a shunt-feedback stage as

$$Z_i(s) \approx -\frac{R_F}{A_V(s)}$$

where, for greater generality, the voltage gain is shown as a function of complex frequency. The voltage gain of a shunt-feedback stage is

$$A_V(s) = Y_T(s)\, R_{L(\text{eff})} = Y_T(s) \left(\frac{R_F \times R_L}{R_F + R_L}\right),$$

where $Y_T(s)$ is the transfer admittance of the device (including the effects of any bypass capacitors in its biasing system). Finally, therefore,

$$Z_i(s) = \frac{R_F + R_L}{Y_T(s)\, R_L} \tag{12.52}$$

showing that the input impedance is a function of the bypass capacitors. At low frequencies $|Y_T|$ falls and $|Z_i|$ rises, and therefore the coupling efficiency into the stage falls. Thus the output voltage from a shunt-feedback stage falls at low frequencies even though its transfer impedance remains constant. From Eq. 11.58

$$\eta(s) = \frac{R_P}{R_P + Z_i(s)}$$

and, substituting from Eq. 12.52,

$$\eta(s) = \frac{1}{1 + \left[\dfrac{R_F + R_L}{Y_T(s)\, R_P R_L}\right]}. \tag{12.53}$$

Equation 12.53 is of the same form as the basic feedback equation (Eq. 10.49), with the following equivalences:

$$A_{lm} \rightarrow -\frac{G_T R_P R_L}{R_F + R_L},$$

$$\psi_\mu(s)\, \psi_\beta(s) \rightarrow \psi \text{ function of } Y_T(s).$$

The singularities of $Y_T(s)$ can be written down from Section 7.4.1, and the singularities of $\eta(s)$ can be then found by a root-locus type of analysis. All poles of $\eta(s)$ remain real if $Y_T(s)$ has only a single pole, or if the poles and zeros of $Y_T(s)$ are interlaced on the negative real axis. The zeros of $\eta(s)$ occur at the zeros of $Y_T(s)$.

12.2.3.2 Feedback Blocking Capacitor

The signal feedback resistor required for a shunt-feedback stage is often too small for convenience in the biasing circuit, and Section 11.6.1 suggests the circuit arrangement shown in Fig. 12.13. At frequencies where the

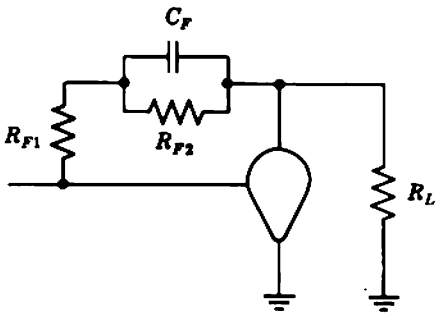


Fig. 12.13 Blocking capacitor in a shunt-feedback network.

reactance of C_F becomes appreciable the stage can be considered as having a frequency-dependent feedback impedance:

$$Z_F(s) = R_{F1}\left(\frac{s - z}{s - p}\right), \tag{12.54}$$

where

$$p = -\frac{1}{R_{F2}C_F}, \tag{12.55}$$

$$z = -\frac{R_{F1} + R_{F2}}{R_{F1}R_{F2}C_F}. \tag{12.56}$$

In the general case the transfer impedance of the stage can be found if R_F in the mid-band equations of Chapter 11 is replaced by the frequency-dependent Z_F. It should be pointed out that, if R_{F2} is much larger than R_{F1}, the input impedance can change so far at low frequencies that the coupling efficiency into the stage is affected; the total singularity pattern depends somewhat on the source resistance.

Section 10.5.1 shows that the closed-loop transfer function of an amplifier has zeros at the poles of its feedback factor. In addition, Section 10.5.3.2 shows that closed-loop poles appear in the vicinity of the zeros of the feedback factor. Because the feedback factor for a shunt stage is $1/Z_F(s)$, the transfer impedance has a zero at

$$s = -\frac{R_{F1} + R_{F2}}{R_{F1}R_{F2}C_F}$$

and a pole near

$$s = -\frac{1}{R_{F2}C_F}.$$

The significance of these results is that, because the zero lies further from the origin than the pole, the transfer impedance rises with decreasing

frequency. Shunt stages with a feedback blocking capacitor are thus potentially useful for compensating the low-frequency response of a multistage amplifier. Blocking capacitors are often introduced into shunt-feedback networks for just this purpose, even though not essential for biasing.

12.2.3.3 The General Case

The ratio of the output voltage from a shunt-feedback stage to the output current from the preceding series-feedback stage depends on the coupling efficiency between the stages and on the transfer impedance of the shunt stage. From the discussion in the two preceding sections it is apparent that both factors can contribute to the total singularity pattern. One method for finding the total singularity pattern is to perform a root-locus analysis on the system loop gain introduced in Section 11.3.2 (Fig. 11.17*b*). This system loop gain in Fig. 11.17*b* depends on

(i) the transfer admittance $Y_T(s)$ of the device (including the effects of any bypass capacitors),
(ii) the load impedance $Z_L(s)$ (largely Z_i of the following stage),
(iii) the feedback impedance $Z_F(s)$,
(iv) the source impedance $Z_S(s)$ (i.e., the biasing supply resistors R_1 and R_2 and the coupling capacitor C_C).

Items (i) and (iii) ordinarily contribute a leading phase shift to the system loop gain at low frequencies, and therefore tend to provoke oscillation (or at least the generation of complex closed-loop poles). Items (ii) and (iv) contribute a lagging phase shift with the opposite tendencies, and can be exploited by the circuit designer to achieve a desired dynamic response.

The source impedance in Fig. 11.17*b* is in shunt with the signal path but, as shown in Fig. 12.14*a*, the coupling capacitor at the input of a shunt-feedback stage is in series. It is therefore necessary to make the transformation to Fig. 12.14*b* before the results of Section 11.3.2 can be applied. The source impedance is

$$Z_S = R_1\left[\frac{1 + sR_2C_C}{1 + s(R_1 + R_2)C_C}\right], \tag{12.57}$$

and the equivalent source current becomes

$$\frac{i_{S(eq)}}{i_{o(series)}} = -\frac{sR_2C_C}{1 + sR_2C_C}. \tag{12.58}$$

Therefore

$$\frac{v_{o(shunt)}}{i_{o(series)}} = -\left(\frac{sR_2C_C}{1 + sR_2C_C}\right)\left[\frac{v_{o(shunt)}}{i_{S(eq)}}\right]$$

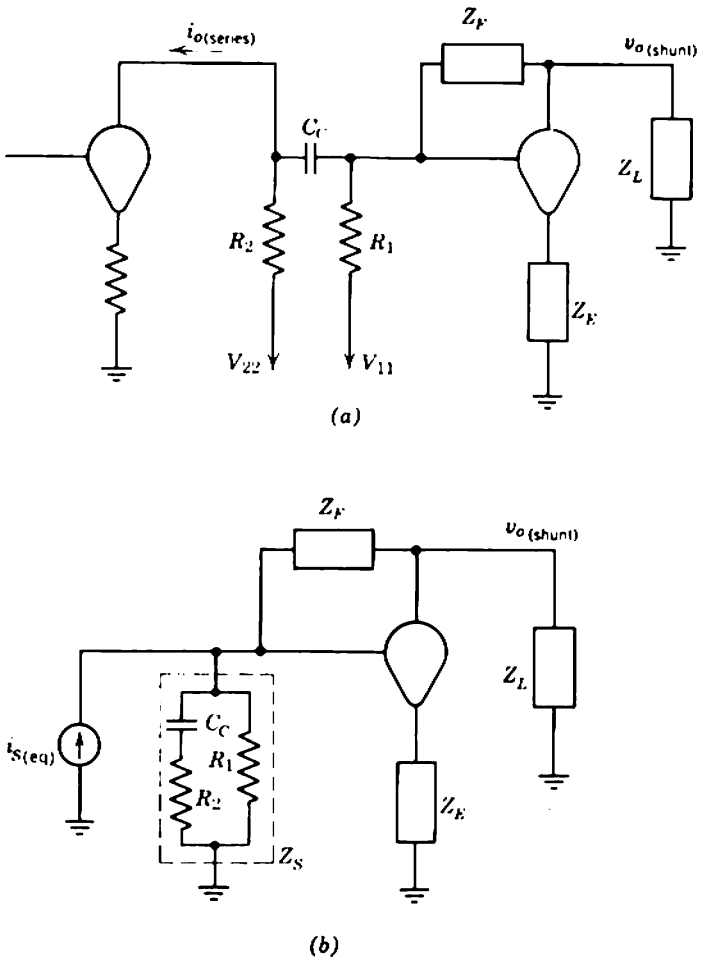


Fig. 12.14 The general shunt-feedback stage: (*a*) circuit diagram; (*b*) equivalent circuit arrangement.

and from either Eq. 11.78 or 11.84 with an obvious change in notation

$$\frac{v_{o(\text{shunt})}}{i_{o(\text{series})}} = \left(\frac{sR_2C_C}{1 + sR_2C_C}\right)\left[\frac{Z_F(s)}{1 - 1/A_i'(s)}\right], \tag{12.59}$$

where $A_i'(s)$ is the system loop gain.

For a vacuum tube, Eq. 11.80 shows that

$$A_i' = Y_T\left(\frac{Z_F \times Z_L}{Z_F + Z_L}\right)\left(\frac{Z_S}{Z_F + Z_S}\right), \tag{12.60}$$

where $Y_T(s)$ is the transfer admittance of the tube (including all bypass

capacitors) which can be found from Section 7.4.1. For a transistor, Eq. 11.81 shows that

$$A_i' = \beta_N\left(\frac{Z_L}{Z_F + Z_L}\right)\left\{\frac{Z_F Z_S}{[r_B + \beta_N(r_E + Z_E)](Z_F + Z_S) + Z_F Z_S}\right\}, \quad (12.61)$$

where $Z_E(s)$ is the biasing impedance in the emitter circuit.

The first term in Eq. 12.59 has a zero at the origin and a pole at $s = -1/R_2C_C$; invariably, however, this pole is canceled by a zero from the second term. This second term represents a transfer impedance with feedback and, from the discussion in Section 10.5.1, it has

zeros at the poles of the feedback factor $\beta(s)$,
zeros at the zeros of the reference variable $\mu(s)$,
poles at the solutions of $A_i'(s) = 1$.

The feedback factor of a shunt-feedback stage is

$$\beta(s) \to \frac{1}{Z_F(s)}.$$

Hence there are poles of $\beta(s)$, and thus closed-loop zeros, at the zeros of $Z_F(s)$. The reference variable is

$$\mu(s) \to \frac{A_i'(s)}{\beta(s)}.$$

Hence there are zeros of $\mu(s)$, and thus closed-loop zeros, at the zeros of $A_i'(s)$ other than the zeros of $\beta(s)$; such zeros of $A_i'(s)$ can be found from Eqs. 12.60 and 12.61. One zero of $A_i'(s)$ for both vacuum tubes and transistors is the zero of $Z_S(s)$ and, as stated, this cancels the pole in the first term of Eq. 12.59. Over-all, the singularity pattern of Eq. 12.59 consists of

(i) a zero at the origin,
(ii) zeros at the zeros of $Z_L(s)$,
(iii) zeros at the zeros of $Y_T(s)$ for a vacuum tube, or the poles of $Z_E(s)$ for a transistor,
(iv) zeros at the zeros of $Z_F(s)$,
(v) poles found from the solution of $A_i'(s) = 1$.

The root-locus solution can be simplified by designing the circuit so that some poles and zeros of $A_i'(s)$ cancel each other, or so that some of the impedances are pure resistance, or so that some impedances are much

smaller than others and can be neglected. The closed-loop singularity pattern can usually be reduced to the following:

(i) A zero at the zero of $Z_F(s)$ plus a pole near the pole of $Z_F(s)$. These are attributable to the feedback blocking capacitor, and are given by Eqs. 12.56 and 12.55, respectively.

(ii) A zero at the origin plus a pole near the zero of $Z_S(s)$. These are attributable to the input coupling capacitor; the pole is given by Eq. 12.37.

(iii) Several more pole-zero pairs which lie very close to each other. These cancel to a good approximation and can be neglected.

All poles can be kept real by interlacing the open-loop poles and zeros on the negative real axis.

The mid-band ratio of the output voltage from a shunt stage to the output current from the preceding series stage can, of course, be calculated from the low-frequency system-loop-gain analysis without calculating the coupling efficiency and stage transfer resistance separately. However, calculating the coupling and bypass capacitors to give a specified low-frequency response is almost the last step in designing a practical amplifier. The mid-band transfer function is required at a much earlier stage in the design process, and is most easily calculated from η and R_T as in Section 11.2.2.2. System-loop-gain analysis is a powerful method for finding the low- and high-frequency response, but tends to be cumbersome when mid-band frequencies alone are of interest.

12.2.4 Decoupling Capacitors

The biasing supply impedances in a nonfeedback amplifier are an essential (often dominant) part of the load impedance for any active device. The *RC* decoupling networks constitute a part of this biasing impedance, and cause the gain to rise at low frequencies. In contrast, biasing impedances in a cascade of series- and shunt-feedback stages in alternation are merely parasitic. The biasing impedances are made as large as practicable and their values do not appear in first-order expressions for gain. Decoupling networks therefore have substantially no effect on the low-frequency response of the alternate cascade.

For similar reasons only a minimal amount of decoupling is required in an alternate cascade; a variation of the dc supply voltage of any stage produces very little signal output from the stage. Therefore, mains ripple voltages on the supply rails do not introduce significant hum unless the signal to be amplified is very small. Further, feedback of signals from the output stage of an amplifier into the input stage via the supply rails is unlikely to cause regeneration unless the total gain is very large.

12.3 CIRCUIT TECHNIQUES FOR IMPROVING LOW-FREQUENCY RESPONSE

Two aspects of low-frequency design are considered in preceding sections of this book: Sections 7.4 and 12.2 derive the poles and zeros of an amplifier as functions of the capacitors and other parameters in the circuit, and Section 12.1 discusses the relation between the dynamic response of an amplifier and its singularity pattern. There remains the engineering problem of obtaining a satisfactory dynamic response with reasonable capacitor sizes. This is essentially a problem in circuit ingenuity. In principle, no limit exists to the low-frequency dynamic response of an amplifier because all coupling, bypass, and decoupling capacitors can be eliminated together with the associated low-frequency singularities. The biasing stability will of course be inadequate if simple circuits are used but special circuit arrangements for dc amplifiers are discussed in Chapter 15. Because no theoretical limits exist, there is no general method for optimizing the design; each problem must be treated on its own merits. The following sections outline some guiding principles, but there is really no substitute for experience. Essentially, the steps in compensating an amplifier at low frequencies are as follows:

1. Eliminate as many bypass capacitors as possible by suitable choice of biasing circuits. This reduces the total number of low-frequency singularities, so that the remaining singularities can be placed further from the origin for a given dynamic response, and smaller capacitors are permissible.
2. For the same reason eliminate as many coupling capacitors as possible by direct coupling some stages.
3. Arrange for the dominant poles that remain to be those that are least dependent on the active device parameters and are therefore most accurately known. An accurately known response is much easier to compensate than an inaccurately known response.
4. Introduce the zeros required to compensate the response from decoupling networks (in nonfeedback stages) or shunt-feedback networks.
5. If the frequency response is to be compensated (as distinct from the time response), it is a convenience if the final singularity pattern falls into one of the simple classes of Section 12.1.2.2—multiple coincident singularities, dominant pole or group, and so on. Pole-zero cancellation can be used to simplify the total pattern.

12.3.1 Vacuum Tubes

Standard practice for compensating vacuum-tube amplifiers is thoroughly unimaginative. Apart from the occasional use of unbypassed

cathode resistors (discussed below), the overwhelming majority of vacuum-tube amplifiers consist of cascaded *RC*-coupled pentodes with the self-bias and screen voltages obtained from *RC* networks. There is room for a great deal of constructive thought and originality by circuit designers.

12.3.1.1 Nonfeedback Stages

For a specified pole position, by far the largest capacitor in a nonfeedback stage is the cathode bypass, whereas the smallest is the interstage coupling. However, no one capacitor persistently dominates the low-frequency response, because different capacitor types are suitable for the different applications. Physically small electrolytic capacitors rated at a few tens of volts are available up to several hundred microfarads, and are suitable for cathode bypasses where the working voltages are small and leakage currents of a few microamperes are tolerable. In contrast, coupling capacitors in vacuum-tube amplifiers must be low-leakage types, because leakage currents of only a few microamperes into a high-impedance grid circuit can disturb the quiescent voltages appreciably. Paper or plastic dielectric types are required, and are available up to about 1 μF with ratings up to about 600 V. Paper or plastic capacitors are also used for screen bypasses up to about 1 μF; electrolytics are used when larger capacitances are required. It is almost always possible to make the screen bypass capacitor so large that its pole is nondominant.

Cathode bypass capacitors are often omitted from self bias circuits (Section 6.1.2.2) to eliminate one low-frequency pole and zero. The cathode resistor applies a small amount of series feedback to the stage and reduces the transfer conductance, but the resultant loss of voltage gain can usually be made up by increasing the load resistance. A possible drawback for low-level stages is that hum may be introduced; induced currents from a mains-powered heater can develop a hum voltage across the resistance in the cathode circuit.

Screen bypass capacitors can be eliminated by feeding the screen directly from a dc supply rail. Because the supply rails in typical vacuum-tube amplifiers are at several hundred volts, there is an advantage in tubes that are designed to operate with high screen voltages—200 to 300 V, rather than 100 to 150 V as is more usual. If a suitable supply rail is available, it is prudent to connect a resistor of about 100 Ω between the screen and the supply. This *stopping resistor* prevents parasitic oscillations at very high frequencies, of the order of 100 MHz. If a voltage dropping resistor must be used, it is not good practice to omit the screen bypass capacitor in the same way that a cathode bypass is often omitted; the voltage gain from the control grid to the screen typically lies between 10 and 100, and the tube develops a large Miller input capacitance like a triode.

If both positive and negative supply voltages are available, the cathode bypass capacitor can be eliminated by using some form of anode or screen-feedback biasing; Section 6.3.3 describes typical circuit arrangements. Screen-feedback biasing systems are particularly suitable for nonfeedback stages as there is no signal feedback via the biasing network because no signal voltage is developed at the screen. Alternatively, if the voltage gain per stage is low (that is, if the load resistances are small), anode-feedback biasing can be used. Essentially, the biasing network applies shunt feedback to the stage. However, the source resistance is the small biasing supply resistor for the preceding tube, and this effectively short-circuits the feedback to ground; for the feedback to be effective, a shunt stage should be fed from a high-impedance source. Anode- and screen-feedback biasing are very much neglected in standard design practice.

Direct coupling of vacuum tubes is entirely feasible, and eliminates the pole and zero due to the coupling capacitor; like anode- and screen-feedback biasing, direct coupling is a very much neglected circuit arrangement. As a general rule, no more than two vacuum-tube stages should be direct coupled together, as very high supply voltages or high-voltage cathode bypass capacitors may be required. The two-stage feedback biasing systems of Section 6.5.3 are examples of satisfactory circuits.

If direct coupling is undesirable for any reason, the coupling pole for a given capacitor size can be moved in closer to the origin by increasing the grid supply resistor R_G. As discussed in Section 6.6.4, manufacturers quote a maximum safe value for R_G when the tube is operated in a specified circuit (usually self biasing). This value of R_G can be exceeded if the tube is operated below maximum dissipation or in a more stable biasing system. More than a factor of 10 increase in R_G is permissible in some cathode-feedback bias circuits (Section 6.1.2) for which the cathode resistor is large. An objection to this arrangement is that a cathode bypass capacitor is almost essential, or the fall in voltage gain becomes too great to be made up by increasing R_L. In addition, the dc voltage across the cathode biasing network may exceed the rating of readily available large-valued electrolytic capacitors. Anode- or screen-feedback biasing systems overcome these objections and are good alternatives.

12.3.1.2 Feedback Stages

The preferred methods for biasing feedback stages are discussed in Section 11.6.1. Anode-feedback biasing is recommended for isolated shunt stages, and a form of cathode-feedback biasing is recommended for isolated series stages; no cathode bypass capacitor is required for either circuit.

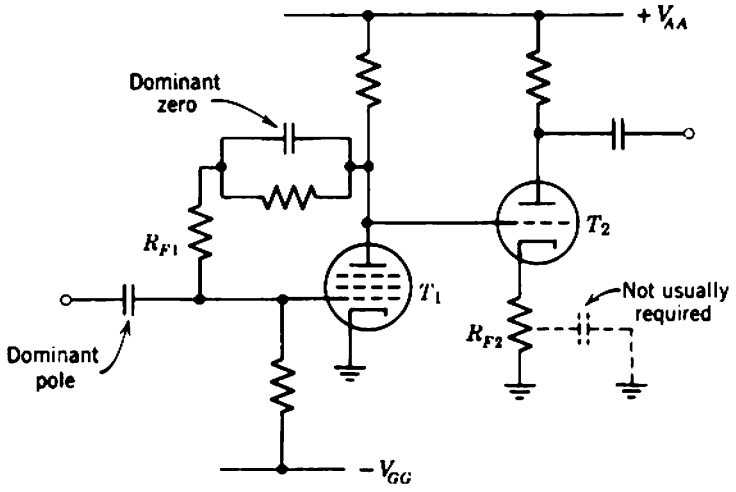


Fig. 12.15 A particularly good arrangement of direct-coupled vacuum-tube feedback stages.

A particularly neat arrangement results from direct coupling a series stage to the preceding shunt stage (Fig. 12.15). Notice that, if a cathode bypass capacitor is required for the series stage, its value is much smaller than for a stage without feedback—compare the pole positions given by Eqs. 7.11 and 12.44. The dominant pole is usually due to the coupling capacitor between groups, although the cathode bypass of the series stage (if it exists) may be significant also. Compensation is achieved via the feedback blocking capacitor of the shunt stage.

Many of the same tricks apply to feedback stages as to nonfeedback stages. Screens can be fed directly from a supply rail to eliminate bypass capacitors; stopping resistors should be included. If no negative supply is available, self bias may be used with shunt-feedback stages; the cathode bypass capacitor may be omitted if the fall in Y_T and consequent rise in input resistance (Section 12.2.3.1) can be tolerated. The reader is referred to Chapter 6 for other biasing circuits which may prove useful in special circumstances.

12.3.2 Transistors

Transistor amplifiers provide more scope for circuit ingenuity than vacuum-tube amplifiers. The lower operating voltages and the existence of both *p-n-p* and *n-p-n* types ease the problems of direct coupling. The lower voltages also ease the problems of obtaining suitable large-valued capacitors. The absence of any electrode corresponding to the screen grid of a pentode reduces the number of poles per stage. The lower

circuit impedances reduce the effects of leakage currents in coupling capacitors so that electrolytics are permissible.

12.3.2.1 Nonfeedback Stages

The dominant pole of a nonfeedback transistor stage is due either to its input coupling capacitor or emitter bypass capacitor, or to a combination of the two. For a given capacitor size, the ratio between the positions of the poles is β_N (Eq. 7.38 as either C_C or C_E becomes infinite) and, since electrolytics can be used for either capacitor, the input-circuit pole can always be made nondominant. The remaining pole is then given by Eq. 7.39. Circuit tricks in nonfeedback transistor amplifiers aim largely at eliminating some of the dominant poles, that is, eliminating some of the emitter bypass capacitors.

It is not satisfactory to follow vacuum-tube practice and simply omit the emitter bypass capacitor from a transistor stage. The value of emitter resistance required for biasing stability is much larger relative to the other circuit impedances than the cathode resistor in a self-biased vacuum-tube stage. Therefore the local series feedback that results from omitting the capacitor is relatively heavy, and the loss of voltage gain is too great to be made up by increasing the load resistance. Effectively, an amplifier with unbypassed emitter resistors is a cascade of series-feedback stages, and Section 11.2.3 shows that this is a poor circuit arrangement. If the pole due to an emitter bypass capacitor must be eliminated, some form of collecting-electrode-feedback biasing should be used. Collector-feedback biasing (with no precautions to reduce the signal feedback via the biasing resistors) is satisfactory if the supply resistors are small so that the gain is stable; the signal feedback is effectively short-circuited to ground as in the vacuum-tube counterpart above. If only one supply voltage is available, the Zener diode circuit of Fig. 6.48 is a good alternative.

Direct coupling of three, four, or even more stages is possible, particularly if *p-n-p* and *n-p-n* transistors are used alternately. Usually, however, the close tolerance required on the biasing resistors (Section 6.5.1) more than offsets the coupling capacitors saved, particularly as their poles can be made nondominant. As with vacuum tubes, a sound working rule is to restrict direct coupling to two stages. Section 6.5.2 suggests basic two-stage circuits, and refined circuits using Zener diodes or *p-n-p* and *n-p-n* transistors are described briefly in Sections 6.6.2 and 6.6.3.

12.3.2.2 Feedback Stages

Section 11.6.1 discusses the preferred methods for biasing feedback stages. Emitter-feedback biasing in any of its forms is suitable for series-feedback stages; an emitter bypass capacitor is required, and usually

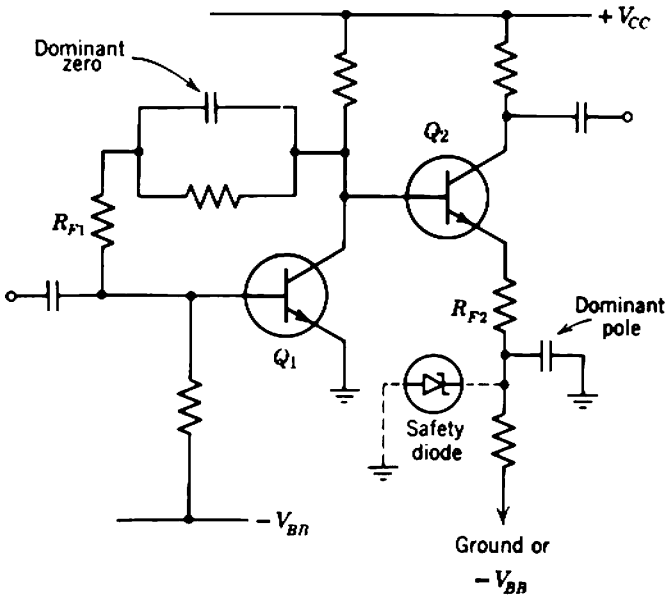


Fig. 12.16 A particularly good arrangement of direct-coupled transistor feedback stages. The function of the safety diode is explained in Section 6.6.2.

contributes the dominant pole to the stage. If this pole cannot be tolerated, collector-feedback biasing may be used; signal feedback via the biasing resistor is negligible because the voltage gain of a series-feedback stage is small. Collector-feedback biasing is recommended for shunt stages if both positive and negative supplies are available; the modified version of Fig. 11.21*c* permits low-frequency compensation. If only one supply is available, combined-feedback biasing or its modification with a Zener diode in the emitter circuit are suggested.

Direct coupling is very commonly employed between transistor feedback stages. Perhaps the most satisfactory arrangement of all is that shown in Fig. 12.16, or the equivalent version using *p-n-p* and *n-p-n* transistors in combination. This circuit is almost a standard building-block for multi-stage transistor amplifiers. It has a stable current gain, with a dominant pole due to the emitter bypass of the series stage, a dominant zero due to the shunt-feedback network (used for compensation), and two each of nondominant poles and zeros.

12.4 TWO WORKED DESIGN EXAMPLES

To illustrate the principles of low-frequency compensation, detailed designs are given for two wide-band amplifiers. The present discussion is restricted to the design of a biasing system to go with a given elemental

circuit, and its low-frequency response; the gain of the circuit and its high-frequency response are discussed in the examples of Section 13.8. This separation of the low-frequency response from the mid-band and high-frequency performance is typical of design problems encountered in practice. The mid-band gain and high-frequency response of an amplifier are fundamentally interrelated through the gain-bandwidth product, and these aspects of the design must be considered together. Once a satisfactory skeleton circuit is obtained, a biasing system can be fitted to it. Chapters 12 and 13 of this book are thus in the reverse order to the steps in practical design. This reversal is intentional; low-frequency compensation is the simpler topic and forms a useful introduction to the more complicated techniques used at high frequencies.

12.4.1 Vacuum-Tube Amplifier

The skeleton circuit diagram for the vacuum-tube amplifier is shown in Fig. 12.17*a*. The nominal quiescent voltages and currents are:

FIRST STAGE

$$I_K = 14 \text{ mA},$$
$$V_{SK} = 170 \text{ V}.$$

Therefore

$$g_m = 15.6 \text{ mA/V},$$
$$I_A = 10 \text{ mA},$$
$$I_S = 4 \text{ mA}.$$

SECOND STAGE

$$I_K = 45 \text{ mA},$$
$$V_{SK} = 300 \text{ V}.$$

Therefore

$$g_m = 10.2 \text{ mA/V},$$
$$I_A = 40 \text{ mA},$$
$$I_S = 5 \text{ mA}.$$

It is left as an exercise for the reader to verify that the biasing system shown in Fig. 12.17*b* does in fact achieve these conditions. The low-frequency response is required to be maximally flat, with 5 Hz cutoff frequency.

The requirements of the first stage are eased by using screen-feedback biasing. This arrangement has two advantages. First, it eliminates the potentially troublesome pole due to the cathode biasing network; a bypass capacitor in excess of 1000 μF is required to make this pole nondominant. In addition, screen-feedback biasing is (for this particular tube) about ten times more stable than self bias; the manufacturers' quoted maximum permissible grid resistor for self bias can therefore be exceeded, allowing

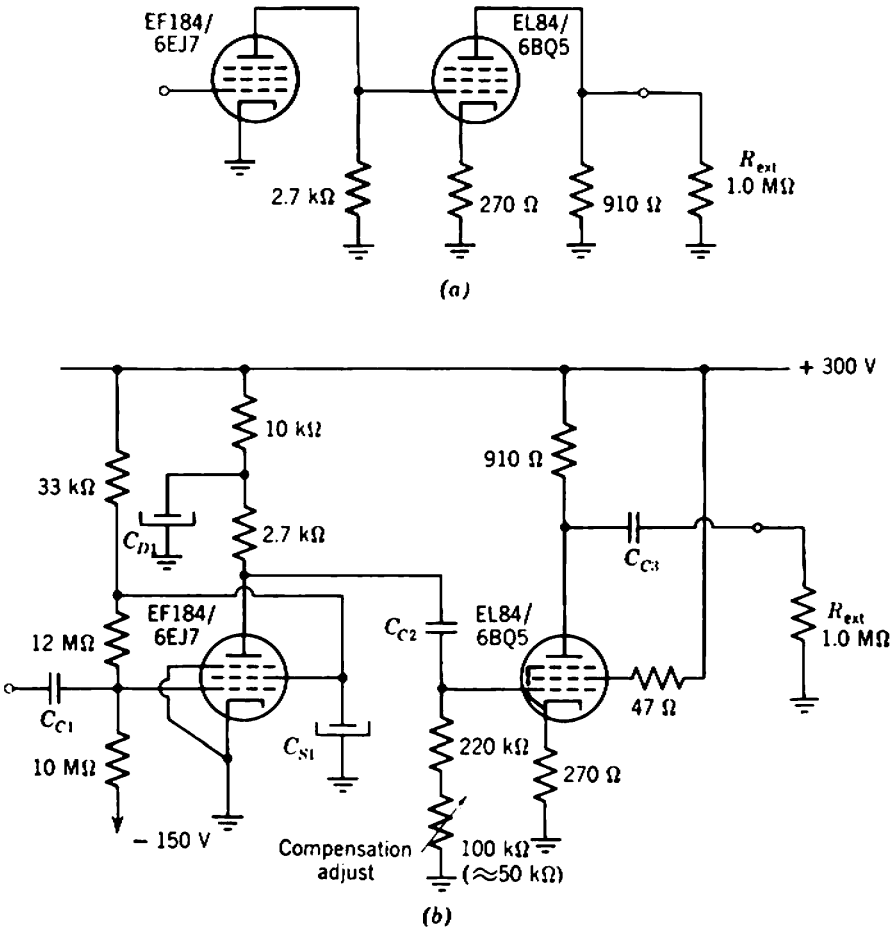


Fig. 12.17 Example of low-frequency compensation: (*a*) elemental mid-band circuit; (*b*) complete amplifier (less high-frequency peaking).

the use of a smaller input coupling capacitor. The poles due to both the input capacitor and the screen bypass capacitor can be made nondominant with reasonable capacitor sizes. The input stage therefore has no dominant poles at all. It has, however, one dominant zero from the decoupling network and this can be used for compensation.

The second stage is self biased with the cathode bypass capacitor omitted to eliminate one pole and one zero. There is a further advantage that the local series feedback reduces nonlinearity in the output stage. The screen is fed directly from the dc supply rail to eliminate another pair of singularities. Neither C_{C2} nor C_{C3} can be made large enough to contribute only nondominant poles without exceeding reasonable values for low-leakage capacitors. The stage therefore has two dominant poles.

In all, the amplifier has the following low-frequency singularities:

2 dominant poles (C_{C2} and C_{C3})
1 dominant zero (C_{D1})
3 nondominant poles (C_{C1}, C_{S1}, and C_{D1})
1 nondominant zero (C_{S1})
3 zeros at the origin (C_{C1}, C_{C2}, and C_{C3})

The chosen singularity pattern is the 2-pole 1-zero maximally-flat arrangement listed in Table 12.2; the nondominant singularities are ignored. For 5 Hz cutoff frequency, the dominant singularities are found from Table 12.2 as:

POLES

$$f = 7.8 \text{ Hz};$$

therefore,

$$\omega_1 = 49 \text{ radian/sec.}$$

ZERO

$$f = 11.0 \text{ Hz};$$

therefore,

$$\omega_2 = 69 \text{ radian/sec.}$$

The nondominant poles due to C_{C1} and C_{S1} are placed a decade outside the passband at 0.5 Hz, that is, $\omega = 3$ radian/sec. The capacitors (corrected to the nearest preferred value) are:

C_{C1} (Eq. 7.33)	68	nF
C_{S1} (Eq. 7.18)	33	μF
C_{D1} (Eq. 7.44)	6.8	μF
C_{C2} (Eq. 7.33)	68	nF
C_{C3} (Eq. 7.33)	22	nF

Some control should be provided for adjusting the compensation and taking up component tolerances. Because variable capacitors are not available of the order of magnitude required for low-frequency compensating circuits, a variable resistor must be used. This resistor should vary a dominant time constant without disturbing either the mid-band gain or the quiescent point of any device. The most satisfactory arrangement for vacuum-tube amplifiers is a potentiometer in series with one of the grid resistors as shown in the second stage of Fig. 12.17*b*.

12.4.2 Transistor Amplifier

The transistor amplifier is shown in outline in Fig. 12.18*a*. Q_1 to Q_3 are required to operate at 5 mA emitter current, with about 5 V between

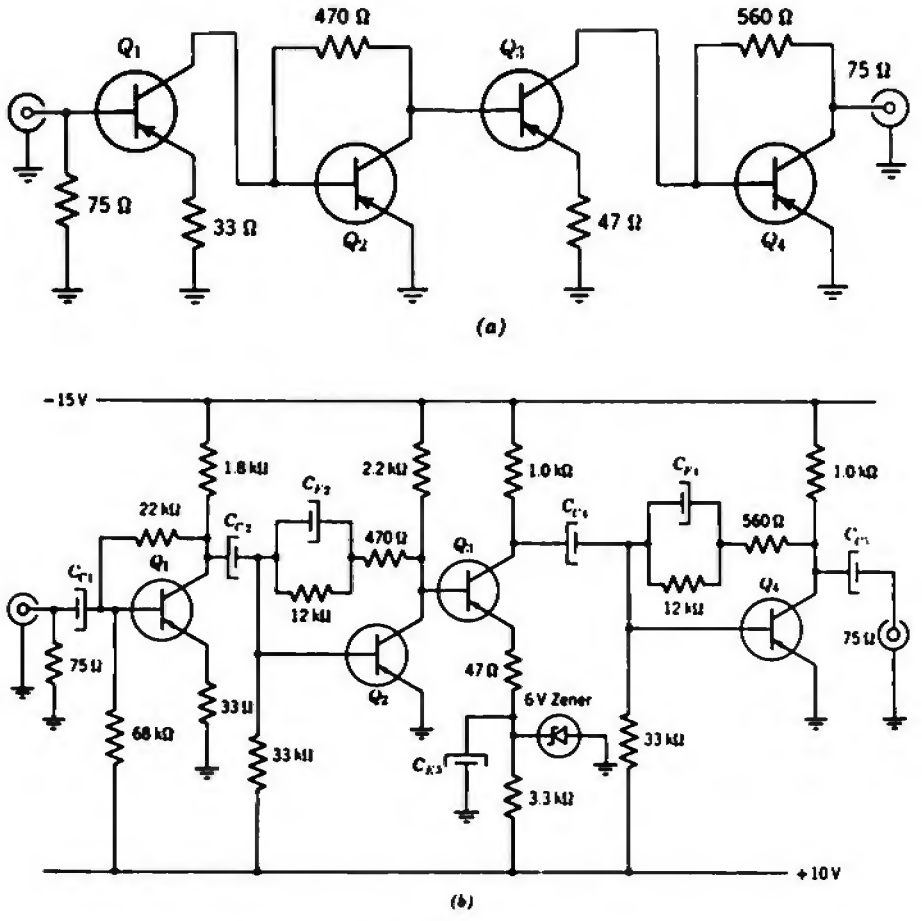


Fig. 12.18 Example of low-frequency compensation: (*a*) elemental mid-band circuit; (*b*) complete amplifier (less high-frequency peaking); (*c*) 50 Hz square-wave response. The transistors are type OC170/2N1516.

collector and emitter. Q_4, the output transistor, operates at a rather large current (10 mA) to increase the signal voltage which can be developed across the 75 Ω external load. It is left for the reader to verify that these quiescent conditions are achieved with the biasing system shown in Fig. 12.18*b*. The amplifier is required to pass a 50 Hz square wave with less than 5% tilt.

The input series-feedback stage has collector-feedback biasing to eliminate the emitter bypass capacitor. This stage can therefore be designed to contribute no dominant singularities at all. The second and third stages are direct coupled, and are a building-block of the form discussed in Section 12.3.2.2. There is a dominant pole due to Q_3 emitter bypass, a dominant zero due to Q_2 feedback network, and four other nondominant singularities. The output stage has a dominant zero due to its feedback network, and three nondominant singularities. In addition, there is a dominant pole due to the load coupling network. In all, there are 14 singularities:

2 dominant poles (C_{E3} and C_{C5})
2 dominant zeros (C_{F2} and C_{F4})
5 nondominant poles (C_{C1}, C_{C2}, C_{F2}, C_{C4}, and C_{F4})
1 nondominant zero (C_{E3})
4 zeros at the origin (C_{C1}, C_{C2}, C_{C4}, and C_{C5})

Equation 12.27 gives the condition for compensating the square-wave response of an amplifier, that the sums of the pole and the zero frequencies should be equal. In the present amplifier, optimum compensation will therefore correspond approximately to canceling the two dominant poles with the zeros. Hence for each of the shunt stages

(i) the zeros of Z_F and Z_L are close together,
(ii) the second bracket on the right-hand side of Eq. 12.61 is approximately unity at all frequencies because $(r_B + \beta_N r_E)$ is small and Z_E is zero.

As a result, A_i' has one pole only, plus a zero at the pole of Z_F. Therefore each shunt-feedback stage has a dominant zero at the zero of Z_F plus a nondominant pole near the pole of Z_F. Because the zeros of Z_F and Z_L are not exactly coincident, a second pole and zero appear close to the dominant zero, but these are very close together and cancel to a good approximation. The second bracket in Eq. 12.61 contributes a nondominant closed-loop pole near the zero of Z_S (attributable to the coupling capacitor and given approximately by Eq. 12.37) and there is a zero at the origin.

If an amplifier is to be capable of adjustment for very small tilts (1 or 2%), the dominant pole frequencies should be less than one tenth the square-wave repetition frequency (Section 12.1.1.2). The necessary values of C_{E3} and C_{C2} are thus 600 and 400 μF, respectively, both of which are rather unreasonably large. The largest capacitors with adequate voltage ratings available to the authors were about half these values, 330 and 220 μF. These capacitors are therefore specified; it is expected that the "bowing" on a 50-Hz square wave will be less than the allowed maximum of 5% tilt. The feedback capacitors are chosen for most near equality between the sums of pole and zero frequencies. The other capacitors in the circuit are chosen so that the nondominant singularities lie two or

Table 12.3 Low-Frequency Compensation of the Circuit Shown in Fig. 12.18

Capacitor	Value (μF)	Pole Time Constant (msec)	Pole Frequency (radian/sec)	Zero Time Constant (msec)	Zero Frequency (radian/sec)
C_{C1}	47	150 (modified Eq. 12.38)	6.7	∞	0.0
C_{C2}	100	180 (Eq. 12.37)	5.6	∞	0.0
C_{F2}	33	396 (Eq. 12.55)	2.5	14.0 (Eq. 12.56)	71.4
C_{E3}	330	15.0 (Eq. 12.44)	66.7	1090 (Eq. 12.43)	0.9
C_{C4}	100	100 (Eq. 12.37)	10.0	∞	0.0
C_{F4}	22	264 (Eq. 12.55)	3.8	11.9 (Eq. 12.56)	84.2
C_{C5}	220	16.5 (modified Eq. 12.38)	60.6	∞	0.0
		Totals:	155.9		156.5

three octaves inside the dominant poles. Table 12.3 lists the pole and the zero frequencies.

There is no really satisfactory place for a low-frequency adjusting potentiometer in Fig. 12.18*b*. This is often the case with transistor circuits because there are dc currents in all electrodes and changing any

circuit resistance changes the dc conditions. There is scope for originality by circuit designers. Rather cumbersome approaches are to select one capacitor (usually an emitter bypass or feedback blocking capacitor), or to add small selected capacitors in shunt with a large unselected capacitor to give optimum square-wave response. Figure 12.18*c* shows the 50-Hz square-wave response of the amplifier. All capacitors were 20% tolerance types with the nominal values given in Table 12.3, although C_{E3} was selected; the "bow" is about 2% peak-to-peak.

Chapter 13

High-Frequency Compensation of Amplifiers

The ultimate limit to the high-frequency performance of an amplifying device is set by inertia of the charge in its control region. Inertia frustrates all attempts to change the charge distribution rapidly, and therefore sets an upper limit to the rate of change of output current. Charge-control theory ignores inertia and assumes that the time taken to rearrange the charge can be neglected. Despite this approximation, the charge-control representations of active devices are valid for low-pass amplifiers. In combination, the mutual conductance and input capacitance of a device determine its current gain-bandwidth product (Section 2.5.1), the frequency beyond which it cannot contribute a gain that is greater than unity. This practical limiting frequency is far below the ultimate limit.

The capacitances in the charge-control representation of a device are of two types; intrinsic capacitances are essential to charge-control theory, whereas extrinsic capacitances are due to technological imperfections. For vacuum tubes the extrinsic capacitances from anode to grid, grid to ground, and anode to ground are dominant and have a controling effect on bandwidth. For transistors both intrinsic and extrinsic capacitances may be important. This chapter discusses techniques that can partially compensate the effect of these capacitances and so increase the realizable bandwidth or improve the transient response.

Compensating techniques fall into two main classes:

(i) *inductance peaking* in which the fall in capacitive reactance at high frequencies is compensated by the addition of inductive reactance which increases with frequency;

(ii) *feedback peaking* in which the tendency for a feedback amplifier to regenerate at high frequencies is used to offset the normal fall in gain.

The term *peaking* enjoys common usage and is quite descriptive of the process of stretching the response to some peak frequency beyond the bandwidth of an uncompensated amplifier. These apparently different approaches affect the dynamic response similarly. Peaked amplifiers usually have complex conjugate pairs of poles whose presence could be ascribed either to the resonant interchange of energy between inductance and capacitance or to the resonant circulation of energy between the input and output of a feedback amplifier. The inclusion of complex pole pairs gives rise to a far greater variety of potentially useful pole-zero patterns (corresponding to potentially useful dynamic responses) than in the simple case of real poles alone. For this reason it is necessary first to develop design techniques that produce accurately defined complex poles as well as real poles, and second to have information available on the correspondence between dynamic response and the pole-zero geometries that are useful in amplifiers. These topics are discussed in Sections 13.1 and 13.2, respectively.

The two introductory sections prepare the way for the main part of the chapter which provides detailed information on the performance of a number of useful compensating schemes. These include both inductance and feedback peaking together with combinations of the two methods. Compensating techniques were developed to increase the gain-bandwidth product $\mathscr{GB}$ that could be realized from available amplifying devices. As a result, it is customary to compare compensating techniques not only by reference to their dynamic response, but also by the gain-bandwidth product that is achieved. The most convenient datum for such comparisons is $\mathscr{GB}$ for a simple unpeaked stage, which is related to the device $\mathscr{GB}$ introduced in Section 2.5.1. Modern devices have high values of $\mathscr{GB}$—up to about 200 MHz for readily available vacuum tubes and $1\frac{1}{2}$ GHz for transistors—and the use of highly efficient compensating schemes is not all-important. The concept of gain-bandwidth, however, does provide a useful basis for comparing otherwise similar designs, and warrants the discussion in Section 13.6.

13.1 AN INTRODUCTION TO THE DESIGN OF COMPENSATED AMPLIFIERS

There are two distinct steps in designing an amplifier (or passive network) that must pass signals and consequently have an acceptable dynamic response:

1. The first step, the *approximation problem*, involves the choice of a physically realizable pole-zero pattern whose dynamic response approximates the "ideal" dynamic response required for a specific system. There is usually no optimum or unique solution to this problem.

2. The second step, the *realization problem*, involves the design of an amplifier that has both a suitable singularity pattern and the desired midband transfer function. There is, likewise, no unique solution to this problem; competing designs may be compared in terms of their gain-bandwidth products, but this is not the sole criterion of quality and other factors may influence the designer's decision.

These two steps are, of course, interdependent and any decision made in one influences the other. It might be expected that some of the formal techniques of modern network synthesis would assist the designer in making decisions; for example, standard realizability criteria could be used to check whether singularity patterns yielding suitable dynamic responses were, in fact, physically realizable. Again, completely formal approaches to the realization problem (corresponding to those used for realizing passive networks) might be thought a practical possibility. These formal mathematical tools, however, find little general application for at least two reasons. First, the singularity patterns of designable amplifiers must be simple and can thus be tested for physical realizability by inspection. Second, formal approaches to active network synthesis appear to be most inflexible and are capable of application only to a few combinations of networks and amplifying devices. Many attempts have been made to produce completely formal approaches to amplifier synthesis, but these appear to have resulted in more theses for higher degrees than useful amplifiers!

The approach to amplifier design suggested in this book is based on the use of two "catalogs" of information:

1. The *approximation catalog* contains information on the dynamic response of useful singularity patterns. Such a catalog does not appear to exist as a complete, readily available entity. The totality of information is spread widely in books, journals, and the knowledge of individual designers. Some relevant information is contained in Section 13.2 of this book, but for much of the vast amount of information available the reader is referred to other sources.

2. The *realization catalog* contains information on the singularity patterns of useful amplifier building-blocks. The main purpose of this chapter is to discuss the details of a number of suitable building-blocks and so present a useful, but incomplete, realization catalog.

Both catalogs are large, continually expanding, and may be extended readily to satisfy special new requirements. The information in this chapter should provide a sound basis for any such extensions. In designing an amplifier, the realization catalog is used to find which building-blocks might be combined to give the over-all singularity pattern chosen from the approximation catalog.

As an example, Fig. 13.1 illustrates the design of a simple amplifier. The "ideal" dynamic response for this example is defined by the voltage-gain/frequency characteristics shown as solid lines in Fig. 13.1*a*. This particular response is often taken as the ideal for wide-band amplifiers

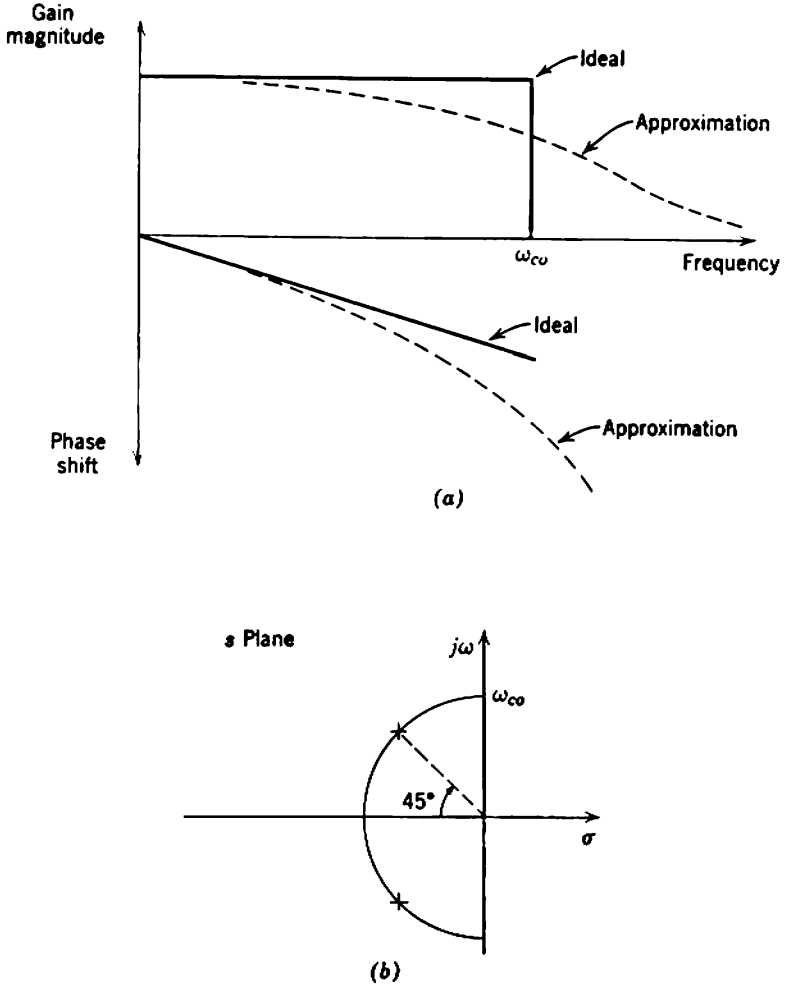
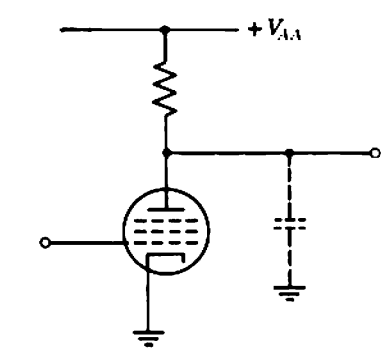
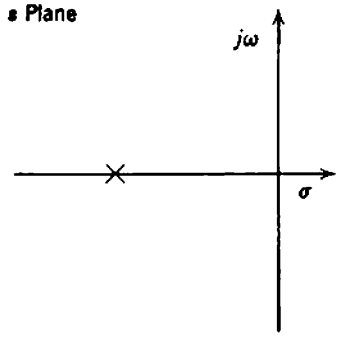
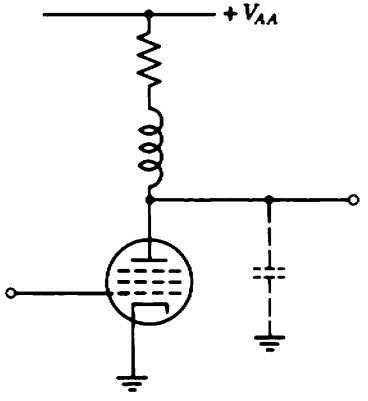


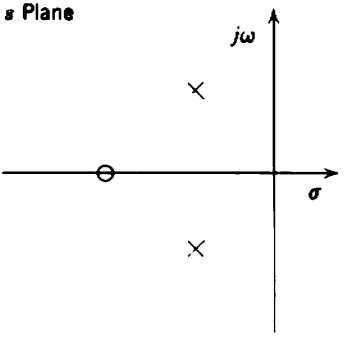
Fig. 13.1 Example of amplifier design: (*a*) ideal and approximating responses; (*b*) approximating singularity pattern; (*c*) single-pole building-block; (*d*) 2-pole, 1-zero building-block; (*e*) complete amplifier.



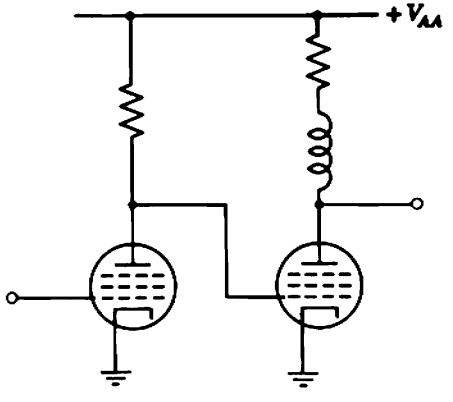


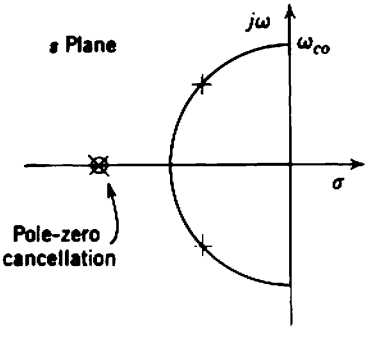
(c)





(d)





(e)

Fig. 13.1 (*continued*)

and is termed the *brick-wall response.* The cutoff frequency ω_{c0} is chosen so that all of the information-bearing Fourier components of the signal are enclosed within the passband; the gain is constant throughout the passband, and the phase delay varies linearly with frequency. Therefore, any signal waveform will be delayed during its transit through the amplifier but its shape will not be distorted. The brick-wall response can never be achieved with realizable circuits but can be approximated in many ways; the most suitable approximation depends on the application.

The realizable singularity pattern of Fig. 13.1*b* yields the response shown as broken lines in Fig. 13.1*a*; it is assumed for the example that this is an adequate approximation. Figures 13.1*c* and *d* show the pole-zero patterns for two voltage-amplifying building-blocks that might be used to realize the required over-all pattern. If these building-blocks can be cascaded without the second reacting on the first, the over-all voltage gain function is the product of the individual voltage gain functions. Consequently, the over-all singularity pattern is obtained by superimposing the patterns of the individual building-blocks. If the building-blocks are cascaded as shown in Fig. 13.1*e*, pole-zero cancellation occurs and the desired over-all response is achieved. As explained in Section 7.3.1, there is no fundamental reason why these building-blocks should not be cascaded in the reverse order; practical considerations such as noise or the capacity to produce undistorted signals often compel the designer to choose a particular order.

In concluding this introduction to design, it should be emphasized that any amplifier requires an acceptable magnitude of mid-band transfer function. Although it is possible to increase the bandwidth over which the transfer function magnitude remains substantially constant and the phase delay linear, gain-bandwidth considerations demand that the mid-band gain fall as the bandwidth is increased. With some modifications (discussed in Section 13.6), the gain of a multistage amplifier is given by an expression of the form

$$\frac{\text{product of the realizable } \mathscr{GB} \text{ for all devices}}{\text{product of the distances of all poles from the origin}}.$$

To facilitate comparison of various singularity pattern types, the responses in Section 13.2 are normalized to the geometric mean of the distances of the poles from the origin; zeros count as reciprocal poles in taking this mean. In general, there is a trade-off between one aspect of the dynamic response and the others as the singularity pattern is varied. The approximation problem is to decide on an acceptable compromise, remembering also the limitation of gain-bandwidth product.

13.2 THE DYNAMIC RESPONSE OF REALIZABLE SINGULARITY PATTERNS

The pole-zero patterns of stable, realizable networks must satisfy the following conditions:

(i) poles and zeros must either be real or occur as complex conjugate pairs,
(ii) poles must have negative real parts,
(iii) the number of zeros cannot exceed the number of poles.

Despite these restrictions, the number of different pattern geometries that correspond to potentially useful dynamic responses is enormous even for systems having quite small numbers of poles and zeros; for numbers of singularities greater than about five, the task of cataloging responses becomes virtually impossible. Therefore it is good practice to design amplifiers so that their singularity patterns are relatively simple. The most useful design data is that pertaining to such simple patterns and, fortunately, much useful information is available.

There are two powerful methods for simplifying what would otherwise be a complicated singularity pattern—*dominant singularities* and *pole-zero cancellation*. Both methods are introduced in earlier chapters, particularly Section 12.1.2.2. Dominant singularities are so situated in the complex frequency plane that they control the main features of the dynamic response, whereas the other singularities are situated so that their effect is negligible. For most purposes a complete singularity pattern can be replaced by its dominant group. As an example, Fig. 13.2 shows a

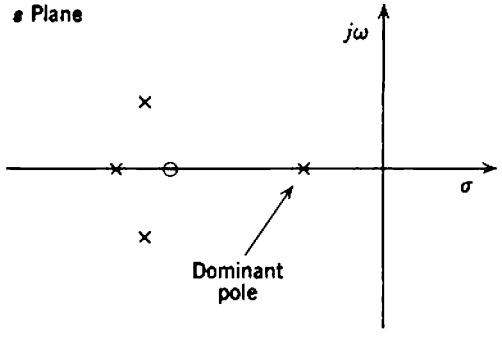


Fig. 13.2 A dominant-pole singularity pattern.

high-frequency dominant-pole pattern; one pole (the dominant pole) lies much closer to the origin than the other singularities, and it can be verified readily that this dominant pole controls the magnitude and phase of the response up to and well beyond the cutoff frequency of the amplifier. Pole-zero cancellation is used in the example discussed in Section 13.1 and illustrated by Fig. 13.1; the pole and zero need not coincide perfectly. Pole-zero cancellation is more subtle than the use of dominant singularities for simplification and can be used to great effect.

Information that relates dynamic response to the pole-zero pattern can be obtained in four ways:

(i) from mathematical analysis,
(ii) by reference to tabular or graphical sets of data,
(iii) by measurements on analogs,
(iv) by digital computation.

The transient response or frequency response for any singularity pattern can be determined by analysis. Solutions based on classical or operational analyses for the step-function response of many simple systems are quoted in the literature, and the response to other forms of transient input can be obtained analytically (for simple input waveforms) or graphically by means of the convolution integral (for unusual input waveforms). Equally, the use of vector algebra and asymptotic plotting allows the frequency response to be calculated. Results based on these analytical procedures are listed in Section 13.2.1. Although individual analyses of this sort are relatively straightforward, the calculations tend to become tedious when a number of slightly different geometrical possibilities are to be investigated. Under such circumstances it is simpler by far to make use of a singularity pattern whose response has been cataloged. A number of workers have performed an invaluable service for amplifier designers by providing extensive tabulations and graphical plots for many types of singularity pattern; some of these sources of information are referenced in Section 13.2.2. For pole-zero patterns of greater complexity or unusual geometry, information on the transient and/or frequency response can be obtained from analog or digital computers; Section 13.2.3 discusses the use of analogs, but readers are referred to the literature for information on digital computations.*

* The following may be of value: M. L. LIOU, "A novel method of evaluating transient response," *Proc. Inst. Elec. Electronics Engrs.*, **54**, 20, January 1966; W. E. THOMPSON, "Evaluation of transient response," *Proc. Inst. Elec. Electronics Engrs.*, **54**, 1584, November 1966; W. EVERLING, "On the evaluation of exp AT," *Proc. Inst. Elec. Electronics Engrs.*, **55**, 413, March 1967.

13.2.1 Mathematical Analysis

Two general types of pole-zero pattern are amenable to simple mathematical analysis. The first is that in which a small number of poles and zeros (one, two, or three at the most) may take up a variety of positions. Examples discussed in this chapter are the single-pole, two-pole, and 2-pole 1-zero functions. The second comprises all-pole patterns in which quite large numbers of poles may exist provided they take up specified positions. Examples are the coincident-pole, maximally-flat, equal-ripple, and linear-phase functions.

13.2.1.1 Single-Pole Function

A single pole on the negative real axis is the simplest low-pass function. Although discussed briefly in Section 7.5.1, this function is considered again here for completeness. The frequency-dependent term $\psi(s)$ in a single-poled transfer function is

$$\psi(s) = \frac{1}{1 + s\tau}. \tag{13.1}$$

Figure 13.3*a* shows the step-function response, an exponential of the form

$$\rho(t) = 1 - \exp\left(-\frac{t}{\tau}\right), \tag{13.1}$$

for which the 10 to 90% rise time is

$$t_r \approx 2.2\,\tau \tag{13.3}$$

and the percentage overshoot is zero. The frequency response is given in magnitude and phase by

$$|\psi(j\omega)| = \frac{1}{\sqrt{1 + \omega^2\tau^2}} \tag{13.4a}$$

and

$$\underline{/\psi(j\omega)} = \arctan(-\omega\tau). \tag{13.4b}$$

These functions are plotted against frequency (normalized to $1/\tau$, the distance of the pole from the origin) in Figs. 13.3*b* and *c*, and combined as a polar plot in Fig. 13.3*d*. The 3-dB radian cutoff frequency, which is equal to the bandwidth, is

$$\omega_{co} = \mathscr{B} = \frac{1}{\tau}. \tag{13.5}$$

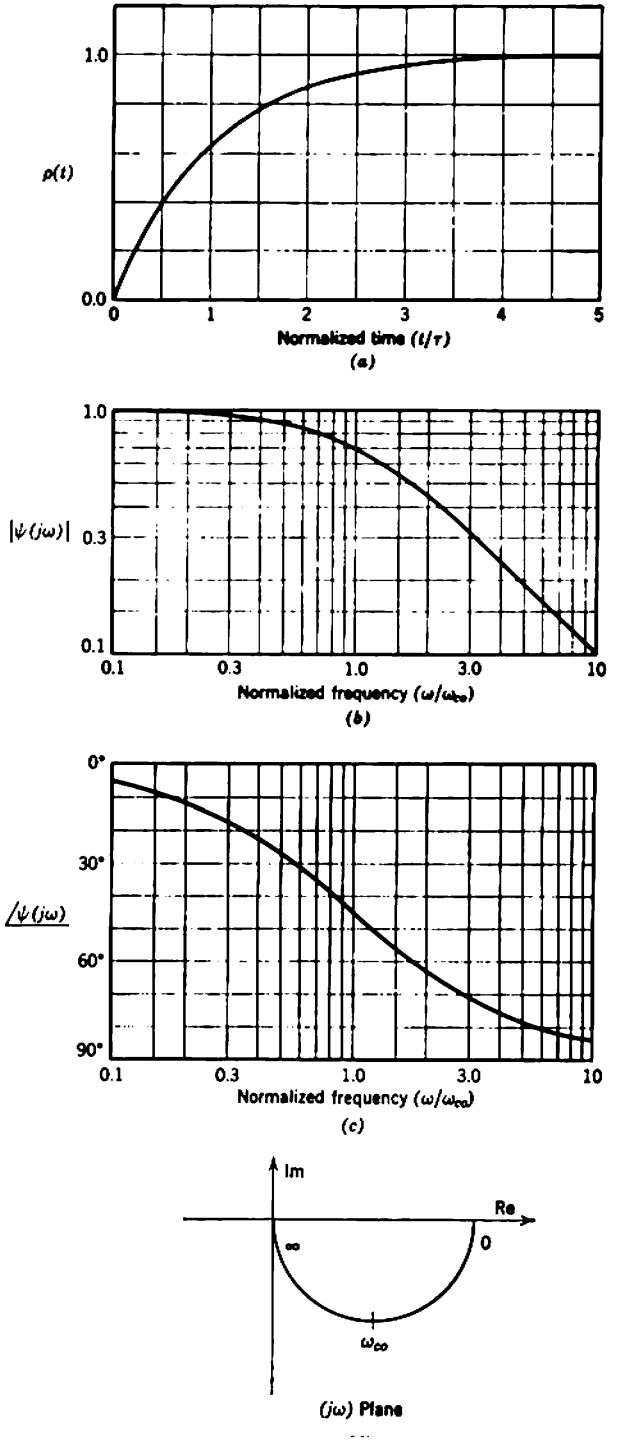


Fig. 13.3 The single-pole function: (*a*) step response; (*b*) magnitude of frequency response; (*c*) phase of frequency response; (*d*) polar plot of frequency response.

13.2.1.2 Multiple Coincident Poles

The simplest multiple-poled transfer function has a multiple coincident pole on the negative real axis. The corresponding frequency-dependent term $\psi(s)$ for an n_p-order pole is

$$\psi(s) = \frac{1}{(1 + s\tau)^{n_p}}. \tag{13.6}$$

The response to a unit step is the inverse transform of

$$\rho(s) = \frac{1}{s(1 + s\tau)^{n_p}},$$

which can be evaluated as

$$\rho(t) = 1 - \sum_{k=0}^{n_p-1} \left[\frac{(t/\tau)^k}{(k)!} \exp\left(-\frac{t}{\tau}\right)\right]. \tag{13.7}$$

This function is plotted in Fig. 13.4*a* for values of n_p from 1 to 10. Notice that as n_p increases, the 10 to 90% rise time $t_r(n_p)$ increases and an initial delay becomes perceptible. If the rise time is plotted against $\sqrt{n_p}$, as in Fig. 13.4*b*, it is observed that the resulting plot is very nearly a straight line:

$$t_r(n_p) \approx t_r(1)\sqrt{n_p}, \tag{13.8}$$

where $t_r(1)$ $(= 2.2\,\tau)$ is the rise time of a single-poled function. This empirical result is very useful and forms the basis for a rule given in Section 13.2.1.8.

The magnitude and phase of the sinusoidal frequency response are given by

$$|\psi(j\omega)| = \frac{1}{(1 + \omega^2\tau^2)^{n_p/2}} \tag{13.9a}$$

and

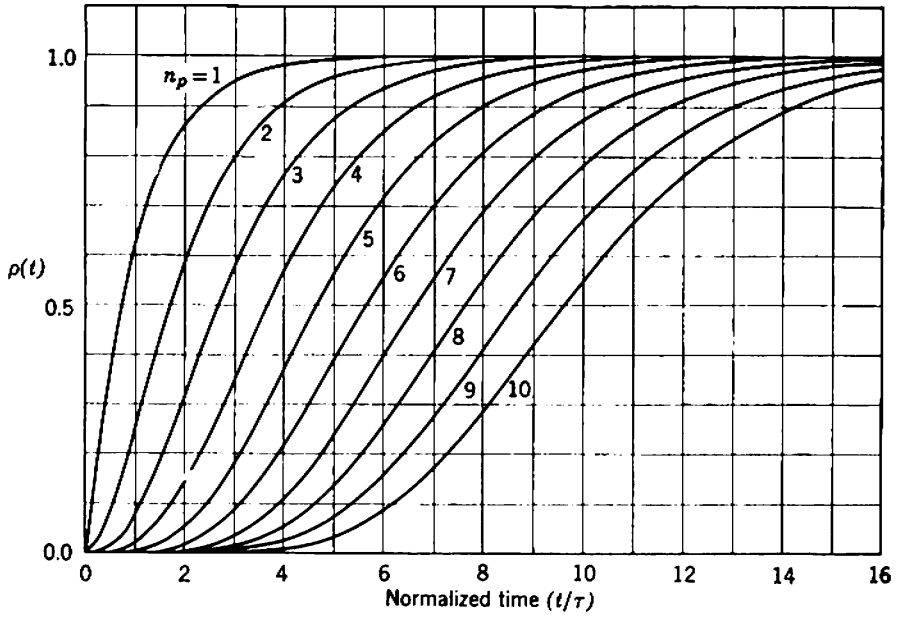
$$\underline{/\psi(j\omega)} = n_p \arctan(-\omega\tau). \tag{13.9b}$$

It follows from Eq. 13.9*a* that the 3-dB radian cutoff frequency or bandwidth is

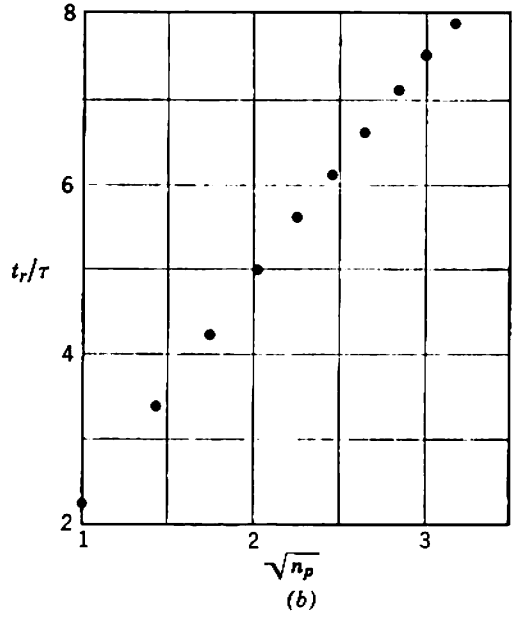
$$\omega_{co}(n_p) = \mathscr{B}(n_p) = \omega_{co}(1)\sqrt{2^{1/n_p} - 1}, \tag{13.10}$$

where $\omega_{co}(1)$ $(= 1/\tau)$ is the cutoff frequency of a single-poled function. The square-root term is known as the *bandwidth reduction factor* and is similar in form to the term introduced in Section 12.1.2.2 in connection with low-frequency response. It is left as an exercise for the reader to show that

$$\sqrt{2^{1/n_p} - 1} \approx \frac{1}{1.2\sqrt{n_p}}. \tag{13.11}$$



(a)



(b)

Fig. 13.4 The multiple-coincident-pole function: (*a*) step response; (*b*) rise time versus number of poles.

Table 13.1 Bandwidth Reduction Factor for Various Order Multiple Coincident Poles

Order of Pole n_p	Reduction Factor $\sqrt{2^{1/n_p} - 1}$	Approximation $\frac{1}{1.2\sqrt{n_p}}$
1	1.000	0.833
2	0.644	0.589
3	0.510	0.481
4	0.435	0.417
5	0.386	0.373
6	0.350	0.341
7	0.323	0.315
8	0.301	0.294
9	0.283	0.278
10	0.268	0.264

The validity of the approximation for typical values of n_p is indicated by the figures in Table 13.1. From Eq. 13.10 and the approximations given by Eqs. 13.8 and 13.11 it follows that

$$\omega_{co}(n_p)\, t_r(n_p) \approx \frac{\omega_{co}(1)\, t_r(1)}{1.2}. \tag{13.12}$$

Thus the product of cutoff frequency and rise time is approximately independent of the number of stages. The magnitude of the right-hand side of Eq. 13.12 can be evaluated from Eq. 13.3 to yield

$$\omega_{co}(n_p)\, t_r(n_p) = 1.8 \tag{13.13a}$$

or

$$f_{co}(n_p)\, t_r(n_p) = 0.3. \tag{13.13b}$$

This, like Eq. 13.8, is a particularly useful result that can be extended to more complicated systems.

13.2.1.3 Maximally-Flat-Magnitude Function

Maximally-flat-magnitude functions are a family of mathematical functions that were investigated by Butterworth and are often named after him. They are based on a somewhat arbitrary choice of approximation to the ideal brick-wall magnitude response, defined mathematically so that the resulting transfer functions are tractable as far as the magnitude of frequency response is concerned, but no attempt is made to approximate

the ideal linear-phase response. A convenient and general statement of the conditions for maximal flatness of $\psi(s)$ can be formulated as follows:

Let the function $\psi(s)$ be a polynomial in s, with n_z zeros and n_p poles, such that

$$\psi(s) = \frac{1 + u_1 s + u_2 s^2 + \cdots + u_{n_z} s^{n_z}}{1 + v_1 s + v_2 s^2 + \cdots + v_{n_p} s^{n_p}}, \tag{13.14}$$

where $u_1 \ldots$ and $v_1 \ldots$ are constants. The real function $|\psi(j\omega)|^2$ is an even function of frequency, and can be written as

$$|\psi(j\omega)|^2 = \psi(j\omega) \times \psi(-j\omega) = \frac{1 + a_2\omega^2 + a_4\omega^4 + \cdots + a_{2n_z}\omega^{2n_z}}{1 + b_2\omega^2 + b_4\omega^4 + \cdots + b_{2n_p}\omega^{2n_p}}, \tag{13.15}$$

where $a_2 \ldots$ and $b_2 \ldots$ are constants. The function $\psi(s)$ is defined as maximally flat if all of the $(n_p - 1)$ partial derivatives of $|\psi(j\omega)|^2$ with respect to ω^2 are zero when ω is zero:

$$\left(\frac{\partial[|\psi(j\omega)|^2]}{\partial[\omega^2]}\right)_{\omega=0} = 0, \tag{13.16a}$$

$$\left(\frac{\partial^2[|\psi(j\omega)|^2]}{\partial[\omega^4]}\right)_{\omega=0} = 0, \tag{13.16b}$$

$$\vdots$$

$$\left(\frac{\partial^{(n_p-1)}[|\psi(j\omega)|^2]}{\partial[\omega^{2(n_p-1)}]}\right)_{\omega=0} = 0. \tag{13.16c}$$

More directly, the condition for maximal flatness can be expressed as

$$\frac{a_2}{b_2} = \frac{a_4}{b_4} = \cdots = \frac{a_{2n_z}}{b_{2n_z}} = 1, \tag{13.17a}$$

$$b_{2(n_z+1)} = b_{2(n_z+2)} = \cdots = b_{2(n_p-1)} = 0, \tag{13.17b}$$

$$b_{2n_p} = \text{finite}. \tag{13.17c}$$

There is an important class of maximally-flat function for which the mathematical definition reduces to a simple geometrical rule. This is the n_p-order all-pole function for which Eq. 13.14 becomes

$$\psi(s) = \frac{1}{1 + v_1 s + v_2 s^2 + \cdots + v_{n_p} s^{n_p}}. \tag{13.18}$$

The definition of maximal flatness (Eq. 13.17) requires that

$$|\psi(j\omega)|^2 = \frac{1}{1 + b_{2n_p}\omega^{2n_p}} \tag{13.19}$$

which can be written in terms of complex frequency s instead of real frequency ω by noting that $s = j\omega$, and therefore

$$|\psi(s)|^2 = \psi(s) \times \psi^*(s) = \frac{1}{1 + (-1)^{n_p} b_{2n_p} s^{2n_p}}. \tag{13.20}$$

The poles of Eq. 13.20 are the $2n_p$ roots of

$$s^{2n_p} = -\frac{(-1)^{n_p}}{b_{2n_p}}, \tag{13.21}$$

which lie equally spaced on a circle of radius $(1/b_{2n_p})^{1/2n_p}$. This is shown in Figs. 13.5*a* and *b* for n_p odd and even, respectively. In either case half the roots lie in the left half-plane and the other half lie in the right half-plane. If $\psi(s)$ is not to be a regenerative function, it can contribute only the roots in the left half-plane; $\psi^*(s)$ contributes the other roots. Therefore the poles of a maximally-flat all-pole function must lie equally spaced on the semicircle of radius

$$|s| = \left(\frac{1}{b_{2n_p}}\right)^{1/2n_p} = \left(\frac{1}{v_{n_p}}\right)^{1/n_p} = \omega_{co} \tag{13.22}$$

in the left half-plane. Figure 13.5*c* shows some examples of maximally-flat all-pole functions, and Fig. 13.5*d* shows normalized plots of frequency response. The 3-dB radian cutoff frequency ω_{co} is equal to the radius of the semicircle, that is, the geometric mean of the distance of the poles

Table 13.2 Maximally-Flat All-Pole Functions

Order	Pole Positions			Cutoff Frequency	Rise Time	Overshoot
n_p	Radius $\|s\|$	Real Part σ	Angle (°)	ω_{co}	t_r (sec)	(%)
1	1.000	−1.000	0.0	1.000	2.20	0.0
2	1.000	−0.707	±45.0	1.000	2.15	4.3
3	1.000	−1.000	0.0	1.000	2.29	8.2
	1.000	−0.500	±60.0			
4	1.000	−0.924	±22.5	1.000	2.43	10.9
	1.000	−0.383	±67.5			
5	1.000	−1.000	0.0	1.000	2.56	12.8
	1.000	−0.809	±36.0			
	1.000	−0.309	±72.0			

These results are normalized so that the geometric mean distance of the poles from the origin is 1 radian/sec. Notice that the 10 to 90% rise time is minimum for $n_p = 2$.

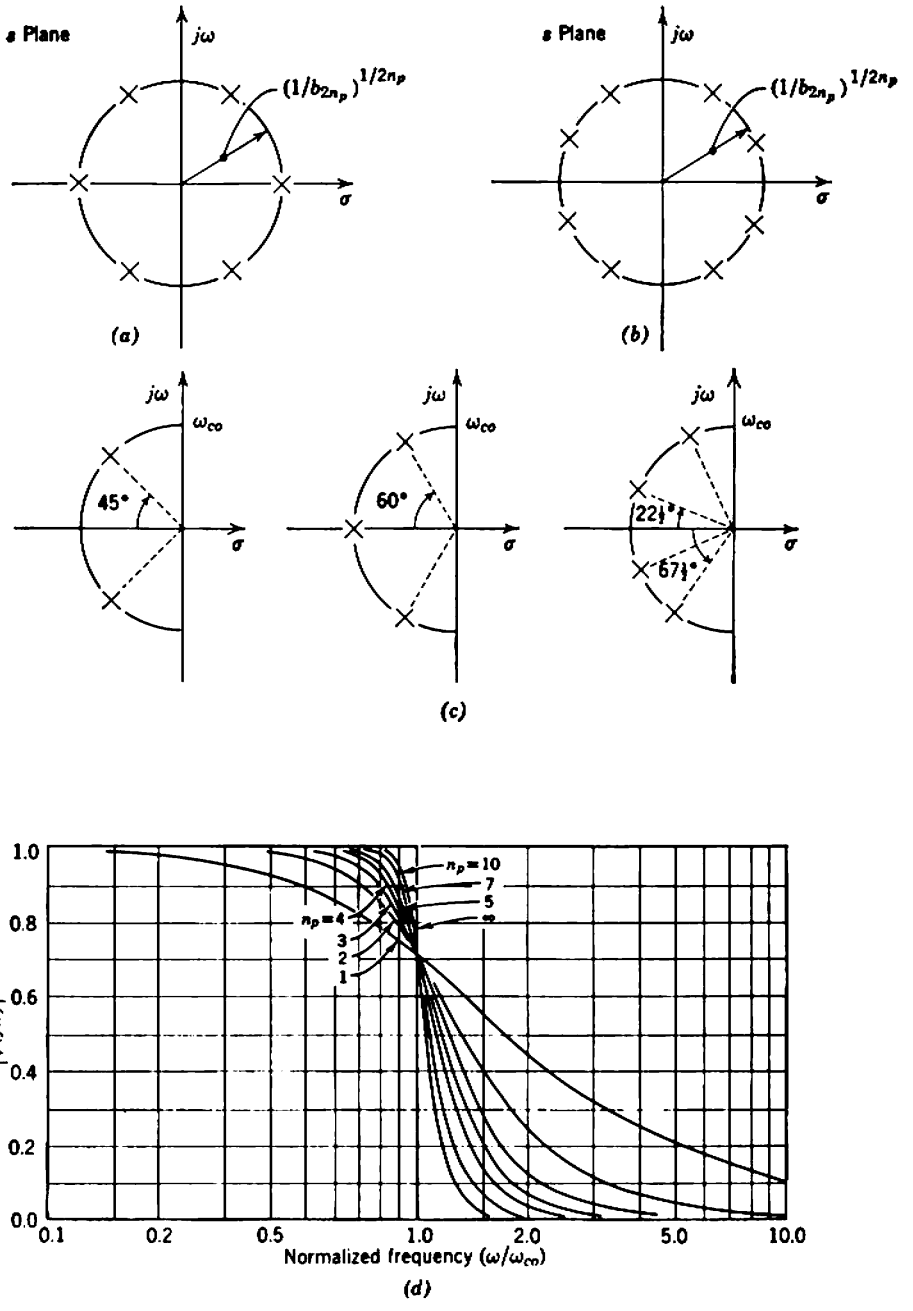


Fig. 13.5 The maximally-flat-magnitude function: (*a*) roots of Eq. 13.21 for $n_p = 3$; (*b*) roots of Eq. 13.21 for $n_p = 4$; (*c*) 2-, 3-, and 4-pole maximally-flat functions; (*d*) magnitude of frequency response.

from the origin. Notice that the accuracy of the approximation to the brick-wall response increases as the number of poles increases. Except for the simple case of a single pole, maximally-flat functions give rise to overshoot in response to a step function; Table 13.2 lists the performance up to five poles.

13.2.1.4 Linear-Phase Functions

Linear-phase functions are a family of mathematical functions that are also known as Thomson functions (after the worker who first investigated their possibilities) or Bessel functions (because this is the convenient mathematical form that the functions take). These functions are based on a maximally-flat approximation to the ideal linear phase delay; no attempt is made to approximate the brick-wall magnitude response. In general, $\psi(j\omega)$ can be written as

$$\psi(j\omega) = \frac{N(j\omega)}{D(j\omega)}, \tag{13.23}$$

where $N(s)$ and $D(s)$ are the numerator and denominator polynomials of $\psi(s)$. The ratio of the imaginary to the real part of $\psi(j\omega)$ is

$$\theta(\omega) = \frac{\text{Im}[N(j\omega) \times D^*(j\omega)]}{\text{Re}[N(j\omega) \times D^*(j\omega)]}, \tag{13.24}$$

where D^* is the conjugate of D, and the phase angle is

$$\varphi(\omega) = \arctan \theta(\omega). \tag{13.25}$$

A linear-phase function should have $\varphi(\omega)$ varying linearly with frequency. Therefore $d\varphi/d\omega$ should be constant; $\psi(s)$ is defined as a linear-phase function if $d\varphi/d\omega$ is maximally flat:

$$\frac{d[\varphi(\omega)]}{d\omega} = \left\{\frac{1}{1 + [\theta(\omega)]^2}\right\} \frac{d[\theta(\omega)]}{d\omega}. \tag{13.26}$$

For the special case of $\psi(s)$ being an all-pole function, it has been shown by many workers that the denominator polynomial $D(s)$ is a Bessel function.

The pole positions for Bessel functions are not given by a simple geometrical rule like those for a maximally-flat function. The poles do, however, lie approximately on a circle centered to the right of the origin (Fig. 13.6). The ratios of real to imaginary parts of the poles are increased from those of a Butterworth function, so linear-phase functions give smaller overshoot in response to a step function. This, however, is offset by a smaller bandwidth for the same geometric mean of the pole distances from the origin. Linear-phase functions are extremely useful in

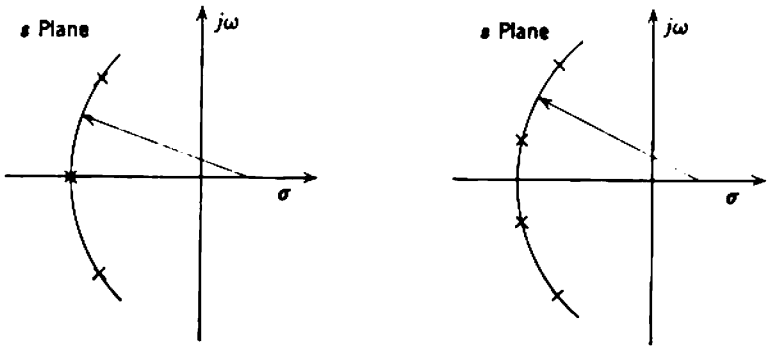


Fig. 13.6 Linear-phase functions having 3 and 4 poles.

the design of pulse amplifiers and Table 13.3 lists their performance up to five poles.

13.2.1.5 Equal-Ripple Functions

Equal-ripple-magnitude functions are a family of mathematical functions that were investigated by Tchebychev and, in a variety of spellings, are often named after him. These functions are based on a type of approximation to the brick-wall magnitude response in which a more

Table 13.3 Linear-Phase All-Pole Functions

Order n_p	Pole Positions			Cutoff Frequency ω_{co}	Rise Time t_r (sec)	Overshoot (%)
	Radius $\lvert s\rvert$	Real Part σ	Angle (°)			
1	1.000	−1.000	0.0	1.000	2.20	0.00
2	1.000	−0.866	±30.0	0.786	2.73	0.43
3	0.942	−0.942	0.0	0.712	3.07	0.75
	1.031	−0.746	±43.7			
4	0.944	−0.905	±16.7	0.659	3.36	0.83
	1.059	−0.657	±51.6			
5	0.927	−0.927	0.0	0.617	3.58	0.75
	0.960	−0.852	±27.5			
	1.083	−0.591	±56.9			

These results are normalized so that the geometric mean distance of the poles from the origin is 1 radian/sec. Notice that the overshoot is maximum for $n_p = 4$.

nearly vertical flank is achieved at the expense of a ripple inside the passband. In general, the magnitude function contains a Tchebychev polynomial of normalized frequency. The important property of Tchebychev polynomials is that their ordinates oscillate between turning points ± 1 when the abscissas lie in the range -1 to $+1$. The precise form of these polynomials is unimportant for this level of treatment.

The real frequency function $|\psi(j\omega)|^2$ for the special case of $\psi(s)$ being an all-pole function is defined by

$$|\psi(j\omega)|^2 = \frac{1}{1 + \epsilon[C_{n_p}(\omega/\omega_0)]^2}. \tag{13.27}$$

where $C_{n_p}(\omega/\omega_0)$ is a Tchebychev polynomial of degree n_p, and ϵ is the ripple parameter. Because the denominator must lie between 1 and $(1 + \epsilon)$, the magnitude response has the typical form shown in Fig. 13.7*a*; the number of ripples in the passband is equal to n_p, and the magnitude of the ripple is controlled by the value of ϵ. Notice that the normalizing frequency ω_0 is not the 3-dB cutoff frequency ω_{co}; rather, ω_0 is the frequency range over which the magnitude stays within the limit set on ripple. In most cases ω_{co} is greater than ω_0.

It is shown in specialized books that the poles of $\psi(s)$ lie on an ellipse (Fig. 13.7*b*) whose foci lie at $\pm j\omega_0$ and whose major axis is very close to $2\omega_{co}$. The poles have the same imaginary parts as the poles of a Butterworth function with the same cutoff frequency and the same number of poles, but their real parts are reduced by a factor k:

$$k = \tanh\left(\frac{1}{n_p}\operatorname{argsinh}\sqrt{\frac{1}{\epsilon}}\right). \tag{13.28}$$

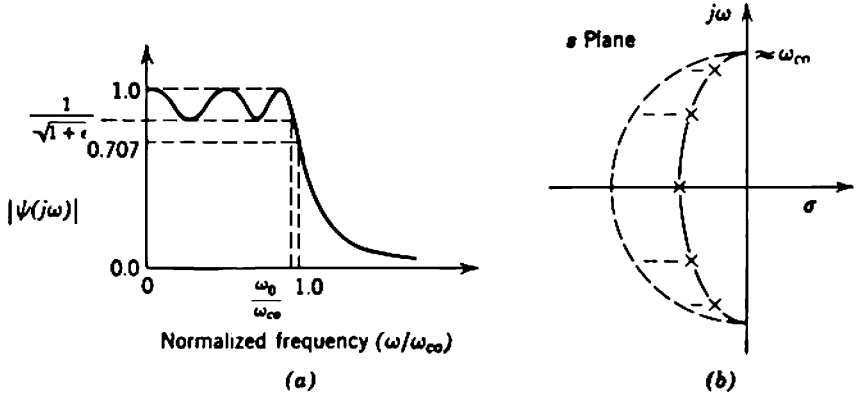


Fig. 13.7 The 5-pole equal-ripple function: (*a*) magnitude of frequency response; (*b*) pole positions.

Because the ratios of real to imaginary parts of the poles are reduced in the Butterworth to Tchebychev transformation, the latter functions give greater overshoot in response to a step function. For this reason equal-ripple functions are little used in amplifiers that must pass pulse signals.

13.2.1.6 Two-Pole Function

A transfer function that is amenable to fairly simple analysis for all values of its parameters is the two-pole function. The frequency-dependent function $\psi(s)$ is of the form

$$\psi(s) = \frac{1}{1 + 2\zeta s/\omega_0 + s^2/{\omega_0}^2}, \tag{13.29}$$

where ζ is the *damping ratio*, and ω_0 is the *undamped natural frequency* of the poles, numerically equal to their geometric mean distance from the origin. When considering the time response, ω_0 is often replaced by its reciprocal τ:

$$\omega_0 = \frac{1}{\tau}. \tag{13.30}$$

The positions of the poles are given for all values of ζ by

$$p_1 \text{ or } p_2 = \omega_0(-\zeta \pm \sqrt{\zeta^2 - 1}). \tag{13.31a}$$

For the limiting case of zero damping ratio, the poles lie on the imaginary axis at $\pm j\omega_0$ and the system oscillates at frequency ω_0.

If ζ is greater than unity, both poles lie on the negative real axis, spaced equally on either side of $-\zeta\omega_0$. If the damping is heavy, that is, if ζ is much greater than unity,

$$p_1 \approx -\frac{\omega_0}{2\zeta}, \tag{13.32a}$$

$$p_2 \approx -2\omega_0\zeta\left(1 - \frac{1}{4\zeta^2}\right). \tag{13.32b}$$

Thus p_1 becomes the dominant pole and the performance is independent of p_2. If ζ lies between zero and unity, Eq. 13.31*a* can be rewritten as

$$p_1 \text{ or } p_2 = \omega_0(-\zeta \pm j\sqrt{1 - \zeta^2}). \tag{13.31b}$$

In this case the poles lie on a circle of radius ω_0 centered at the origin, and their real part is $-\zeta\omega_0$; the angle of the poles in the left half-plane is arccos ζ.

If the poles are real, the step-function response is

$$\rho'(t) = 1 - \frac{1}{2}\left(1 + \frac{\zeta}{\sqrt{\zeta^2 - 1}}\right)\exp(p_1 t) - \frac{1}{2}\left(1 - \frac{\zeta}{\sqrt{\zeta^2 - 1}}\right)\exp(p_2 t), \tag{13.33a}$$

whereas the step-function response for complex poles is

$$\rho''(t) = 1 - \frac{\exp(-\zeta t/\tau)}{\sqrt{1-\zeta^2}} \sin\left(\frac{t}{\tau}\sqrt{1-\zeta^2} + \arccos\zeta\right). \quad (13.33b)$$

Comparison with Eq. 13.2 shows the increase in analytical complexity associated with the change from one to two poles. As with more complicated systems, the step-function response is conveniently presented as a family of curves (Fig. 13.8*a*). The time scale is normalized to τ and thus the form of response depends only on the value of ζ. No simple general expression equivalent to Eq. 13.3 is available for rise time, but approximate values and the percentage overshoot can be obtained from Fig. 13.8*a*. Notice that the rise time increases as the damping ratio is increased. The overshoot is zero so long as ζ is greater than unity and the poles are real, but becomes finite and increases monotonically with reduction of ζ below unity. The overshoot, however, does not exceed 5% until ζ falls below 0.7. Table 13.4 lists exact data for a few special cases.

Table 13.4 Two-Pole Function

Damping Ratio ζ	Cutoff Frequency ω_{co}	Rise Time t_r (sec)	Over-shoot (%)	Remarks
1.0	0.644	3.37	0.0	Coincident poles
0.866	0.786	2.73	0.4	Linear-phase (30° poles)
0.707	1.000	2.15	4.3	Maximally-flat-magnitude (45° poles)
0.5	1.272	1.62	16.4	Half-critical damping (60° poles)

These results are normalized so that the geometric mean distance of the poles from the origin is 1 radian/sec (that is, $\omega_0 = 1$ radian/sec, $\tau = 1$ sec).

The frequency response of the two-pole function is given in magnitude and phase by

$$|\psi(j\omega)| = \frac{1}{\sqrt{[1-(\omega/\omega_0)^2]^2 + [2\zeta(\omega/\omega_0)]^2}} \quad (13.34a)$$

and

$$\underline{/\psi(j\omega)} = \arctan\left[\frac{-2\zeta(\omega/\omega_0)}{1-(\omega/\omega_0)^2}\right]. \quad (13.34b)$$

These functions are plotted against frequency normalized to ω_0 in Figs. 13.8*b* and *c*. The cutoff frequency ω_{co} decreases with increase in ζ, but

the relation between them is complicated. A point of some significance is that

$$\omega_{co} = \omega_0 \tag{13.35}$$

when the poles lie at 45° in the left half-plane, that is, when the response is maximally flat.

13.2.1.7 Two-Pole, One-Zero Function

The two-pole one-zero transfer function is of some importance because it results from a particularly useful inductive peaking circuit. In the general case the function $\psi(s)$ is

$$\psi(s) = \frac{1 + s/\sigma}{1 + 2\zeta s/\omega_0 + s^2/\omega_0^2}, \tag{13.36}$$

where ζ and ω_0 are the damping ratio and undamped natural frequency of the poles, and $-\sigma$ is the position of the zero on the negative real axis. There is complete freedom in positioning the singularities.

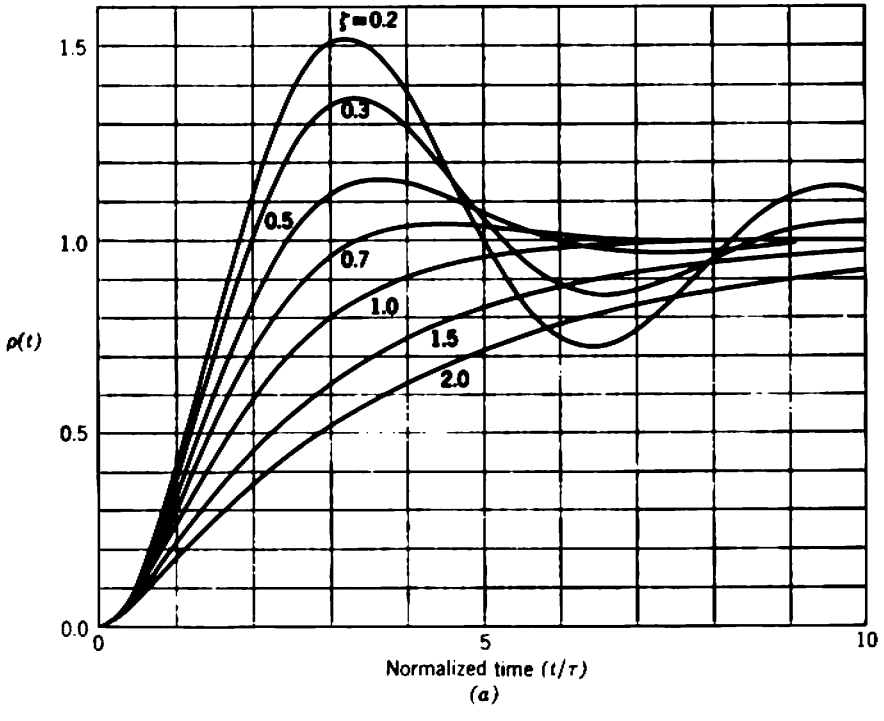
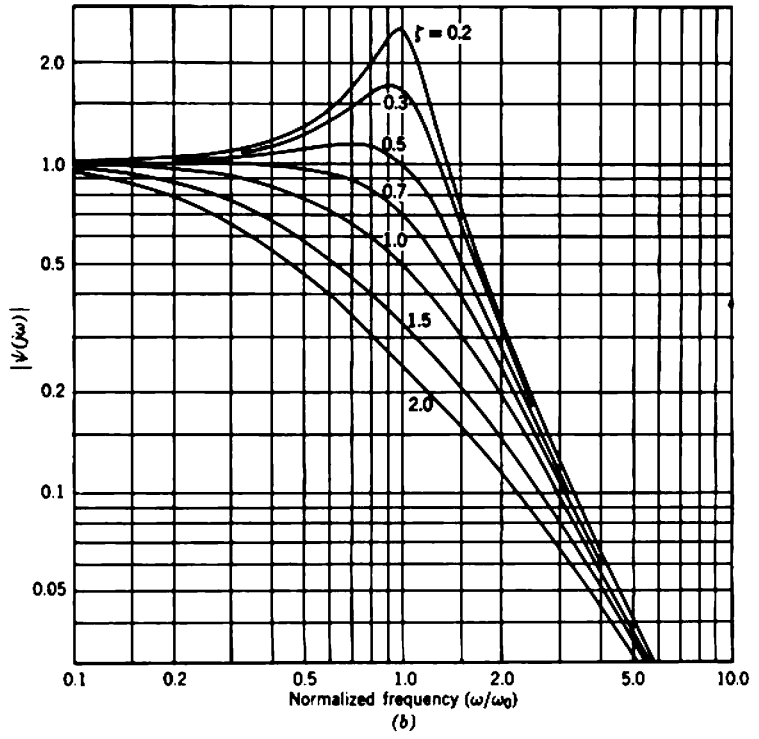
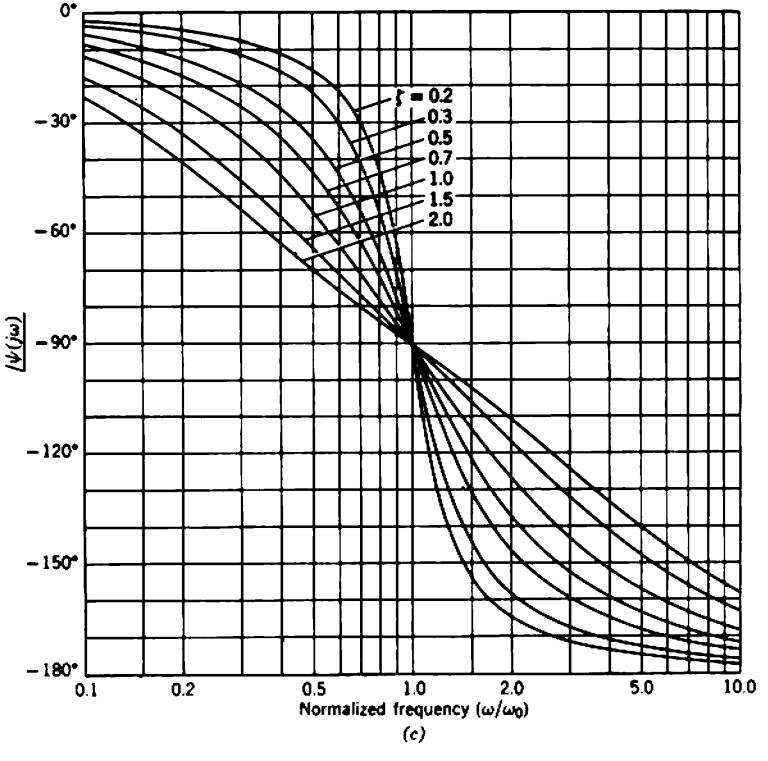


Fig. 13.8 The 2-pole function: (*a*) step response; (*b*) magnitude of frequency response; (*c*) phase of frequency response.



(b)



(c)

Fig. 13.8 *(continued)*

In the special case of interest there is a relation between σ, ζ, and ω_0 such that $\psi(s)$ can be written as

$$\psi(s) = \frac{1 + sm\tau}{1 + s\tau + s^2m\tau^2} \tag{13.37}$$

and the geometric mean of the distances of the singularities from the origin is $1/\tau$. The step-function and frequency response normalized to τ can be calculated as in the preceding two-pole case and are completely determined by the parameter m. Figure 13.9 shows a family of response curves, and Table 13.5 gives exact data for a few special cases.

Table 13.5 Two-Pole, One-Zero Function

m	Cutoff Frequency ω_{co}	Rise Time t_r (sec)	Overshoot (%)	Remarks
0.0	1.000	2.20	0.0	Single-pole function
0.25	1.414	1.54	0.0	Coincident poles
0.322	1.573	1.35	0.8	Linear-phase
0.414	1.722	1.21	3.1	Maximally-flat-magnitude
0.5	1.800	1.14	6.7	$\lvert\psi(j\omega)\rvert = 1$ at $\omega = 1$

These results are normalized so that the geometric mean distance of the effective poles from the origin is 1 radian/sec (that is, $\tau = 1$ sec).

0.322 is a solution of $m^3 + 3m - 1 = 0$,

$0.414 = \sqrt{2} - 1$.

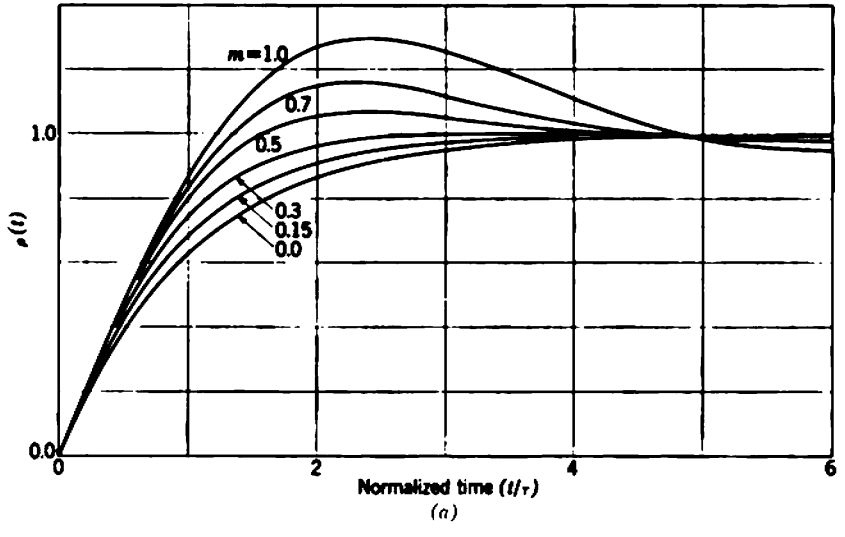
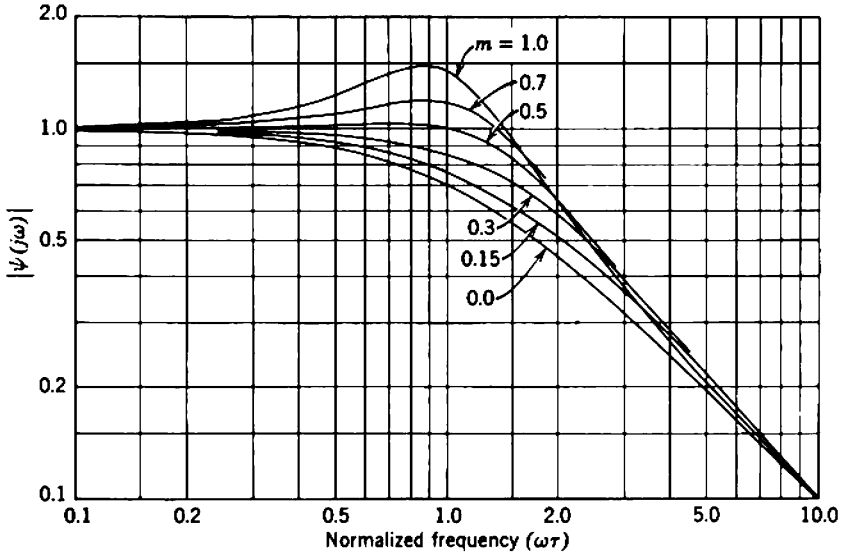
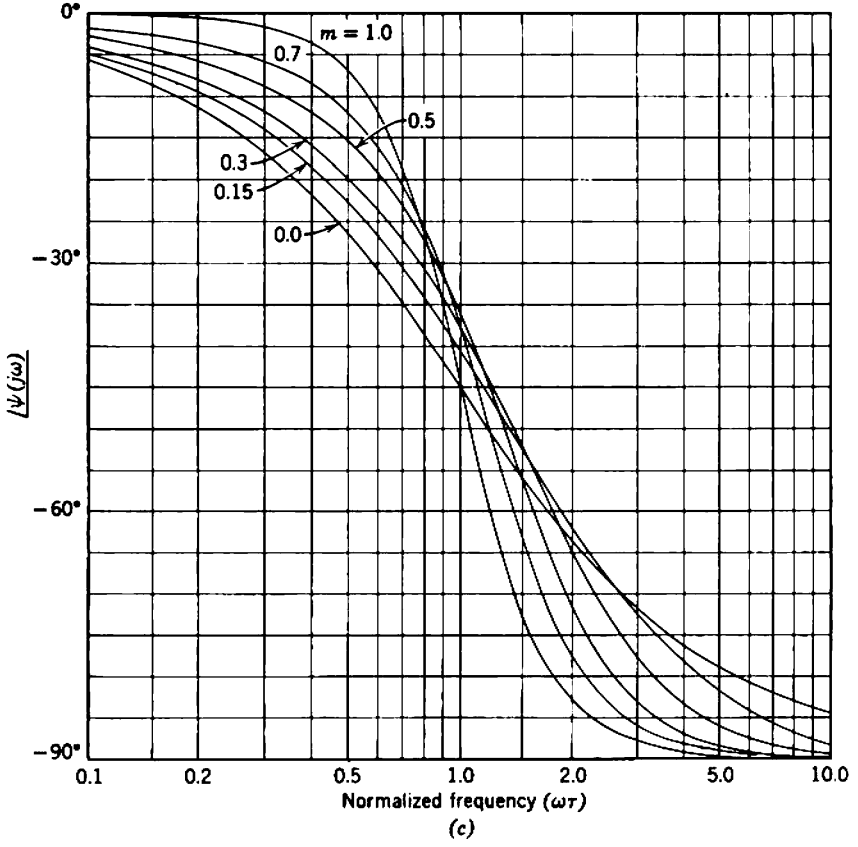


Fig. 13.9 A special 2-pole, 1-zero function, defined by Eq. 13.37: (*a*) step response; (*b*) magnitude of frequency response; (*c*) phase of frequency response.



(b)



(c)

Fig. 13.9 *(continued)*

13.2.1.8 Empirical Rules

When an amplifier's singularity pattern does not fall into any of the above categories, the following four empirical rules may be of assistance in estimating its dynamic response. These rules fail when the overshoot exceeds about 10%, but this is not a real drawback because amplifiers with more than 10% overshoot are of limited practical application. Empirical rules are often regarded with suspicion, so it is worth pointing out that these rules have some mathematical basis.*

The rules are given in terms of the number of building-blocks in an amplifier. In simple cases each building-block is just one stage—the combination of one active device and its associated passive components. More generally, a building-block is any group of active and passive components that has a known singularity pattern and therefore whose dynamic response is in principle known. Moreover, because singularities are commutative, a building-block can be made up from widely separated parts of the complete amplifier.

1. As the number of building-blocks in an amplifier is increased, the rise time increases. Based on Eq. 13.8 which applies for identical stages, the rise time of a number of nonidentical building-blocks is

$$t_r \approx \sqrt{t_{r1}^2 + t_{r2}^2 + t_{r3}^2 + \cdots}. \tag{13.38a}$$

Equation 13.38*a* holds very well for building-blocks having individual overshoots less than about 5%, but see 3 and 4 below.

2. Following Eq. 13.13 which applies for a number of identical stages, the product of the cutoff frequency and rise time of any amplifier is approximately constant:

$$\omega_{co} t_r = 1.8 \text{ to } 2.7, \tag{13.39a}$$

$$f_{co} t_r = 0.3 \text{ to } 0.45. \tag{13.39b}$$

For amplifiers having only 1 or 2% overshoot, the constants are 1.8 and 0.3. As the overshoot increases, the constants increase and become 2.7 and 0.45 at about 10% overshoot. Beyond 10% overshoot their values are unpredictable. In combination, Eqs. 13.38*a* and 13.39*a* give the cutoff frequency of number of building-blocks as

$$\frac{1}{\omega_{co}} \approx \left(\frac{1}{\omega_{co1}^2} + \frac{1}{\omega_{co2}^2} + \frac{1}{\omega_{co3}^2} + \cdots\right)^{1/2}. \tag{13.38b}$$

3. For an amplifier consisting of a number of building-blocks, each of which has a very small overshoot (1 or 2%), the total overshoot increases

* A number of references are listed in J. M. PETTIT and M. M. MCWHORTER, *Electronic Amplifier Circuits*, Chapter 4 (McGraw-Hill, New York, 1961).

very slowly or even not at all as the number of building-blocks is increased. The rise time increases in accordance with Eq. 13.38.

4. For an amplifier consisting of a number of similar building-blocks each of which has 5 to 10% overshoot, the total overshoot of n building-blocks is

$$\eta(n) \approx \eta(1)\sqrt{n}. \tag{13.40}$$

The rise time increases substantially less rapidly than the overshoot; that is,

$$t_r(n) < t_r(1)\sqrt{n}. \tag{13.41}$$

13.2.2 Tabular and Graphical Sets of Data

The seven types of transfer function discussed in the preceding section are all amenable to some form of analytical treatment; one- and two-pole functions can be analyzed fairly generally because of their simplicity, whereas all-pole functions allow certain types of general analysis because of the special form of their polynomials. These functions are useful to the engineer, not because they are necessarily the optimum functions for any application, but rather because they serve as points of reference for other functions which are possibly more useful but not so amenable to analysis. If, for example, the third-order maximally-flat and third-order linear-phase functions are taken as points of reference, their normalized rise time and percentage overshoot are:

Function	Rise Time	Percentage Overshoot
Maximally-flat	2.29	8.15
Linear-phase	3.07	0.75

In an application calling for less than 5% overshoot, it would be worth investigating the performance of three-pole functions in which the singularities lie somewhere between the known positions for maximal flatness and linear phase. This could lead to a significant decrease in rise time from that of the linear-phase function before reaching the excessive overshoot of the maximally-flat function. A familiarity with a few reference functions such as those listed in the preceding section is almost essential as a starting point for this type of more general investigation of the relation between singularity pattern and dynamic response. Below are listed a few sources of numerical information for some of the better-known reference functions and useful derived functions. The list is not

intended to be a complete bibliography; the reader may well find additional useful references.

1. ALL-POLE REFERENCE FUNCTIONS. A number of plots of step and impulse response of maximally-flat, linear-phase, and equal-ripple functions are given in K. W. HENDERSON and W. H. KATZ, "Transient response of conventional filters," *Trans. Inst. Radio Engrs.*, **CT-5**, 333, December 1958.

2. PERTURBATIONS ABOUT THE TWO-POLE FUNCTION. The following two references indicate the perturbation produced in a two-pole system by the addition of a third singularity—a zero in the first case and a pole in the second: R. N. CLARK, *Introduction to Automatic Control Systems*, Chapter 4 (Wiley, New York, 1962). P. CLEMENT, "A note on third-order linear systems," *Trans. Inst. Radio Engrs.*, **AC-5**, 151, June 1960.

3. DERIVED ALL-POLE FUNCTIONS. Plots and tabulations of the step and frequency response of a family of all-pole functions lying between the maximally-flat and linear-phase functions are given in Y. PELESS and T. MURAKAMI, "Analysis and synthesis of transitional Butterworth-Thompson filters and band-pass amplifiers," *RCA Rev.*, **18**, 60, March 1957.

13.2.3 Measurements on Analogs

Information on the dynamic response of an amplifier can be obtained by direct measurement on an analog. The use of analogs has the two-fold advantage of simplicity and flexibility; the simplicity stems from the ease of performing direct measurements, whereas the flexibility stems from the ease of changing systems or system parameters. Of the various types of analog available, the electronic analog computer is the most useful tool for this type of investigation and there must be very few workers in the electronics field who do not have access to such equipment. Computers should be regarded as an aid to thought and not a substitute for thought. At first sight it might appear that the use of analog computers could reduce design to a trial-and-error process which is carried out on a different time scale to ease the problems of instrumentation. If this were true, the use of computers would probably only result in bad designs being produced more quickly. In actual fact this sort of situation cannot (or at least should not) occur. The common language for design problems and analog computers is the transfer function (or correspondingly the integro-differential equation), so a problem can be transferred to the computer only if the transfer function is known—that is, if the design problem itself is understood. It may be rewarding to use an analog computer for the following types of problem:

1. To investigate the performance of complicated pole-zero patterns. Very often this amounts to determining the perturbations due to non-dominant singularities from the performance expected of a simplified system.

2. To investigate the performance of pole-zero patterns that have unusual geometry. Occasionally, unusual types of source or load contribute singularities that cause the pole-zero patterns to depart from familiar types.

3. To investigate the effect of small changes in singularity position on performance. The positions of some singularities vary as device or circuit parameters change. Often worst-case calculations of the effects of these changes become most tedious and an analog solution is much simpler.

Newcomers are more likely to benefit from the use of analogs than experienced designers, and should gain confidence and insight by performing analog computations. In short, analogs may be regarded as teaching aids for both students and practicing engineers.

13.2.3.1 Electronic Analog Computers

An electronic analog computer consists essentially of a combination of linear amplifiers and passive network elements that are interconnected in such a way that the transfer function of the analog is related to the transfer function of the system being investigated by one or more scale factors. A complete discussion of analog computers is beyond the scope of this book, but the following outline provides sufficient background for tackling the linear problems associated with amplifier design.

The basic unit of an analog computer is the *operational amplifier* shown in Fig. 13.10*a*. It can be shown that if

$$-Z_T(s) \gg Z_2(s)$$

and

$$-A_V(s) \gg \frac{Z_2(s)}{Z_1(s)},$$

the transfer function of the operational amplifier is

$$\frac{v_o(s)}{v_i(s)} \approx -\frac{Z_2(s)}{Z_1(s)}. \tag{13.42}$$

The accuracy of Eq. 13.42 depends on the extent to which the above inequalities are satisfied. Most modern computers achieve accuracies better than 1% and, for many purposes, Eq. 13.42 may be assumed exact.

The functional relation between v_o and v_i can be controlled by appropriate choice of the passive elements Z_1 and Z_2. In the general case shown in Fig. 13.10*b*, both Z_1 and Z_2 are parallel combinations of

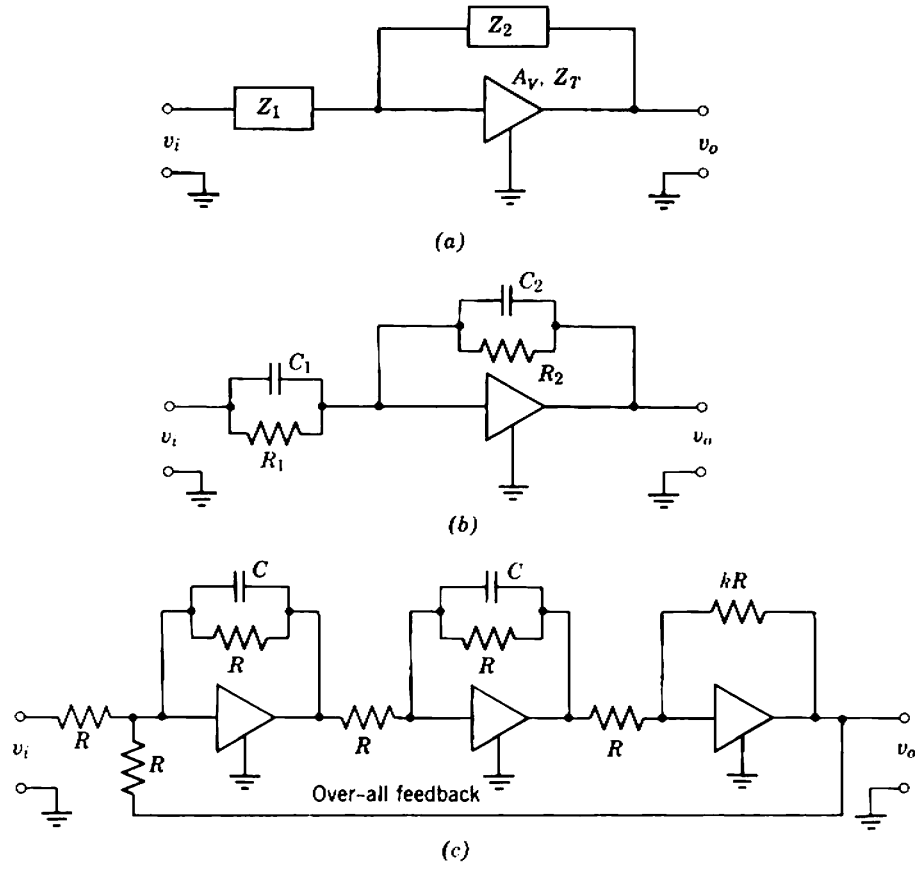


Fig. 13.10 The electronic analogue computer: (*a*) the basic operational amplifier; (*b*) 1-pole, 1-zero building-block; (*c*) 2-complex-pole building block.

resistance and capacitance and the transfer function has one real pole and one real zero:

$$\frac{v_o(s)}{v_i(s)} = -\frac{R_2}{R_1}\left(\frac{1 + sR_1C_1}{1 + sR_2C_2}\right). \tag{13.43}$$

Pairs of complex poles can be produced by arrangements such as Fig. 13.10*c* in which three operational amplifiers are cascaded with an additional over-all feedback loop. The transfer function is

$$\frac{v_o(s)}{v_i(s)} = -\frac{k}{1 + k}\left[\frac{1}{(1 - s/p_1)(1 - s/p_2)}\right], \tag{13.44a}$$

where the poles p_1 and p_2 are given by

$$p_1 \text{ or } p_2 = \frac{1}{RC}(-1 \pm j\sqrt{k}). \tag{13.44b}$$

Any transfer function can be synthesized by cascading the appropriate operational amplifiers and scaling the time so that the pole frequencies fall within the computer's operating range.

13.2.3.2 Mechanical and Electrical Analogs

A number of mechanical and electrical systems can be used as analogs of the complex frequency plane. These are useful for determining the magnitude and phase of the response of a system to sinusoidal signals but give no indication of the transient response. All these analogs have the following mathematical basis:

In general, $\psi(s)$ can be written as

$$\psi(s) = |\psi(s)| \times \angle\psi(s).$$

If natural logarithms are taken of both sides, then

$$\ln [\psi(s)] = \ln |\psi(s)| + j\angle\psi(s).$$

Now, the function $\ln [\psi(s)]$ is analytic at all points except the singularities of $\psi(s)$ and therefore the Cauchy–Riemann conditions are satisfied:

$$\frac{\partial[\ln |\psi(s)|]}{\partial\sigma} = \frac{\partial[\angle\psi(s)]}{\partial\omega}, \tag{13.45a}$$

$$\frac{\partial[\ln |\psi(s)|]}{\partial\omega} = -\frac{\partial[\angle\psi(s)]}{\partial\sigma}. \tag{13.45b}$$

By taking second derivatives it follows that both $\ln |\psi(s)|$ and $\angle\psi(s)$ satisfy Laplace's equation:

$$\nabla^2[\ln |\psi(s)|] = \left(\frac{\partial^2}{\partial\sigma^2} + \frac{\partial^2}{\partial\omega^2}\right)[\ln |\psi(s)|] = 0, \tag{13.46a}$$

$$\nabla^2[\angle\psi(s)] = \left(\frac{\partial^2}{\partial\sigma^2} + \frac{\partial^2}{\partial\omega^2}\right)[\angle\psi(s)] = 0. \tag{13.46b}$$

Thus any two-dimensional system that satisfies Laplace's equation can be used as an analog.

It is left as an exercise* for the reader to show that the following systems may be used as analogs and that the amplitude and phase response at any point on the $j\omega$ axis (that is, at any sinusoidal frequency ω) is related to the analog quantities as in Table 13.6:

(i) An *elastic membrane analog* is a sheet of uniform tensioned material, with unit positive force applied at the poles and unit negative force applied at the zeros.

* An excellent discussion appears in J. M. PETTIT and M. M. MCWHORTER, *Electronic Amplifier Circuits*, Chapter 8 (McGraw-Hill, New York. 1961).

(ii) A *conduction analog* is a uniform sheet of conducting material or a tank of electrolyte, with unit current supplied at the poles and withdrawn at the zeros.

Strictly, these analogs should be infinite in extent, but adequate results are obtained from finite models if only the central portion is used—say, the central third in each direction.

Table 13.6 Analogs of the Complex Frequency Plane for Finding the Response to a Sinusoidal Signal ω

Response	Membrane Analog	Conduction Analog
$\ln \lvert\psi(j\omega)\rvert$	Displacement	Voltage
$\underline{/\psi(j\omega)}$	Integral from 0 to ω of the gradient in the σ direction	Integral from 0 to ω of the current density in the σ direction

13.3 CHOICE OF DEVICE TYPE AND QUIESCENT POINT

Section 7.5 shows that the voltage gain of a resistance-coupled amplifier stage is a single-poled function. The gain can be written in terms of the realizable gain-bandwidth product as

$$A_V(s) = A_{Vm}\left[\frac{1}{1 + s(|A_{Vm}|/\mathscr{GB})}\right]. \tag{13.47}$$

This section describes simple methods for realizing a large gain-bandwidth product, that is, simple methods for increasing the gain while retaining constant bandwidth, or for increasing the bandwidth while maintaining a constant gain.

The first step is to choose an active device that has a large intrinsic $\mathscr{GB}$. In terms of charge-control parameters, the device should have a small transit time τ_1 (Section 2.5.1) because

$$\text{intrinsic } \mathscr{GB} = \frac{1 + u_1}{\tau_1} = \frac{g_m}{c_1}.$$

Notice that $\mathscr{GB}$ is not changed by connecting a number of devices in parallel; the mutual conductances add, but the capacitances add also. Second, the device must be used to the best advantage, so that its realizable $\mathscr{GB}$ approaches its intrinsic $\mathscr{GB}$. This implies that

(i) the stray wiring capacitance external to the device should be kept to a minimum by suitable construction techniques,

(ii) the device should be operated at an optimum (or at least well chosen) quiescent point.

Here the discussions for a vacuum tube and transistor diverge.

13.3.1 Vacuum-Tube Circuits

The realizable $\mathscr{GB}$ of a tube is shown in Section 7.5.1.1 to be

$$\mathscr{GB} = \frac{g_m}{C_o + C_S + C_i}, \tag{13.48}$$

where C_i and C_o are the input and output capacitance of the tube, and C_S is the stray wiring capacitance. Pentodes are preferred to triodes in wide-band amplifiers that do not employ negative feedback, because their realizable $\mathscr{GB}$ tends to be considerably greater. Often this is not apparent from manufacturers' data. Manufacturers specify an "input" and "output" capacitance for a tube, c_{in} and c_{out}, together with its anode-to-grid capacitance. Although the "input" and "output" capacitances of a pentode may be larger than the corresponding capacitances of a triode that has the same g_m, the true input and output capacitances depend on the anode-to-grid capacitance through Miller effect (Section 7.5.2.2). The Miller input capacitance is very much larger for a triode than for a pentode, and more than outweighs the larger c_{in} and c_{out} of the pentode.

According to first-order theory (Section 3.4.2), the capacitances in the equivalent circuit of a vacuum tube are independent of the quiescent conditions, whereas g_m varies as $I_K^{1/3}$. Therefore $\mathscr{GB}$ is maximized by operating a tube at the largest cathode current consistent with its ratings. According to second-order theory, there is a slight advantage in reducing the anode voltage (or screen voltage if the tube is a pentode) because g_m at constant cathode current tends to increase. There is the further practical advantage that reducing the anode and screen voltages allows the cathode current to be increased without exceeding the tube's rated power dissipation. The process of reducing the anode and screen quiescent voltages can be carried too far, however. As these voltages are reduced, the magnitude of the negative grid bias must also be reduced to maintain or increase the cathode current. If the grid becomes less negative than about -1 V, grid current becomes appreciable and the tube develops a nonlinear input conductance. This nonlinear conductance reacts on the preceding stage and causes gross distortion. In the absence of specific data from a vacuum-tube manufacturer it is prudent to restrict the extreme positive excursion of the grid to 1 V negative with respect to the cathode.

Having decided on a device type and its quiescent point, we can vary the

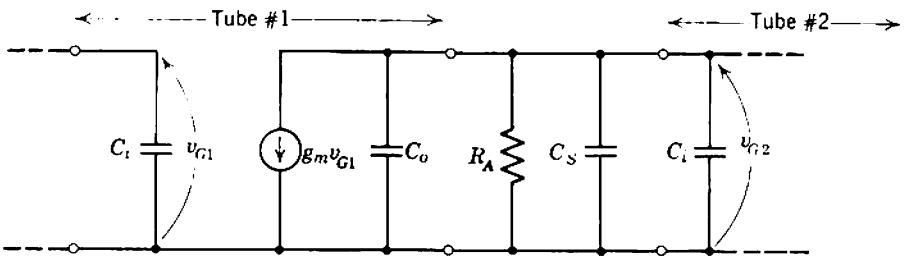


Fig. 13.11 Simplified equivalent circuit of a vacuum-tube interstage network at high frequencies.

load resistance R_L to exchange gain for bandwidth. In vacuum-tube circuits R_L is the parallel combination of the anode and grid supply resistors, R_A and R_G. Almost invariably R_G is chosen to be as large as possible in order to improve the low-frequency response (Eq. 7.33) and can be neglected except at very low frequencies. Typically, R_G is of the order of 1 MΩ, whereas R_A may range from a few hundred ohms to perhaps 50 kΩ, depending on the bandwidth requirements. The anode resistance r_A of a pentode is also of the order of 1 MΩ and it, like R_G, can be neglected. Figure 13.11 shows the simplified equivalent circuit; the mid-band gain is

$$A_{Vm} = -g_m R_A \tag{13.49}$$

and the frequency-dependent gain $A_V(s)$ is given by Eq. 13.47. Notice that the inequalities assumed in Section 7.4.3.1 are satisfied; the decoupling network can be used for compensating the low-frequency response.

13.3.2 Transistor Circuits

Transistor circuits are rather more complicated than vacuum-tube circuits in that the realizable gain-bandwidth product depends on the circuit configuration. Figure 13.12*a* shows the approximate mid-band equivalent circuit of a transistor to which the capacitances c_B, c_{tE}, and c_{tC} have been added. Continuing the discussion of Miller effect in Section 7.5.2, we can replace the bridging capacitance c_{tC} with two shunt capacitances (Fig. 13.12*b*). The components of input capacitance are combined in Fig. 13.12*c* and

$$c_C = c_B + c_{tE} + c_{tC}(1 - A_{Vm}). \tag{13.50}$$

Figure 13.12*c* is often termed the *unilateral equivalent circuit* for a transistor at high frequencies. Subject to the assumption that c_{tC} is much less than

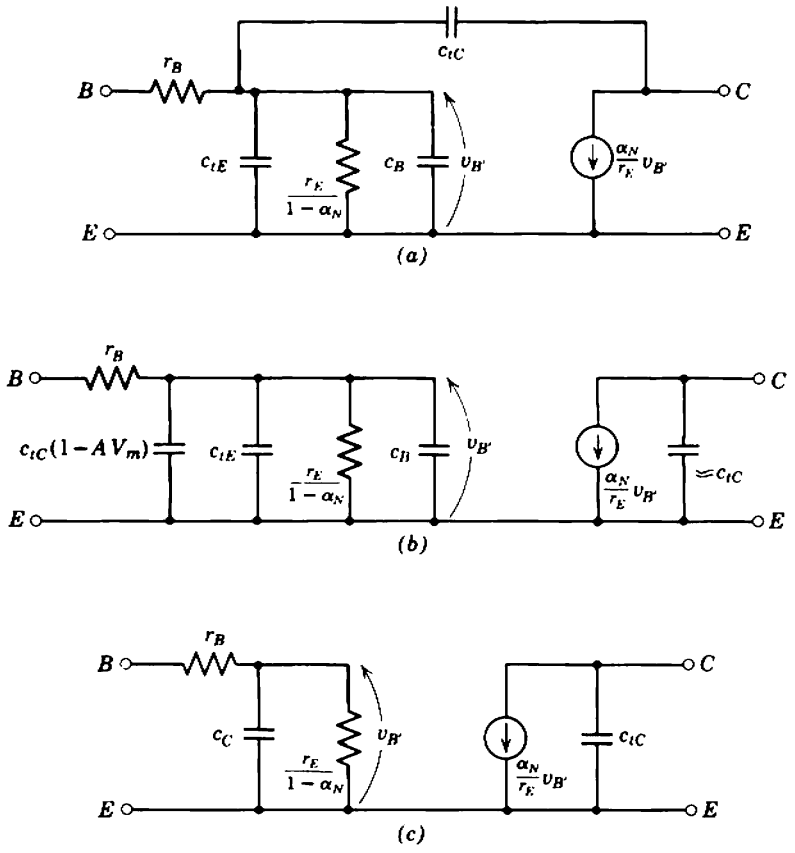


Fig. 13.12 High-frequency equivalent circuit of a transistor: (*a*) complete; (*b*) with shunt Miller capacitances; (*c*) unilateral equivalent circuit.

c_C, the realizable gain-bandwidth product for stages without feedback is shown in Section 7.5.1.2 to be

$$\mathscr{G}\mathscr{B} = \left(\frac{R_P}{R_P + r_B}\right)\omega_C, \tag{13.51}$$

where

$$\frac{1}{\omega_C} = \tau_C = \frac{r_E c_C}{\alpha_N} = \tau_1 + \frac{r_E}{\alpha_N}[c_{tE} + c_{tC}(1 - A_{Vm})]. \tag{13.52}$$

In comparison, Section 13.5.3.1 shows that the use of stages with alternate series and shunt feedback substantially eliminates the effects of r_B and Miller input capacitance. Accordingly, the realizable gain-bandwidth product is

$$\mathscr{G}\mathscr{B} = \omega_T, \tag{13.53}$$

where

$$\frac{1}{\omega_T} = \tau_T = \tau_1 + \frac{r_E c_{tE}}{\alpha_N}. \tag{13.54}$$

First-order theory (Section 4.6.2) shows that ω_T (and ω_C also, if A_{Vm} is constant) increase monotonically with emitter current because r_E is inversely proportional to I_E whereas τ_1, c_{tE}, and c_{tC} are constants. Second-order theory shows that τ_1 falls with increasing collector voltage, and also depends slightly on emitter current. There are therefore optimum emitter currents at which ω_T and ω_C are maximized. The transistors in wide-band amplifiers are normally operated near the peak of the $\mathscr{GB}$ versus I_E curve. In feedback amplifiers this corresponds to the peak of the ω_T versus I_E curve, but in nonfeedback amplifiers it does not correspond to the peak of ω_C. As I_E is increased, the biasing resistor R_P must be reduced to keep A_{Vm} constant because g_m is proportional to I_E. It follows from Eq. 13.51 that the realizable $\mathscr{GB}$ is maximum at a much smaller current than that at which ω_C is maximum. The following theory is more of academic interest than practical importance because the realizable $\mathscr{GB}$ of a feedback transistor amplifier is so much greater than $\mathscr{GB}$ of a non-feedback amplifier that the majority of wide-band transistor amplifiers employ feedback.

13.3.2.1 Optimum Emitter Current for Nonfeedback Stages

Figure 13.13*a* is the unilateral equivalent circuit for two stages in an amplifier. If the bandwidth is to be appreciably greater than (ω_C/β_N), R_P must be much smaller than $r_E/(1 - \alpha_N)$. In addition, c_{tC} of one transistor is much less than the c_C of the following transistor. Therefore the equivalent circuit can be simplified to Fig. 13.13*b*. This simplified

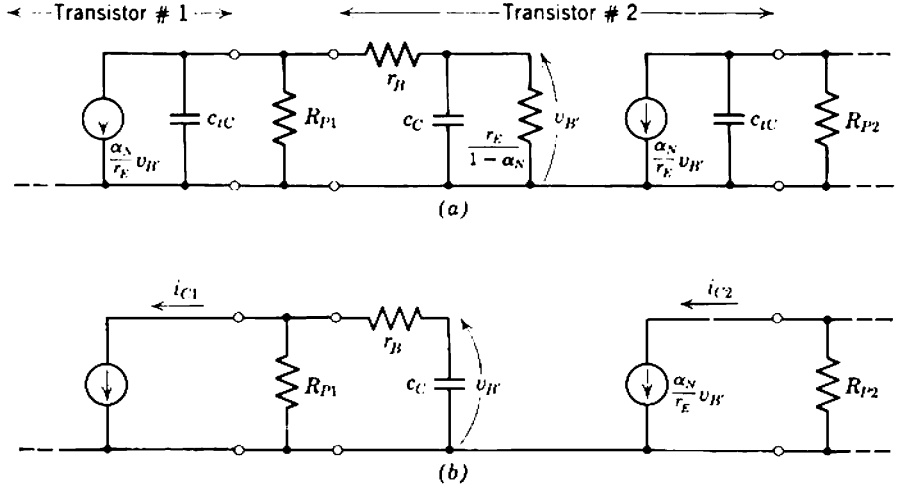


Fig. 13.13 Transistor interstage network at high frequencies: (*a*) unilateral equivalent circuit; (*b*) simplified for large bandwidths.

circuit is adequate for calculating the position of the circuit pole, but it yields a slightly high value for the mid-band gain. The problem is to find the optimum emitter current for the second transistor. This implies maximizing the current gain $|i_{C2}/i_{C1}|$ between the collectors of the two transistors, given by

$$A_I(s) = \frac{i_{C2}(s)}{i_{C1}(s)} = -\frac{\alpha_N R_{P1}}{r_E}\left[\frac{1}{1 + sc_C(R_{P1} + r_B)}\right]. \tag{13.55}$$

Suppose the stage is required to have a 3-dB cutoff frequency ω_{co}. It follows directly from Eq. 13.55 that

$$R_{P1} + r_B = \frac{1}{\omega_{co} c_C}. \tag{13.56}$$

Now, c_C is given by Eq. 13.50, and the voltage gain of the second transistor is

$$A_{Vm} = -\alpha_N\left(\frac{R_{P2}}{r_E}\right).$$

Substitution in Eq. 13.56 gives

$$R_{P1} = \frac{1}{\omega_{co}\{\alpha_N(\tau_1/r_E) + c_{tE} + c_{tC}[1 + \alpha_N(R_{P2}/r_E)]\}} - r_B. \tag{13.57}$$

With the exception of r_E, all parameters in Eq. 13.57 are independent of emitter current (to a first order). The optimum emitter current (or r_E) for given values of ω_{co}, τ_1, R_{P2}, c_{tE}, and c_{tC} is that which maximizes the mid-band current gain between the collectors; it is necessary to maximize $|A_{Im}|$, where

$$A_{Im} = -\alpha_N\left(\frac{R_{P1}}{r_E}\right). \tag{13.58}$$

Substituting Eq. 13.57 into 13.58 gives

$$|A_{Im}| = \frac{1}{\omega_{co}\{\tau_1 + (r_E/\alpha_N)c_{tE} + c_{tC}[(r_E/\alpha_N) + R_{P2}]\}} - \frac{r_B}{r_E/\alpha_N},$$

and differentiation gives the optimum value of r_E as

$$\frac{r_{E(\text{opt})}}{\alpha_N} = \frac{\tau_1 + c_{tC}R_{P2}}{(c_{tE} + c_{tC})\{[\omega_{co}r_B(c_{tE} + c_{tC})]^{-1/2} - 1\}}. \tag{13.59}$$

This expression for r_E can be substituted into Eqs. 13.58 and 13.57, respectively, to yield general explicit expressions for A_{Im} and the optimum value of R_{P1}, but it is simpler to carry out a numerical evaluation in each case.

In practice, R_P is made up of two resistors in parallel—the collector supply resistor R_C for the preceding stage and the base supply resistor R_B. By analogy with the vacuum-tube circuit considered above, R_B is chosen as large as possible so that $R_P \approx R_C$ and the decoupling network can be used for compensating the low-frequency response. The approximation of neglecting R_B in comparison with R_C, however, is not often so good as neglecting R_G in a vacuum-tube circuit in comparison with R_A. Hence R_P rather than R_C is used throughout the analysis.

13.4 INDUCTANCE PEAKING

Inductors can be connected into the load network of a resistance-coupled amplifier to increase the bandwidth or reduce the rise time. These inductors introduce additional poles, and may introduce zeros; a number of different circuit configurations are in use, including some in which there is mutual inductance between the inductors. Unfortunately, all but the simplest (and least efficient) are not designable because their characteristic equations are of high order so that the poles and zeros cannot be found readily. Modern devices, however, have such large gain-bandwidth products that adequate performance can usually be obtained with the simpler peaking networks. Except perhaps for amplifiers that are to be produced in quantity, the labor involved in designing a complicated peaking network is out of all proportion to the improvement realized; many high-grade commercially available amplifiers use only simple peaking circuits. The reader is referred to the literature for tables of the response of the more complicated peaking networks.*

13.4.1 Vacuum-Tube Circuits

Figure 13.14 is the high-frequency equivalent circuit for the junction between two vacuum-tube stages, including two peaking inductors. The shunt capacitance is split into two components: C_1 represents the sum of the output capacitance of the first tube and the stray capacitance to the left of L_2, whereas C_2 represents the sum of the input capacitance of the second tube and the stray capacitance to the right of L_2. Qualitatively, peaking inductor L_1 forms a parallel resonant circuit with R_A and C_1, and the resulting resonant peak tends to lift the response at high frequencies.

* G. E. VALLEY and H. WALLMAN, *Vacuum-Tube Amplifiers*, Chapter 2 (McGraw-Hill, New York, 1948). Also F. A. MULLER, "High-frequency compensation of *RC* amplifiers," *Proc. Inst. Radio Engrs.*, **42**, 1271, August 1954.

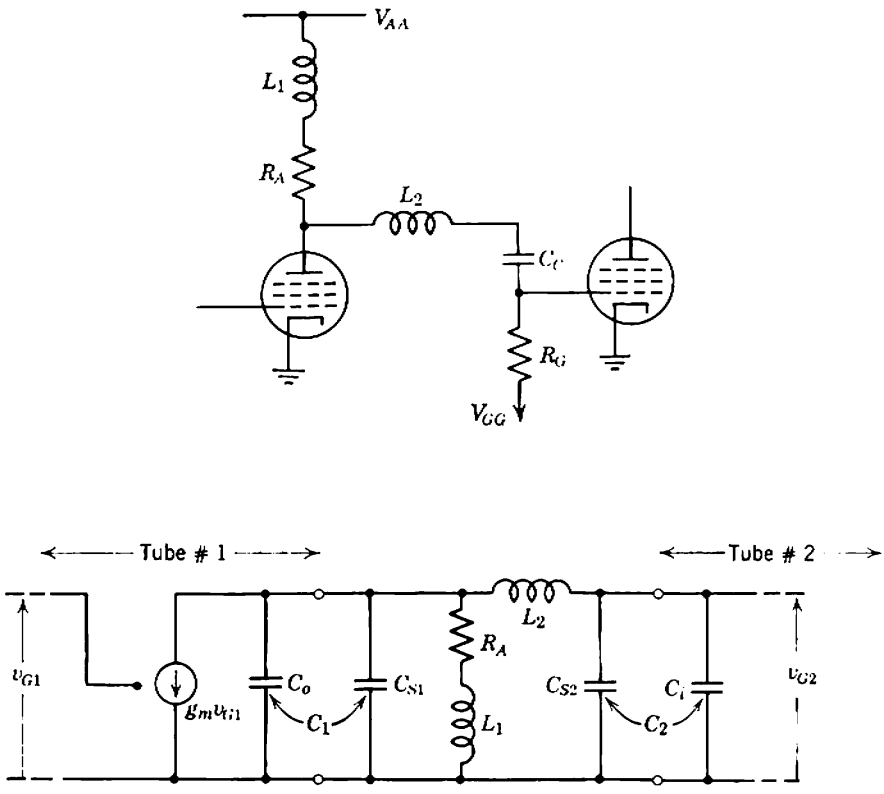


Fig. 13.14 Vacuum-tube interstage network with peaking inductors. Many practical circuits include additional resistors or capacitors in shunt with L_1 and L_2.

In addition, peaking inductor L_2 forms a series resonant circuit with C_2 and there is a peak in the ratio v_{G2}/v_{A1}. The voltage gain is

$$A_V(s) = \frac{v_{G2}(s)}{v_{G1}(s)} = -g_m R_A \psi(s), \tag{13.60}$$

where

$$\psi(s) = \frac{1 + s(L_1/R_A)}{1 + s[R_A(C_1 + C_2)] + s^2[L_1(C_1 + C_2) + L_2C_2] + s^3[R_A L_2 C_1 C_2] + s^4[L_1 L_2 C_1 C_2]}, \tag{13.61}$$

and in the general case there are four poles and one zero. Although it is impracticable to find the positions of the poles, a number of solutions for the transient response have been obtained for specific ratios of C_1 to C_2

and are available in the literature.* Two special cases are of considerable practical importance.

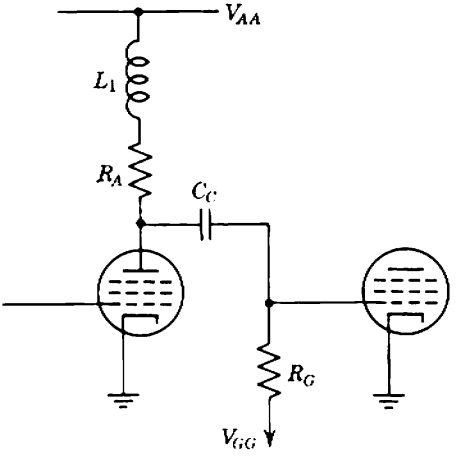
13.4.1.1 Shunt Peaking

If the series inductor L_2 is reduced to zero, the circuit reduces to the well-known shunt-peaked circuit shown in Fig. 13.15. Two of the poles vanish, and the frequency dependent term becomes

$$\psi(s) = \frac{1 + s(L_1/R_A)}{1 + s[R_A(C_o + C_S + C_i)] + s^2[L_1(C_o + C_S + C_i)]},$$

which is of the form

$$\psi(s) = \frac{1 + s(m\tau)}{1 + s(\tau) + s^2(m\tau^2)}; \tag{13.62}$$



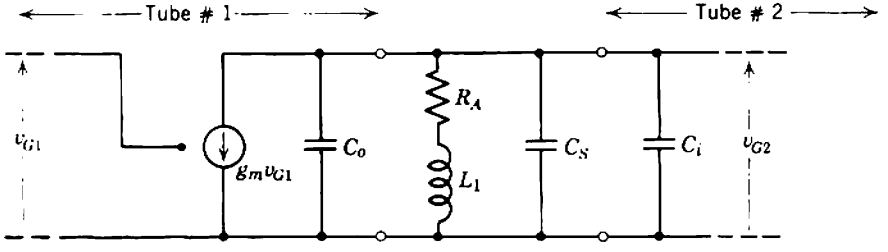


Fig. 13.15 Shunt-peaked vacuum-tube interstage network.

* For example, G. E. VALLEY and H. WALLMAN, *loc. cit.*

τ is the time constant of the pole which results from removing L_1 (i.e., if the circuit is not peaked),

$$\tau = R_A(C_o + C_S + C_i), \tag{13.63}$$

and

$$m = \frac{L_1}{R_A^2(C_o + C_S + C_i)}. \tag{13.64}$$

This is the form of the special two-pole, one-zero function considered in Section 13.2.1.7. Shunt peaking is by far the most common method for peaking a vacuum-tube amplifier.

13.4.1.2 Special Case $C_1 \ll C_2$

This special case occurs sometimes when a vacuum tube is coupled to an external load. If C_1 can be neglected entirely, the frequency-dependent term reduces to

$$\psi(s) = \frac{1 + s(L_1/R_A)}{1 + s[R_AC_2] + s^2[(L_1 + L_2)C_2]}. \tag{13.65}$$

The positions of all singularities can be controlled independently. In the usual practical case the poles are a complex conjugate pair; they lie on a circle centered at the origin with radius

$$|s| = \left[\frac{1}{(L_1 + L_2)C_2}\right]^{1/2}, \tag{13.66}$$

and their real part is

$$\sigma = -\frac{R_A}{2(L_1 + L_2)}. \tag{13.67}$$

There is a zero on the negative real axis at

$$s = -\frac{R_A}{L_1}, \tag{13.68}$$

which vanishes if L_1 is reduced to zero. The poles lie on the negative real axis if

$$(L_1 + L_2) \leq \tfrac{1}{4}R_A^2C_2$$

and their positions can be found from Eq. 13.65.

13.4.1.3 Accuracy of Inductance Peaking

The gain of an inductively peaked amplifier depends in part on the parameters of its vacuum tubes, and these have appreciable production tolerances. First, the spread in the mid-band gain is as great as it can possibly be for given tolerances on the tube parameters. The gain is directly proportional to g_m and virtually independent of r_A, so there is none

of the stabilizing effect that occurs when tubes are operated with load resistances significantly larger than their anode resistances (Section 5.2.1). Against this, the positions of the singularities are quite accurately reproducible. The positions of the zeros are determined entirely by passive circuit elements—R_A and the peaking inductors—and are therefore predictable to any accuracy required. The positions of the poles are determined by the passive elements and the stray capacitances. The intrinsic capacitances associated with charge-control theory have quite large tolerances (perhaps from 70 to 140% of their nominal values) but are almost completely masked by the extrinsic capacitances. Both the extrinsic capacitances within the tubes and interstage wiring capacitances depend only on geometry and, although they cannot be calculated, they are reproducible and can be measured. If the capacitances are measured in a prototype model of an amplifier and the appropriate peaking inductors are calculated, the peaking should hold to within $\pm 10\%$ if the design is produced in quantity and random tubes are used. There is certainly no cause for providing adjustments to all peaking inductors, if for no other reason than that it would be intolerably difficult to adjust them. Some manufacturers provide no adjustments at all; others provide one adjustment for every two or three stages to smooth out any small irregularities in the response.

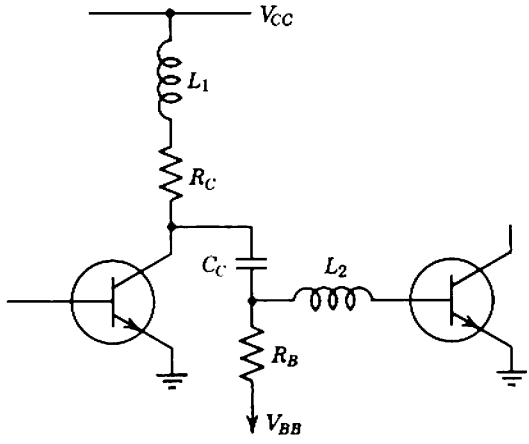
13.4.2 Transistor Circuits

Inductive peaking is little used in transistor amplifiers because the results are far inferior to those obtainable with feedback peaking. The general analysis of a transistor interstage network with inductive peaking virtually reduces to the special case for vacuum tubes considered in Section 13.4.1.2, because the output capacitance of a transistor and the stray wiring capacitance are negligible in comparison to the input capacitance. Figure 13.16 shows the unilateral high-frequency equivalent circuit with two peaking inductors added; the voltage gain between the internal bases is

$$A_V(s) = \frac{v_{B'2}(s)}{v_{B'1}(s)} = -\frac{\alpha_N R_P}{r_E}\left\{\frac{1 + s(L_1/R_P)}{1 + s[(R_P + r_B)c_C] + s^2[(L_1 + L_2)c_C]}\right\}. \tag{13.69}$$

In the usual practical case a zero is on the negative real axis at

$$s = -\frac{R_P}{L_1}, \tag{13.70}$$



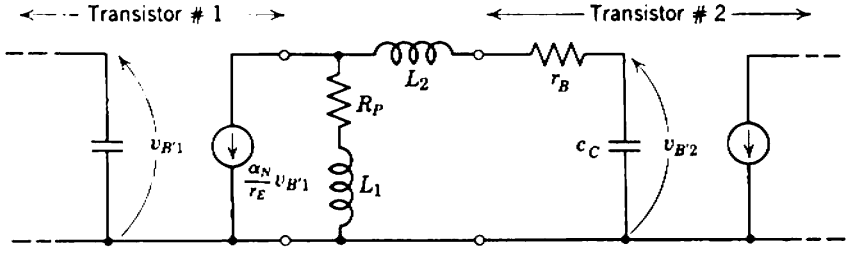


Fig. 13.16 Simplified transistor interstage network with peaking inductors.

and a pair of complex conjugate poles lies on a circle of radius

$$|s| = \left[\frac{1}{(L_1 + L_2)c_C}\right]^{1/2} \tag{13.71}$$

and their real part is

$$\sigma = -\frac{(R_P + r_B)}{2(L_1 + L_2)}. \tag{13.72}$$

Other less common configurations follow from Eq. 13.69. The zero vanishes if L_1 is reduced to zero.

An optimum emitter current exists for an inductively peaked transistor amplifier and can be deduced from the analysis of unpeaked amplifiers in Section 13.3.2. In this prior analysis the emitter current is derived in terms of the 3-dB cutoff frequency

$$\omega_{co} = \frac{1}{(R_P + r_B)c_C}.$$

For a peaked amplifier

$$\frac{1}{(R_P + r_B)c_C} = \frac{|s|^2}{-2\sigma} = \omega_{\text{eff}}. \tag{13.73}$$

Thus the optimum emitter current for a desired singularity pattern is found by replacing ω_{c0} in Eq. 13.59 by the effective frequency ω_{eff} determined from the positions of the poles.

13.5 FEEDBACK PEAKING

Application of negative feedback to an amplifier has a profound effect upon its singularity pattern. Section 10.5.3 describes the root-locus method of analysis, and shows that pairs of complex poles can be produced even though all open-loop singularities are real. In addition, closed-loop zeros can be produced on the negative real axis. Thus the singularity patterns attainable through the use of feedback are similar in nature to those attainable with inductance peaking. As a bonus the use of feedback for peaking stabilizes the mid-band gain. This section deals only with the high-frequency response of single-stage feedback amplifiers; Chapter 14 discusses feedback loops that enclose more than one stage.

13.5.1 Series Feedback

Figures 13.17*a* and *b* show the circuit diagram and complete equivalent circuit for the charge-control model with series feedback. As explained in Section 11.3.1, the conductances g_2 and g_f can be omitted at mid-band frequencies provided the load resistance is not too large. The capacitances c_f (which represents the sum of $-c_{12}$ in Fig. 2.7*f* plus any extrinsic capacitance) and c_{22} can ordinarily be omitted at high frequencies, although c_f is shown broken in the simplified circuit of Fig. 13.17*c*.

A high-frequency peaking capacitor C_F is shown connected across R_F. Qualitatively, this capacitor bypasses the feedback resistor at high frequencies and, if the capacitor is properly chosen, the associated rise in transfer admittance partially compensates for the normal high-frequency drop; the bandwidth is extended. The total feedback impedance is

$$Z_F(s) = \frac{1}{Y_F(s)} = \frac{R_F}{1 + sR_FC_F} = \frac{R_F}{1 + s\tau_F} \tag{13.74}$$

and network analysis shows that the transfer admittance of the stage (neglecting c_f) is

$$Y_T(s) = \frac{i_o}{v_i} = \frac{g_m Y_F}{g_m + g_1 + sc_1 + Y_F}. \tag{13.75}$$

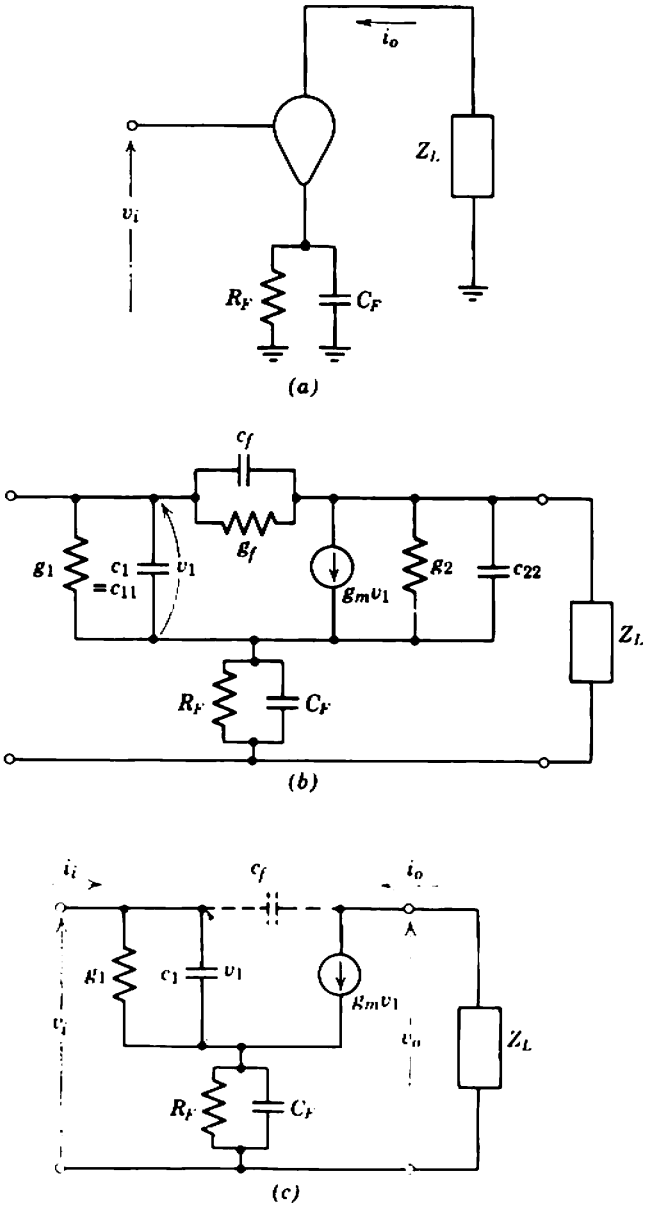


Fig. 13.17 General series-feedback stage with high-frequency peaking: (*a*) elemental circuit diagram; (*b*) complete equivalent circuit; (*c*) simplified equivalent circuit.

The input admittance is

$$Y_i(s) = Y_T(s)\left(\frac{g_1 + sc_1}{g_m}\right). \tag{13.76}$$

From the discussion of Miller effect in Section 7.5.2.1, the bridging capacitance c_f can be replaced by shunt capacitances across the input and output circuits. The Miller input capacitance, which appears in shunt with Y_i given by Eq. 13.76, is

$$c_M \approx c_f(1 - A_{Vm}) = c_f(1 + G_T R_L),$$

where R_L is the resistive component of the load impedance. Since the mid-band transfer conductance of a series-feedback stage is

$$G_T \approx \frac{1}{R_F},$$

it follows that the Miller input admittance is

$$[Y_i(s)]_{c_f} \approx sc_f\left(1 + \frac{R_L}{R_F}\right). \tag{13.77}$$

In a well-designed circuit $|A_{Vm}|$ is small because R_L is the small input resistance of the following shunt-feedback stage. Hence $c_M \approx c_f$, and the Miller component of Y_i is small compared with the main component.

The contribution of c_f to the load capacitance would ordinarily be approximated by

$$C_o \approx c_f\left(1 - \frac{1}{A_{Vm}}\right),$$

but this approximation is poor for a series-feedback stage because $|1/A_{Vm}|$ tends to be large. A more satisfactory approach is to consider the direct feedthrough of signal from input to output via c_f. Provided the voltage gain is small (that is, provided $|v_i| \gg |v_o|$) the component of current in Z_L due to c_f is

$$\left(\frac{i_o}{v_i}\right)_{c_f} \approx -sc_f.$$

In other words c_f contributes a term which adds directly to the transfer admittance given by Eq. 13.75:

$$[Y_T(s)]_{c_f} \approx -sc_f. \tag{13.78}$$

This Miller component of transfer admittance, like the Miller component of input admittance, can often be neglected.

13.5.1.1 Vacuum Tubes

Vacuum tubes are atypical of charge-controlled devices, in that the intrinsic input capacitance associated with the mobile charge is small compared with the extrinsic capacitance. In addition, the extrinsic capacitance is predominantly from grid to ground rather than from grid to cathode. The equivalent circuit of a series-feedback stage is shown in Fig. 13.18; to a good approximation

$$c_1 = 0$$

and the extrinsic capacitance c_{1e} is identified with the "input" capacitance c_{in} quoted by tube manufacturers (Section 7.5.2.2). This approximation

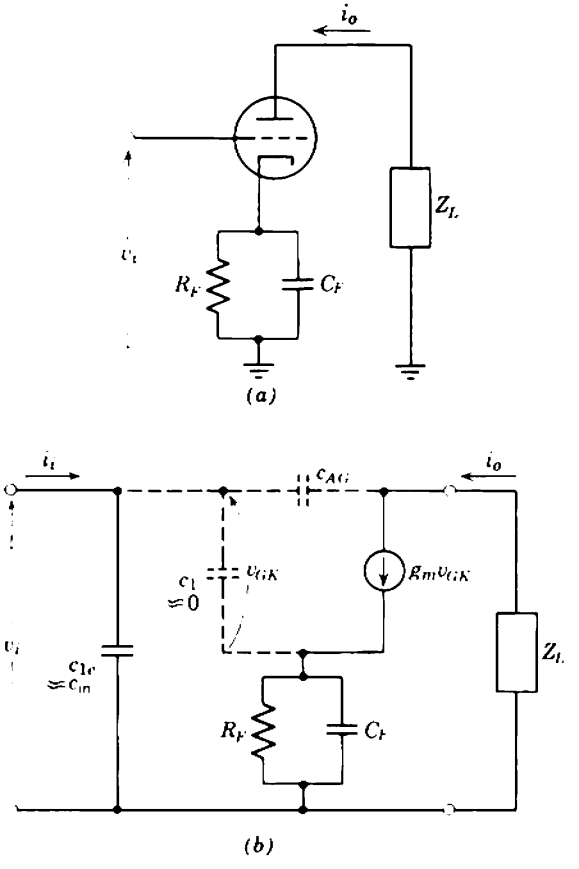


Fig. 13.18 Vacuum-tube series-feedback stage: (*a*) elemental circuit diagram; (*b*) equivalent circuit.

very greatly simplifies the analysis of a vacuum-tube series-feedback stage. The input admittance of the stage is purely capacitive:

$$C_i = c_{in}. \tag{13.79}$$

Substitution into Eq. 13.75 gives the transfer admittance of the stage as

$$Y_T(s) = \frac{g_m}{1 + g_m R_F} \left\{ \frac{1 + s\tau_F}{1 + s[\tau_F/(1 + g_m R_F)]} \right\}. \tag{13.80}$$

The first term is the mid-band transfer conductance

$$G_T = \frac{g_m}{1 + g_m R_F} \tag{13.81}$$

and the second is the ψ function. This ψ function is unity at low and mid-band frequencies but has a high-frequency pole and zero on the negative real axis. Because the pole lies further from the origin, $|\psi(j\omega)|$ rises with increasing frequency and ultimately becomes $(1 + g_m R_F)$; the transfer admittance increases to g_m, the value expected when the peaking capacitor C_F completely bypasses R_F. If the tube is a pentode with its screen bypassed to ground, R_F in Eqs. 13.80 and 13.81 should be replaced by $(1 + k)R_F$ (see Eq. 11.15).

The rising high-frequency response of a series-feedback stage is very useful for compensating the falling response of other stages in a multi-stage amplifier. A simple but effective design technique is to cancel a pole from some other stage with the zero from the series stage. The position of this zero is defined accurately by the passive elements R_F and C_F, but the pole of a series-feedback stage depends on g_m and has an appreciable tolerance.

The Miller components of Y_T and Y_i are given by Eqs. 13.78 and 13.77, with the anode-to-grid capacitance c_{AG} replacing c_f. These components are negligible for a pentode because c_{AG} is so small but may be significant for a triode. The Miller input capacitance $c_{AG}(1 + R_L/R_F)$ adds directly to C_i given by Eq. 13.79.

13.5.1.2 Transistors

A bipolar transistor is more complicated than the general charge-control model in that the internal base is isolated from the so-called base terminal by the extrinsic base resistance r_B. In the series-feedback stage whose circuit is shown in Fig. 13.19, r_B forms a voltage divider at the input of the transistor. The transfer admittance of the stage is therefore

$$Y_T(s) = \frac{[Y_T(s)]_i}{1 + r_B[Y_i(s)]_i}$$

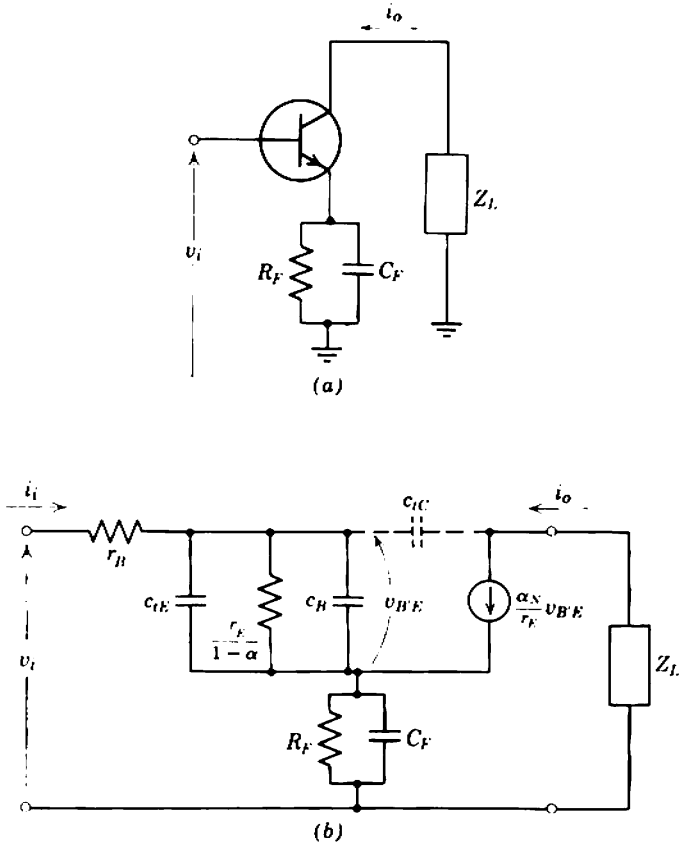


Fig. 13.19 Transistor series-feedback stage: (*a*) elemental circuit diagram; (*b*) equivalent circuit.

and the input impedance is

$$Z_i(s) = r_B + \frac{1}{[Y_i(s)]_i},$$

where $[Y_T(s)]_i$ and $[Y_i(s)]_i$ are parameters of the stage without r_B. Making the substitutions

$$g_m = \frac{\alpha_N}{r_E},$$

$$g_1 = \frac{1 - \alpha_N}{r_E},$$

$$(c_B + c_{tE})r_E = \alpha_N \tau_T,$$

the transfer admittance and input impedance follow from the general analysis as

$$Y_T(s) = \frac{\alpha_N}{r_B(1-\alpha_N)+r_E+R_F} \times \left\{ \frac{1+s\tau_F}{1+s\left[\tau_F+\frac{\alpha_N r_B\tau_T+R_F(\alpha_N\tau_T-\tau_F)}{r_B(1-\alpha_N)+r_E+R_F}\right]+s^2\left[\frac{\alpha_N r_B\tau_T\tau_F}{r_B(1-\alpha_N)+r_E+R_F}\right]} \right\}, \tag{13.82}$$

$$Z_i(s) = r_B + (\beta_N+1)\left(\frac{r_E}{1+s\alpha_N\tau_T}+\frac{R_F}{1+s\tau_F}\right)\left(\frac{1+s\tau_F}{1+s\beta_N\tau_T}\right). \tag{13.83}$$

In the general case Y_T has two poles and one zero whereas Z_i has two poles and two zeros.

A great simplification is achieved by designing the circuit with

$$\tau_F = R_F C_F = \alpha_N\tau_T. \tag{13.84}$$

Both Y_T and Z_i become single-poled functions:

$$Y_T(s) = \frac{\alpha_N}{r_B(1-\alpha_N)+r_E+R_F}\left\{\frac{1}{1+s\left[\frac{\alpha_N r_B\tau_T}{r_B(1-\alpha_N)+r_E+R_F}\right]}\right\}, \tag{13.85}$$

$$Z_i(s) = r_B + \frac{r_E+R_F}{1-\alpha_N}\left(\frac{1}{1+s\beta_N\tau_T}\right). \tag{13.86}$$

The first term in Eq. 13.85 is the mid-band transfer conductance

$$G_T = \frac{\alpha_N}{r_B(1-\alpha_N)+r_E+R_F} \approx \frac{\alpha_N}{r_B/\beta_N+r_E+R_F} \tag{13.87}$$

and the second term is the ψ function. Equation 13.85 can be rewritten as

$$Y_T(s) = G_T\left[\frac{1}{1+s(r_B G_T\tau_T)}\right]. \tag{13.88}$$

To a very good approximation the input impedance is a parallel combination of resistance and capacitance:

$$R_i = r_B + \frac{r_E+R_F}{1-\alpha_N} \approx r_B + \beta_N(r_E+R_F), \tag{13.89}$$

$$C_i = G_T\tau_T. \tag{13.90}$$

Series feedback stabilizes the transfer conductance G_T of a stage. Therefore, when τ_T has its nominal value, C_i of a series-feedback stage is also known exactly, whereas the pole of Y_T depends only on r_B. Typically, the tolerance on r_B is 70 to 140% of its nominal value, and this becomes the tolerance on the pole. If the peaking is adjusted for a nominal transistor but τ_T is off tolerance, both Y_T and Z_i develop additional poles and zeros. It can be verified from Eq. 13.82 that Y_T has one zero and two poles, but all lie close to the nominal pole position so that the net effect is very nearly equivalent to the one ideal pole. Similarly, from Eq. 13.83, the dominating component of input impedance at high frequencies is a capacitance close to the ideal value given by Eq. 13.90.

Because of the presence of r_B, Eqs. 13.78 and 13.77 are not so good approximations for the Miller components of Y_T and Y_i of a transistor as they are for the general charge-control model. Nevertheless, the approximations are adequate because the Miller terms are merely small corrections. The Miller input capacitance $c_{tC}(1 + R_L/R_F)$ adds directly to C_i given by Eq. 13.90.

13.5.2 Shunt Feedback

As with the low-frequency response (Section 12.2.3), one convenient approach to the high-frequency response of a shunt-feedback stage is to combine the coupling efficiency and stage transfer impedance in a system-loop-gain analysis. This analysis gives the poles and zeros and, while it can also give the mid-band transfer function magnitude, the simple approach described in Section 11.2.2.2 is more generally satisfactory; η and R_T should be evaluated separately.

Figures 13.20*a* and *b* show the circuit diagram and equivalent circuit for the charge-control model with shunt feedback. As explained in Section 11.3.2, the conductances g_2 and g_f can usually be omitted or, if not omitted, combined with the load and feedback resistances. The capacitances c_{22} and c_f can likewise be absorbed into the load and feedback impedances. Finally, the source impedance and input impedance of the device can be combined and the equivalent circuit simplified to Fig. 13.20*c*.

In the absence of feedback the transfer impedance from source current to load voltage has two poles on the negative real axis—one each due to C_1 and C_2. If frequency-independent feedback is applied, the poles come together and then separate and move parallel to the imaginary axis (see Section 10.5.3.3). Even for quite modest values of loop gain, the damping ratio of the poles is small and the transient response of the system is

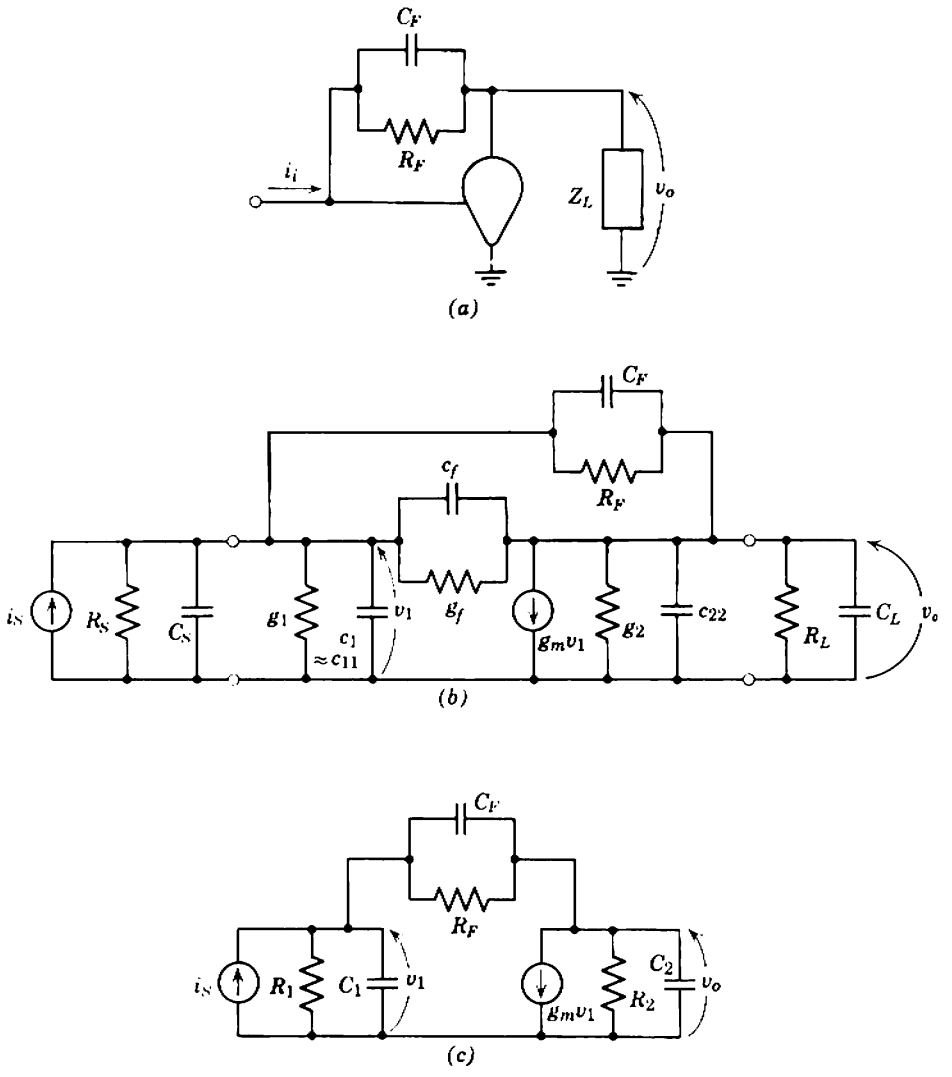


Fig. 13.20 General shunt-feedback stage with high-frequency peaking: (*a*) elemental circuit diagram; (*b*) complete equivalent circuit; (*c*) simplified equivalent circuit.

unacceptable. Therefore a capacitor C_F is connected in shunt with R_F in order to contribute a zero to the feedback factor; the closed-loop poles can be moved into useful positions by suitably positioning this zero, even at high values of loop gain. The feedback impedance is

$$Z_F(s) = \frac{R_F}{1 + sR_FC_F} = \frac{R_F}{1 + s\tau_F}. \tag{13.91}$$

Provided the mid-band loop gain is large, the system transfer function is

$$\frac{v_o(s)}{i_S(s)} = -\frac{R_F}{1 + \dfrac{(R_F + R_1)(R_F + R_2)}{g_m R_F R_1 R_2}} \times \left\{ \frac{1}{\begin{aligned} 1 + s\left[R_F C_F + \frac{(R_F + R_1)(C_F + C_2)}{g_m R_1} + \frac{(R_F + R_2)(C_F + C_1)}{g_m R_2}\right] \\ + s^2\left[\left(\frac{R_F}{g_m}\right)(C_F + C_1)(C_F + C_2)\right] \end{aligned}} \right\}. \tag{13.92}$$

The first term is the system mid-band transfer function and the second is the ψ function. As explained in connection with Fig. 11.17*b* (Section 11.3.2), the first term represents the combined effects of coupling efficiency and stage transfer resistance:

$$\frac{v_o}{i_S} = \eta \times R_T = -\frac{R_F}{1 + \dfrac{(R_F + R_1)(R_F + R_2)}{g_m R_F R_1 R_2}}. \tag{13.93}$$

The second term gives the combined singularity pattern. In order to study this term, it is convenient to make two simplifying approximations:

(i) $R_F C_F$ is the dominating term in the coefficient of s;
(ii) C_F is much smaller than either C_1 or C_2.

Both approximations are good for practical circuits. Hence

$$\psi(s) \approx \frac{1}{1 + s(R_F C_F) + s^2(R_F C_1 C_2 / g_m)}. \tag{13.94}$$

The two closed-loop poles lie on a circle of radius

$$|s| = \left(\frac{g_m}{R_F C_1 C_2}\right)^{1/2}, \tag{13.95}$$

and their real part is

$$\sigma = -\frac{g_m C_F}{2 C_1 C_2}. \tag{13.96}$$

For a given device the pole-circle radius is (to a first order) determined by R_F alone, whereas the position of the poles on the circle is determined by C_F alone. Figure 13.21 illustrates the behavior of the singularity pattern as C_F is increased. The poles become real when C_F is large; also, the mean distance of the poles from the origin (equal to $|s|$) falls if C_F becomes comparable with C_1 or C_2.

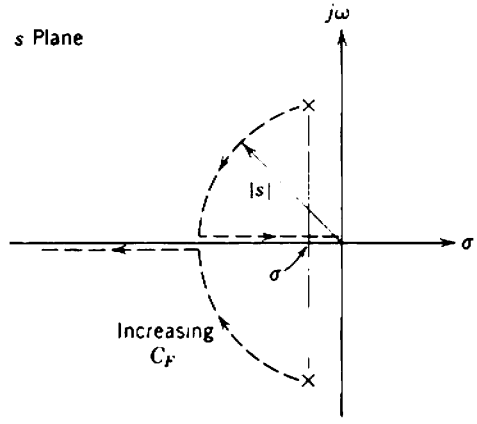


Fig. 13.21 Behavior of the poles of a shunt-feedback stage as C_F is increased.

13.5.2.1 Vacuum Tubes

Section 11.2.2.2 shows that the vacuum tube in a shunt-feedback stage should be a pentode. Hence r_A and c_{AG} can be neglected and the general equivalent circuit reduces to Fig. 13.22*b*. With the equivalences

$$C_1 = C_S + c_{in},$$

$$C_2 = c_{out} + C_L,$$

the ψ function for the stage becomes

$$\psi(s) = \frac{1}{1 + s\left[R_F C_F + \dfrac{(R_F + R_S)(C_F + C_2)}{g_m R_S} + \dfrac{(R_F + R_L)(C_F + C_1)}{g_m R_L}\right] + s^2\left[\left(\dfrac{R_F}{g_m}\right)(C_F + C_1)(C_F + C_2)\right]} \tag{13.97}$$

The poles lie on a circle of radius

$$|s| = \left[\frac{g_m}{R_F(C_F + C_1)(C_F + C_2)}\right]^{1/2}, \tag{13.98}$$

and their real part is

$$\sigma = -\frac{g_m C_F}{2(C_F + C_1)(C_F + C_2)} - \frac{R_F + R_S}{2R_F R_S(C_F + C_1)} - \frac{R_F + R_L}{2R_F R_L(C_F + C_2)} \tag{13.99}$$

Often some of the terms in these equations can be neglected; Eqs. 13.94 to 13.96 give approximate forms.

The stray capacitances in a vacuum-tube circuit depend largely on the geometry of its construction and are, therefore, reproducible with quite

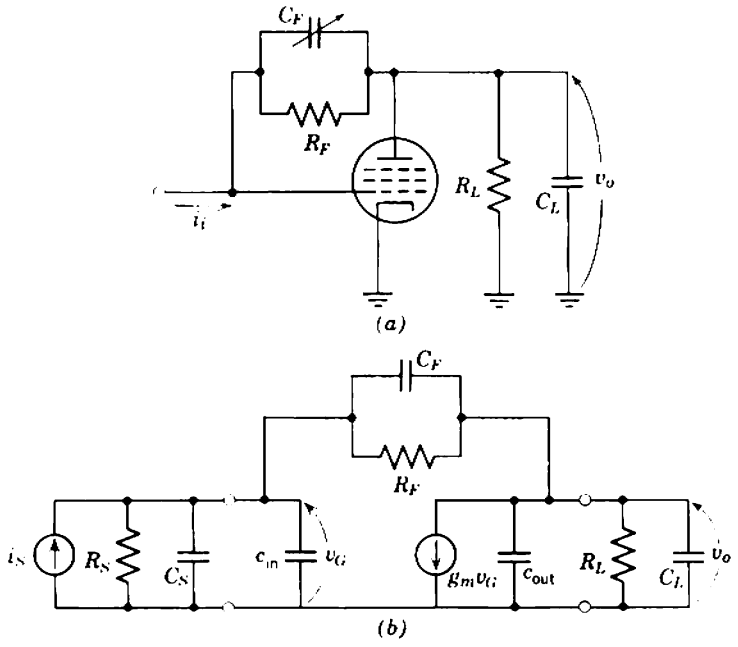


Fig. 13.22 Vacuum-pentode shunt-feedback stage: (*a*) elemental circuit diagram; (*b*) equivalent circuit.

small tolerances. Hence the mutual conductance g_m is the only uncertain term in Eqs. 13.97 to 13.99. Typically, the value of g_m for a tube lies between 60 and 150% of the nominal value. The pole circle radius $|s|$ can therefore vary by $\pm 20\%$ about its nominal value, whereas the real part σ can vary by $\pm 40\%$. This relatively large tolerance on σ results in a large tolerance on the damping ratio of the poles, with consequent large changes in the response. Usually the peaking capacitor C_F in a shunt-feedback stage is made adjustable so that the poles can be given the desired damping ratio.

13.5.2.2 Transistors

Figures 13.23*a* and *b* show the circuit diagram and equivalent circuit for a transistor shunt-feedback stage with peaking. As with many other transistor circuits, the base resistance r_B adds formal complication to the analysis but, unlike most other circuits, r_B (as well as several other elements) can be neglected:

1. For input currents r_B is in series with Z_S, whereas for feedback currents r_B is in series with Z_F. Therefore r_B can be neglected if both Z_S and Z_F are three or four times as large; these conditions are invariably satisfied in practical circuits.

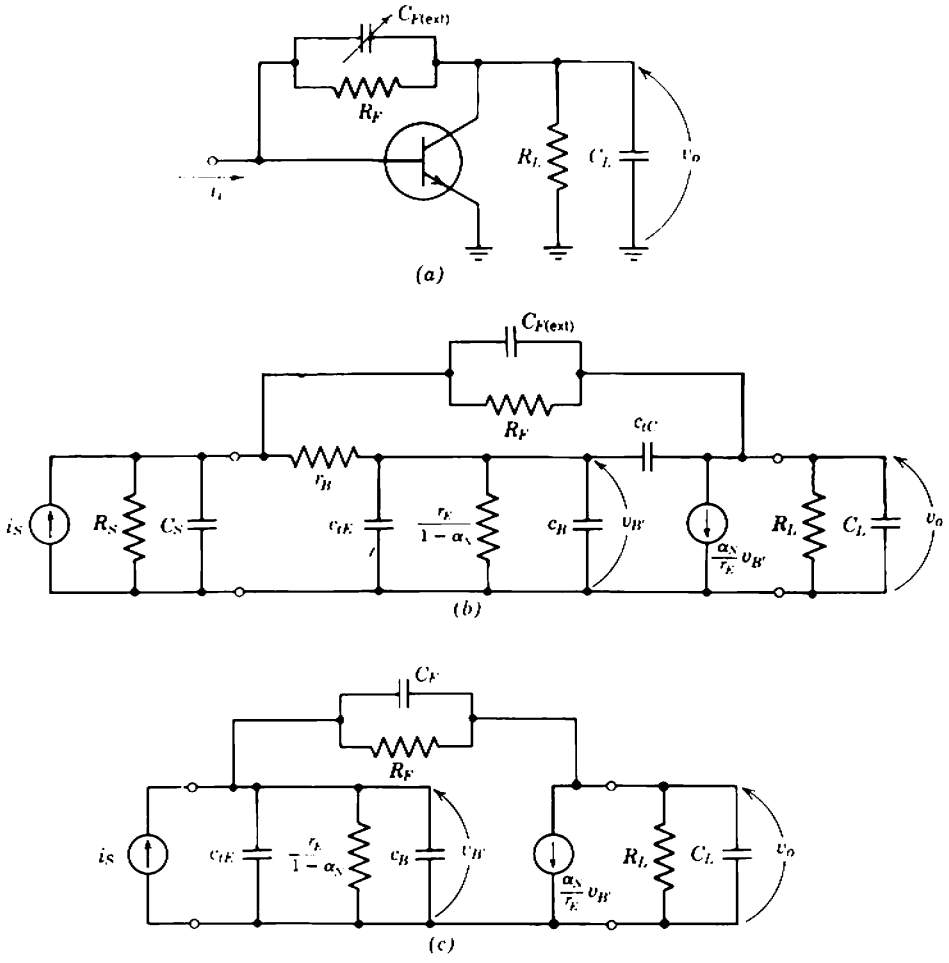


Fig. 13.23 Transistor shunt-feedback stage: (*a*) elemental circuit diagram; (*b*) complete equivalent circuit; (*c*) simplified equivalent circuit.

2. The input resistance of the intrinsic transistor $r_E/(1 - \alpha_N)$ is much smaller than any source resistance that should be used with a shunt-feedback stage. The feedback is short-circuited to ground if the source resistance is too small.

3. The input capacitance $(c_B + c_{tE})$ is usually much larger than the source stray capacitance C_S. When this is not the case, C_S can be combined with the transistor capacitances, and an effective frequency is defined as

$$\omega_{T(\text{eff})} = \frac{\alpha_N}{r_E(c_B + c_{tE} + C_S)}. \tag{13.100}$$

4. The total feedback capacitance is

$$C_F = C_{F(\text{ext})} + c_{tC}. \tag{13.101}$$

Thus the equivalent circuit reduces to Fig. 13.23*c*, and the ψ function for the simplified equivalent circuit follows from Eq. 13.92 as

$$\psi(s) = \frac{1}{1 + s\left[R_F C_F + \left(\frac{R_F + R_L}{R_L}\right)\tau_T + \frac{R_F(C_F + C_L)}{\beta_N}\right] + s^2[R_F(C_F + C_L)\tau_T]}. \tag{13.102}$$

The poles lie on a circle of radius

$$|s| = \left[\frac{1}{R_F(C_F + C_L)\tau_T}\right]^{\frac{1}{2}}, \tag{13.103}$$

and their real part is

$$\sigma = -\frac{C_F}{2(C_F + C_L)\tau_T} - \frac{R_F + R_L}{2R_F R_L(C_F + C_L)} - \frac{1}{2\beta_N \tau_T}. \tag{13.104}$$

Often some of the terms in these equations can be neglected; σ can usually be approximated by its first term.

The only term in Eqs. 13.102 to 13.104 that has a significant tolerance is the time constant τ_T. (The terms involving β_N are so small that they can almost invariably be neglected.) Typically, the value of τ_T lies between 70 and 140% of its nominal value. Thus $|s|$ can vary by $\pm 20\%$ about its nominal value, and σ can vary by $\pm 40\%$. As with vacuum-tube shunt-feedback stages, C_F is usually made adjustable to set the pole damping ratio.

13.5.3 Miscellaneous Topics

13.5.3.1 Gain-Bandwidth Product

The realizable gain-bandwidth product of an amplifying device has the dimensions of frequency and is defined as the product of the mid-band voltage gain with the 3-dB cutoff frequency in an unpeaked circuit:

$$\mathscr{GB} = |A_{Vm}| \times \omega_{co}.$$

A problem with feedback stages is that the mid-band transfer functions have the dimensions of admittance or impedance. The realizable $\mathscr{GB}$ for devices with feedback can be defined as in the following paragraph:

When two identical nonfeedback stages are cascaded, their combined voltage gain has a double pole on the negative real axis at

$$s_0 = -\frac{\mathscr{GB}}{\sqrt{|A_{Vm}|}},$$

where A_{Vm} is the total gain. Rearranging, the gain-bandwidth product for each device is

$$\mathscr{G}\mathscr{B} = -s_0\sqrt{|A_{Vm}|}.$$

When two feedback stages are cascaded, the stable transfer function is a voltage gain or current gain depending on the order of the stages. By suitable design the over-all singularity pattern can always be reduced to a double pole on the negative real axis and, by analogy with the nonfeedback stages, the realizable $\mathscr{G}\mathscr{B}$ for each device is defined as

$$\mathscr{G}\mathscr{B} = -s_0\sqrt{|A_{Vm}|} \quad \text{or} \quad -s_0\sqrt{|A_{Im}|}. \tag{13.105}$$

The two values are the same if the two amplifying devices are identical. For vacuum tubes the gain-bandwidth product defined in this way is (to a first order) the same as $\mathscr{G}\mathscr{B}$ without feedback. For transistors, however, $\mathscr{G}\mathscr{B}$ increases from less than ω_C without feedback to almost ω_T; the losses due to the base resistance r_B and collector transition capacitance c_{tC} are substantially eliminated.

Figure 13.24*a* shows a vacuum-tube shunt-feedback stage followed by a series-feedback stage. Single primes are appended to the parameters of the first stage and double primes to the second. Neglecting the feedback capacitor, the mean distance of the poles of the shunt stage from the origin follows from Eq. 13.98 as

$$|s| = \left(\frac{g_m'}{R_F'C_1'C_2'}\right)^{1/2}$$

and $|s|$ is independent of the actual configuration of the poles. For the series-feedback stage the ratio of the distances of the pole and zero from the origin follows from Eq. 13.80 as

$$\left|\frac{p}{z}\right| = 1 + g_m''R_F''.$$

Therefore, if the stages are so designed that the zero from the series stage cancels any one of the poles, the mean distance of the remaining poles from the origin is

$$|s| = \left[\frac{(1 + g_m''R_F'')g_m'}{R_F'C_1'C_2'}\right]^{1/2}.$$

When the poles are set coincident (for evaluating $\mathscr{G}\mathscr{B}$), their position on the negative real axis is

$$s_0 = -\left[\frac{(1 + g_m''R_F'')g_m'}{R_F'C_1'C_2'}\right]^{1/2}.$$

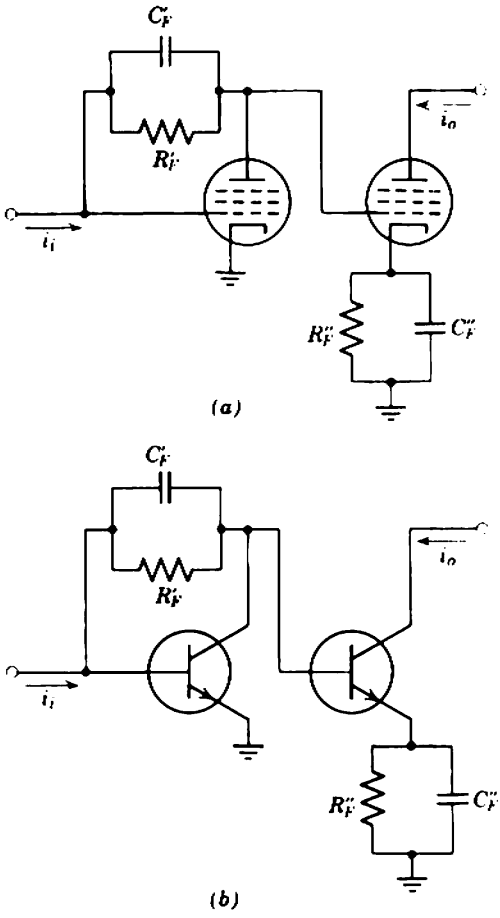


Fig. 13.24 Elemental circuit diagrams for $\mathscr{G}\mathscr{B}$ calculations on feedback stages.

But the mid-band current gain of the two stages is

$$|A_{Im}| = R_T' \times G_T'' \approx R_F'\left(\frac{g_m''}{1 + g_m'' R_F'}\right).$$

Hence the realizable gain-bandwidth product follows from Eq. 13.105 as

$$\mathscr{G}\mathscr{B} = \left(\frac{g_m' g_m'}{C_1' C_2'}\right)^{1/2}.$$

The capacitances C_1' and C_2' are the sums of the input capacitance of one tube, the output capacitance of the preceding tube, and the stray capacitance C_S. If the tubes are identical, then

$$g_m' = g_m'' = g_m,$$
$$C_1' = C_2' = C_o + C_S + C_i,$$

and the realizable gain-bandwidth product becomes

$$\mathscr{G}\mathscr{B} = \frac{g_m}{C_o + C_S + C_i}, \tag{13.106}$$

as given by Eq. 13.48 for stages without feedback.

Figure 13.24*b* shows a transistor shunt-feedback stage followed by a series-feedback stage. With the same notation and approximations as for the vacuum tube, the mean distance of the poles of the shunt stage from the origin follows from Eq. 13.103 as

$$|s| = \left(\frac{1}{R'_F C'_L \tau'_T}\right)^{1/2}.$$

But C'_L is the input capacitance of the following series stage, and using Eq. 13.90

$$|s| = \left(\frac{1}{R'_F G''_T \tau'_T \tau''_T}\right)^{1/2}.$$

In addition to the two poles from the shunt stage, the current gain of the two stages has a third pole due to the series stage. With typical values for r_B and G_T, however, Eq. 13.88 shows that this pole is in the neighborhood of $s = -1/\tau_T$ (subject, of course, to the validity of the charge-control representation at such frequencies). This is so far out along the negative real axis that the pole is always nondominant and may be neglected. If the remaining two poles are set coincident, their position is

$$s_0 = -\left(\frac{1}{R'_F G''_T \tau'_T \tau''_T}\right)^{1/2}$$

and, since the mid-band current gain is

$$A_{Im} = R'_T \times G''_T \approx R'_F \times G''_T,$$

the realizable gain-bandwidth product is

$$\mathscr{G}\mathscr{B} = \left(\frac{1}{\tau'_T \tau''_T}\right)^{1/2} = \sqrt{\omega'_T \omega''_T}.$$

For identical transistors

$$\mathscr{G}\mathscr{B} = \omega_T. \tag{13.107}$$

The approximations in the derivations of Eqs. 13.106 and 13.107 are such that the true realizable $\mathscr{G}\mathscr{B}$ falls somewhat below the ideal value:

(i) The mid-band transfer resistance R_T of the shunt stage is approximated by the feedback resistor R_F.

(ii) Miller effect in the series stage is ignored.
(iii) The feedback capacitor of the shunt stage is ignored in comparison with C_1 and C_2.

The last of these approximations sets an upper limit to the frequency at which feedback peaking is efficient. If all devices are the same so that

$$C_1 = C_2 = C$$

and

$$\mathscr{G}\mathscr{B} = \frac{g_m}{C},$$

then Eq. 13.96 gives the real part of the poles as

$$\sigma = -\mathscr{G}\mathscr{B}\left(\frac{C_F}{2C}\right).$$

If the realizable gain-bandwidth product is to approach its ideal limit, C_F must be much smaller than C_1 or C_2, say

$$C_F \leq 3C.$$

By substitution, therefore, the real part of the poles should satisfy

$$|\sigma| \leq \frac{\mathscr{G}\mathscr{B}}{6}. \tag{13.108}$$

Although feedback peaking is usable at larger values of $|\sigma|$, the realizable $\mathscr{G}\mathscr{B}$ falls away rapidly.

13.5.3.2 Cathode Peaking

A very useful circuit arrangement for vacuum tubes is a cascade of self-biased RC-coupled pentodes with their cathode bypass capacitors omitted. Essentially, the arrangement is a cascade of series-feedback stages, and it can be shown that if a suitable high-frequency peaking capacitor is connected across each cathode resistor the performance is identical to that of an unpeaked amplifier with perfectly bypassed cathode resistors. The advantage of the cathode-peaking circuit stems from two practical considerations:

(i) The low-frequency pole and zero associated with the cathode-bypass capacitor are eliminated, thus simplifying the low-frequency compensation.
(ii) The small amount of series feedback provided by the cathode resistor stabilizes the mid-band gain and reduces the nonlinearity.

Although Section 11.2.3 shows that the alternate cascade has no theoretical advantage over cascaded series-feedback vacuum-tube stages, the same is

not true for transistors. Emitter peaking, the transistor counterpart of cathode peaking, is a very unsatisfactory circuit arrangement.

Figure 13.25*a* shows a simple pentode amplifier stage at high frequencies, and Fig. 13.25*b* is the same stage modified to include cathode peaking. The stray capacitance C is the same for each circuit, and represents the sum of the output capacitance for one tube, the wiring capacitance, and the input capacitance of the following tube. The voltage gain of the simple circuit is

$$A_V(s) = -g_m \times Z_L(s);$$

that is,

$$A_V(s) = -g_m\left(\frac{R_A}{1 + sR_AC}\right). \tag{13.109}$$

For the modified circuit

$$A_V(s) = -Y_T(s) \times Z'_L(s)$$

and, using Eq. 13.80,

$$A_V(s) = -\left(\frac{g_m}{1 + g_mR_F}\right)\left\{\frac{1 + s(R_FC_F)}{1 + s[R_FC_F/(1 + g_mR_F)]}\right\}\left(\frac{R'_A}{1 + sR'_AC}\right). \tag{13.110}$$

Equation 13.110 reduces to Eq. 13.109 if the circuit is designed with

$$R'_A = R_A(1 + g_mR_F), \tag{13.111}$$

$$R_FC_F = R'_AC. \tag{13.112}$$

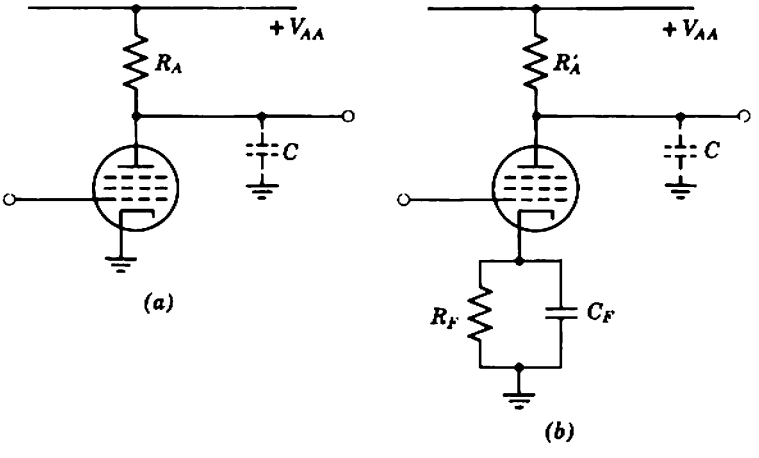


Fig. 13.25 Cathode peaking: (*a*) simple amplifier; (*b*) modified amplifier.

13.5.3.3 Transistor Shunt Stage with a Resistive Load

A transistor shunt-feedback stage is sometimes used to drive a small, purely resistive load such as a terminated coaxial cable. Because the load is resistive, there is only one pole in the open-loop transfer impedance and, consequently, only one real pole in the closed-loop response. Complex poles can be produced by adding a small peaking inductor in series with the feedback resistor (Fig. 13.26); the total feedback impedance is

$$Z_F(s) = R_F + sL_F = R_F(1 + s\tau_F). \tag{13.113}$$

The zero in Z_F appears as a pole in the feedback factor (and hence in the loop gain) and the closed-loop poles move off the negative real axis. Notice particularly that this form of peaking is quite unsuitable when the load is capacitive, for the loop gain then has three poles.

With the same approximations as for the stage in Section 13.5.2.2 which has a capacitive load, the transfer impedance of the stage becomes

$$Z_T(s) = -\frac{R_F}{1 + (R_F + R_L)/\beta_N R_L} \times \left\{ \frac{1 + s\tau_F}{1 + s\left[\left(\frac{R_F + R_L}{R_L}\right)\tau_T\right] + s^2\left[\left(\frac{R_F}{R_L}\right)\tau_T\tau_F\right]} \right\}. \tag{13.114}$$

The two poles lie on a circle of radius

$$|s| = \left(\frac{R_L}{R_F\tau_T\tau_F}\right)^{1/2}, \tag{13.115}$$

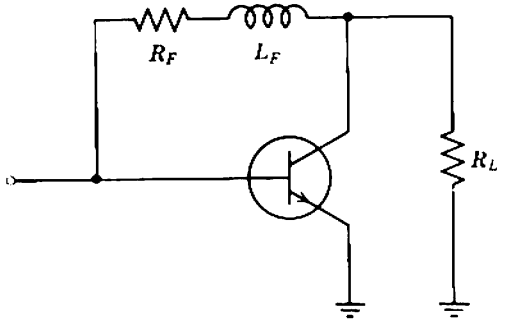


Fig. 13.26 High-frequency peaking circuit for a transistor shunt stage with a small, purely resistive load.

and their real part is

$$\sigma = -\frac{R_F + R_L}{2R_F\tau_F}. \tag{13.116}$$

There is a zero on the negative real axis at

$$s = -\frac{1}{\tau_F}. \tag{13.117}$$

13.6 THE BASIS OF PEAKING

When all stages in an amplifier are the same, the realizable gain-bandwidth product for each device is defined as

$$\mathscr{GB} = |A_{Vm}| \times \omega_{co}, \tag{13.118a}$$

where A_{Vm} is the mid-band voltage gain per stage and ω_{co} is the 3-dB cutoff frequency per stage (in the absence of peaking). The realizable $\mathscr{GB}$ is the quotient of the device g_m with the total shunt capacitance (multiplied, in the case of a transistor, by a constant). Strictly speaking, $\mathscr{GB}$ defined in this way is the voltage gain-bandwidth product; a current gain-bandwidth product can be defined as the product of the mid-band current gain between collecting electrodes and the 3-dB cutoff frequency:

$$\mathscr{GB} = |A_{Im}| \times \omega_{co}. \tag{13.118b}$$

As examples, transistors are treated on a voltage basis in Section 7.5.1.2 and on a current basis in Section 13.3.2. The voltage and current gain-bandwidth products are, of course, equal when all stages are the same.

When the stages in an amplifier are not the same, $\mathscr{GB}$ for each device must be interpreted as a suitably defined effective value. Allowance must be made for the fact that the shunt capacitances at the input and output of a device are not the same. The voltage gain-bandwidth product is related to the device g_m and the total shunt capacitance at the output of the device; an example is the calculation for a vacuum tube in Section 7.5.1.1. In contrast, the current gain-bandwidth product is related to the device g_m and the total shunt capacitance at the input of the device; an example is the calculation for the charge-control model in Section 2.5.1.

Formal complications arise with feedback amplifiers in which the individual stage transfer functions have the dimensions of admittance or impedance. When an amplifier consists of an even number of stages it can be split into groups of two stages; each group has either voltage gain or current gain stable, and the voltage or current gain-bandwidth product

of each device can be defined as in Section 13.5.3.1. When there are an odd number of stages, however, the last stage must be considered separately. If the last stage has series feedback, its voltage gain-bandwidth product is

$$\mathscr{GB} = |G_T \times R_L| \times \omega_{co}, \tag{13.118c}$$

where R_L is the resistive component of the load impedance, and the cutoff frequency ω_{co} depends principally on the load capacitance. For a shunt-feedback stage the current gain-bandwidth product is

$$\mathscr{GB} = \left|\frac{R_T}{R_L}\right| \times \omega_{co}. \tag{13.118d}$$

With these conventions, the mid-band voltage or current gain of an amplifying device with high-frequency peaking but without feedback can always be manipulated into the form

$$TF_{m(1\,\text{stage})} = \xi\left[\frac{\mathscr{GB}}{\Pi(p_{\text{eff}})}\right], \tag{13.119a}$$

where $\mathscr{GB}$ = the realizable gain-bandwidth product of the device (Eqs. 13.48, 13.51, or 13.53),

$\Pi(p_{\text{eff}})$ = the product of the effective pole distances from the origin; as zeros count as reciprocal poles,

$$\Pi(p_{\text{eff}}) = \frac{\Pi(p)}{\Pi(z)}. \tag{13.120a}$$

ξ = a constant of the peaking circuit.

The same equation applies to one feedback stage in combination with its load impedance, but for a group of two feedback stages

$$TF_{m(2\,\text{feedback stages})} = \xi\left[\frac{\mathscr{GB}' \times \mathscr{GB}''}{\Pi(p_{\text{eff}})}\right]. \tag{13.119b}$$

Examination of the preceding sections on practical high-frequency peaking circuits reveals two important facts. Except for peaking circuits that include some elements in series with the signal path:

(i) There is only one effective pole per device (or two poles for a group of two devices with feedback); the number of effective poles is the difference between the actual numbers of poles and zeros:

$$n_{\text{eff}} = n_p - n_z. \tag{13.120b}$$

(ii) The constant ξ is unity.

Peaking networks that do not include series elements are described as one-port or more commonly as *two-terminal*, because the input and output ports are the same. Networks that do include series elements are described as two-port or *four-terminal*. Four-terminal peaking networks are relatively uncommon; the undesignable vacuum-tube circuit of Fig. 13.14 is one example, and the very inefficient nonfeedback transistor circuits are another group.

By multiplication, the mid-band voltage or current gain of a multistage amplifier with two-terminal peaking networks is

$$TF_m = \frac{\text{product of } \mathscr{G}\mathscr{B} \text{ for all devices}}{\text{product of distances of effective poles from origin}} = \frac{\Pi(\mathscr{G}\mathscr{B})}{\Pi(p_{\text{eff}})}. \tag{13.121}$$

Equation 13.121 is the mathematical statement of a very important theorem for two-terminal peaking networks:

> For an amplifier using given amplifying devices [i.e., known $\Pi(\mathscr{G}\mathscr{B})$] and having a specified dynamic response [i.e., is a specified singularity pattern and hence $\Pi(p_{\text{eff}})$], the maximum attainable mid-band gain is independent of the peaking system used to realize the response.

Poles and zeros can be introduced and moved at will by modifying the peaking of individual stages but, provided pole-zero cancellation occurs between stages so that the total singularity pattern is constant, the maximum attainable gain is constant also. The actual gain can, of course, be reduced by means of an attenuator.

13.6.1 Bandwidth Reduction Factor

The 3-dB bandwidth of a transfer function that has n_{eff} effective poles can be written as

$$\mathscr{B} = \chi[\Pi(p_{\text{eff}})]^{1/n_{\text{eff}}}, \tag{13.122}$$

where $[\Pi(p_{\text{eff}})]^{1/n_{\text{eff}}}$ is the geometric mean distance of the effective poles from the origin, and χ is a constant of the singularity pattern known as the *bandwidth reduction factor*. The bandwidths of the singularity pattern types discussed in Section 13.2 are normalized to the geometric mean effective pole distance, and hence χ for these singularity patterns can be written down from tabulated data.

Rearrangement of Eq. 13.121 yields

$$\Pi(p_{\text{eff}}) = \frac{\Pi(\mathscr{G}\mathscr{B})}{TF_m},$$

and substitution into Eq. 13.122 gives the bandwidth of an amplifier with two-terminal peaking networks as

$$\mathscr{B} = \chi\left[\frac{\Pi(\mathscr{G}\mathscr{B})}{TF_m}\right]^{1/n_{eff}}. \tag{13.123}$$

Thus *the bandwidth of an amplifier that uses given amplifying devices and has a specified mid-band gain is determined entirely by the bandwidth reduction factor of its singularity pattern.* In essence, the process of peaking an amplifier at high frequencies is nothing more than choosing a singularity pattern for which χ is adequately large and for which the dynamic response is suitable. Realizing this singularity pattern is largely a mechanical process, based on the equations of Sections 13.4 and 13.5.

Table 13.7 compares the values of χ for three all-pole functions, and Table 13.8 compares three linear-phase functions. The two entries under

Table 13.7 Bandwidth Reduction Factor χ for Various All-Pole Singularity Patterns

Order n_p	Coincident	Linear-Phase	Maximally-Flat
2	0.644	0.786	1.000
3	0.510	0.712	1.000
4	0.435	0.659	1.000
5	0.386	0.617	1.000

cascaded 2-pole groups in Table 13.8 represent a single 2-pole linear-phase group and two such groups in cascade. The normalized frequencies 0.79 and 0.55 are read from Fig. 13.8*b* as the 3-dB and 1.5-dB points for a single group, and correspond to 3-dB total attenuation for one and two groups, respectively. Similarly, the entries under *cascaded 2-pole, 1-zero groups* represent from one to five 2-pole, 1-zero linear-phase groups in

Table 13.8 Bandwidth Reduction Factor χ for Various Linear-Phase Functions

$n_{eff} = n_p - n_z$	All-Pole	Cascaded 2-Pole Groups	Cascaded 2-Pole, 1-Zero Groups
1	1.00	—	1.57
2	0.79	0.79	1.07
3	0.71	—	0.94
4	0.66	0.55	0.81
5	0.62	—	0.70

cascade. The normalized frequencies 1.57, 1.07, 0.94, 0.81, and 0.70 are read from Fig. 13.9*b* as the 3-dB, 1.5-dB, 1-dB, 0.75-dB, and 0.6-dB points and correspond to 3 dB total for one to five groups, respectively. Some general qualitative trends for χ are apparent from these tables:

1. Table 13.7 illustrates that χ for a given number of effective poles falls as the overshoot in the step response is reduced. Coincident-pole functions have zero overshoot, linear-phase functions have less than 1%, and maximally-flat have overshoots that increase monotonically with increasing order.

2. Also from Table 13.7, χ for a given singularity pattern type falls as the number of effective poles is increased. The fall is most rapid for coincident-pole functions, and zero for maximally-flat functions. In general, χ for large numbers of effective poles can be written as

$$\chi = \chi_0 n_{\text{eff}}^r, \tag{13.124}$$

where $r = -\frac{1}{2}$ for coincident-pole functions,
$r \approx -\frac{1}{4}$ for linear-phase functions,
$r = 0$ for maximally-flat functions.

For pattern types such as equal-ripple functions which give more overshoot than maximally-flat functions, χ may increase with increasing n_{eff}, but always remains finite.

3. Table 13.8 illustrates that χ is greater for one large group of poles than for a number of smaller groups in cascade. For example, the normalized cutoff frequency of a 4-pole linear-phase group is 0.66, compared with 0.55 for two 2-pole groups in cascade. Section 13.6.1.1 below contains a relevant theorem.

4. Table 13.8 also illustrates that χ for n_{eff} effective poles tends to be larger when the actual number n_p of poles is greater than n_{eff}, and there are $(n_p - n_{\text{eff}})$ zeros in the singularity pattern. For small numbers of effective poles, cascaded 2-pole, 1-zero linear-phase groups yield more bandwidth than one large linear-phase group. The single group always takes the lead for large numbers of poles (seven in this particular case).

5. A rule not illustrated by the tables is that χ tends to be small for a singularity pattern in which some poles are relatively close to the origin and dominate the response, whereas other poles are relatively far from the origin. Consider any singularity pattern for which χ is known. Two effects occur if another pole is added a long way from the origin; the bandwidth falls (or at best remains unchanged), and the mean distance of the poles from the origin increases. There is a reduction in χ.

Although large singularity groups that contain some zeros have the theoretical advantage of the largest values of χ for specified overshoot,

these groups have the disadvantage of being the least practicable. Invariably, some of the poles in a large group lie close to the imaginary axis and such poles are difficult to control; four (or perhaps five) is a practical limit to the number of poles in one group. If zeros are present (except as in the 2-pole, 1-zero function), the pattern is not one of the tabulated types for which the response and value of χ are known. Hence practical amplifiers having more than four or five effective poles are usually built up by cascading a number of building-blocks, each of which may have up to about four poles.

13.6.1.1 The Cost of Subdivision*

Although large amplifiers are of necessity built up by cascading a number of building-blocks, paragraph 3 above suggests that this practical simplification is achieved only at the cost of a reduction in χ. The reduction in χ can be found exactly for large numbers of poles (for which Eq. 13.124 applies) and approximately for smaller numbers.

Consider an amplifier that has n_{eff} effective poles and whose response is to be of some specified type (linear-phase, maximally-flat, etc). If the poles are arranged to form one group, Eq. 13.124 gives the bandwidth reduction factor as

$$\chi_{\text{large group}} = \chi_0 \, n_{\text{eff}}^r. \tag{13.125}$$

The exponent r depends on the nature of the group but is never more negative than $-\frac{1}{2}$.

If the amplifier is divided into k identical building-blocks, each of these will contain (n_{eff}/k) poles. The bandwidth reduction factor for each group is

$$\chi_{\text{building-block}} = \chi_0 \left(\frac{n_{\text{eff}}}{k}\right)^r.$$

When a number of identical building-blocks are cascaded, Eq. 13.38*b* shows that the cutoff frequency falls as the square root of the number of blocks provided their overshoot is small. The over-all bandwidth reduction factor for the k building-blocks in cascade is therefore

$$\chi_{\text{cascaded blocks}} = \frac{\chi_{\text{building-block}}}{\sqrt{k}} = \chi_0 \left[\frac{n_{\text{eff}}^r}{k^{r+\frac{1}{2}}}\right]. \tag{13.126}$$

Dividing Eq. 13.125 by Eq. 13.126, the bandwidth advantage of the single group of poles is

$$\frac{\chi_{\text{large group}}}{\chi_{\text{cascaded blocks}}} = k^{r+\frac{1}{2}}. \tag{13.127}$$

* Section 13.6.1.1 may be omitted on a first reading.

Since r is not more negative than $-\frac{1}{2}$, the exponent is positive and the bandwidth advantage of the single group increases monotonically with increasing k. In other words, the over-all bandwidth falls as an amplifier is split into more building-blocks.

The result of this theorem is expected from intuitive reasoning. Bandwidth shrinkage for cascaded building-blocks is obviously reduced as the response of the individual building-blocks approaches closer to the brick-wall magnitude function; in the limit there is no shrinkage at all for building-blocks that have the ideal brick-wall response with a square corner. Maximally-flat, linear-phase, and equal-ripple functions all approach closer to the brick-wall magnitude response as the number of their poles is increased. Hence the optimum amplifier is that in which each building-block has the largest possible number of poles, that is, when the complete amplifier consists of just one group.

13.6.1.2 Maximum Bandwidth with Specified Gain*

There is an upper limit to the bandwidth that can be achieved from an amplifier with a specified mid-band gain, even if the number of active devices can be increased without limit.

In general, the bandwidth of an n-stage amplifier with two-terminal peaking is given by Eq. 13.123, with $n_{eff} = n$. For simplicity assume that all devices have the same gain-bandwidth product, so that Eq. 13.123 reduces to

$$\mathscr{B} = \frac{\chi\, \mathscr{G}\mathscr{B}}{(TF_m)^{1/n}}.$$

Also, for simplicity, assume that the amplifier's singularity pattern is a single large group for which Eq. 13.124 applies; if the amplifier consists of a number of cascaded building-blocks, the over-all bandwidth will be less. Hence the maximum bandwidth available with the specified singularity pattern type is

$$\mathscr{B} = \chi_0\, \mathscr{G}\mathscr{B}\left[\frac{n^r}{(TF_m)^{1/n}}\right]. \tag{13.128}$$

By differentiation with respect to n there is an optimum number of stages for which the bandwidth is maximum:

$$n_{opt} = -\frac{\ln (TF_m)}{r}. \tag{13.129}$$

n_{opt} is a finite positive number for all singularity pattern types for which r

* Section 13.6.1.2 may be omitted on a first reading.

is negative. Substitution into Eq. 13.128 gives the maximum bandwidth as

$$\mathscr{B}_{max} = \chi_0 \mathscr{G}\mathscr{B}\left\{\frac{[\ln (TF_m)]^r}{[-r]^r[\exp(-r)]}\right\}. \tag{13.130}$$

For the special case of maximally-flat functions, r is zero, n_{opt} is infinite, χ_0 is unity, and

$$\mathscr{B}_{max} = \mathscr{G}\mathscr{B}.$$

As an example, consider an amplifier that has n coincident poles. From Eq. 13.11

$$\chi_0 \approx \frac{1}{1.2} = 0.83,$$

and the exponent r is $-\frac{1}{2}$. Hence,

$$n_{opt} = 2 \ln (TF_m)$$

and

$$\mathscr{B}_{max} = \frac{\mathscr{G}\mathscr{B}}{2.8\sqrt{\ln (TF_m)}}.$$

If the devices have a realizable gain-bandwidth product of 100 MHz and the mid-band gain is 1000, the optimum number of stages is 14 and the bandwidth is 13.6 MHz. Very nearly the same performance can be obtained more economically with a much smaller number of stages. The optimum condition is not at all sharp.

13.6.2 Four-Terminal Peaking Networks*

The voltage or current gain of an amplifying device with high-frequency peaking can be expressed in the form of Eq. 13.119. For two-terminal peaking networks there is only one effective pole and the constant ξ is unity. For four-terminal networks, however, there is more than one pole and ξ has the dimensions of frequency raised to some integral power. Rearranging Eq. 13.119, the mean distance from the origin of the effective poles of one device with a four-terminal peaking network is

$$[\Pi(p_{eff})]^{1/n_{eff}} = \left(\frac{\xi\,\mathscr{G}\mathscr{B}}{TF_m}\right)^{1/n_{eff}}$$

compared with $(\mathscr{G}\mathscr{B}/TF_m)$ for one device with a two-terminal peaking network. The ratio of the mean pole distances for one device with four- and two-terminal peaking networks is thus

$$Y = \frac{\{[\Pi(p_{eff})]^{1/n_{eff}}\}_{4\text{-terminal peaking}}}{\{[\Pi(p_{eff})]^{1/n_{eff}}\}_{2\text{-terminal peaking}}} = \left[\xi\left(\frac{TF_m}{\mathscr{G}\mathscr{B}}\right)^{n_{eff}-1}\right]^{1/n_{eff}}. \tag{13.131}$$

* Section 13.6.2 may be omitted on a first reading.

This ratio can be made greater than unity by designing the four-terminal network so ξ is large, and Eq. 13.122 predicts an increase in bandwidth over two-terminal peaking provided that the bandwidth reduction factor χ remains constant. Unfortunately, networks that yield a large value for Y are not often useful for compensating wide-band amplifiers; either the overshoot is so large as to be quite unacceptable or, if the response is acceptable, there is an almost direct trade-off between χ and Y. Although it is a well-established fact that four-terminal peaking can be more efficient than two-terminal, to the authors' knowledge the advantage has not been established quantitatively. In the following section, Y is shown to be 2 for one useful singularity pattern (the three-pole maximally-flat for which $\chi = 1$). Empirically, it is known that the advantage of four-terminal peaking over two-terminal can be rather more than two-fold. The advantage arises because the stray capacitance is split into two components by the series peaking elements; the sum of the two components is the capacitance that defines the realizable $\mathscr{GB}$ for the device, but either component is less than the total.

With two simple modifications, the results of Section 13.6.1 apply to amplifiers with four-terminal peaking networks:

(i) n_{eff}, the effective number of poles, is more than one per device;
(ii) $\mathscr{GB}$ is replaced by $\xi\,\mathscr{GB}$.

13.6.2.1 Example in which $Y > 1$

Consider the interstage network shown in Fig. 13.27. This network is a reduced form of the vacuum-tube circuit of Fig. 13.14, in which L_1 is set at zero:

$$\psi(s) = \frac{1}{1 + s[R_A(C_1 + C_2)] + s^2(L_2C_2) + s^3(R_AL_2C_1C_2)}. \quad (13.132)$$

The mid-band gain and realizable $\mathscr{GB}$ are

$$A_{Vm} = -g_mR_A,$$

$$\mathscr{GB} = \frac{g_m}{C_1 + C_2},$$

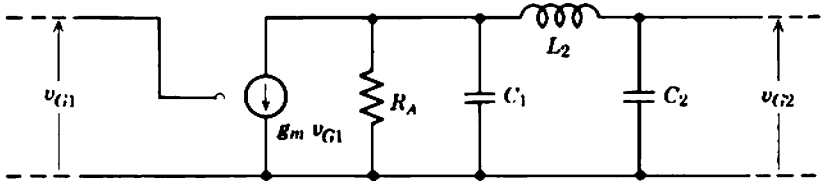


Fig. 13.27 Vacuum-tube interstage network with a series peaking inductor.

and the magnitudes of the sum and product of the poles follow from Eq. 13.132 as

$$\Sigma(p) = \frac{1}{R_A C_1},$$

$$\Pi(p) = \frac{1}{R_A L_2 C_1 C_2}.$$

Hence, from Eq. 13.119*a*,

$$\xi = \frac{C_1 + C_2}{L_2 C_1 C_2}.$$

Notice that, although ξ can be increased without limit by reducing the value of L_2, χ must become very small because the performance approaches that of an unpeaked amplifier. Substitution into Eq. 13.131 gives the ratio of the mean pole distances with four- and two-terminal peaking as

$$Y = \left[\frac{R_A^2(C_1 + C_2)^3}{L_2 C_1 C_2}\right]^{1/3}$$

and, after manipulation,

$$Y = \frac{[\Pi(p)]^{1/3}(1 + C_2/C_1)}{\Sigma(p)}. \tag{13.133}$$

It is easy to show that if the stray capacitances are in the ratio

$$C_2 = 3C_1,$$

the response becomes three-pole maximally-flat for a particular value of L_2. The relation between the poles follows from the geometry of a three-pole maximally-flat pattern as

$$[\Pi(p)]^{1/3} = \tfrac{1}{2}\Sigma(p),$$

and substituting into Eq. 13.133 gives

$$Y = 2.$$

Thus Y is greater than unity for this useful singularity pattern, illustrating that four-terminal peaking can under certain circumstances be an advantage.

13.6.3 Concluding Comments

As stressed in the introduction to this chapter, modern amplifying devices have large realizable gain-bandwidth products. Usually, therefore, adequate mid-band gain and bandwidth can be achieved without a prohibitively large number of stages, even if the design is far from the most efficient. The bandwidth reduction factor χ and the mean pole distance

ratio Y are thus not the sole figures of merit for a wide-band amplifier's design; gain stability, linearity, and the circuit's general designability are also important considerations. Simplifying design techniques are therefore acceptable, although they reduce the values of χ and Y:

(i) Most large amplifiers consist of a cascade of building-blocks. Each of these building-blocks normally has about the same bandwidth and a similar transient response; often some of the blocks are identical.
(ii) Two-terminal peaking circuits are more common than the less-designable four-terminal peaking circuits.

On all accounts, feedback peaking of transistor amplifiers is the most satisfactory arrangement. The realizable $\mathcal{GB}$ increases from less than ω_C to very nearly the ideal limit ω_T derived from charge-control theory in Section 4.2.3.2. Four-terminal inductance peaking offers no advantage over two-terminal peaking, because the shunt capacitance is predominantly the input capacitance of the following transistor and this cannot be split by a series inductance. Gain stability and linearity accrue as bonuses attendant on the use of feedback for peaking. Feedback peaking is, of course, mandatory in integrated-circuit amplifiers because inductors cannot be realized.

The optimum arrangement for vacuum tubes is not so clear cut. Four-terminal inductance peaking offers about a 2:1 increase in bandwidth over any two-terminal peaking system, and is almost mandatory in very-wide-band amplifiers. Feedback peaking offers gain stability, but usually at the price of a second-order reduction in bandwidth below two-terminal inductance peaking. One very satisfactory arrangement is a combination of feedback and inductance peaking—cathode peaking to apply a small amount of feedback and to eliminate one troublesome low-frequency bypass capacitor per stage, and shunt peaking in the anode circuit for a further increase in bandwidth.

13.7 COAXIAL CABLES

Wide-band amplifiers are usually interconnected with coaxial cables. In the classification of Section 7.2.2, coaxial cables are described as matched sources or loads, which should be matched at each end by the driving-point impedances of the amplifier. Single-ended termination, however, is often acceptable; the cable is matched at one end (preferably the receiving end) but fed from either a low-impedance or high-impedance source.

There are two methods for producing an amplifier with an accurately defined driving-point impedance, such as is required for matching a coaxial cable. Either the amplifier is designed to have a very high driving-point impedance and a small terminating resistor is connected in shunt, or else the amplifier has a very low driving-point impedance and a resistor is connected in series. Section 7.2.3 contains the details of these methods.

Typically, the characteristic impedance of a coaxial cable lies between 50 and 100 Ω. This is so small that, with the circuit techniques described so far in this book, it is almost impossible to design a vacuum-tube amplifier whose driving-point impedance is small in comparison. Hence, the sending-end and receiving-end circuit arrangements for a vacuum-tube amplifier are usually those shown in Fig. 13.28. There is a 2:1 increase in gain if either terminating impedance is omitted, but against this reflections occur at the unterminated end of the cable. Notice that the low-frequency cutoff is determined largely by R_A and C_C, not by Z_0 and C_C.

Notice that the terminating impedances shown in Fig. 13.28 as resistors are often *RLC* networks and should be designed to give the best cable match after allowing for the driving-point impedance of the amplifier proper. Figure 13.29 shows a particularly useful receiving-end circuit. The input impedance is purely resistive at all frequencies if

$$R_1 = R_2 = Z_0 \tag{13.134}$$

and

$$L = Z_0^2 C_1. \tag{13.135}$$

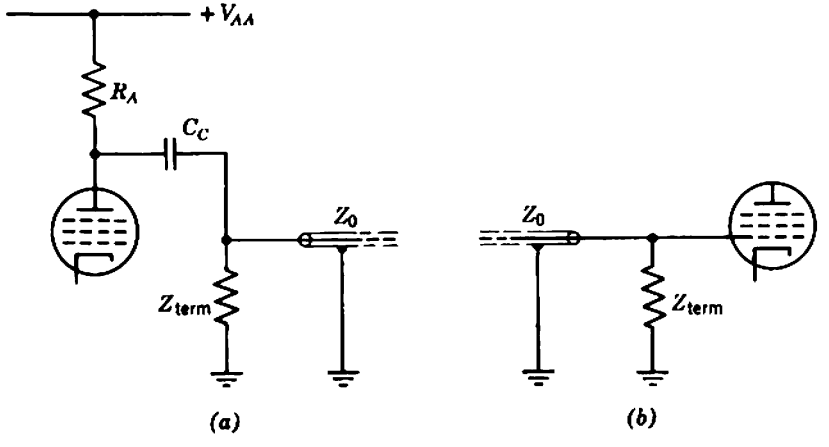


Fig. 13.28 Circuit arrangements for coaxial cables and vacuum-tube amplifiers: (*a*) sending-end; (*b*) receiving-end.

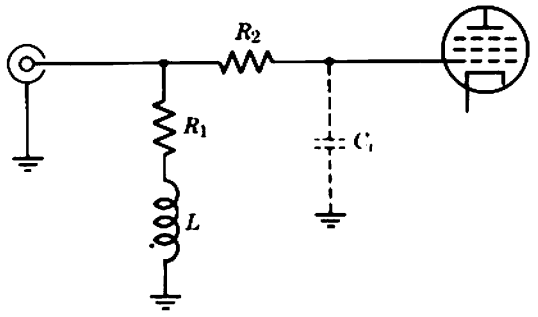


Fig. 13.29 A particularly useful receiving-end terminating network for a coaxial cable.

In the transistor counterpart of this circuit, the base resistance r_B forms a part of R_2.

With nonfeedback or series-feedback transistor stages, it is again difficult to realize a low driving-point impedance, so the circuit arrangements shown in Figs. 13.30*a* and *b* are used. These are analogous to the

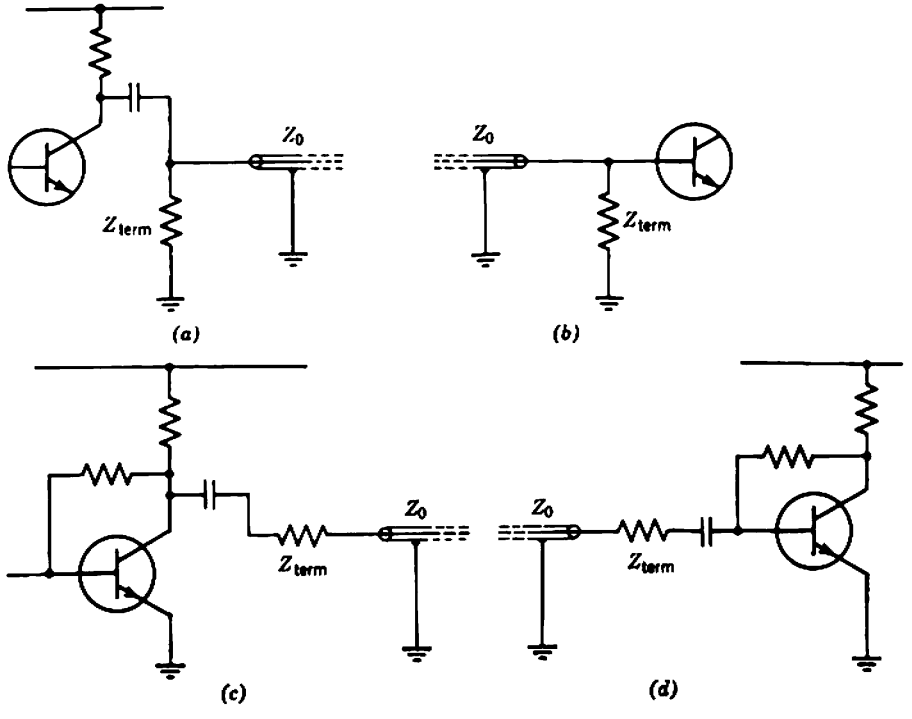


Fig. 13.30 Circuit arrangements for coaxial cables and transistor amplifiers.

vacuum-tube circuits; either of the terminations may be omitted if single-ended matching is acceptable. With shunt-feedback stages, however, driving-point impedances as low as a few tens of ohms can be obtained and the circuit arrangements in Figs. 13.30*c* and *d* are feasible. The cable matches attainable with these shunt-feedback circuits tend to be unstable, because the unstabilized driving-point impedances of the amplifier proper are comparable in magnitude with the series terminating impedance (see Section 7.2.3). A very useful sending-end circuit results from omitting the terminating impedance from Fig. 13.30*c* and using a shunt-feedback stage to drive the cable directly with a voltage. Section 13.5.3.3 contains the analysis of a transistor shunt-feedback stage with a resistive load.

13.8 TWO WORKED DESIGN EXAMPLES

This section gives two examples of the practical application of the theory of compensated wide-band amplifiers. The discussion is restricted to the mid-band and high-frequency aspects of the design; the biasing and low-frequency response are considered in Section 12.4.

The essential steps in designing a wide-band amplifier are as follows; the order in which they are taken depends on the data supplied and the exact specification of what is required of the amplifier.

1. Choose a singularity pattern type for which the dynamic response is suitable and for which the bandwidth reduction factor χ is known or can be estimated.
2. Choose a number of amplifying devices for which the product of the realizable gain-bandwidth products $\Pi(\mathscr{G}\mathscr{B})$ is such that adequate mid-band gain and bandwidth can be achieved with the chosen value of χ (Eq. 13.123); if four-terminal peaking is used, include the value of ξ.
3. Go through the mechanics of positioning the singularities as desired. If the original estimates of χ or $\Pi(\mathscr{G}\mathscr{B})$ are too large, the attainable mid-band gain or bandwidth will fall short of specifications and the design must be started again.

13.8.1 Vacuum-Tube Amplifier

(i) mid-band voltage gain	100,
(ii) bandwidth	> 5 MHz,
(iii) overshoot	unspecified,
(iv) load impedance	1 MΩ + 30 pF,
(v) output voltage capability	> 50 V peak-to-peak.

As the sinusoidal response alone is specified, the first step in the design is to choose an all-pole maximally-flat singularity pattern, for which χ is unity. Hence from Eq. 13.123 the required mid-band gain can be achieved from two stages whose mean realizable $\mathscr{GB}$ is 50 MHz or from three stages for which $\mathscr{GB}$ is 23.2 MHz. Very-high-slope pentodes can better the former specification, so it appears that two such stages are sufficient; the second stage requires a power tube operating at about 40-mA anode current to achieve the specified 50-V p-p output (with some reserve) across the 30-pF load capacitance at 5 MHz.

13.8.1.1 $\mathscr{GB}$ Considerations

Tube types EF184/6EJ7 and EL84/6BQ5 are chosen for the first and second stages, respectively, and both are operated near maximum ratings to achieve the largest possible mutual conductance (Section 13.3.1):

FIRST STAGE

$$I_K = 14 \text{ mA},$$
$$V_{SK} = 170 \text{ V}.$$

Therefore

$$g_m = 15.6 \text{ mA/V},$$
$$I_A = 10 \text{ mA},$$
$$I_S = 4 \text{ mA}.$$

SECOND STAGE

$$I_K = 45 \text{ mA},$$
$$V_{SK} = 300 \text{ V}.$$

Therefore

$$g_m = 10.2 \text{ mA/V},$$
$$I_A = 40 \text{ mA},$$
$$I_S = 5 \text{ mA}.$$

The realizable gain-bandwidth products are now calculated from Eq. 13.48:

FIRST STAGE

C_o (c_{out} of EF184)	3.0 pF
C_S (an intelligent guess)	15.0 pF
C_i (c_{in} of EL84)	10.8 pF

Therefore

$$\mathscr{GB} = 5.42 \times 10^8 \text{ radian/sec} = 86.2 \text{ MHz}.$$

SECOND STAGE

C_o (c_{out} of EL84)	6.5 pF
C_S (an intelligent guess)	7.5 pF
C_L (data)	30.0 pF

Therefore

$$\mathcal{GB} = 2.32 \times 10^8 \text{ radian/sec} = 36.9 \text{ MHz}.$$

Hence the bandwidth attainable with two-terminal peaking and a maximally-flat response follows from Eq. 13.123 as

$$\mathcal{B} = 3.56 \times 10^7 \text{ radian/sec} = 5.64 \text{ MHz}.$$

As there are two devices and two-terminal peaking is to be used, the response has two poles. For a maximally-flat response, the poles lie on a circle of radius $|s| = 3.56 \times 10^7$ radian/sec, and are at an angle of 45° in the left half-plane.

13.8.1.2 Detailed Design

The output circuit uses shunt inductance peaking to give the desired poles at 45° in the left half-plane, plus a zero on the negative real axis. In addition, the output stage has cathode peaking to apply feedback and reduce the nonlinearity. The cathode-circuit pole is positioned to cancel the output-circuit zero, and the cathode-circuit zero is canceled by a pole from the unpeaked load network of the first stage. Figure 13.31*a* shows the elemental circuit diagram, and Fig. 13.31*b* shows the individual singularity patterns. There is no particular reason for inductance peaking the output circuit rather than the interstage load network. The component values and reference equations are:

OUTPUT CIRCUIT

m in Eq. 13.62 (for 45° poles)	0.5
τ in Eq. 13.62 (for 3.56×10^7 radian/sec pole-circle radius)	39.7 nsec

Hence

R_{A2} (modified Eq. 13.63)	903 Ω
L (modified Eq. 13.64)	17.9 μH

and the zero in Eq. 13.62 lies at

$$\sigma = -5.02 \times 10^7 \text{ radian/sec}.$$

CATHODE CIRCUIT

R_F (for self biasing)	270 Ω
C_F (pole in Eq. 13.80)	277 pF

Hence the cathode-circuit zero is at

$$\sigma = -1.34 \times 10^7 \text{ radian/sec}.$$

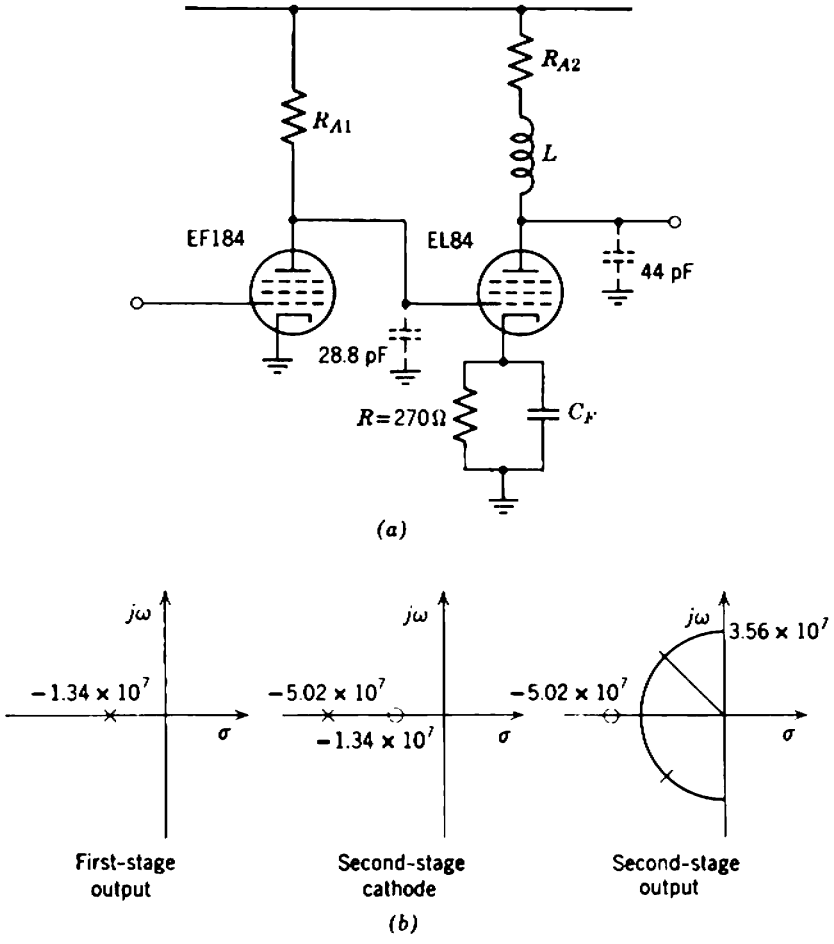


Fig. 13.31 Wide-band vacuum-tube amplifier: (*a*) elemental circuit diagram; (*b*) singularity patterns.

INTERSTAGE CIRCUIT

$$R_{A1} \text{ (pole in Eq. 7.55)} \qquad 2.59 \text{ k}\Omega.$$

The mid-band voltage gain of the first stage is

$$A_{Vm} = -g_m \times R_{A1} = -40.4,$$

whereas the mid-band gain of the second stage is

$$A_{Vm} = -G_T \times R_{A2}$$

and, using Eq. 13.81,

$$A_V = -2.45.$$

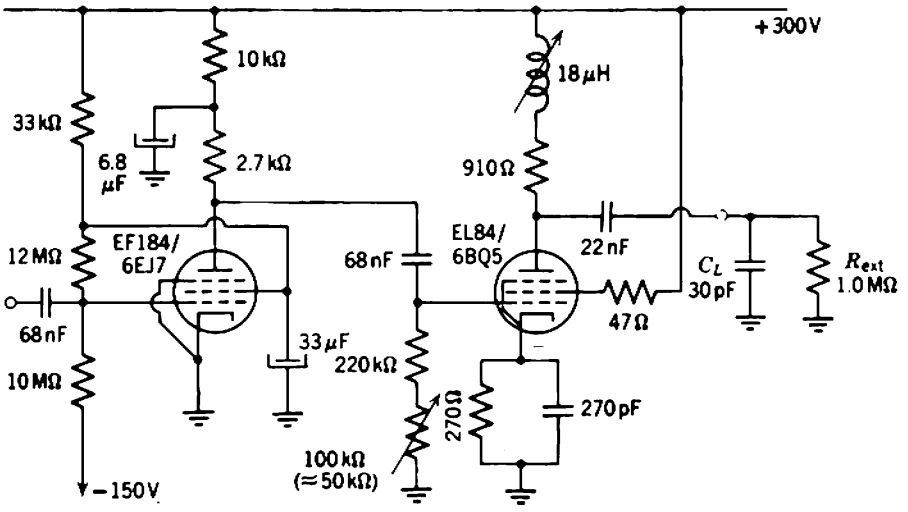


Fig. 13.32 Complete wide-band vacuum-tube amplifier.

By multiplication, the over-all mid-band gain 100, as it must of necessity be for the specified 5.64 MHz bandwidth and the device realizable gain-bandwidth products. Figure 13.32 shows the complete circuit diagram with components corrected to the nearest preferred value; the biasing system and low-frequency capacitors are designed in Section 12.4.1.

13.8.2 Transistor Amplifier

(i) 40-dB gain between 75-Ω resistive terminations,
(ii) overshoot < 1%,
(iii) bandwidth as large as possible, using four type OC170/2N1516 transistors.

As there are four devices and two-terminal feedback peaking is to be used (because it is the most satisfactory for transistors), the first step in the design is to choose a four-pole linear-phase transfer function with several other nondominant singularities. Figure 13.33*a* shows the proposed elemental circuit diagram. Q_1 to Q_3 are operated at 5-mA emitter current for which τ_T is near its minimum value 1.2 nsec. Q_4 is operated at a larger current, 10 mA, to increase the output voltage available into the 75-Ω load, and from manufacturer's data (in Problem 13.3), τ_T is again 1.2 nsec. The corresponding value of ω_T is 8.3×10^8 radian/sec for all transistors.

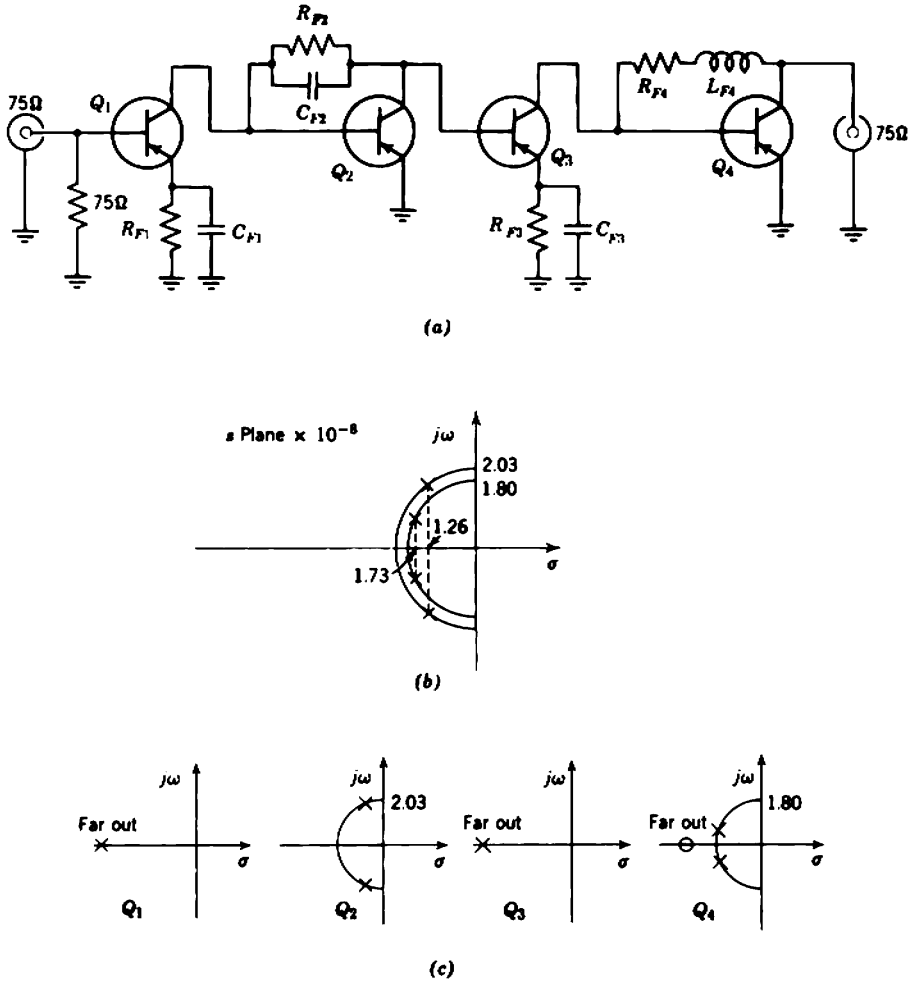


Fig. 13.33 Wide-band transistor amplifier: (*a*) elemental circuit diagram; (*b*) dominant singularity pattern; (*c*) stage singularity patterns.

13.8.2.1 $\mathscr{GM}$ Considerations

The realizable $\mathscr{GM}$ for transistors in the alternate cascade is somewhat less than ω_T, but the exact value depends on the detail of the design and cannot be found until the design is completed. Therefore the attainable bandwidth cannot be predicted exactly from the mid-band gain as in the vacuum-tube amplifier described above. An intelligent guess is that the realizable $\mathscr{GM}$ is about 70% of ω_T, that is, about 6.0×10^8 radian/sec. The bandwidth reduction factor χ for a four-pole linear-phase function is

read from Table 13.3 as 0.659, and therefore from Eq. 13.123 the attainable bandwidth is

$$\mathscr{B} \approx 1.26 \times 10^8 \text{ radian/sec} = 20.0 \text{ MHz}.$$

Figure 13.33*b* shows the proposed pattern for the dominant singularities.

The inductively peaked output transistor Q_4 should produce the pole pair closer to the real axis, as the associated zero is then further from the origin. The capacitively peaked second stage Q_2 should produce the pole pair closer to the imaginary axis (i.e., having the smaller real part) as the discussion leading to Eq. 13.108 shows that the realizable $\mathscr{GB}$ is thereby increased. The individual stage singularity patterns are shown in Fig. 13.33*c*; the poles from the two series stages and the zero from the output shunt stage are nondominant, and are ignored in a first-order design.

13.8.2.2 Detailed Design

The output stage has $R_L \approx 60\ \Omega$ (the parallel combination of the 75-Ω external load with the collector supply resistor) and τ_T is 1.2 nsec. The required poles are

$$|s| = 1.80 \times 10^8 \text{ radian/sec},$$

$$\sigma = -1.73 \times 10^8 \text{ radian/sec}.$$

Therefore, from Section 13.5.3.3,

$$\tau_F = 2.90 \text{ nsec},$$

$$R_F = 532\ \Omega,$$

$$L_F = 1.54\ \mu\text{H},$$

$$R_T = 507 \text{ V/A}.$$

There is a zero on the negative real axis at

$$s = -3.46 \times 10^8 \text{ radian/sec}.$$

For the third stage the feedback resistor is chosen as 47 Ω. This choice is somewhat arbitrary. A larger value reduces the transfer conductance, but also reduces the input capacitance. Therefore a larger feedback resistor can be used in the preceding shunt stage to make up the combined current gain, without loss of bandwidth. To a first order the attainable

current gain-bandwidth product of a shunt stage followed by a series stage is independent of R_F for the series stage. Second-order effects, however, reduce the realizable $\mathscr{GB}$ if R_F is either very large or very small. For large values of R_F, the input capacitance becomes so small that Miller effect is significant. For small values of R_F, the series-stage pole moves in close to the origin and is not nondominant. A rule of thumb is that R_F should be of the same order as r_B for the transistor. With R_F as 47 Ω the parameters of the stage follow from Section 13.5.1.2 as:

$$C_F = 25.5 \text{ pF},$$

$$G_T = 18.2 \text{ mA/V},$$

$$C_i = 28.0 \text{ pF}.$$

There is a pole on the negative real axis at

$$s = -8.8 \times 10^8 \text{ radian/sec}.$$

Notice that the value given for C_i includes the Miller capacitance.

For the second stage the poles are required at

$$|s| = 2.03 \times 10^8 \text{ radian/sec},$$

$$\sigma = -1.26 \times 10^8 \text{ radian/sec}.$$

Therefore, from the first term in Eq. 13.104,

$$\frac{C_F}{C_F + C_L} \approx 0.30.$$

But C_L is about 30 pF (including the stray wiring capacitance), and using the accurate form of the equations in Section 13.5.2.2 the parameters of the stage are

$$C_F = 12.8 \text{ pF},$$

$$C_{F(\text{ext})} = 10.8 \text{ pF},$$

$$R_F = 472 \text{ Ω},$$

$$R_T = 467 \text{ V/A}.$$

Neglecting coupling efficiency, the combined mid-band transfer resistance of the second, third, and fourth stages is

$$R_T = R_{T2} \times G_{T3} \times R_{T4} = 4.26 \text{ V/mA}.$$

To make up the over-all voltage gain of 100, the transfer conductance of

the first stage must be 23.4 mA/V. Therefore from Section 13.5.1.2 the components are

$$R_F = 36.0\ \Omega,$$

$$C_F = 33.3\ \text{pF};$$

hence

$$C_i = 36.0\ \text{pF}.$$

The 75-Ω resistive input impedance is obtained with a network of the type shown in Fig. 13.29. The 50-Ω base resistance r_B of the transistor is built up to about 75 Ω by a 27-Ω series resistor, and a 75-Ω resistor and series inductor are connected in shunt with the signal path. With r_B effectively 75 Ω, the pole in the transfer admittance of the first stage is at

$$s = -4.69 \times 10^7 \text{ radian/sec}.$$

Figure 13.34 shows the pattern for the three nondominant singularities. Figures 13.3*b* and *c* can be used to verify that these singularities contribute a total gain increase of 0.12 dB at 20 MHz (the edge of the passband), and 3° phase lag. The total fall in gain due to these singularities does not reach 3 dB until $\omega \approx 10^9$, that is, at about 150 MHz. The approximation in neglecting the nondominant singularities should therefore be excellent.

Figure 13.35*a* shows a practical realization of the amplifier using preferred value components, and adjusted to give 40-dB gain after allowing for coupling efficiency. The biasing system and low-frequency capacitor values are designed in Section 12.4.2. Figure 13.35*b* shows the step response, using unselected production transistors and with $C_{F(\text{ext})}$ for Q_2 set at its nominal value, 10 pF; there is a small overshoot. With C_F increased to 12 pF, the response is not measurably different from that of an ideal 20-MHz four-pole linear-phase function.

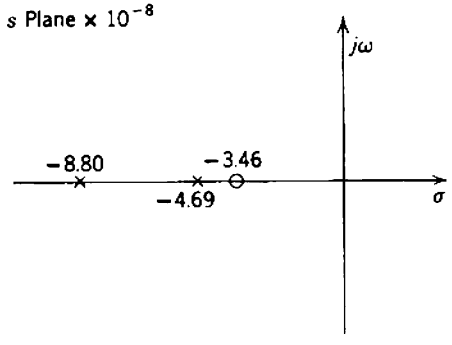


Fig. 13.34 Nondominant singularity pattern for Fig. 13.33.

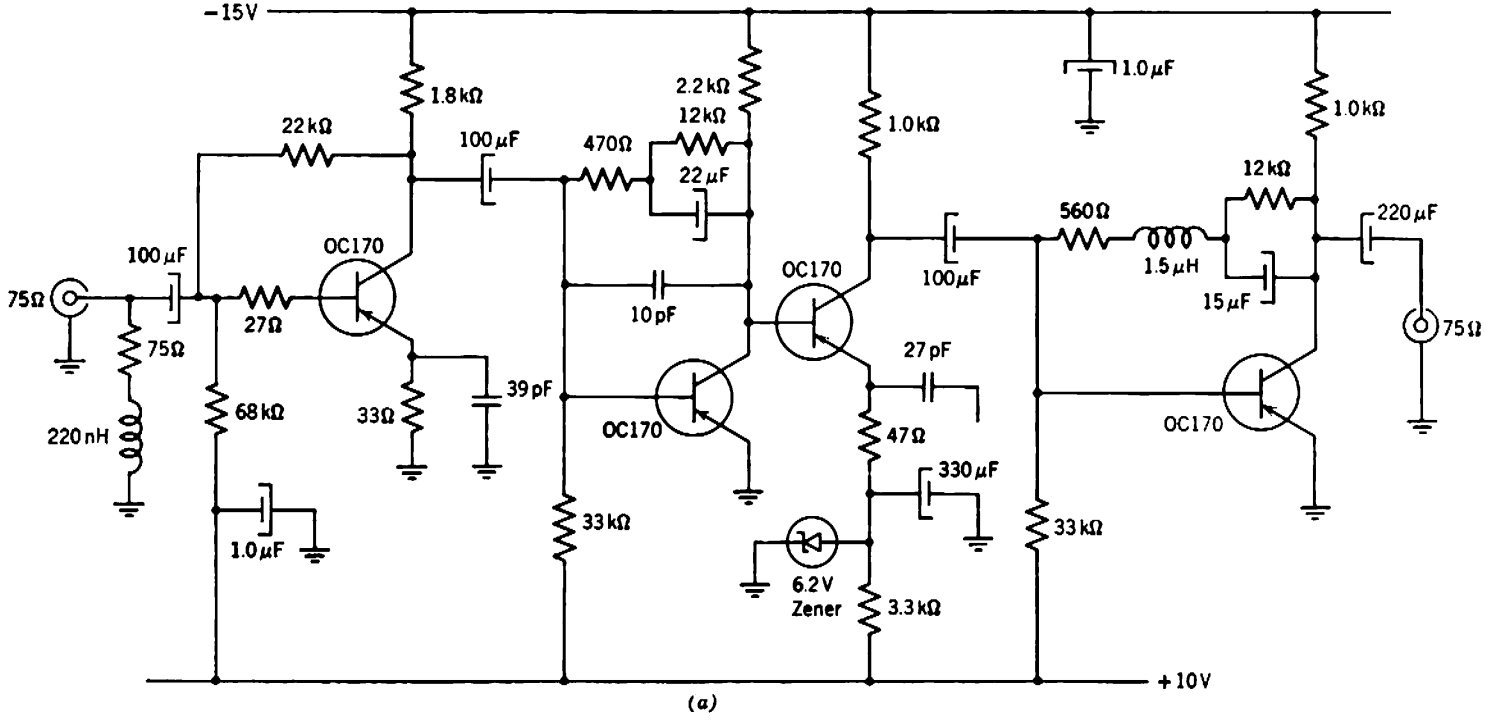


Fig. 13.35 Complete wide-band transistor amplifier: (*a*) circuit diagram; (*b*) step response as designed; (*c*) step response as modified. The horizontal scale in parts (*b*) and (*c*) is 20 nsec/div.

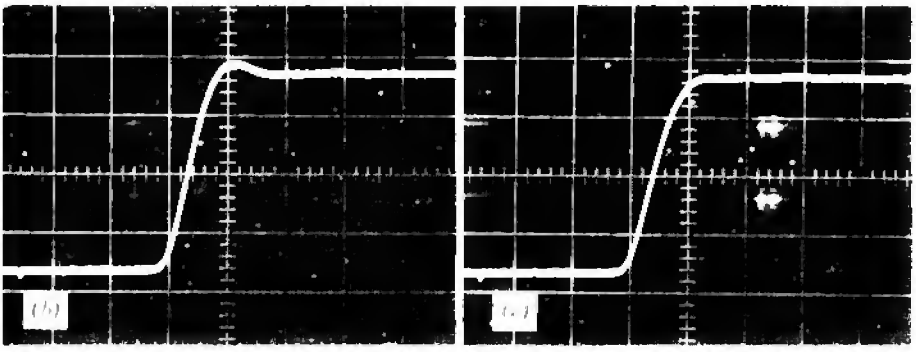


Fig. 13.35 (*continued*)

13.9 VERY-WIDE-BAND AMPLIFIERS

The circuit techniques described in Sections 13.3 to 13.5 are suitable for amplifiers whose individual stage bandwidths do not exceed about one-fifth of the device gain-bandwidth product. This limit arises from two quite separate considerations:

1. At larger bandwidths, the gain per stage becomes very small and a prohibitive number of stages may be required to achieve a specified over-all gain. Indeed, Section 13.6.1.2 shows that it may not be possible to achieve the required combination of gain and bandwidth even if the number of stages can be increased without limit.
2. Approximations in the analysis become poor, so that the bandwidth actually attained falls far short of expectations. This is particularly true of feedback peaking.

As an order of magnitude, amplifiers using readily available tubes can be designed for high gain and bandwidths up to about 20 MHz. For transistors the corresponding limit is about 100 MHz. Although larger bandwidths can be achieved (particularly if the over-all gain is small), the same or better performance can often be obtained from a smaller number of devices in the circuit configurations to be described in this section. However, these new configurations require more active devices when the bandwidth is not large; because of this inefficiency they are unsuitable for the majority of practical amplifiers. This explains why there is no discussion of (say) cathode followers or emitter followers in the main part of the chapter.

13.9.1 Other Stage Configurations

The realizable gain-bandwidth product of an active device always falls short of the intrinsic gain-bandwidth product calculated from charge-control considerations (Section 2.5.1). Losses are attributable to the stray wiring capacitances, the Miller contributions to input and output capacitances, and in the case of a transistor to the base resistance r_B. As a general rule the realizable $\mathscr{GB}$ in an amplifier circuit falls further below the intrinsic $\mathscr{GB}$ as the bandwidth is increased.

A number of useful wide-band circuits combine one or more common-control-electrode stages or common-collecting-electrode stages (Section 5.7) with each common-emitting-electrode stage. These common-control-electrode and common-collecting-electrode stages have unity current gain and voltage gain respectively and so cannot add to the total gain, but they can increase the bandwidth. The philosophy underlying their use is that it can be more efficient to increase the realizable $\mathscr{GB}$ of one common-emitting-electrode stage by means of several unity-gain auxiliary stages, rather than to achieve gain but only a small $\mathscr{GB}$ from all stages. Obviously, this will be the case only if the attainable gain per stage is small, because the increase in realizable $\mathscr{GB}$ cannot be very large. Therefore, the use of unity-gain stages cannot possibly be efficient at small bandwidths where the gain per stage can be large.

As an introduction to this type of circuit, Fig. 13.36 shows the transistor version of the *cascode*, a circuit which substantially eliminates the loss of realizable $\mathscr{GB}$ due to Miller effect in triodes and transistors. Quite apart from this application to very-wide-band circuits, the vacuum-tube cascode is extremely useful in low-noise amplifiers at moderate bandwidths (see Problem 8.1). In Fig. 13.36*a*, Q_1 is a common-emitter amplifying stage and Q_2 is an auxiliary common-base stage. The two devices in a cascode are usually of the same type; the two halves of a twin triode are often used in a vacuum-tube cascode. Moreover, the two devices are usually connected in series for dc as well as for signals (Fig. 13.36*b*) so the quiescent conditions and small-signal parameters are identical.

The load resistance for the common-emitting-electrode stage of a cascode is the input resistance of the common-control-electrode stage. Section 5.7.1 shows that this input resistance is small. Therefore the voltage gain and Miller input capacitance of the common emitting-electrode stage are small also. From Eq. 5.106 the input resistance of Q_2 in Fig. 13.36 is

$$R_{i2} = \frac{r_B}{\beta_N} + r_E.$$

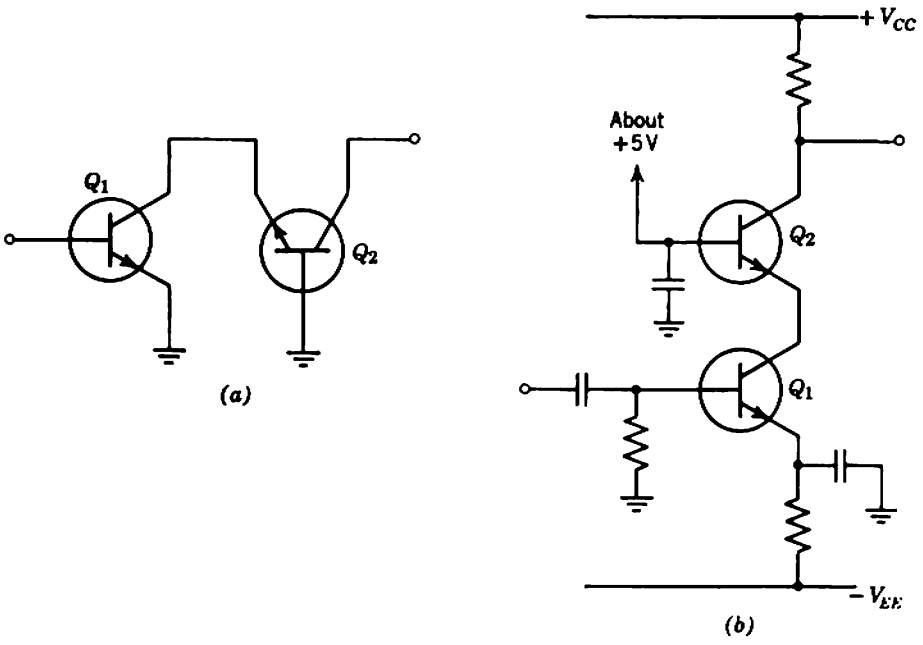


Fig. 13.36 The cascode: (*a*) elemental circuit diagram; (*b*) complete circuit, showing the usual biasing arrangement.

The voltage gain of Q_1 is

$$A_{V1} = -G_{T1} \times R_L,$$

and substituting

$$A_{V1} = -\left(\frac{\alpha_N}{r_B/\beta_N + r_E}\right)(r_B/\beta_N + r_E) = -\alpha_N \approx -1.$$

Hence the Miller input capacitance is

$$c_M = c_{tC}(1 - A_V) \approx 2c_{tC}.$$

The transfer conductance of the complete cascode is

$$G_{T(\text{tot})} = G_{T1} \times A_{I2} = \frac{{\alpha_N}^2}{r_B/\beta_N + r_E},$$

which is almost equal to G_T of a simple common-emitter stage; there is no significant loss of gain through using the cascode.

Many circuits using common-base stages and emitter followers have appeared in the literature. Possibly the most ingenious is Rush's feedback

current amplifier* shown in Fig. 13.37. In outline, the design philosophy and operation are:

Ideally, a transistor has a short-circuit current gain-bandwidth product ω_T, with losses occurring due to r_B, c_{tC}, and stray wiring capacitances.

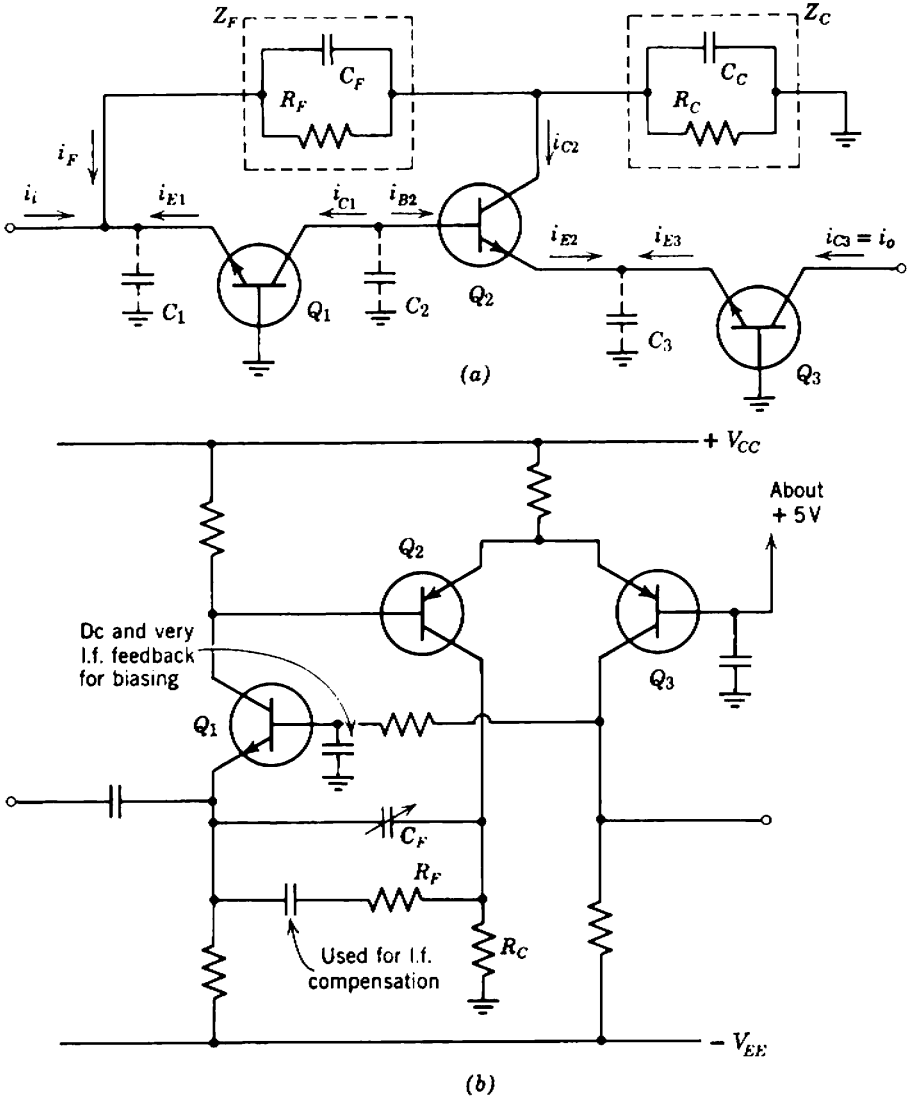


Fig. 13.37 Rush's feedback current amplifier: (*a*) elemental circuit diagram; (*b*) one convenient practical realization.

* C. J. RUSH, "New technique for designing fast-rise transistor pulse amplifiers," *Rev. Sci. Instr.*, **35**, 149, February 1964.

The current amplifying stage in Fig. 13.37 is Q_2 operating in the common-collector configuration for which the mid-band gain is

$$\frac{i_{E2}}{i_{B2}} = \beta_N + 1.$$

However, Q_2 has a low output resistance at its emitter (Eq. 5.121) and is therefore not a good current source; Q_3 is added as a near-unity-current-gain common-base stage to raise the output resistance. Assuming that Q_2 and Q_3 are identical, their collector currents are equal but opposite; the current gain from the base of Q_2 to either collector is therefore

$$\frac{i_{C2}}{i_{B2}} = -\frac{i_{C3}}{i_{B2}} = \beta_N.$$

Loss of realizable $\mathscr{GB}$ due to r_B of Q_2 is minimized by feeding Q_2 from a high-resistance source; Q_1 is a common-base buffer stage with current gain α_N. Loss of $\mathscr{GB}$ due to Miller effect is minimized by operating Q_2 with a small collector load resistance (R_F and R_C in shunt). Loss of $\mathscr{GB}$ due to stray wiring capacitances is small because there is a low impedance to ground at all points; stray capacitances C_1, C_2, and C_3 are in shunt with the relatively large input capacitances of Q_1, Q_2, and Q_3, respectively. Over-all, the current gain from the emitter of Q_1 to the collector of either Q_2 or Q_3 is

$$\frac{i_{C2}}{i_{E1}} = -\frac{i_{C3}}{i_{E1}} = -\alpha_N\beta_N,$$

and the gain-bandwidth product is

$$\mathscr{GB} \approx \omega_T.$$

In addition to the small losses in $\mathscr{GB}$ mentioned above, there are losses due to the α cutoff frequencies of Q_1 and Q_3; these losses are small also, because ω_α is nearly an octave above ω_T for a high-frequency transistor (Sections 4.3.2.1 and 4.7.2.1).

Current gain can be exchanged for bandwidth by means of negative feedback derived from the collector of Q_2 and returned to the emitter of Q_1. The feedback factor is

$$\beta = \frac{i_F}{i_o} = -\frac{i_F}{i_{C2}} = +\frac{Z_C}{Z_F + Z_C}.$$

If the loop gain is large, the mid-band closed-loop current gain is $-1/\beta_m$:

$$A_{Im} = \frac{i_o}{i_i} = -\frac{R_F + R_C}{R_C}.$$

In Fig. 13.37*a* C_C represents the stray capacitance from the collector of Q_2 to ground, where it contributes a lagging phase shift to the feedback loop. C_F is added in shunt with R_F to contribute a leading phase shift to the loop and modify the damping ratio of the closed-loop poles.

Figure 13.37*b* shows a convenient practical form of the circuit using *p-n-p* and *n-p-n* transistors in combination. In his original paper Rush describes a stage with a gain of 4.5 and a rise time (without overshoot) of 1.2 nsec.

In concluding this outline of other useful stage configurations it is worth enlarging on a statement made in the introduction. The use of common-base stages or emitter followers at moderate bandwidths cannot result in a more efficient amplifier than one designed in accordance with Sections 13.3 to 13.5. Indeed there is almost always a loss of efficiency. There is also a loss of designability—designability being a measure of the extent to which a circuit can be designed on paper and its performance predicted. These facts are not at all widely recognized. Notice, however, that common-base stages and emitter followers are useful in integrated-circuit amplifiers; transistors occupy less chip area than resistors and capacitors, and it follows that an inefficient circuit which uses many transistors but few passive components is more economical than an efficient circuit with few transistors but elaborate *RC* peaking.

13.9.2 Distributed Amplifiers

Section 13.6.1.2 shows that there is an upper limit to the bandwidth which can be obtained from a conventional type of amplifier having a specified gain, even if the number of stages can be increased indefinitely. This limit is not of much practical importance, because transistors with gain-bandwidth products of 1 GHz are available readily, and at the time of writing (1967) 3 GHz is available on a limited production basis. Nevertheless, it is an interesting philosophical point whether the $\mathscr{GB}$ limit is truly fundamental. Percival's* distributed amplifier is a major invention because it does allow the realizable $\mathscr{GB}$ to be increased without limit. The operation depends on the propagation of electromagnetic waves along a transmission line.

Ideally, the lumped *LC* transmission line shown in Fig. 13.38*a* has a resistive characteristic impedance

$$R_0 = \sqrt{\frac{L}{C}}, \tag{13.136}$$

* W. C. PERCIVAL, British patent 460, 562, July 24, 1935.

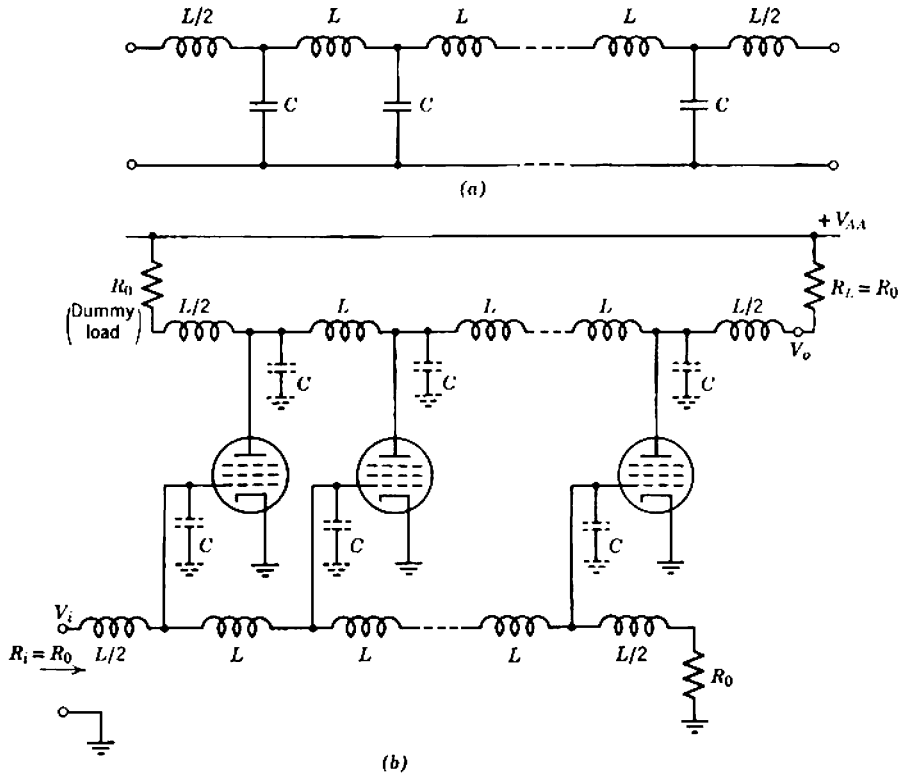


Fig. 13.38 The distributed amplifier: (*a*) lumped *LC* delay line; (*b*) elemental circuit of a distributed amplifier stage.

and a cutoff frequency

$$\omega_{co} = \frac{2}{\sqrt{LC}}. \tag{13.137}$$

A simple lumped line of the type shown is not well behaved for an octave or so below the cutoff frequency, but sophisticated lines with added elements and mutual inductances approach close to the ideal. The vital point on which a distributed amplifier depends is that the cutoff frequency is independent of the line length, and therefore of the total capacitance shunting the terminating resistors.

Figure 13.38*b* shows a distributed amplifier stage. The grids and anodes of a number of tubes are connected together by means of two lumped delay lines, the tube and stray capacitances constituting the shunt capacitances of the lines. If a voltage step is applied to the input, the step wave propagates along the input line without loss and is applied to

each tube in turn; at the end of the line the wave is absorbed in the terminating resistor R_0. Therefore a step occurs in the anode current of each tube in order from the left. Half the signal current from each tube propagates to the left along the output transmission line, until it is absorbed by the dummy load termination. The other half propagates to the right and, if the propagation velocities of the input and output delay lines are equal, all the current steps arrive at the load together; the load current is

$$i_o = v_i \Sigma(\tfrac{1}{2}g_m).$$

Therefore the transfer conductance of the stage is

$$G_T = \frac{i_o}{v_i} = \Sigma(\tfrac{1}{2}g_m) = \left(\frac{n}{2}\right)g_m, \tag{13.138}$$

where the last form applies for n identical tubes. The voltage gain is

$$A_V = -G_T \times R_L = -ng_m\left(\frac{R_0}{2}\right), \tag{13.139}$$

which can be made as large as desired by increasing the number of tubes.

The bandwidth of the amplifier is simply the cutoff frequency of the transmission lines, and the gain-bandwidth product is

$$\mathscr{GB} = |A_V \times \omega_{co}| = n\omega_{co}g_m\left(\frac{R_0}{2}\right).$$

Substituting from Eqs. 13.136 and 13.137,

$$\mathscr{GB} = n\left(\frac{g_m}{C}\right). \tag{13.140}$$

Thus the gain-bandwidth product can be increased without limit by increasing the number of tubes.

As explained in Section 13.3, $\mathscr{GB}$ of a conventional amplifier stage is not changed by connecting a number of devices in parallel. The mutual conductances add, but the capacitances add also and the quotient remains constant. In a distributed amplifier stage the mutual conductances add just as they do for parallel devices in a conventional amplifier, but the capacitances do not add. The devices in a distributed amplifier are spread out along the transmission lines, and their capacitances are separated from each other.

The optimum distributed amplifier for a given gain and bandwidth does not ordinarily consist of just one stage. There is a definite optimum number of stages with an optimum number of devices in each for which the total number of devices is minimum. For this derivation and the practical details of designing a distributed amplifier the reader is referred to specialized literature.*

* A good introductory discussion with a number of references is J. M. PETTIT and M. M. MCWHORTER, *Electronic Amplifier Circuits*, Chapter 6 (McGraw-Hill, New York, 1961).

Chapter 14

Amplifiers with Multistage Feedback

Chapter 11 considers the application of negative feedback to the individual stages of an amplifier, and shows that much greater gain stability can be achieved at mid-band frequencies than can be achieved without feedback. As an order of magnitude, absolute gain stability up to 5% per stage can be achieved fairly easily. Consequent on the increase in gain stability with local feedback is a reduction in nonlinearity. In addition to these advantages at mid-band frequencies, Chapters 12 and 13 show that local feedback can be used to modify and improve the small-signal dynamic response.

Although they are in many ways a great improvement over amplifiers with no feedback at all, amplifiers with local feedback around the individual stages have the disadvantage of poor noise performance. Also, the gain stability and nonlinearity remain inadequate for some purposes. Much improved performance in these respects can be attained through the use of feedback loops that enclose more than one stage, but the price paid is a considerable increase in design complexity, and a reduction in the mid-band gain that can be realized with a given 3-dB bandwidth. In general, there is a trade-off between the advantages and disadvantages as a feedback loop is expanded to enclose more stages. Gain stability and nonlinearity improve, but more design effort is required and the realizable $\mathscr{G}\mathscr{B}$ falls. Section 14.1 contains a discussion.

There are a great many ways of applying feedback to an amplifier. All can be analyzed by the general methods described in Chapter 10, but most feedback amplifiers fall into one of four distinct classes:

1. In maximum feedback amplifiers, the object is to apply as much feedback as possible over a specified frequency band. Such amplifiers are used when the ultimate in gain stability is required. Invariably, the

product of the mid-band gain and 3-dB bandwidth falls well below the value attainable with single-stage feedback or the other high-frequency peaking techniques described in Chapter 13.

2. Amplifiers in which over-all feedback is applied around a group of stages, each of which has local feedback, represent a compromise between the high gain stability of maximum feedback amplifiers and the high gain-bandwidth product of single-stage feedback amplifiers.

3. Feedback pairs are groups of two stages with feedback around them. Some of these groups are useful building-blocks for multistage amplifiers, in many ways akin to the single-stage feedback amplifiers of Chapter 11.

4. Audio power amplifiers are typical of a class in which a relatively small amount of feedback is applied over several stages. No attempt is made to maximize the feedback or to approach the limiting gain-bandwidth product.

These classes of feedback amplifier are discussed in Section 14.2 to 14.5. respectively.

14.1 ADVANTAGES AND DISADVANTAGES OF MULTISTAGE FEEDBACK

This section contains three theorems relating to the gain stability, noise, and gain-bandwidth product of multistage feedback amplifiers. The given proofs are in some respects idealized (for example, they assume identical stages) and they lack rigor. These simplifications are introduced to highlight the factual results and avoid obscuring the physical principles in a mass of algebra. Mathematically inclined readers may make good any deficiencies.

14.1.1 Maximum Gain Stability

If a given selection of active devices is to be used in a feedback amplifier to achieve a specified over-all gain, the greatest gain stability occurs when there is just one feedback loop that encloses all stages, and the least gain stability occurs when local feedback is applied to the individual stages. More generally, the stability increases monotonically as the number of feedback loops is reduced, and therefore as the number of stages within each loop is increased.

The proof depends on Section 10.4.3. Consider an amplifier consisting of n identical stages each of gain A_0, and suppose there is no interaction between stages. Feedback is applied to the amplifier in configurations such as those shown in Fig. 14.1; in each case the total gain A_{tot} is to be

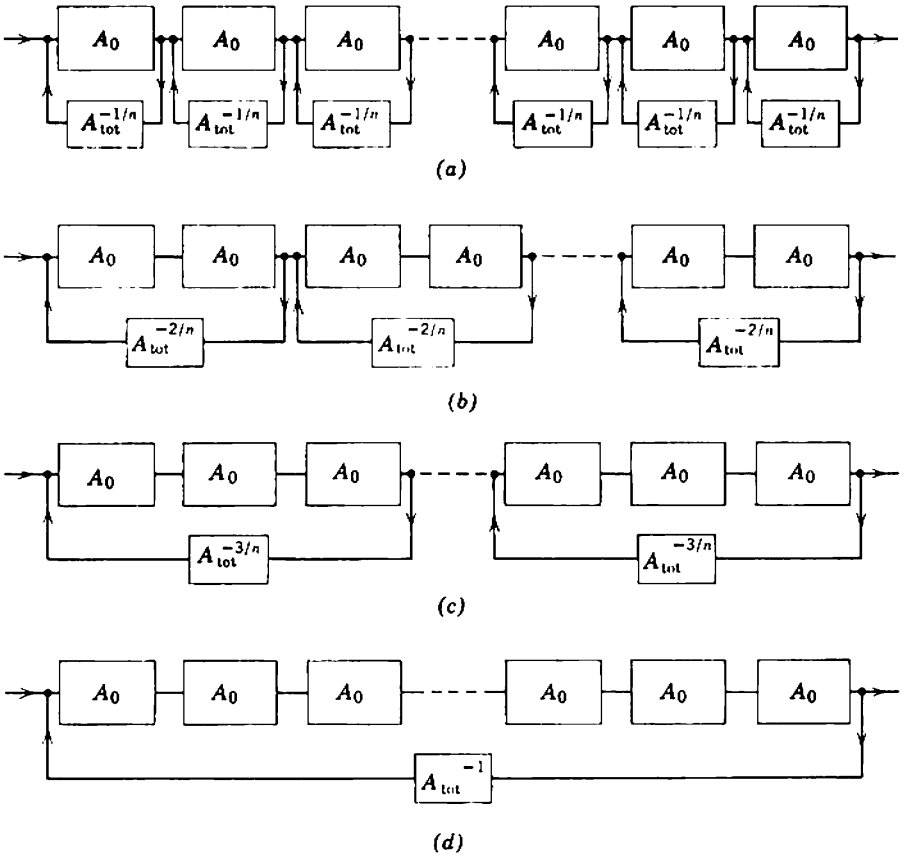


Fig. 14.1 Multistage feedback amplifiers: (*a*) feedback around the individual stages; (*b*) feedback around groups of two stages; (*c*) feedback around groups of three stages; (*d*) feedback around the complete amplifier.

the same. It is required to prove that the sensitivity of the total gain to a change in the gain of any one of the stages falls as the circuit progresses from part (*a*) to part (*d*) of the diagram. Section 10.4.3 defines sensitivity S as the ratio of the fractional change in closed-loop transfer function to the change in any reference variable μ that produces it, and shows that

$$S_\mu = \frac{\Delta TF/TF}{\Delta\mu/\mu} = \frac{1}{1 - A_l} = \frac{1}{F}. \tag{14.1}$$

To prove the theorem it is necessary to evaluate S when the gain A_0 of any one of the stages is chosen as the reference variable μ.

Suppose the amplifier is divided into k identical groups, each containing n/k stages and each having a feedback loop around it. The open-loop gain of each of the k groups is ${A_0}^{n/k}$, and the closed-loop gain is

$$A_k = \frac{{A_0}^{n/k}}{1 - A_l} = \frac{{A_0}^{n/k}}{F_k},$$

where F_k is the return difference. As the total gain is to be independent of k, it is necessary that

$${A_k}^k = A_{\text{tot}}.$$

Hence

$$\frac{{A_0}^n}{{F_k}^k} = A_{\text{tot}}$$

and by rearranging

$$\frac{1}{F_k} = \left(\frac{A_{\text{tot}}}{{A_0}^n}\right)^{1/k} \tag{14.2}$$

Because A_{tot} (the total gain with feedback) is less than ${A_0}^n$ (the total gain of n stages with no feedback at all), the value of $1/F_k$ decreases with increasing k; that is, the feedback around each group increases as the amplifier is divided into fewer groups.

Comparison of Eqs. 14.1 and 14.2 shows that the sensitivity of each group in the amplifier falls as the number of groups is reduced. Now, the sensitivity of an amplifier that consists of a number of groups of stages (such as Figs. 14.1*a* to 14.1*c*) to changes within any one of the groups is equal to the sensitivity of that group alone. This follows because both ΔTF and TF in Eq. 14.1 as applied to the group are multiplied by the combined gain of all other groups when Eq. 14.1 is applied to the complete amplifier. Therefore the sensitivity of the complete amplifier falls as k, the number of groups, is reduced. This completes the proof.

14.1.1.1 Some Common Fallacies

Notice that the foregoing result has been established subject to two assumptions:

(i) regeneration does not occur.
(ii) the stages do not interact.

These assumptions are often overlooked and the theorem is misquoted as:

> Feedback should always be applied over the largest possible number of stages.

When the possibility of regeneration is admitted, it follows that a loop gain and gain stability of the order predicted in Section 14.1.1 can be maintained over a small bandwidth only. Maximum feedback in a single-loop

amplifier of the type shown in Fig. 14.1*d* is, in fact, maintained over a specified bandwidth when the amplifier contains a finite, small, optimum number of stages; Section 14.2.1.2 contains the derivation. Section 14.2.2.3 shows that more feedback can be maintained over the same bandwidth if the amplifier is split into a cascade of feedback loops as in Figs. 14.1*a* to *c*. Finally, Section 14.1.3 shows that the largest possible closed-loop gain-bandwidth product is realized from a cascade of stages with individual feedback loops (Fig. 14.1*a*). Thus the result of Section 14.1.1 is not valid for wide-band amplifiers.

Interaction in the form of capacitive loading occurs between the stages of an amplifier at high frequencies. Again the result of Section 14.1.1 is invalidated, with a suggestion that an optimum wide-band amplifier might not have its feedback confined to a single over-all loop. Interaction can also occur at low frequencies. The outstanding example is the alternate cascade of Chapter 11 (a cascade of stages with local feedback applied to each, Fig. 14.1*a*), for which the gain stability is very nearly as good as for cascaded feedback pairs (Fig. 14.1*b*). Interaction in the alternate cascade occurs as follows:

Consider the transistor stages shown in Fig. 14.2, and suppose the supply resistor R_C is very large (or bootstrapped, Section 11.6.1.1). The loop gain around Q_1 is, from Eq. 11.35,

$$|A_{l1}| = \beta_{N1}\left(\frac{R_{L1}}{R_{F1} + R_{L1}}\right),$$

whereas Eq. 11.28 gives the loop gain around Q_2 as

$$|A_{l2}| = \frac{R_{F2}}{r_{B2}/\beta_{N2} + r_{E2}}.$$

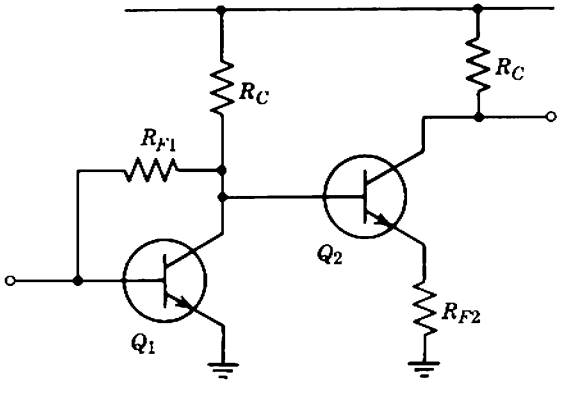


Fig. 14.2 The alternate cascade, an example of interaction between the stages of a feedback amplifier.

At fixed operating conditions the loop gain around Q_2 can be increased only by increasing R_{F2}, that is, by increasing the feedback factor. Consequent on any increase in loop gain is an increase in input resistance:

$$R_{i2} = r_{B2} + \beta_{N2}(r_{E2} + R_{F2}).$$

Now, the input resistance of Q_2 is the load resistance for Q_1:

$$R_{L1} = R_{i2}$$

Increasing the load for a shunt-feedback stage not only increases its loop gain and improves its gain stability, but actually increases the closed-loop transfer resistance. The situation in the alternate cascade is thus very different from that assumed in Fig. 14.1; increasing the feedback around one stage increases both the feedback and the gain of another stage.

14.1.2 Noise

Section 11.5 shows that the series- and shunt-feedback stages suffer from rather poor noise performance. The total mean-square noise voltage at the input of a series stage is the sum of the device noise voltage generator and the thermal noise voltage of the feedback resistor (Eq. 11.91). Similarly, the total mean-square noise current at the input of a shunt stage is the sum of the device noise current generator and the noise current of the feedback resistor (Eq. 11.99). If the loop gain around either stage is other than small, the feedback resistors must be so chosen that their thermal noise is the dominating component of the noise.

Section 10.6 (Fig. 10.26) describes four basic configurations for applying feedback to an amplifier. The present section demonstrates that, provided the amplifiers have enough stages, the noise at the input of these four configurations can be reduced approximately to the noise voltage generator or noise current generator at the input of the first device in the amplifier.

The theory of noise in Fig. 10.26 follows from previous chapters:

(i) Section 10.8 shows that the noise generators at the input of an amplifier are not changed by applying feedback to it.
(ii) The effects of thermal noise R_{F1} in Fig. 10.26*a*, and thermal noise of R_{F2} in Fig. 10.26*d* are discussed in Section 8.3.1 (Eqs. 8.42 to 8.45). The effects of noise of R_{F1} and R_{F2} in Figs. 10.26*b* and *c* follow from straightforward extensions.

The general expressions for the total noise generators at the input are unwieldly but, fortunately, some simplifying approximations are valid.

The feedback at the input of the amplifiers shown in Figs. 10.26*a* and *b* is series and therefore stabilizes a transfer function defined in terms of an input voltage. These amplifiers should therefore be fed from a low-impedance (approximate voltage) source. Under these circumstances, the total noise voltage at the input is substantially equal to the noise voltage generator at the amplifier input. To a good approximation, this generator (like the noise voltage generator of the simple series-feedback stage) is the sum of the noise voltage generator of the first device in the amplifier and the thermal noise of R_{F1}:

$$d(v_{Ni}^2) \approx d(v_N^2) \approx d(v_N^2)_D + 4kTR_{F1}df. \tag{14.3a}$$

A dual conclusion applies to Figs. 10.26*c* and *d*. These amplifiers should be fed from a high-impedance (approximate current) source, and the total noise current at their input is

$$d(i_{Ni}^2) \approx d(i_N^2) \approx d(i_N^2)_D + \frac{4kT}{R_{F2}}df. \tag{14.3b}$$

The difference between the noise performance of multistage feedback amplifiers and series- and shunt-feedback stages stems from the fact that more gain is available in the forward path. Therefore the feedback factor can be reduced while still maintaining a high loop gain. Reducing the feedback factor implies reducing R_{F1} and increasing R_{F2}; either change reduces the thermal noise and, in the limit, the total noise falls to the noise of the first stage in the amplifier. Obviously, the first stage of a low-noise amplifier should never have a local feedback resistor, even in amplifiers of the type that combine local and over-all feedback (Section 14.3); local shunt capacitive feedback, however, does not degrade the noise performance, and may prove useful for locating a dominant pole.

14.1.3 Gain-Bandwidth Product

As a feedback loop is expanded to enclose more stages, the product of the mid-band gain and the attainable 3-dB bandwidth falls further below the product obtainable from single stages with inductance or feedback high-frequency peaking.

The proof depends on the theory of root-locus plotting (Section 10.5.3) and the theory of high-frequency peaking (Section 13.6). The essential concepts are simple:

1. Expansion of the basic feedback equation (Eq. 10.49) shows that the product of the distances from the origin of the effective poles of a multi-

stage feedback amplifier is equal to product of the realizable device gain-bandwidth products divided by the mid-band gain. In other words, Eq. 13.121 applies for multistage feedback amplifiers. Therefore the bandwidth reduction factor χ of the closed-loop singularity pattern is the determining factor in the bandwidth of a feedback amplifier.

2. Paragraph 5 in Section 13.6.1 shows that the bandwidth reduction factor χ tends to be small for singularity patterns in which some poles dominate while others are relatively far from the origin.

3. Rule 5 for root-locus plotting (Eq. 10.64) shows that at large values of loop gain the poles radiate outward at the angles

$$\theta = \frac{(2n + 1)\pi}{n_p - n_z}.$$

Some of the poles move to the left of the complex frequency plane where they become nondominant. The singularity pattern becomes a type for which χ is small.

It is left as an exercise for readers to complete the details of this proof.

14.2 MAXIMUM FEEDBACK AMPLIFIERS

Maximum feedback amplifiers are used when the greatest possible gain stability is required. In these amplifiers a very high loop gain is maintained over a specified signal bandwidth. The closed-loop 3-dB bandwidth of the amplifier is actually much larger than the signal bandwidth, but the upper and lower regions where the loop-gain magnitude falls away are not used. Peaks in this part of the frequency response are of no consequence, so the criterion of satisfactory dynamic response is merely that the amplifier should not regenerate (with some safety margin for component tolerances and device aging).

The subject of maximum feedback amplifiers is extremely complicated. For a specified signal bandwidth, relations exist between the attainable loop gain, the number of stages, and the device gain-bandwidth products. There is, in fact, an optimum number of stages at which the attainable feedback is maximum. This section sets out the design principles whereby near-maximum feedback can be obtained simply. For a complete treatment, readers are referred to Bode's famous book.*

* H. W. Bode, *Network Analysis and Feedback Amplifier Design* (van Nostrand, Princeton, N.J., 1945).

14.2.1 Elementary High-Frequency Considerations

A relation much used in this section is that between the rate of fall-off of loop gain with frequency and the excess loop phase shift. This result is a special case of a general theorem for minimum-phase networks*:

At frequencies where the rate of change of $|\psi(j\omega)|$ with frequency is itself not changing rapidly, the phase shift (in degrees) is

$$\underline{/\psi(j\omega)} \approx \frac{[d|\psi(j\omega)|/d\omega]\text{dB/decade}}{20} \times 90°. \tag{14.4}$$

As an illustration of Eq. 14.4, the response at high frequencies of a n-pole system is asymptotic to $-20n$ dB/decade; the phase shift approaches $-90n°$.

The ultimate aim in a maximum feedback amplifier is to apply as much feedback as possible before the onset of regeneration is reached. Some safety margin is allowed, as illustrated in the Nyquist diagram of Fig. 14.3; the excess loop phase shift is kept below

$$\theta = (2 - y)90°$$

until the loop-gain magnitude has fallen to $1/x$. The parameters x and y are known as the *magnitude* and *phase margins* against regeneration. Intuition suggests, and mathematical analysis confirms, that the amount

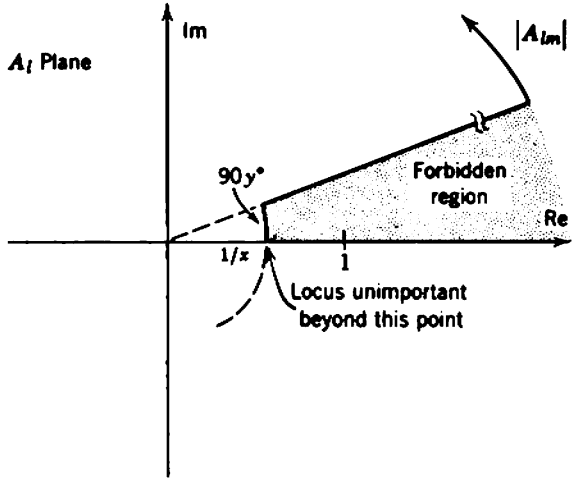


Fig. 14.3 Ideal loop-gain locus for a maximum feedback amplifier, illustrating the margins x and y.

* H. W. Bode, *op. cit.*, Chapter 14.

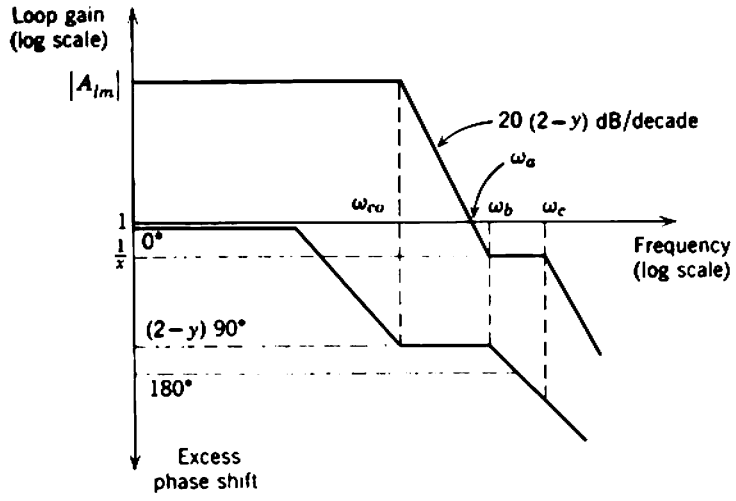


Fig. 14.4 Bode plots for an ideal maximum feedback amplifier. Notice that these are the actual curves, not the asymptotes. The shape of the phase curve below ω_{co}, and the shapes of either curve above ω_c are not important. The magnitude curve between ω_{co} and ω_b is uniquely determined by the phase curve and approximates a $20(2 - y)$-dB/decade slope.

of feedback available becomes larger as the margins x and y are reduced. As a corollary, the available feedback increases as the actual loop-gain locus approaches more closely to the idealized limiting form shown in Fig. 14.3. The immediate aim in a maximum feedback amplifier is therefore to approximate as closely as possible to this idealized locus. Corresponding Bode plots of loop-gain magnitude and excess phase appear as in Fig. 14.4; the loop-gain cutoff frequency ω_{co} is specified together with the margins x and y, and $|A_{lm}|$ is to be as large as possible. Relations between $|A_{lm}|$, x, y, and the various corner frequencies follow from the geometry of the diagram and Eq. 14.4 as:

$$x = \left(\frac{\omega_b}{\omega_a}\right)^{2-y}, \tag{14.5}$$

$$|A_{lm}| = \left(\frac{\omega_a}{\omega_{co}}\right)^{2-y} = \frac{1}{x}\left(\frac{\omega_b}{\omega_{co}}\right)^{2-y} \tag{14.6}$$

14.2.1.1 Outline of Design

The design of a maximum feedback amplifier is outlined in Fig. 14.5; the ideal gain magnitude curve in Fig. 14.4 is taken as the asymptote for the actual curve. Two stages in the amplifier are designed to cut off at ω_{co}, and between them these stages give the loop-gain attenuation rate

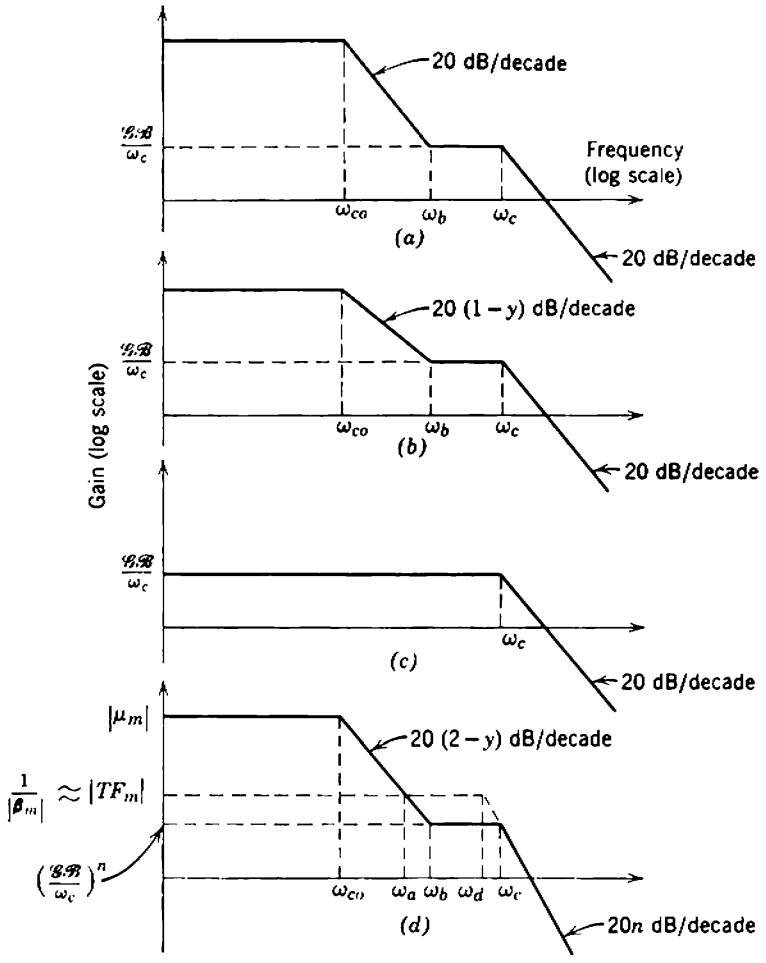


Fig. 14.5 Gain asymptotes for the stages in a maximum feedback amplifier: (*a*) 20 dB/decade stage; (*b*) 20(1 − *y*) dB/decade stage; (*c*) all other stages; (*d*) total gain μ (drawn to a smaller scale).

demanded by the specified phase margin. For example, if $y = \frac{1}{3}$ corresponding to 30° phase margin, the cutoff rate is about −34 dB/decade. The precise detail of this tapered gain roll-off is unimportant for the present purpose and is taken up in Section 14.2.1.3. Suppose one stage has a 20-dB/decade asymptote between ω_{co} and ω_b, whereas the other has a 20(1 − *y*)-dB/decade asymptote. Both stages have horizontal asymptotes between ω_b and ω_c, but beyond ω_c their gains fall off with a single-pole 20-dB/decade slope as in Figs. 14.5*a* and *b*.

More stages are added, each having a single-pole cutoff at ω_c as in Fig. 14.5*c*; the total forward transfer function of the amplifier is the product of the stage gains, and appears as in Fig. 14.5*d*. The frequency ω_c is so chosen that the excess phase shift at frequencies below ω_b, due to all stages in the amplifier, is equal to the specified limiting value $(2 - y)90°$. As more stages are added, the order of the pole at ω_c increases and the ratio between ω_b and ω_c must be increased to keep the total excess phase shift within the prescribed bound; for specified ω_{co} and ω_c, ω_b must be reduced and, from Eq. 14.6, the feedback is reduced. Increasing ω_c, however, increases the amount of feedback that can be applied.

At ω_a, the frequency of unity loop gain,

$$|A_l(j\omega_a)| = |\mu(j\omega_a)| \times |\beta(j\omega_a)| = 1.$$

Hence,

$$|\mu(j\omega_a)| = \left|\frac{1}{\beta(j\omega_a)}\right|.$$

If the feedback factor β is independent of frequency (the usual situation for maximum feedback amplifiers), then

$$|\mu(j\omega_a)| = \left|\frac{1}{\beta_m}\right| \approx |TF_m|. \tag{14.7}$$

In other words, the forward transfer function μ at ω_a is equal in magnitude to the mid-band transfer function with feedback applied. The corner frequency ω_a in Fig. 14.5*d* is therefore related to the closed-loop gain through the number of active devices in the amplifier and their realizable gain-bandwidth products.

Large amounts of feedback require that the order of the pole at ω_c should be low (implying few stages in the amplifier) but that ω_c itself should be large. Gain-bandwidth considerations show that, for the same closed-loop transfer function, ω_a and hence ω_c can be increased only by increasing the number of stages. There are thus two conflicting requirements. The following section derives an optimum number of stages for which the feedback is maximum.

14.2.1.2 Optimum Number of Stages

Consider an n-stage amplifier in which all stages have the same realizable gain-bandwidth product $\mathscr{GB}$. The gain at the horizontal asymptote passing through ω_c is the same for each stage, namely, $\mathscr{GB}/\omega_c$. This value is the mid-band gain of all stages except those associated with the tapered roll-off between ω_{co} and ω_b. Hence the mid-band forward

transfer function of the complete amplifier, including the tapered roll-off $(\omega_b/\omega_{co})^{2-y}$, is

$$|\mu_m| = \left(\frac{\omega_b}{\omega_{co}}\right)^{2-y} \left(\frac{\mathscr{GH}}{\omega_c}\right)^n. \tag{14.8}$$

At any frequency ω somewhat less than ω_b, the phase shift associated with each of the n poles at ω_c is

$$\phi_{\text{pole}} = -\arctan\left(\frac{\omega}{\omega_c}\right) \approx -\left(\frac{\omega}{\omega_c}\right),$$

whereas the phase shift associated with the tapered roll-off is

$$\phi_{\text{taper}} = -(2-y)\left[\frac{\pi}{2} - \arctan\left(\frac{\omega}{\omega_b}\right)\right] \approx -(2-y)\left(\frac{\pi}{2} - \frac{\omega}{\omega_b}\right).$$

Hence the total phase shift at ω is

$$\phi = -n\left(\frac{\omega}{\omega_c}\right) - (2-y)\left(\frac{\pi}{2} - \frac{\omega}{\omega_b}\right). \tag{14.9}$$

The phase margin against regeneration requires that the excess loop phase shift (in degrees) at frequencies below ω_b should be $(2-y)90°$; that is,

$$\phi = -(2-y)\frac{\pi}{2},$$

and substituting,

$$\frac{\omega_b}{\omega_c} = \frac{2-y}{n}. \tag{14.10}$$

The mid-band closed-loop transfer function of any feedback system is

$$|TF_m| = \frac{|\mu_m|}{1 + |A_{lm}|}.$$

If the mid-band loop gain is large (as it is in a maximum feedback amplifier), then

$$|TF_m| \approx \left|\frac{\mu_m}{A_{lm}}\right|. \tag{14.11}$$

Equations 14.6, 14.8, 14.10, and 14.11 can be solved to express $|A_{lm}|$ in terms of the specified parameters TF_m, ω_{co}, x, and y, together with the number of stages n:

$$|A_{lm}| = \frac{1}{x}\left[\left(\frac{\mathscr{GH}}{\omega_{co}}\right)\left(\frac{x}{|TF_m|}\right)^{1/n}\left(\frac{2-y}{n}\right)\right]^{2-y}. \tag{14.12}$$

Differentiation with respect to n yields the optimum number of stages for which $|A_{lm}|$ is maximum as

$$n_{opt} = \ln\left(\frac{|TF_m|}{x}\right). \tag{14.13}$$

14.2.1.3 The Tapered Roll-Off

Over a finite frequency range a gain roll-off at a rate other than a multiple of 20 dB/decade can be obtained from a suitable array of poles and zeros on the negative real axis. For rates between 0 and 20 dB/decade, poles and zeros are interlaced with logarithmic spacing, as in Fig. 14.6. Writing the pole frequencies as

$$\omega_0 \quad p\omega_0 \quad p^2\omega_0 \quad p^3\omega_0 \quad \cdots$$

and the zeros as

$$q\omega_0 \quad pq\omega_0 \quad p^2q\omega_0 \quad p^3q\omega_0 \quad \cdots,$$

the frequency-dependent factor in the gain is

$$\psi(s) = \frac{(1 + s/q\omega_0)(1 + s/pq\omega_0)(1 + s/p^2q\omega_0)\cdots}{(1 + s/\omega_0)(1 + s/p\omega_0)(1 + s/p^2\omega_0)\cdots}. \tag{14.14}$$

If both p and q approach unity so that the number of singularities in a finite frequency range becomes infinite, the asymptotic attenuation rate becomes

$$\frac{d|\psi(j\omega)|}{d\omega} = -20\left(\frac{\log_{10} q}{\log_{10} p}\right) \text{ dB/decade} \tag{14.15}$$

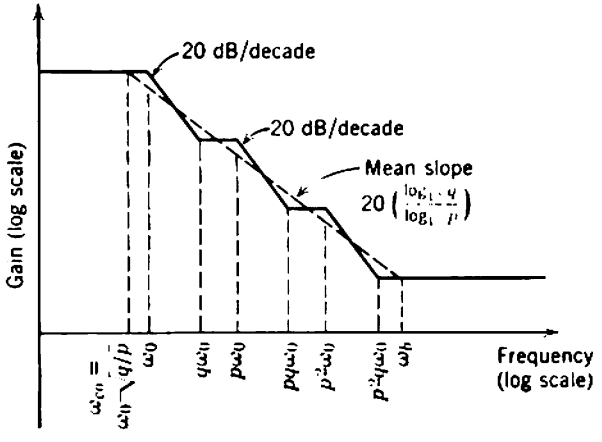


Fig. 14.6 One method for approximating a gain roll-off at a rate between 0 and 20 dB/decade.

and the asymptotic phase shift is

$$\underline{/\psi(j\omega)} = -90\left(\frac{\log_{10} q}{\log_{10} p}\right) \text{ degrees.} \tag{14.16}$$

Equations 14.14 and 14.15 hold quite well for p as large as 10, corresponding to poles one decade apart. Notice that the magnitude asymptote does not pass through ω_0 (the frequency of the first pole), but through $\omega_0\sqrt{q/p}$. Figure 14.7 shows a plot of the magnitude, for three poles and three zeros with

$$p = 10, \qquad q = 5.$$

For rates of attenuation greater than 20 dB/decade, more poles are added at $\omega_0\sqrt{q/p}$. Each additional pole contributes 20 dB/decade and 90° to the asymptotes.

The voltage gain of any amplifier stage can be expressed as

$$A_V(s) = -g_m \times Z_L(s).$$

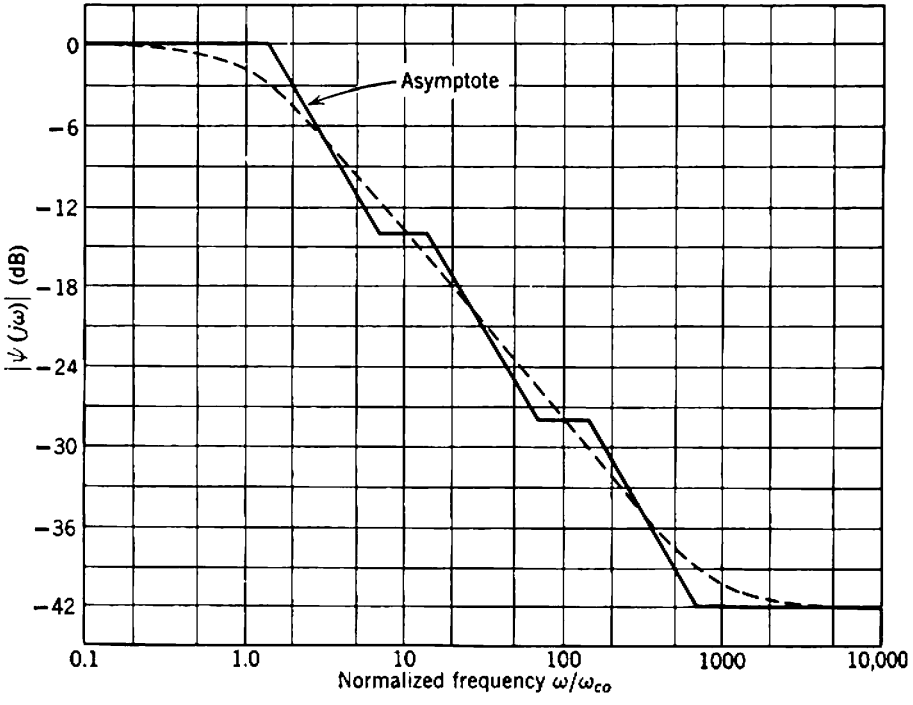


Fig. 14.7 Example of a tapered roll-off with poles a decade apart:

$$\omega_0 = \sqrt{2} \qquad p = 10$$
$$\omega_{co} = 1 \qquad q = 5$$

Because g_m for a charge-controlled device is sensibly independent of frequency, the singularities of A_V are merely those of Z_L. Hence the gain roll-off at $20(2 - y)$ dB/decade for the tapered stages in a maximum feedback amplifier can be attained by suitable design of their passive load impedances. One network whose poles and zeros interlace on the negative real axis and which is treated in many elementary books is the *RC* ladder.

From a practical viewpoint a long *RC* ladder is not a convenient network for achieving the tapered gain roll-off, because finding its singularities involves the solution of a high-order equation. The two-section ladder, which has two poles and one zero, is the largest that can be handled easily. Fortunately, this simple network is sufficient to achieve the roll-off in most practical amplifiers. Although Section 14.2.1.1 describes the tapered roll-off over two stages of a maximum feedback amplifier, there is no reason why more stages should not be involved. Each additional stage makes one more pole and zero available for the taper. As an example, Fig. 14.8 shows how the tapered roll-off might be

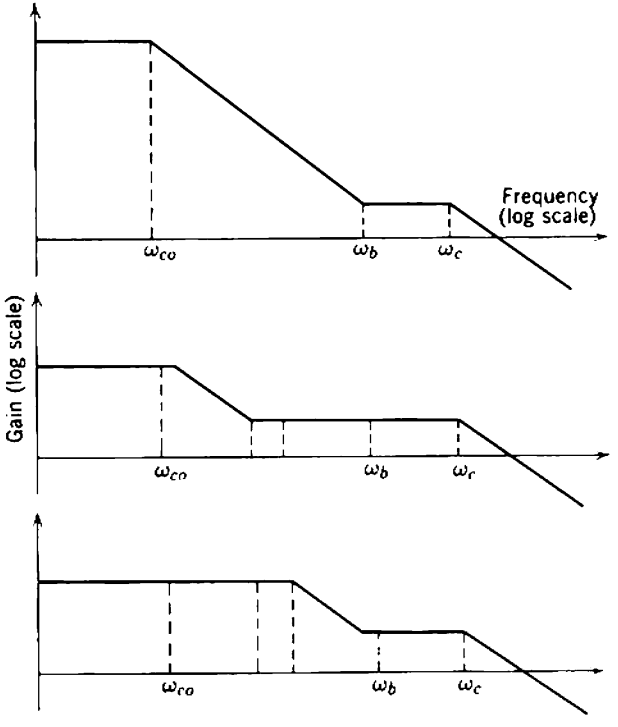


Fig. 14.8 Tapered roll-off in a three-stage maximum feedback amplifier; gain asymptotes of the individual stages. The sloping asymptotes are at 20 dB/decade.

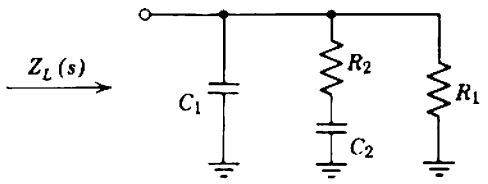


Fig. 14.9 Two-section RC ladder load network, having two poles and one zero interlaced on the negative real axis.

achieved in a three-stage amplifier. One stage contributes a 20-dB/decade roll-off over the entire frequency range from ω_{co} to ω_b. The other two stages approximate the remaining $20(1 - y)$ dB/decade by means of two poles and two zeros; this is an adequate approximation over a two-decade frequency range within which the loop gain falls by $40(2 - y)$ dB. Figure 14.9 shows the form of the two-section ladder load network used in each stage; C_1 represents the irreducible stray capacitance, R_1 represents the parallel combination of the dc supply resistors with device output and input resistances, and R_2 and C_2 are added components. The total load impedance is

$$Z_L(s) = R_1\left[\frac{1 + s(R_2C_2)}{1 + s(R_1C_1 + R_1C_2 + R_2C_2) + s^2(R_1R_2C_1C_2)}\right] \quad (14.17)$$

from which the poles and zero follow immediately. Designating the singularities by p_1, p_2, and z, the values of the components are given in terms of the stray capacitance C_1 by:

$$C_2 = \left[\frac{(z - p_1)(p_2 - z)}{z^2}\right]C_1, \quad (14.18a)$$

$$R_1 = -\left(\frac{z}{p_1p_2}\right)\frac{1}{C_1}, \quad (14.18b)$$

$$R_2 = -\left[\frac{z}{(z - p_1)(p_2 - z)}\right]\frac{1}{C_1}. \quad (14.18c)$$

14.2.1.4 Some Numerical Values

Two approximations are made in the theoretical development of Section 14.2.1.2: actual values of loop-gain magnitude are replaced by their asymptote, and arctangents of angles are replaced by the angles themselves. This section makes a numerical comparison between the actual loop-gain locus of an amplifier and the idealized locus shown in Fig. 14.3.

Suppose an amplifier is designed with magnitude and phase margins

$$x = 2, \qquad y = \tfrac{1}{3},$$

against regeneration. Ideally, the loop-gain locus in the vicinity of $|A_l| = 1$ should appear as shown in Fig. 14.10. Suppose further that the amplifier has

$$\frac{\mathscr{GB}}{\omega_{co}} = 100, \qquad |TF_m| = 50,$$

although these numerical choices have little effect on the actual locus. From Eq. 14.13 the optimum number of stages is

$$n_{opt} = 3$$

and substituting into Eqs. 14.12, 14.10, and 14.6 gives

$$|A_{lm}| = 68.0,$$

$$\frac{\omega_c}{\omega_b} = 1.80,$$

$$\frac{\omega_b}{\omega_{co}} = 19.1.$$

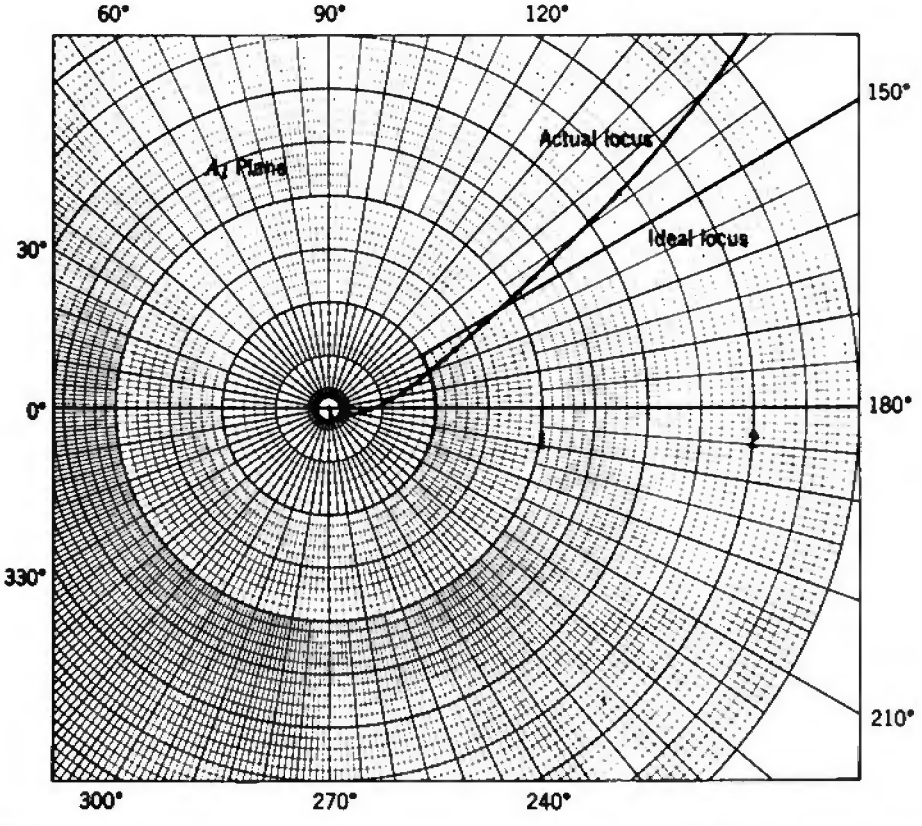


Fig. 14.10 Comparison of the actual and ideal loop-gain loci for a maximum feedback amplifier. Drawn for three stages, with $|A_{lm}| = 68$, $x = 2$, $y = \frac{1}{2}$.

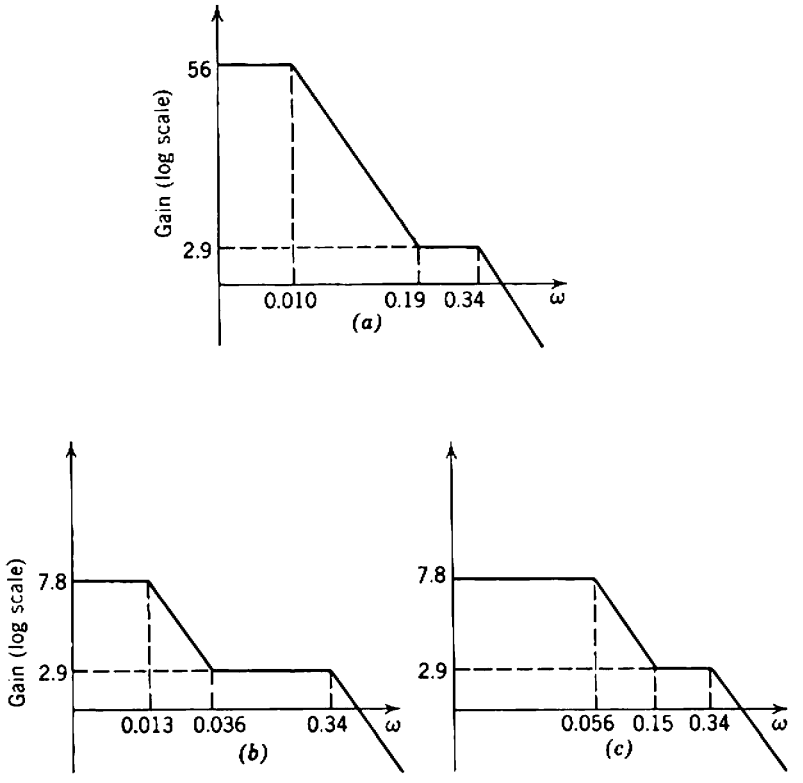


Fig. 14.11 Stage gain asymptotes for the loop-gain locus shown in Fig. 14.10. The frequency scale is normalized with respect to $\mathscr{G}\mathscr{B}$.

Figure 14.11 shows the gain asymptotes for the stages, and computed values of loop gain are plotted in Fig. 14.10. On the whole the actual locus is an adequate approximation to the ideal, although it does just enter the forbidden region.

14.2.2 High-Frequency Peaking

The available feedback for an amplifier can be increased substantially through the use of high-frequency peaking. Broadly, three things can be accomplished:

1. In simple amplifiers of the type discussed in Section 14.2.1, the loop gain is down by about 6 dB at the nominal cutoff frequency ω_{co}: 3 dB from the 20-dB/decade stage, and slightly less than 3 dB from the other stages in the taper. High-frequency peaking can "square up" the corner of the loop-gain cutoff.

2. The mid-band gain of any stage can be raised by increasing R_1 in the ladder load network of Fig. 14.9, although this alters the positions of both poles. The pole at ω_c can be restored by changing the values of R_2 and C_2, but the pole near ω_{co} then falls in direct proportion to the increase in mid-band gain. More efficient peaking schemes than those implied in 1 above can restore the bandwidth at the increased gain.

3. At frequencies below ω_c the phase shift due to the ultimate high-frequency roll-off can be reduced. Therefore, ω_c can be reduced relative to ω_b while maintaining the loop phase shift within prescribed bounds and, from gain-bandwidth considerations, more gain becomes available.

Traditionally, inductance peaking is used in preference to feedback peaking (although a fair case can be made in favor of the latter), and two-terminal peaking is preferable to four-terminal. The disadvantage of four-terminal peaking is that there is more than one effective pole per stage, so that the asymptotic phase shift at high frequencies exceeds 90°. This excess phase shift tends to provoke oscillation in a feedback amplifier, and reduces the amount of feedback that can safely be applied. The advantage of two-terminal inductance peaking over feedback peaking is that the singularity pattern is defined more precisely. In an inductance peaked amplifier, the singularity pattern is substantially independent of device parameters although the mid-band gain is proportional to g_m. The opposite is true of feedback peaking; the mid-band gain is known within close limits, but the positions of the poles depend on g_m. The former situation is clearly the more desirable in the forward path of a feedback amplifier, because a fall in $|A_{lm}|$ due to device aging can only increase the magnitude margin x against regeneration. In contrast, a change in the singularity pattern can reduce the phase margin y and provoke regeneration. For this reason the present section deals only with inductance peaked amplifiers. The advantage of feedback peaking is that it improves the stability of the mid-band gain; Section 14.3 discusses amplifiers having both local and over-all feedback.

14.2.2.1 The Corner at ω_{co}

Adding a shunt peaking inductor (Section 13.4.1.1) to the RC ladder load network can give a worthwhile improvement in the shape of the loop-gain cutoff near ω_{co}, although it does not increase the mid-band gain. Only two stages are peaked: the 20-dB/decade stage whose unpeaked pole is at ω_{co}, and the first of the stages involved in the taper whose pole lies near ω_{co}. Nothing is gained by peaking the other stages in the taper because their combined response is already an excellent approximation to the desired $20(1 - y)$-dB/decade roll-off.

Without peaking, the two relevant stages have only one pole each in the vicinity of ω_{co}. Under these conditions Eq. 14.17 shows that the lower-frequency pole p_1 in Fig. 14.9 (the pole near ω_{co}) is

$$p_1 \approx -\frac{1}{R_1(C_1 + C_2)}, \tag{14.19}$$

so that the RC ladder load network can be approximated near ω_{co} by the simplified form shown in Fig. 14.12*a*. A peaking inductor can be added to this simplified network as shown in Fig. 14.12*b*, and the peaking parameter m in Eq. 13.64 is

$$m = \frac{L}{R_1^2(C_1 + C_2)}. \tag{14.20}$$

The complete load network appears as in Fig. 14.12*c*. Notice that R_2 and C_2 may require slight changes to relocate the second pole at its desired position; the peaking inductor open-circuits the R_1 branch of the circuit at high frequencies.

The response of a shunt peaked stage is given by Eq. 13.62 as

$$\psi(s) = \frac{1 + sm\tau}{1 + s\tau + s^2m\tau^2},$$

where $\omega_{co} = 1/\tau$ is the 3-dB cutoff frequency in the absence of peaking.

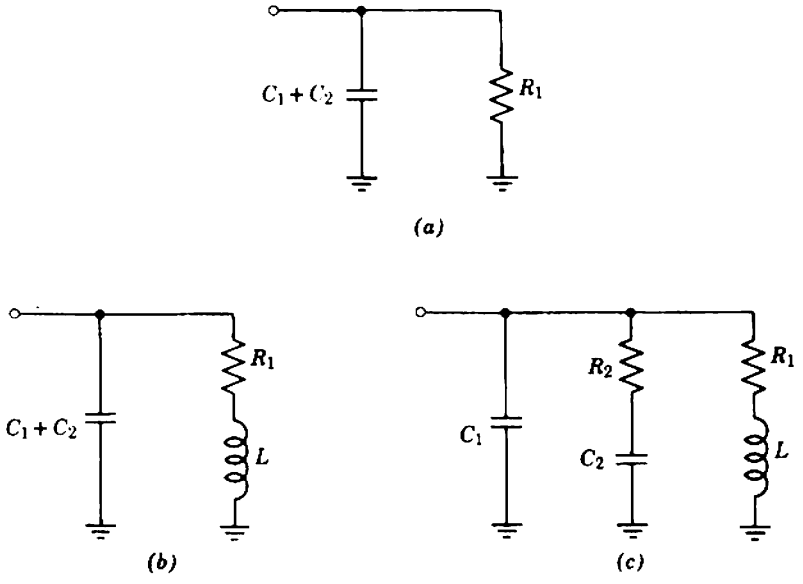


Fig. 14.12 Shunt peaking an RC ladder network near ω_{co}: (*a*) approximate form of the network without peaking; (*b*) approximate form of the network with peaking added; (*c*) complete load network.

Optimum peaking for the two stages in a maximum feedback amplifier corresponds to

$$m = 0.5,$$

for which the response at ω_{co} is

$$|\psi(j\omega_{co})| = 1,$$

with a very small peak at lower frequencies. With this peaking, the loop gain remains constant to ± 0.5 dB for all frequencies up to ω_{co}; beyond this, $|A_l|$ falls rapidly toward the asymptote for an unpeaked amplifier, so no changes are required in the theory of Section 14.2.1.2.

With more efficient peaking, the loop gain can be increased by a factor g over the entire frequency range up to ω_{co}. Equation 14.6 becomes

$$|A_{lm}| = g\left(\frac{\omega_a}{\omega_{co}}\right)^{2-y} = \frac{g}{x}\left(\frac{\omega_b}{\omega_{co}}\right)^{2-y}. \tag{14.21}$$

Bode* performs an exhaustive analysis of the corner in the loop gain characteristic when there is a prescribed effective shunt capacitance C and shows that g has an upper bound:

$$g_{max} = 2^{2-y}. \tag{14.22}$$

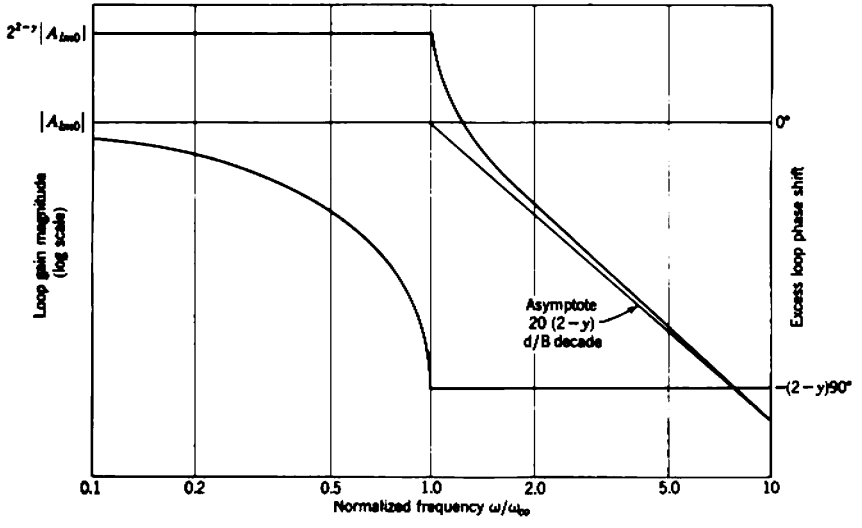


Fig. 14.13 Theoretical upper bound to the increase in loop gain at high frequencies, through high-frequency peaking.

* H. W. Bode, *op. cit.*, Chapters 17 and 18.

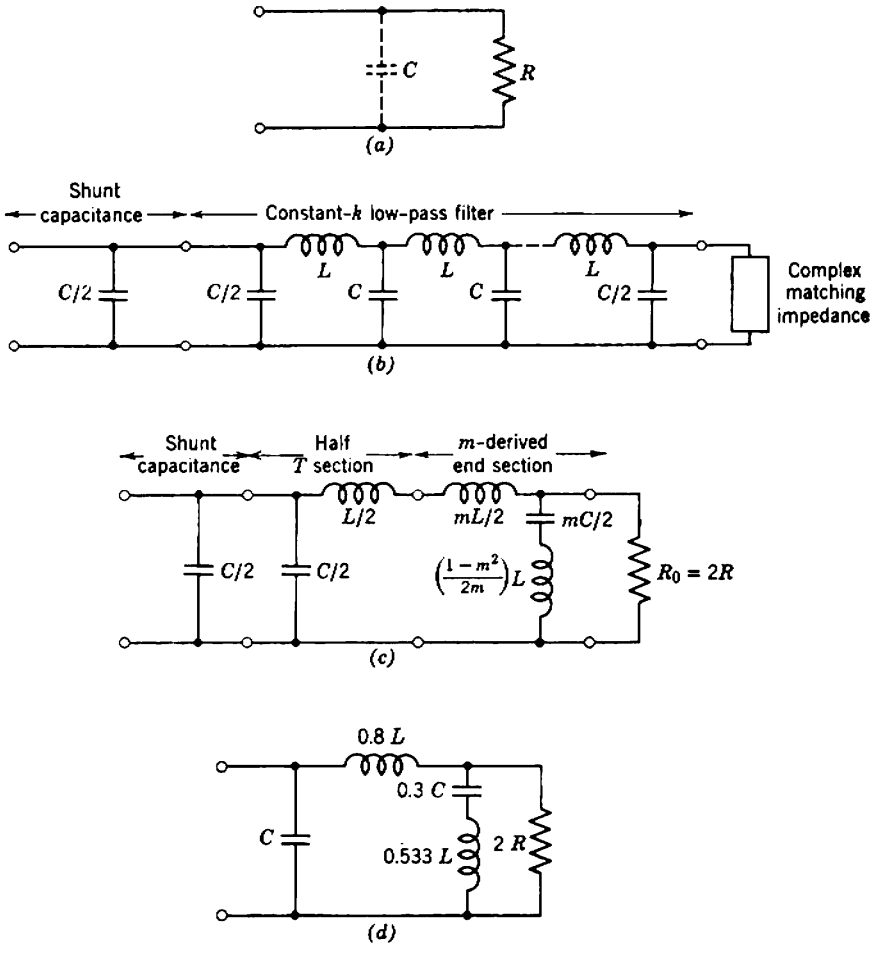


Fig. 14.14 Peaking network for one stage in a maximum feedback amplifier: (*a*) load network without peaking; (*b*) ideal load network involving a correctly matched constant-k filter; (*c*) approximate matching with an m-derived end section; (*d*) practical form of the load network; (*e*) normalized response, with $C = 1\text{F}$, $\omega_{co} = 1$ radian/sec. C represents the total capacitance in a simple stage, or the capacitance effective near ω_{co} in a stage with a ladder load (namely, $C_1 + C_2$).

The loop-gain characteristic with peaking that realizes this advantage is

$$\ln [A_l(j\omega)] = \ln (A_{lm0}) + (2 - y) \ln \left[\frac{2}{\sqrt{1 - (\omega/\omega_{co})^2} + j(\omega/\omega_{co})}\right], \tag{14.23}$$

where A_{lm0} is mid-band loop gain attainable with the same effective capacitance, but without peaking. Figure 14.13 shows this ideal charac-

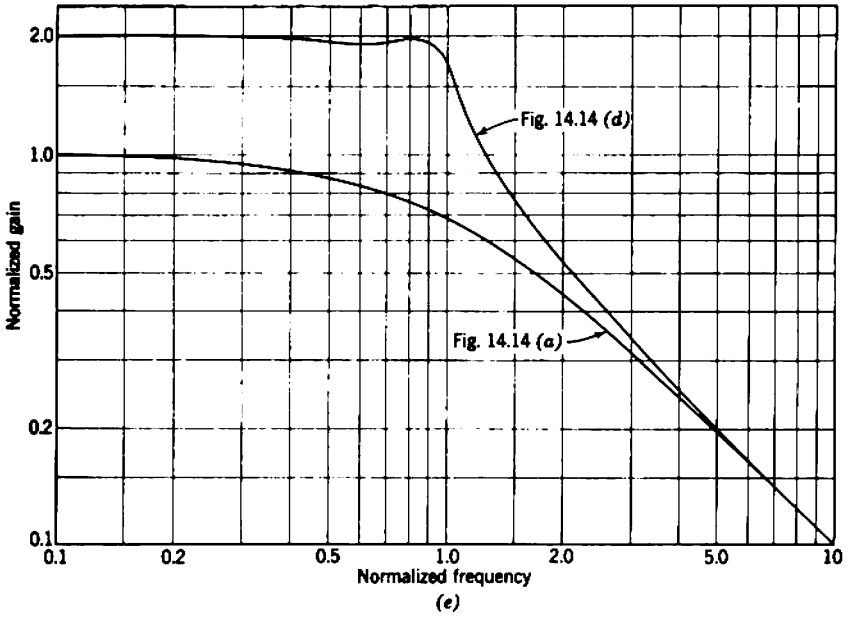


Fig. 14.14 (*continued*)

teristic; an indication of the type of peaking network required to achieve it can be obtained as follows:

From the form of Eq. 14.23 it is apparent that the gain of the 20-dB/decade stage must be

$$\ln [A_V(j\omega)] = \ln (A_{Vm0}) + \ln \left[\frac{2}{\sqrt{1 - (\omega/\omega_{co})^2} + j(\omega/\omega_{co})}\right].$$

A_{Vm0} and ω_{co} are parameters of the stage without peaking, whose load network appears in Fig. 14.14*a*:

$$A_{Vm0} = -g_m R,$$

$$\omega_{co} = \frac{1}{RC},$$

where R and C are the load resistance and shunt capacitance effective near ω_{co}. Substitution and rearrangement yields

$$A_V(j\omega) = \frac{-g_m}{\dfrac{\omega_{co}C}{2}\left[1 - \left(\dfrac{\omega}{\omega_{co}}\right)^2\right]^{1/2} + \dfrac{j\omega C}{2}}.$$

Hence the load admittance for the stage must be

$$Y_L(j\omega) = \frac{\omega_{co}C}{2}\left[1 - \left(\frac{\omega}{\omega_{co}}\right)^2\right]^{1/2} + \frac{j\omega C}{2}. \tag{14.24}$$

The first term in Eq. 14.24 is recognized as the characteristic admittance of a Π-section constant-k low-pass filter in which the first element is $C/2$; the second term is the admittance of a shunt capacitance $C/2$. Therefore the load must be the network shown in Fig. 14.14*b*; the inductances in a constant-k filter are

$$L = \frac{4}{\omega_{co}^2 C} = 4R^2C. \tag{14.25}$$

A simple approximation to the matching impedance for the filter can be obtained through the use of an m-derived end section, with $m = 0.6$. The first of the constant-k Π sections is split in two to yield a half T section, and an m-derived end section is added as in Fig. 14.14*c*. The terminating resistance is

$$R_0 = \sqrt{\frac{L}{C}} = 2R. \tag{14.26}$$

With appropriate elements combined, the complete interstage network appears as in Fig. 14.14*d*, and Fig. 14.14*e* shows the normalized response. Bode describes several other terminating networks that are better approximations to the ideal.

14.2.2.2. The corner at ω_c

The shape of the corner in the loop gain at ω_c can be improved through the use of high-frequency peaking. The object of such peaking is not to "square up" the corner as in the preceding section. Rather the object is to reduce the phase shift at frequencies below ω_c, so that ω_b and hence the loop gain can be increased in accordance with Eq. 14.6.

Although the concepts involved are simple, there are considerable analytical difficulties in finding the optimum number of stages unless two conditions are satisfied:

(i) The peaking must not introduce a peak into the frequency response. Otherwise the loop gain, which falls below unity at ω_a in Fig. 14.4, may rise above unity in the vicinity of ω_c and regeneration can occur.
(ii) The phase shift must approximate a linear function of frequency:

$$\phi_{\text{pole}} \approx h\left(\frac{\omega}{\omega_c}\right). \tag{14.27}$$

The relation between ω_b and ω_c (corresponding to Eq. 14.10 of the analysis without peaking) then becomes

$$\frac{\omega_b}{\omega_c} = \frac{2 - y}{hn}. \tag{14.28}$$

Shunt peaking offers most of the advantages of more complicated peaking systems, and in simple cases satisfies the above two conditions. Provided the higher-frequency pole p_2 in Fig. 14.9 (corresponding to ω_c) is well separated from the other singularities, Eq. 14.17 gives

$$p_2 \approx -\frac{1}{R_2}\left(\frac{C_1 + C_2}{C_1 \times C_2}\right). \tag{14.29}$$

The *RC* ladder load network can therefore be approximated near ω_c by the simple form shown in Fig. 14.15*a*. A peaking inductor can be added as shown in Fig. 14.15*b*, and the peaking parameter is

$$m = \frac{L}{R_2^2\left[\frac{C_1 \times C_2}{C_1 + C_2}\right]}. \tag{14.30}$$

Figure 14.15*c* shows the complete load network. If m is chosen as 0.322, corresponding to a linear-phase response, the constant h in Eqs. 14.27 and 14.28 becomes

$$h_{\text{linear-phase}} = 0.678.$$

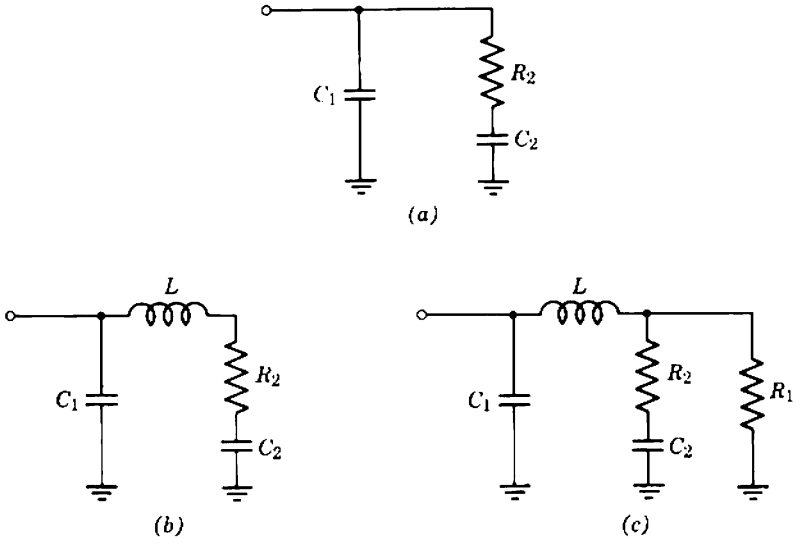


Fig. 14.15 Shunt peaking an *RC* ladder network near ω_c: (*a*) approximate form of the network without peaking; (*b*) approximate form of the network with peaking added; (*c*) complete load network.

In the 20-dB/decade stage, and in the last of the stages involved in the $20(1 - y)$-dB/decade taper, the zero (corresponding to ω_b) lies close to p_2 (corresponding to ω_c) and Eq. 14.29 becomes a poor approximation. Formal analysis becomes tedious. Numerically, however, the value

$$h \approx \tfrac{2}{3}$$

can be achieved in Eq. 14.28 by peaking only the stages that do not have a zero near ω_b. For example, in a three-stage amplifier the 20-dB/decade stage and the last of the $20(1 - y)$-dB/decade stages are left unpeaked; the first of the $20(1 - y)$-dB/decade stage is peaked with $m = 1$. The normalized response curves given in Section 13.2.1.7 are useful design aids for numbers of stages other than three.

14.2.2.3 Optimum Number of Stages

With the addition of high-frequency peaking, Eq. 14.12 for the loop gain in a maximum feedback amplifier becomes

$$|A_{lm}| = \frac{g}{x}\left[\left(\frac{\mathscr{G}\mathscr{B}}{\omega_{co}}\right)\left(\frac{x}{|TF_m|}\right)^{1/n}\left(\frac{2-y}{hn}\right)\right]^{2-y} \tag{14.31}$$

Differentiation shows that the optimum number of stages is still that given by Eq. 14.13, namely,

$$n_{\text{opt}} = \ln\left(\frac{|TF_m|}{x}\right),$$

but that the maximum loop gain is increased by the factor

$$\frac{g}{h^{2-y}}.$$

Substituting Bode's value for $g_{\max}$ (Eq. 14.22) and $h = \frac{2}{3}$, the loop gain increase in an amplifier with 30° phase margin ($y = \frac{1}{3}$) amounts to 6.0. In a simpler amplifier in which shunt peaking alone is used the increase amounts only to 2.0.

Maximum feedback amplifiers that include high-frequency peaking achieve the highest possible gain stability over a specified frequency band of all single-loop feedback amplifiers; that is, maximum feedback amplifiers achieve the highest stability of all amplifiers in which the feedback is restricted to a single over-all loop. However, further increases in stability, can be achieved by building up an amplifier as a cascade of several groups of stages; each group is in itself a maximum feedback amplifier, but there is no feedback around the entire amplifier. This fact does not contradict Section 14.1.1:

(i) Section 14.1.1 assumes that the total number of stages is constant, whereas the number of stages in the present case is varied to optimize the amplifier.

(ii) Section 14.1.1 takes no account of the bandwidth over which the feedback is to be maintained.

That more feedback can be obtained in a multisection amplifier is apparent from Eq. 14.31. Consider the special case in which n, the total number of stages, is constant. Suppose the amplifier is split into k identical groups, each of n/k stages and each with transfer function $TF_m^{1/k}$. Substitution into Eq. 14.31 gives the loop gain around each group as

$$|A_{lm}| = \frac{g}{x}\left[\left(\frac{\mathscr{G}\mathscr{B}}{\omega_{co}}\right)\left(\frac{x}{|TF_m|^{1/k}}\right)^{k/n}\left(\frac{2-y}{hn/k}\right)\right]^{2-y};$$

that is,

$$|A_{lm}| = \frac{g}{x}\left[\left(\frac{\mathscr{G}\mathscr{B}}{\omega_{co}}\right)\left(\frac{x^{k/n}}{|TF_m|^{1/n}}\right)\left(\frac{2-y}{hn/k}\right)\right]^{2-y}, \tag{14.32}$$

where it is assumed that the same margins x and y apply to each group as apply to the complete single-loop amplifier. Evidently, the available feedback increases monotonically with increasing k, until the limit of $k = n$ when local feedback is applied to each stage. This conclusion should not be taken too seriously, because approximations in the derivation of Eq. 14.31 become poor when n (corresponding to n/k in the multiloop amplifier) is less than 3. Nevertheless, the point is made that a single-loop amplifier is not necessarily the optimum structure. Further increases in feedback are available if the total number of stages is varied so that each group has its number of stages optimized; the price paid is an increase in the number of devices over the optimum number for a single-loop maximum feedback amplifier. The subject of multisection maximum feedback amplifiers does not appear to be well covered in the literature; there is scope for further research.

14.2.3 Low-Frequency Considerations

Sections 14.2.1 and 14.2.2 show that the amount of feedback that can be maintained over a specified bandwidth is limited ultimately by the gain-bandwidth product of available active devices. High-frequency oscillation occurs if too much feedback is applied. No corresponding theoretical limit exists to the amount of feedback that can be applied at low frequencies, although available coupling and bypass capacitors may impose practical limits.

The first part of designing a maximum feedback amplifier is to determine the amount of feedback that can be applied at high frequencies. The second, and usually much the easier, part is to fit coupling and bypass capacitors to the design so that the feedback is maintained down to the lowest frequency of interest. The roll-off in loop gain at low frequencies

is merely the inverse to that at high frequencies. Moreover, as no peaking is available at low frequencies, the low-frequency design is inverse to the simpler high-frequency case. Figure 14.16 shows the low-frequency roll-off, to which the following equations apply:

$$|A_{lm}| = \frac{1}{x}\left(\frac{\omega_{co}}{\omega_b}\right)^{2-y}, \tag{14.33}$$

$$\frac{\omega_c}{\omega_b} = \frac{2-y}{n}. \tag{14.34}$$

These equations are inverse to Eqs. 14.6 and 14.10 of the high-frequency analysis. The value of $|A_{lm}|$ determined by high-frequency considerations is substituted into Eq. 14.33 to find ω_b, and ω_c then follows from Eq. 14.34. As the loop gain is actually down by about 6 dB at ω_{co}, it is good practice to choose ω_{co} one or two octaves below the frequency to which the loop gain should be maintained.

The loop-gain cutoff at $20(2 - y)$ dB/decade is obtained by interlacing poles and zeros on the negative real axis. Section 7.4 gives equations for the low-frequency singularities due to coupling, bypass, and decoupling capacitors. Pole-zero cancellation is a useful design technique for reducing the number of effective singularities. Notice that n, equivalent to the number of stages in the high-frequency analysis, is equal to the slope in multiples of 20 dB/decade of the final asymptote. In other words n is equal to the number of zeros at or close to the origin and, in practice, n is usually equal to the number of coupling capacitors in the amplifier. Exceptions can occur when there are zeros very close to the origin due to bypass capacitors.

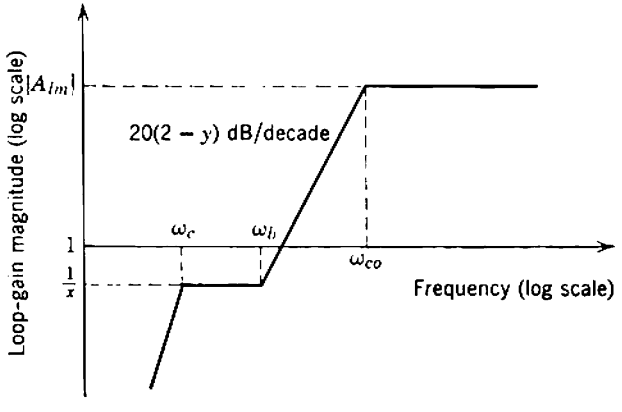


Fig. 14.16 Gain asymptotes for the low-frequency roll-off in a maximum feedback amplifier.

14.2.4 Practical Example

As an illustration of the principles developed above, this section describes in outline a maximum feedback amplifier. The amplifier forms part of a carrier-frequency telephony system and is to meet the following specifications:

(i) 34-dB power gain between 21 and 530 kHz; the corresponding voltage gain between equal impedances is 50;
(ii) tubes type EF184/6EJ7 are to be used throughout; for these tubes $g_m = 15$ mA/V and the total capacitance (including strays) is 25 pF;
(iii) as much feedback as possible is to be applied, consistent with the margins $x = 2$ and $y = \frac{1}{3}$.

14.2.4.1 High-Frequency Considerations

As explained in Section 14.2.3, the high-frequency aspects of the design are considered first. In order, the pertinent data are

$$\begin{aligned} TF_m &= 50, & g &= 1.8, \\ f_{co} &= 530 \text{ kHz}, & h &= \tfrac{2}{3}, \\ x &= 2, & g_m &= 15 \text{ mA/V}, \\ y &= \tfrac{1}{3}, & C &= 25 \text{ pF}. \end{aligned}$$

The value of g corresponds to a high-frequency peaking system that is appreciably less efficient than the theoretical limit and will be described shortly. Substitution into the relevant equations gives

ω_{co}	3.33×10^6 radian/sec
$\mathscr{G}\mathscr{B}$	6.0×10^8 radian/sec
n_{opt} (Eq. 14.13)	3
$\lvert A_{lm}\rvert$ (Eq. 14.31)	666
ω_b (Eq. 14.21)	1.75×10^8 radian/sec
ω_c (Eq. 14.28)	2.10×10^8 radian/sec

Before peaking is added the poles and zeros of the three stages are:

20-DB/DECADE STAGE

Poles: -3.33×10^6, -2.10×10^8 radian/sec
Zero: -1.75×10^8 radian/sec

FIRST STAGE IN TAPER

Poles: -4.66×10^6, -2.10×10^8 radian/sec
Zero: -1.75×10^7 radian/sec

SECOND STAGE IN TAPER

Poles: -3.33×10^7, -2.10×10^8 radian/sec
Zero: -1.25×10^8 radian/sec

These poles and zeros give a 34-dB/decade roll-off between ω_{co} and ω_b and, of course, a triple pole at ω_c.

The peaking in the vicinity of ω_{co} aims to increase the mid-band loop gain by the factor $g = 1.8$. The chosen peaking is as follows:

1. The 20-dB/decade stage is shunt peaked with $m = 0.5$; this squares up the corner, but does not increase $|A_{lm}|$.
2. The first of the stages in the taper has the peaking network shown in Fig. 14.14*d* for which the gain increase is 2 up to a frequency slightly less

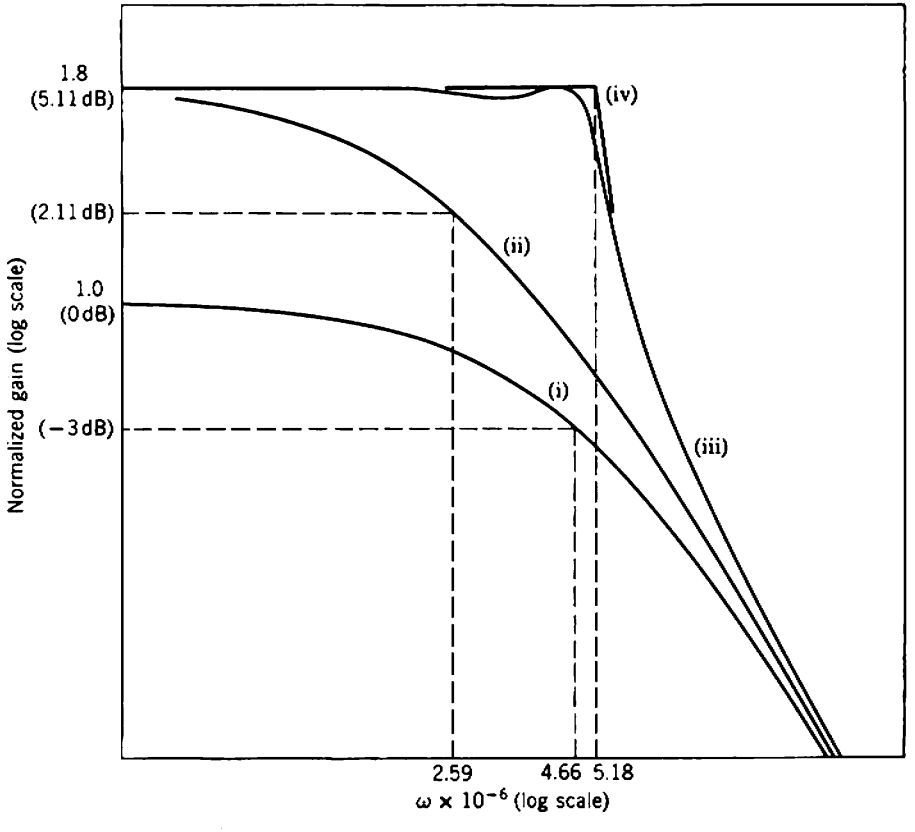


Fig. 14.17 Peaking near ω_{co} in an illustrative amplifier: curve (i)—normalized response without peaking; curve (ii)– response with load resistance increased by a factor 1.8; curve (iii)– as (ii), but with peaking added; curve (iv)– as (ii), but with ideal peaking.

than the network's nominal cutoff. The nominal cutoff is therefore chosen 10% above the desired cutoff; consequently, the gain increase is 10% less than 2, namely, $g = 1.8$. Figure 14.17 shows the form of the response. In accordance with Eq. 14.19, the capacitance effective near ω_{co} is taken as $(C_1 + C_2)$.

The peaking in the vicinity of ω_c aims at reducing the phase shift. Only one stage is peaked—the first stage in the taper because it is the only stage whose other singularities lie well away from ω_c. Shunt peaking is used with $m = 1$.

14.2.4.2 Low-Frequency Considerations

The pertinent data for the low-frequency design are

$$|A_{lm}| = 666, \qquad x = 2,$$
$$f_{co} = 21 \text{ kHz}, \qquad y = \tfrac{1}{3}.$$

Remembering that the specified cutoff frequency is the frequency to which $|A_{lm}|$ should be maintained, the actual cutoff frequency is chosen about a factor 5 lower:

$$\omega_{co} = 3 \times 10^4 \text{ radian/sec.}$$

Substitution into the relevant equations gives

ω_b (Eq. 14.33)	400 radian/sec
ω_c (Eq. 14.34)	222 radian/sec

The singularities required to give 34-dB/decade roll-off between ω_{co} and ω_b are

20-DB/DECADE STAGE

Poles: -3×10^4, -222 radian/sec
Zero: -400 radian/sec

FIRST STAGE IN TAPER

Poles: -2.09×10^4, -222 radian/sec
Zero: -4.96×10^3 radian/sec

SECOND STAGE IN TAPER

Poles: -2.42×10^3, -222 radian/sec
Zero: -574 radian/sec

14.2.4.3 Details

Figure 14.18 shows the elemental circuit diagram for the complete amplifier. Three pentode stages have cathode-feedback biasing and are *RC* coupled together. Signals are fed into and taken from the amplifier by

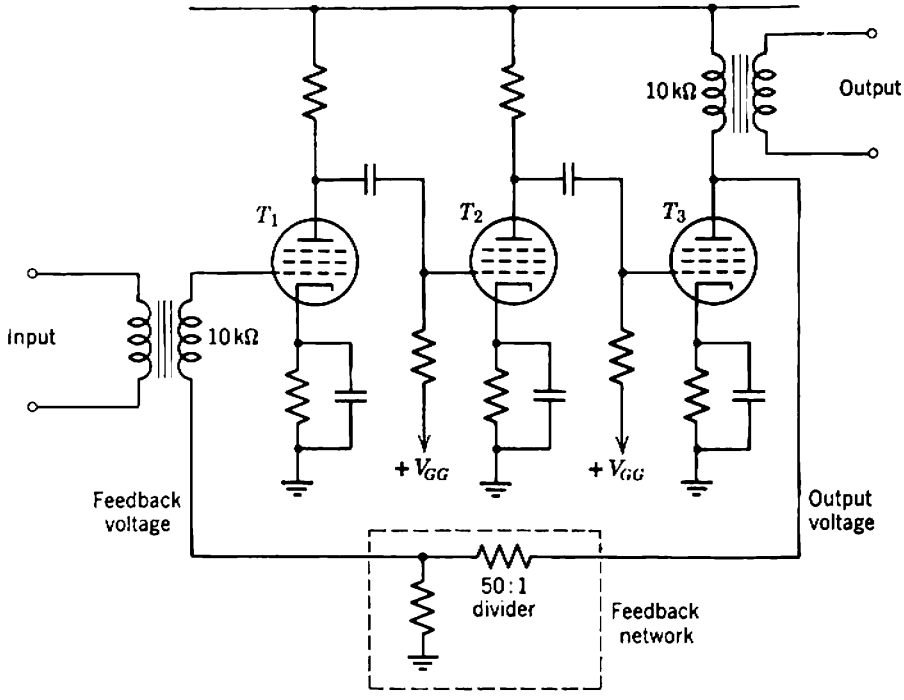


Fig. 14.18 Elemental circuit diagram for a three-stage maximum feedback amplifier.

means of transformers. The input transformer secondary impedance and the output transformer primary impedance are equal (at 10 kΩ); therefore 34-dB power gain corresponds to $\times 50$ voltage gain. Notice that these transformers are not within the feedback loop, although the primary impedance of the output transformer does constitute the load for the third stage. Irregularities in the transformer response, together with irregularities in the loss of the telephone transmission lines are corrected by passive networks external to the amplifier.

Figure 14.19 shows the complete circuit diagram. At both low and high frequencies the distribution of stages is

Stage 1 → second stage in taper,
Stage 2 → first stage in taper,
Stage 3 → 20-dB/decade stage.

The output transformer primary inductance is chosen to give a low-frequency pole at -2.73×10^3 radian/sec and the leakage inductance is chosen for shunt peaking with $m = 0.5$. The balancing capacitors at the input are to give a feedback factor that is independent of frequency,

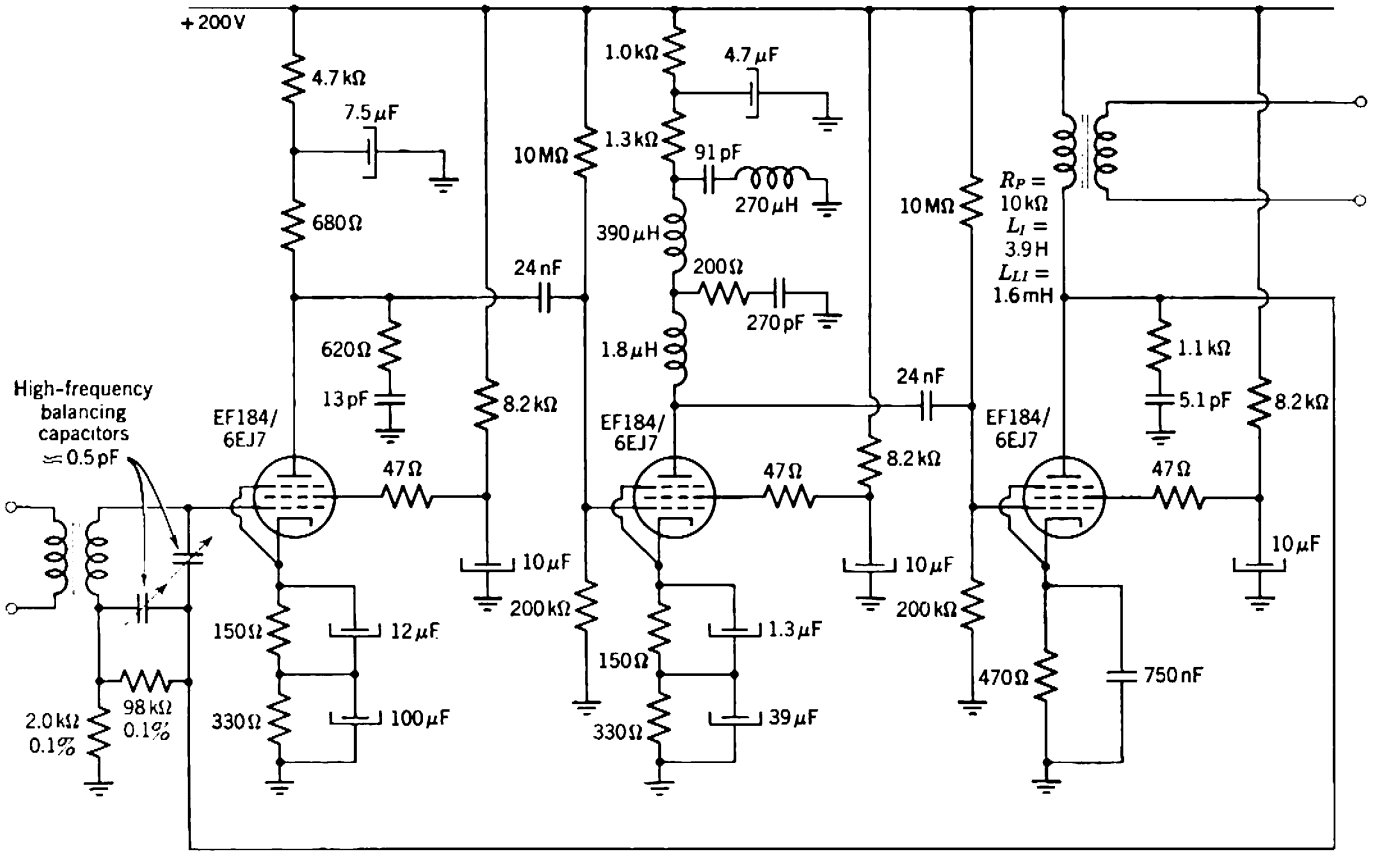


Fig. 14.19 Circuit diagram for a carrier-frequency telephone repeater; line equalizing networks external to the feedback loop are not included. The feedback is 56 dB from 21 to 530 kHz.

allowing for the stray capacitances associated with the input transformer and first tube. The screen bypass capacitors are so large that they do not enter into considerations of loop phase shift; 47-Ω stopping resistors prevent v.h.f. parasitic oscillation. The high-frequency singularities are as set out in Section 14.2.4.1, and Table 14.1 gives a complete list of low-frequency singularities; pole-zero cancellation occurs many times to reduce the total low-frequency singularity pattern to that set out in Section 14.2.4.2.

Table 14.1 Low-Frequency Singularities for Fig. 14.19

Circuit	Poles	Zeros
T_1 cathode	-2.42×10^3 radian/sec	-574 radian/sec
T_1 cathode	-79.3	-29.7
T_1 decoupling	-29.7	-222
T_2 input coupling	-222	0
T_2 cathode	-2.09×10^4	-4.96×10^3
T_2 cathode	-222	-79.3
T_2 decoupling	-222	-400
T_3 input coupling	-222	0
T_3 cathode	-3.00×10^4	-2.73×10^3
T_3 output transformer	-2.73×10^3	0

For all stages the screen circuit poles lie in the vicinity of -15 radian/sec. The exact position cannot be found without an elaborate calculation, because cross-coupling occurs due to the impedances in the cathode circuits.

14.3 OVER-ALL FEEDBACK AROUND CASCADED FEEDBACK STAGES

Amplifiers in which over-all feedback is applied around a cascade of feedback stages represent a compromise between the high gain-bandwidth product of series- and shunt-feedback stages and the high gain stability of single-loop maximum feedback amplifiers.* Figure 14.20 shows a block diagram. The ratio of the feedback around the local loops to the feedback around the over-all loop controls the trade-off between gain stability and gain-bandwidth product. For a specified number of stages and total

* This circuit technique was first proposed in E. M. CHERRY and D. E. HOOPER, "The design of wide-band transistor feedback amplifiers," *Proc. Inst. Elec. Engrs. (London)*, **110**, 375, February 1963. Notice that the discussion of noise on p. 386 is wrong.

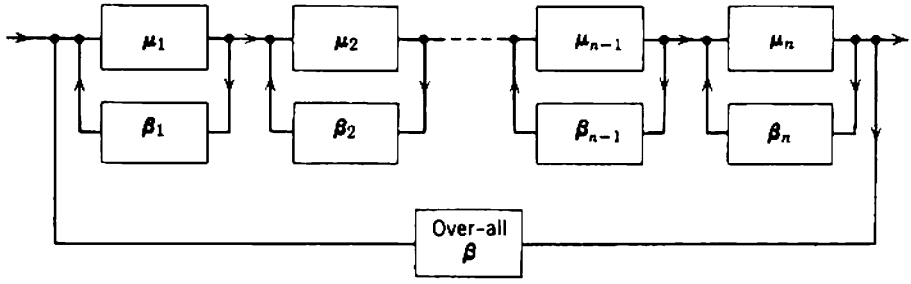


Fig. 14.20 Block diagram showing over-all feedback around a complete amplifier as well as local feedback around the stages.

gain, increasing the local feedback increases $\mathscr{GB}$ but reduces the gain stability. The noise performance of these amplifiers is poor; thermal noise of the local feedback resistor of the first stage is usually the dominant component of noise at the input. For this reason local feedback is sometimes omitted from the first stage.

The cascade of feedback stages should be that which has the highest inherent gain stability and gain-bandwidth product, namely, a cascade of series and shunt stages in alternation. Figure 14.21 shows circuit diagrams for realizing the transfer functions Y_T, A_V, A_I, and Z_T. These correspond to the block diagrams in Fig. 10.26, for which the closed-loop transfer functions are

$$Y_T \approx -\frac{1}{\beta} = \frac{1}{Z_{F1}}, \tag{14.35a}$$

$$A_V \approx -\frac{1}{\beta} = \frac{Z_{F1} + Z_{F2}}{Z_{F1}}, \tag{14.35b}$$

$$A_I \approx -\frac{1}{\beta} = -\frac{Z_{F1} + Z_{F2}}{Z_{F1}}, \tag{14.35c}$$

$$Z_T \approx -\frac{1}{\beta} = -Z_{F2}. \tag{14.35d}$$

Typically, the gain around the over-all feedback loop is of the order of 10. At lower values the gain stability is not significantly better than that of the simple alternate cascade. At higher values the amplifier approaches a single-loop maximum feedback type and the loss of $\mathscr{GB}$ becomes excessive.

The local feedback in Fig. 14.21 has the very desirable feature of modifying the input and output immitances of the amplifier in the direction required for separability of the over-all feedback loop into its μ and β

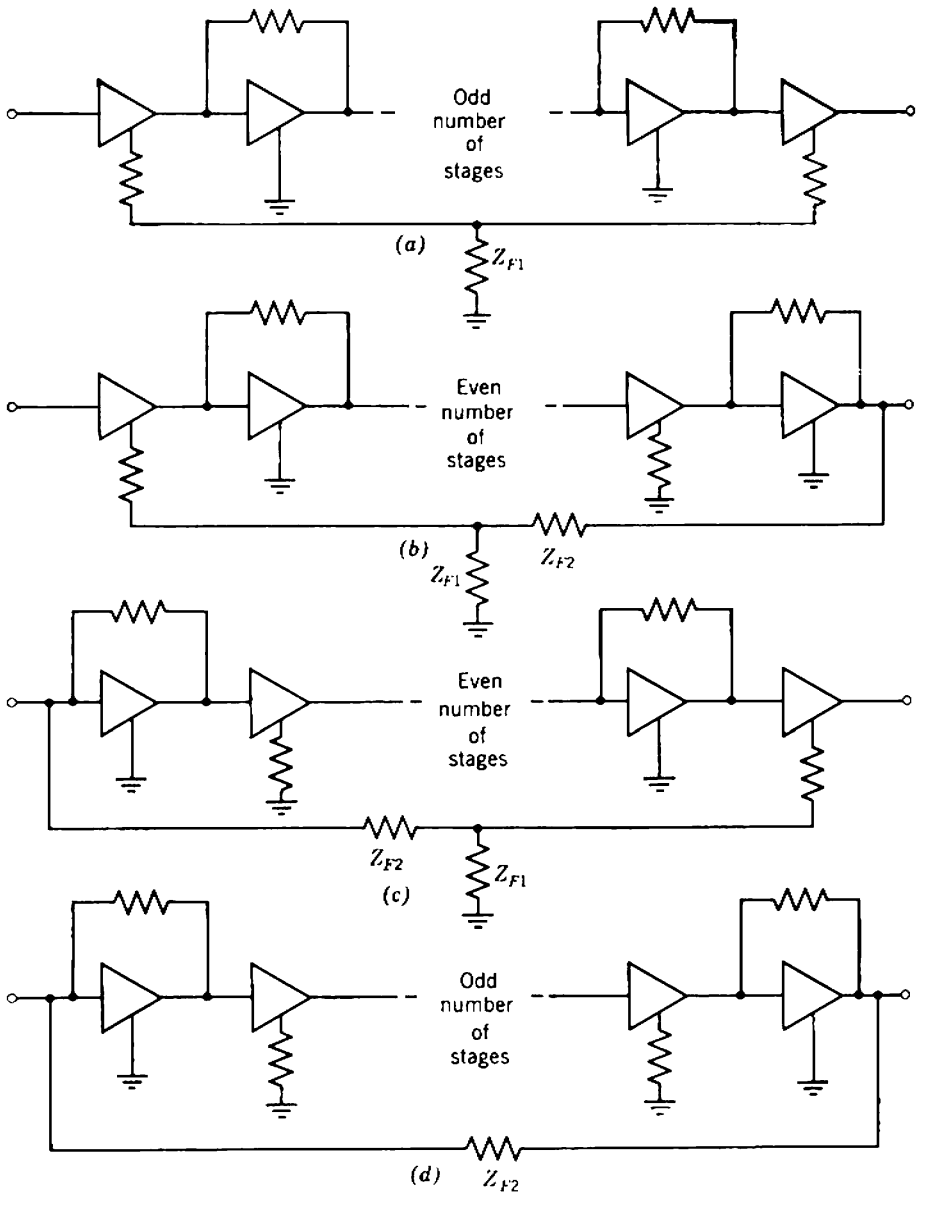


Fig. 14.21 Realization of transfer functions by means of combined local and over-all feedback: (*a*) transfer admittance Y_T; (*b*) voltage gain A_V; (*c*) current gain A_I; (*d*) transfer impedance Z_T.

components. As explained in Section 10.6.1, the input admittance y_{11} in parts *a* and *b* of Fig. 10.26 should be small, whereas the input impedance z_{11} in parts *c* and *d* should be small. Because the input immitances of the first stages in Fig. 14.21 satisfy these conditions, the reference variable μ can be approximated very closely by the transfer function with the over-all feedback network removed. The singularities of $\mu(s)$ are thus simply and directly under the designer's control; Sections 12.2 and 13.5 discuss the ultimate low- and high-frequency singularities of series- and shunt-feedback stages, whereas Section 11.6.2.2 discusses singularities in the mid-band range.

The type of singularity pattern chosen for $\mu(s)$ depends on system requirements. At one extreme is an array of poles and zeros interlaced on the negative real axis. The over-all feedback loop is a maximum feedback type in which the local feedback may be regarded as peaking. Figure 14.22 shows a typical high-frequency root-locus diagram for a three-stage amplifier; two dominant closed-loop poles lie close to the imaginary axis, so there is considerable overshoot in the transient response. At the other extreme is the dominant-pole design technique described in Section 10.5.2.4. One high-frequency pole of $\mu(s)$ lies so much closer to the origin than all others that the excess loop phase shift scarcely exceeds 90° before the loop-gain magnitude falls below unity. The dominant closed-loop pole remains on the negative real axis as shown in Fig. 14.23, and the transient response is monotonic. Between these extremes is a whole range of patterns for which some dominant closed-loop poles are

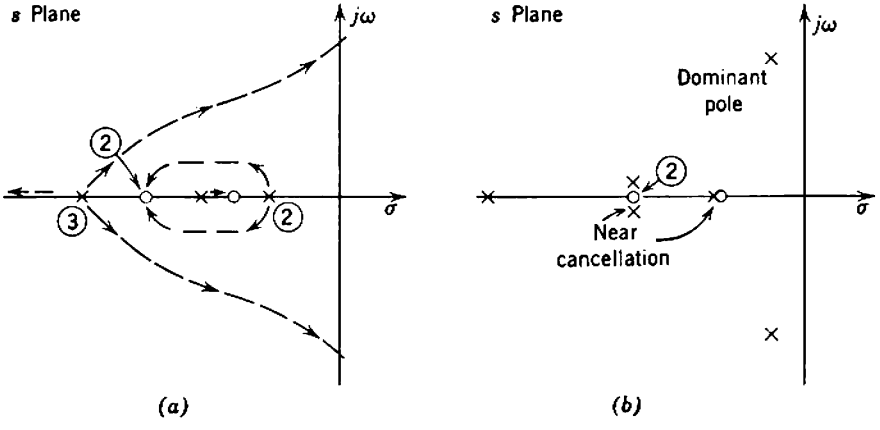


Fig. 14.22 Typical high-frequency root-locus diagram for a maximum-feedback over-all loop: (*a*) open-loop singularity pattern and root loci; (*b*) closed-loop singularity pattern.

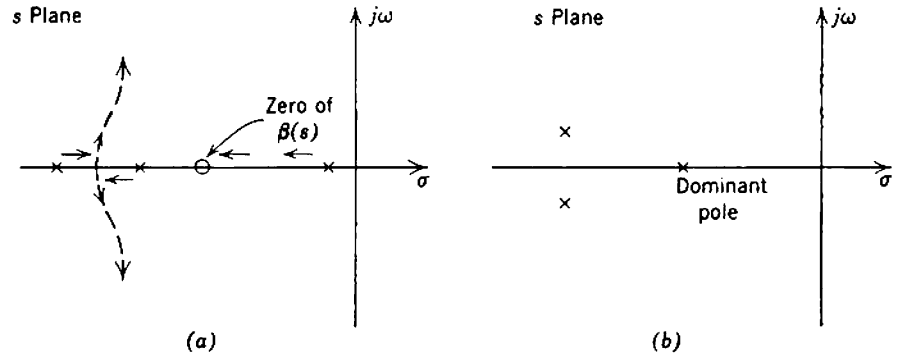


Fig. 14.23 Typical high-frequency root-locus diagram for a dominant-pole over-all loop: (*a*) open-loop singularity pattern and root loci; (*b*) closed-loop singularity pattern.

real while others are complex. Root-locus plotting is by far the most useful aid to selecting a pattern for $\mu(s)$.

A powerful design method for locating the dominant closed-loop poles accurately is described in Section 10.5.3.4: phantom zeros of the feedback factor $\beta(s)$ are placed near the desired closed-loop pole positions. In Fig. 14.21 real high-frequency zeros of $\beta(s)$ result from connecting a capacitance in shunt with R_{F2} or an inductance in series with R_{F1}. Low-frequency zeros result from connecting a capacitance in series with R_{F1} or an inductance (usually prohibitively large) in shunt with R_{F2}. Complex zeros can be produced by bridged-T or lattice structures. Zeros of the feedback factor have the added feature of reducing the excess loop phase shift, and so steering the loop-gain locus away from the critical point in a Nyquist diagram. Even when not required to locate the dominant closed-loop poles, zeros of $\beta(s)$ are often introduced to increase the safety margin against regeneration.

14.3.1 Shunt Stage with a Dominant Pole

For most applications a shunt-feedback stage is designed so that its two poles occur as a complex conjugate pair or lie close together on the negative real axis. When over-all feedback is applied to a cascade of feedback stages, however, at least one of the stages must have a dominant pole which causes the loop-gain magnitude to fall away at high frequencies without the excessive phase shift associated with a group of poles. This stage is usually a shunt stage (because the pole of a series stage is not accurately defined and is at an inconvenient frequency) so its two poles

must be widely separated on the negative real axis. Although the equations of Section 13.5.2 give the dominant pole correctly, some of the approximations become poor and the second pole may be seriously in error. This nondominant pole is of vital interest in the over-all feedback loop; therefore the error is intolerable.

In Section 13.5.2 the voltage gain of a shunt-feedback stage such as that shown in Fig. 14.24 is assumed to be

$$\begin{aligned} A_V(s) &= -g_m\left(\frac{Z_F \times Z_L}{Z_F + Z_L}\right) \\ &= -g_m\left(\frac{R_F \times R_L}{R_F + R_L}\right)\left\{\frac{1}{1 + s[R_F \times R_L/(R_F + R_L)](C_F + C_L)}\right\}. \end{aligned} \tag{14.36}$$

From the form of Eq. 14.36, the voltage gain falls away at high frequencies with a 20-dB/decade asymptote. The implication of Eq. 14.36 is that Z_F merely appears in shunt with Z_L so far as its loading effect on the device is concerned. In fact, Eq. 7.68 in the general discussion of Miller effect shows that the contribution of Z_F to the load is

$$Z_{\text{out}} = \frac{Z_F}{1 - 1/A_V}.$$

Equation 14.36 is a good approximation over the frequency range normally of interest, because $|A_V|$ is large. When a shunt-feedback stage has a dominant pole, however, the voltage gain may be small at the frequency of the second pole, and is the solution of

$$A_V(s) = -g_m\left\{\frac{\left[\dfrac{Z_F(s)}{1 - 1/A_V(s)}\right]Z_L(s)}{\dfrac{Z_F(s)}{1 - 1/A_V(s)} + Z_L(s)}\right\}. \tag{14.37}$$

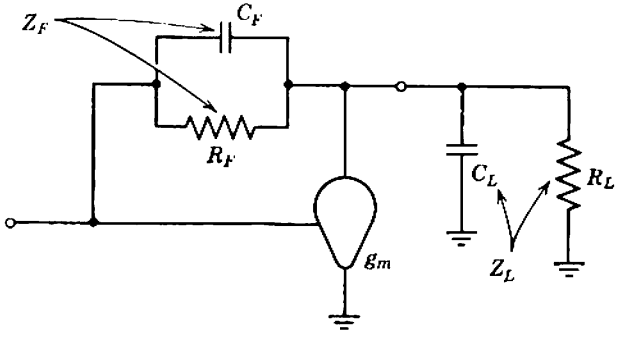


Fig. 14.24 Elemental circuit diagram for a shunt-feedback stage.

There is no longer a 20-dB/decade asymptote, so the conclusions of Section 13.5.2 are invalidated. It is now shown that connecting a small resistance R in series with C_F corrects the response.

At high frequencies both Z_F and Z_L are dominated by their capacitive components. Therefore, the equivalent circuit (with R included) reduces to Fig. 14.25. It is required that the voltage gain should be

$$A_V(s) = -\frac{g_m}{s(C_F + C_L)} \tag{14.38}$$

implying that the effective load is C_F and C_L in shunt. Substituting this value of voltage gain into Eq. 14.37, together with the values of Z_F and Z_L in Fig. 14.25, we obtain

$$\frac{-g_m}{s(C_F + C_L)} = -g_m \left\{ \frac{\left[\dfrac{R + 1/sC_F}{1 + s(C_F + C_L)/g_m}\right]\left(\dfrac{1}{sC_L}\right)}{\dfrac{R + 1/sC_F}{1 + s(C_F + C_L)/g_m} + \dfrac{1}{sC_L}} \right\}$$

from which it follows that R should satisfy

$$R = \frac{1}{g_m}\left(\frac{C_F + C_L}{C_F}\right). \tag{14.39}$$

The above analysis for the general charge-control model applies well to vacuum tubes but may be poor for transistors because of the presence of r_B. Formally, g_m in the analysis should be replaced by the transfer admittance from base voltage to collector current:

$$g_m \to Y_T(s) = \frac{\alpha(s)}{r_B/\beta(s) + r_E}$$

and $Y_T(s)$ has a phase delay at high frequencies. This phase delay can be reduced by applying series feedback to the transistor; quite a small loop

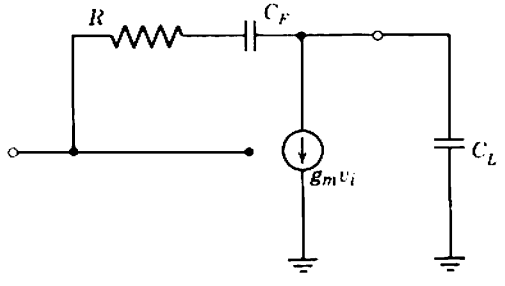


Fig. 14.25 High-frequency equivalent circuit of a shunt-feedback stage, modified to include a resistance R in series with C_F.

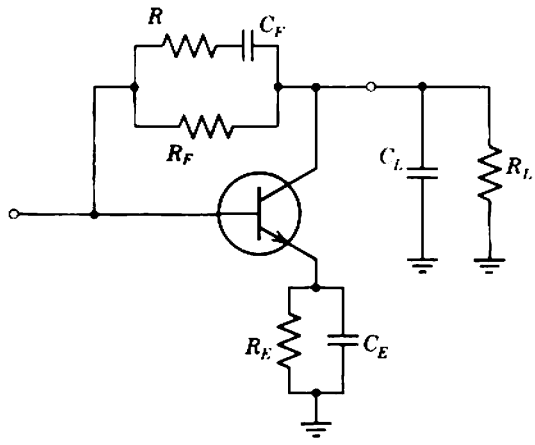


Fig. 14.26 Elemental circuit diagram of a transistor shunt-feedback stage with a dominant pole, modified so that both poles are located accurately.

gain is sufficient—R_E in Fig. 14.26 is typically one-half r_B. The other component values are

$$C_E = \frac{\tau_T}{R_E}, \tag{14.40}$$

$$R = \left(\frac{r_B/\beta_N + r_E + R_E}{\alpha_N}\right)\left(\frac{C_F + C_L}{C_F}\right) \approx R_E\left(\frac{C_F + C_L}{C_F}\right). \tag{14.41}$$

and the poles are

$$p_1 \approx -\frac{1}{R_F C_F}, \tag{14.42}$$

$$p_2 \approx -\frac{C_F}{(C_F + C_L)(\tau_T + R_E C_F)}. \tag{14.43}$$

More accurate expressions for the poles can be deduced from Section 13.5.2.2; R_F may become infinite (i.e., be omitted from the circuit) if a very large ratio is required between the poles.

14.3.2 Example

As an illustration, Fig. 14.27 shows the circuit of an amplifier that realizes a very stable, wide-band transfer impedance; the amplifier is based on Fig. 14.21*d*. Three stages have shunt, series, and shunt feedback, respectively. The mid-band transfer functions are 14.5 V/mA, 33 mA/V, and 0.33 V/mA, to give a combined transfer resistance 160 V/mA.

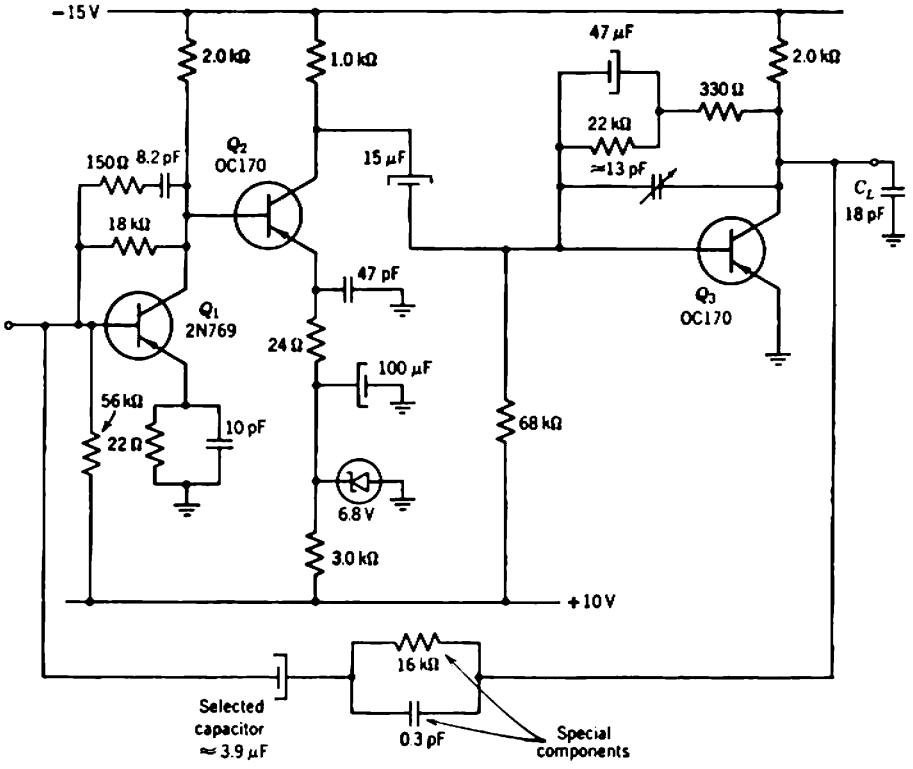


Fig. 14.27 Circuit diagram of an amplifier that realizes a very stable wide-band transfer impedance of 15 V/mA.

Over-all feedback with loop gain 10 is provided by a 16 kΩ resistor, and the closed-loop transfer resistance is 14.4 V/mA.

Figures 14.28*a* to *c* show the high-frequency singularity patterns of the individual stages, and Fig. 14.28*d* shows the pattern for the feedback factor. Figure 14.28*e* is the root-locus diagram, and it can be verified that the poles take up the positions in Fig. 14.28*f* when the loop gain is 10. The three dominant poles constitute a linear-phase group for which the rise time is 22 nsec with less than 1% overshoot.

At low frequencies the loop gain has four poles and two zeros on the negative real axis, and two zeros at the origin. However, the circuit constants are such that the pole from Q_2–Q_3 coupling capacitor cancels the zero from Q_3 shunt-feedback network; in addition, the zero from Q_2 emitter capacitor and the pole from Q_3 feedback network lie close to the origin and can be neglected. Thus the loop gain has two significant poles plus two zeros at the origin. The pole from Q_2 emitter capacitor

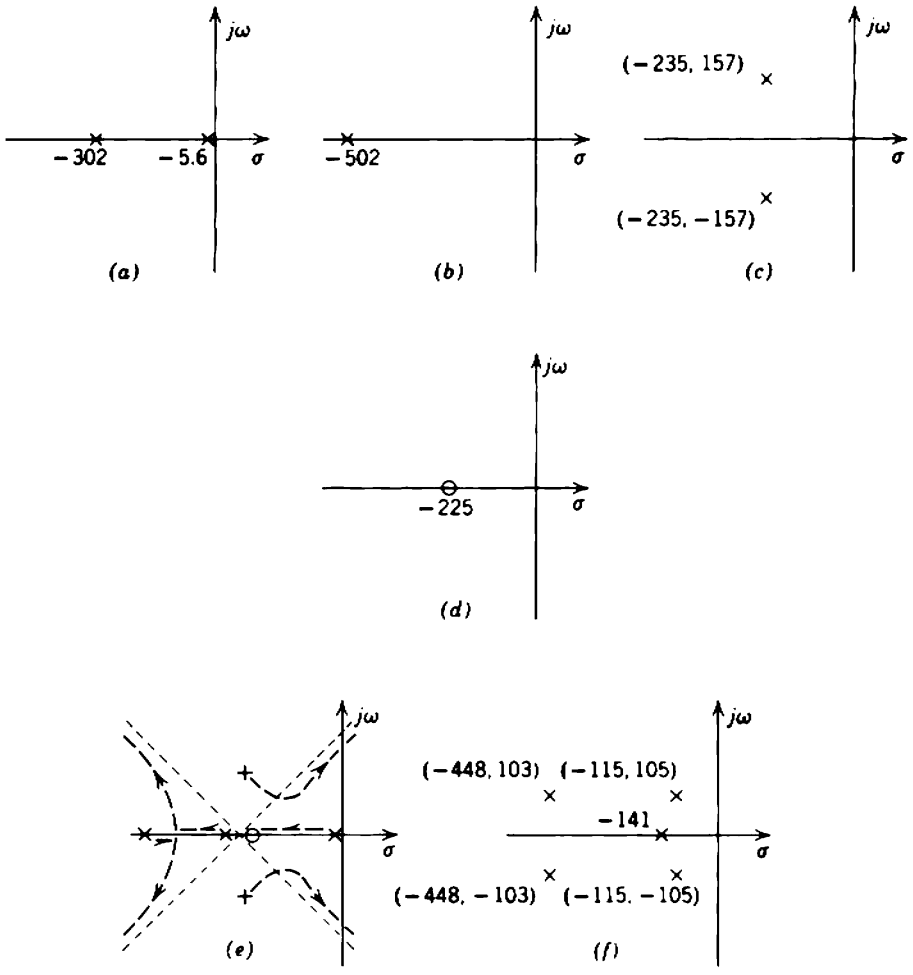


Fig. 14.28 High-frequency root-locus diagram for Fig. 14.27: (*a*) first-stage singularity pattern; (*b*) second-stage singularity pattern; (*c*) third-stage singularity pattern; (*d*) feedback network singularity pattern; (*e*) root-locus diagram; (*f*) closed-loop pattern for $|A_{lm}| = 10$. Numerical values are multiples of 10^6 radian/sec.

is at 50 Hz whereas the pole from the selected feedback blocking capacitor lies near 2.5 Hz. With a loop gain of 10, the poles come together and then separate to form a complex conjugate pair near $s = (-5, \pm j6)$ Hz. In addition, the closed-loop singularity pattern contains a zero at the pole of the feedback network, near 2.5 Hz. The feedback blocking capacitor is selected so that the total phase shift at frequencies of the order of 50 Hz is as small as possible, giving minimum tilt on a square wave (Section 12.1.2.1).

14.3.2.1 The Root-Locus Diagram

A word is in order about the derivation of the high-frequency open-loop singularity pattern. Like all root-locus plotting, the process is essentially trial-and-error, but an enormous number of the possibilities can be eliminated by inspection. Because the forward path contains two shunt stages and one series stage, its singularity pattern must contain four poles plus a fifth pole far out on the negative real axis. Moreover, as one pole must dominate, and as simple over-all feedback networks can give only one zero, the individual stage patterns must appear qualitatively as in Fig. 14.28. (Alternative patterns, with disadvantages and no obvious advantage, are that the complex pole pair could be exchanged for two real poles and that the zero could be moved further from the origin.) Root-locus considerations show that there are three dominant closed-loop poles; it is desired that these should form a linear-phase group, a compromise between small overshoot and an excessively small bandwidth reduction factor χ. Figure 14.29 shows the normalized pole positions for a linear-phase group.

Figure 14.30*a* is the authors' introductory guess at a normalized open-loop singularity pattern:

(i) The dominant pole is shown at the origin, although its actual position is close to the origin at $-p_1$.

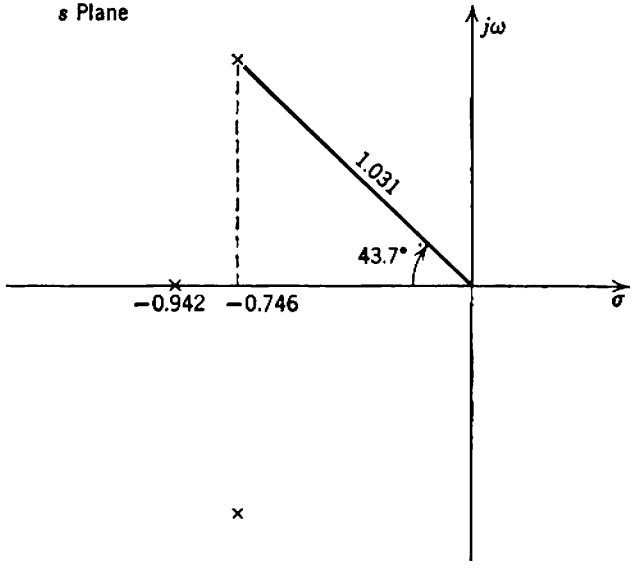


Fig. 14.29 Pole positions for the three-pole linear-phase function, normalized so that the geometric mean distance from the origin is 1.

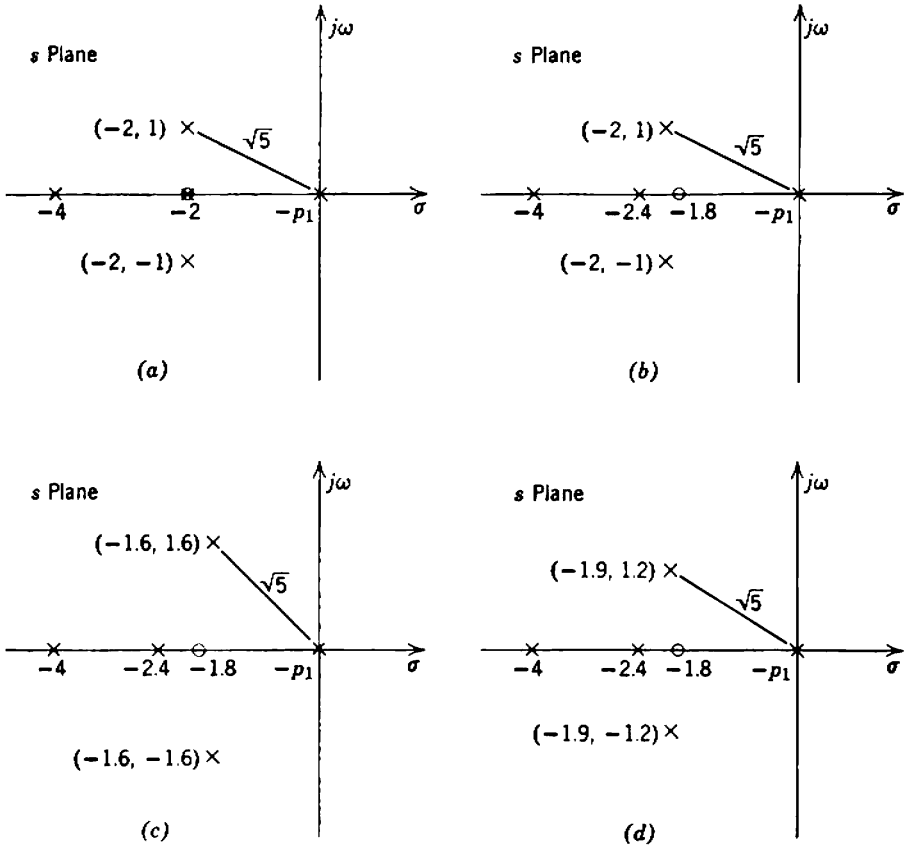


Fig. 14.30 Derivation of the open-loop singularity pattern, Fig. 14.28*e*. The normalization aims at: (i) ease in Spirule calculations; (ii) mean closed-loop dominant pole distance from the origin approximately (but certainly not exactly) 1.

(ii) The zero, toward which the real-axis closed-loop pole moves, is placed at twice the nominal mean dominant closed-loop pole distance from the origin. Because the loop gain is only 10, the pole does not approach close to the zero.

(iii) The far-out pole due to the series stage is placed at four times the nominal mean closed-loop pole distance.

(iv) The other poles have a real part -2, equal to that of the zero; the complex poles lie on a circle of radius $\sqrt{5}$, so their imaginary parts are ± 1.

Before constructing the loci accurately, the coincident pole and zero are separated slightly. The factors 1.8 and 2.4 shown in Fig. 14.30*b* are arbitrary; others would certainly prove satisfactory.

The loci are now constructed with a Spirule, paying particular attention to the points at $\pm 43.7°$ as these are the angles at which the closed-loop poles should lie. If the distance of the points from the origin is designated by d, it follows from Fig. 14.29 that the real-axis closed-loop pole should lie at

$$s = -d\left(\frac{0.942}{1.031}\right).$$

The loop gain at these closed-loop pole positions can be found in terms of p_1 from Eqs. 10.59 and 10.69; if the open-loop pattern is correct, the loop gain should be the same at all closed-loop poles.

In fact, the loop gain at the complex poles is higher than at the real pole. Therefore the open-loop complex poles are moved on the circle of radius $\sqrt{5}$ to a position of lower damping ratio (Fig. 14.30*c*), and the loci are again constructed. The loop gain now is higher at the real-axis closed-loop pole than at the complex poles, so the correct open-loop pattern must be between Figs. 14.30*b* and *c*. Figure 14.30*d* shows the correct pattern, obtained at the next attempt.

Two final steps remain:

(i) Choose p_1 (from Eqs. 10.59 and 10.69) so that the loop gain is 10 at the desired pole positions.
(ii) Convert the normalized pattern to the actual pattern shown in Fig. 14.28.

In passing, it may be remarked that the damping ratio of the complex poles in the open-loop pattern is a convenient adjustment to the response of a practical amplifier; the feedback capacitor for Q_3 has a preset adjustment.

14.4 FEEDBACK PAIRS

Feedback can be applied around a group of two stages in 36 different ways: all combinations of series and shunt feedback at the input and output with all combinations of the devices in the common-emitting-electrode, common-control-electrode, and common-collecting-electrode configurations. In half the circuits the feedback is positive and the response regenerative. These positive feedback configurations belong to the general class of multivibrators and, although not useful in amplifiers, some are useful in switching circuits. Of the remaining 18 configurations, two are useful building-blocks for general purpose amplifiers and warrant the discussion in Section 14.4.1. At least a further 11 are useful in not-too-specialized applications, and are described briefly in Section 14.4.2.

Readers are advised to investigate the remaining 5, and either find why these circuits achieve nothing that cannot be done better a simpler way or else find applications.

14.4.1 Voltage- and Current-Feedback Pairs

Voltage- and current-feedback pairs are groups of two common-emitting-electrode stages with over-all feedback. Figure 14.31 shows elemental

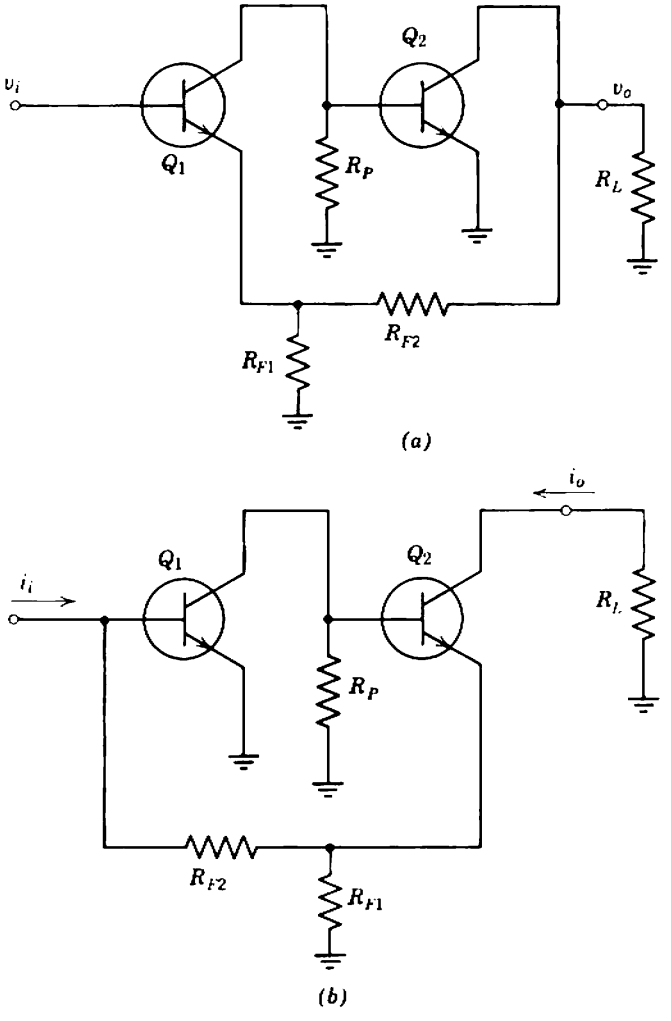


Fig. 14.31 General purpose two-stage building-blocks: (*a*) voltage-feedback pair; (*b*) current-feedback pair.

circuit diagrams for the transistor realizations. In these diagrams R_P represents the total of the biasing resistors shunting the input of the second stage. Similarly, R_L represents the total load, comprising all biasing resistors and more especially the input resistance of any following amplifier.

Voltage- and current-feedback pairs are dual circuits. The voltage-feedback pair is based on Fig. 10.26*b* and takes its name from the fact that the feedback is series (voltage) at the input and shunt (voltage) at the output. The input resistance is large, whereas the output resistance is small, and voltage gain is the stable transfer function being given approximately by

$$A_V \approx -\frac{1}{\beta} = \frac{R_{F1} + R_{F2}}{R_{F1}}. \tag{14.44}$$

In order to make use of this stable transfer function, a voltage-feedback pair should be fed from a low-resistance (approximate voltage) source and should look into a high-resistance (ideally infinite) load. Its own driving-point resistances are such that the voltage-feedback pair will cascade directly to yield a stable voltage gain over several stages. The total gain is the product of the voltage gains of the individual pairs.

The current-feedback pair is based on Fig. 10.26*c*. The feedback is shunt (current) at the input and series (current) at the output, so current gain is the stable transfer function:

$$A_I \approx -\frac{1}{\beta} = -\frac{R_{F1} + R_{F2}}{R_{F1}}. \tag{14.45}$$

The input resistance is small and the output resistance is large, whereas the source resistance should be large and the load resistance small. Current-feedback pairs will cascade directly to yield a stable current gain over several stages and, to a good approximation, the total gain is the product of the current gains of the individual pairs. In precise calculations there should be a small correction for the coupling efficiency between pairs; part of the output current from one pair is shunted by the biasing resistors and does not flow on into the next pair.

Voltage- and current-feedback pairs are useful building-blocks for multistage amplifiers, in many ways similar to the cascade of series- and shunt-feedback stages in alternation. When moderate gain stability is required (up to about 5% per stage), feedback pairs can be cascaded directly. When greater precision is called for, over-all feedback may be applied around a cascade of feedback pairs in arrangements generally similar to those in Fig. 14.21. Notice that cascaded voltage-feedback pairs can realize only a

stable voltage gain, whereas cascaded current-feedback pairs can realize only a stable current gain. If a stable transfer conductance or transfer resistance is required, it is necessary to interpose a series- or shunt-feedback stage at some point in the amplifier and thereafter change from voltage- to current-feedback pairs or vice versa as appropriate. Figure 14.32 shows block diagrams for realizing all four transfer functions.

As building-blocks, feedback pairs have the advantage of marginally greater gain stability than the alternate cascade. For specified tolerance on the gain, an optimized feedback pair yields about 10% more gain than an optimized combination of one series stage with one shunt stage. Against this, feedback pairs have the disadvantage of being difficult to design at high frequencies. Moreover, feedback pairs are apt to oscillate unless preventive measures are taken; it is therefore impossible to divorce their mid-band performance from their high-frequency performance, so there is no simple set of design rules corresponding to Chapter 11.

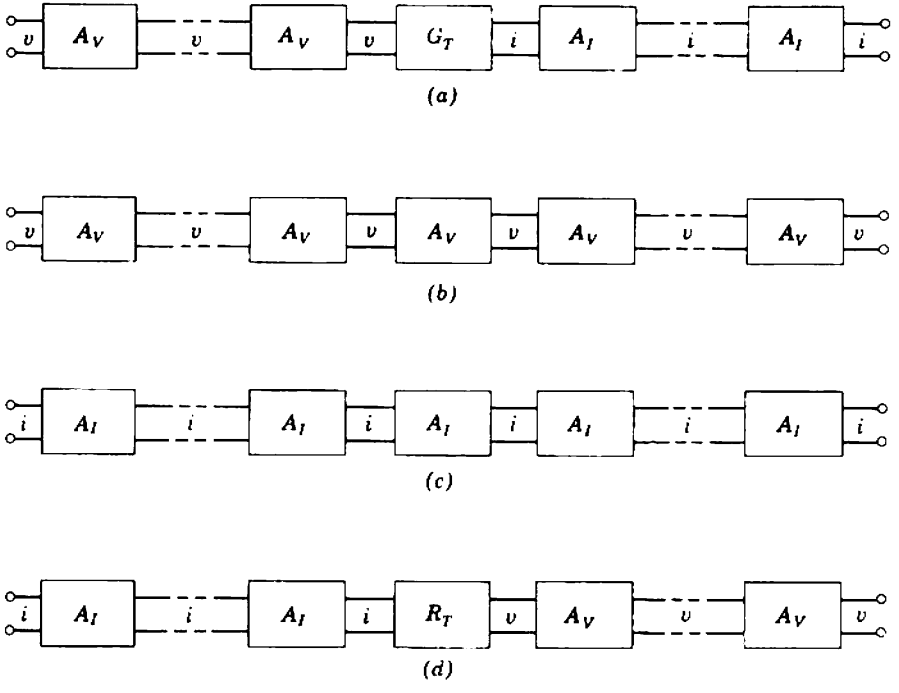


Fig. 14.32 Realization of transfer functions from voltage- and current-feedback pairs, and, in parts (*a*) and (*d*), series- and shunt-feedback stages: (*a*) transfer admittance Y_T; (*b*) voltage gain A_V; (*c*) current gain A_I; (*d*) transfer impedance Z_T. Notice the pertinent signal type (v or i) between blocks.

The authors prefer series- and shunt-feedback stages to feedback pairs as building-blocks, for two reasons:

(i) Even if feedback pairs are chosen, series- and shunt-feedback stages are still required sometimes, as shown in Figs. 14.32*a* and *d*. The circuit designer has, therefore, to familiarize himself with both types of building-block.

(ii) In designing a large amplifier, it is an advantage if the performance of the component building-blocks can be predicted simply and accurately.

Feedback pairs are undoubtedly useful in audio-frequency amplifiers, and despite the above objections are used by many designers of wide-band amplifiers. The following sections consider the mid-band design of transistor feedback pairs.*

14.4.1.1 Voltage-Feedback Transistor Pair

Although voltage- and current-feedback pairs can be analyzed easily enough, a conceptual difficulty is that they cannot readily be separated into their μ and β components. Section 10.6.1 contains a general discussion of the conditions necessary for separability. Consider, for example, the voltage-feedback pair which is redrawn in Fig. 14.33 to emphasize its similarity to the basic voltage gain structure (Fig. 10.26*b*). The condition under which the forward transfer function μ in Fig. 10.26*b* is the voltage gain of the amplifier without its feedback network is that the component of v_F directly attributable to i_1 should be negligible compared with the component attributable to v_2. This condition certainly is not satisfied in Fig. 14.33; the current into the feedback network is $(\beta_N + 1)i_1$, not merely i_1, and β_N is a large number.

The voltage-feedback pair can be analyzed by the superposition method described in Section 10.3.1, Eqs. 10.18 and 10.19 in particular. Figure 14.34 shows the equivalent circuit. One assumption is made to simplify writing down the superposition constants:

$$R_{F2} \gg R_{F1}.$$

The transfer conductance of Q_2 is chosen as the reference variable

$$\frac{i_k}{v_j} = G_{T2} = \frac{\alpha_{N2}}{r_{B2}/\beta_{N2} + r_{E2}}$$

* A high-frequency analysis of the current-feedback pair appears in M. S. GHAUSI, "Optimum design of the shunt-series feedback pair with maximally-flat-magnitude response," *Trans. Inst. Radio Engrs.*, **CT-8**, p. 448, December 1961; also M. S. GHAUSI, *Principles and Design of Linear Active Circuits*, Chapter 14 (McGraw Hill, New York, 1965).

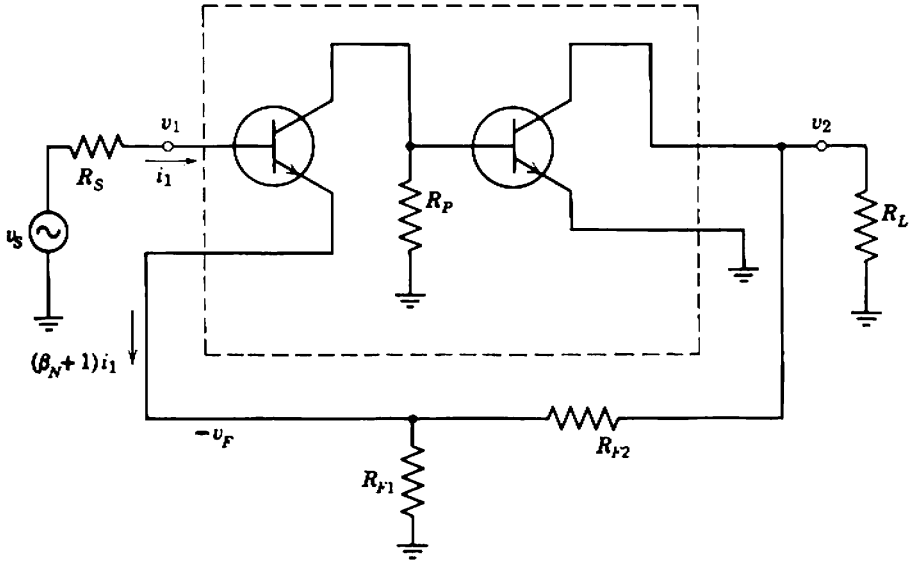


Fig. 14.33 Voltage-feedback pair, redrawn to emphasize its relation to Fig. 10.26*b*.

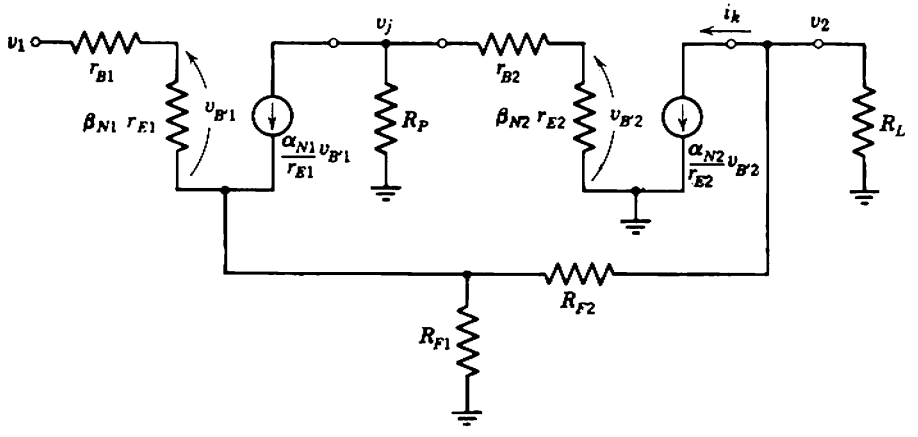


Fig. 14.34 Mid-band equivalent circuit of the transistor voltage-feedback pair.

and the superposition constants defined in Eqs. 10.19 become:

$$a_{12} = \left.\frac{v_2}{v_1}\right|_{i_k = 0} \approx 0,$$

$$z_{k2} = \left.\frac{v_2}{i_k}\right|_{v_1 = 0} \approx -\frac{R_{F2} \times R_L}{R_{F2} + R_L},$$

$$a_{1j} = \left.\frac{v_j}{v_1}\right|_{i_k = 0} \approx -\left(\frac{\alpha_{N1}}{r_{B1}/\beta_{N1} + r_{E1} + R_{F1}}\right)\left[\frac{R_P \times (r_{B2} + \beta_{N2} r_{E2})}{R_P + (r_{B2} + \beta_{N2} r_{E2})}\right],$$

$$z_{kj} = \left.\frac{v_j}{i_k}\right|_{v_1 = 0} \approx \left(\frac{R_L}{R_{F2} + R_L}\right)\left(\frac{R_{F1}}{R_{F1} + r_{B1}/\beta_{N1} + r_{E1}}\right) \times \alpha_{N1}\left[\frac{R_P \times (r_{B2} + \beta_{N2} r_{E2})}{R_P + (r_{B2} + \beta_{N2} r_{E2})}\right].$$

Substitution into Eq. 10.18 gives the closed-loop transfer function v_2/v_1, from which the feedback factor $-1/\beta$ can be extracted to yield an expression in the form of Eq. 10.36:

$$\frac{v_2}{v_1} = \left(\frac{R_{F1} + R_{F2}}{R_{F1}}\right)\left[\frac{1}{1 + (R_{F1} + R_{F2})/\mu R_{F1}}\right], \tag{14.46}$$

where

$$\mu \approx \left(\frac{\alpha_{N1}}{r_{B1}/\beta_{N1} + r_{E1} + R_{F1}}\right)\left(\frac{R_P}{R_P + r_{B2} + \beta_{N2} r_{E2}}\right)\beta_{N2}\left(\frac{R_{F2} \times R_L}{R_{F2} + R_L}\right). \tag{14.47}$$

The terms in μ can be associated with definite processes:

1. $\alpha_{N1}/(r_{B1}/\beta_{N1} + r_{E1} + R_{F1})$ is the transfer conductance of Q_1 with R_{F1} as a series-feedback resistor.
2. $R_P/(R_P + r_{B2} + \beta_{N2} r_{E2})$ is the fraction of the current output from Q_1 that flows into Q_2 rather than into R_P.
3. β_{N2} is the current gain of Q_2.
4. $(R_{F2} \times R_L)/(R_{F2} + R_L)$ is the effective load resistance seen by Q_2.

Thus μ is the voltage gain of the amplifier without the over-all feedback, but with account taken of the local series feedback provided around Q_1 by the network in its emitter circuit, and with account taken of the loading on Q_2 provided by the over-all feedback network. As an alternative, the loop gain can be written down directly if the feedback loop is broken at a point which satisfies the conditions set out in Problem 10.3; obvious points are the collector terminals of Q_1 or Q_2.

Three useful conclusions can be drawn from Eq. 14.47. First, if the transistors are types with unequal values of β_N, more feedback will result if the higher gain transistor is used for Q_2; μ is almost proportional to

β_{N2} but depends on β_{N1} only as a small correction. Second, if $|A_I|$ is large, the transfer function approaches $-1/\beta$. The ratio $R_{F2}:R_{F1}$ is therefore determined by the closed-loop gain demanded of the amplifier. Substitution into Eq. 14.47 and differentiating shows that $|\mu|$ (and hence $|A_I|$) is maximum when the product of the feedback resistors satisfies

$$R_{F1}R_{F2} = \left(\frac{r_{B1}}{\beta_{N1}} + r_{E1}\right)R_L. \tag{14.48}$$

Third, R_P should be as large as possible.

The input resistance of a voltage-feedback pair is required in estimating the load for any preceding amplifier:

$$R_i \approx \beta_{N1}\left(\frac{R_P}{R_P + r_{B2} + \beta_{N2}r_{E2}}\right)\beta_{N2}\left(\frac{R_{F2} \times R_L}{R_{F2} + R_L}\right)\left(\frac{R_{F1}}{R_{F1} + R_{F2}}\right). \tag{14.49}$$

14.4.1.2 Current-Feedback Transistor Pair

A current-feedback pair can similarly be analyzed using Eqs. 10.22 and 10.23, with the transfer conductance of Q_1 as reference variable. The closed-loop current gain is

$$A_I = -\alpha_{N2}\left(\frac{R_{F1} + R_{F2}}{R_{F1}}\right)\left[\frac{1}{1 - (R_{F1} + R_{F2})/\mu R_{F1}}\right], \tag{14.50}$$

where

$$\mu \approx -\left(\frac{R_{F2}}{R_{F2} + r_{B1} + \beta_{N1}r_{E1}}\right)\beta_{N1}\left[\frac{R_P}{R_P + r_{B2} + \beta_{N2}(r_{E2} + R_{F1})}\right]\beta_{N2}. \tag{14.51}$$

The input resistance (required in an estimate of coupling efficiency) is

$$R_i \approx (r_{B1} + \beta_{N1}r_{E1})\left(\frac{R_{F2} + r_{B1} + \beta_{N1}r_{E1}}{R_{F2}}\right)\frac{1}{\beta_{N1}} \times \left[\frac{R_P + r_{B2} + \beta_{N2}(r_{E2} + R_{F1})}{R_P}\right]\frac{1}{\beta_{N2}}\left(\frac{R_{F1} + R_{F2}}{R_{F1}}\right). \tag{14.52}$$

Notice the term α_{N2} in Eq. 14.50; the feedback stabilizes the ratio of the input current i_1 to the emitter current of Q_2. The terms in Eq. 14.51 have the following interpretation:

1. $R_{F2}/(R_{F2} + r_{B1} + \beta_{N1}r_{E1})$ is the fraction of the input current that flows into Q_1 rather than into R_{F2}.
2. β_{N1} is the current gain of Q_1.
3. $R_P/[R_P + r_{B2} + \beta_{N2}(r_{E2} + R_{F1})]$ is the fraction of the output current from Q_1 that flows into Q_2; R_{F1} applies series feedback to Q_2 and raises its input resistance.
4. β_{N2} is the current gain of Q_2.

The condition for maximum loop gain is that

$$R_{F1}R_{F2} = (r_{B1} + \beta_{N1}r_{E1})\left(\frac{R_P + r_B}{\beta_{N2}} + r_{E2}\right), \qquad (14.53)$$

and R_P should be as large as possible.

14.4.1.3 Example

As an example, Fig. 14.35 shows an audio-frequency current amplifier building-block, in which the current-feedback pair is reduced to its barest essentials. Two transistors are direct coupled together, with the signal feedback resistors also serving as biasing components. The biasing circuit is described briefly in Section 6.5.2.1; it is a modified form of self bias in which dc feedback is applied over two stages. Biasing stability is poor, but adequate for this sort of application. Q_1 operates at 0.5 mA with about 0.5 V between emitter and collector. Q_2 nominally operates at 2 mA, but the current can lie anywhere between 1 mA and 4 mA. The circuit constants, found by substitution into the foregoing equations are

$$\mu \approx -1440,$$

$$\beta = 1/81,$$

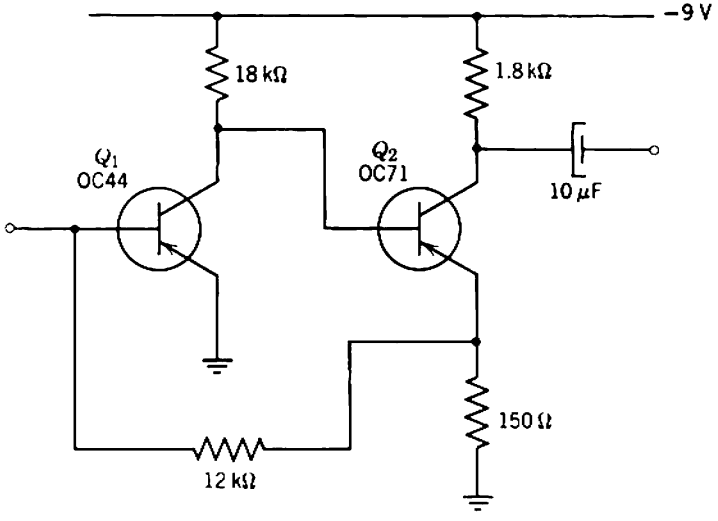


Fig. 14.35 Circuit diagram for an audio-frequency current-feedback pair. The gain is about 75, and the bandwidth about 8 Hz to 200 kHz. OC44: alloyed germanium transistor, $\beta_N = 60$, $f_T = 10$ MHz; OC71: alloyed germanium transistor, $\beta_N = 40$, $f_T = 500$ kHz.

from which

$$A_l \approx -17.8$$

and

$$A_l \approx -76.7.$$

The input resistance is 160 Ω, so that if the amplifiers are cascaded the coupling efficiency is

$$\eta = 91.8\%.$$

Because of the widely different f_T of the two transistors, no additional components are required to prevent high-frequency regeneration; Q_2 contributes a dominant pole to the loop gain.

14.4.2 Other Two-Stage Configurations

This section describes 11 two-stage feedback circuits that are useful in rather specialized applications. Table 14.2 explains the relation of the circuits one to another.

Table 14.2 Two-Stage Negative-Feedback Transistor Circuits

Feedback Type at Input	First-Stage Configuration*	Second-Stage Configuration*	Feedback Type at Output	Circuit Diagram
Series	CE	CE	Shunt	Fig. 14.31*a*
Shunt	CE	CE	Series	Fig. 14.31*b*
Series	CE	CB	Series	Fig. 14.37*a*
Shunt	CE	CB	Shunt	Fig. 14.36*a*
Series	CE	CC	Series	Fig. 14.38*b*
Shunt	CE	CC	Shunt	Fig. 14.36*c*
Series	CB	CE	Series	Fig. 14.37*c*
Shunt	CB	CE	Shunt	Fig. 14.36*b*
Series	CB	CB	Shunt	Not shown
Shunt	CB	CB	Series	Not shown
Series	CB	CC	Shunt	Not shown
Shunt	CB	CC	Series	Fig. 14.39*a*
Series	CC	CE	Series	Fig. 14.38*a*
Shunt	CC	CE	Shunt	Not shown
Series	CC	CB	Series	Fig. 14.37*b*
Shunt	CC	CB	Series†	Fig. 14.39*b*
Series	CC	CC	Shunt	Fig. 14.38*c*
Shunt	CC	CC	Series	Not shown

* CE: common emitter; CB: common base; CC: common collector.

† This amplifier breaks the pattern of the table.

Figure 14.36 shows three modified forms of the shunt-feedback stage. In each case the transfer resistance is

$$R_T = \frac{v_o}{i_1} \approx -R_F.$$

Figure 14.36*a* is a simple shunt-feedback stage using a cascode rather than a single device; the vacuum triode version of this circuit is useful in low-noise applications. Section 11.2.2.2 shows that a shunt-feedback stage should use a pentode rather than a triode because of the latter's small amplification factor. Pentodes, however, are noisy compared with triodes because of partition of the electron stream. The cascode, introduced in Problem 8.1, combines the low noise of a triode with the high gain of a pentode. Figures 14.36*b* and *c* are modifications of the shunt stage that achieve one very low driving-point resistance each; the input resistance in Fig. 14.36*b* and the output resistance in Fig. 14.36*c* are typically less than 1 Ω.

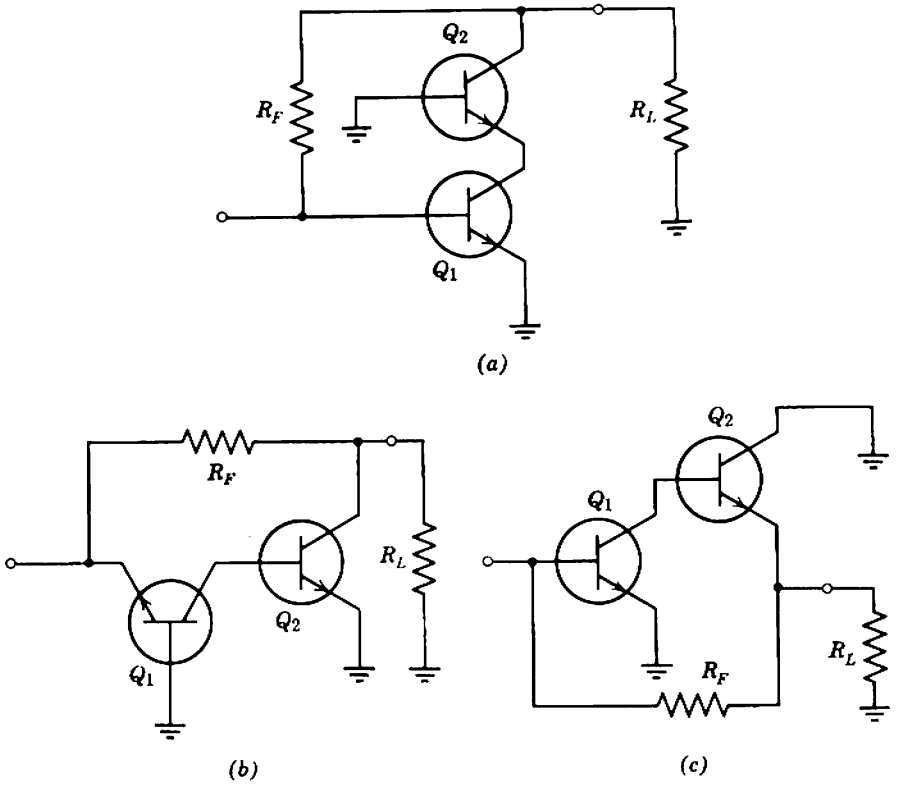


Fig. 14.36 Elemental circuits of three two-stage derivatives of the shunt-feedback stage.

Figure 14.37 shows three modified forms of the series-feedback stage. For each circuit

$$G_T = \frac{i_o}{v_i} \approx \frac{1}{R_F},$$

$$R_i = \frac{v_i}{i_i} \approx \beta_N R_F.$$

Figure 14.37*a* is the cascode series stage, not a particularly useful circuit because the simple triode series stage is entirely satisfactory so the noise problem with pentodes need not be encountered. Figure 14.37*b* is a series stage whose phase shift is 180° different from that of a conventional stage. This circuit is sometimes useful for obtaining the correct phase in an over-all feedback loop. Finally, Fig. 14.37*c* shows a circuit that can realize a very high transfer conductance; the loop gain remains large until R_F is made very small. This circuit is useful as the first stage of an audio amplifier where the noise performance of a simple series stage is inadequate.

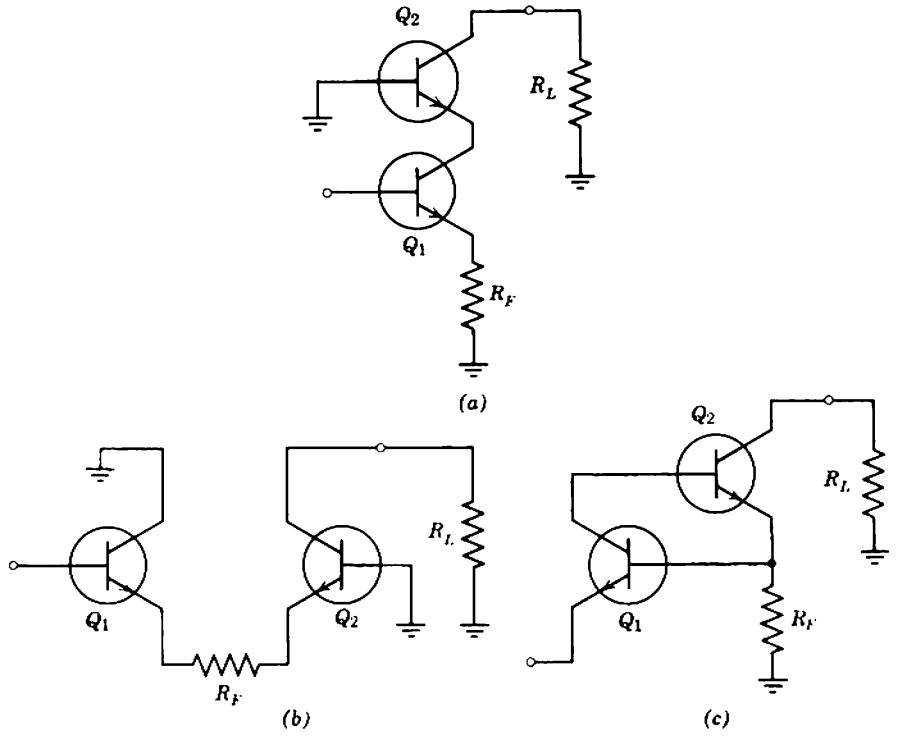


Fig. 14.37 Elemental circuits of three two-stage derivatives of the series-feedback stage.

In Fig. 14.37*c* the equivalent mean-square noise voltage referred to the input is approximately the sum of the noise voltage generator of Q_1 and the thermal noise of R_F; the latter component can be reduced to negligible proportions, because R_F may be made small.

Figure 14.38 shows three circuit arrangements that are useful in large-signal amplifiers (Section 16.3.1). In each a low-power transistor Q_1 is combined with a high-power transistor Q_2 to form a composite transistor having high dissipation capabilities plus a very high effective β_N. Figures 14.38*a* and *b* realize stable transfer conductances; the transistors in Fig. 14.38*a* are normally of the same polarity, either *p-n-p* or *n-p-n*, whereas those in Fig. 14.38*b* are normally of opposite polarity. Figure 14.38*c* is the compound emitter follower, not strictly a feedback amplifier as there

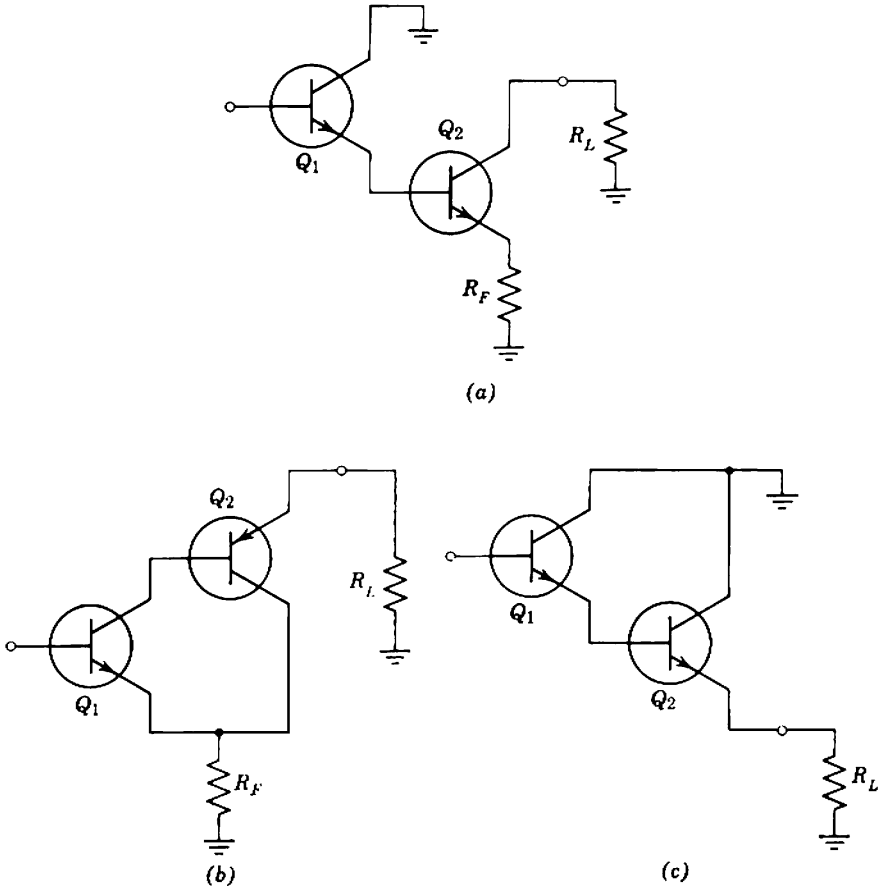


Fig. 14.38 Elemental circuits of three composite transistor circuits, useful in large-signal amplifiers.

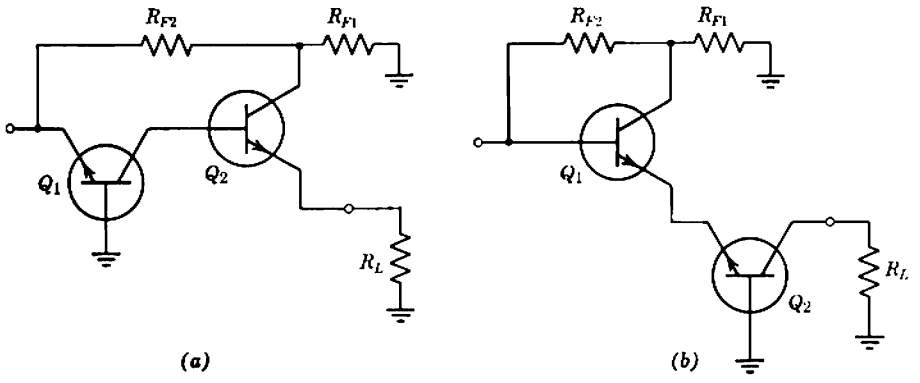


Fig. 14.39 Elemental circuits of two two-stage current-gain building blocks.

is no feedback network per se; gain magnitude cannot be exchanged for gain stability. Compound emitter followers using low-power transistors are sometimes used as the input stage of an amplifier to obtain a very high input resistance (Section 15.5).

Figure 14.39 shows two current-gain building-blocks, for which

$$A_I \approx \frac{R_{F1} + R_{F2}}{R_{F1}}.$$

These building-blocks are not particularly useful in their own right, but in conjunction they form the basis of the very-wide-band feedback current amplifier described in Section 13.9.1.

14.5 AMPLIFIERS WITH MODERATE OVER-ALL FEEDBACK

Amplifiers often have a moderate amount of feedback applied to them in order to realize some of the potential advantages of feedback without the design complexity associated with a large amount; typical values of loop gain lie between 5 and 50. If regeneration is a problem with loop gain less than 5, the whole circuit can almost certainly be improved upon. With loop gain greater than 50, the design difficulties approach those of maximum feedback amplifiers. Advantages obtained through the use of moderate feedback are

(i) reduced nonlinearity,
(ii) favorable changes in driving-point impedances,
(iii) good noise performance (as long as there is no local feedback around the first stage, Section 14.1.2),
(iv) (possibly) extended bandwidth.

Gain stability is rarely a significant advantage, because the amount of feedback is not large. Amplifiers of this type often contain a transformer or some other element for which the singularity pattern is not known, so that a complete root-locus analysis cannot be performed. As no attempt is made to optimize any one aspect of the performance, it is difficult to formulate any general design rules except in the very broadest terms:

1. The singularities of $\mu(s)$ are so chosen that the closed-loop poles take up desirable positions. Inductance peaking and *RC* ladder load networks are useful for modifying the high-frequency singularity pattern, whereas local feedback is useful at both low and high frequencies.
2. Zeros of $\beta(s)$ can be exploited to increase the margins against regeneration.
3. A dominant pole of $\mu(s)$ can reduce the loop gain magnitude below unity before the phase shift associated with the other poles is excessive. This has the disadvantage of reducing the frequency band over which the mid-band feedback is maintained.

A practical example is perhaps the best way of illustrating these principles.

14.5.1 An Audio Power Amplifier

The example chosen is an audio power amplifier for use in a high-quality sound system. The amplifier must be capable of delivering 6 W sine wave power into a 15-Ω loudspeaker, and the voltage gain is to be about 30.

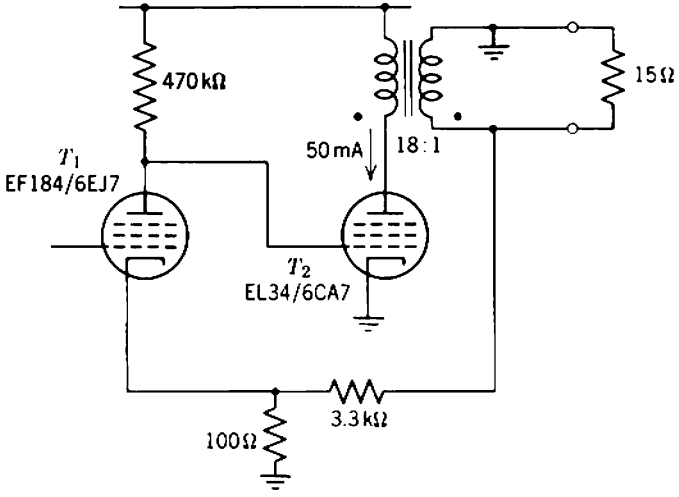


Fig. 14.40 Elemental circuit diagram of an audio power amplifier.

Figure 14.40 shows the elemental circuit diagram, basically a voltage-feedback pair. There are two pentode stages, with an 18:1 transformer to couple the 15-Ω load into the anode circuit of the second stage. This second stage uses a power tube operating at 50-mA quiescent anode current; the maximum symmetrical peak-to-peak current in the transformer primary is therefore 100 mA, corresponding to slightly more than 6-W power output. The quiescent anode-to-cathode voltage is 300 V, so that cutoff clipping occurs before saturation clipping and the full current swing can be utilized. The first tube is a very-high-slope pentode operating with a large anode load resistance in order to achieve a high gain in the forward path, hence a high loop gain. The over-all feedback resistors are approximately in the ratio of 30:1, the desired closed-loop gain; R_{F1} is 100 Ω, small enough not to apply appreciable local feedback to the first stage.

14.5.1.1 Biasing and Low-Frequency Response

The biasing system is based on Fig. 6.45; Fig. 14.41 shows the component values. Two stages are direct coupled together, with biasing feedback taken from T_2 cathode to T_1 screen. At dc the gain around the

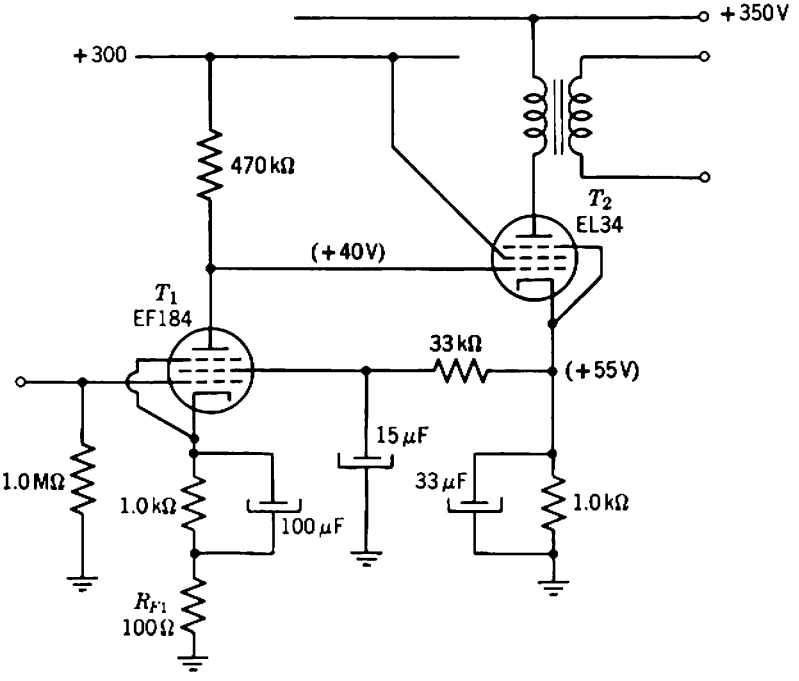


Fig. 14.41 Biasing circuit for Fig. 14.40.

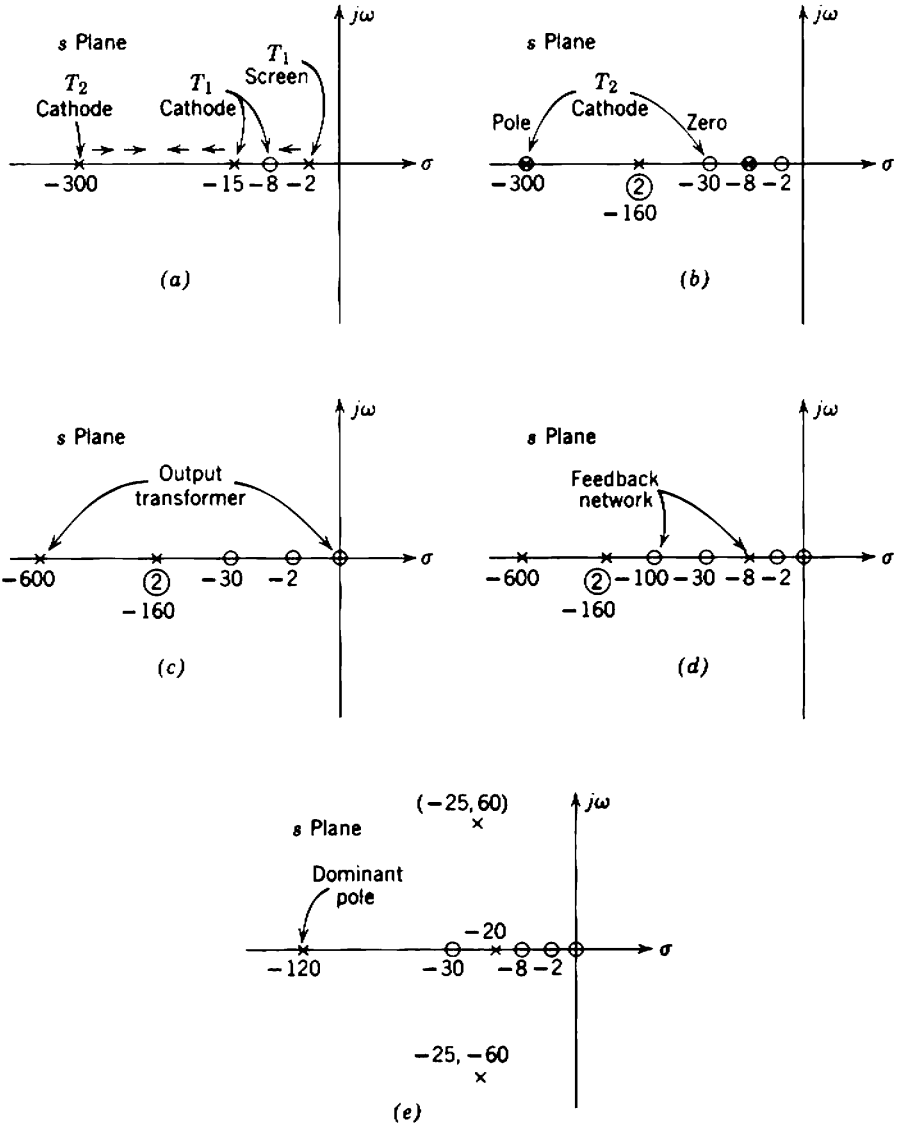


Fig. 14.42 Low-frequency singularity patterns for the audio power amplifier: (*a*) root-locus diagram for the biasing feedback loop; (*b*) singularity pattern for the transfer admittance i_{A2}/v_{G1}; (*c*) singularity pattern for the amplifier without feedback; (*d*) complete singularity pattern for the loop gain; (*e*) closed-loop singularity pattern.

biasing feedback loop is about -15. At signal frequencies, feedback is suppressed by T_2 cathode bypass and the RC network in T_1 screen; Fig. 14.42*a* shows the open-loop singularity pattern for the biasing feedback loop. Direct coupling the stages eliminates the grid supply resistor for T_2, thereby increasing the load resistance and voltage gain of T_1. Direct coupling also eliminates one low-frequency pole from the feedback loop. The quiescent conditions are

EF184		EL34	
I_K	$= 0.7$ mA,	I_K	$= 55$ mA,
V_{SK}	$= 50$ V,	V_{SK}	$= 245$ V,
V_{AK}	$= 40$ V,	V_{AK}	$= 295$ V,

from which the mid-band transfer conductance from T_1 grid to T_2 anode follows as

$$G_T = \frac{i_{A2}}{v_{G1}} \approx 5.0 \text{ A/V}.$$

The low-frequency singularities, determined by root-locus methods, are shown in Fig. 14.42*b*.

The forward gain of the amplifier can be found from G_T given above, the transformer turns ratio, and the 15-Ω load resistance:

$$\frac{v_o}{v_{G1}} \approx 1350.$$

With the over-all feedback network shown in Fig. 14.40, the mid-band loop gain is 40.

The transformer in a practical version of this amplifier had 9 H primary inductance. This transformer contributes a pole at $s = -600$ radian/sec, together with a zero at the origin; Fig. 14.42*c* shows the complete singularity pattern for the voltage gain without feedback. It can be verified that the dominant poles move close to the imaginary axis if feedback is applied with loop gain 40, so the response is unacceptable. One convenient method for improving the response is to add a phase lag to the feedback loop by connecting a capacitor in series with R_{F1}; in the final circuit (Fig. 14.43), T_1 cathode biasing network is rearranged so that the bypass capacitor serves the purpose. Figure 14.42*d* shows the complete singularity pattern for the loop gain, and Fig. 14.42*e* shows the closed-loop pattern. Although close to the imaginary axis, the nondominant poles definitely lie in the left half-plane and remain there for substantial variations in loop gain and component values. The dominant pole at 120 radian/sec masks the frequency response peak associated with the complex poles; the 3-dB cutoff frequency is about 15 Hz, adequate for a high-quality sound system.

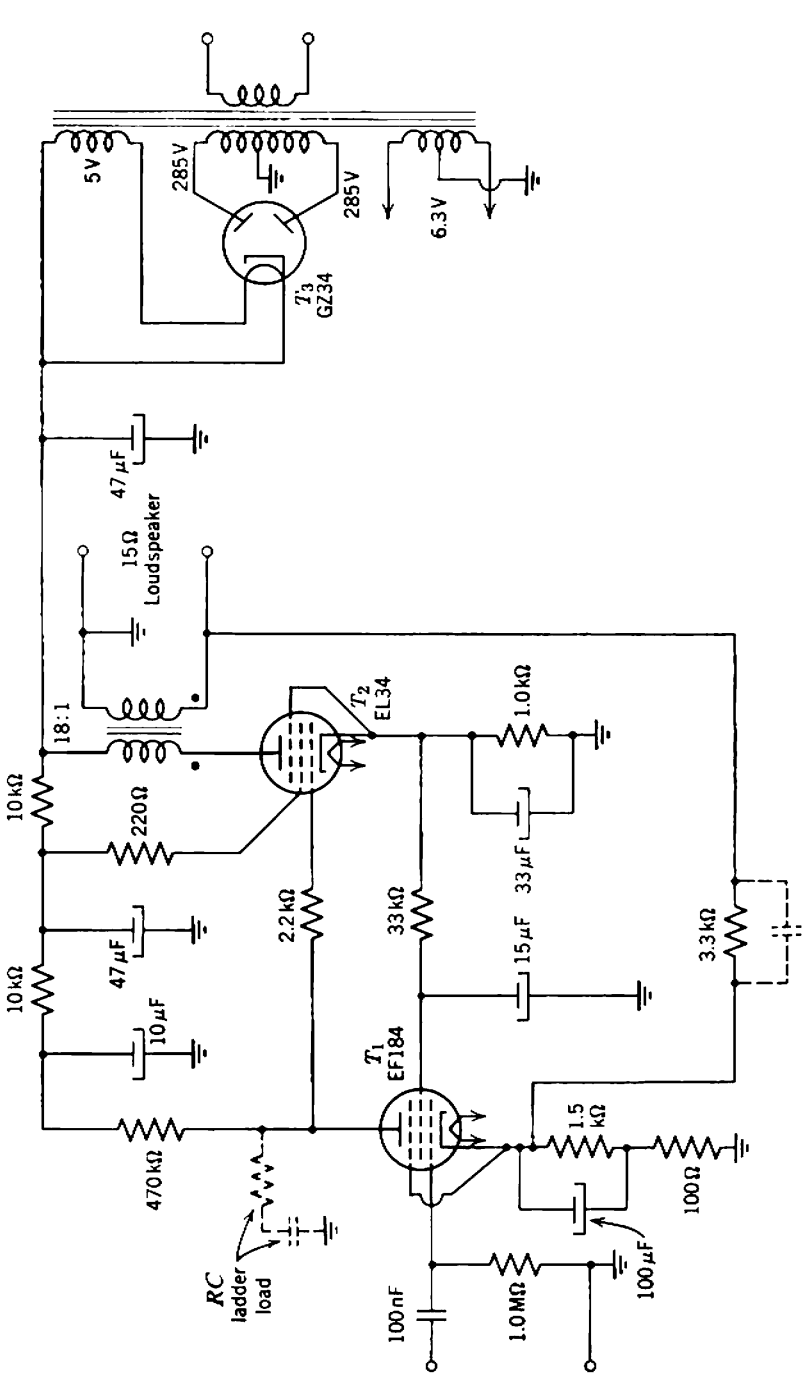


Fig. 14.43 Complete circuit diagram for the audio power amplifier, including the rectifier. The components shown broken may be required to prevent high-frequency regeneration; the values depend on the output transformer.

14.5.1.2 High-Frequency Response

At high frequencies the basic amplifier has a single pole due to the interstage network, plus a cluster of singularities due to the transformer. There is no simple way of finding the transformer singularities, so the design must at this point become empirical. Modifications to the basic amplifier that may prevent regeneration or improve the step response are the following:

(i) Connecting a small capacitor in shunt with R_{F2} contributes a zero to the feedback factor $\beta(s)$.
(ii) The load network for T_1 can be changed to an RC ladder by adding two components (Fig. 14.9). Proper choice of the ladder constants can reduce the interstage phase lag at the frequency of unity loop gain.
(iii) If all else fails, a relatively large capacitor can be connected from the anode of T_1 to ground, to contribute a dominant pole to $\mu(s)$.

In a practical version of the amplifier using a moderately priced transformer, it was sufficient to connect a 1 nF capacitor in shunt with R_{F2}. The dominant pole in the closed-loop response lies near the resulting zero at $s = -3 \times 10^5$; the closed-loop 3-dB bandwidth is 40 kHz, the rise time is 8 μsec without overshoot or ringing.

14.5.1.3 Nonlinearity

Nonlinearity in the amplifier is dominated by contributions from the second stage, and therefore can be investigated by means of the anode characteristics for tube type EL34/6CA7. First an ac load line of gradient corresponding to

$$R_L = 4860\ \Omega$$

is drawn through the quiescent point

$$I_A = 50\ \text{mA}, \qquad V_{AK} = 300\ \text{V}$$

on the manufacturers' set of characteristics corresponding to $V_{SK} = 250$ V (near enough to the actual value 245 V). Next, the I_A/V_{GK} transfer characteristic is constructed from the load line, and from this transfer characteristic values of differential error γ_0 can be found for the amplifier without feedback. With feedback applied, the differential error can be found from the exact form of Eq. 10.78:

$$\gamma = \frac{\gamma_0}{1 - A_{IQ}(1 + \gamma_0)}.$$

In calculating the harmonic distortion it is formally necessary to find the values γ' and γ'' of the differential error at the peaks of a sinusoidal input waveform. Because the amplifier is a feedback type for which the nonlinearity is small, Eqs. 9.23 become valid approximations and it is permissible to use the values of γ at the peaks of a symmetrical output waveform.

At output amplitude 10% below the clipping point, the peak-to-peak anode current swing is

$$I_{pp} = 90 \text{ mA}.$$

The differential error (without feedback) is read from the transfer characteristic as

$$\gamma_0' \approx \gamma_0(I_{AQ} + \tfrac{1}{2}I_{pp}) = -40.5\%,$$

$$\gamma_0'' \approx \gamma_0(I_{AQ} - \tfrac{1}{2}I_{pp}) = -72.3\%.$$

The differential error with feedback follows as

$$\gamma' = -1.7\%,$$

$$\gamma'' = -6.5\%,$$

and substitution into Eqs. 9.19 and 9.3 gives the second, third, and total harmonic distortions as

$$D_2 = 0.60\%,$$

$$D_3 = 0.34\%,$$

$$D_T = 0.69\%.$$

In a practical version of the amplifier, the measured total harmonic distortion at 1 kHz is 0.6% at output amplitude just less than that for which clipping becomes visually detectable on an oscilloscope.

14.5.2 Concluding Comments

The above example is typical of many practical situations, in that empirical high-frequency design is a necessity. All too often a feedback loop contains a transformer or other component whose singularity pattern is not known, so a root-locus analysis cannot be performed. Nyquist methods, however, are useful because the only information required is the measurable loop gain (in magnitude and phase) as a function of frequency. Plotting the Nyquist diagram for a system shows the frequency range over which the loop phase shift becomes excessive, and therefore suggests the frequencies at which corrective zeros willbe most beneficial.

Feedback is often regarded as a panacea, a universal means for converting a bad amplifier into a good one. Undoubtedly the forward path of a feedback amplifier is "bad" in the sense that its 3-dB bandwidth falls far short of the closed-loop bandwidth, and its frequency response may have odd kinks. The open-loop response, however, is "bad" in a controlled way; the kinks are an essential part of the design. To apply feedback indiscriminately to an amplifier is to invite trouble.

Chapter 15

Differential and Dc Amplifiers

The amplifier stages considered in earlier chapters of this book are *single-ended* types in which one each of the input and output terminals is at ground potential; in other words, the stages are three-terminal, two-port networks. This chapter introduces an important type of stage called the *long-tailed pair* which responds to the difference between the signals applied to two nongrounded input terminals. Long-tailed pairs are used extensively in differential and dc amplifiers, whose analysis and design constitute the bulk of the chapter.

A *differential amplifier* has three terminals at both input and output, with one terminal of each group being grounded. Whenever there is a difference between the signals applied to the two nongrounded input terminals, the amplified response appears as a difference between the nongrounded output terminals. Difference signals of this type are known as *differential-mode signals*, whereas any component common to both terminals is described as *common-mode*. The special case of a differential-mode signal that has no associated common-mode is said to be *push-pull*; push-pull operation is important in large-signal amplifiers (Section 16.3.2). An ideal differential amplifier does not give a differential-mode output in response to a common-mode input, but no practical differential amplifier is perfect. A useful measure of performance is the *common-mode rejection factor* λ, defined as the ratio of the common-mode input to the differential-mode input that produces the same differential-mode output:

$$\lambda = \frac{\text{common-mode input for a certain differential-mode output}}{\text{differential-mode input for the same differential-mode output}}. \quad (15.1)$$

The higher λ, the better the amplifier.

An important application for differential amplifiers is in test instruments such as oscilloscopes; an oscilloscope with a differential input can be used

as in Fig. 15.1 to examine the voltage across some component in a circuit, even though neither side of the component is grounded. Another important application is in signal transmission paths that use a balanced twin-lead or transmission line; balanced transmission paths are less susceptible to interference than unbalanced paths (Section 7.2.1).

A *dc amplifier* has a useful dynamic response that extends to zero frequency; ideally there are no low-frequency poles or zeros associated with the circuit. These amplifiers are used when the signal has useful components at very low frequencies, so that the coupling and bypass capacitors required in a conventional amplifier are prohibitively large. Typical

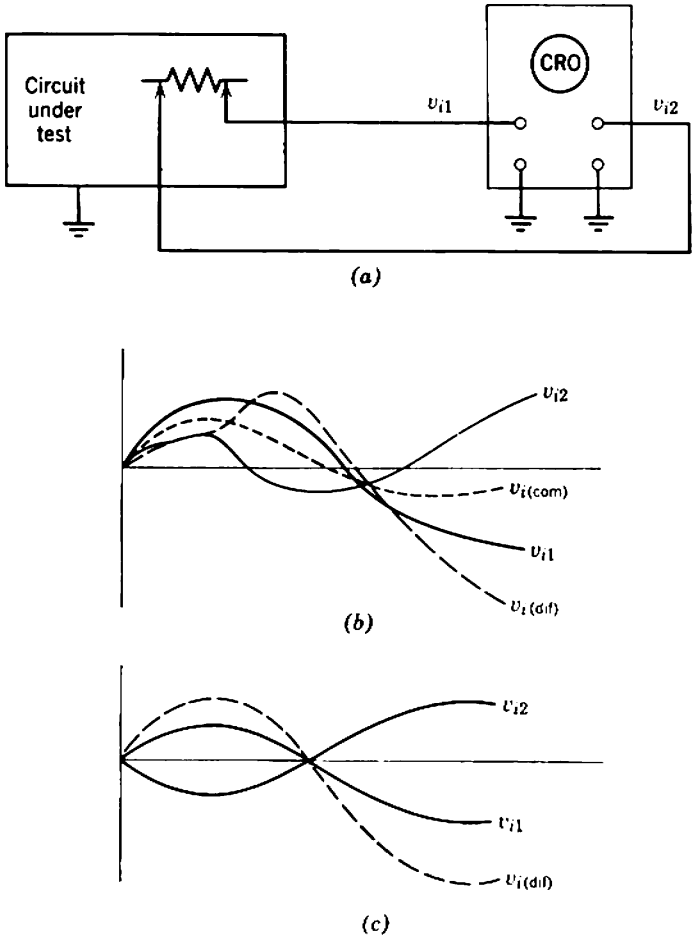


Fig. 15.1 Input signals to a differential amplifier: (*a*) typical oscilloscope application; (*b*) common-mode and differential-mode input signals; (*c*) push-pull input signal (zero common-mode component).

examples are amplifiers for the small currents from photocells and ionization chambers, or the small voltages from thermocouples, strain gauges, and physiological probes. Because there are no restrictions to the dynamic response at low frequency, any change in the quiescent current of an active device (due to a change in the device itself or the associated biasing components) propagates through the amplifier, and is quite indistinguishable from a signal. Such changes are known as *drift*. Drift is a corrupting signal, similar to noise, and design procedures for dc amplifiers aim at increasing the signal-to-drift ratio. Another useful feature of dc amplifiers is their instantaneous recovery from an overload (neglecting high-frequency charge-storage effects in the active devices and stray capacitances). In contrast, the average electrode currents and voltages in a conventional *RC*-coupled amplifier change from their quiescent values when the circuit is overloaded, and excess charge accumulates in the coupling and bypass capacitors. After the overload is removed, this charge must leak away via the supply resistors before the amplifier reverts to normal operation. Overload and recovery are discussed in Section 9.7. The absence from dc amplifiers of capacitors and other energy-storage elements that contribute low-frequency poles eliminates this *paralysis* or *dead-time* effect. Test instruments with dc amplifiers can therefore be used to examine small portions of a large waveform that overloads the amplifier.

The emphasis in preceding chapters is on the design of circuits whose performance is substantially unaffected by changes of devices and components within their normal tolerance limits. Such designable circuits do not require selection of devices (from within a given type number) or components (from within a group of given nominal value), and there is seldom need for preset adjustments. In contrast, adjustments are usually required on differential and dc amplifiers to balance the circuits before they are put into service, and for subsequent rebalancing as the occasion demands. In a sense these circuits are not designable but, if the sources of unbalance are understood, preset controls can be provided and operated in a consistent and logical manner so that optimum performance is maintained for periods of months or years between readjustments.

15.1 SIMPLE DIFFERENTIAL AMPLIFIERS

The long-tailed pair amplifier stage is an interconnection of two active devices; Fig. 15.2 shows the circuit diagrams for triode and transistor realizations. The notation of differential-mode and common-mode

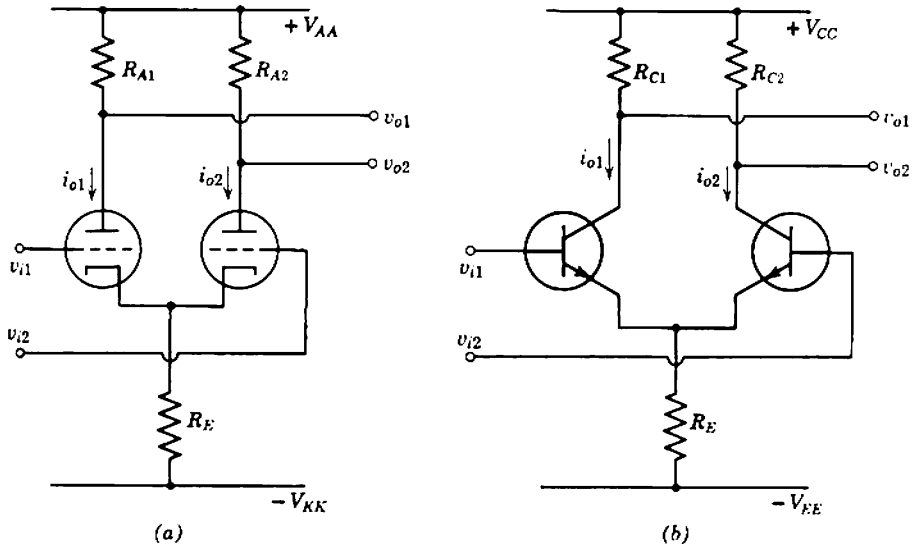


Fig. 15.2 Long-tailed pair differential amplifier: (*a*) vacuum-tube realization; (*b*) transistor realization.

voltages and gains* are introduced consistent with Fig. 15.1 in Eqs. 15.2 to 15.5. For either input or output quantities, distinguished below by additional subscripts *i* or *o*, respectively,

$$v_{\text{dif}} = v_1 - v_2, \tag{15.2}$$

$$v_{\text{com}} = \tfrac{1}{2}(v_1 + v_2). \tag{15.3}$$

Thence,

$$A_{V(\text{dif})} = \frac{v_{o(\text{dif})}}{v_{i(\text{dif})}} = \frac{v_{o1} - v_{o2}}{v_{i1} - v_{i2}}, \tag{15.4}$$

$$A_{V(\text{com})} = \frac{v_{o(\text{com})}}{v_{i(\text{com})}} = \frac{v_{o1} + v_{o2}}{v_{i1} + v_{i2}}. \tag{15.5}$$

The factor $\frac{1}{2}$ appears in Eq. 15.3 for consistency with the special case of v_{i1} and v_{i2} being equal.

If the two devices in a long-tailed pair are identical, the differential-mode voltage gain is equal to the gain of either device used in the normal single-ended configuration with the same load resistance. For vacuum tubes

$$A_{V(\text{dif})} = -\frac{g_m r_A R_L}{r_A + R_L} \approx -g_m R_L, \tag{15.6}$$

* R. D. MIDDLEBROOK, *Differential Amplifiers* (Wiley, New York, 1963).

where the approximate form applies for pentodes, or triodes with small load resistors. For transistors

$$A_{V(\text{dif})} = -\frac{\alpha_N R_L}{r_B/\beta_N + r_E}, \tag{15.7}$$

provided that the load resistance is not too large. In either case the differential gain is independent of the value of R_E because no differential-mode voltage appears across R_E. In the degenerate case in which R_E is set equal to zero, the circuit reduces to two quite independent amplifier stages and, if the devices are identical, the gains are identical also. Therefore the difference between the output signal voltages is just the difference between the input voltages multiplied by the gain of either stage. The collective name for both the long-tailed pair and its degenerate form is the *balanced stage*. Although first-order theory predicts identical performance for the long-tailed pair and its degenerate form, the long-tailed pair is, for reasons given in Section 15.1.1, much the more satisfactory differential amplifier.

A differential-mode transfer conductance can be defined for a balanced stage, but care must be exercised because either of two concepts of differential current might be used. In response to a push-pull signal input (that is, no common-mode component) the signal output currents of the two devices are equal and opposite. If the differential-mode output current is defined as the current which flows right around the load circuit, it is equal to the push-pull output current of either device and needs to be taken as half the difference of device currents, not just the difference. For consistency this definition can be extended to the more general case in which both differential- and common-mode input signals are present, and is used throughout this book:

$$i_{\text{dif}} = \tfrac{1}{2}(i_1 - i_2). \tag{15.8}$$

(The alternative definition omits the factor $\frac{1}{2}$.) Hence the differential transfer conductance of an amplifier is

$$G_{T(\text{dif})} = \frac{i_{o(\text{dif})}}{v_{i(\text{dif})}} = \frac{\frac{1}{2}(i_{o1} - i_{o2})}{v_{i1} - v_{i2}}, \tag{15.9}$$

and its value is half the transfer conductance of either device in a single-ended circuit:

$$G_{T(\text{dif})} = \tfrac{1}{2} g_m \left(\frac{r_A}{r_A + R_L}\right), \tag{15.10}$$

$$G_{T(\text{dif})} = \tfrac{1}{2} \left(\frac{\alpha_N}{r_B/\beta_N + r_E}\right), \tag{15.11}$$

for vacuum tubes and transistors, respectively. It is easy to see the consistency of these expressions for the special case of a single input (say, v_{i1} finite and v_{i2} zero). Then device 1 operates as a series stage having a feedback resistor of $[(r_A + R_L)/g_m r_A]$ or $(r_E + r_B/\beta_N)$, and Eqs. 15.10 and 15.11 give the equal and opposite output currents. The differential-mode voltage gain is

$$A_{V(\text{dif})} = -G_{T(\text{dif})} \times 2R_L, \tag{15.12}$$

where the factor 2 arises because the differential output voltage is developed across the series combination of two load resistances.

Common-mode currents are similarly defined with the factor $\frac{1}{2}$ as

$$i_{\text{com}} = \tfrac{1}{2}(i_1 + i_2). \tag{15.13}$$

In terms of this definition the transfer conductance becomes

$$G_{T(\text{com})} = \frac{i_{o(\text{com})}}{v_{i(\text{com})}} = \frac{i_{o1} + i_{o2}}{v_{i1} + v_{i2}}, \tag{15.14}$$

and the common-mode voltage gain is

$$A_{V(\text{com})} = -G_{T(\text{com})} \times R_L. \tag{15.15}$$

No factor 2 is associated with R_L for common-mode signals, because each device develops the common-mode output voltage across its own load resistor, not the two in series. For common-mode inputs a long-tailed pair can be regarded as two series-feedback stages with common input voltage and feedback resistor. The effective feedback resistance per device is $2R_E$, and the common-mode transfer conductance follows as

$$G_{T(\text{com})} = \frac{1}{1/G_T + 2R_E} \approx \frac{1}{2R_E}, \tag{15.16}$$

where G_T is the transfer conductance of either device in a single-ended common-emitting-electrode circuit. The approximate form of Eq. 15.16 is usually adequate because R_E is ordinarily quite large.

15.1.1 Unbalance

Ideally, any balanced amplifier will serve as a differential amplifier because the differential output voltage depends only on the differential input voltage. A multistage differential amplifier can be formed by cascading a number of balanced stages; each stage amplifies the differential output of the preceding stage. The usual design rules for the signal circuits of an amplifier apply, although biasing of balanced stages is not straightforward; some of the problems are discussed in the example of Section

15.1.2. In practice, the common-mode rejection of a balanced amplifier stage is never infinite (as the simple explanation suggests), but it is considerably greater for the long-tailed pair than for its degenerate form. Possible sources of unbalance are the following:

1. Two devices never have identical characteristics, so the transfer conductances of the two sides of a balanced amplifier are not equal. There is, therefore, a differential-mode component of output current in response to a common-mode input voltage.
2. Inequality of the load resistances for the two sides of a balanced amplifier results in a differential component of output voltage, even if the two sides have equal transfer conductance. Inequality of load resistances is more troublesome in transistor amplifiers than in vacuum-tube amplifiers, because the load resistance for a transistor may involve the unreproducible β_N of the following transistor.
3. Unbalance between coupling capacitors and the stray shunt capacitance causes the common-mode rejection to fall at low and high frequencies, respectively.

The balance and common-mode rejection of elementary circuits can be improved up to a point by selecting active devices for close matching of parameters. Almost perfect balance can be achieved at any one quiescent point by adding resistors and capacitors in series or shunt with the device terminals to trim out the differences. However, the quiescent point shifts dynamically under signal conditions and the hard-won balance is disturbed. Thus, apart from the obvious necessity for balancing the circuit, the design of a differential amplifier should aim at minimizing the dynamic shift of quiescent point under common-mode signal conditions, and this implies reducing the common-mode transfer conductance. A major advantage of the long-tailed pair in differential amplifier circuits is thus apparent; $G_{T(\text{com})}$ and $A_{V(\text{com})}$ can be made as small as desired by increasing R_E, without affecting the values of $G_{T(\text{dif})}$ and $A_{V(\text{dif})}$.

A second advantage of the long-tailed pair is that the common-mode rejection of a multistage amplifier can be made to depend almost entirely on its first stage. Consider a two-stage amplifier in which the stages have common-mode rejection factors λ_1 and λ_2. If the common-mode input signal to the first stage is $v_{i(\text{com})1}$, the input to the second stage consists of a common-mode component and a differential-mode component:

$$v_{i(\text{com})2} = v_{o(\text{com})1} = v_{i(\text{com})1} \times A_{V(\text{com})1},$$

$$v_{i(\text{dif})2} = v_{o(\text{dif})1} = v_{i(\text{com})1} \times \frac{A_{V(\text{dif})1}}{\lambda_1}.$$

The output from the second stage contains one common-mode component and two differential-mode components:

$$v_{o(\text{com})2} = v_{i(\text{com})2} \times A_{V(\text{com})2} = v_{i(\text{com})1} \times A_{V(\text{com})1} \times A_{V(\text{com})2},$$

$$v'_{o(\text{dif})2} = v_{i(\text{dif})2} \times A_{V(\text{dif})2} = v_{i(\text{com})1} \times \frac{A_{V(\text{dif})1}}{\lambda_1} \times A_{V(\text{dif})2},$$

$$v''_{o(\text{dif})2} = v_{i(\text{com})2} \times \frac{A_{V(\text{dif})2}}{\lambda_2} = v_{i(\text{com})1} \times A_{V(\text{com})1} \times \frac{A_{V(\text{dif})2}}{\lambda_2}.$$

In the worst case the two differential components of output add and the total is

$$v_{o(\text{dif})2} = v_{i(\text{com})1}\left[\frac{A_{V(\text{dif})1} \times A_{V(\text{dif})2}}{\lambda_1} + \frac{A_{V(\text{com})1} \times A_{V(\text{dif})2}}{\lambda_2}\right].$$

The same differential output is produced by a differential input

$$v_{i(\text{dif})1} = \frac{v_{o(\text{dif})2}}{A_{V(\text{dif})1} \times A_{V(\text{dif})2}}.$$

Therefore, the reciprocal of the common-mode rejection factor for a two-stage amplifier is

$$\frac{1}{\lambda} = \frac{v_{i(\text{dif})1}/v_{o(\text{dif})2}}{v_{i(\text{com})1}/v_{o(\text{dif})2}} = \frac{1}{\lambda_1} + \frac{1}{\lambda_2}\left[\frac{A_{V(\text{com})1}}{A_{V(\text{dif})1}}\right].$$

The second term can be made very small by designing the circuit so that $A_{V(\text{com})1}$ is much less than $A_{V(\text{dif})1}$, and λ approaches λ_1.

The argument can be extended to the general case of a multistage amplifier to show that the reciprocal of the common-mode rejection factor is

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{A_{V(\text{com})1}}{A_{V(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \left[\frac{A_{V(\text{com})2}}{A_{V(\text{dif})2}}\right]\left\{\frac{1}{\lambda_3} + \left[\frac{A_{V(\text{com})3}}{A_{V(\text{dif})2}}\right]\left\{\cdots\left\{\frac{1}{\lambda_n}\right\}\cdots\right\}\right\}\right\}. \tag{15.17}$$

By designing the individual stages with R_E so large that their common-mode gain is much smaller than their differential gain, λ can be increased approximately to λ_1. Progressively less careful balancing is therefore acceptable for stages beyond the first. Indeed, after two or three long-tailed pairs it is usually permissible to revert to single-ended stages. The common-sense interpretation is that the later stages are subjected to relatively smaller common-mode signals, so that any unbalance is less significant. In principle, some of the signs in Eq. 15.17 may be negative because unbalance in one stage may partially compensate unbalance in

another. Nevertheless, it is prudent to use the pessimistic form with all signs positive.

Often, reversion of a differential amplifier to single-ended stages is mandatory. In the classification of Section 7.2.1 loads are divided into two categories: single-ended loads in which one terminal is grounded and the signal is applied to the other, and balanced or floating loads which respond merely to the voltage impressed across them. Single-ended loads are the more common; these must be connected between one output terminal of a differential amplifier and ground, where the voltage impressed on them contains common-mode as well as differential-mode components. In the special case of a perfectly balanced n-stage differential amplifier with a single-ended load, the output currents in response to differential-mode and common-mode input voltages are

$$i'_o = v_{i(\mathrm{dif})} \times A_{V(\mathrm{dif})1} \times A_{V(\mathrm{dif})2} \times \cdots \times A_{V(\mathrm{dif})(n-1)} \times G_{T(\mathrm{dif})n},$$

$$i''_o = v_{i(\mathrm{com})} \times A_{V(\mathrm{com})1} \times A_{V(\mathrm{com})2} \times \cdots \times A_{V(\mathrm{com})(n-1)} \times G_{T(\mathrm{com})n}.$$

The common-mode rejection factor due to the single-ended load follows as

$$\lambda = \frac{v_{i(\mathrm{com})}/i''_o}{v_{i(\mathrm{dif})}/i'_o} = \frac{A_{V(\mathrm{dif})1}}{A_{V(\mathrm{com})1}} \times \frac{A_{V(\mathrm{dif})2}}{A_{V(\mathrm{com})2}} \times \cdots \times \frac{A_{V(\mathrm{dif})(n-1)}}{A_{V(\mathrm{com})(n-1)}} \times \frac{G_{T(\mathrm{dif})n}}{G_{T(\mathrm{com})n}}.$$

More generally, an n-stage differential amplifier consists of k balanced stages, $(n - k)$ single-ended stages and, finally, a single-ended load. The expression for the total common-mode rejection factor (corresponding to Eq. 15.17 for an amplifier with a balanced load) is

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{A_{V(\mathrm{com})1}}{A_{V(\mathrm{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \left[\frac{A_{V(\mathrm{com})2}}{A_{V(\mathrm{dif})2}}\right]\left\{\cdots\left\{\frac{1}{\lambda_k} + \frac{G_{T(\mathrm{com})k}}{G_{T(\mathrm{dif})k}}\right\}\cdots\right\}\right\}. \tag{15.18}$$

Compared with Eq. 15.17 for an amplifier consisting entirely of balanced stages, Eq. 15.18 contains less terms. Equation 15.17 can be written as

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{A_{V(\mathrm{com})1}}{A_{V(\mathrm{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \left[\frac{A_{V(\mathrm{com})2}}{A_{V(\mathrm{dif})2}}\right]\left\{\cdots\left\{\frac{1}{\lambda_k} + \left[\frac{G_{T(\mathrm{com})k}}{G_{T(\mathrm{dif})k}}\right]\right.\right.\right.$$
$$\left.\left.\left. \times \overline{\left\{\left[\frac{R_{L(\mathrm{com})k}}{R_{L(\mathrm{dif})k}}\right]\left\{\frac{1}{\lambda_{k+1}} + \left[\frac{A_{V(\mathrm{com})(k+1)}}{A_{V(\mathrm{dif})(k+1)}}\right]\left\{\cdots\right\}\right\}\right\}\cdots\right\}\right\},$$

and the term under the bar is the increase over Eq. 15.18. As the value of this term is usually less than one, the common-mode rejection factor is greater for the completely balanced amplifier than for the amplifier that reverts to single-ended stages.

The order of common-mode rejection that can be obtained depends on the balance of the circuit. In the center of the mid-band frequency range, values of 10^4 can be obtained quite easily, 10^5 can be obtained with reasonable care, and values in excess of 10^6 are not impossible. The attainable rejection ratio falls off at both low and high frequencies, reaching a small value at the lower and upper 3-dB frequencies. Sources of unbalance and means for correcting them are discussed in the following sections.

15.1.1.1 Mid-Band Unbalance

The dynamic shift in quiescent current and consequent unbalance in a long-tailed pair with a common-mode input can be reduced by increasing the emitting-electrode supply resistor R_E. With reasonable supply voltages and operating currents, the upper bound to R_E is quite small—perhaps a few tens of kilohms; larger values are often desirable. A technique for realizing a very large effective supply resistance is to use another tube or transistor as a current source for the long-tailed pair, Fig. 15.3. This extra device has a stable emitting-electrode-feedback

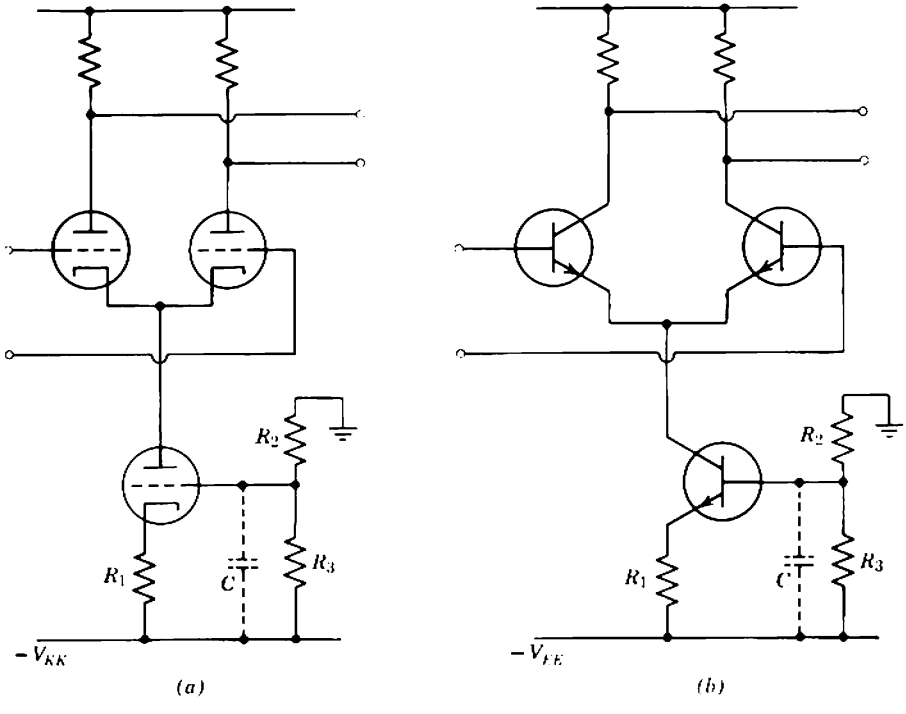


Fig. 15.3 Realization of current source: (*a*) vacuum tubes; (*b*) transistors.

biasing system which fixes its dc output current at the total value required by the pair. From a signal point of view the source is a series-feedback stage which has a very large output resistance. It is thus a very good approximation to a dc current source; output resistances of several megohms are obtained easily. Notice that the emitting-electrode resistor R_1 should not be bypassed, as this removes the series feedback and lowers the output impedance. The capacitor (shown broken) between the control electrode and negative supply reduces the source impedance at medium and high frequencies, thereby increasing the loop gain and reducing the noise voltage developed at the junction of R_2 and R_3. A pentode is preferable to a triode as a current source, because of its inherently larger output (anode) resistance; its screen should be fed directly from a dc supply rail, not via an RC supply network. With vacuum tubes the output resistance can be made as large as desired by applying enough series feedback. With transistors the output resistance cannot exceed the collector-to-base resistance of the transistor, $\beta_N r_C$ in shunt with the collector junction leakage resistance.

A matter of some importance is the way in which unbalance between the two devices of a long-tailed pair degrades the common-mode rejection. Once the mechanism is understood, optimum methods for trimming out the unbalance can be formulated. Consider first the hypothetical charge-controlled device without recombination; Fig. 15.4 shows the equivalent

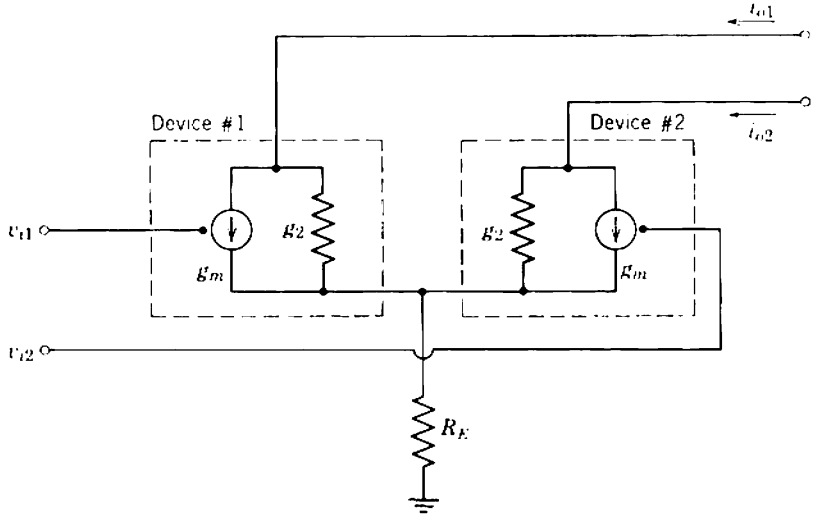


Fig. 15.4 Unbalance in a long-tailed pair.

circuit. The condition for balance between short-circuit output currents i_{o1} and i_{o2} is

$$\frac{1}{g_{m2}} - \frac{1}{g_{m1}} = 2\left(\frac{\mu_2 - \mu_1}{\mu_1 \mu_2}\right) R_E, \tag{15.19}$$

where μ is the voltage amplification factor of the device:

$$\mu = \frac{g_m}{g_2}.$$

Two important conclusions can be drawn from Eq. 15.19. First, since μ for a device is almost independent of the quiescent current whereas g_m is critically dependent on the current, a long-tailed pair can be balanced by altering the ratio of the currents in the two devices. Second, R_E should not be too large, or an excessive inequality between g_{m1} and g_{m2} may be required at the balance point. Depending on the matching between μ_1 and μ_2 (which is usually good because μ is a reproducible device parameter) the value of R_E may be a few times $1/g_2$; the largest convenient value should be used, to keep the common-mode transfer conductance small. Usually the output resistance of a current-source circuit is too large, whereas a simple resistor for R_E is too small. A satisfactory arrangement is a current-source to provide the dc current, with a resistor connected between the source output terminal and ground to reduce the effective resistance (Fig. 15.5).

Another source of unbalance is inequality of the load resistances seen by the two sides of a long-tailed pair. Inequalities of a few percent are tolerable, even in the first stage of a differential amplifier, because the common-mode signal at this point is attenuated relative to the differential-mode signal by the ratio of $G_{T(\text{dif})}$ to $G_{T(\text{com})}$. Moreover, the inequality

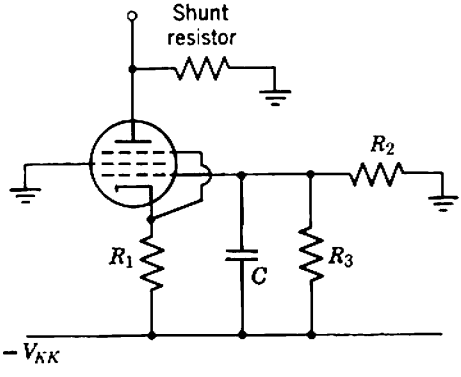
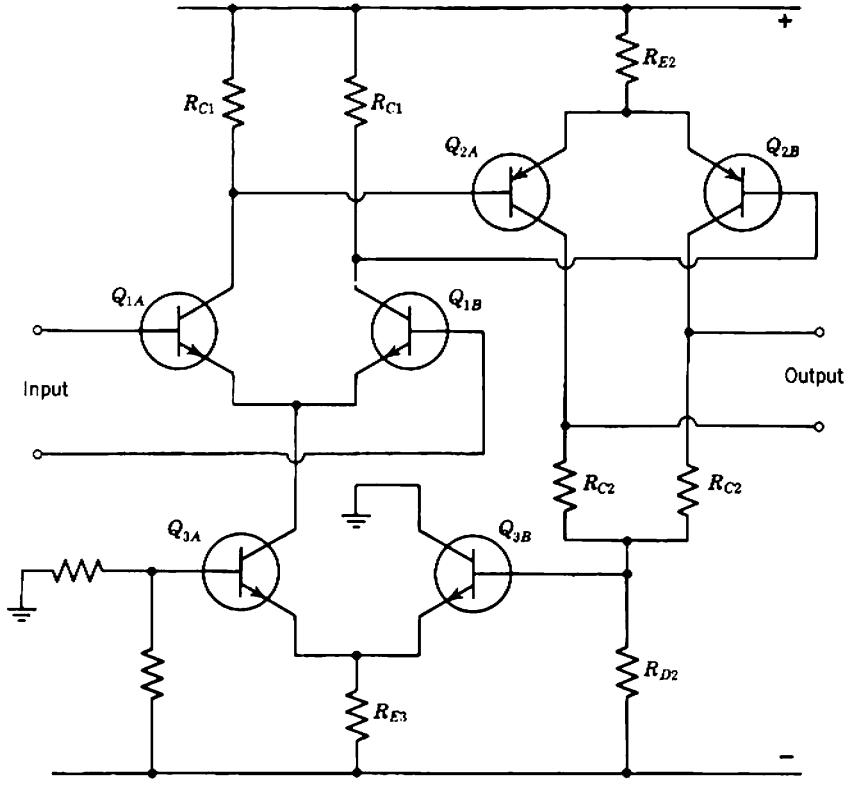
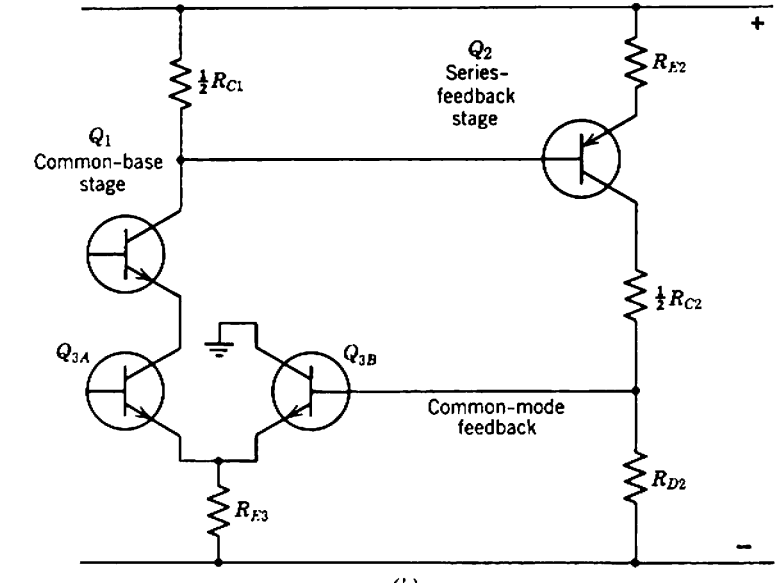


Fig. 15.5 Method of controlling R_E to assist balancing.



(a)



(b)

Fig. 15.6 Common-mode negative feedback to increase common-mode rejection: (*a*) complete circuit; (*b*) common-mode feedback path.

can be compensated by setting the quiescent currents of the preceding stage slightly away from the ideal balance point. The supply resistors in the load network of a differential amplifier stage should be high-stability types, so that any inequality is stable.

If recombination in the devices is considered, a further source of unbalance is signal feedthrough via the device feedback conductance g_f. Even for transistors this conductance is so small that the possible unbalance is often negligible and, in any case, it can be taken up by adjusting the ratio of the quiescent currents. Notice that inequality between device input conductances does not unbalance a stage. However, the input conductance may unbalance the preceding stage as g_1 forms part of its load resistance. With vacuum-tube amplifiers, or with transistor amplifiers which have small supply resistors so that the gain is stable, this is not a serious problem because the load is dominated by the stable supply resistors. In transistor amplifiers having large supply resistors, however, unbalance of the loads is a source of considerable analytic complexity. The load resistances are unstable because they depend on the unreproducible β_N of the following transistor, so compensation is not satisfactory. In addition, the differential-mode and common-mode input resistances of a transistor are not equal; in general

$$R_i = \frac{\beta_N}{G_T}$$

and $R_{i(com)}$ is much larger than $R_{i(dif)}$. This reduces the ratio of differential-mode to common-mode voltage gain for the preceding stage, further degrading the total common-mode rejection. Figure 15.6*a* shows a satisfactory approach, in which common-mode negative feedback is taken from the output of the second stage and applied to the current source of the first stage. Figure 15.6*b* shows the common-mode signal path. Because of the over-all feedback, the effective output resistance of the current source can be raised well above $\beta_N r_C$; this minimizes the common-mode transfer conductance and signal output of the first stage, and reduces the significance of unbalance in the load. By use of selected dual transistors, common-mode rejection ratios as high as 10^6:1 have been claimed.*

15.1.1.2 Low-Frequency Unbalance

In a well-designed differential amplifier, the only source of unbalance peculiar to low frequencies (i.e., additional to the mid-band unbalance) is inequality between interstage coupling time constants. Decoupling

* For example, S. W. HOLCOMB, "Dual transistors in low-level circuits," *Texas Instruments, Inc. Report*, July 1963.

capacitors and screen bypass capacitors can be arranged so that they do not contribute unbalance, and there is of course no emitting-electrode bypass capacitor in a long-tailed pair. Figure 15.7 illustrates suitable circuit arrangements for a pentode. The decoupling capacitor should be common to both sides of a long-tailed pair, so that any low-frequency signal voltage developed across it is common to both sides (i.e., it is a common-mode signal) and is rejected by the next stage. Desirably, screen grids should be fed directly from a supply rail so that no signal voltage can be developed on them. If this is not possible, the screens should be tied together and fed from a low-impedance supply network as in Fig. 15.7. Ideally, no bypass capacitor is required because a differential input produces equal and opposite screen currents, although a small capacitor is usually included to suppress any voltage due to high-frequency unbalance.

Inequality between the coupling time constants at the input of a long-tailed pair unbalances the gain at low frequencies, resulting in noninfinite common-mode rejection. Progressively more unbalance can be tolerated in the later stages. The input circuit to the first stage of an amplifier is the most critical point of all; if possible, the signal should be applied directly

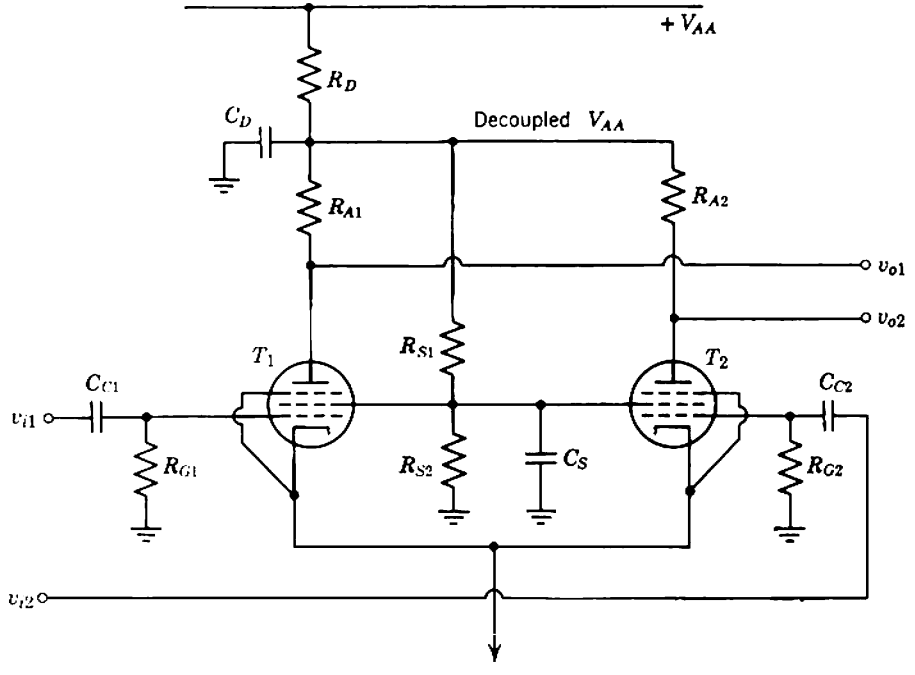


Fig. 15.7 Decoupling and screen supply for a pentode differential amplifier.

to the control electrodes of the first stage. Low-frequency unbalance is a particular problem in transistor circuits, because the time constants depend on the unreproducible β_N; further, the coupling capacitors are usually electrolytic types, and these have large production tolerances.

The unbalance in the transmission through two coupling networks depends on the signal frequency, and obviously falls to zero at high frequencies where the reactance of the capacitors is small. At frequencies well inside the passband, the equivalent common-mode rejection due to unbalance in coupling time constants is

$$\lambda_{\text{coupling}} = \text{const}\left(\frac{\omega}{\omega_{co}}\right), \tag{15.20}$$

where ω_{co} is the nominal 3-dB frequency, and the constant depends on the ratio of time constants. In vacuum-tube amplifiers using unselected 10% tolerance components, the constant is unlikely to be less than 10; selection or adjustment of components can raise its value by one or two orders of magnitude. The constant is smaller in transistor amplifiers due to the effects mentioned above; 3 is a likely minimum value for unselected components. Equation 15.20 (or the equation for any other frequency-dependent common-mode rejection factor) can be derived as follows:

Suppose the actual cutoff frequencies of the two sides of the circuit are ω'_{co} and ω''_{co}. If a 1-V common-mode signal is present on the input side of the coupling capacitors, the difference between the output voltages is

$$v_{\text{error}} = \frac{j\omega/\omega'_{co}}{1 + j\omega/\omega'_{co}} - \frac{j\omega/\omega''_{co}}{1 + j\omega/\omega''_{co}}.$$

A differential input equal to v_{error} will produce the same differential output. But λ is defined as the ratio of common-mode and differential inputs that produce the same output. Thus

$$\frac{1}{\lambda_{\text{coupling}}} = \frac{1}{1 - j\omega'_{co}/\omega} - \frac{1}{1 - j\omega''_{co}/\omega}.$$

If ω is much greater than ω_{co} (i.e., if the signal frequency is well within the passband) then

$$\frac{1}{\lambda_{\text{coupling}}} \approx \left(1 + \frac{j\omega'_{co}}{\omega}\right) - \left(1 + \frac{j\omega''_{co}}{\omega}\right) = j\left(\frac{\omega'_{co} - \omega''_{co}}{\omega}\right).$$

Inverting, and dropping the j to give magnitudes only,

$$\lambda_{\text{coupling}} \approx \frac{\omega}{\omega'_{co} - \omega''_{co}} = \text{const}\left(\frac{\omega}{\omega_{co}}\right),$$

and the constant becomes larger as ω'_{co} approaches ω''_{co}.

15.1.1.3 High-Frequency Unbalance

Unbalance peculiar to high-frequencies stems from inequality between the stray capacitances associated with the two sides of a long-tailed pair. The capacitance to ground at any point forms part of the load for the preceding device, and any inequality unbalances the voltage gain at high frequencies. The equivalent common-mode rejection due to unbalance of the stray capacitance obviously increases with decreasing frequency, becoming infinite at zero frequency. Its value is

$$\lambda_{\text{stray}} = \text{const}\left(\frac{\omega_{co}}{\omega}\right), \tag{15.21}$$

where ω_{co} is the nominal cutoff frequency due to the stray capacitance concerned. The derivation of Eq. 15.21 is similar to that for Eq. 15.20, and the constant is unlikely to be less than 3 if unselected components are used. The use of selected components or the addition of balancing capacitors can raise the constant by one or two orders of magnitude.

The output-to-input capacitance C_F of the first stage in a differential amplifier may contribute appreciably to the high-frequency unbalance. This capacitance contributes a direct-feedthrough component to the transfer admittance (like g_f at mid-band frequencies) and the common-mode rejection is not infinite unless balance is maintained:

$$\lambda_{\text{feedthrough}} = \text{const}\left(\frac{G_{T(\text{dif})}}{\omega C_F}\right). \tag{15.22}$$

The constant is unlikely to be less than 3 with unselected devices; addition of the balancing capacitors shown in Fig. 15.8 can raise its value by an order of magnitude. Balancing is usually required for triodes and transistors, but not for pentodes because their anode-to-grid capacitance is small.

Direct signal feedthrough via C_F contributes a component to $Y_{T(\text{com})}$ which increases with increasing frequency:

$$|Y_{T(\text{com})}|_{\text{direct}} = \omega C_F. \tag{15.23}$$

If $G_{T(\text{dif})}$ is rather small, the ratio of $G_{T(\text{dif})}$ to $Y_{T(\text{com})}$ may not be large at high frequencies; the second stage may be subjected to large common-mode signals, and careful balancing may be required. As a better alternative, the null capacitor shown in Fig. 15.8 can be used to reduce $Y_{T(\text{com})}$ to its mid-band value $G_{T(\text{com})}$; if

$$C_{\text{null}} = 2C_F, \tag{15.24}$$

the direct and normal high-frequency components of $Y_{T(\text{com})}$ are equal

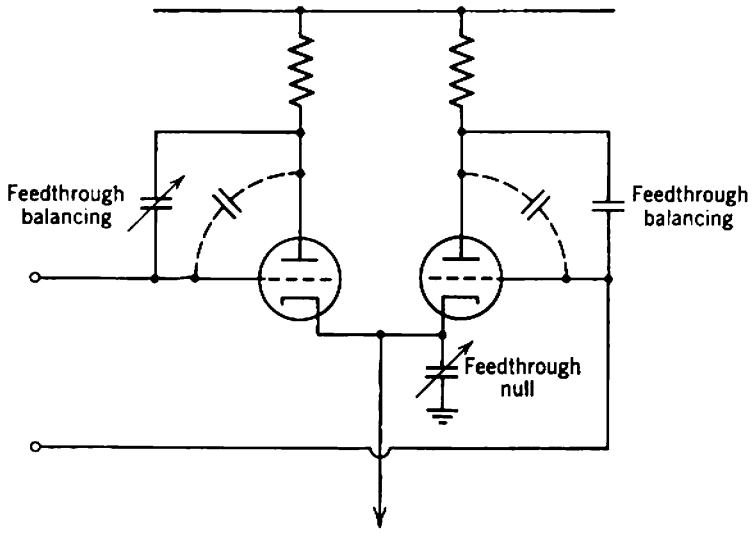


Fig. 15.8 Capacitors for high-frequency balance.

and opposite. It is only in very rare instances that balanced high-frequency feedthrough is significant; C_{null} is seldom required.

15.1.2 Illustrative Example

In general, there are many designs of differential amplifiers that fulfill given specifications. The differences between these designs are largely in circuit arrangement, and the choice of the best design cannot be reduced to a study of formal equations. The circuit designer is largely on his own initiative. As an example, the following paragraphs discuss the features of an amplifier which is to meet the following specifications:

(i) mid-band differential gain: 2×10^4;
(ii) common-mode rejection: 10^5:1 over the frequency range 10 Hz to 10 kHz;
(iii) output capability at least 10 V rms into a high resistance.

The circuit shown in Fig. 15.9 consists of three *RC*-coupled long-tailed pair stages—one triode stage followed by two pentode stages. A single-ended output signal is taken from one side of the final stage; the signal output current from the other side is disregarded, the anode being connected directly to the supply rail. The critical first stage uses a twin triode tube rather than separate tubes, because balance between the sections of a twin tube is likely to be better than balance between two separate tubes.

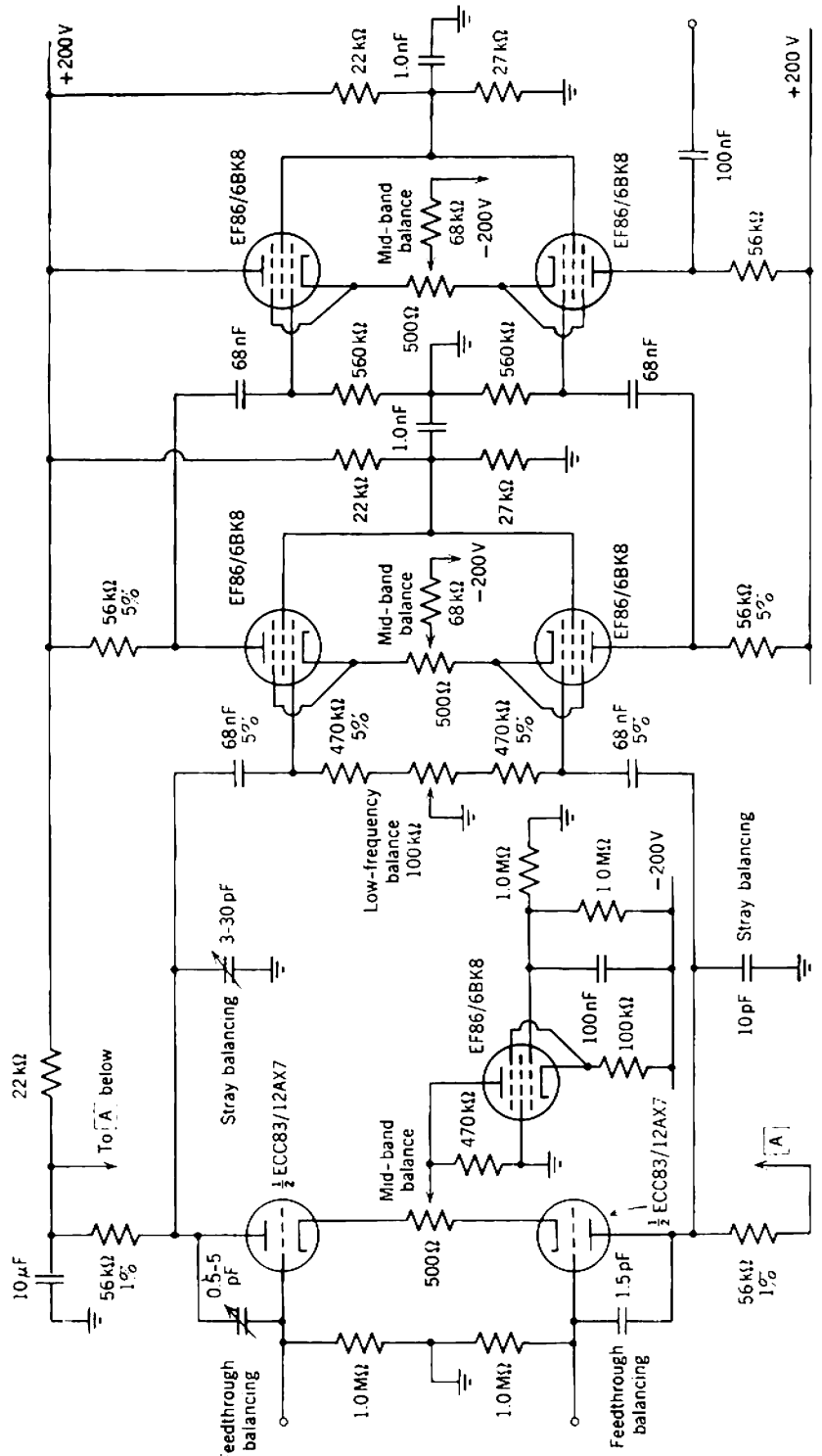


Fig. 15.9 Circuit diagram for a vacuum-tube differential amplifier.

In addition, a current-source tube is used in the cathode circuit. The output resistance of the current source is padded down by a 470-kΩ resistor, this value being about five times the anode resistance of the twin triode type ECC83/12AX7. Separate pentodes are used in the second and third stages where balance is less critical, to give increased gain and improved high-frequency response; simple current-feed resistors are used in their cathode circuits, and their screens are fed from (relatively) low-resistance voltage dividers.

Balancing of the stages is accomplished by a small potentiometer in the cathode circuit. Adjustment of this potentiometer has two effects. First, it unbalances the negative grid bias applied to two tubes, thereby altering the ratio of their quiescent currents and mutual conductances; this is the balancing method discussed in Section 15.1.1. Second, it applies series feedback to the two tubes. If the potentiometer is set away from its mid position, the tube with the higher proportion of resistance has the larger grid bias and the more feedback. Both effects reduce G_T; thus the two balancing mechanisms aid each other, rather than oppose. Balancing of a long-tailed pair by a potentiometer in the cathode circuit is usually the most satisfactory method. The total resistance of the potentiometer should be about

$$R_F \approx \frac{1}{2g_m} \tag{15.25}$$

to give adequate range of control without excessive loss of differential-mode gain due to the feedback. An alternative balancing circuit is shown in Fig. 15.10; adjustment of R_G''' alters the voltages at the grids and changes the ratio of the quiescent currents. The disadvantage is that the balancing potentiometer also changes the input coupling time constant, possibly unbalancing the stage at low frequencies. Notice that the biasing arrangement shown in Fig. 15.11 should not be used. Inequalities of a few percent between R_{G1}' and R_{G2}' or R_{G1}'' and R_{G2}'' may give so great a difference between the grid voltages that one tube is always cut off. Grid resistors for the two sides of a long-tailed pair should be returned to the same supply potential. Needless to say, all the above comments on the balancing and biasing of vacuum tubes apply equally to transistors.

The nominal operating conditions, small-signal parameters, and transfer functions of the stages are

FIRST STAGE:

$$V_{AK} = 150 \text{ V},$$
$$I_K = 0.5 \text{ mA},$$

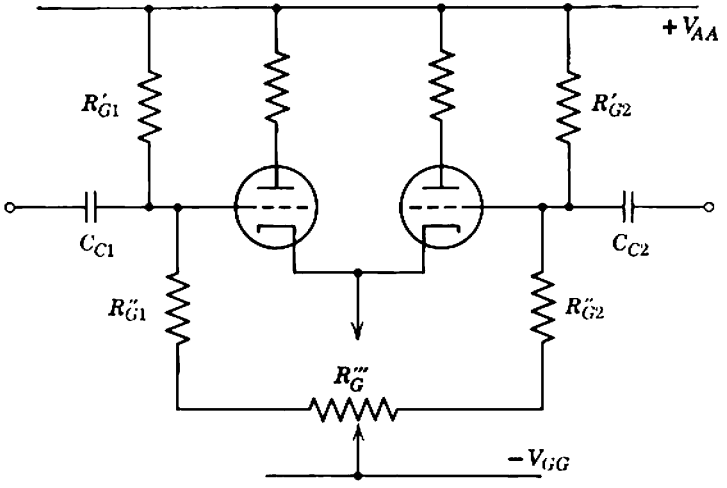


Fig. 15.10 Alternative balancing circuit to that in Fig. 15.9.

for which

$$g_m = 1 \text{ mA/V},$$
$$r_A = 100 \text{ k}\Omega.$$

Therefore

$$G_{T(\text{dif})1} = 290 \ \mu\text{A/V}, \qquad A_{V(\text{dif})1} = 29,$$
$$G_{T(\text{com})1} = 1.06 \ \mu\text{A/V}, \qquad A_{V(\text{com})1} = 0.021.$$

Notice that the differential-mode feedback provided by the balancing potentiometer is taken into account in calculating $G_{T(\text{dif})1}$.

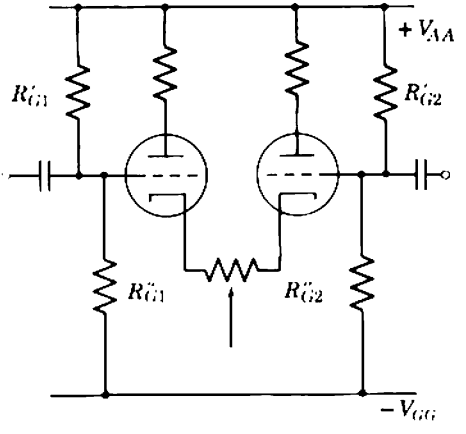


Fig. 15.11 A balancing circuit that should never be used.

SECOND STAGE

$$V_{SK} = 100 \text{ V},$$
$$I_K = 1.5 \text{ mA},$$

for which

$$g_m = 1 \text{ mA/V}.$$

Therefore

$$G_{T(\text{dif})2} = 380\ \mu\text{A/V}, \qquad A_{V(\text{dif})2} = 38,$$
$$G_{T(\text{com})2} = 5.8\ \mu\text{A/V}, \qquad A_{V(\text{com})2} = 0.29.$$

Notice that the dividing factor $(1 + k)$ due to the screen current of a pentode is taken into account when considering the feedback provided by the resistances in the cathode circuit (see Eq. 11.15, for example).

THIRD STAGE

$$V_{SK} = 100 \text{ V},$$
$$I_K = 1.5 \text{ mA},$$

for which

$$g_m = 1 \text{ mA/V}.$$

Therefore

$$G_{T(\text{dif})3} = 380\ \mu\text{A/V}, \qquad A_{V(\text{dif})3} = 19,$$
$$G_{T(\text{com})3} = 5.8\ \mu\text{A/V}, \qquad A_{V(\text{com})3} = 0.29.$$

If the over-all common-mode rejection ratio of 10^5:1 is to be achieved, Eq. 15.18 shows that the performance of the individual stages must be at least

$$\lambda_1 \geq 10^5,$$
$$\lambda_2 \geq 72,$$
$$\lambda_3 \geq 0.55.$$

and also

$$\left[\frac{A_{V(\text{dif})1}}{A_{V(\text{com})1}}\right] \times \left[\frac{A_{V(\text{dif})2}}{A_{V(\text{com})2}}\right] \times \left[\frac{G_{T(\text{dif})3}}{G_{T(\text{com})3}}\right] \geq 10^5.$$

Desirably, the inequalities except that for λ_1 should be satisfied adequately, so that λ_1 makes the dominant contribution to the total.

All reasonable care is taken in balancing the first stage at mid-band frequencies: a current source with padding resistor is used in the cathode circuit, a twin tube is used, and a balancing potentiometer is provided in the cathode circuit. The two sides should balance to better than $1:10^5$, and the specification for common-mode rejection is met. Selection of the tube may be necessary.

The second stage must satisfy $\lambda_2 \geq 72$. This is a small value, and can be met by quite a crude circuit. Nevertheless, good circuit practice is followed as it costs little; a balancing potentiometer is provided in the cathode

circuit and the screens are tied together and fed from a low-impedance supply network. A cathode current-source tube is not warranted, and even the balancing potentiometer could probably be omitted. Certainly, the balancing procedure for this stage should not be elaborate—it would be sufficient to equalize the quiescent currents in the two tubes without checking for maximum common-mode rejection. Any unbalance in this stage can be taken up by deliberately unbalancing the first stage.

The load resistors for the first stage are at the input of the second stage, and should be stable to 1 part in λ_2. High-stability 1% resistors are specified. Notice, however, that the absolute values of the loads need not come within this ratio, as the output currents from the first stage can be deliberately unbalanced to take up the difference in the loads and equalize $A_{V(\text{com})}$ for the two sides; a ratio 3:2 in the loads can be accommodated if Eq. 15.25 is satisfied. The 1% resistors are specified to achieve stability of the load, not equality.

The third stage may be very simple indeed as the required value of λ_3 would be met even if the two tubes had quite dissimilar characteristics. Still, good practice costs little. Standard-grade 10% tolerance resistors are specified for all positions.

After three stages

$$\left[\frac{A_{V(\text{dif})1}}{A_{V(\text{com})1}}\right] \times \left[\frac{A_{V(\text{dif})2}}{A_{V(\text{com})2}}\right] \times \left[\frac{G_{T(\text{dif})3}}{G_{T(\text{com})3}}\right] = 1.2 \times 10^7.$$

This exceeds the minimum allowed value by a factor 120, so there would be no purpose at all in using balanced circuits for any subsequent stages. In point of fact, two balanced stages would very nearly suffice, because

$$\left[\frac{A_{V(\text{dif})1}}{A_{V(\text{com})1}}\right] \times \left[\frac{G_{T(\text{dif})2}}{G_{T(\text{com})2}}\right] = 9.0 \times 10^4,$$

which only just falls short of the required value, 10^5.

Inequality between the coupling time constants causes the common-mode rejection to fall at low frequencies. The minimum permissible values of common-mode rejection for the coupling networks between the first and second and second and third stages are λ_2 and λ_3 of the mid-band analysis above, namely, 72 and 0.55, respectively. Notice that there are no coupling capacitors in the critical input circuit, where the value of λ_1 ($= 10^5$) would apply. A balancing potentiometer is provided in the grid circuit of the second stage; with adjustment, the value of λ_2 can be achieved down to the cutoff frequency (Eq. 15.20). The minimum value of λ_3 is so small that the specification will be met at all frequencies down to the nominal cutoff frequency of the network even if 10% tolerance com-

ponents are used. Both interstage coupling networks are designed for 5 Hz cutoff frequency, one octave below the specified system cutoff frequency. No other capacitors contribute to the low-frequency common-mode rejection.

The stray capacitance from the two sides of the amplifier to ground can be kept down to about 30 pF between each of the stages, provided that reasonable care is taken with the physical layout and construction. The capacitances should be reasonably well balanced if the layout is physically symmetrical. In conjunction with the total shunting resistance of about 50 kΩ, each stage has a cutoff frequency around 100 kHz. The permissible common-mode rejection at the inputs of the second and third stages are λ_2 and λ_3 respectively, and the upper frequencies at which these values will be achieved follow from Eq. 15.21 as:

$$f_2 = 4 \text{ kHz}, \qquad f_3 = 500 \text{ kHz}.$$

The latter is far beyond the specified 10 kHz so no balancing capacitors are provided. The former is rather low; accordingly, a 10 pF fixed capacitor is added to one side of the amplifier, and a 3 to 30 pF trimmer to the other.

Another source of high-frequency unbalance is direct feedthrough via the anode-to-grid capacitance of the first stage. This capacitance totals about 2 pF, including the tube socket and leads. From Eq. 15.23, the frequency at which the direct component of $Y_{T(\text{com})}$ becomes comparable with the normal component is about 80 kHz; this is so high that no null capacitor need be provided in the cathode circuit. Unless balancing capacitors are provided, however, Eq. 15.22 shows that the frequency at which unbalance in the direct common-mode transmission exceeds the permissible $1:10^5$ is about 1 kHz. This can be raised to the desired value of 10 kHz by adding a fixed 1.5 pF capacitor and a 0.5 to 5.0 pF trimmer as shown.

15.2 FEEDBACK DIFFERENTIAL AMPLIFIERS

Negative feedback is often applied to differential amplifiers to stabilize their differential gain. Local feedback can be applied to each stage and, as with single-ended amplifiers, a cascade of alternate series- and shunt-feedback stages is the most satisfactory arrangement (Section 15.2.1). Over-all feedback finds little application for amplifiers with differential-mode input signals, but useful amplifiers for single-ended inputs can be realized by feeding the input signal to one input of a differential amplifier and the feedback signal to the other (Section 15.2.2).

15.2.1 Single-Stage Feedback

Series-feedback stages require little comment. Indeed, the normal method for balancing a long-tailed pair by means of a potentiometer in the emitting-electrode circuit is in fact a series-feedback arrangement. The differential transfer conductance of the stage shown in Fig. 15.12 is

$$G_{T(\text{dif})} = \tfrac{1}{2}G_T, \tag{15.26}$$

where G_T is the transfer conductance of a single-ended stage, with $(R'_F + R''_F)$ as feedback resistor. The common-mode transfer conductance is

$$G_{T(\text{com})} \approx \frac{1}{2R_E}. \tag{15.27}$$

Both results are as expected from considerations of the long-tailed pair without feedback. High-frequency peaking and balancing capacitors may be added as shown.

The shunt-feedback stages in a differential amplifier should not be connected as long-tailed pairs. Rather, the two sides of a balanced shunt-feedback stage should be completely isolated from each other as in Fig.

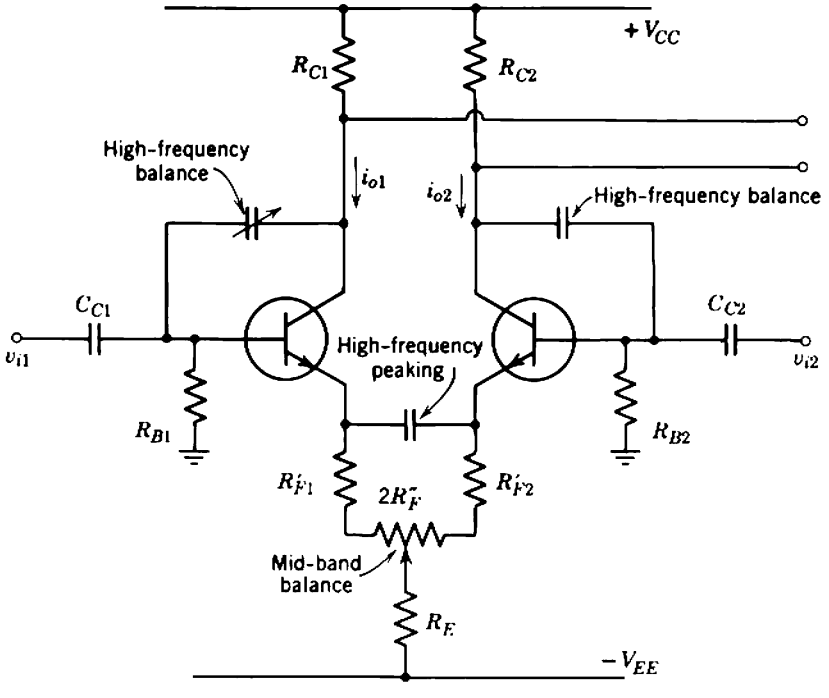


Fig. 15.12 Series-feedback differential amplifier stage.

15.13, with the emitting electrodes grounded for signals either by connecting them directly to ground, connecting them to an appropriate dc supply rail, or connecting them together and bypassing them to ground with a capacitor. If the emitting electrodes are not grounded, but instead are connected as a long-tailed pair, the common-mode input resistance is large. Any common-mode output current from the preceding series stage develops a large common-mode voltage, and the effect of unbalance in either of the stages is magnified. For grounded emitting electrodes, the differential and common-mode transfer resistances differ only by the factor of two that occurs in the definition of differential input current:

$$R_{T(\mathrm{dif})} = 2R_T, \tag{15.28}$$

$$R_{T(\mathrm{com})} = R_T, \tag{15.29}$$

where R_T is the transfer resistance of a single-ended stage. High-frequency peaking capacitors may be added to a balanced shunt-feedback stage.

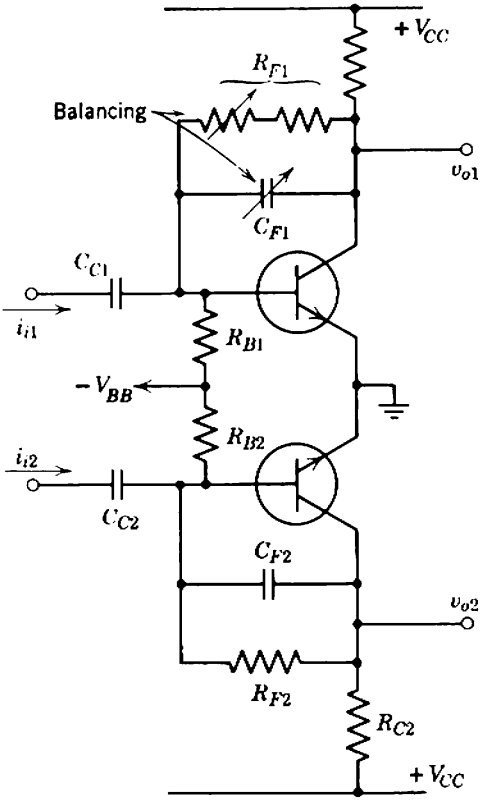


Fig. 15.13 Shunt-feedback differential amplifier stage.

For an alternate cascade that uses balanced stages throughout and has a balanced load, the total common-mode rejection factor (corresponding to Eq. 15.17) is

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{G_{T(\text{com})1}}{G_{T(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \left[\frac{R_{T(\text{com})2}}{R_{T(\text{dif})2}}\right]\left\{\frac{1}{\lambda_3} + \left[\frac{G_{T(\text{com})3}}{G_{T(\text{dif})3}}\right]\left\{\cdots\right\}\right\}\right\},$$

where $\lambda_1, \lambda_3, \ldots$ are the common-mode rejection factors for the series-feedback stages:

$$\lambda_{(2m+1)} = \frac{\text{common-mode input voltage for a certain differential-mode output current}}{\text{differential-mode input voltage for the same differential-mode output current}},$$

and $\lambda_2, \lambda_4, \ldots$ are the common-mode rejection factors for the shunt-feedback stages:

$$\lambda_{2m} = \frac{\text{common-mode input current for a certain differential-mode output voltage}}{\text{differential-mode input current for the same differential-mode output voltage}}.$$

Using Eqs. 15.28 and 15.29,

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{G_{T(\text{com})1}}{G_{T(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \frac{1}{2}\left\{\frac{1}{\lambda_3} + \left[\frac{G_{T(\text{com})3}}{G_{T(\text{dif})3}}\right]\left\{\cdots\right\}\right\}\right\}. \quad (15.30)$$

Notice that the contributions for λ_2 and λ_3 to the total λ differ only by a factor of two, the ratio of the differential-mode to the common-mode transfer resistance of the second stage. Balance of the third (series) stage is very nearly as critical as balance of the second (shunt) stage. This contrasts with a nonfeedback amplifier, in which the differential-mode and common-mode gains of the second stage differ by several orders of magnitude, and balance of the third stage is much less critical.

There are two equations for the common-mode rejection when the amplifier reverts to single-ended stages after the kth stage. If the kth stage is a series stage

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{G_{T(\text{com})1}}{G_{T(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \frac{1}{2}\left\{\cdots + \frac{1}{2}\left\{\frac{1}{\lambda_k} + \frac{G_{T(\text{com})k}}{G_{T(\text{dif})k}}\right\}\cdots\right\}\right\}. \quad (15.31a)$$

If the kth stage is a shunt stage

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{G_{T(\text{com})1}}{G_{T(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \frac{1}{2}\left\{\cdots + \frac{1}{2}\left\{\frac{1}{\lambda_{k-1}} + \left[\frac{G_{T(\text{com})(k-1)}}{G_{T(\text{dif})(k-1)}}\right]\left\{\frac{1}{\lambda_k} + \frac{1}{2}\right\}\right\}\cdots\right\}\right\}.$$

From the definitions of differential-mode and common-mode transfer resistance (Eqs. 15.28 and 15.29) it is obvious that the common-mode

rejection factor of a balanced shunt-feedback stage with a single-ended load is

$$\lambda_{\text{shunt}} = \tfrac{1}{2}.$$

Therefore

$$\frac{1}{\lambda} = \frac{1}{\lambda_1} + \left[\frac{G_{T(\text{com})1}}{G_{T(\text{dif})1}}\right]\left\{\frac{1}{\lambda_2} + \frac{1}{2}\left\{\cdots + \frac{1}{2}\left\{\frac{1}{\lambda_{k-1}} + \left[\frac{G_{T(\text{com})(k-1)}}{G_{T(\text{dif})(k-1)}}\right]\right\}\cdots\right\}\right\}. \tag{15.31b}$$

Comparison of Eqs. 15.31*a* and 15.31*b* show that the *k*th balanced shunt-feedback stage accomplishes nothing. There is no point at all in using a balanced shunt stage to drive a single-ended load. The last balanced stage in a feedback differential amplifier should always be a series-feedback stage.

Some general principles for the design of feedback differential amplifiers can be formulated as follows:

1. The input (series) stage requires the usual, careful balancing. An emitting-electrode balancing potentiometer should always be provided, and a current source is desirable. High-frequency balancing capacitors may be required also.
2. The second (shunt) stage has its emitting electrodes grounded for signals. A balancing control is usually required, and takes the form of an adjustable feedback resistor for one side. If high-frequency peaking is provided, one peaking capacitor should be made adjustable.
3. The third (series) stage is more critical than the third stage in non-feedback amplifiers. A long-tailed pair is almost always required, but balancing adjustments can usually be omitted.
4. Further balanced stages are necessary only when a very high common-mode rejection is required. Additional balanced stages must be added in pairs, so that the last balanced stage has series feedback.
5. If a small common-mode rejection is sufficient, it may be permissible to use single-ended stages after the first (series) stage.
6. Close balancing of the supply resistors is not required at mid-band frequencies, because these resistors do not occur in the first-order expressions for the gain of an alternate cascade. However, the supply resistors enter into the coupling time constants, so balancing may be required at low frequencies.

15.2.1.1 Illustrative Example

Figure 15.14 shows a transistor feedback differential amplifier which illustrates most of the above points. Its mid-band common-mode

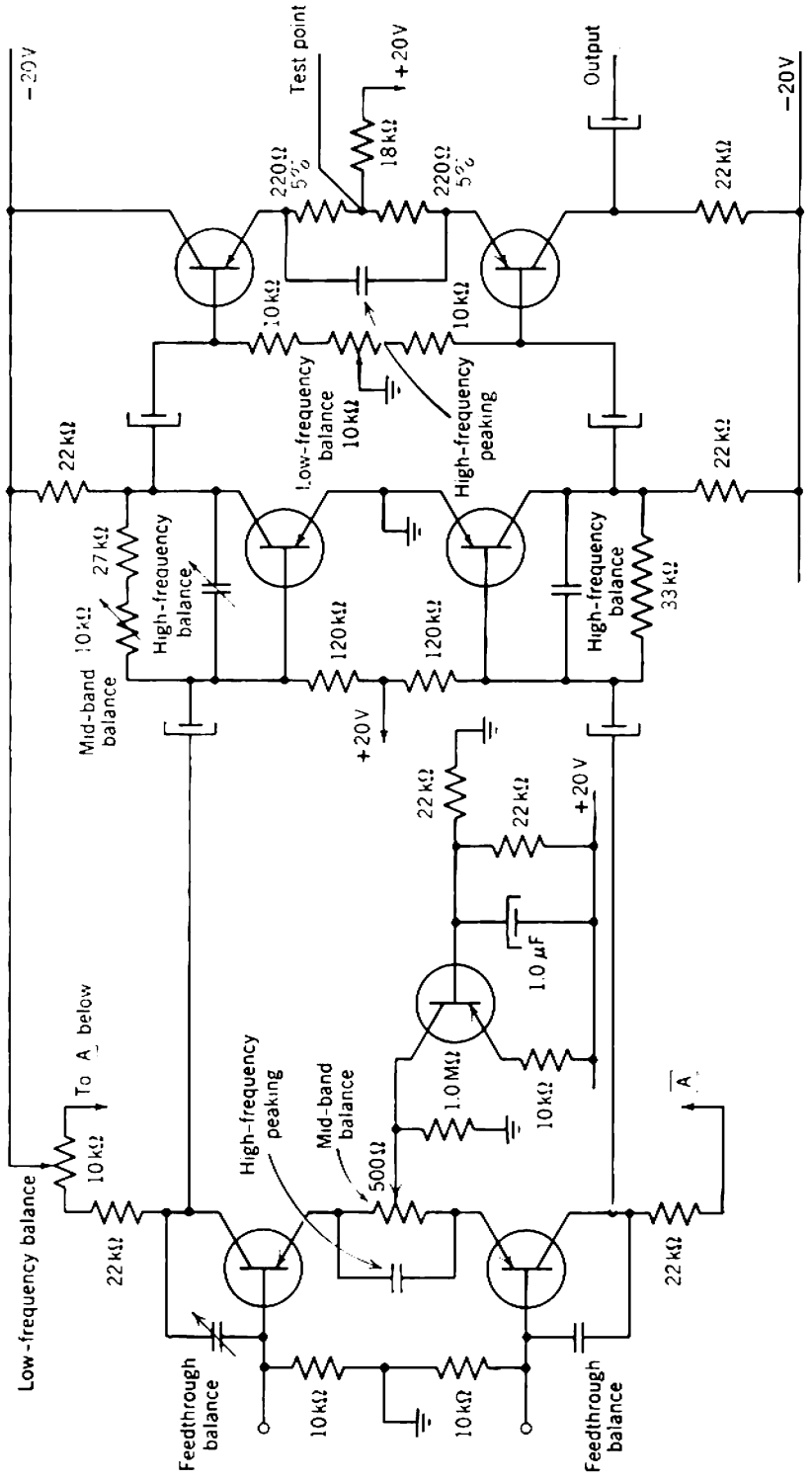


Fig. 15.14 Circuit diagram for a transistor feedback differential amplifier.

rejection is required to be 10^5:1. The differential and common-mode transfer functions of the stages are:

$$\begin{aligned} G_{T1(\text{dif})} &= 1.6 \text{ mA/V}, & G_{T1(\text{com})} &= 0.5\ \mu\text{A/V}, \\ R_{T2(\text{dif})} &= 60 \text{ V/mA}, & R_{T2(\text{com})} &= 30 \text{ V/mA}, \\ G_{T3(\text{dif})} &= 1.6 \text{ mA/V}, & G_{T3(\text{com})} &= 30\ \mu\text{A/V}. \end{aligned}$$

If the specification of 10^5:1 is to be met, the performance of the individual stages must be

$$\begin{aligned} \lambda_1 &\geq 10^5, \\ \lambda_2 &\geq 30, \\ \lambda_3 &\geq 15. \end{aligned}$$

In addition, Eq. 15.31*a* shows it is necessary that

$$\left[\frac{G_{T(\text{dif})1}}{G_{T(\text{com})1}}\right] \times 2 \times \left[\frac{G_{T(\text{dif})3}}{G_{T(\text{com})3}}\right] \geq 10^5.$$

This condition is satisfied by a factor 3.5.

At mid-band frequencies, the first stage will probably meet specifications using unselected transistors. The second stage requires adjustment to accommodate the spread in β_N of unselected transistors, but the third stage requires no adjustment at all. Low-frequency balancing potentiometers are provided in the collector circuit of the first stage and the base circuit of the third stage. Because the electrolytic coupling capacitors between stages have a large production tolerance, a large range of adjustment must be provided — 1.5:1 in the first stage and 2:1 in the third stage. The larger value is required in the third stage because β_N occurs in the expression for coupling time constant. High-frequency peaking is added to all stages, and is adjustable in the second stage. Feedthrough balancing capacitors are added to the first stage. The balancing procedure is:

1. Apply a mid-band differential signal to the input, and adjust the mid-band balance of the second stage for minimum signal at the test point. (If a balanced signal generator is not available, it is sufficient to ground one input terminal and apply a signal to the other.)
2. Join the bases of the second stage, apply a low-frequency differential signal to the input, and adjust the first-stage low-frequency balance for minimum output signal. Remove the connection between the transistor bases.
3. Apply a low-frequency differential signal to the input, and adjust the third-stage low-frequency balance for minimum signal at the test point.
4. Repeat steps 1–3 until all balance conditions are satisfied simultaneously.

5. Apply a high-frequency differential signal to the input, and adjust the second-stage high-frequency balance for minimum signal at the test point.

6. Apply a mid-band common-mode signal to the input and adjust the first-stage mid-band balance for minimum output signal.

7. Apply a high-frequency common-mode signal to the input, and adjust the first-stage feedthrough balance for minimum output signal.

8. Repeat the entire balancing procedure until all balance conditions are satisfied simultaneously.

In a production model of the amplifier, the first-stage mid-band balance is a front-panel screwdriver adjustment; all other adjustments are preset and substantially never touched.

15.2.2 Over-all Feedback

Over-all feedback finds little application in differential amplifiers, largely because the common-mode rejection of differential amplifiers is more important than absolute precision in differential-mode gain. To a first order, over-all feedback has no effect on the common-mode rejection. In realizing over-all feedback differential amplifiers it is essential to use balanced stages throughout, even if the ultimate load is single-ended.

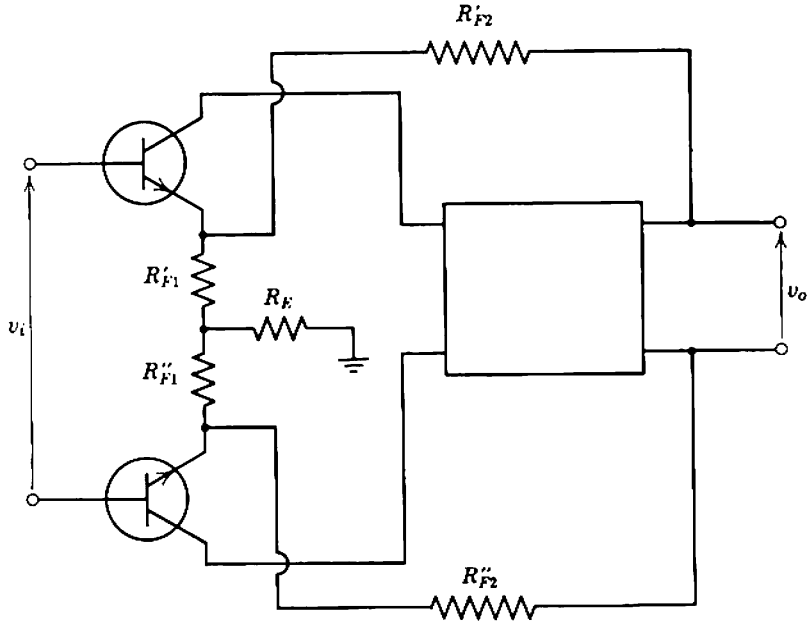


Fig. 15.15 Differential amplifier with over-all voltage feedback.

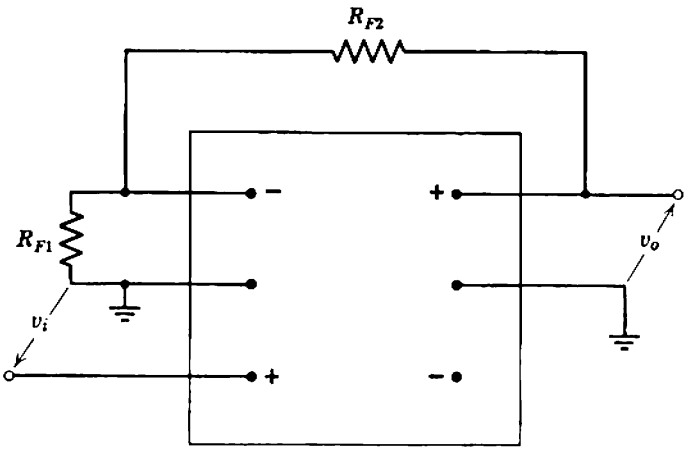


Fig. 15.16 Differential amplifier used as a single-ended feedback amplifier.

Otherwise, the feedback voltage has a large common-mode component which appears in the input circuit mesh.

As an example, Fig. 15.15 shows the block diagram for an amplifier with stable voltage gain. Ideally, the differential voltage gain into a balanced load is

$$A_{V(\text{dif})} = \frac{R_{F1} + R_{F2}}{R_{F1}}, \tag{15.32}$$

whereas the gain into a single-ended load is half this value. Notice that the resistors R'_{F2} and R''_{F2} form part of the "tail" resistance R_E for the first stage.

Differential amplifiers often find application in the forward path of single-ended over-all feedback amplifiers. Figure 15.16 shows the block diagram for a single-ended voltage amplifier having gain very close to $(R_{F1} + R_{F2})/R_{F1}$. Notice that the voltage gain has a positive sign rather than the negative sign of a purely single-ended amplifier. Examples of this technique are given in Sections 16.6.2 and 16.7.2 of the next chapter.

15.3 SIMPLE DC AMPLIFIERS

The discussion of amplifiers up to this point of the book (including that of differential amplifiers) assumes the presence of artificial limits to the low-frequency response. Such limits are associated with the bypass and decoupling capacitors in the biasing system and, with the few exceptions of instances when stages are coupled together directly, limits are also associated with coupling networks. These conventional amplifiers are

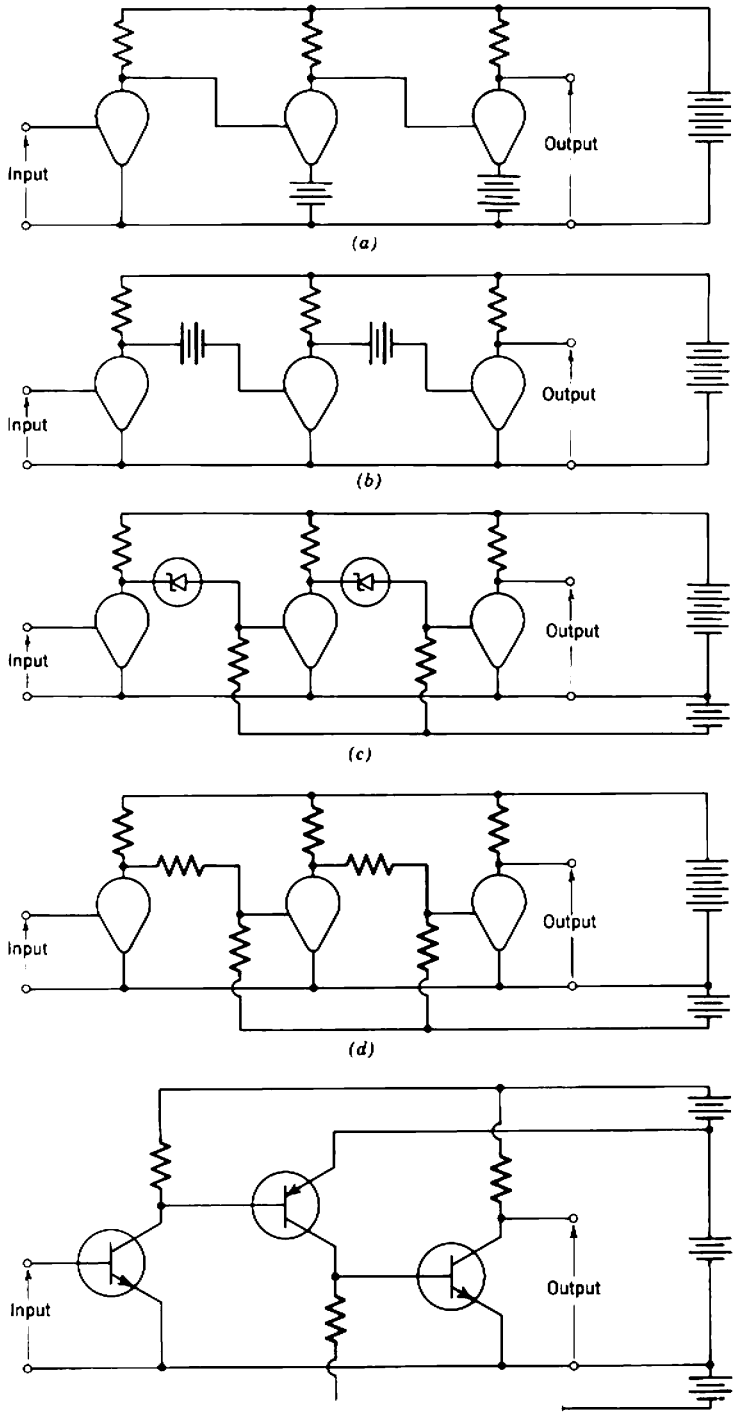


Fig. 15.17 Methods for direct coupling stages: (*a*) several power supplies, one terminal common; (*b*) several power supplies, no terminals common; (*c*) constant voltage devices; (*d*) resistive voltage dividers; (*e*) devices of alternate carrier types.

quite unsuitable for amplifying slowly varying (near zero frequency) signals, because prohibitively large capacitances and inductances are required. They are *ac amplifiers*, in the sense that they amplify alternating (time-varying) signals. There is a need for a new class of amplifier, the *dc amplifier*, in which the low-frequency response has no limit.

The elimination of energy-storage devices such as capacitors and inductors from dc amplifiers makes their design more difficult than that of ac amplifiers. Design difficulties are of two types:

1. The absence of coupling capacitors places constraints on the design of biasing circuits. An introduction to the problems is given in Sections 6.5 and 6.6.1, where it is shown that groups of two or more direct-coupled stages can be biased adequately if dc feedback is applied between a later stage and the first stage. However, all of the direct-coupled circuits considered in these earlier sections employ at least one bypass capacitor to separate bias and signal feedback and so avoid signal degeneration. In the context of purely dc amplifiers, the preceding examples illustrate techniques by which direct *coupling* can be combined with satisfactory biasing of individual stages; *biasing* of dc amplifier stages is yet to be considered. The coupling techniques of earlier chapters are summarized in the elemental circuits of Fig. 15.17 and involve the use of:

(a) a multiplicity of power supplies, either with one common terminal (Fig. 15.17*a*) or, less desirably, floating power supplies (Fig. 15.17*b*);
(b) voltage reference diodes of semiconductor or gas-filled types (Fig. 15.17*c*) to give a more satisfactory realization of Fig. 15.17*b*;
(c) resistive voltage dividers between stages (Fig. 15.17*d*) which have the disadvantage of attenuating the signal;
(d) semiconductor devices of alternating *p*-carrier or *n*-carrier types, for example, *p-n-p* and *n-p-n* transistors or an *n*-channel f.e.t. with a *p-n-p* transistor (Fig. 15.17*e*).

2. There is no way of distinguishing between a signal and a slow change or *drift* in the quiescent point of an amplifying device. Drift occurs because of changes with age, temperature, and so on, in the devices, passive elements, or supply voltages. In contrast, changes in the quiescent conditions in ac amplifiers cannot propagate to the output and have only a second-order effect on small-signal gain. This chapter considers biasing methods and other circuit details for minimizing drift.

15.3.1 Drift

Drift is an undesired disturbance in the quiescent point of an amplifying device. Drift in the early stages of a dc amplifier is a special problem,

because it propagates through the later stages and is amplified along with the signal. A dc amplifier that meets all other specifications may be quite useless if its ratio of signal to drift at the output falls below a certain value. Noise and drift have much in common, and it is convenient to adopt the procedure used to represent noise and employ *drift generators at the input* to characterize the drift performance of an amplifier.

In Fig. 15.18 the voltage drift generator v_D accounts for all drift at the output when the input terminals are short circuited. Its value can be found by measuring the output drift over a specified time interval and dividing by the transfer function defined in terms of an input voltage:

$$v_D = \frac{v_{Do(sc)}}{A_V} = \frac{i_{Do(sc)}}{G_T}, \tag{15.33}$$

where $v_{Do(sc)}$ and $i_{Do(sc)}$ are the drifts in output voltage and current when the input is short-circuited. Similarly, the current drift generator i_D accounts for all drift when the input terminals are open-circuited:

$$i_D = \frac{v_{Do(oc)}}{R_T} = \frac{i_{Do(oc)}}{A_I}. \tag{15.34}$$

Physically, v_D has the significance of the voltage increment that must be applied to the control electrode to restore the collecting-electrode current to its original value, whereas i_D is the current that must be forced into the control electrode.

The output drift for any finite source resistance R_S is a combination of the effects of v_D and i_D. In many instances, for example when the dominant source of drift is a temperature change or a supply voltage change, v_D and i_D are correlated. The drift output is therefore found by simple addition of voltage or current, not addition of power as required in noise summations:

$$v_{Do} = \left[\frac{(v_D + i_D R_S)R_i}{R_S + R_i}\right] A_V = \left(\frac{v_D + i_D R_S}{R_S + R_i}\right) R_T, \tag{15.35}$$

$$i_{Do} = \left[\frac{(v_D + i_D R_S)R_i}{R_S + R_i}\right] G_T = \left(\frac{v_D + i_D R_S}{R_S + R_i}\right) A_I. \tag{15.36}$$

Occasionally, the v_D and i_D generators are completely independent and an rms type of summation is valid but, in view of the usual imprecise knowledge of their values, it is desirable to base calculations on the pessimistic assumption of complete correlation.

The output signal-to-drift ratio for either Fig. 15.18*a* or *b* is

$$\frac{v_{So}}{v_{Do}} = \frac{i_{So}}{i_{Do}} = \frac{v_S}{v_D + i_D R_S} = \frac{i_S}{i_D + v_D/R_S}. \tag{15.37}$$

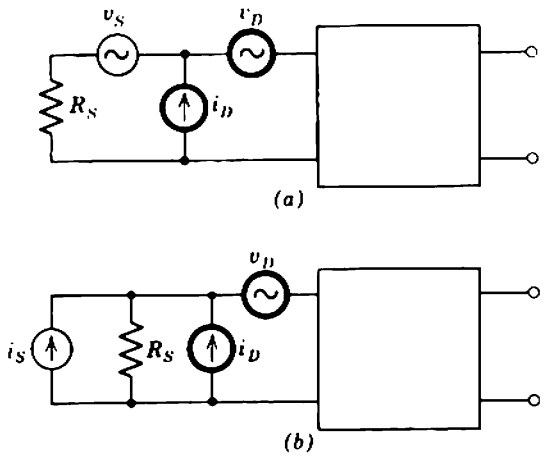


Fig. 15.18 Signal-to-drift ratio (*a*) voltage signal source; (*b*) current signal source.

Often, only one of the drift generators is significant; drift due to v_D dominates if R_S is small, whereas drift due to i_D dominates if R_S is large.

All stages of an amplifier contribute to the over-all drift, but stages other than the first (and possibly the second) have little influence on the over-all signal-to-drift ratio. The proof is similar to that for noise in multistage amplifiers (Section 8.4.2).

15.3.1.1 Types of Drift

Drift can arise from a number of sources, including supply voltage changes, temperature changes, and device aging. The basic irreducible source of drift, however, is associated with flicker noise. This can be seen from the ideal example of an EF86/6BK8 pentode operated so that screen voltage and grid voltage are constant. For the operating conditions in Section 8.4.4.1 the input noise voltage generator contains shot noise and flicker terms and is given by

$$d(v_N^2) = \left(1.05 \times 10^{-16} + \frac{4.0 \times 10^{-13}}{f}\right) df.$$

In a dc amplifier, drift can be equated with low-frequency noise so that the mean-square drift voltage referred to the input is

$$v_D^2 = \int_{f_1}^{f_2} d(v_N^2) = 1.05 \times 10^{-16}(f_2 - f_1) + 4.0 \times 10^{-13} \ln\left(\frac{f_2}{f_1}\right), \tag{15.38}$$

where f_2 and f_1 are the upper and lower limits of a "brick-wall" response.

Somewhat crudely, the lower frequency f_1 can be associated with the period T over which a measurement is made:

$$f_1 = \frac{1}{2\pi T}.$$

The upper frequency f_2 is assumed to be 10 kHz; this numerical choice has little effect on the results. Experimental evidence* indicates that the $1/f$ character of flicker noise persists over observation periods of several months and thus Eq. 15.38 is likely to be obeyed for normal observation periods; drift voltage is plotted against time in Fig. 15.19. If, for example, the drift due to flicker noise is to be kept below 3 μV, the operating point requires resetting about once an hour. This idealized example shows that the period of observation can be a significant parameter in specifying drift performance. It is customary to provide at least an approximate indication of observation time by referring specifically to either *long-term drift* (measured over several weeks or months) or *short-term drift* (measured over several minutes or hours).

In practical situations the long-term drift is greater than can be attributed to flicker noise alone and is usually dominated by long-term changes in device characteristics, supply voltages, and resistors. Long-term drift in amplifying devices is a consequence of the gradual deterioration of all materials: the characteristics of vacuum tubes change as the cathode material deteriorates, whereas the drift of semiconductor devices is usually due to slow changes in surface conditions. The influence of supply

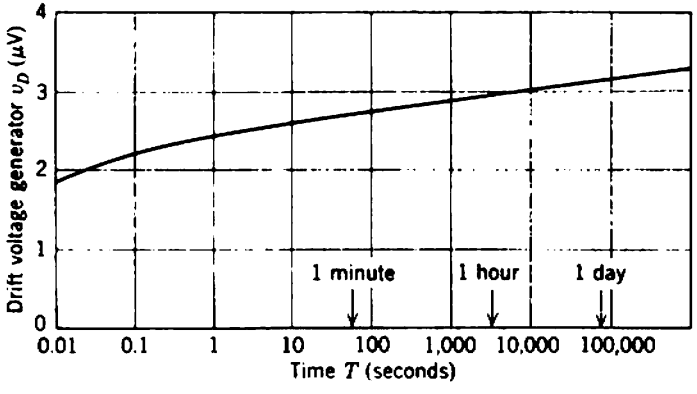


Fig. 15.19 Voltage drift expected at the input of type EF86/6BK8 pentode from flicker noise considerations.

* C. A. HOGARTH (ed.), *Noise in Electronic Devices*: Chapter 2, "The physical basis of noise," by F. J. HYDE (published on behalf of the Physical Society by Chapman and Hall, London, 1961).

voltage and resistor changes can be illustrated by further reference to the pentode in Sections 7.8.1 and 8.4.4.1.

First, if the supply voltage V_{AA} changes by a small increment ΔV_{AA}, the output (anode) voltage changes by an amount

$$\Delta V_A = \Delta V_{AA} \frac{r_A}{r_A + R_L}.$$

Dividing ΔV_A by the voltage gain shows that supply voltage variation contributes a drift generator

$$v_D = -\frac{\Delta V_{AA}}{g_m R_L} \tag{15.39}$$

to the input. Because $g_m R_L$ is about 100, the voltage drift due to supply voltage variation exceeds that due to flicker noise (taken as a maximum of 3.5 μV) whenever ΔV_{AA} exceeds 350 μV. This order of change in a 260-V power supply is only 1.35 parts in 10^6. With extremely careful design and the use of a standard cell as a voltage reference, long-term power-supply stabilities of about 1 part in 10^4 can be achieved in practice but 1 part in 10^6 is virtually unobtainable. Thus drift due to power-supply variations exceeds that expected from flicker noise considerations by about two orders of magnitude.

Second, if a change ΔR_L occurs in the load resistor, it contributes a drift generator

$$v_D = \frac{\Delta R_L}{g_m R_L} I_A \tag{15.40}$$

to the input. For the example considered this source of drift exceeds that due to noise if the change $\Delta R_L/R_L$ exceeds 2.6 parts in 10^6. High-stability resistors seldom have long-term drift figures better than 0.01% so this source of drift exceeds that due to flicker noise in all practical circuits.

Usually, short-term drift also exceeds that expected from flicker noise. Changes of device temperature are the dominant additional source—cathode temperature in vacuum tubes and crystal temperature in semiconductor devices. Changes in dc supply voltages are not often a problem, nor is the temperature coefficient of resistors which can be less than 10 ppm/°C.

15.3.2 Short-Term Drift Generators at the Input

Graphical information on parameter variations with temperature is provided by some device manufacturers. Alternatively, values of the short-term drift generators can be estimated from the first-order theory of

active devices given in Chapters 2 to 4. When available, manufacturers' data should obviously be used in preference to the idealized expressions derived in the following sections.

15.3.2.1 Vacuum Tubes

Second-order vacuum-tube theory takes into account the finite emission velocity of electrons by postulating the existence of a zero-field barrier of potential $(-V_m)$ with respect to the cathode, situated a distance x_m from the cathode (Section 3.1.1.3). The magnitudes of both V_m and x_m increase with cathode temperature due to the increased mean emission velocities of the electrons; a consequent change in the anode voltage of the equivalent diode V_e' causes the cathode current to vary. The influence of V_m and x_m on I_K follows from Eqs. 3.33 and 3.34 as

$$I_K = G(V_e')^{3/2} = G(V_G + V_A/\mu + V_m)^{3/2}$$

and

$$G = \frac{\text{constant}}{(W_G - x_m)^2}.$$

Usually x_m is considerably smaller than the cathode-grid spacing W_G; hence changes in x_m have little effect on perveance G. Consequently, the effect of cathode temperature changes can be estimated by considering only the change in V_m. For a small change in cathode temperature (due to a small change ΔV_H in heater voltage)

$$\Delta V_m \propto \Delta T_K = (\text{constant}) \frac{\Delta V_H}{V_H}. \tag{15.41}$$

Consequently, the anode current change is

$$\Delta I_A = g_m \Delta V_m,$$

and the drift generator at the input is

$$v_D = \frac{\Delta I_A}{g_m} = (\text{constant}) \frac{\Delta V_H}{V_H}. \tag{15.42}$$

This drift voltage generator is usually between 10 and 20 mV per 1% change in heater voltage. Because it is extremely difficult to keep the heater voltage constant better than 0.5%, this is a very substantial source of drift and an extreme problem in vacuum-tube dc amplifiers.

Changes in heater voltage also alter the grid current and give rise to current drift i_D. However, this contribution to signal-to-drift ratio is small compared with that of v_D unless the source resistance is abnormally

high. In normal operation of a tube, only the positive ion current is significant and the approximation of Eq. 8.94

$$I_G \approx \nu V_A^{3/2} I_A$$

may be assumed. Differentiation then yields

$$i_D = \Delta I_G \approx I_G g_m v_D \left(\frac{1}{I_A} - \frac{3}{2}\frac{R_L}{V_A}\right).$$

Because of the difference in sign of the two terms in the bracket, it appears possible for the grid current drift to be positive, negative, or zero. The result is a little dubious but it is reasonable to assume that, as an order of magnitude,

$$|i_D| \approx g_m v_D \frac{I_G}{I_A}. \tag{15.43}$$

15.3.2.2 Transistors

First-order theory for transistors (Section 4.6.3) indicates that $V_{B'E}$ varies linearly with temperature, I_{CO} varies exponentially, and β_N varies roughly as the square of temperature. These are the important sources of short-term drift.

The base-emitter voltage drop of a transistor is

$$V_{BE} = V_{B'E} + I_B r_B = V_{B'E} + \left(\frac{I_E}{\beta_N} - I_{CO}\right) r_B. \tag{15.44}$$

The voltage drift generator v_D has the significance of the voltage increment that must be applied to the base to maintain the collector current constant when the temperature is changed; that is,

$$v_D = \left(\frac{\partial V_{BE}}{\partial T}\right)_{I_C} \Delta T$$

and, substituting from Eq. 15.44,

$$v_D = \left(\frac{\partial V_{B'E}}{\partial T} - \frac{I_E r_B}{\beta_N^2}\frac{\partial \beta_N}{\partial T} + I_B \frac{\partial r_B}{\partial T} - r_B \frac{\partial I_{CO}}{\partial T}\right)_{I_C} \Delta T.$$

Using the results of Table 4.6,

$$v_D \approx \left[\frac{\partial V_{B'E}}{\partial T} - r_B \frac{\partial I_{CO}}{\partial T}\right]_{I_C} \Delta T = \left[\left(\frac{\partial V_{B'E}}{\partial T}\right)_{I_C} - r_B \frac{I_{CO}}{a}\right] \Delta T, \tag{15.45}$$

where there is almost perfect cancellation between the temperature-dependent components due to β_N and r_B [both have magnitudes close to $2r_B(I_B/T)$]. $\partial V_{B'E}/\partial T$ has substantially the same value for all transistors

(about -2.5 mV/°C) and is usually the dominant term. The constant a is in the vicinity of 12°C.

The base current of a transistor is

$$I_B = \frac{I_E}{\beta_N} - I_{CO}.$$

The drift current generator i_D has the significance of the current that must be forced into the base to maintain the collector current constant when temperature is changed. Hence

$$i_D = \left(\frac{\partial I_B}{\partial T}\right)_{I_C} \Delta T = -\left(\frac{\partial I_{CO}}{\partial T} + \frac{I_E}{\beta_N{}^2}\frac{\partial \beta_N}{\partial T}\right)_{I_C} \Delta T = -\left(\frac{I_{CO}}{a} + b\frac{I_E}{\beta_N T}\right)\Delta T, \tag{15.46}$$

where, from Table 4.6, the constant b is in the vicinity of 2.

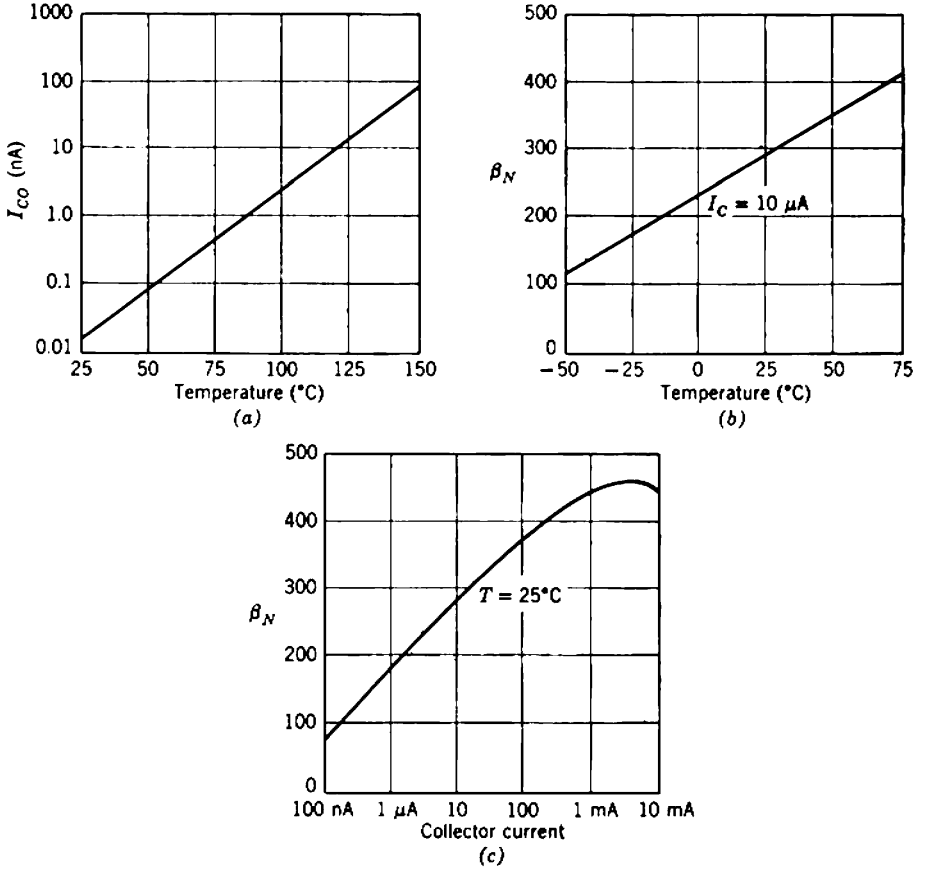


Fig. 15.20 Typical characteristics of 2N2484 transistor.

Since all transistors have comparable magnitudes for $\partial V_{BE}/\partial T$, transistors of good drift performance have small drift current generators; that is, they require low I_{CO} and high β_N at low I_E. Much effort has been expended in the development of transistors with these characteristics and planar silicon types are very satisfactory. The characteristics of a typical transistor (type 2N2484) are summarized by the curves of Fig. 15.20: the constants a and b have magnitudes of 14.5°C and 2.5, respectively. For typical operating conditions

$$I_C = 10\ \mu\text{A}, \qquad T = 25°\text{C},$$

the current drift from Eq. 15.45 is

$$\frac{i_D}{\Delta T} = -0.35\ \text{nA/°C}.$$

The contributions of the drift current generator equals that of the drift voltage generator when $R_S = 7\ \text{M}\Omega$. The current drift can be reduced by lowering the collector current, but this often lowers the stage gain so much that the drift of the second stage becomes important.

When the source resistance is low, it is quite unnecessary to use silicon transistors or to operate at extremely low currents. Small germanium transistors having I_{CO} of the order of 1 μA achieve current drifts of 200 nA/°C when operated at 1 mA emitter current. The resulting component of output drift does not exceed that due to v_D until R_S is larger than 10 kΩ.

15.3.2.3 Field-Effect Transistors

The use of field-effect transistors in the first stage of dc amplifiers has some interesting possibilities. As for vacuum tubes, the input-electrode current is so small that it scarcely contributes to drift except for abnormally high values of source resistance. It is very difficult to envisage circumstances under which slight changes in a gate current of 10^{-16} A (for an insulated-gate device) would produce significant drift. Thus input voltage drift is the significant quantity for field-effect transistors.

If an f.e.t. is operated with constant electrode voltages, the drain current can either decrease or increase with temperature and it is theoretically possible to adjust the bias conditions to give zero drift (Section 4.8.6). This contrasts with tubes and bipolar transistors, for which the various components of temperature-induced drift invariably add. A general expression for temperature-induced drift of a pinched-off f.e.t. can be developed by rewriting Eq. 4.141 to include a contact potential term:

$$I_D = I_{DO}\left[1 + \frac{(V_G - V_I)}{V_P}\right]^2.$$

Partial differentiation with respect to the three temperature-dependent parameters I_{DO}, V_i, and V_P leads to

$$v_D = \left[\frac{I_D}{g_m I_{DO}}\left(\frac{\partial I_{DO}}{\partial T}\right) - \frac{\partial V_i}{\partial T} - \frac{(V_G - V_i)}{V_P}\left(\frac{\partial V_P}{\partial T}\right)\right]\Delta T. \quad (15.47)$$

For a *p-n* junction f.e.t., V_P is relatively independent of temperature and only the first two terms of the right-hand side of Eq. 15.47 are important. Equation 4.136 shows that the only term in I_{DO} that has a significant temperature coefficient is mobility μ and, from Eq. 4.23,

$$\mu \propto T^{-b}$$

where the constant b is about 2. Therefore

$$\frac{1}{I_{DO}}\left(\frac{\partial I_{DO}}{\partial T}\right) = -\frac{b}{T} \approx -7 \times 10^{-3} \text{ per °C}$$

at room temperature. The temperature coefficient of the contact potential is

$$\frac{\partial V_i}{\partial T} \approx -2 \text{ mV/°C}.$$

Thus the two components of drift have opposite signs, and cancellation occurs when

$$\frac{I_D}{g_m} = \frac{T}{b}\left(\frac{\partial V_i}{\partial T}\right) \approx 0.3 \text{ V}.$$

Replacing I_D and g_m from Eqs. 4.141 and 4.142, the gate voltage at which v_D is zero follows as

$$V_G = -\left[V_P - 2\frac{T}{b}\left(\frac{\partial V_i}{\partial T}\right)\right] \approx -(V_P - 0.6). \quad (15.48)$$

For typical field-effect transistors, V_P is of the order of several volts. Therefore temperature compensation occurs only when the device is biased near cutoff where g_m and the stage gain are very small, and drift in the second stage dominates performance. In practice, it is impossible to achieve exact cancellation of temperature-dependent drifts, but values less than 50 μV/°C have been claimed.*

For an enhancement-mode i.g.f.e.t., the contact potential term disappears from Eq. 15.47 but the temperature-dependence of V_P becomes important:

$$v_D = \left[\frac{I_D}{g_m I_{DO}}\left(\frac{\partial I_{DO}}{\partial T}\right) - \frac{V_G}{V_P}\left(\frac{\partial V_P}{\partial T}\right)\right]\Delta T. \quad (15.49)$$

* J. C. S. RICHARDS, "Field effect transistors in dc amplifiers," *Electronic Eng.*, **37**, 302, May 1965.

Because increasing the temperature results in a higher concentration of minority carriers throughout the crystal, the inversion layer is produced at a lower value of V_P. Thus, $\partial V_P/\partial T$ and $\partial I_{DO}/\partial T$ are both negative and it is possible to adjust the operating point so that the net temperature drift is zero. At high currents the first (negative temperature coefficient) term on the right-hand side of Eq. 15.49 dominates, whereas at low currents the second (positive temperature coefficient) term dominates. Notice that the drain current at zero drift is sufficiently high to ensure useful stage gain.

For a depletion mode i.g.f.e.t. the variation of I_{DO} is dominant and inherent temperature compensation appears to be unlikely.

Although the possibility of adjustment for zero drift in field-effect devices is interesting, it may be impractical in designable circuits which require a minimum of zero-setting adjustments. A more convenient approach is to use field-effect transistors as input devices only, with bipolar transistors for the following stages. An f.e.t.—bipolar combination can be designed so that the drifts compensate by operating the f.e.t. near zero gate voltage (where mobility is the dominant temperature coefficient) and deriving a voltage from the drain current drift which just compensates for the V_{BE} drift in the following transistor.* Figure 15.21 shows an

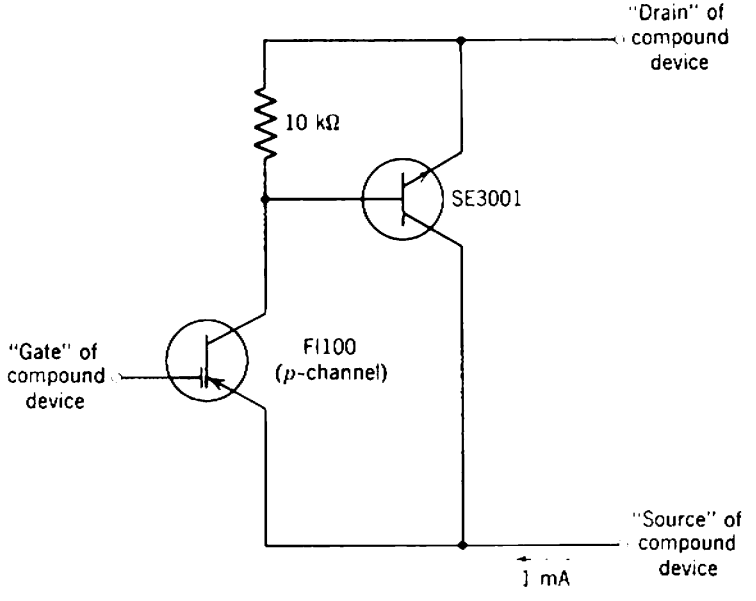


Fig. 15.21 Compounding of field-effect and bipolar transistors to give low voltage drift.

* W. Gosling, "A drift-compensated f.e.t.–bipolar hybrid amplifier," *Proc. Inst. Elec. Electronics Engrs.*, **53**, 323, March 1965.

example. The technique results in designable circuits, because two separate and reproducible temperature variations are used. In addition, the input f.e.t. operates at high g_m.

15.3.3 Concluding Comments on Simple Dc Amplifiers

Vacuum tubes, silicon planar bipolar transistors, and field-effect transistors all have their drifts dominated by the voltage component v_D provided the source resistance is less than a few megohms. Excepting the special case of very high source impedance, the relative capabilities of the various devices as low-drift amplifiers can be assessed by comparing the orders of input voltage drift that are achievable in practice. For vacuum tubes, heater voltage variation is the major source of drift. Few power authorities maintain voltage to narrower tolerances than $\pm 10\%$ and, if simple heater supplies are used, the expected drift is about ± 200 mV. It is possible to reduce this by about a factor of 20 by stabilizing the heater supply. For semiconductor devices the absolute magnitude of short-term drift is controlled by ambient temperature. In air-conditioned laboratories it is unlikely that the drift of bipolar transistors will exceed ± 10 mV. Because of their small size, it is quite practicable to enclose transistors in a controlled oven which keeps temperature constant to within a fraction of a degree.* This implies that drifts of less than 1 mV can be achieved by introducing a constant-temperature system of comparable complexity to that required for regulating vacuum-tube heater supplies. It seems fair to argue that vacuum tubes have no advantage insofar as voltage drift is concerned and it is not surprising that transistor dc amplifiers are used extensively. The examples considered in later sections of this chapter are limited to transistor amplifiers.

The situation is less clear when R_S exceeds about ten megohms. Bipolar transistors are certainly inferior to both tubes and insulated-gate field-effect transistors, but the choice between a tube and an i.g.f.e.t. is complicated by the relative lack of knowledge of long-term drift effects in the latter device. Some i.g.f.e.t.s are extremely susceptible to damage; their drift performance may change radically if they are subjected even to overload signals.

---

* H. KEMHADJIAN, "A simple temperature control system for transistors," *Mullard Technical Communications*, **4**, 186, December 1958. An example of a transistor with a temperature-controlled oven incorporated into the same silicon chip is Fairchild type μA726.

15.4 BALANCED DC AMPLIFIERS

Drift can be reduced to some extent by selecting devices, by choosing a near-optimum operating point, by using highly stable supplies, and by using high-stability components. Nevertheless, a certain finite, irreducible drift always remains. A powerful technique for reducing the effect of this residual drift is to use balanced differential stages in the amplifier. Ideally, the drift generators for the two sides of each stage are equal; the drift therefore appears only as a common-mode signal and is rejected. In practice, there is some unbalance so a finite (but relatively small) differential-mode component of drift remains; with reasonable care in matching devices it is possible to obtain a one- or two-order-of-magnitude reduction in drift compared with an unbalanced amplifier. In short, although the drift generators of a single-ended amplifier cannot be eliminated, they can be compensated by having almost equal generators on the other side of a long-tailed pair.

After passing through several long-tailed pairs, the amplified drift of the first stage becomes comparable with that of a single-ended stage. Consequently, from the viewpoint of drift performance, it is permissible to revert to single-ended operation for the later stages of an amplifier.

Although the balanced arrangement is employed for drift reduction, this type of amplifier is not restricted to balanced input signals. A long-tailed pair which is adjusted to minimize drift and maximize common-mode rejection produces a differential output of magnitude

$$v_{o1} - v_{o2} = A_{V(\text{dif})} v_{i1}$$

in response to a single-ended input voltage v_{i1} when no signal is applied to the other input. The majority of dc amplifiers are used with single-ended input signals and the unused input of the long-tailed pair is grounded, or connected to a balancing control, or connected to a feedback circuit as in Fig. 15.16.

The main drift in a long-tailed pair is associated with the devices so it is necessary to choose devices in which the parameters controlling drift are as closely balanced as possible. This calls for a match in the absolute value and temperature coefficients of device parameters, and equality of temperature. For vacuum tubes, near-equality of cathode temperature is assured by using twin triodes for the first long-tailed pair. For transistors, almost perfect equality of junction temperature and close matching of parameters can be achieved by using twin devices fabricated in a single chip of silicon; the temperatures of two discrete devices can be matched approximately by mounting them together in a common heat sink or constant temperature enclosure.

15.4.1 Balancing Details for Transistors

The need for close thermal bonding of transistors can be seen from the fact that a temperature differential of 0.1°C between the sides of a long-tailed pair produces a relative input drift voltage of 250 μV, which is considerably higher than the minimum values of drift achieved in well-designed circuits.

For low or moderate values of source resistance, $\partial V_{BE}/\partial T$ is the dominant source of input drift. Although this has a magnitude close to −2.5 mV/°C, the precise value is a function of temperature and varies somewhat from device to device. It has been pointed out that the production spread in $\partial V_{BE}/\partial T$ is considerably less at constant V_{BE} than at constant I_E.* Figure 15.22 shows a reduction in spread for the 2N2484 from 100 μV/°C (at constant I_E) to 5 μV/°C (at constant V_{BE}). Thus, even without any selection or compensation of devices, drifts no more than 5 μV/°C can be expected from voltage-driven long-tailed pairs having thermal bonding between input transistors. If the device temperatures are held constant to better than 1°C, the short-term drift is almost down to the irreducible level of flicker noise. With increased circuit complexity and setting-up procedures, the voltage drift can be reduced below 5 μV/°C. One approach is to slightly unbalance the base-to-emitter

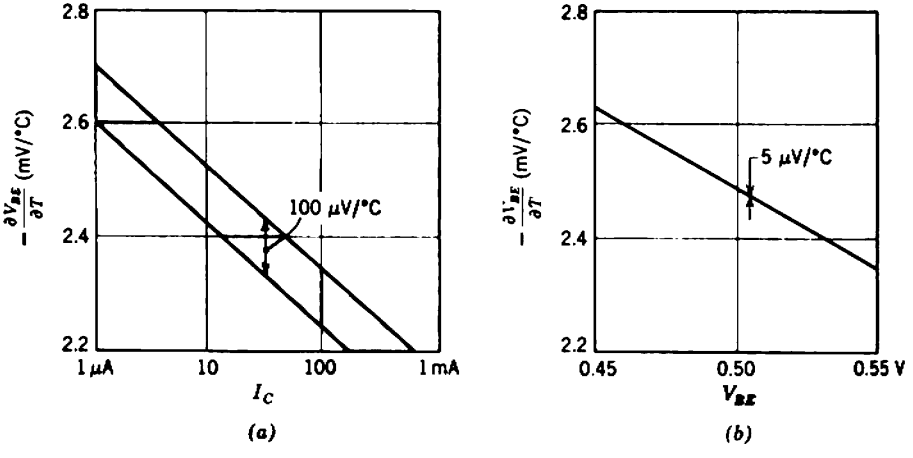


Fig. 15.22 Production spread in $\partial V_{BE}/\partial T$ for 2N2484 transistor: (*a*) at constant emitter current; (*b*) at constant base voltage.

* A. H. HOFFAIT and R. D. THORNTON, "Limitations of transistor dc amplifiers," *Proc. Inst. Elec. Electronics Engrs.*, **52**, 179, February 1964. See also A. H. HOFFAIT and R. D. THORNTON, "Nanovolt transistor dc amplifiers," *Proc. Inst. Elec. Electronics Engrs.*, **51**, 1147, August 1963.

voltages to equalize $\partial V_{BE}/\partial T$; under these circumstances the over-all voltage drift is certainly dominated by flicker noise. The balancing potentiometers are usually brought out as screwdriver adjustments which require attention at intervals of perhaps a year.

The collector currents of the two sides of a drift-balanced long-tailed pair are unlikely to be equal, so additional adjustments are necessary to ensure that there is no difference between the quiescent collector voltages. Figure 15.23 shows a useful circuit arrangement in which a potentiometer in the emitter circuit is adjusted to minimize voltage drift, and two ganged potentiometers in the collector circuit are adjusted to give equal output voltages. This arrangement has the advantage of not unbalancing the load resistance. In addition to changing the emitter currents, the drift balancing potentiometer R_E affects the common-mode rejection, and it is unlikely that the setting for minimum drift and maximum common-mode rejection will coincide. If a very high value of λ is required, as well as low drift, it is necessary to provide an additional adjustment so that the emitter currents can be unbalanced for minimum drift and, in addition, the gain of the two halves can be balanced for maximum common-mode rejection. A fairly typical circuit having provision for drift balancing,

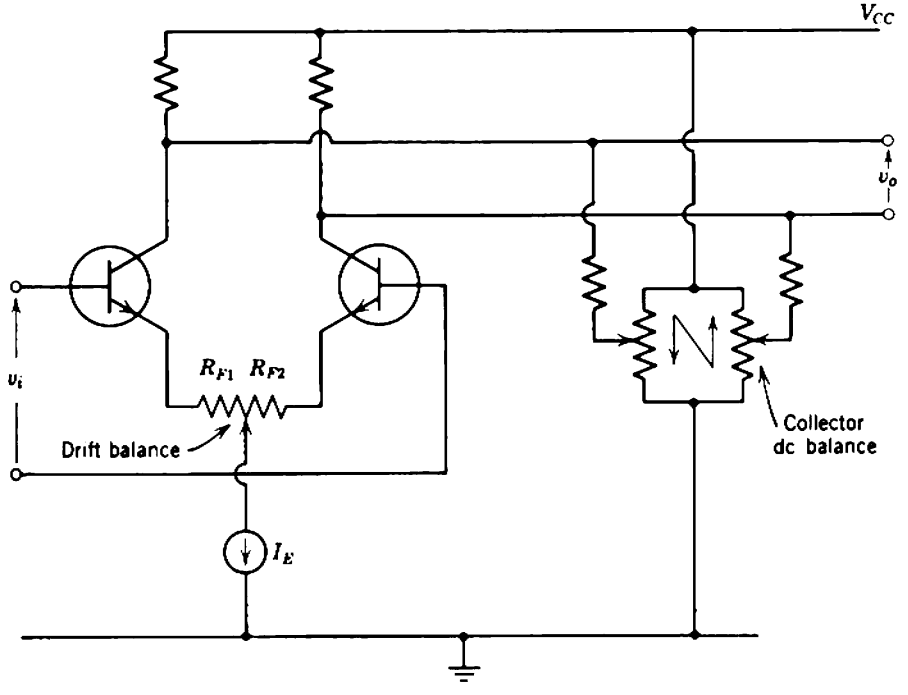


Fig. 15.23 Drift balancing details for a long-tailed pair.

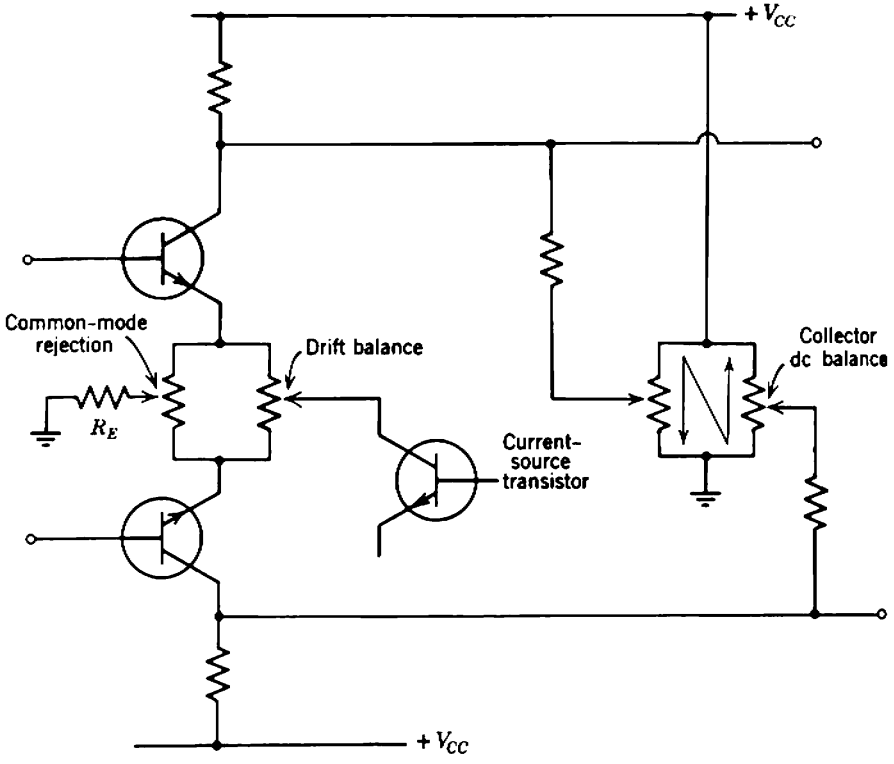


Fig. 15.24 Simultaneous balancing of drift, common-mode rejection, and collector dc voltage.

collector voltage balancing, and common-mode rejection adjustment is shown in Fig. 15.24. The presence of a drift balance adjustment complicates the balancing procedure, because a change in the drift balance control requires subsequent changes in all other controls, except possibly the high-frequency balancing. Thus for ultimate performance the balancing procedure becomes a tedious trial-and-error one which involves at least two (and possibly more) runs over some specified heating and cooling cycle. It is unusual, however, for an amplifier specification to call for the ultimate of performance in both drift and common-mode rejection. Very often an adjustment for minimum drift alone results in acceptable common-mode rejection. Equally, one adjustment to maximize λ and another to equalize collector voltages often yields acceptable drift.

15.4.2 Illustrative Example

The preceding discussion emphasizes the drift reducing properties of the long-tailed pair and suggests that the design of a low-drift dc amplifier

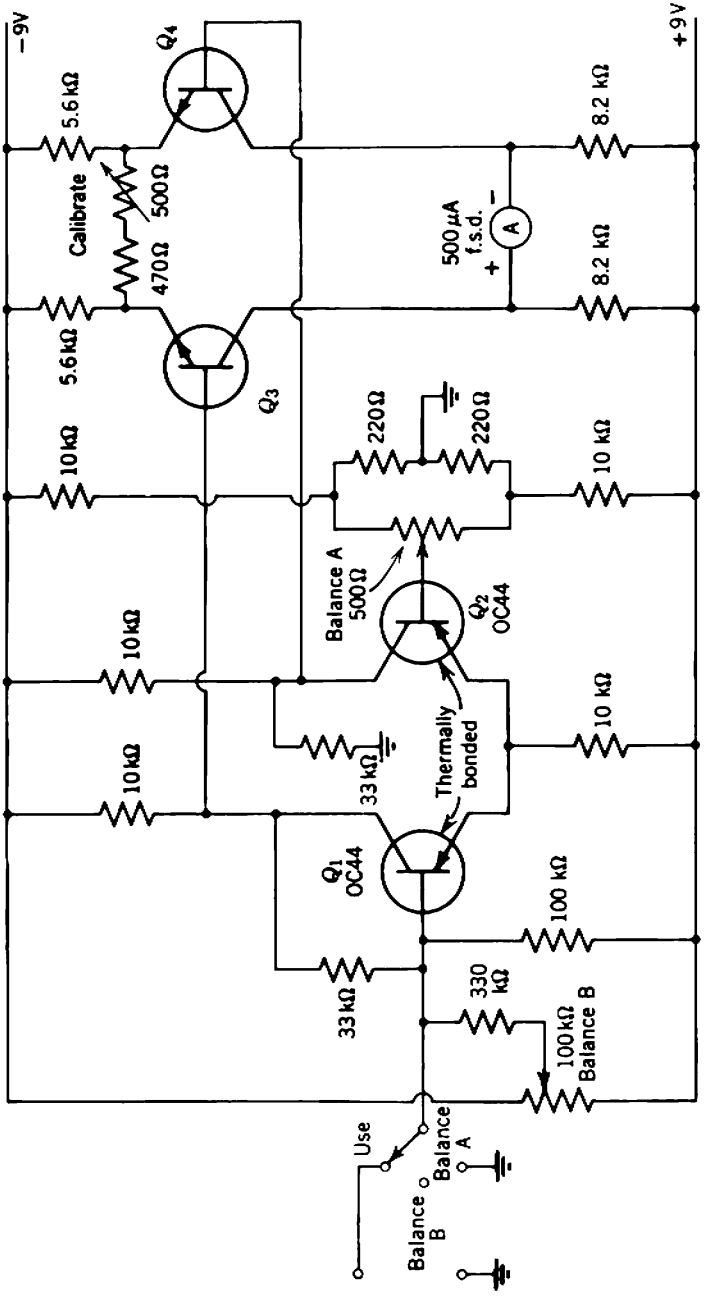


Fig. 15.25 Circuit diagram for a $\times 100$ dc current amplifier.

involves first designing a differential amplifier, and then taking the necessary steps to balance the drift generators on the two sides of each stage. As an example, Fig. 15.25 shows a $\times 100$ current amplifier for converting a ± 500 μA meter into one of ± 5 μA sensitivity. Local shunt feedback is provided around the input side of the first stage, whereas series feedback is applied to both halves of the second stage. Notice that feedback is not required around the second side of the input stage; the 33-kΩ resistor merely serves to equalize the load resistances seen by Q_1 and Q_2. Feedback is not required because the signal current in Q_2 is constrained by the 10-kΩ tail resistor to be equal and opposite to that in Q_1. Transistor Q_2 in fact provides a quasi signal ground for the emitter of Q_1, as required for correct operation of a shunt-feedback stage (Section 15.2.1). The two balancing potentiometers ensure that the meter reads zero when the input terminals are open-circuited and short-circuited, and therefore when any finite resistance is connected between them. No drift balancing is provided, but Q_1 and Q_2 are bonded thermally.

15.5 HIGH-INPUT-IMPEDANCE DC AMPLIFIERS

There are many applications which call for the amplification of voltage signals from sources with an impedance of several megohms or higher. Under such circumstances the input impedance of simple transistor long-tailed pairs is too low and different solutions must be sought. The design problems associated with high source resistance hinge on the requirements of achieving high input resistance, low drift, and adequate gain stability.

For source resistances up to a few megohms, balanced series-feedback stages can be modified by compounding two or more transistors* to give increased current gain (Fig. 15.26); the current gain is approximately the product of β_N for each of the component transistors, and the differential input resistance is $\beta_{N1}\beta_{N2}(R_{F1} + R_{F2})$. The *n-p-n* transistors are operated at low current to minimize the effects of input current drift. The use of three-transistor compounds permits further improvement and leads to excellent designs.† Further examples of compounding are given in Sections 14.4.2 and 16.6.1.

---

* P. J. Beneteau, "The design of high stability dc amplifiers," *Semiconductor Products*, 27, February 1961.

† D. F. Hilbiber, "A new dc transistor differential amplifier," *Trans. Inst. Radio Engrs.*, **CT-8**, 434, June 1961.

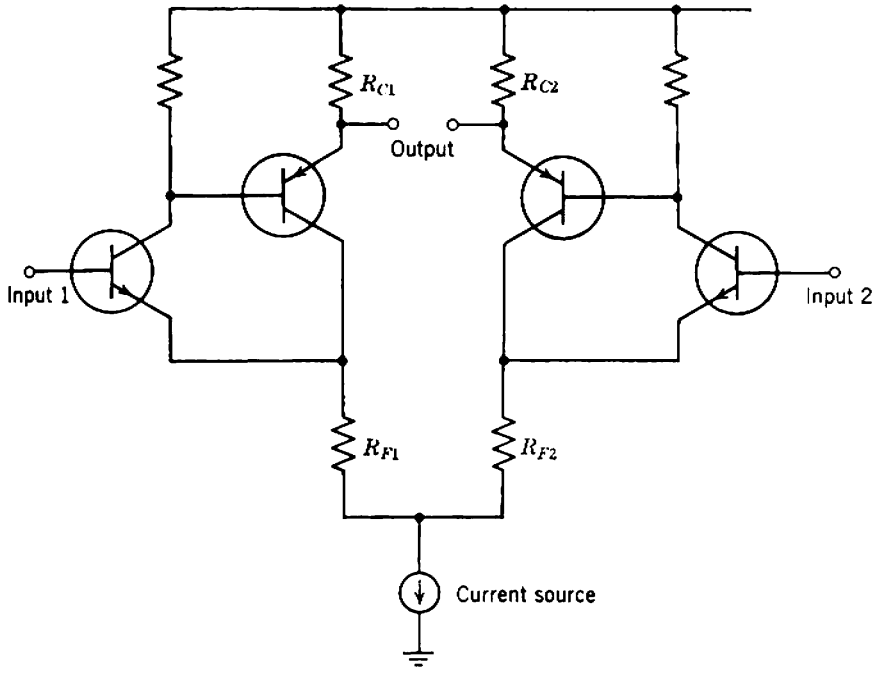


Fig. 15.26 Compound transistors to give high input impedance.

For source impedances of tens or hundreds of megohms it is mandatory to use vacuum tubes or field-effect transistors at the input. For low-level applications the devices need to be matched for drift, and to achieve adequate gain stability and input resistance the common-collecting-electrode (unity voltage gain) configuration is often used. The input device may be compounded with a bipolar transistor, as in Fig. 15.21. Special *electrometer tubes*, with grid currents of the order of picoamperes or less, have been developed for use in high-input-impedance circuits.

Workers in the biological fields have contributed significantly to the development of dc amplifiers with extremely high input resistance (greater than 100 MΩ); such a high input impedance is essential in these applications, since living cells are destroyed by the passage of quite minute currents. The basic technique uses series feedback to achieve an extremely high input resistance and a voltage gain which is usually less than 10. Compounded devices can be used in unity voltage gain cathode- or source-followers in which the input impedance is dominated by the reactance of any stray capacitance C_S (which is unlikely to be less than 10 pF), and the bandwidth is dominated by the R_SC_S time constant. If R_S is 10 GΩ

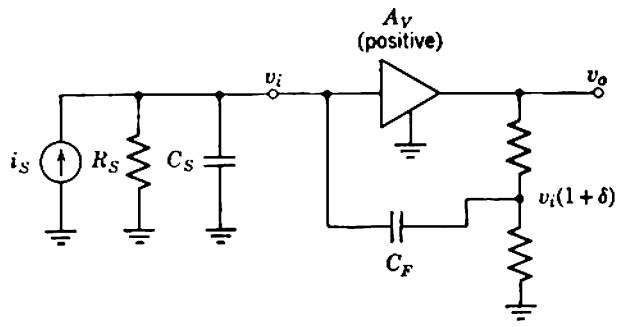


Fig. 15.27 Positive capacitance feedback for very-high-input-impedance applications.

(10^{10} Ω), the bandwidth is only 1.6 Hz. If the voltage gain exceeds unity, capacitive positive feedback can be used as shown in Fig. 15.27 to neutralize the stray capacitance and extend the bandwidth to useful limits. The reader is referred to the literature* for further information on these specialized high-input-impedance techniques.

15.6 PARALLEL-PATH DC AMPLIFIERS

It is shown in Chapter 13 that the transistors in small-signal wide-band amplifiers should operate at emitter currents in the vicinity of 5 mA in order to maximize f_T. Unfortunately, this gives the drift current generator i_D a relatively large value; the emitter current required in the low-drift input stages is small, typically 10 μA. This conflict often cannot be resolved except by using *additive amplification* in a *parallel-path amplifier*, thereby separating the drift problem from the bandwidth problem. Essentially, a parallel-path amplifier consists of two (or occasionally more) input amplifiers, each of which amplifies only a portion of the frequency spectrum. The high-frequency channel is designed primarily

* The following set of papers is especially recommended: E. F. MACNICHOL, "Negative impedance electrometer amplifiers—introduction," *Proc. Inst. Radio Engrs.*, **50**, 1909, September 1962; C. GULD, "Cathode follower and negative capacitance as high input impedance circuits," *Proc. Inst. Radio Engrs.*, **50**, 1912, September 1962; J. W. MOORE and J. H. GEBHART, "Stabilized wide-band potentiometric preamplifiers," *Proc. Inst. Radio Engrs.*, **50**, 1928, September 1962; R. L. SCHOENFIELD, "Bandwidth limits for neutralized input capacity amplifiers," *Proc. Inst. Radio Engrs.*, **50**, 1942, September 1962.

with regard to achieving large bandwidth, and is *RC* coupled so that it cannot contribute to the total drift. The direct-coupled low-frequency channel is designed primarily for low drift. The outputs from the two channels are combined, and usually amplified further to give the output signal, as in Fig. 15.28. Over-all feedback is usually applied to parallel-path amplifiers so that close matching of the gains of the two channels is not required.

For a current-input transfer function the arrangement of low-frequency and high-frequency channels shown in Fig. 15.29*a* is quite straightforward. The open-loop current gains of the two channels are made approximately equal. Resistor R is considerably larger than the input resistance of the high-frequency channel so that, at high frequencies, almost all of the input current flows into the high-frequency channel. The break frequency of R and C is chosen to lie roughly midway between the overlapping high-frequency break of the low-frequency channel and low-frequency break of the high-frequency channel. One possible approach for a voltage-input transfer function is shown in Fig. 15.29*b*; *RC* filters are used for band splitting and recombining. An alternative solution shown in Fig. 15.29*c* uses a transformer to couple in the high-frequency channel. Because the transformer can be designed for satisfactory high-frequency response but extended low-frequency response is not important, this is an entirely practical realization.

An important exception to the above-mentioned conflict between emitter current requirements occurs when the source resistance is very low (of the order of 100 Ω or less) so that the current drift contribution is negligible at the high levels of emitter current required for maximum f_T. In these special cases wide-band long-tailed pairs can be balanced to give drifts of less than 10 μV/°C and there is no need to have separate parallel paths.

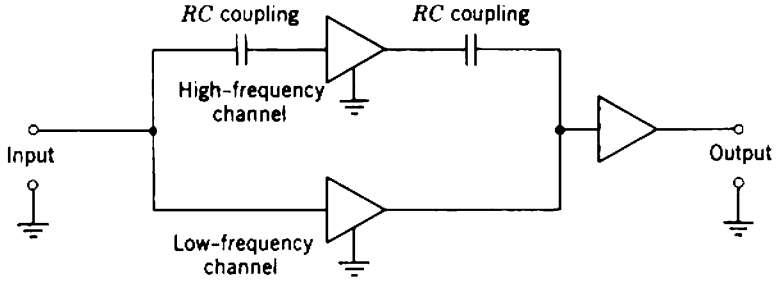


Fig. 15.28 Elements of a parallel-path amplifier.

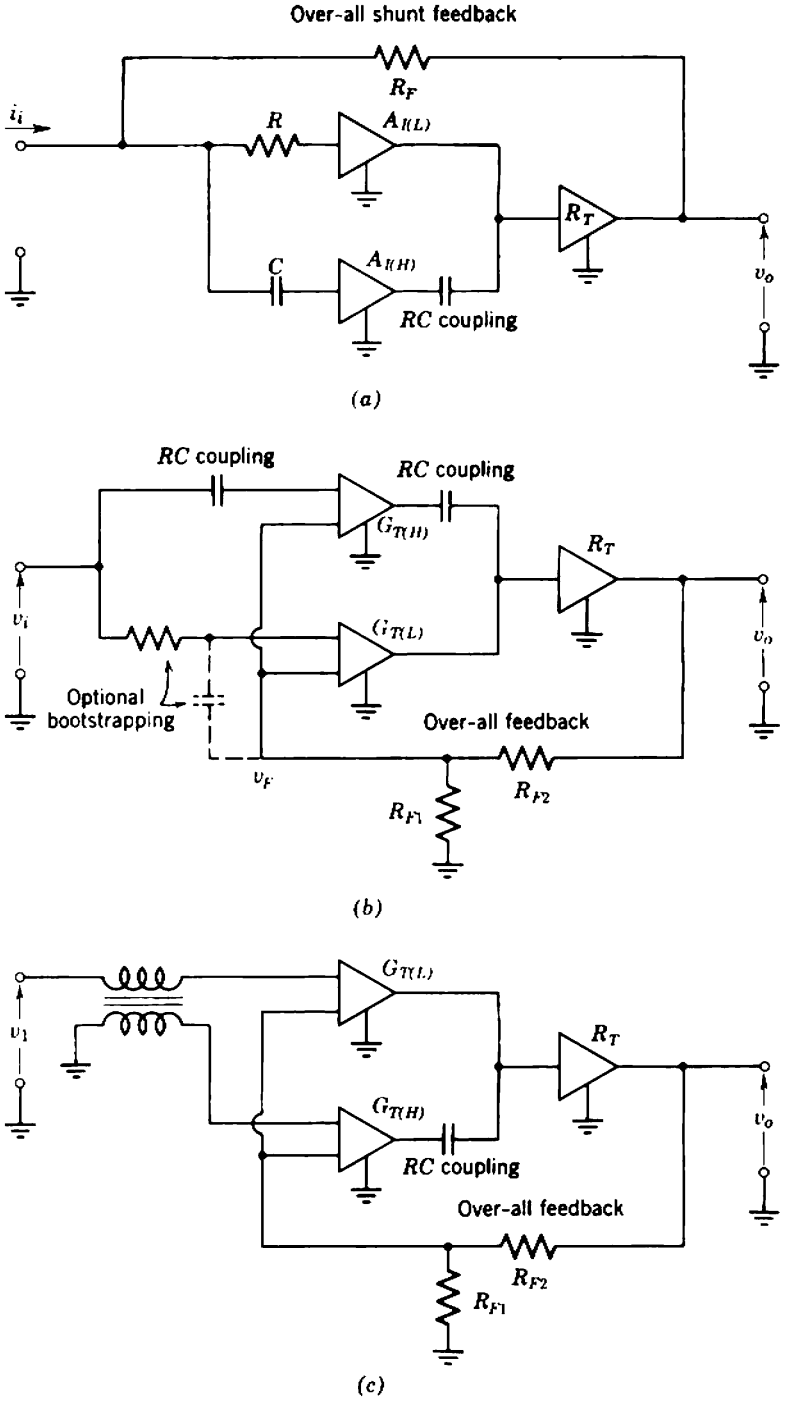


Fig. 15.29 Practical parallel-path amplifiers: (*a*) current-input amplifier; (*b*) voltage-input amplifier, *RC* band-splitting; (*c*) voltage-input amplifier, transformer band-splitting.

15.7 CHOPPER DC AMPLIFIERS

A quite different type of amplifier, suitable either for the dc path of a parallel-path amplifier or for use alone when the bandwidth is small, is the *chopper amplifier.** The incoming base-band signal is converted into a modulated carrier-frequency signal by some form of precise low-level modulator. Highly stable amplification of the modulated signal can then be achieved in a conventional ac amplifier (which is free of drift) prior to precise high-level demodulation to recover the base-band information.

Most standard modulators (such as diode ring modulators and parametric amplifiers) employ the relatively imprecise "soft" nonlinearities of device characteristics. For simplicity and precision, carrier-type dc amplifiers use the "hard" nonlinearity of switching elements for accurate pulse-height modulation. The switching-mode pulse-height modulator is called a *chopper*. The amplified pulse-height-modulated signal is synchronously demodulated.

The advantages of this system all center around the design of an amplifier section which is simple, free from drift, calls for modest orders of power supply stability, and does not require tedious balancing adjustments. These advantages are achieved mainly at the expense of new difficulties in the precise modulation process. Figure 15.30*a* shows an elemental system in which the synchronous modulating and demodulating choppers are represented as a two-pole, two-position switch. Whenever the switches open-circuit the signal path, the coupling capacitors are discharged to ground. Switches in the alternative elemental system of Fig. 15.30*b* short-circuit the signal path and are generally less satisfactory.

Ideally, the resistance across the chopper contacts is zero when they are closed and infinite when they are open. Practical choppers have finite magnitudes for one or both of these quantities and these result in a form of drift in the modulation process. An important additional source of error arises from the finite sampling rate. Nyquist's sampling theorem shows that the maximum bandwidth of a chopper is half the chopping rate; should the input signal contain components of frequency greater than half the chopping rate, the frequency spectrum will be "folded over," resulting in gross distortion in both amplitude and frequency. The input low-pass filter of Fig. 15.30 is essential to band limit the signal; it also prevents the chopper from cyclically changing the input resistance of the amplifier, an important consideration in parallel-path amplifiers which would otherwise

* E. A. GOLDBERG, "Stabilization of wide-band direct current amplifiers for zero and gain," *RCA Rev.*, **11**, 291, June 1950.

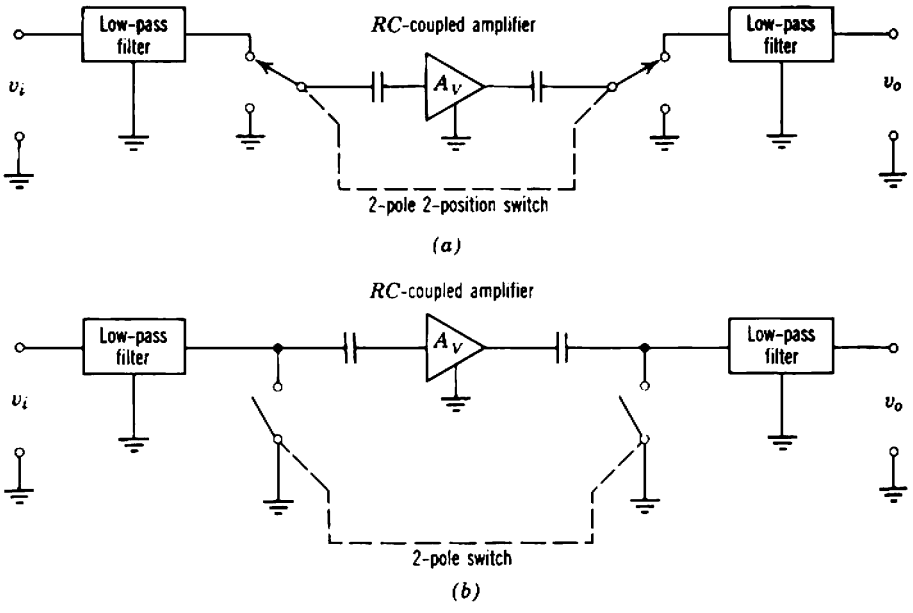


Fig. 15.30 Elements of a chopper amplifier: (*a*) two-pole, two-position chopper; (*b*) simple two-pole chopper.

pass the chopper signal through the high-frequency channel. The low-pass output filter removes components at the chopper frequency and its harmonics. Often both filters are simple *RC* networks with 20-dB/decade roll-off; their 3-dB cutoff frequency must therefore be well below half the chopping rate. In addition to electrical filtering, it may be necessary to use electrostatic or electromagnetic shielding around the chopper to avoid induction of chopper signals into other low-level circuits.

15.7.1 Types of Chopper

The three main types of chopper at present in use may be arranged in order of increasing drift as

(i) mechanical choppers,
(ii) photoconductive choppers,
(iii) transistor choppers (both bipolar and unipolar types).

This list, however, is also roughly in order of increasing operating speed and decreasing complexity, so the decision as to the best chopper for a given application is not simple. The following paragraphs summarize the main features of the various types.

Mechanical choppers are the best approximation to an ideal switch with an "on" resistance of about 50 mΩ (0.05 Ω) and an "off" resistance of about 1 TΩ (10^{12} Ω). Drift arises mainly from differential thermal voltages across the closed contacts; values less than 1 μV are attainable. The maximum chopping speed is limited to about 500 Hz and the power requirements for driving a mechanical chopper are relatively high.

Photoconductive choppers use materials such as cadmium selenide which become conducting when exposed to light. The "on" and "off" resistances are about 2 kΩ and 200 kΩ, respectively. Often, the flashing light source is provided by neon lamps (Fig. 15.31). Drift arises mainly from differential changes in the "on" resistance and finite leakage current across the "off" resistance; these sources contribute voltage and current drift, respectively. Voltage drifts less than 0.5 μV/°C and less than 1 μV/week are attainable. A practical advantage of photoconductive chapters is the absence of transient spikes from the switching waveform. The principal disadvantage is the comparatively slow chopping rate, set by the photoconductor response time; operation is limited to a few hundred hertz. Operating temperature is limited to about 55°C.

Transistor choppers are based on controlling the effective resistance of the conducting channel by a switching signal applied to the control electrode. Because of their small size, modest drive requirements, and

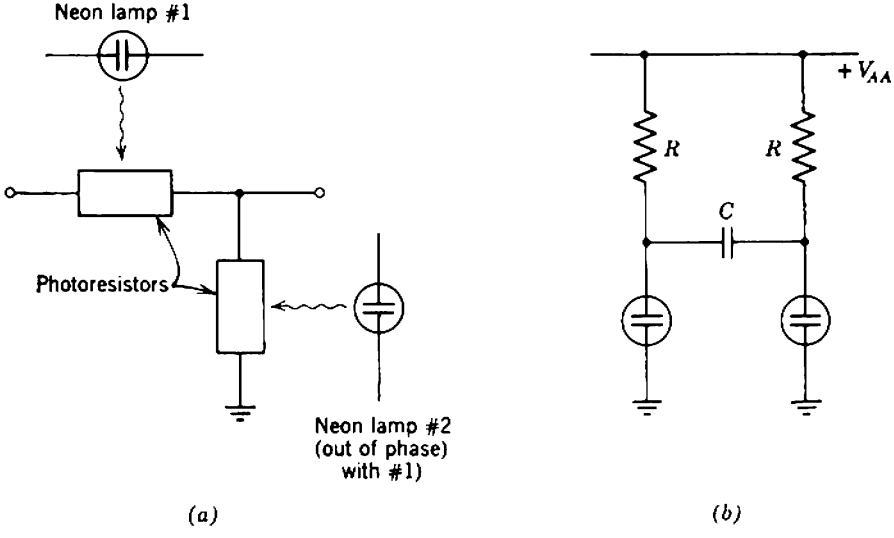


Fig. 15.31 Photoconductive chopper: (*a*) arrangement as two-position switch; (*b*) oscillating circuit for neon lamps.

relatively high operating speeds, bipolar transistors have been used extensively in choppers. They have, however, the great disadvantage that the "on" current flows through the two forward-biased *p-n* junctions in series only when a finite offset voltage (of the order of 1 mV or more) is applied. The temperature coefficient of this offset is about 5 μV/°C. Considerable improvement results from using two transistors with their collectors connected as shown in Fig. 15.32. Special integrated circuit versions of this type of chopper are available (for example, type 3N87) and yield the following performance figures:

offset voltage	50 μV,
"on" resistance	100 Ω
"off" resistance	500 GΩ
drift	1 μV/°C

The field-effect transistor, like the photoconductive chopper, has no offset and therefore shows considerable promise. It follows from Section 4.8.1.2 that the drain characteristics pass through the origin as almost straight lines whose slope is controlled by the gate voltage. Values of "on" and "off" resistance about 150 Ω and 5 GΩ are possible, and circuit packages containing two f.e.t. choppers have been produced.

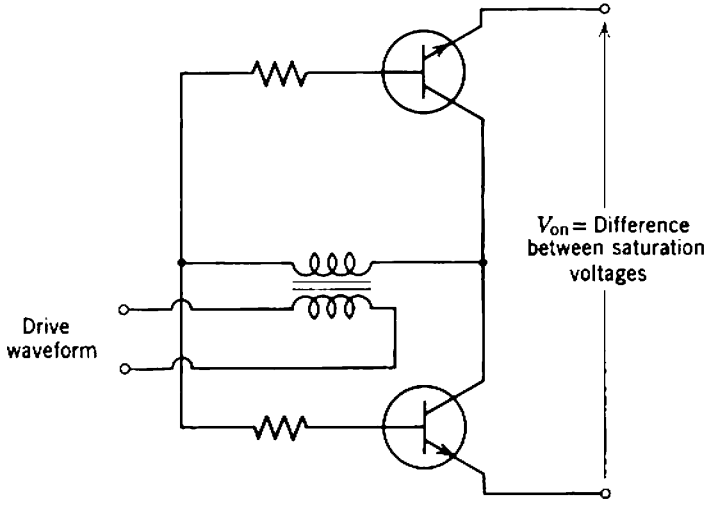


Fig. 15.32 Transistor chopper.

In general, the drift performance of a chopper amplifier is of the same order as a well-designed long-tailed pair. Each approach to drift reduction has its own problems and no clear-cut general decision in favor of any one technique is possible.

15.8 OVERLOAD RECOVERY OF LONG-TAILED PAIRS

The problem of recovery from overload signals is introduced in Section 9.7 where it is shown that paralysis times of dc amplifiers are short because there are no coupling or bypass capacitors to accumulate excess charge during the overload. Single-ended direct-coupled stages, however, can be driven into heavy conduction by overload signals. This at best slows the recovery and at worst can cause damage to the transistors; amplifiers having *p-n-p* and *n-p-n* transistors in adjacent stages are particularly susceptible to damage, because the currents shown in Fig. 15.33 can increase virtually without limit.

The overload currents in a long-tailed pair are limited by the tail current source. Consequently, when severe overload signals are expected, it is advantageous to use long-tailed pairs throughout the amplifier, including the later stages where drift is not a problem. Some economies can be effected in later stages by replacing one transistor of a pair with a diode of comparable switching speed (Fig. 15.34).

An example of a quick-recovery amplifier used for examining the middle 10 mV of a 2-V signal is shown in Fig. 15.35. The three stages have a combined transfer conductance of 600 mA/V with a bandwidth of 30 MHz.

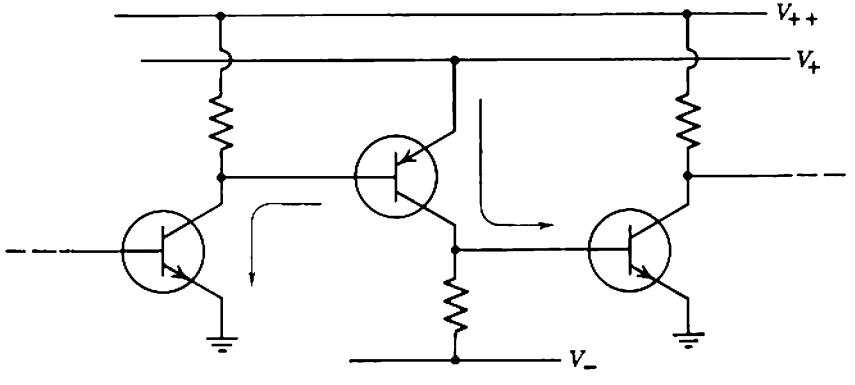


Fig. 15.33 Large overload currents in alternating *n-p-n* and *p-n-p* stages.

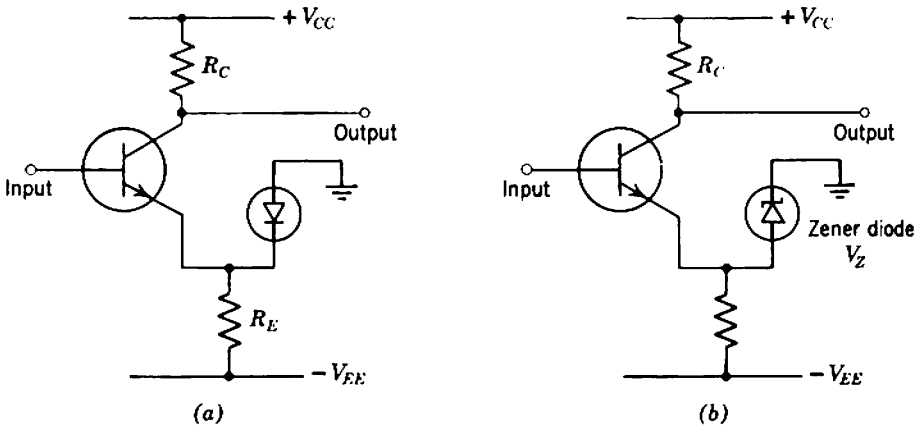


Fig. 15.34 Overload protection with diodes: (*a*) only the portion of the input near ground voltage is amplified; (*b*) only the portion near $-V_Z$ is amplified.

The first low-drift stage is a complete pair, whereas the second and third stages use diodes to simulate the emitter characteristic of a transistor. Additional diodes are placed in shunt with the inputs of the latter stages to reduce signal excursions under extreme overloads, and so keep average dissipation constant and signal-induced drift small.

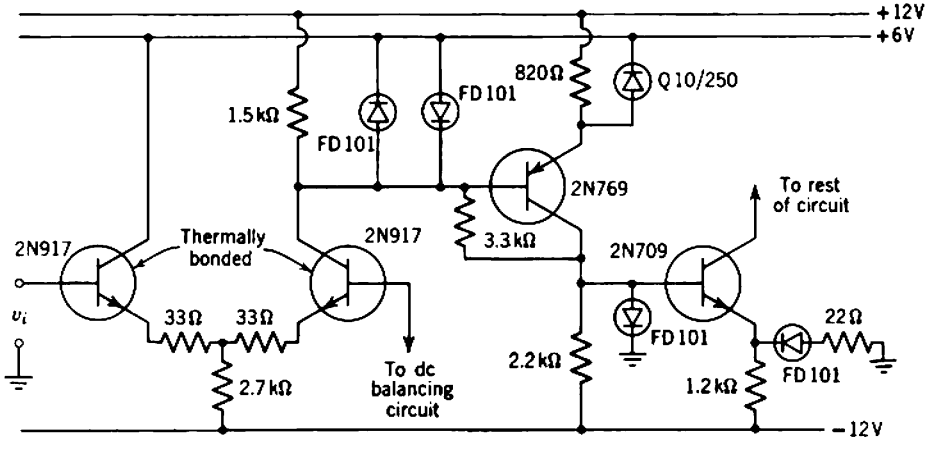


Fig. 15.35 Circuit diagram for a quick-recovery amplifier.

Chapter 16

Large-Signal Amplifiers

In most of the small-signal amplifiers discussed in preceding chapters the onset of moderate nonlinearity limits the useful excursion of the output signal to a relatively small fraction of the range between clipping points. This range depends primarily on the voltage, current, and power ratings of the amplifying device, the type of load, and the voltage and current available from the power supply. If output signals of more than modest magnitude are required, simple amplifiers of the small-signal type may require extravagantly large values of quiescent voltage and/or current. Compelling technical reasons often prohibit such a realization, and a demand arises for amplifiers that make more efficient use of devices and power supplies. Signal excursions in these *large-signal amplifiers* may occupy a large fraction of the range between clipping points.

There is no sharp line of demarkation between small-signal and large-signal amplifiers. Indeed, each must have a nearly linear transfer characteristic to avoid excessive signal distortion. Thus the design of a large-signal amplifier is essentially only an extension of small-signal design practice. This extension, however, is not trivial; the design of large-signal amplifiers calls for more knowledge, ability, and experience than are normally required for most other types of amplifier.

Large-signal amplifiers are required to drive such loads as matched transmission lines, electromechanical transducers such as loudspeakers, and the field or armature windings of rotating machines. These loads are predominantly resistive and much of the theoretical development which follows relies on this assumption. It must be kept in mind, however, that reactive components of the load impedance may become dominant at low and high frequencies. Furthermore, some large-signal amplifiers are used to drive almost perfectly reactive loads such as the deflecting coils or plates of a cathode-ray tube.

One new problem with large-signal amplifiers arises because of the relatively large powers that are dissipated in the amplifying devices; the alternative name *power amplifier* recognizes this fact. It is self-evident that large signal powers can be supplied to a load only if the signal voltage and/or signal current swings are large. These swings may be limited at cutoff by the device voltage rating and at saturation by the device current rating. The ultimate limit, however, is almost always imposed by the power dissipating capability of the device, and this has led to the development of devices which are capable of high power dissipation. Even so, it is still necessary to give very careful consideration to the thermal performance of the devices in a large-signal amplifier. Improper design can lead to overheating and rapid failure of equipment. Because of the large signal powers, a related consideration for large-signal amplifiers is the efficiency of converting dc power from the supply to useful signal power in the load. Quite apart from the effect of internal power losses on device life, it is all too easy for power losses to become an embarrassingly high proportion of the power supplied; this can severely affect the initial cost and subsequent running costs of equipment. The thermal performance of power devices and power conversion efficiency of output circuits are considered in Sections 16.1 and 16.2, respectively.

A second new problem arises from the relatively large nonlinearity of device transfer characteristics when signal swings are large. The major sources of nonlinearity may differ somewhat from those considered for small-signal amplifiers in Chapter 9. Nonlinearity can be reduced either by using devices in combination so that their output signal currents add but their nonlinearities oppose, or by using negative feedback. Section 16.3 outlines several important circuit techniques for reducing distortion but does not pretend to be all-embracing. Undoubtedly, there is still much scope for creative thought.

A third new problem arises in the design of biasing circuits. The simple circuits discussed in Chapter 6 depend for their operation on the development of relatively large voltages (usually greater than one volt) across resistors in series with the collecting or emitting electrodes. These arrangements lead to gross power losses in large-signal amplifiers and may be unsatisfactory for other reasons also. This is particularly the case for circuits in which the quiescent current in the devices is small compared with the signal current (the so-called Class AB or B operations described in Section 16.2). Some suitable biasing techniques are discussed in Section 16.4.

Certain types of output stage consisting of two or more devices must be driven by a balanced push-pull signal. The input signal to the amplifier is usually single-ended and it is necessary to employ a *driving circuit* to

provide the balanced signal required by the output stage. Driver circuits are discussed in Section 16.5. The interdependence of the foregoing considerations is illustrated by the practical design examples of Section 16.6. Finally, Sections 16.7 and 16.8 consider the special problems of large-signal dc amplifiers and regulated power supplies.

16.1 THERMAL PERFORMANCE OF AMPLIFYING DEVICES

The basic temperature limits for vacuum tubes and transistors are discussed in Section 3.3 and 4.5, respectively. In short, the power dissipation rating of most tubes is set by the allowable glass temperature, whereas the limitation for transistors is the temperature of the semiconductor crystal itself. Assuming for simplicity that all of the power dissipation occurs at the collecting electrode, the instantaneous power $P_2(t)$ dissipated within a device consists of two components: an average or dc component P_{20} and a superimposed time-dependent component $P_{2S}(t)$. Thermal performance depends, in general, on the signal magnitudes, their rates of change, and the precise nature of the complete thermal circuit. The apparent thermal time constant of a vacuum tube is of the order of several minutes but the longest period for components of $P_{2S}(t)$ is unlikely to be greater than a fraction of a second. Consequently, the contribution from $P_{2S}(t)$ is usually negligible and it is customary to base designs only on the average power P_{20}. If the magnitude of P_{20} falls within the tube power rating, reliable operation can be expected. Temperature changes are much more rapid in a transistor; both components of dissipated power can influence junction temperature, and it is necessary to consider the thermal characterization of the transistor in some detail. The treatment which follows is based explicitly on the transistor but, if necessary, can be reapplied to other devices.

It is shown in Section 2.7 that a lumped approximation to a thermal circuit can be constructed by considering each distinct region of the heat path separately and relating the thermal power flow to the temperature of the region via the *thermal impedance*. The thermal impedance consists of two terms, a *thermal resistance* θ °C/W through which thermal power flows, and a *thermal capacitance* C J/ C in which thermal energy is stored. There is a direct analogy between a thermal circuit and a topologically similar electric circuit; the analogous quantities are

thermal power — electrical current
temperature — voltage
thermal resistance — electrical resistance
thermal capacitance — electrical capacitance

A typical transistor in combination with its cooling system may have its thermal circuit split up into the sections shown in Fig. 16.1*a*, namely,

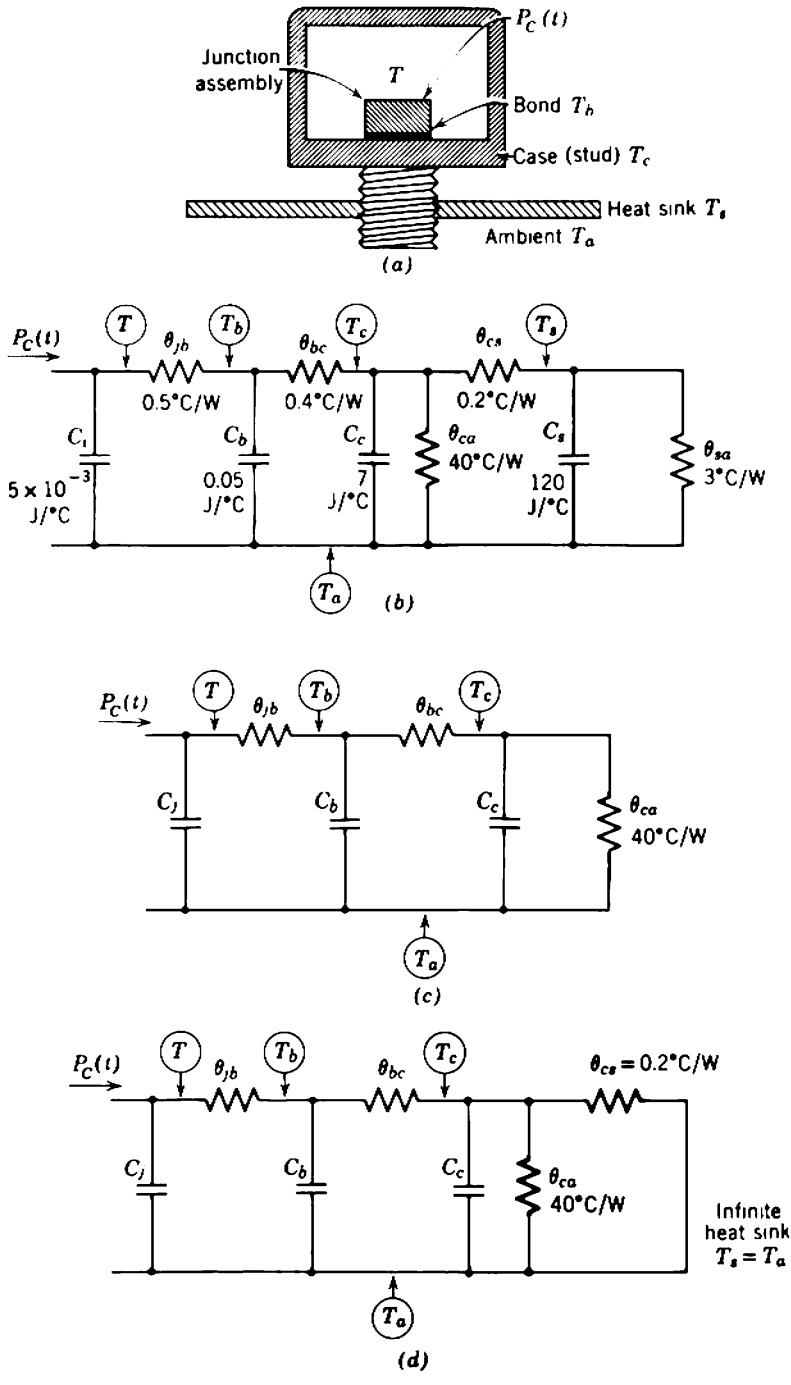


Fig. 16.1 Thermal circuit of a transistor: (*a*) sections of thermal circuit; (*b*) lumped equivalent circuit with heat sink; (*c*) lumped equivalent circuit without heat sink; (*d*) lumped equivalent circuit with infinite heat sink.

(i) the semiconductor junction assembly, temperature T,
(ii) the bond, temperature T_b,
(iii) the metal case or stud, temperature T_c,
(iv) the external heat sink, temperature T_s,
(v) the ambient atmosphere, temperature T_a.

The equivalent lumped thermal circuit is shown in Fig. 16.1*b* with parameter values included for a typical power transistor with a 100 in.^{2} heat sink; these values are assumed in the following discussion. The thermal performance can be changed to a limited extent by changing the accessible heat sink. If the heat sink is absent, the power flow must take place directly (and inefficiently) from the case to the air, and the thermal circuit is changed to that shown in Fig. 16.1*c*. The value 40°C/W for θ_{ca} corresponds to cooling by radiation and natural convection only; it can be reduced substantially by forced air cooling. If the heat sink is infinitely large, the power flow from sink to air is unimpeded and the thermal circuit simplifies to Fig. 16.1*d*.

In principle, the time response of junction temperature can be determined for any time-varying collector power dissipation $P_C(t)$:

$$P_2(t) = P_C(t) = P_{C0} + P_{CS}(t) = V_{CE}(t)\, I_C(t). \tag{16.1}$$

In practice only the simplest time variations of $P_C(t)$ such as sinusoids or pulse trains can conveniently be investigated analytically; more complicated examples can be solved by direct measurements on the electrical analogue. In many instances, however, even this is unnecessary and the following simplifications can be used.

For steady conditions the capacitive reactances approach infinity and, in general,

$$\frac{T_0 - T_a}{P_{C0}} = (\theta_{jb} + \theta_{bc} + \theta_{cs} + \theta_{sa}) = \theta, \tag{16.2}$$

where θ is the total thermal resistance:

$\theta = 4.1$°C/W for the finite heat sink,
$\theta = 40.9$°C/W for no heat sink,
$\theta = 1.1$°C/W for an infinite heat sink.

The designer only has control of the thermal resistance between the case and the air so that the absolute minimum attainable value of θ is

$$\theta_{\min} = \theta_{jb} + \theta_{bc} + \theta_{cs}$$

for the ideal of an infinite heat sink.

Because of the thermal energy shunted into the capacitances, the component of junction temperature rise due to a time-varying power dissipation

$P_{CS}(t)$ is relatively less than that due to P_{C0}. In the numerical case considered, the reactance of C_C becomes negligible for signals having frequencies above 0.1 Hz, and

$$\frac{T_S(t) - T_a}{P_{CS}(t)} \approx \theta_{jb} + \theta_{bc}. \tag{16.3}$$

At higher frequencies the thermal impedance from junction to ambient falls further. Thus it is often permissible to combine Eqs. 16.2 and 16.3 and assume that the total junction temperature is given by

$$T(t) \approx T_a + \theta P_{C0} + (\theta_{jb} + \theta_{bc})P_{CS}(t). \tag{16.4}$$

For example, with the finite heat sink,

$$T(t) \approx T_a + 4.1P_{C0} + 0.9P_{CS}(t),$$

and the instantaneous rise of junction temperature above ambient can be dominated by either the average or the time-varying components or power dissipation.

A very conservative value for maximum junction temperature is given by assuming that the time variation of junction temperature is exactly in step with the time variation of total dissipated power, so that

$$T_{max} \approx T_a + \theta P_C(t)_{max}. \tag{16.5}$$

Table 16.1 Thermal Resistance of Some Common Transistor Packages

Type of Package	Approximate* Junction-to-Case Thermal Resistance θ_{jc} (°C/W)	Approximate Junction-to-Ambient Thermal Resistance θ_{ja} (°C/W)
Jedec TO-18		
(i) Metal	200	600
(ii) Epoxy	200	500
Jedec TO-5		
(i) Metal	60	250
(ii) Epoxy	140	330
Jedec TO-3	1 to 7	†

* This value varies with the detail of the internal thermal circuit. For large devices, substantial variations of θ_{jc} occur.

† Large devices are used with heat sinks which control θ_{ja}.

This is a useful design simplification for some low-power amplifiers. When high powers are to be dissipated, however, the absolute safety of this conservative estimate is not a sufficient repayment for the extravagance of using more expensive transistors whose dissipation ratings are much higher than necessary.

Manufacturers rarely supply precise thermal information, but transistors of similar chip and case size have very similar thermal circuits and it is relatively easy to estimate parameter values for any transistor from the information available on a few basic types. Approximate values of the thermal resistance for standard transistor packages are given in Table 16.1.

16.1.1 Regenerative Thermal Systems

Because transistor parameters depend on temperature, any change in temperature affects the quiescent operating point and, consequently, the power dissipation. There is therefore a feedback path not apparent in the thermal circuits of Fig. 16.1. As with most feedback systems, there is the possibility of regeneration.

For the types of biasing circuit considered in Chapter 6, the feedback is usually provided by the dependence of leakage current I_{CO} on junction temperature. The quiescent collector current can therefore be expressed in the form

$$I_C = I_0 + SI_{CO},$$

where I_0 is independent of temperature, S is a function of the circuit design (its magnitude usually lies between unity and β_N), and

$$I_{CO} = (\text{const}) \exp\left(\frac{T}{a}\right),$$

where the constant a lies between 10 and 14°C. Figure 16.2 shows the thermal feedback loop; the loop gain can be evaluated as follows:

For small changes in temperature,

$$\frac{\Delta I_{CO}}{\Delta T} = \frac{I_{CO}}{a}. \tag{16.6}$$

Also, for a transistor operating under no-signal conditions the power dissipated at the collector junction is

$$P_C = P_{C0} = V_{CE}I_C = V_{CE}(I_0 + SI_{CO})$$

from which

$$\frac{\Delta P_C}{\Delta I_{CO}} = SV_{CE}. \tag{16.7}$$

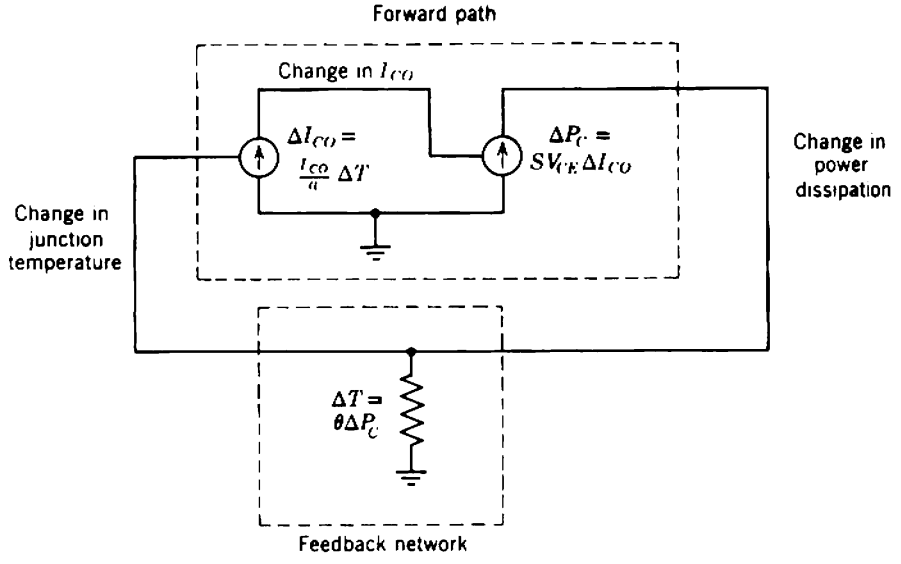


Fig. 16.2 Complete thermal equivalent circuit showing thermal feedback.

Finally, if the system is in thermal equilibrium, the relation between a change in power dissipation and the resultant change in junction temperature is

$$\frac{\Delta T}{\Delta P_C} = \theta.$$

Hence the mid-band thermal loop gain is

$$A_t = \frac{\Delta I_{CO}}{\Delta T} \frac{\Delta P_C}{\Delta I_{CO}} \frac{\Delta T}{\Delta P_C} = \frac{SV_{CE} I_{CO} \theta}{a} \tag{16.8}$$

and is positive; regeneration will occur if $A_t > 1$. Regeneration in a thermal system is known as *runaway* and usually leads to self-destruction.

A more detailed explanation of runaway is provided by Fig. 16.3, which combines plots of power dissipation versus junction temperature with each quantity, in turn, as the independent variable. The family of straight lines represents the junction temperature response to changes in power dissipation considered as an independent variable. There is a different line for each value of ambient temperature, and the incremental response in temperature to an incremental change in dissipated power corresponds to the feedback path in the equivalent circuit of Fig. 16.2. The single curved line of Fig. 16.3 shows the power dissipation response to changes in junction temperature considered as an independent variable. The incremental response in power dissipation to an incremental change in temperature results from the

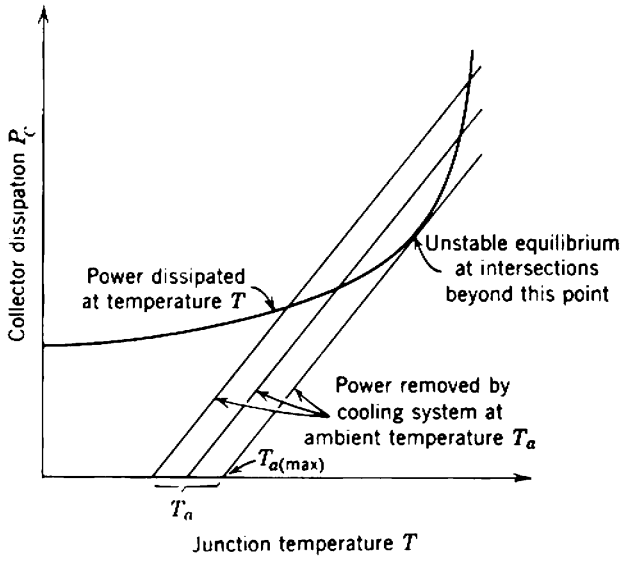


Fig. 16.3 Thermal performance of a transistor when I_{CO} is the dominant temperature-dependent parameter.

change in I_{CO} and corresponds to the forward path in the equivalent circuit of Fig. 16.2. Thermal equilibrium exists when the power dissipated in the device just equals the power removed by the cooling system; the operating power dissipation and junction temperature are at the intersection of the power dissipation curve with the appropriate cooling system line. At an intersection at which the gradient of the power dissipation curve is less than the gradient of the cooling line, the thermal loop gain given by Eq. 16.8 is less than unity and the equilibrium is stable. When the gradient of the power dissipation curve is greater than that of the cooling line, $A_l > 1$ and the equilibrium is unstable. At junction temperatures for which equilibrium does not occur (or, indeed, if no intersection exists for any junction temperature), the temperature will rise if the power dissipation curve lies above the cooling line and fall if the power dissipation curve lies below the cooling line. Thus thermal runaway will occur if either

(i) the ambient temperature T_a exceeds $T_{a(max)}$ in Fig. 16.3,
(ii) $T_a \leq T_{a(max)}$, but the junction temperature T starts from a high enough value.

Clearly, a thermally satisfactory amplifier must be designed so that $T_{a(max)}$ lies above any likely ambient temperature, and so that the operating junction temperature T lies below the rated maximum for the transistor.

The foregoing result applies only under ideal circumstances, but is not too unrealistic for many of the circuits described in Chapter 6. In most instances these biasing circuits are unsuitable for large-signal amplifiers because of the large power losses in the resistors. Consequently, other biasing arrangements are used (Section 16.4) in which the dependence of ΔP_C on ΔT may not be dominated by I_{CO}. In another extreme case the only temperature-dependent term in the collector current is the base-emitter voltage V_{BE}. If the emitter and collector junctions are at the same temperature T, it follows from Eqs. 4.2, 4.30*a*, and 4.37 that

$$I_C \approx (\text{constant}) \exp\left[\frac{q(V_{BE} - V_{\text{gap}})}{kT}\right],$$

where V_{gap} is the gap energy of the semiconductor in electron volts ($V_{\text{gap}} \approx 0.8$V for germanium, $V_{\text{gap}} \approx 1.2$V for silicon). For small changes in temperature

$$\frac{\Delta I_C}{\Delta T} = \frac{qI_C}{kT}\left(\frac{V_{\text{gap}} - V_{BE}}{T}\right)$$

and, at constant collector voltage V_{CE}, the change in power dissipation is

$$\frac{\Delta P_C}{\Delta T} = V_{CE}\frac{qI_C}{kT}\left(\frac{V_{\text{gap}} - V_{BE}}{T}\right). \tag{16.9}$$

The term $(V_{\text{gap}} - V_{BE})/T$ represents the rate of reduction of base-emitter voltage with temperature, and for both germanium and silicon transistors it lies close to 2 mV/°C. Thermal regeneration occurs when

$$A_t = \theta\frac{\Delta P_C}{\Delta T} > 1.$$

Each of the two cases considered above assumes that one temperature-dependent parameter dominates thermal performance. In practice the collector current depends on a number of the transistor parameters and very often additional temperature-dependent biasing elements are used to compensate partly for the dominant temperature effects in the transistors. Thus the general case is not mathematically tractable, but it is always possible to use some combination of experiment and numerical analysis to obtain a plot of power dissipation as a function of junction temperature for any given circuit. The runaway limit is reached at power and temperature coordinates for which the slope of the P_C—T curve is just $1/\theta$. Notice

particularly that the magnitudes of temperature-dependent biasing elements are usually functions of ambient temperature (or, at best, heat sink temperature) but not junction temperature. Thus there may be a different plot of P_C versus T for every ambient temperature.

The foregoing considerations of thermal runaway are based only on quiescent conditions in the absence of signals. Often, but not always, this represents the most critical situation. The details and more general consequences of thermal feedback are too complicated for inclusion in this book and the reader is referred to the literature* for further information.

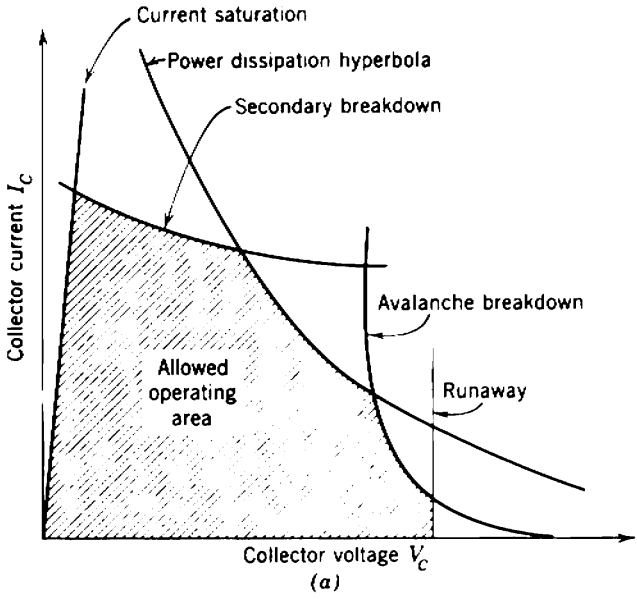
A more important limitation of the foregoing development is that it assumes an idealized situation in which the power is dissipated uniformly across the collector junction; this assumption is implicit in ascribing a constant value to θ_{jb}. It is correct for most transistors of small junction area, but in large-area narrow-base power transistors excessive power dissipation can occur at isolated *hot spots*. Hot spots can develop near the middle of planar junctions because of the poor cooling efficiency, or they can occur near a localized thinning of the base region. Irrespective of the basic causes, the localized increase of power dissipation increases the temperature which, in turn, reduces V_{BE} for the given element of junction area. The consequent localized increase in current (at the expense of current from remote regions of the junction) further increases power dissipation and temperature. Thus hot spots are regions of *current crowding* or *current hogging* and it is easy to see that a localized thermal runaway is quite possible. Because of the minute volume associated with the hot spot, this form of runaway is very much more rapid than normal runaway; the heating is virtually an adiabatic process in which the temperature depends hn the total thermal energy supplied rather than on the time rate of change of energy. Hot spots are most likely to develop when a transistor is passing a large collector current. If runaway occurs at a local hot spot, the collector current increases to a value limited by the external circuit and, since this behavior is similar to that observed for avalanche breakdown at the collector, the phenomenon is usually termed *secondary breakdown*.† If sufficient thermal energy is supplied, secondary breakdown proceeds until the hot spot becomes a molten, short-circuited zone which bridges the base region and destroys the device.

Manufacturers take into account the effects of runaway, avalanche breakdown, and secondary breakdown in the rating of power transistors and specify allowed operating areas of the form shown in Fig. 16.4. Notice

* O. MUELLER, "Internal thermal feedback in four-poles, especially in transistors," *Proc. Inst. Elec. Electronics Engrs.*, **52**, 924, August, 1964.

† H. MELCHIOR and M. J. O. STRUTT, "Secondary breakdown in transistors," *Proc. Inst. Elec. Electronics Engrs.*, **52**, 439, April, 1964.

that there is some relaxation of the limits due to secondary breakdown when pulse signals are employed, because there is less time for local heating.



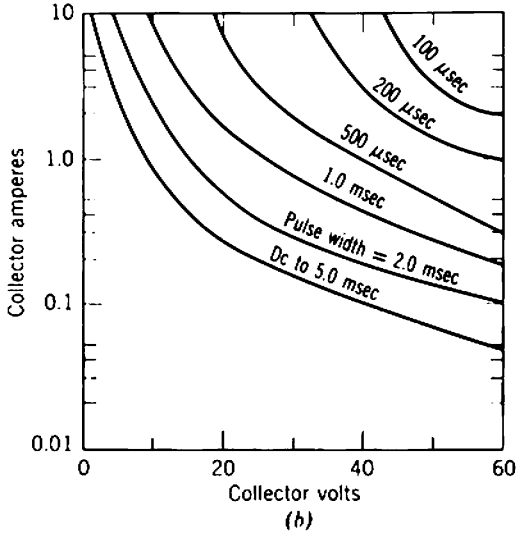


Fig. 16.4 Allowed operating areas for junction transistors: (*a*) boundaries under quiescent conditions; (*b*) manufacturers' pulse-signal ratings for type SE3030 (Courtesy Fairchild Camera and Instrument Corporation).

16.2 POWER RELATIONS IN AMPLIFIERS

It is usual to base an elementary discussion of power relations on the assumption of sinusoidal signals. This has the merits of standardization and analytical simplicity, but the results obtained must be interpreted with care as most actual signals are not sinusoidal.

16.2.1 Ideal Class A Operation

Consider a device operating at mid-band frequencies and suppose that the load seen by it is resistive and of magnitude R_L. In practice the load network may take a variety of forms, of which the most common are illustrated in Fig. 16.5:

(a) a simple load resistance $R_2 = R_L$ in series with the collecting electrode,
(b) RC coupling, for which $R_L = R_2 \parallel R_{ext}$,

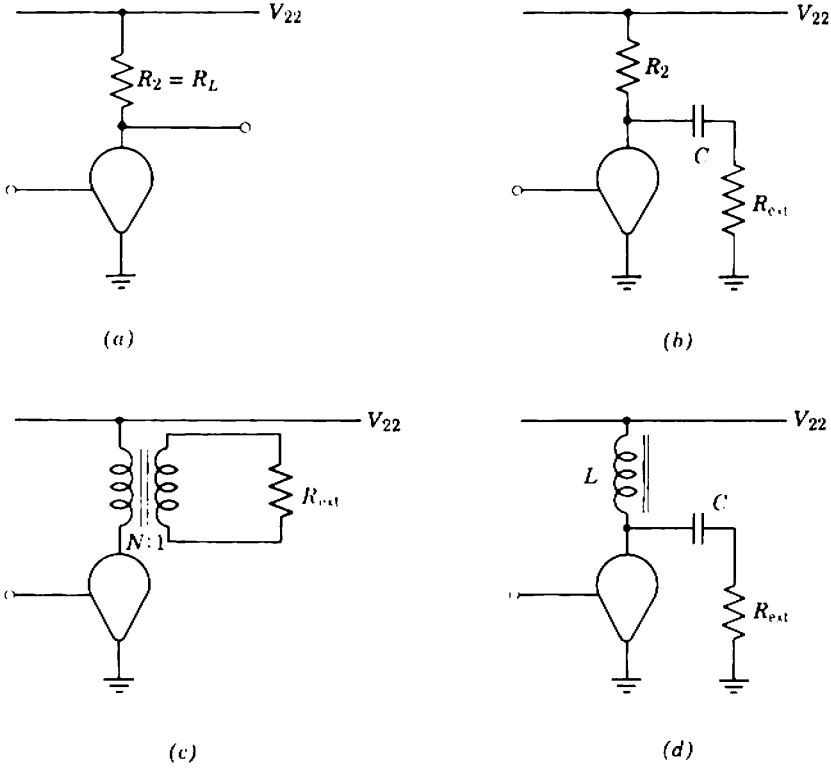


Fig. 16.5 Typical load networks: (*a*) direct coupling; (*b*) RC coupling; (*c*) transformer coupling; (*d*) inductor coupling.

(c) ideal transformer coupling, for which $R_L = N^2 R_{ext}$,
(d) ideal inductor coupling, for which $R_L = R_{ext}$.

An ac load line representing R_L can be drawn through the quiescent point on the collecting-electrode characteristics, as in Fig. 16.6. Assume that the output voltage swings sinusoidally about the quiescent point V_{2Q} and that its peak amplitude is kV_{2Q}. The parameter k must lie between zero (for no signal) and unity (at maximum signal), this limit being set by zero instantaneous collecting-electrode voltage. In practice the maximum value of k can never reach unity, but this limitation is neglected in the ideal theory. It is further assumed that waveform distortion is negligible for k less than unity. Consequently, the operating locus is terminated at output voltages of zero and $2V_{2Q}$ as shown for various values of R_L in Fig. 16.6.

The corresponding current waveform swings about the quiescent point I_{2Q} with peak amplitude $k'I_{2Q}$, where $k' \leq k$ because the collecting-electrode current cannot be negative. The signal power supplied to the load follows as

$$P_{LS} = \left(\frac{kV_{2Q}}{\sqrt{2}}\right)\left(\frac{k'I_{2Q}}{\sqrt{2}}\right).$$

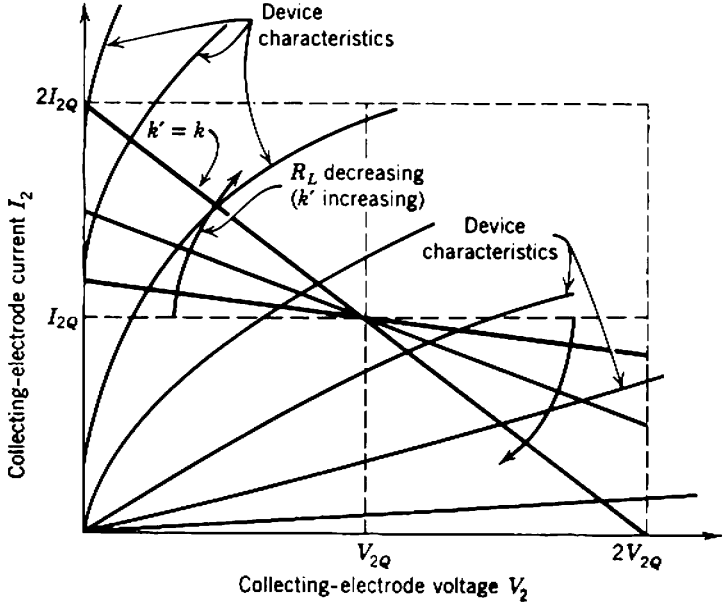


Fig. 16.6 Mid-band Class A operating locus (ac load line) showing the effect of increasing k' from 0 to k.

For given drive factor k, the load power will be as large as possible when $k' = k$ so that

$$R_L = \frac{V_{2Q}}{I_{2Q}} \tag{16.11}$$

and

$$P_{LS} = \tfrac{1}{2}k^2 V_{2Q} I_{2Q}. \tag{16.12a}$$

Maximum output then occurs at full drive when $k = 1$:

$$P_{LS(\max)} = \tfrac{1}{2} V_{2Q} I_{2Q}. \tag{16.12b}$$

Equation 16.11 is of extreme importance in Class A amplifier design and must be satisfied if maximum signal power is to be delivered to the load. The condition is assumed to be satisfied in the following development. Notice that maximum available signal power does not correspond to matching the load to the amplifier output resistance.

The instantaneous device voltage and current are

$$V_2(t) = V_{2Q}(1 + k \cos \omega t),$$
$$I_2(t) = I_{2Q}(1 - k \cos \omega t),$$

as shown in Fig. 16.7. The instantaneous power dissipation is

$$P_2(t) = V_{2Q}(1 + k \cos \omega t)\, I_{2Q}(1 - k \cos \omega t);$$

that is,

$$P_2(t) = V_{2Q} I_{2Q}[(1 - \tfrac{1}{2}k^2) - \tfrac{1}{2}k^2 \cos 2\omega t]. \tag{16.13}$$

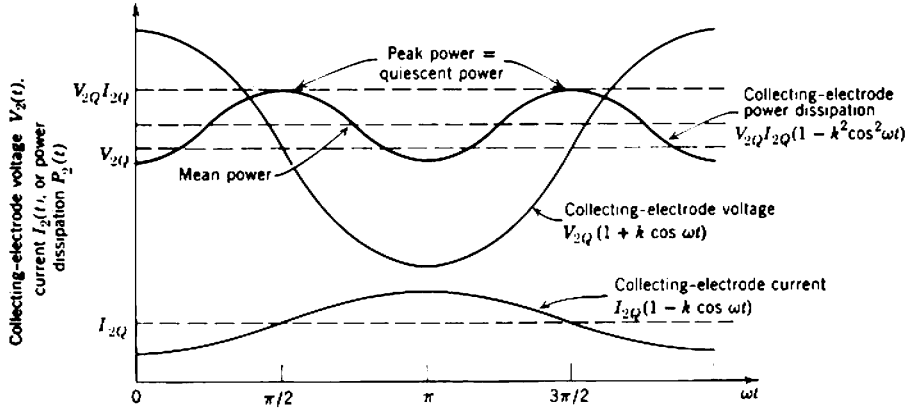


Fig. 16.7 Instantaneous voltage, current, and power waveforms in a Class A amplifier.

Maximum instantaneous (peak) power dissipation (of interest in the conservative design of transistor amplifiers, Section 16.1) occurs when

$$\omega t = \frac{\pi}{2}, \frac{3\pi}{2}, \frac{5\pi}{2}, \cdots$$

and its value is

$$\hat{P}_2 = V_{2Q}I_{2Q} = 2P_{LS(\max)}, \tag{16.14}$$

quite independent of the value of k. The average power dissipation is

$$\bar{P}_2 = V_{2Q}I_{2Q}(1 - \tfrac{1}{2}k^2). \tag{16.15a}$$

Maximum average power dissipation (of interest in the design of vacuum-tube amplifiers) occurs when k is zero, that is, under no-signal conditions, for which

$$\bar{P}_{2(\max)} = V_{2Q}I_{2Q} = 2P_{LS(\max)}. \tag{16.15b}$$

The mean power drawn from the dc supply is shared between the device and load and is given by

$$\bar{P}_{\text{tot}} = V_{22}I_{2Q}. \tag{16.16}$$

The *power conversion efficiency* of the circuit is defined as

$$\eta = \frac{P_{LS}}{\bar{P}_{\text{tot}}} \tag{16.17}$$

and the value of η depends on the detail of the circuit as well as on the drive parameter k.

Conversion efficiency is as high as possible in Figs. 16.5*c* or *d*, because there is (ideally) no dc voltage drop and power wastage in the supply impedance; hence $V_{22} = V_{2Q}$, and substitution into Eqs. 16.12 and 16.16 yields

$$\eta = \tfrac{1}{2}k^2. \tag{16.18a}$$

Maximum conversion efficiency of 50% occurs when $k = 1$. In Figs. 16.5*a* and *b*, the conversion efficiency is less because of the dc power wastage in R_C. For the simpler case (Fig. 16.5*a*) $V_{22} = 2V_{2Q}$, so that

$$\eta = \tfrac{1}{4}k^2 \tag{16.18b}$$

and the maximum conversion efficiency is only 25%. These results can be given a geometrical interpretation. Figure 16.8*a* shows the load line corresponding to Fig. 16.5*a*; the shaded areas correspond to signal power in the load (at $k = 1$) and the power drawn from the supply. Similarly, Fig. 16.8*b* corresponds to either Fig. 16.5*c* or *d*.

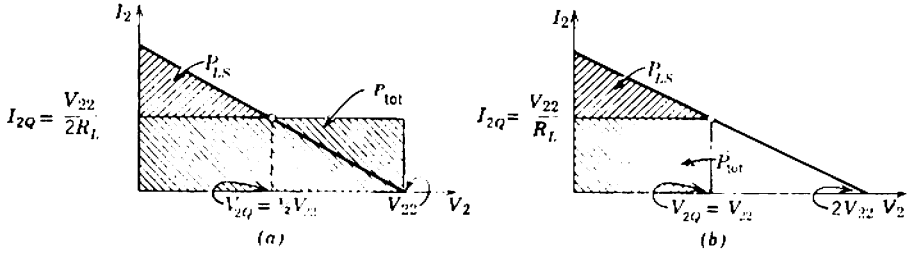


Fig. 16.8 Geometrical interpretation of maximum conversion efficiency: (*a*) direct coupling in which $\eta_{max} = 25\%$; (*b*) transformer or inductor coupling in which $\eta_{max} = 50\%$.

16.2.2 Actual Class A Operation

The foregoing theory assumes an ideal device, that is, one in which k can approach unity. This is a good approximation for transistors provided that the supply voltage is much larger than the relatively low saturation voltage defined by the knee of the collector characteristic, and provided that the load line does in fact pass through the knee. A similar situation exists with the pentode, but in this instance there is an additional power loss due to the screen dissipation:

$$P_S = V_S I_S.$$

For a triode, however, the situation is far from ideal. Consider the transformer-coupled triode stage of Fig. 16.9 for which the piecewise-linear anode characteristics (Section 6.2.1) have mean slope $1/r_A$. It is obvious from the geometry of the figure and can be verified analytically that a maximum conversion efficiency of only 25% is achieved when

$$R_L = 2r_A. \tag{16.19}$$

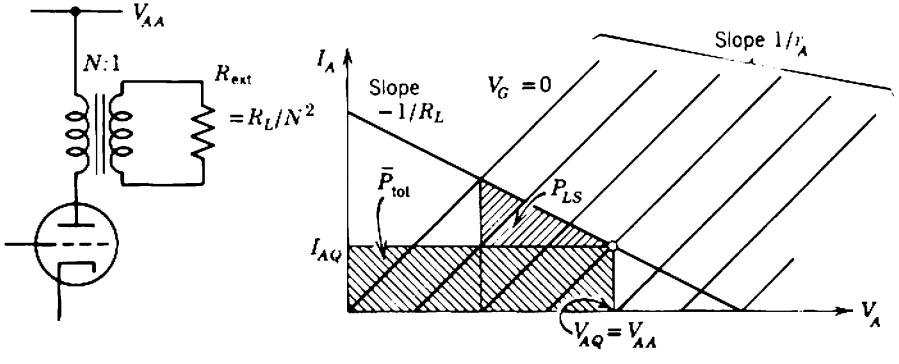


Fig. 16.9 Conversion efficiency of triode with transformer coupling: $\eta_{max} = 25\%$.

Thus, in order to approach maximum conversion efficiency, there are definite limits on the slope of the ac load line; for transistors and pentodes R_L must be sufficiently large to ensure that the load line intersects the knee of the output characteristics, whereas R_L for triodes must have a value close to twice r_A. In addition, the load line must remain within the allowable operating areas of the device; in the simplest case, these are bounded by the power dissipation hyperbola

$$P_2 = V_2 I_2 \leq P_{2(\max)}$$

and the maximum ratings for voltage and current (Fig. 16.10). Transistors usually have more complicated shapes for the operating area, for example, Fig. 16.4. Suitable combinations of magnitudes for R_L and the quiescent voltage and current must be chosen for achieving efficient power conversion and high power output without exceeding device ratings. In addition to removing the power loss in R_L due to quiescent current flow, output transformers provide a convenient means for changing the resistance of some arbitrary external load R_{ext} to the optimum value required by the device.

In conclusion, it should be emphasized that the foregoing development makes one further idealization in assuming that the output signals lie in a frequency range for which the load line (or, more generally, operating locus) is straight. For reactively coupled loads the operating locus at low frequencies becomes an ellipse and the instantaneous power dissipation can exceed $V_{2Q} I_{2Q}$ (see Problems 2.3 and 16.3).

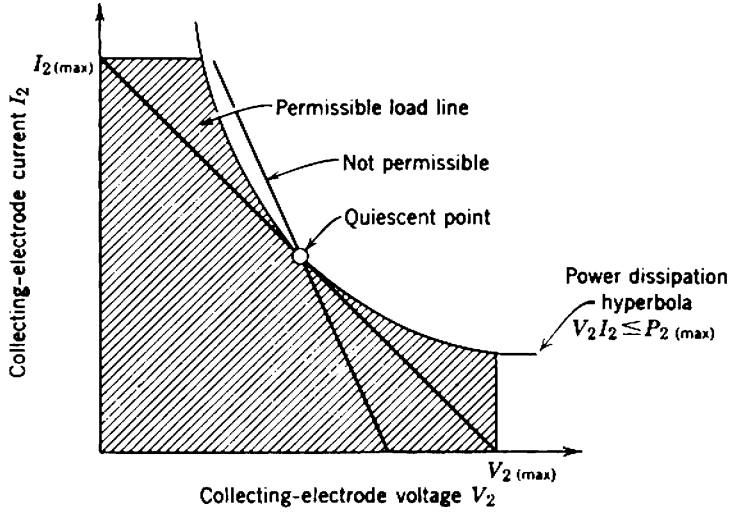


Fig. 16.10 Permissible load line for Class A operation.

16.2.3 Ideal Class B Operation

Provided that two output devices are arranged to deliver a push-pull output signal, higher conversion efficiency and lower peak dissipation are obtained when each device conducts only for the angle of 180° corresponding to the low-voltage portion of its cycle. This is Class B operation in which the quiescent point for each device lies at the end of the load line on the output voltage axis. Figures 16.11*a* and *b* show two possible arrangements—*push-pull* and *single-ended push-pull*, respectively. Notice that the latter arrangement does not necessarily require an output transformer although one may be desirable for impedance transformation. It is also possible to make use of *n-p-n* and *p-n-p* transistors so that both transformers can be omitted; Section 16.3.4 contains a discussion.

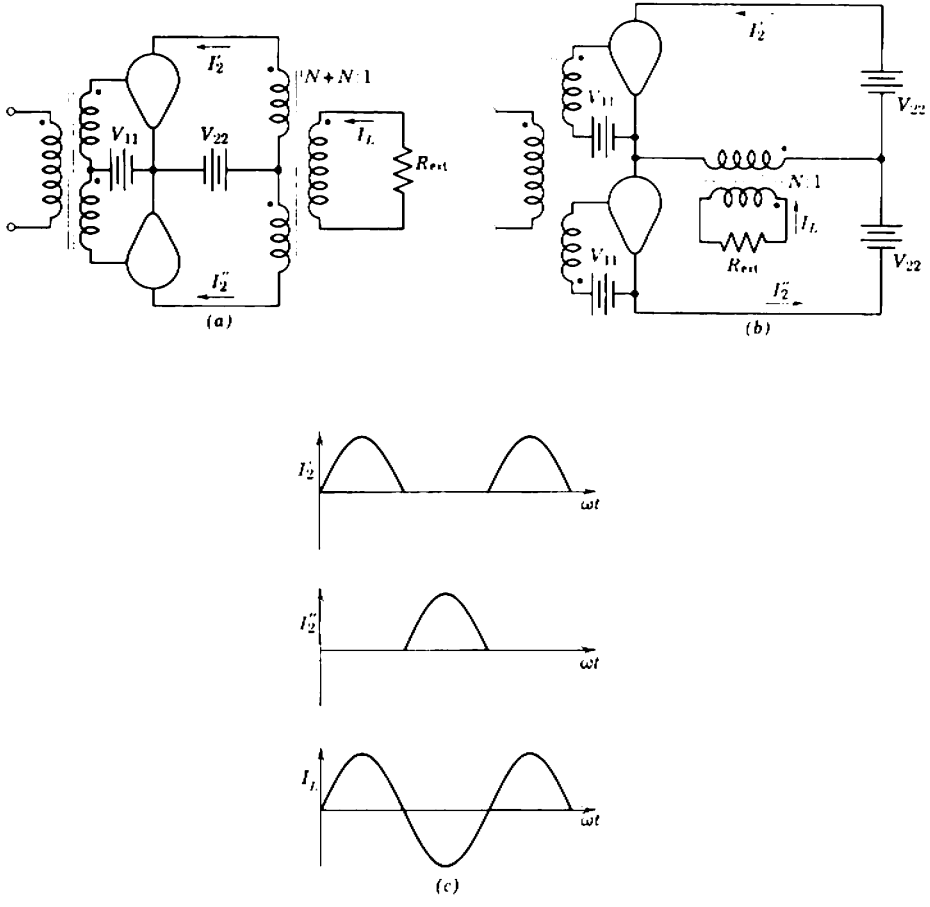
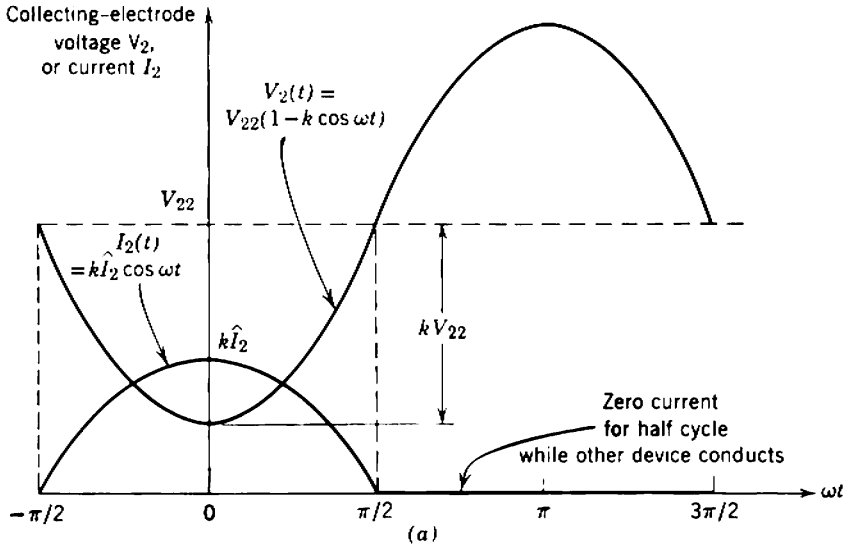


Fig. 16.11 Circuits for Class B operation: (*a*) push-pull; (*b*) single-ended push-pull; (*c*) current waveforms.

Consider the power relations for a Class B amplifier in which the voltage and current waveforms for one device are shown in Fig. 16.12*a*. The



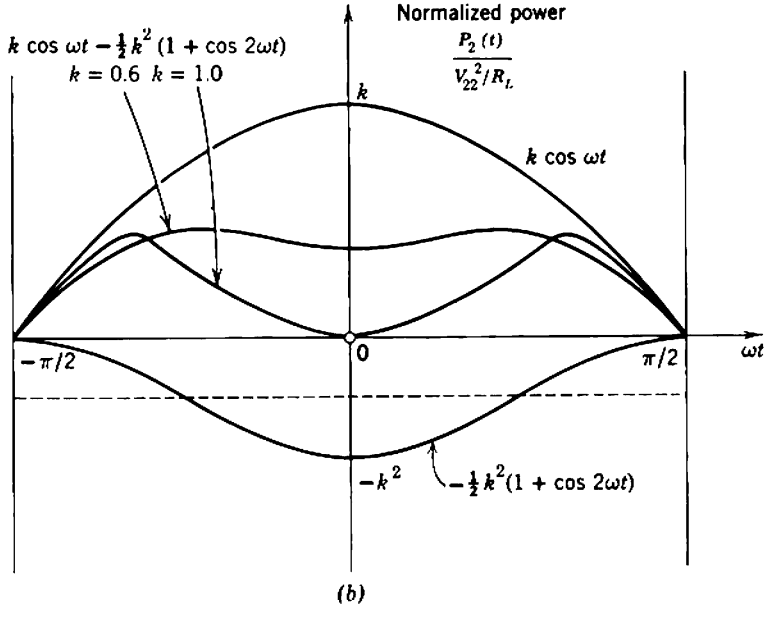


Fig. 16.12 Voltage, current, and power waveforms for one device in a Class B amplifier.

quiescent collecting-electrode voltage is equal to the supply voltage V_{22}, and $\hat{I}_2$ is the peak current. The signal power in the load is

$$P_{LS} = \tfrac{1}{2}k^2 V_{22}\hat{I}_2.$$

To eliminate $\hat{I}_2$ notice that

$$\hat{I}_2 = \frac{V_{22}}{R_L}, \tag{16.20}$$

so that

$$P_{LS} = \frac{k^2 V_{22}^2}{2R_L}. \tag{16.21a}$$

Maximum output occurs at full drive for which $k = 1$:

$$P_{LS(\max)} = \frac{V_{22}^2}{2R_L}. \tag{16.21b}$$

The instantaneous power dissipated in each device is

$$P_2(t) = V_{22}(1 - k\cos\omega t)\hat{I}_2 k\cos\omega t,$$

that is,

$$P_2(t) = \frac{V_{22}^2}{R_L}[k\cos\omega t - \tfrac{1}{2}k^2(1 + \cos 2\omega t)]. \tag{16.22}$$

Waveforms are shown in Fig. 16.12*b*. The maximum peak dissipation (of interest in transistor amplifiers) is independent of the drive factor k and occurs for any type of signal whenever both the voltage and current swing have half their peak values:

$$\hat{P}_2 = \frac{V_{22}^2}{4R_L} = \tfrac{1}{2}P_{LS(\max)}. \tag{16.23}$$

The average power dissipation per device over a complete cycle is

$$\bar{P}_2 = \frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} P_2(\omega t)\, d\omega t$$

and, from Eq. 16.22,

$$\bar{P}_2 = \frac{V_{22}^2}{R_L}\left(\frac{k}{\pi} - \frac{k^2}{4}\right) = P_{LS(\max)}\left(\frac{2k}{\pi} - \frac{k^2}{2}\right). \tag{16.24a}$$

By differentiation, maximum average dissipation (of interest in vacuum-tube amplifiers) occurs when

$$k = \frac{2}{\pi} \approx 0.64$$

for which

$$\bar{P}_{2(\max)} = \frac{V_{22}^2}{R_L}\left(\frac{1}{\pi^2}\right) = P_{LS(\max)}\left(\frac{2}{\pi^2}\right) \approx 0.20P_{LS(\max)}. \tag{16.24b}$$

Thus maximum average dissipation occurs at 64% of full signal amplitude, corresponding to 41% of maximum signal power. By comparison of Eqs. 16.23 and 16.24*b* maximum average dissipation is 40% of peak dissipation.

The power conversion efficiency follows from Eq. 16.17 as

$$\eta = \frac{P_{LS}}{P_{LS} + 2\bar{P}_2},$$

where the factor 2 appears in the denominator because there are two devices. Substitution from Eqs. 16.21*a* and 16.24*a* yields

$$\eta = \tfrac{1}{4}k\pi \tag{16.25}$$

Maximum efficiency of 78% occurs at full drive for which $k = 1$; this represents a significant improvement over Class A operation for which η is limited to 50%.

For a given output signal power, Class B operation allows the use of devices of lower power ratings. If the thermal properties of the devices are such that average power $\bar{P}_2$ tends to be more significant than peak power $\hat{P}_2$, two devices in Class B are capable of 10 times as much power as one device in Class A. If $\hat{P}_2$ is more significant than $\bar{P}_2$, the advantage is reduced to a factor 4.

True Class B operation (with devices biased to the cutoff point) is rarely desirable, because g_m vanishes when the current falls to zero and this leads to excessive distortion. Thus in practice it is necessary to compromise between Class A and Class B operations so that the angle of current flow in any device is slightly greater than 180°. This is discussed further in the following sections. Furthermore, all devices have a finite knee voltage and this reduces the maximum Class B conversion efficiency below the ideal 78%. Actual values of η greater than 70% can be achieved with transistors. Baker* has analysed Class B operation with imperfect devices.

16.3 NONLINEARITY

The discussion of static nonlinearity in Chapter 9 is based largely on the idea of differential error γ, that is, on the variation of the small-signal mid-band transfer function with input signal level. This specification of nonlinearity is also appropriate for large-signal amplifiers. For signal excursions which remain in the general vicinity of the quiescent operating point the differential error can usually be predicted accurately from the theoretical laws of small-signal parameter variation with operating point (as in Section

* L. Baker, "Power dissipation in Class B amplifiers," *Trans. Inst. Radio Engrs.*, **AU-10**, 139, September–October, 1962.

9.5). In large-signal amplifiers, however, the extremities of signal excursions approach the clipping points so that the character of the nonlinearity changes. This change is shown for a pentode stage in Fig. 16.13:

As the signal approaches the current-saturation (knee) portion of the characteristic, the transfer function changes from a $\frac{3}{2}$-power law (shown by a dashed line) to the more complicated curve shown by the full line. The differential error changes from an approximately antisymmetric function of $(V_G - V_{GQ})$ to an approximately symmetric function, and second-harmonic distortion falls at the expense of some increase in third. It is no longer possible to express γ as a simple, general mathematical function, but recourse may be made to graphical or experimental techniques for determining its value.

The discussion of mid-band small-signal amplifiers in preceding chapters concentrates on minimizing the uncertainty in transfer functions. In

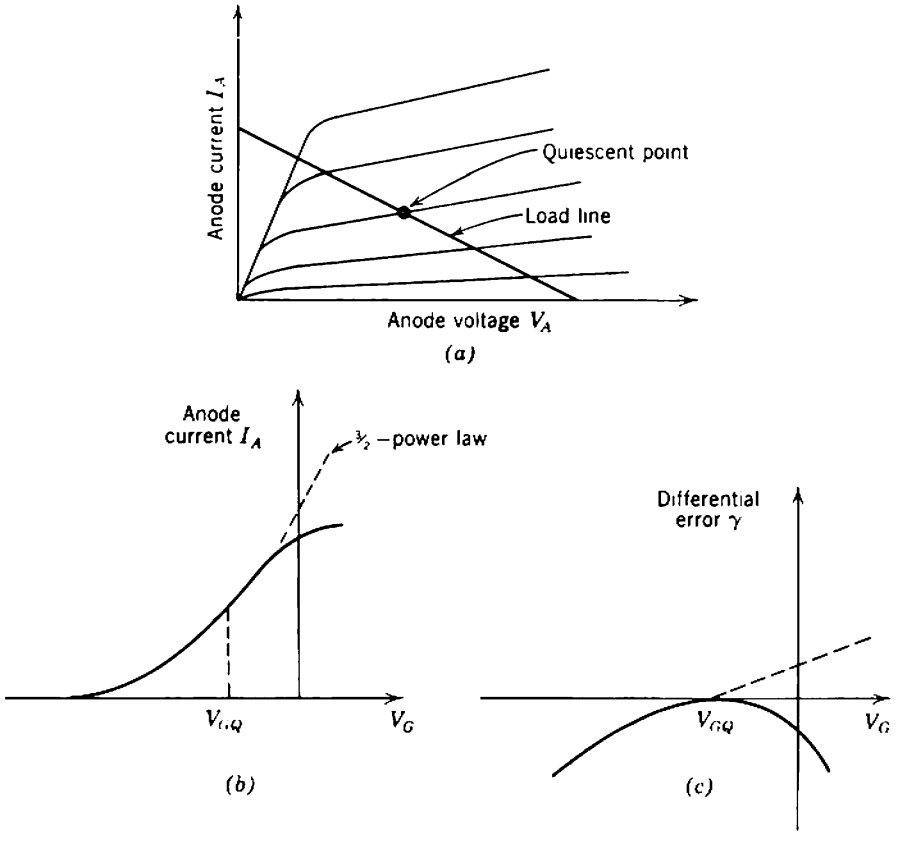


Fig. 16.13 Effect of gross nonlinearity on transfer characteristic: (*a*) output characteristics and load line; (*b*) transfer characteristic; (*c*) differential error.

general, the transfer function between source and load depends on source resistance R_S, input resistance R_i, and the transfer function TF_{amp} between input and output of the amplifier. In designable circuits, TF_{amp} is a relatively predictable transfer function, and the values of R_S and/or R_i are such that there is no significant uncertainty in the division of signal between them. This latter condition is not always easy to achieve in power amplifiers, particularly when an approximate voltage drive is required. Furthermore, it may well be desirable to trade some predictability in small-signal transfer function for an improvement in linearity. If the over-all nonlinearity is a function of both the amplifier and R_S, it is sometimes possible to use the nonlinearity associated with the signal division between R_S and R_i to compensate partially for some nonlinearity associated with TF_{amp}. One example of this technique is considered in Section 9.5.2, where it is shown that increasing R_S of a common-emitter stage decreases the differential error at the expense of increased uncertainty in the small-signal transfer function. Such possibilities emphasize the need for a fresh approach to the detailed design of large-signal amplifiers, an approach which emphasizes linearity of the transfer function as much as its predictability.

Information on the shape of an amplifier's transfer characteristic is usually obtained from manufacturers' graphical data. Sometimes these data need to be checked by reference to other sources or by experiment so that the designer is sure of both the general form of nonlinearity in the transfer characteristic and of the spread in this characteristic within devices of a given type number. If graphical data of acceptable accuracy are available, it is relatively easy to obtain graphical solutions for the various combinations of nonlinear equations (device characteristics) and linear equations (representing external circuit elements) which define the transfer characteristic.

Vacuum-tube data are almost always given as a set of average anode characteristics (Section 6.2.1), and the transfer characteristic of an amplifier stage can be obtained simply by superimposing an ac load line on them. Vacuum-tube data scarcely ever indicate the order of spread in the characteristics, so the designer must use experience or experimental information to make an estimate.

Transistor data are given in a variety of graphical forms. Usually, there is more than one graph per transistor because of the necessity of providing information on variation of the finite base current with operating conditions. Transistor data usually indicate the spread of the particular characteristics displayed; this may in part account for the common misconception that transistor characteristics are less reproducible than those of vacuum tubes.

The general approach to designing a low-distortion amplifier using a given device is to make a trial-and-error study of the effects of changing

circuit parameters. In particular, the slope of the load line is of critical importance. Although Eqs. 16.11 and 16.19 define optimum magnitudes of R_L for maximum signal power output, a small change about the optimum does not result in an excessive reduction of signal power and can lead to a very significant reduction of distortion. Tube manufacturers usually specify optimum values of load resistance which represent a reasonable compromise between power handling ability and nonlinearity. For transistors, resistance R_L in the collector circuit has little effect on nonlinearity.

Local feedback is another powerful tool for reducing nonlinearity and has the useful side effect of changing the input resistance to a more convenient value. Series feedback is particularly useful with transistors because it raises the input resistance and a closer approximation to voltage drive can be obtained. Shunt feedback (with either tubes or transistors) is relatively less useful. Notice that, when a feedback amplifier is overloaded and the output devices are driven beyond their clipping points, the loop gain falls to zero and no linearization is achieved. Consequently, the effect of a soft nonlinearity in device characteristics is sharpened considerably when feedback is applied to an amplifier: the static transfer characteristic of a feedback amplifier is very nearly a straight line joining the clipping points.

If careful choice of device and operating conditions and the use of local feedback do not reduce the static nonlinearity to an acceptable value, over-all feedback must be used in addition. With so many design variables involved, there is little wonder that the designers of very-low-distortion audio amplifiers achieve fame and often have standardized amplifiers named after them!

Dynamic nonlinearity can be visualized as changes in the positions of the high- and low-frequency singularities of the small-signal transfer function with signal level. It is usually desirable to ensure that the dominant singularities are fixed with fair accuracy by passive elements. This is achieved most conveniently at both low and high frequencies by using negative feedback and ensuring that one or more phantom zeros of the feedback factor have a controlling influence on dynamic response (Section 10.5.3.4).

16.3.1 Single-Ended Circuits

Vacuum-tube power amplifiers are invariably voltage driven and the initial choice of device is between triodes and pentodes. For the same anode dissipation, pentodes handle higher signal power and have higher gain at the expense of greater nonlinearity. However, the choice need not be a straightout decision in favor of a triode or a pentode because a compromise

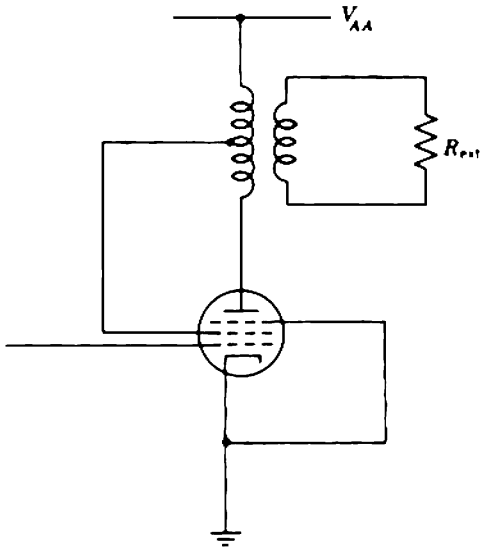


Fig. 16.14 Ultralinear connection of a pentode.

is possible; the so-called *ultralinear circuit** of Fig. 16.14 uses a pentode with its screen connected to a tap on the transformer primary. If the tapping point approaches the supply end of the transformer, pentode-like behavior results. If the tapping point approaches the anode end of the transformer, the tube behaves as a triode. The information contained in Table 16.2 is extracted from one manufacturer's information on the relative performance of a pair of tubes (operating under Class A conditions in a push-pull circuit) when the tubes are connected as triodes, ultralinear pentodes, and normal pentodes; the tapping points are specified as a percentage of the winding from the supply voltage end. Ultralinear pentodes are used extensively in low-distortion audio amplifiers.

Transistor power amplifiers can be voltage driven or current driven, the relevant transfer characteristics being of the form shown in Fig. 16.15. The voltage-driven transfer characteristic is markedly nonlinear, except at high currents when ohmic voltage drops in the base region are significant, but the spread of the characteristic is not excessive. The current-driven transfer characteristic is more linear but usually has a greater spread.

There are several methods of combining two or more devices (transistors particularly) so that their signal outputs add and increase the signal in the load or so that the input impedance is changed to a more convenient value.

* This circuit, along with many other outstanding contributions to electronics, is a product of the fertile genius of A. D. Blumlein; British patent 496,883 (1937).

Table 16.2 Comparison of Triode, Ultralinear, and Pentode Operation of Two Type EL34/6CA7 in Push Pull. (Courtesy Mullard Limited, London)

Mode of Operation	Available Power Output at Clipping	Distortion at 10 W Output
Triode	19 W	0.5%
Ultralinear (43% Tapping)	34 W	0.6%
Ultralinear (20% Tapping)	40 W	0.7%
Pentode	54 W	1.5%

Specific circuit requirements can usually be met by more than one of these combinational circuits, so the choice depends in part on the designer's personal preference. Some of the more useful techniques follow:

1. *Compound transistors* (Section 14.4.2) are produced in principle by interconnecting two or more transistors so that they are correctly biased and the resulting three-terminal combination has a very high current gain (approximately the product of β's for each of the component transistors). The two most useful circuits are the Darlington circuit and the synthetic *n-p-n* shown in Figs. 16.16*a* and 16.17*a*, respectively. Either of these compound transistors behaves as a high-gain *n-p-n* device and can be used as shown in Figs. 16.16*b* and 16.17*b* to realize a series-feedback stage of

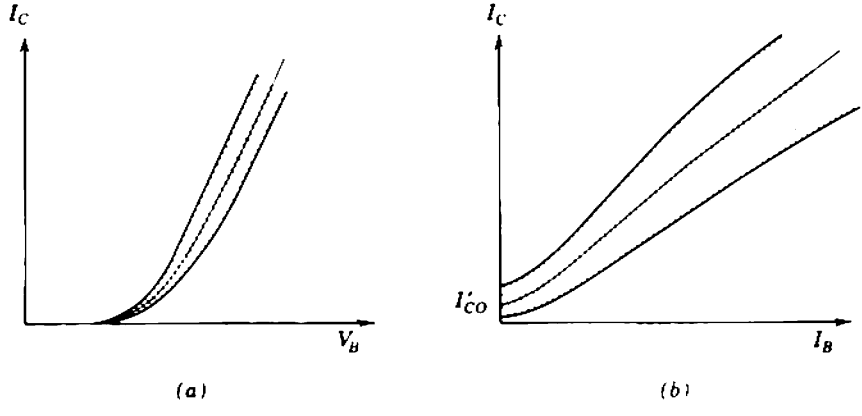


Fig. 16.15 Transfer characteristics for power transistors: (*a*) base voltage drive; (*b*) base current drive.

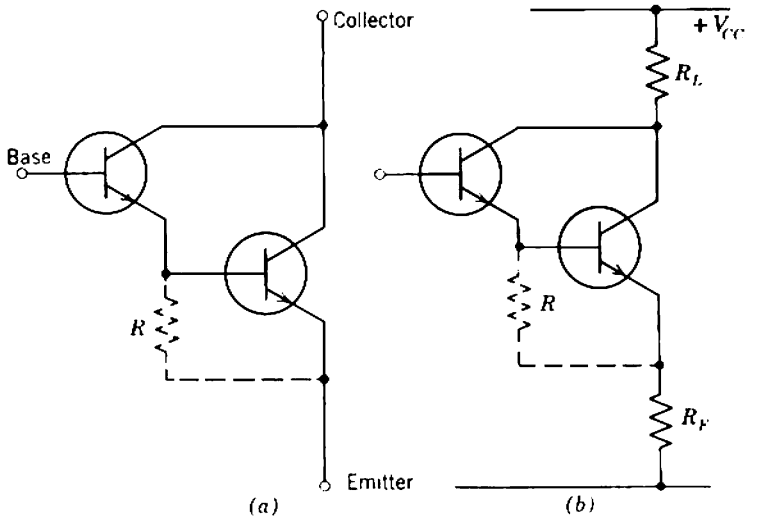


Fig. 16.16 Darlington compound *n-p-n* transistor: (*a*) basic circuit; (*b*) compound series-feedback stage.

high input resistance. A *p-n-p* Darlington circuit and a synthetic *p-n-p* are also possible. In practice the high-frequency response is much improved by including the resistors R (shown broken) whose function is to remove stored minority carrier charge from the base of the second transistor

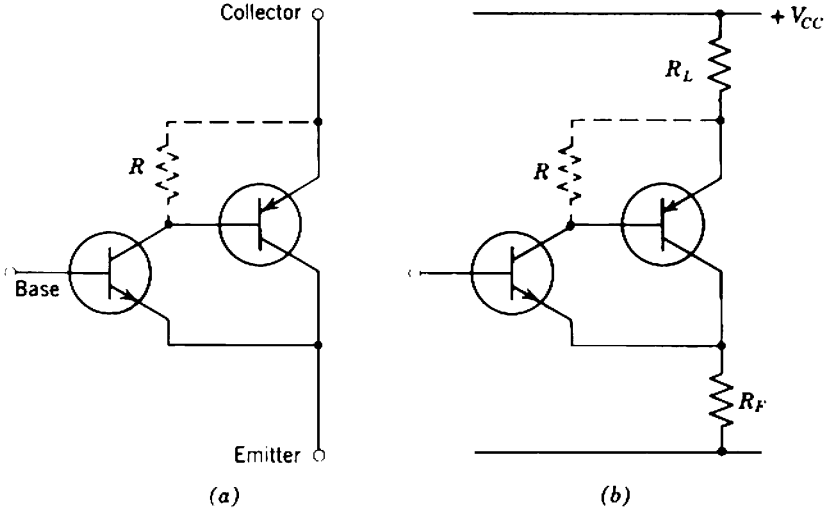


Fig. 16.17 Synthetic compound *n-p-n* transistor: (*a*) basic circuit; (*b*) compound series-feedback stage.

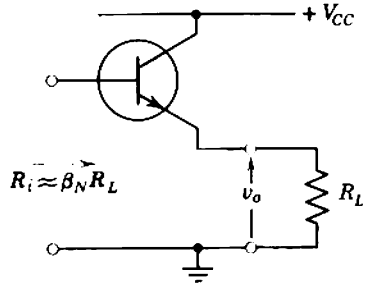


Fig. 16.18 Emitter-follower stage.

when the current in the compound transistor is instantaneously falling in magnitude.

2. *Emitter-follower stages* are not discussed with enthusiasm in this book because of their less than unity voltage gain, poor isolation of input from output, inflexible fixing of loop gain, and often unsatisfactory transient response. However, these stages do find application as output stages in audio amplifiers where the dynamic response requirements are not critical. In Fig. 16.18, the emitter-follower stage has a relatively large but unpredictable input resistance $\beta_N R_L$; it is much easier to develop the required output voltage across this larger resistance than across R_L itself.

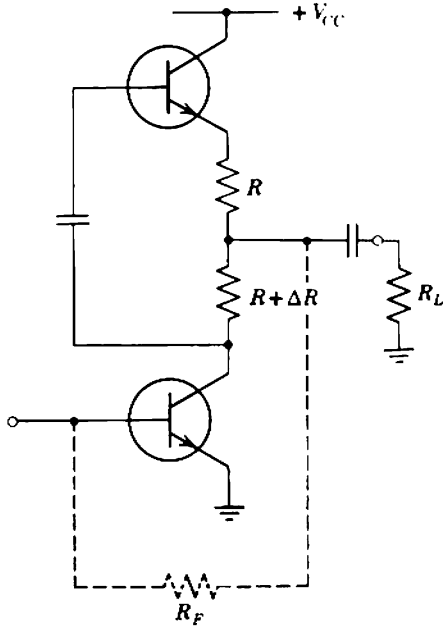


Fig. 16.19 Totem-pole amplifier.

3. *Totem-pole amplifiers* use a second transistor to double the possible output current swing. In the circuit of Fig. 16.19 the quiescent current flows through both transistors. Any change in the collector current of the lower transistor produces an equal and opposite change in the collector current of the upper transistor; thus a net signal current of twice that in either transistor flows into the load. Local shunt feedback can be employed to stabilize the transfer resistance. This circuit is especially useful when the dynamic response specification is critical.

4. *Bean-stalk amplifiers* consist of a number of transistors arranged in series so that their individual collector voltage ratings are not exceeded. The common-emitter input transistor Q_1 has a number of transistors placed in series between its collector and the load resistance (Fig. 16.20). The resistors R_0 to R_n ensure a predictable division of collector-to-emitter voltage, provided that the bleed current is considerably larger than the individual base currents. The bean-stalk technique enables large signal voltage swings to be obtained from transistors of low collector voltage ratings.

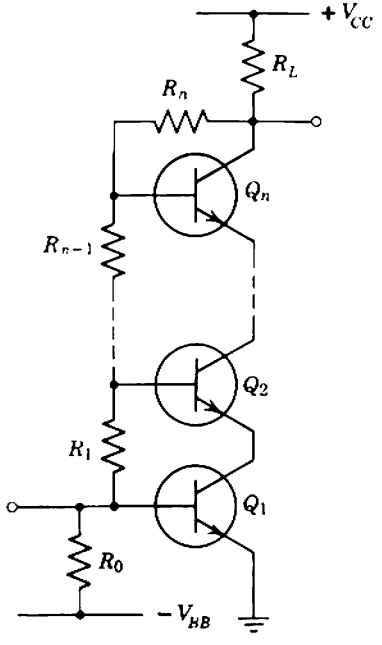


Fig. 16.20 Bean-stalk amplifier.

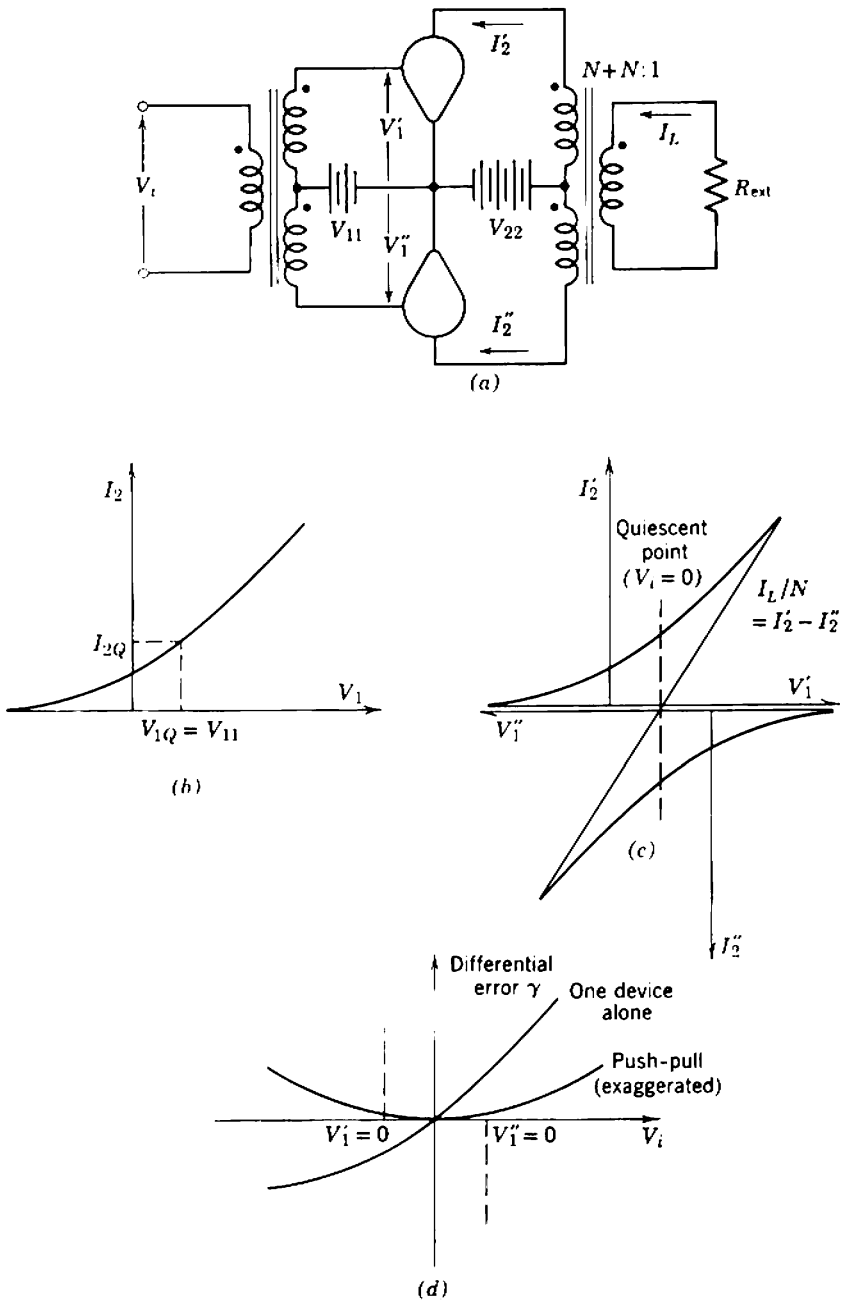


Fig. 16.21 Reduction of nonlinearity in a Class A amplifier by the use of push-pull: (*a*) circuit; (*b*) transfer characteristic of individual devices; (*c*) over-all transfer characteristic; (*d*) differential error.

16.3.2 Push-Pull Class A Amplifiers

Push-pull operation is mandatory with Class B amplifiers, in which each device contributes only one half of the output current waveform. In this context, push-pull operation is an extreme case of cancelling the nonlinearity of one device by the nonlinearity of another. Push-pull can also be used in Class A amplifiers to partially cancel the nonlinearity of two devices and, if an output transformer is used, to cancel polarizing dc fields in the core. Figure 16.21 shows how the nonlinearity of a single device is reduced when two identical devices are used in push-pull. The individual devices are biased at the same quiescent point (Fig. 16.21*b*), the input signal is applied in antiphase, and the output transformer is phased so that device output currents subtract. Thus the over-all transfer characteristic is obtained by summing the current ordinates with one of the device transfer characteristics inverted (Fig. 16.21*c*). It is obvious from the form of the characteristic (or the corresponding differential error, Fig. 16.21*d*) that no even-harmonic distortion exists, although the odd-harmonic distortion is unchanged. In a more practical situation, the two devices are not identical and finite even-harmonic distortion remains, but it is readily seen that a substantial reduction in differential error is achieved even with devices at the opposite limits of production spreads. This can be reduced further by the application of negative feedback. Indeed, it is quite common for some emitting-electrode resistance to be included as local series feedback for the individual devices to reduce the spread in their transfer characteristics. The highest quality audio amplifiers employ Class A push-pull circuits.

16.3.3 Class B Amplifiers

An ideal Class B amplifier can similarly be represented by Fig. 16.21*a*, but in this case the quiescent operating point for each device is set at the current cutoff point (Fig. 16.22*a*). Even with identical devices, the complete transfer characteristic is quite nonlinear in the vicinity of the cutoff or crossover points; this results in gross *crossover distortion*. Negative feedback can do little to improve the linearity, because the slope of the transfer characteristic (and hence the small-signal loop gain) falls to zero at cutoff. This inherent defect of Class B operation can be overcome, with little loss in power conversion efficiency, by applying a slight forward bias to each device as shown in Fig. 16.23. The mode of operation is not true Class B and, in falling between Class B and Class A operations, should perhaps be termed Class AB. It is customary, however, to refer to operation as being in Class B provided that the finite quiescent current is a small proportion of the peak current.

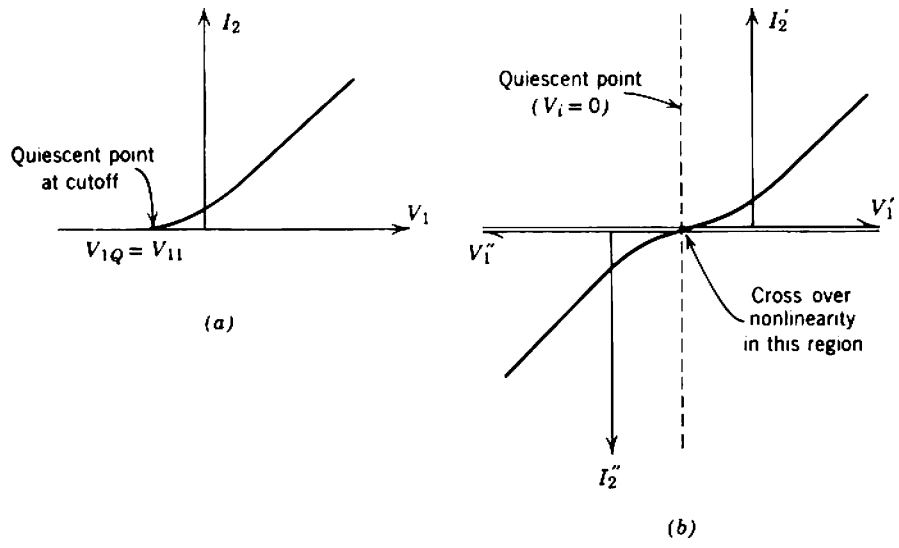


Fig. 16.22 Transfer characteristic of a Class B amplifier: (*a*) transfer characteristic of individual devices; (*b*) over-all transfer characteristic.

The other factor of critical importance in distortion performance is the mismatch between transfer characteristics of the two devices. Figure 16.24

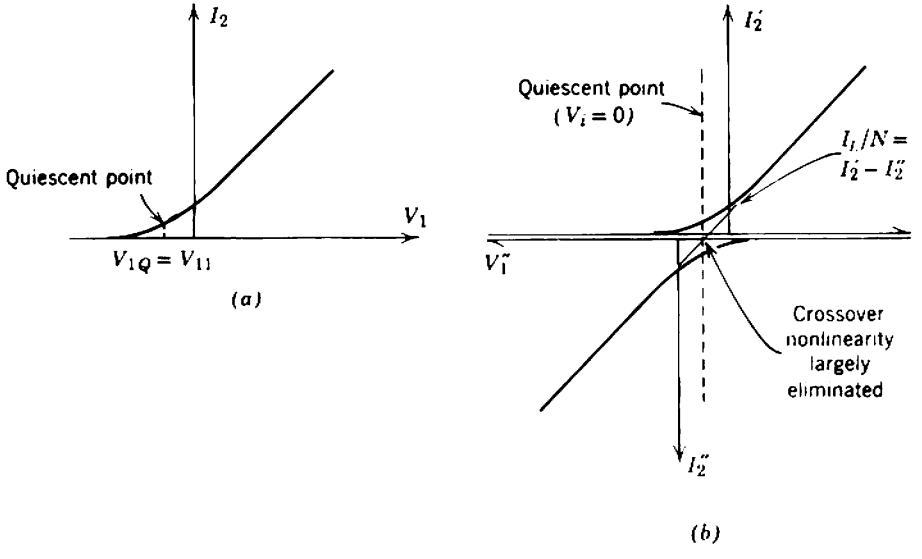


Fig. 16.23 Reduction of crossover distortion by small forward bias: (*a*) device characteristic; (*b*) over-all transfer characteristic.

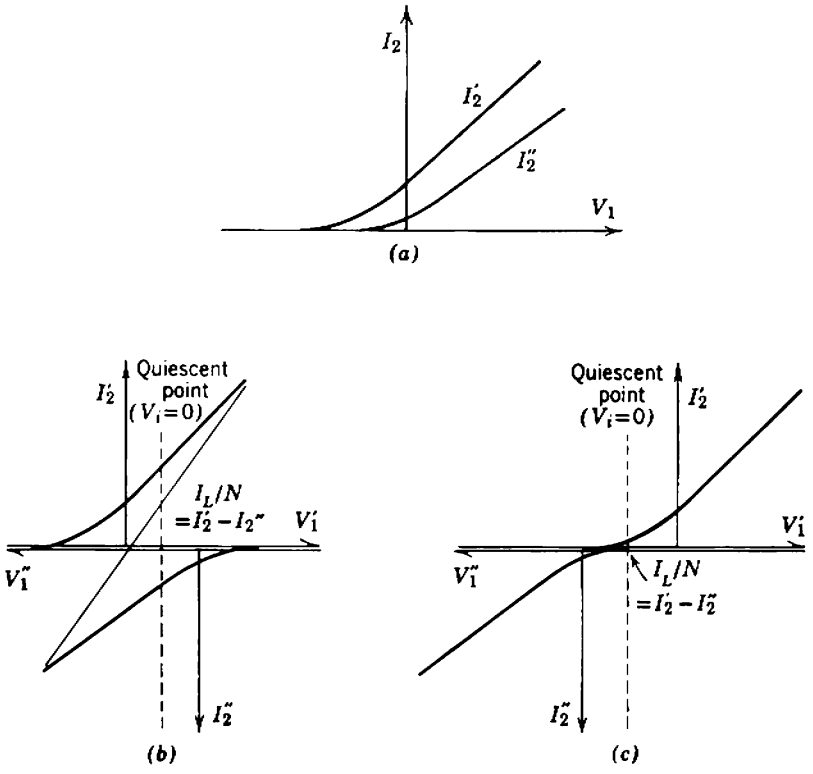


Fig. 16.24 Effects of device mismatch on over-all transfer characteristic: (*a*) individual transfer characteristics; (*b*) Class A; (*c*) Class B.

shows how mismatch produces greater nonlinearity in Class B operation than in Class A operation. Nonlinearity due to device mismatch can be reduced by using matched pairs of devices or, more desirably, by local feedback; compound devices can be used within the local feedback loop.

The signal currents in individual halves of Class AB or B stages are very rich in harmonics and the various phenomena which limit high-frequency performance attenuate the higher-order harmonics relative to the fundamental. Thus the individual device currents are deformed asymmetrically; this results in high-frequency distortion of the load current as shown in Fig. 16.25. With transistors the high-frequency response is dominated by the nonlinear capacitors c_B, c_{tE}, and c_{tC} (which represent the charge requirements of the conducting channel and emitter and collector depletion layers, respectively). The calculation of the current transient for any given voltage or current driving waveform can be based on a linearized

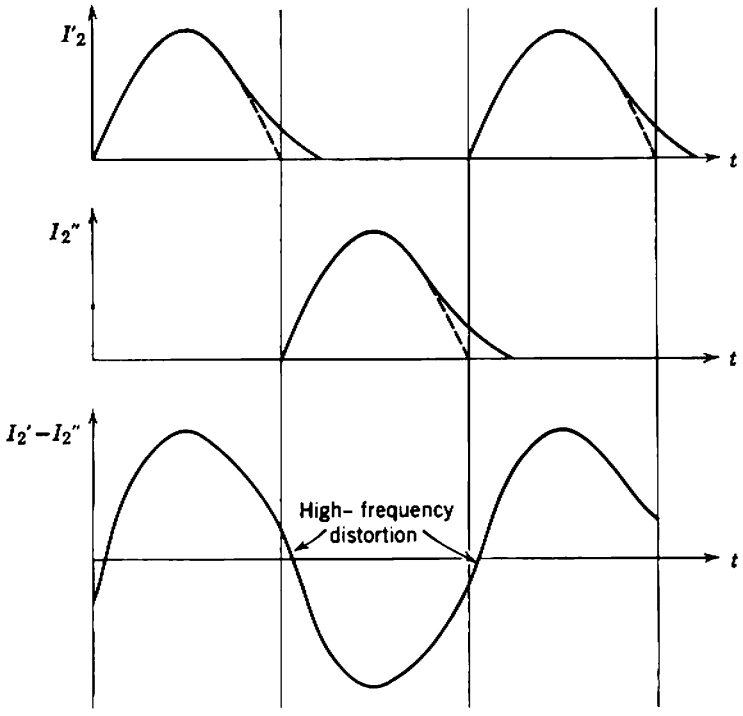


Fig. 16.25 High-frequency distortion in a Class B amplifier.

application of charge-control theory to transistor switching circuits.* The details are straightforward but beyond the scope of this book. Physically, the relatively slow turn-off of collector current arises from the finite time required to remove all of the minority carrier charge from the base. Other factors being equal, this charge is removed more quickly if a voltage (low-impedance) drive is used which is capable of forcing the base-emitter junction into reverse bias. A less effective arrangement is to connect a small bleed resistor between the base and emitter of each transistor, as in Figs. 16.16 and 16.17.

16.3.4 Some Push-Pull Amplifier Configurations

Sections 16.3.2 and 16.3.3 suggest that an effective method for reducing nonlinearity in large-signal amplifiers is to use devices in push-pull. There are several useful push-pull configurations and a selection of these is discussed below. For simplicity, the main concern in this section is the

* R. BEAUFOY, "Transistor switching circuit design using the charge-control parameters," *Proc. Inst. Elec. Engrs. (London)*, **106B**, 1085, Suppl. 17, 1959.

topology and general character of the transfer characteristics for various configurations. Several assumptions are made in order to clarify the main argument:

1. Transistor realizations are assumed because of the greater generality and flexibility which result from the existence of *p-n-p* and *n-p-n* types. Each single transistor shown in the circuit diagrams can, in fact, be synthesized by compounding two or more actual transistors and can incorporate local feedback loops; all such complications are omitted from this section.
2. The details of driving circuits are omitted and it is assumed that appropriate single-ended or push-pull signal sources are available. Driver stages are discussed in Section 16.5.
3. Some details of biasing circuits are omitted. Although collector voltage supplies are included, base supplies are not shown. Details of some approaches to biasing circuit design are given in Section 16.4.
4. The output signal is ac coupled to the load to avoid the usually undesirable consequences of heavy quiescent currents flowing through the load. The more specialized case of dc-coupled loads is considered in Section 16.7.
5. The input signal is *RC* coupled to the transistor bases. This simple form of feed is suitable for Class A operation, but cannot normally be used with transistor Class AB or B output stages because the base current has an average component which depends on signal level. In the latter cases it is necessary to modify the detail of the circuit, but not the topology.
6. In practical amplifiers over-all feedback is usually applied from the output to the driver or some preceding stage; all such feedback loops are omitted in this section.

The circuits described in the following paragraphs do not represent a complete survey of possibilities but should serve as a useful introduction. Quite often circuits of particular topologies can be developed from quite differing starting points. This obvious fact has led to much argument!

The circuit of Fig. 16.26*a* is the basic common-emitting-electrode push-pull amplifier which can also use pentodes or triodes. A push-pull input is required and the output signals from the two devices are combined in the correct phase by means of a center-tapped transformer. The transformer can also be used for impedance conversion if the actual load resistor R_{ext} has an inconvenient magnitude. If R_{ext} has a suitable magnitude and it is not necessary to ground one side of the load, the transformer secondary winding need not be used and R_{ext} can be inserted directly between the two collectors so that inductor coupling (rather than transformer coupling) is used. The current gain and voltage gain can both exceed unity. Because

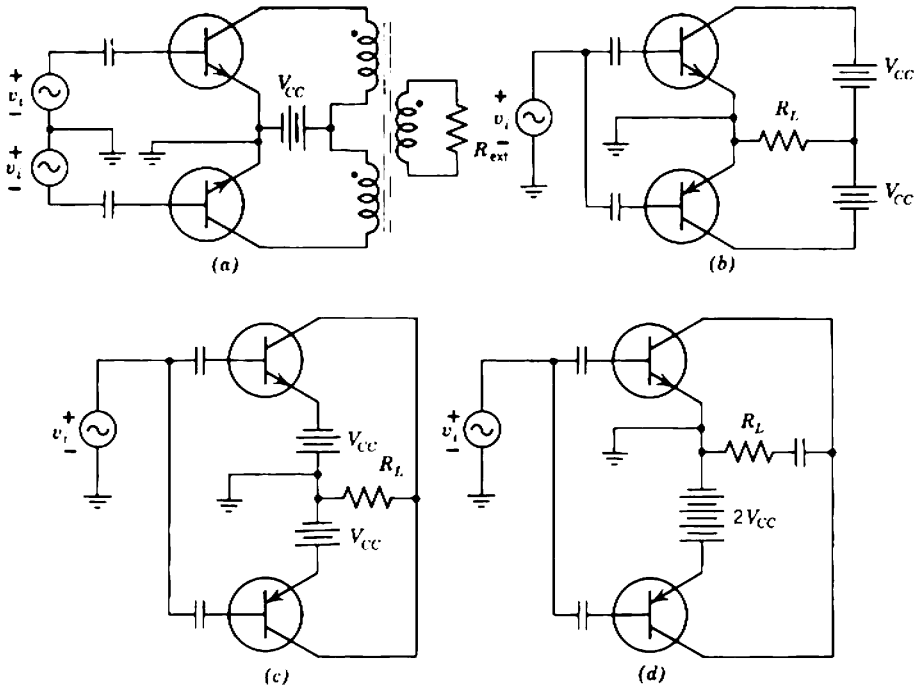


Fig. 16.26 Common-emitter push-pull amplifiers: (*a*) transformer coupling; (*b*) complementary symmetry, two floating power supplies; (*c*) complementary symmetry, two power supplies; (*d*) complementary symmetry, single power supply.

of the transformer, the peak collector voltage of either transistor rises to twice the supply voltage V_{CC}.

When transistors are used, the circuit can be simplified by the use of *complementary symmetry* to the transformerless circuit of Fig. 16.26*b*. A single-ended input is required and the output signals are combined in the correct phase in the load resistor R_L. Notice that a dc blocking capacitor is not required if equal quiescent currents flow in the two devices. The circuit of Fig. 16.26*b* has the practical disadvantage of requiring two floating power supplies (i.e., power supplies that do not have one terminal at ground potential); this undesirable arrangement can be modified to the more practical circuits of Figs. 16.26*c* or *d*. Both the current gain and voltage gain of all circuits can exceed unity. Because there is no output transformer, the peak collector voltage on either transistor is equal to the *total* supply voltage $2V_{CC}$, so that there is no advantage over Fig. 16.26*a*. Notice that R_L (in the sense of Figs. 16.26*b* to *d*) is the reflected value of R_{ext} looking into either one of the transformer primary windings in Fig. 16.26*a*; the resistance between collectors is $4R_L$.

Figure 16.27*a* represents a push-pull amplifier that differs from Fig. 16.26*a* only in the placing of the reflected load in series with the emitter rather than the collector. The individual devices act as common-collecting-electrode amplifiers which have a maximum voltage gain slightly less than unity. The arrangement is generally not convenient for vacuum-tube output stages because of limitations in the heater-to-cathode voltage rating. It is interesting to note that some quite successful vacuum-tube amplifiers combine the arrangements of Fig. 16.26*a* and 16.27*a* by using a transformer with four primary windings to couple some of the load into the anode circuit and some into the cathode circuit. When transistors are used as common-collector output stages, it is convenient to employ the complementary symmetry arrangement shown in the ideal circuit of Fig. 16.27*b*, or

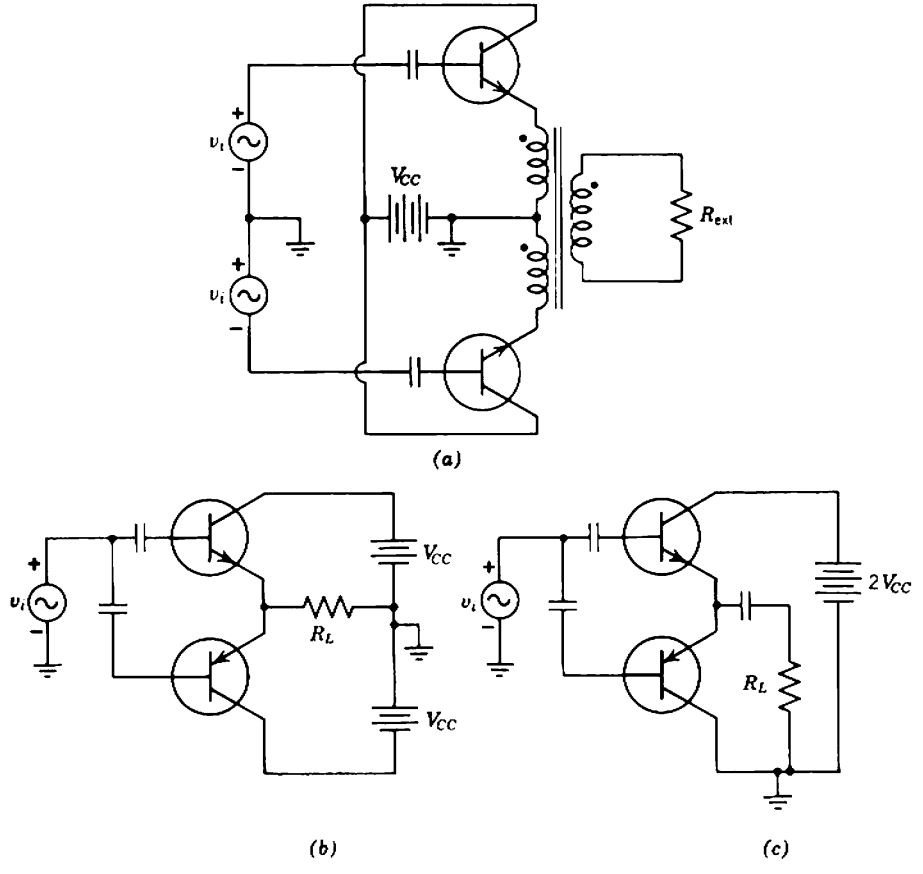


Fig. 16.27 Common-collector push-pull amplifiers: (*a*) transformer coupling; (*b*) complementary symmetry, two power supplies; (*c*) complementary symmetry, single power supply.

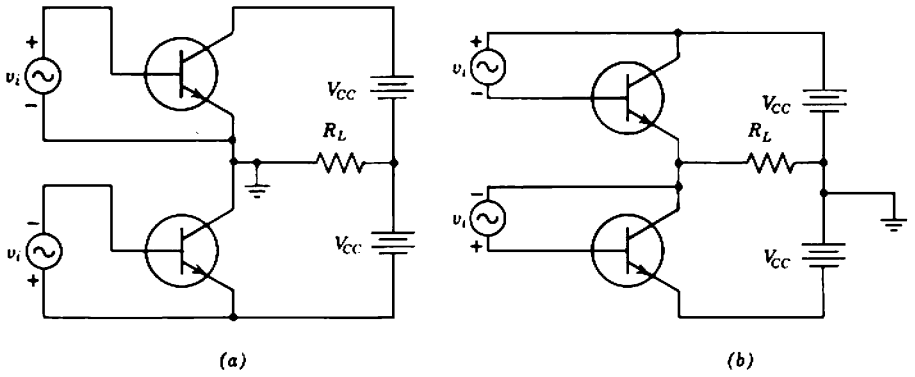


Fig. 16.28 Single-ended push-pull circuits: (*a*) common emitter; (*b*) common collector.

the more practical circuit of Fig. 16.27*c*. Similar advantages accrue to those for common-emitter stages.

Noncomplementary devices can also be connected to give a push-pull output without an output transformer, although an input transformer or equivalent circuitry is required. Figures 16.28*a* and *b* show two possible *single-ended push-pull* circuits, with the common electrode being the emitter and collector, respectively. These circuits require a push-pull driving signal. Their output portions can be rearranged readily to use a single power supply with one side grounded.

The final circuit (Fig. 16.29) is a push-pull version of the totem-pole

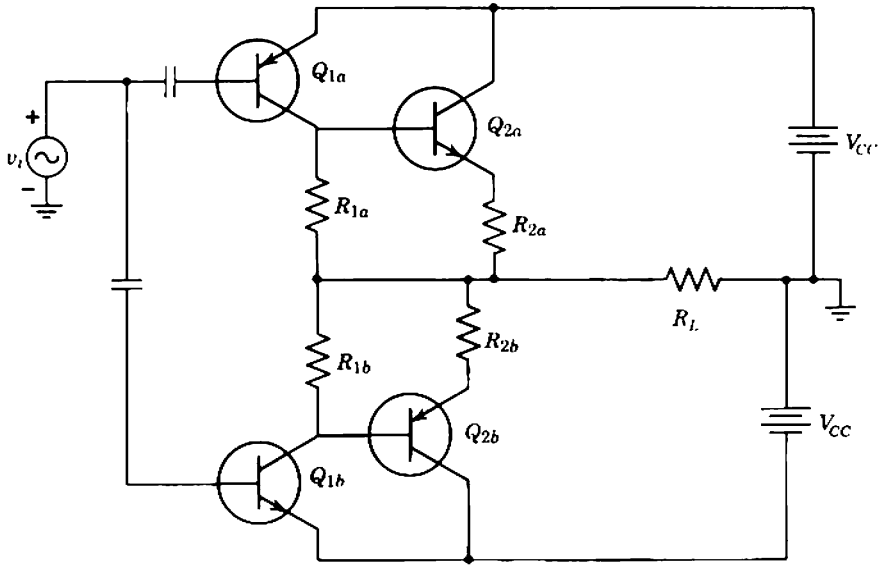


Fig. 16.29 Push-pull folded totem-pole amplifier.

amplifier that employs complementary devices; each half of the circuit is a "folded" totem-pole amplifier. Unlike the single-ended totem-pole amplifier (Fig. 16.19), the input transistor Q_1 and the output transistor Q_2 need not share the same quiescent current and a useful current gain (largely determined by resistors R_1 and R_2) can be realized between Q_1 and Q_2.

16.3.5 Dynamic Range with a Capacitive Load

In general, a voltage-output transfer function of an amplifier can be written as

$$A_V(s) = -Y_T(s) \times Z_L(s). \tag{16.26}$$

If the output resistance of the amplifier is infinite, Y_T in Eq. 16.26 can be replaced by g_m, and the poles of A_V become the poles of Z_L, where Z_L is the parallel combination of load resistance R_L with load capacitance C_L. When the output resistance is finite, however, the pole of Z_L is canceled by a zero of Y_T; the load capacitance appears in the expression for A_V as a new pole of Y_T, and this pole lies beyond $s = -1/R_LC_L$. As an example, Section 7.5.1.1 shows that the pole of a triode amplifier stage is at

$$s = -\frac{r_A + R_L}{r_A R_L C_L}.$$

Outstanding examples of circuits whose poles lie well beyond the poles of the load impedance are the shunt-feedback stage and the common-collecting-electrode stage.

Equation 16.26 can be rearranged as

$$Y_T(s) = -\frac{A_V(s)}{Z_L(s)},$$

from which it follows that Y_T has a zero at the pole of Z_L. Because the pole of Z_L lies closer to the origin than the pole of A_V (unless the output resistance is infinite), it follows that $|Y_T(j\omega)|$ must rise at moderately high frequencies. This mathematical result has a common-sense interpretation. Consider an amplifier driving a capacitive load with a sinusoidal voltage. As the frequency is raised, the magnitude of the load impedance falls, and a larger output current from the amplifier is required to maintain the load voltage.

Because of the current required by the load capacitance, the high-frequency dynamic range of an amplifier is less than the range at mid-band frequencies predicted from the clipping points. The clipping points define the available output current swing, and the available output voltage at any frequency is proportional to the magnitude of the load impedance at that frequency. (For a shunt-feedback stage the total load includes the feedback

impedance.) An adequate approximation to the symmetrical peak-to-peak output current swing available from most amplifiers is $2I_{2Q}$; this value is set by cutoff of the active device.

It is instructive to calculate the largest input voltage that can be applied to an amplifier before dynamic overload occurs, that is, before the collecting-electrode current falls instantaneously to zero. Therefore we need to calculate the input voltage for which the signal output current $i_o(t)$ satisfies

$$|\hat{i}_o| = I_{2Q}. \tag{16.27}$$

For simplicity suppose that the voltage gain has a single pole at $s = -\omega_{co}$. The signal output current in response to a signal input voltage $v_i(s)$ is then

$$i_o(s) = v_i(s)\left(\frac{A_{vm}}{R_L}\right)\left(\frac{1 + sR_LC_L}{1 + s/\omega_{co}}\right). \tag{16.28}$$

1. TIME RESPONSE. If $v_i(t)$ is a negative step $-\hat{V}_i$, Eq. 16.28 becomes

$$i_o(s) = -\frac{\hat{V}_i A_{vm}}{R_L}\left[\frac{1}{s(1 + s/\omega_{co})} + \frac{R_LC_L}{1 + s/\omega_{co}}\right].$$

The corresponding time function is

$$i_o(t) = -\frac{\hat{V}_i A_{vm}}{R_L}[1 + (\omega_{co}R_LC_L - 1)\exp(-\omega_{co}t)]$$

whose maximum magnitude (when $t = 0$) is

$$|\hat{i}_o| = \hat{V}_i A_{vm}\omega_{co}C_L.$$

Figure 16.30 shows the corresponding collecting-electrode current waveform. If dynamic overload is not to occur, the amplitude of the input step must satisfy

$$\hat{V}_i \leq \frac{I_{2Q}}{A_{vm}\omega_{co}C_L}. \tag{16.29a}$$

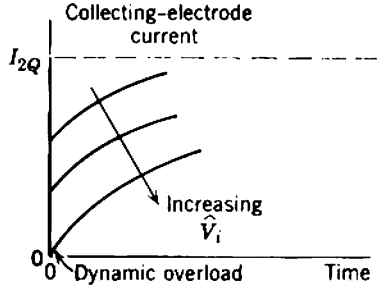


Fig. 16.30 Dynamic overload on a step waveform.

For a shunt-feedback stage the corresponding limit on input current is

$$\hat{I}_i \leq \frac{I_{2Q}}{R_T\omega_{co}(C_F + C_L)} \approx \frac{I_{2Q}C_F}{C_F + C_L}. \tag{16.29b}$$

If the drive to an amplifier is increased beyond the limit suggested by Eqs. 16.29, a deterioration in the turn-off edge can be observed readily.

2. FREQUENCY RESPONSE. If $V_i(t)$ is a sinusoid of peak amplitude $\hat{V}_i$, it follows from Eq. 16.28 that the peak amplitude of the output signal current is

$$\hat{i}_o = \frac{\hat{V}_i A_{Vm}}{R_L}\left[\frac{1 + (\omega R_L C_L)^2}{1 + (\omega/\omega_{co})^2}\right]^{1/2}.$$

As the frequency increases toward infinity, $\hat{i}_o$ approaches its maximum value $\hat{V}_i A_{Vm}\omega_{co}C_L$. Hence dynamic overload will never occur if

$$\hat{V}_i \leq \frac{I_{2Q}}{A_{Vm}\omega_{co}C_L}, \tag{16.30}$$

which is numerically the same result as Eq. 16.29.

Equations 16.29 and 16.30 suggest only the obvious expedient of increasing the quiescent current to increase the dynamic range. Often this is not acceptable. A preferable alternative for increasing the dynamic range into a capacitive load is to use two output devices in push-pull; one device turns on irrespective of the direction in which the output voltage is moving, and this device can supply the peak current required by the load capacitance.

16.4 BIASING CIRCUITS

Design of the biasing circuits is important for any amplifier and particularly so for the output stage of a large-signal amplifier. Ideally, the quiescent point of a Class A output stage should lie mid-way between the clipping points of its transfer characteristic; if the quiescent point shifts, the available signal output swing is reduced. In Class B amplifiers the quiescent point is close to cutoff; a shift results either in gross crossover distortion or in reduced power conversion efficiency. The essential requirement of a biasing circuit is that it should hold the quiescent current close to its intended value, despite reasonable changes in the parameters of a device or its environment, and despite the presence of signals in the amplifier.

16.4.1 Vacuum-Tube Output Stages

Class A vacuum-tube output stages very often use some variation of the self biasing circuit described in Section 6.1.2.2. The cathode resistor may

be bypassed (Fig. 16.31*a*) or partly or totally unbypassed if some series feedback is desired (Figs. 16.31*b* and *c*). When the resistor is completely bypassed, the pole due to the bypass capacitor lies in the vicinity of

$$s \approx -\frac{g_m}{C_K}.$$

Because g_m is unlikely to exceed a few tens of milliamperes per volt (even at the peak cathode current), there is seldom a problem in obtaining an electrolytic capacitor of sufficiently large value to locate the pole at any desired position.

When greater stability is required, simple self biasing can be replaced by cathode-feedback biasing (Section 6.1.2) or any of the two-stage feedback circuits (Section 6.5.3). Figure 14.43 shows a practical example.

Push-pull Class A vacuum-tube stages can also use self biasing and one obvious circuit uses separate cathode components for each device as shown in Fig. 16.32*a*. This circuit is not economic in components, but has the virtue of approximately equalizing the quiescent currents for two dissimilar devices. A simpler arrangement is to use a single resistor as shown in Fig. 16.32*b*; this eliminates three components, but does not equalize the quiescent currents in the two tubes unless a preset adjustment is included.

The essential difficulty in biasing Class B (and also Class AB) stages is that the average and quiescent currents are not equal. The dc voltage developed across a bypassed cathode resistor is proportional to the average current, which equals the quiescent current only for the special case of Class A operation. Thus, shunt capacitors cannot be used to separate quiescent and signal components of currents in Class B amplifiers. One possible approach is to return the grids to a supply voltage of about the

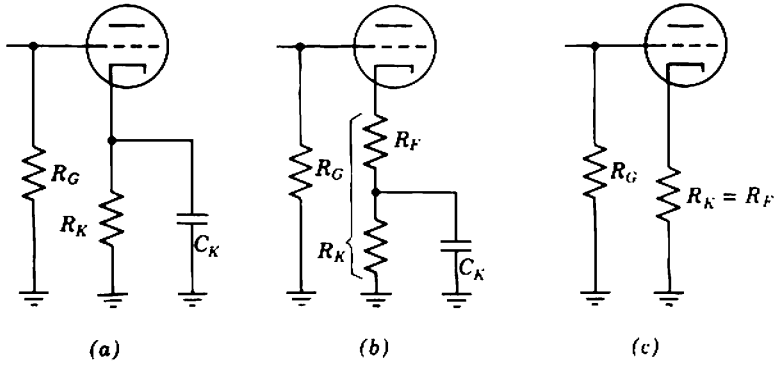


Fig. 16.31 Self-biased single-ended vacuum-tube Class A stages.

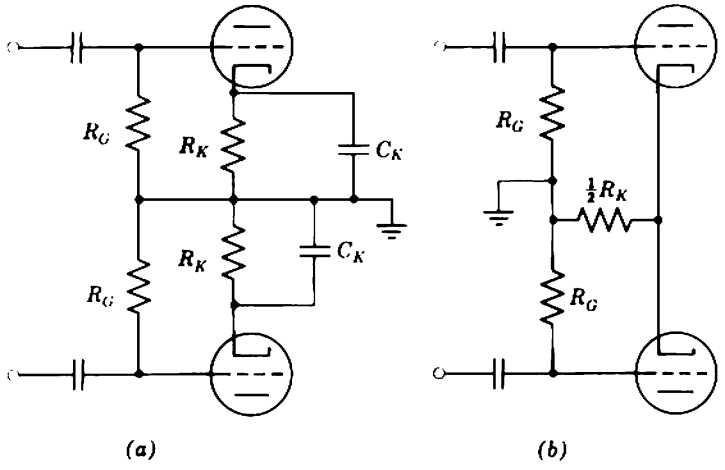


Fig. 16.32 Self-biased push-pull vacuum-tube Class A stages.

cut-off value. A small value unbypassed cathode resistor for each tube assists in equalizing the quiescent currents.

Vacuum-tube Class B amplifiers can be subdivided into two categories:

(i) Class B_1 in which the grid voltage is always negative and no significant grid current flows;

(ii) Class B_2 in which the grid is driven positive in order to increase the available peak anode current and grid current flows for part of each half-cycle.

In Class B_1 the driving signals can be capacitively coupled to the grids of the output tubes because their input resistance is virtually infinite. In Class B_2, however, the input resistance is low and nonlinear (like a transistor); it is normal to use transformer coupling between the driver and output stages. Class B_2 amplifiers often employ high-μ triodes (such as type 6N7) which are designed to operate predominantly in the positive-grid region and are biased at zero quiescent grid voltage. Because the positive-grid anode characteristics are like those of a pentode, yet there is no power loss due to screen dissipation, very high conversion efficiencies can be attained.

16.4.2 Transistor Output Stages

The design of biasing circuits for transistor output stages is complicated by three factors:

1. The finite base current necessitates a low-resistance supply network if the quiescent base voltage is to be independent of current (hence of signal).

2. If a bypassed emitter resistor is to be used, the very high mutual conductance of a transistor calls for a huge bypass capacitor. It is not good practice to remove all resistance from the emitter circuit; biasing stability is improved significantly if the quiescent voltage drop across R_E exceeds about 250 mV.

3. The temperature dependence of V_{BE} may require some temperature-sensitive compensating elements in the biasing circuit to prevent runaway.

Figure 16.33 shows two basic arrangements for biasing transistor output stages which operate in any of Classes A, AB, or B. The base voltage is supplied from a low-resistance divider which incorporates a nonlinear element having a negative temperature coefficient of resistance. The temperature-dependent element is a forward-biased diode in Fig. 16.33*a* and a thermistor in Fig. 16.33*b*; in either case, the magnitudes assigned to the other components ensure that base voltage varies with ambient temperature in step with V_{BE}, and so keeps the emitter current constant. Because of the tolerances on transistors and nonlinear elements, one of the resistors may need a preset adjustment; this is clearly undesirable for designable circuits. An additional disadvantage of these and most other compensating circuits is that they respond to changes of ambient temperature or, at best, heat-sink temperature, and not to the true junction temperature of the transistor. Thus, unless the emitter resistor is large, there is limited direct sensing of the changes in junction temperature and it is difficult to ensure thermal stability in high-power, high-voltage applications.

A more designable circuit which provides greater thermal stability than the simple compensating schemes is shown in Fig. 16.34*a*. Resistor R_E is chosen sufficiently large to define the quiescent current with reasonable

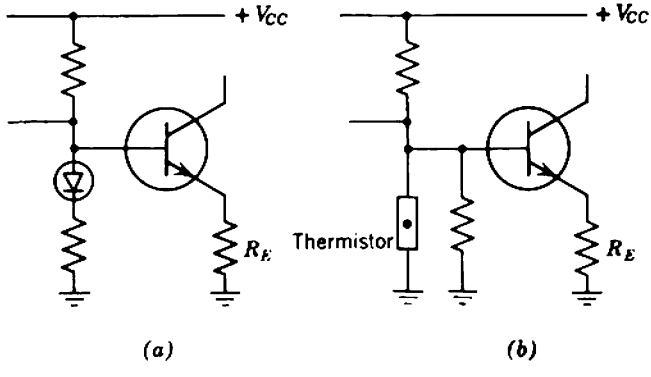


Fig. 16.33 Temperature-compensated base voltage supply networks.

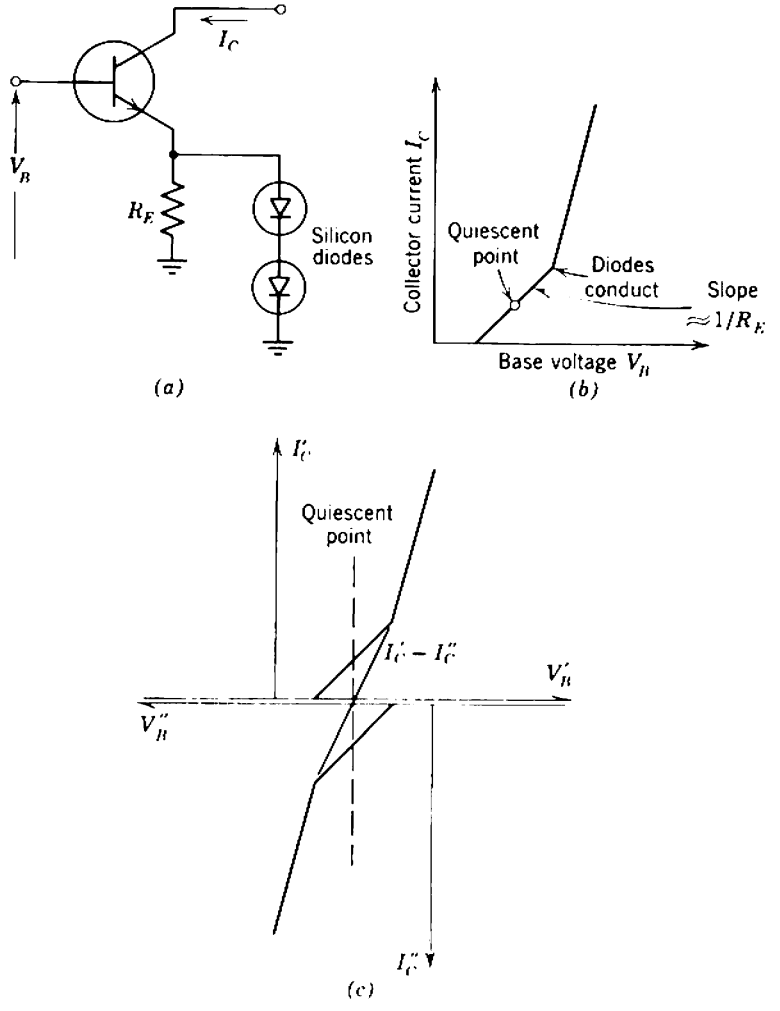


Fig. 16.34 Nonlinear emitter-feedback biasing network: (*a*) circuit; (*b*) transfer characteristic; (*c*) composite transfer characteristic for a push-pull Class B circuit.

accuracy provided that the silicon diodes do not conduct. In practice the quiescent emitter-to-ground voltage can be about 0.5 V for each silicon diode used. When the collector current increases from its quiescent value, the emitter-to-ground voltage increases and the diodes conduct to give a lower dynamic resistance from emitter to ground. The transfer characteristic thus has the nonlinear form shown in Fig. 16.34*b*; the price of better biasing is increased nonlinearity. This nonlinearity is reduced substantially in a push-pull Class B amplifier (Fig. 16.34*c*), but the over-all transfer characteristic still exhibits appreciable discontinuities at the conduction

points of the diodes. Notice, however, that the slope of the transfer characteristic never falls to zero as it does in a simple Class B circuit (Fig. 16.22*a*), so feedback can be used to reduce the nonlinearity.

With any emitter-feedback biasing circuit, the dc impedance of the base supply should be low, yet the ac impedance should be high to avoid loss of signal current. These conflicting requirements can be met readily by using a coupling transformer (Fig. 16.35*a*). However, because of the difficulty and cost of realizing wide-band transformers, this solution is seldom practicable and *RC* coupling is preferred. Loss of signal current down the low-resistance shunt path can be reduced by bootstrapping (Fig. 16.35*b*.)

Class AB and B stages require a base network which can supply large peak currents. Biasing network impedances which are satisfactorily low as far as quiescent base current requirements are concerned may not be low enough to satisfy the peak current demands. In amplifiers which are used for speech, music, or other signals having a "burst" character, it is permissible to use a relatively high-resistance base supply (such as a diode or thermistor network) if some capacitor in the circuit can provide the peak current; there will be a slight shift in the quiescent point following a signal burst until the excess charge accumulated on the capacitor leaks away. Examples are the bypass capacitor shown broken in Fig. 16.35*a*, and the coupling capacitor of an *RC* coupling network. The same technique is applicable to

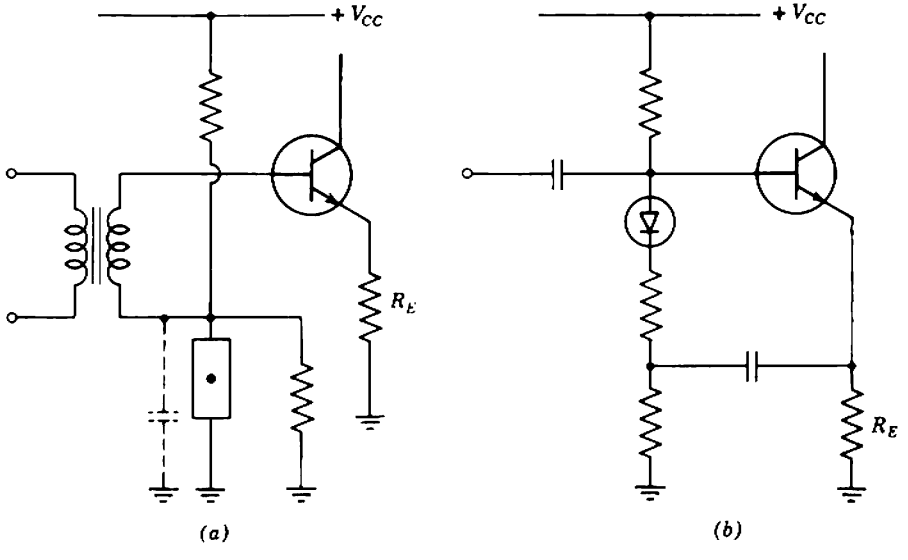


Fig. 16.35 Methods for achieving low dc base supply resistance without excessive loss of signal current: (*a*) transformer coupling; (*b*) bootstrapping.

amplifiers designed for continuous power output, provided that the excess charge can be removed from the capacitor on the half-cycles for which the transistor does not conduct. One approach is shown in Fig. 16.36. Both the transistor and the diode network half-wave rectify the drive but, if the network matches the nonlinear input resistance of the transistor on negative half-cycles, there is no net charge deposited on the capacitor.

16.5 DRIVING AND PHASE-SPLITTING CIRCUITS

The penultimate or *driver stage* of a large-signal amplifier is required to deliver a (substantially) undistorted signal of appropriate magnitude to the output stage, despite nonlinearities in the latter's input impedance. In addition, the driver stage may be required to

(i) provide biasing voltages and currents to the input electrode of the output stage,
(ii) provide identical two-phase signals for a push-pull output stage,
(iii) have satisfactory recovery characteristics after an overload.

Often the disposition of driver and output stages in a circuit is quite obvious. With many transistors amplifiers employing direct-coupled or compound devices, however, the decision on which devices are grouped as the output stage and which constitute the driver is somewhat arbitrary.

Phase-splitting stages provide the two-phase signals required by some push-pull amplifiers and can operate at sufficiently large signal levels to be considered as driver stages (Fig. 16.37*a*) or can operate at lower signal levels and be followed by a pair of driver stages (Fig. 16.37*b*).

Except for the special case of Class B_2 operation, the load on driver stages for vacuum-tube amplifiers is dominated by passive elements and is substantially linear. In contrast, the load on driver stages for transistor

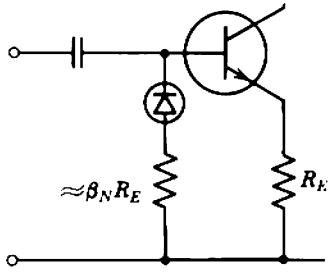


Fig. 16.36 Network for preventing dynamic shift of quiescent point in Class AB or B transistor amplifiers.

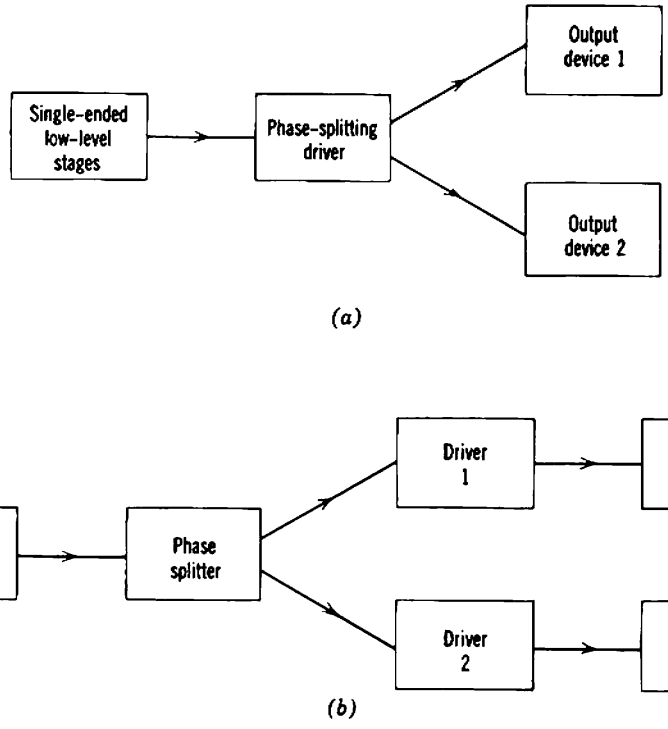


Fig. 16.37 Block diagrams for amplifiers with push-pull output: (*a*) high-level phase splitter; (*b*) low-level phase splitter.

amplifiers can be substantially nonlinear because of the nonlinear input characteristic of the output transistors and the nonlinearities associated with diodes and/or thermistors in the biasing circuits. At least three methods are used to reduce the over-all nonlinearity that results from nonlinear loading on the driver. In order of increasing sophistication, these are the following:

1. Choose the collector supply resistor for the driver so that it masks the nonlinear input resistance of the output stage. A small resistor is required for voltage drive and a large resistor for current drive. The former arrangement is not very satisfactory because the driver stage must be operated at a large quiescent current in order to achieve the desired output voltage swing.
2. Apply local feedback to the driver to raise or lower its output resistance as appropriate for current or voltage drive to the output stage.
3. Design the driver and output stages as a feedback pair.

Control of the driver output impedance may also be desirable to improve the high-frequency response. Large, high-power transistors tend to have

relatively low β cutoff frequencies, and poor high-frequency performance can result unless the driver approximates to a voltage source.

Phase splitters are essential for push-pull vacuum-tube amplifiers but they can be avoided in transistor amplifiers by using appropriate arrangements of *n-p-n* and *p-n-p* devices. The simplest phase splitter is a transformer which has a center-tapped secondary winding. Transformers simplify biasing circuit design, but are expensive and have poor frequency response. The three purely electronic phase splitters shown in Fig. 16.38 use tubes, but transistor versions are also possible. The *concertina phase splitter* (Fig. 16.38*a*) has identical voltage gains close to unity from grid to both anode and cathode, but the output impedance at the cathode is low whereas that at the anode is high. This circuit has the considerable advantage of allowing the grid to be coupled directly to the anode of a preceding

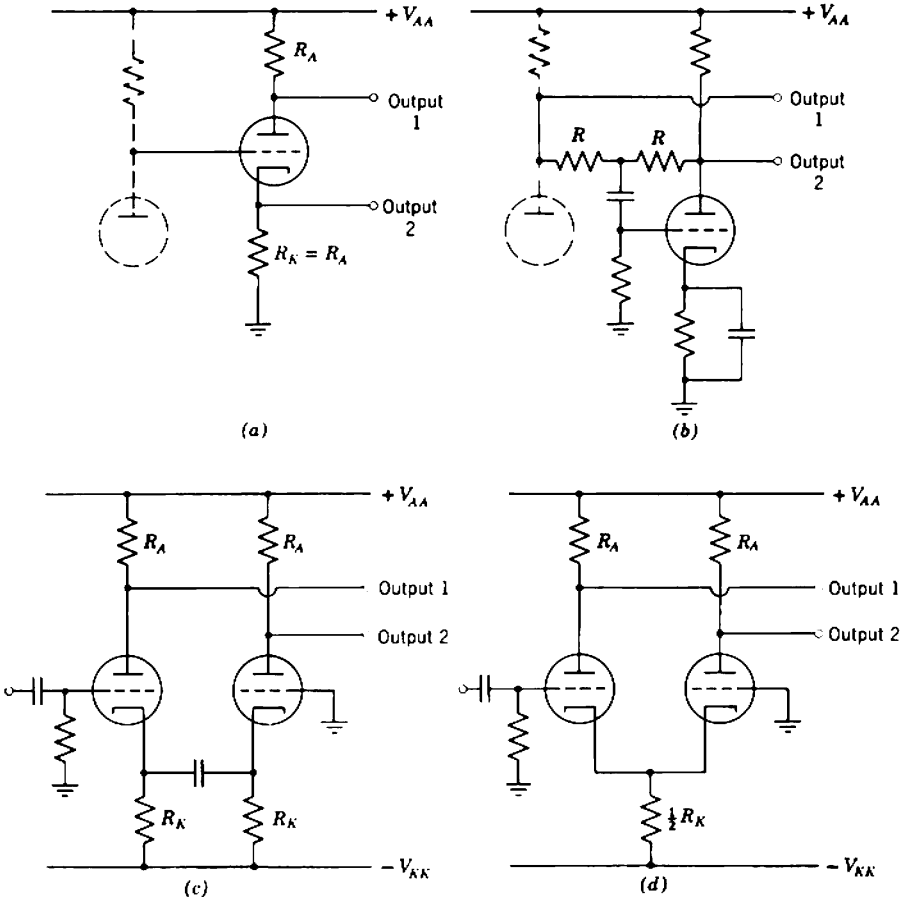


Fig. 16.38 Phase splitters: (*a*) concertina; (*b*) seesaw; (*c*) ac-coupled long-tailed pair; (*d*) dc-coupled long-tailed pair.

tube, but the designer must ensure that the heater-cathode voltage rating is not exceeded. The *seesaw phase splitter* of Fig. 16.38*b* also has unity gain but the cathode is only a few volts above ground. This circuit is more complex than the concertina arrangement and has the disadvantage of different frequency response to the two outputs. The *long-tailed pair phase splitter* of Figs. 16.38*c* and *d* provides a gain well in excess of unity and produces output signals at the same impedance level. The circuit shown in Fig. 16.38*c* has separate cathode resistors to ensure equality of the cathode currents; the alternative arrangement of Fig. 16.38*d* is better balanced at low frequencies. It should be pointed out that the relatively large gain of the long-tailed pair is achieved with two tubes. A fair comparison of the relative gains of the three phase splitters should be based on the use of an additional voltage amplifier stage (shown broken) for Figs. 16.38*a* and *b*. Assuming identical tubes, the first two circuits give about twice the gain of the third. Nevertheless, the long-tailed pair is probably the most useful phase-splitting circuit for most applications, because the balance is better at high frequencies.

16.6 ILLUSTRATIVE PRACTICAL DESIGNS

The preceding sections outline the problems (often conflicting problems) which can occur in large-signal amplifiers and show some useful solutions. Even though serious attempts are made to present only the more useful techniques and to develop some system in presenting these ideas, the reader may well be confused by the very descriptive treatment of a large catalog of potentially useful ideas. This emphasizes the point made in the introduction to the chapter that the design of large-signal amplifiers requires far more knowledge, ability, and experience than are required for most other types of amplifier. Certainly this knowledge, ability and experience can come only from performing a number of actual designs. However, it is instructive first to look critically at the decisions made and circuit techniques used in extant designs. Two large-signal transistor amplifiers are described in the following sections; a vacuum-tube amplifier is described in Section 14.5.1.

16.6.1 A 5-Watt Class A Amplifier

The amplifier shown in Fig. 16.39 uses alloyed-junction germanium transistors in a simple single-ended voltage-feedback pair. The second "transistor" is compounded from Q_2(OC81) and Q_3(AD140), the compound being modified by 270-Ω and 150-Ω resistors in series with Q_2 to define the collector current and limit its value under overload conditions.

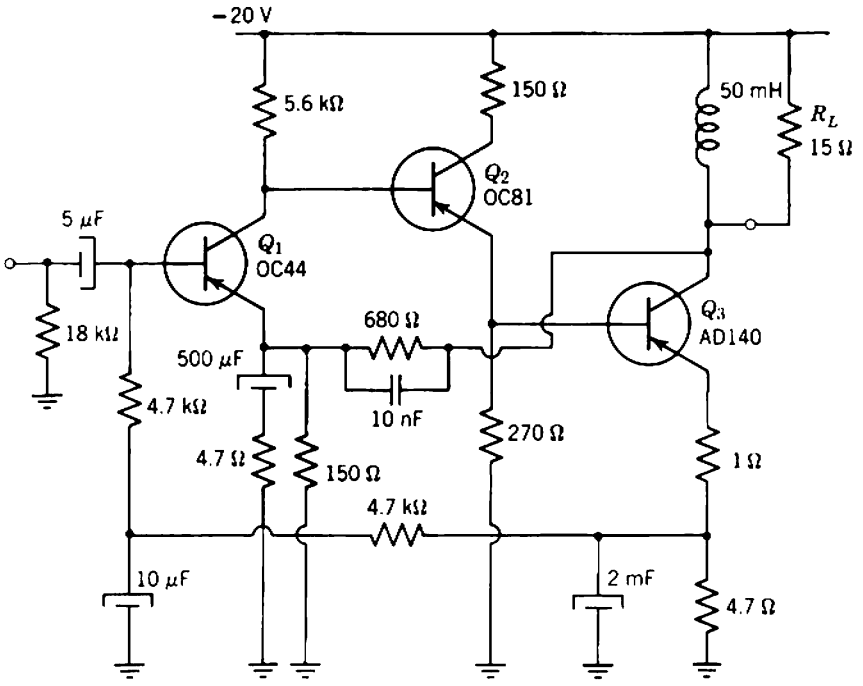


Fig. 16.39 A 5-W Class A audio power amplifier.

For a signal power of 5 W into the 15-Ω load the rms and peak values of load current are 580 and 820 mA, respectively. The corresponding values of load voltage are 8.7 and 12.3 V. Thus the quiescent current in Q_3 should be about 1 A and the quiescent collector to emitter voltage should be about 13 V. Inductance coupling is used to increase the conversion efficiency.

Separate feedback paths are used for signal and biasing; signal feedback is taken from the collector of Q_3 to the emitter of Q_1, whereas biasing feedback is taken from the emitter of Q_3 to the base of Q_1. With the −20-V supply shown, the collector-to-emitter voltage for Q_3 is −14 V; the remaining 6 V appears across resistors external to the emitter. Transistor dissipation is about 12 W. Maximum permissible junction temperature for type AD140 is 90°C, and the thermal resistance from junction to heat sink is 1.7°C/W. If the maximum expected ambient temperature is 40°C, the allowable thermal resistance from junction to ambient is

$$\theta \leq 4.2°\text{C/W},$$

and the heat sink must have a thermal resistance of

$$\theta_{sa} \leq 2.5°\text{C/W}$$

which can be realized with a blackened heat sink of surface area about

100 in.^{2} (Fig. 4.21). Runaway is no problem because of the heavy biasing feedback.

The manufacturers of type AD140 provide minimal information on nonlinearity by merely specifying the ratio

$$\frac{\beta_N \text{ at } 1 \text{ A}, 0 \text{ V}}{\beta_N \text{ at } 100 \text{ mA}, -14 \text{ V}} = 0.5.$$

From this figure and from experience, it can be assumed that a 2:1 change in β_N might be expected over the operating conditions envisaged. In the absence of any local feedback r_B is likely to be the dominant term in input resistance for Q_3. Consequently, the operation approximates to current feed, and β_N is the most significant source of nonlinearity. Thus, in order to bring the differential error to about 5% (corresponding to 0.5% total harmonic distortion) the total loop gain around Q_3 must be about -10.

Local feedback for Q_3 is provided by the 1-Ω resistor which adds to the comparable emitter spreading resistance to give a stage voltage gain of about 6 and an input resistance of about 100 Ω. Q_2 has almost unity voltage gain and an input resistance of about 7 kΩ; the latter is not sufficiently high for its contribution to the load resistance of Q_1 to be neglected. Thus, the voltage gain of the first stage (including local series feedback provided by the 4.7-Ω resistor) is

$$G_{T1}R_{L1} \approx 60 \text{ mA/V} \times 3 \text{ k}\Omega = 180$$

and the over-all forward voltage gain is approximately 1100. The loop gain has the satisfactory value of

$$|A_l| = \mu\beta = 7.5$$

and the closed-loop gain is

$$A_V = \frac{R_{F1} + R_{F2}}{R_{F1}} \left(\frac{1}{1 + 1/|A_l|} \right) = 118.$$

For full output, the rms input voltage required (the *voltage sensitivity*) is 66 mV. The input resistance is dominated by the 4.7-kΩ resistor. Measurements confirm that the total harmonic distortion lies at about 0.6% over most of the power output range and approaches 1% only at full power.

For the modest values of loop gain employed, the dynamic response is dominated by the high-frequency poles of Q_2 and Q_3. To eliminate overshoot and to increase the passband to beyond 20 kHz, a 10-nF capacitor is connected across the 680-Ω signal feedback resistor to provide a phantom zero at -23 kHz.

16.6.2 A 10-Watt Class B Amplifier

The preceding circuit is simple and conventional but is uneconomical in power-supply requirements and in the use of components. The circuit of

Fig. 16.40 is an attempt to reduce power losses by employing Class B operation and to minimize the number of passive components (particularly electrolytic capacitors) by the use of a single over-all feedback loop with minimal local feedback. These complications increase the design effort required but, when quantity production is envisaged, may lead to the most economical realization.

Complementary symmetry is employed for the Class B compound output transistors Q_4Q_5 and Q_6Q_7. A suitable value of forward bias on Q_4 and Q_6 is obtained from the voltage drop across the diodes in the collector circuit of Q_3, these diodes being thermally bonded to Q_5 and Q_7. The nominal collector current is 50 mA; over a temperature range of 0 to 50°C, a mains voltage variation of $\pm 10\%$, and a range of transistors, the actual collector current varies from 90 to 200% of this nominal value. The output voltage from the collectors of Q_5 and Q_7 swings over the range ± 18.0 V before clipping occurs; this corresponds to a maximum power output of 11.0 W into the 15-Ω load at a conversion efficiency of 70%. Peak collector current is 1.2 A. Maximum instantaneous power dissipation per device is 5.8 W, maximum average dissipation is about 2 W; this power can be dissipated readily with a heat sink of moderate size. The circuit is thermally stable and will not run away.

Because the output transistors Q_5 and Q_7 are driven from current sources, the dominant nonlinearity results from relatively large mismatches in the values of β_N for the germanium 2N2147 and the silicon AY8103. To

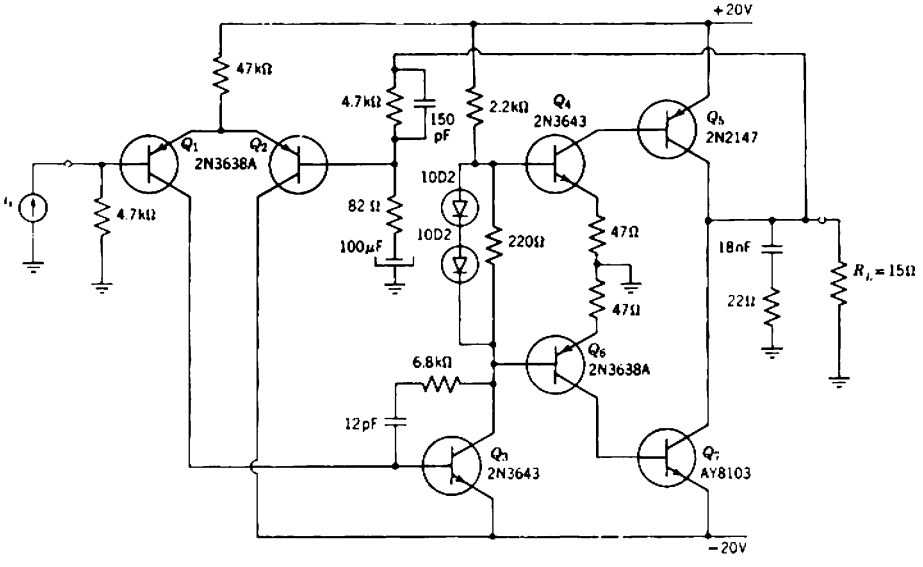


Fig. 16.40 A 10-W complementary Class B audio power amplifier.

minimize the effects of mismatch and avoid the problems of individually matching devices the nominal mid-band loop gain is 100. Measurements yield the following values of total harmonic distortion at 1 kHz:

$$< 0.2\% \text{ at } 10 \text{ W},$$
$$0.07\% \text{ at } 7.5 \text{ W},$$
$$0.05\% \text{ at } 5 \text{ W}.$$

Notice that the 100-μF blocking capacitor increases the dc loop gain to about 10,000 and thus ensures that the quiescent output voltage is close to zero.

The forward voltage gain is given approximately by

$$\mu_m = G_T(Q_1, Q_2) \times \beta_N(Q_3) \times (2.2\text{k}\Omega) \times G_T(Q_4, Q_6) \times \beta_N(Q_5, Q_7) \times R_L.$$

Because this depends on two β_N factors, the likely tolerance is large. Similarly, at least two singularities of $\mu(s)$ can occupy a wide range of positions. Consideration of the extreme conditions shows that a dynamic response without excessive overshoot can be achieved by shaping the forward-path frequency response with local feedback around Q_3 and employing a phantom zero in the feedback path at 106 kHz; this singularity defines the closed-loop bandwidth. In addition, the high-frequency transfer admittance from the base of Q_2 to the collector of Q_1 is controlled by the 4.7-kΩ base resistor. In order to prevent the resistor from being short circuited by the source, it is necessary to drive the amplifier from a current source. This is an example of the technique outlined in Section 7.2.3 by which a current source is converted to what is essentially a voltage drive by a low-value shunt input resistor. The input signal current required for 10-W output is 23 μA and the input resistance is obviously 4.7 kΩ.

16.7 SPECIAL PROBLEMS OF DC AMPLIFIERS

The circuit techniques of Sections 16.3 to 16.5 are discussed in the context of ac amplifiers for which the useful response need not extend to zero frequency. Capacitors and inductors are used for separating signal and bias currents and voltages. In applications such as driving dc machines, indicating instruments or recorders, and in the important special case of regulated power supplies the above-mentioned circuits are unsuitable because appreciable dc power must be supplied to the load. Similar constraints are imposed on the design of integrated-circuit power amplifiers because large capacitors cannot be realized.

The long-tailed pair is extremely useful in low-level dc amplifiers because it combines accurate fixing of the quiescent conditions with minimal signal degeneration. Dc offset voltages between output and input of an

amplifier can be eliminated by circuit techniques such as those illustrated for single-ended stages in Fig. 15.17. For high-level stages, however, power losses in the tail resistor or current source are excessive. Single-ended output stages are therefore used in conjunction with long-tailed pairs for the low-level stages, and over-all feedback is applied to reduce the output drift. If sufficient gain is included in front of the drift-prone single-ended stages, the drift generators at the input of the amplifier can be reduced to those of the long-tailed pair input stage; the proof is similar to that in Section 8.4.2 for noise in a multistage amplifier. Like noise generators at the input, drift generators are not changed by negative feedback, although the gain and hence the output drift are reduced.

16.7.1 Transition Circuits

Transition from long-tailed pairs to single-ended stages may take place at a point in the circuit at which the signal level is high and drift is of no concern or at a point at which the signal is relatively small. In the latter case the mere utilization of one output from the long-tailed pair may result in excessive drift due to changes in the supply voltage. Figure 16.41 shows two transition circuits that are insensitive to supply-voltage fluctuations.

In Fig. 16.41*a** Q_3 is a shunt-feedback stage of transfer resistance R. The signal voltage at point X in response to i_{o1} is

$$v'_X = +Ri_{o1},$$

and the response to i_{o2} is

$$v''_X \approx 0$$

because the output resistance of a shunt-feedback stage is small. Over-all, the signal voltage at point Y is

$$v_Y = R(i_{o1} - i_{o2}), \tag{16.31}$$

which depends only on the differential output of the long-tailed pair. Changes in V_{CC} are highly attenuated at the output because of the low output resistance of Q_3 as a shunt feedback stage:

$$\frac{\Delta V_Y}{\Delta V_{CC}} = \frac{R_o(Q_3)}{R_C + R_o(Q_3)} \approx \frac{1}{1 + \beta_{N3}R_C/R}. \tag{16.32}$$

As the circuit stands, the output is near ground potential and the bases of Q_1 and Q_2 are several volts negative; the circuit can, of course, be translated to any desired dc level.

* This arrangement appears first to have been used in an integrated circuit application. R. J. WIDLAR, "A monolithic operational amplifier," *Fairchild Application Bulletin*, April 1964.

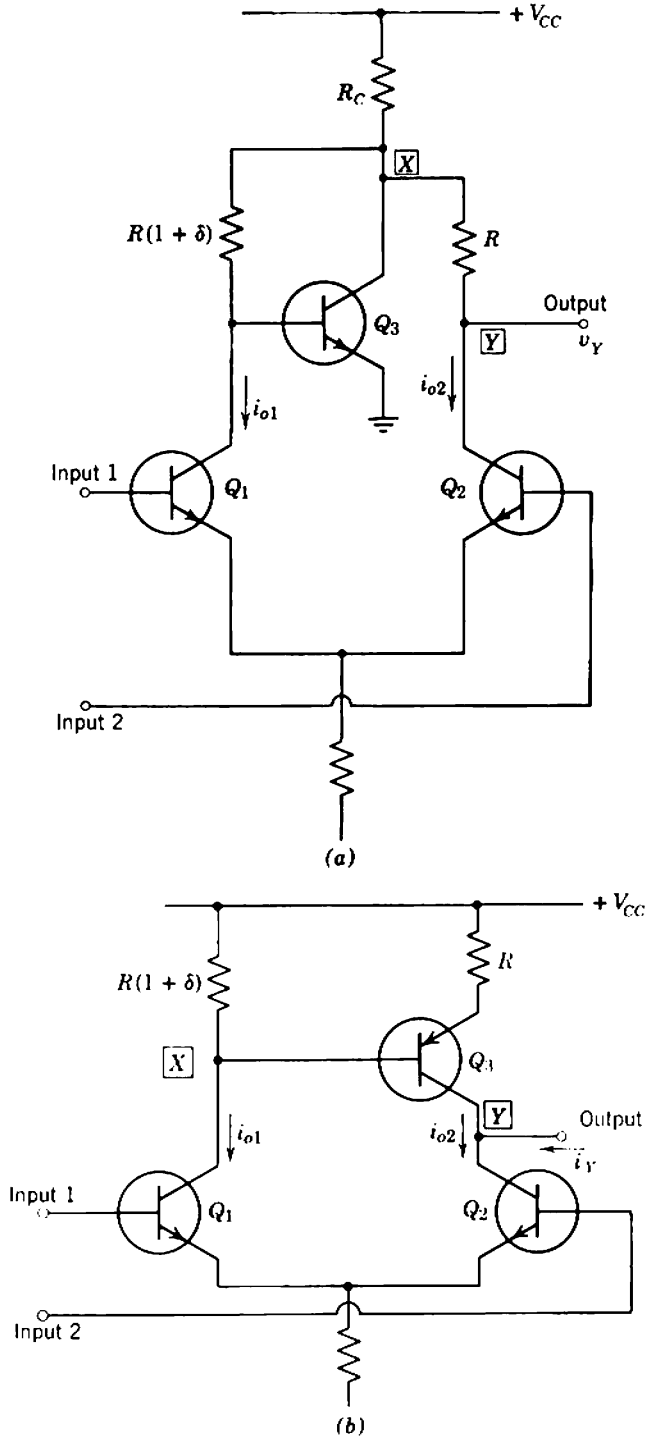


Fig. 16.41 Transition circuits from long-tailed pairs to single-ended stages: (*a*) voltage output; (*b*) current output.

Figure 16.41*b** is a dual circuit that uses a series-feedback stage to produce a current output that depends only on the differential input:

$$i_Y = -(i_{o1} - i_{o2}). \tag{16.33}$$

16.7.2 Illustrative Example

Figure 16.42 is a simple dc power amplifier that uses a long-tailed pair input stage, a totem-pole output stage, and over-all negative feedback. It can deliver rather more than 1 W of sinewave power to a 20-Ω load.

The quiescent current in the output stage is set at about 250 mA by the 2-V drop across three silicon diodes and the 2.7- and 4.7-Ω resistors. The ratio of these resistors is chosen so that there is a signal current gain of two between the collector of Q_3 and the load. Under these circumstances the mid-band forward voltage gain is

$$\mu = G_T(Q_1, Q_2) \times 2\beta_N(Q_3) \times R_L \approx 150.$$

The feedback factor is set by the 100-Ω and 2.2-kΩ resistors:

$$\beta = -\frac{1}{23}.$$

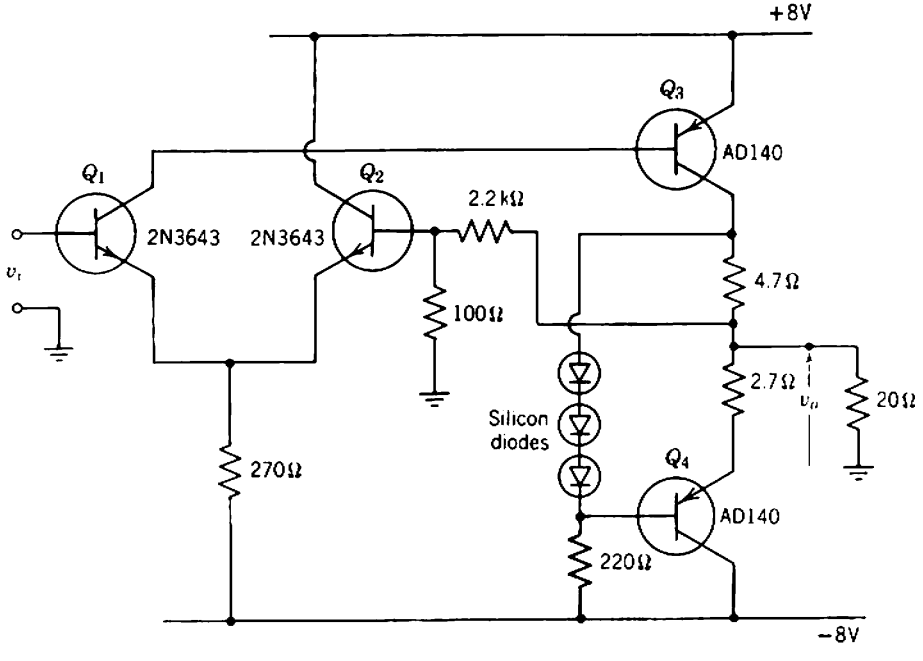


Fig. 16.42 A 1-W dc power amplifier.

* Suggested to the authors by W. H. BERRYMAN.

Hence the loop gain is

$$A_l = \mu\beta = -6.5$$

and the closed-loop voltage gain is

$$A_V = -\frac{1}{\beta}\left(\frac{1}{1 - 1/A_l}\right) = 20.$$

16.8 REGULATED POWER SUPPLIES

All preceding sections of this book assume the existence of stable dc supply voltages which remain reasonably constant despite variations in the current being drawn from them and changes in the mains voltage. For dc amplifiers the stability requirements are very stringent indeed: accuracies better than 1 part in 10^4 are often sought. Hence an understanding of *voltage-regulated power supplies* (or, more simply, *voltage regulators*) is extremely important. Circuit techniques for voltage regulators can be reapplied with little change to *current regulators*; this is left as an exercise for the reader.

The basic aim in designing a voltage regulator is to produce a dc power amplifier (as shown in Fig. 16.43) in which the regulated load voltage V_L depends only on the magnitude of a precise reference voltage V_{ref}; the large-signal voltage gain

$$\bar{A}_V = \frac{V_L}{V_{ref}} \tag{16.34}$$

should be stable and independent of changes in the unregulated supply

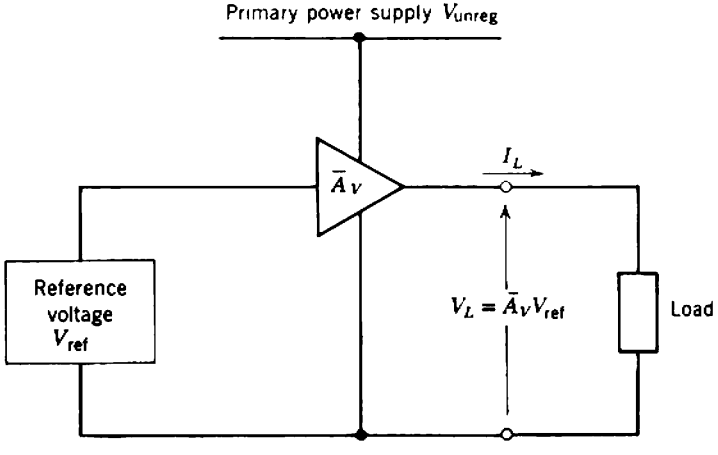


Fig. 16.43 Block diagram of a voltage-regulated power supply.

voltage V_{unreg}, the load current I_L, and temperature T. Usually, the unregulated voltage is supplied from a rectifier-filter system so that V_{unreg} contains ripple components at harmonics of the mains frequency and its average value varies with changes in mains voltage. Furthermore, V_{unreg} is the primary source of power for the regulating system and, because an unregulated supply always has a finite internal impedance, increases in load current must reduce the unregulated voltage. The load current typically varies over a large range from virtually zero to a full-load value that depends on the number and types of circuits supplied. Under special circumstances (e.g., Class B amplifiers) I_L may vary rapidly with time; more often, the presence of decoupling capacitors in the circuits supplied by the regulator keeps the rate of change of I_L relatively low. Temperature changes introduce drift in any dc amplifier and voltage regulators are no exception.

The reference voltage is obtained from a precisely known and stable voltage source such as a glow discharge tube, a Zener diode, or a standard cell. The long-term stability of the output voltage cannot be better than that of the reference, and in critical applications it is necessary to select the reference device carefully, operate it at optimum current, and maintain the temperature constant. Table 16.3 indicates the performance achievable with the best examples of each type. Notice that the ultimate in performance is achieved only at a particular value of voltage; devices having other nominal voltages are available but are inferior in tolerance, temperature coefficient, and internal resistance to those listed.

Table 16.3. Performance of Voltage Reference Elements

Type	Nominal Voltage	Tolerance on Voltage (%)	Temperature Coefficient (ppm/°C)	Usual Operating Current (mA)	Internal Resistance (Ω)
Glow discharge tube	83.7	±1	30	4.5	110
Zener diode	5.6	±5	50	7.5	10
Standard cell	1.01895	±0.02	5	0	100

16.8.1 Basis of Regulator Design

A satisfactory regulator can be realized by designing a good quality dc power amplifier. This amplifier is operated unusually, in that the input signal

$$V_i = V_{ref}$$

is derived from a source which, ideally, is constant for all time. For a given

reference voltage changes in output voltage V_L result only from errors in the amplifier. In any useful regulator the changes in V_L must be small and it is appropriate to assume linearity by writing

$$\Delta V_L = \left(\frac{\partial V_L}{\partial I_L}\right)\Delta I_L + \left(\frac{\partial V_L}{\partial V_{\text{unreg}}}\right)\Delta V_{\text{unreg}} + \left(\frac{\partial V_L}{\partial T}\right)\Delta T, \qquad (16.35)$$

where the terms on the right-hand side represent errors from changes in load current alone (V_{unreg} and T constant), unregulated voltage alone, and temperature alone. These sources of error are not completely independent because ΔV_{unreg} contains a component that depends on the change in load current. Regulators are usually designed so that amplifier errors are comparable with or smaller than the tolerance of V_{ref}. Occasionally, it is possible to achieve some compensation between errors in the reference voltage and errors in the amplifier so that V_L has a smaller tolerance than V_{ref}; an obvious example of this technique is in compensating temperature-induced errors.

The partial derivative $\partial V_L/\partial I_L$ in Eq. 16.35 represents the negative of the output resistance of the regulator and, for this source of error to be small, R_o needs to be minimized. Unlike the other amplifiers discussed in this book, the fundamentally important small-signal quantity of a regulator is the output resistance rather than a transfer function. This, of course, is a consequence of V_{ref} being constant. Nevertheless, it is useful to deduce the output resistance from the mid-band small-signal voltage gain:

$$A_V = \frac{\Delta V_L}{\Delta V_{\text{ref}}}, \qquad (16.36)$$

and R_o is the value of R_L that reduces A_V to half its open-circuit value (Section 5.1.5). In calculating R_o for a regulator, R_L is taken as an independent small-signal parameter which is in no way related to the dc load current:

$$R_L \neq \frac{V_L}{I_L}. \qquad (16.37)$$

Notice, however, that the small-signal parameters used in calculating R_o are functions of the load current; R_o usually decreases with increasing I_L.

Negative feedback is a powerful method for reducing the output resistance of a voltage amplifier. For such an amplifier

$$A_V = -\frac{1}{\beta}\left(\frac{1}{1 - 1/A_l}\right) \qquad (16.38)$$

where A_l is, in general, a function of the load resistance. Assuming that the open-circuit loop gain is large, the maximum value of $|A_V|$ is $1/\beta$, and it follows from Eq. 16.38 that $|A_V|$ falls to half this value if

$$A_l|_{R_L = R_o} = -1. \qquad (16.39)$$

Thus the output resistance of a feedback amplifier is equal to the load resistance for which the small-signal loop gain falls to unity.

As with any feedback amplifier, care must be exercised in designing regulators so that the dynamic response is satisfactory. Often regulator performance is considered acceptable if oscillation does not occur for any condition of load current and load impedance. Regeneration can be investigated by means of the complex loop gain $A_l(s)$ of the regulator considered as a small-signal amplifier. In this instance the load impedance is not an independent variable as assumed above but the small-signal impedance

$$Z_L(s) = \frac{v_L(s)}{i_L(s)} \tag{16.40}$$

seen by the regulator output. The impedance $Z_L(s)$ is usually a time-varying parallel RC combination in which only the general ranges of resistance and capacitance are known. It is standard practice to define the minimum load capacitance seen by the regulator by shunting a large capacitor across its output.

Under some special circumstances the dynamic response needs much more careful control; for example, the reflector supply for a modulated reflex klystron needs to be a square wave without overshoot and of precise amplitude. This time-varying regulated voltage can be obtained from a conventional voltage regulator by switching V_{ref} on and off, provided that the dynamic response of $A_l(s)$ is satisfactory. It is easy to play with words and say that the klystron supply is a precision amplifier rather than a static regulator, but, in truth, it is extremely difficult to distinguish between a voltage regulator and a precision dc power amplifier. Indeed, some commercial manufacturers supply equipment that can be used for either function, depending on whether the input is static or time-varying.

16.8.2 Practical Details

In the simplest regulators the amplifier is just a cathode follower or emitter follower as shown in Fig. 16.44. The large-signal voltage gain is approximately unity, so that the output voltage is fixed at approximately the reference voltage. Further disadvantages of this simple regulator are its relatively large output resistance (approximately $r_h/\beta_N + r_e$), and its output drift resulting from changes in V_{BE} with changes in temperature or unregulated supply voltage.

More sophisticated regulators achieve large-signal voltage gains different from unity (and therefore output voltages different from the reference voltage) by feeding back a constant fraction β of the output voltage. Multistage amplifiers are used to increase the loop gain and reduce the

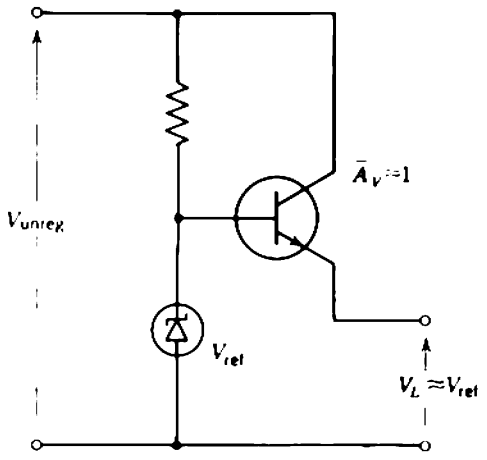


Fig. 16.44 The simplest regulator: an emitter follower.

output resistance. Better dc amplifiers (better comparators, particularly) such as long-tailed pairs reduce the drift. These and other ideas are perhaps best explained by reference to practical circuits.

16.8.2.1 A Simple Common-Emitter Regulator

Figure 16.45 shows a simple regulator in which the common-emitter output transistor Q_3 is fed from a long-tailed pair comparator. The feedback factor β is 0.63, determined by the 330- and 560-Ω divider chain, and the regulated output with 6.2-V reference is close to 10 V. To some degree the transition from a long-tailed pair to a single-ended output stage negates the good drift performance of the comparator. However, changes in V_{CE} for Q_1 do not change its collector current significantly, so that changes in V_{unreg} do not cause significant drift. Notice, too, that the reference diode is supplied from the regulated output rather than V_{unreg}. Temperature-induced drift is kept within acceptable limits by the long-tailed pair comparator.

The loop gain, hence the output resistance, can be calculated by breaking the feedback loop at the base of Q_2 and assuming that a small-signal load resistance R_L exists across the output terminals:

$$A_l \approx -G_7(Q_1, Q_2) \times \beta_N(Q_3) \times R_L \times \beta. \qquad (16.41)$$

Three approximations are implicit in Eq. 16.41:

(i) the Zener diode impedance is small so that no signal is fed back to the base of Q_1,
(ii) the input resistance of Q_2 is large compared with 560 Ω,

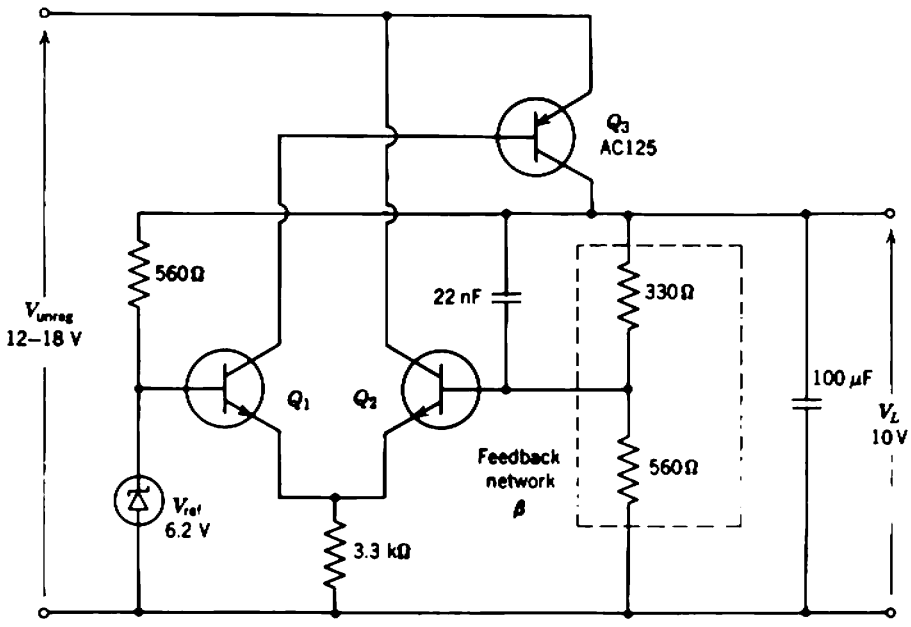


Fig. 16.45 Basic common-emitter regulator with a long-tailed pair comparator. The output is 10 V rated at 100 mA.

(iii) R_L is small compared with both the 560-Ω Zener supply resistor and the 330- and 560-Ω divider chain.

The output resistance is the value of R_L for which $|A_l|$ given by Eq. 16.41 is unity:

$$R_o \approx \frac{1}{G_T(Q_1, Q_2) \times \beta_N(Q_3) \times \beta} \tag{16.42}$$

For the components shown R_o is about 1 Ω. Regeneration of the feedback loop is prevented by two capacitors: 100 μF contributes a dominant pole and 22 nF contributes a zero of the feedback factor at 20 kHz.

The output transistor Q_3 must be adequately rated to dissipate a power

$$P_C = (V_{unreg} - V_L)I_L. \tag{16.43}$$

Maximum dissipation usually occurs when I_L has its maximum value, and it is tempting to rate Q_3 on the basis of this dissipation. Laboratory supplies, however, are subject to accidental overloads or short circuits and should therefore be protected by fast-acting fuses (of thermal time constant less than that of Q_3) or, preferably, by an electronic switch that turns the supply off when the load current exceeds its rated value. It is left as an

exercise for the reader to show that some measure of self-protection is built into Fig. 16.45; the load current cannot exceed about 200 mA, which current must be maintained for several seconds to damage Q_3.

16.8.2.2 A Compensated Common-Collector Regulator

The output transistor in Fig. 16.45 is operated in common emitter, the configuration that normally yields the highest power gain. Often regulator designers are forced to use the output device in the common-collecting-electrode configuration. For example, a very-high-power *p-n-p* transistor might not be available for Q_3 in a scaled-up version of Fig. 16.45. The nonexistence of tubes of polarity corresponding to *p-n-p* transistors often precludes the use of anything but a cathode-follower output stage.

Other factors being equal, a common-collecting-electrode regulator has the same output resistance as a common-emitting-electrode regulator. Figure 16.46 shows, in outline, a common-collector regulator, in which, for simplicity, the comparator is a single-ended stage rather than a long-tailed pair. As in Fig. 16.45, the output resistance is

$$R_o \approx \frac{1}{G_T(Q_1) \times \beta_N(Q_2) \times \beta},$$

where there is one new approximation:

$$R_C \gg \beta_N(Q_2) \times R_o.$$

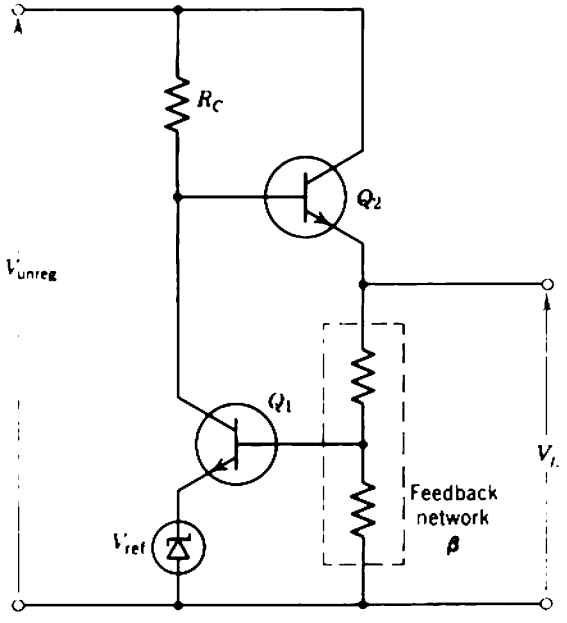


Fig. 16.46 Basic common-collector regulator.

What is inferior in common-collecting-electrode regulators is the drift resulting from a change in unregulated supply voltage. If V_{unreg} changes, then, in the absence of feedback, the output voltage in Fig. 16.46 will change by almost the same amount because the collector current of Q_1 is independent of V_{CE}. The change in V_{unreg} propagates to the base of Q_2 via R_C, thence to the output. When the feedback loop is closed,

$$\frac{\Delta V_L}{\Delta V_{\text{unreg}}} = \frac{1}{G_T(Q_1) \times R_C \times \beta} \tag{16.44}$$

In principle this source of drift can be compensated by applying a voltage

$$\frac{\Delta V_{\text{unreg}}}{G_T(Q_1) \times R_C}$$

to the base of Q_1; a resistor of suitable magnitude is connected between V_{unreg} and the base of Q_1. In practice the compensation needs to be a preset adjustment because of tolerances on components.

If preset adjustments are acceptable, a worthwhile improvement to both common-emitting-electrode and common-collecting-electrode regulators is a positive current feedback loop that reduces the output resistance (ideally) to zero. Positive current feedback is considered in Problem 10.8.

Figure 16.47 is the complete circuit for a compensated common-emitter regulator. Notice the following:

1. Temperature-induced drift in the single-ended comparator is held within acceptable limits by using a Zener reference diode whose temperature coefficient approximately matches that of V_{BE} of the comparator transistor.
2. A compound output transistor is used to increase the loop gain.
3. Electronic overload protection is provided by two silicon diodes and the 2.5-Ω resistor in the emitter circuit of Q_3. If the load current exceeds 250 mA, the diodes become forward biased and pull down the base of Q_2, thereby preventing further increase in current.
4. Experience in developing these regulators indicated that compensation for changes in V_{unreg} should be a screwdriver adjustment, whereas the positive current feedback could be provided by a fixed 10%-tolerance component (the 0.15-Ω resistor).
5. A production batch of these regulators met the specification shown in Fig. 16.48. After one year, the six units tested were all within specifications. After four years, the 2.5-mV regulation of some units had deteriorated by an order of magnitude; it could be restored by adjustment of the preset controls.

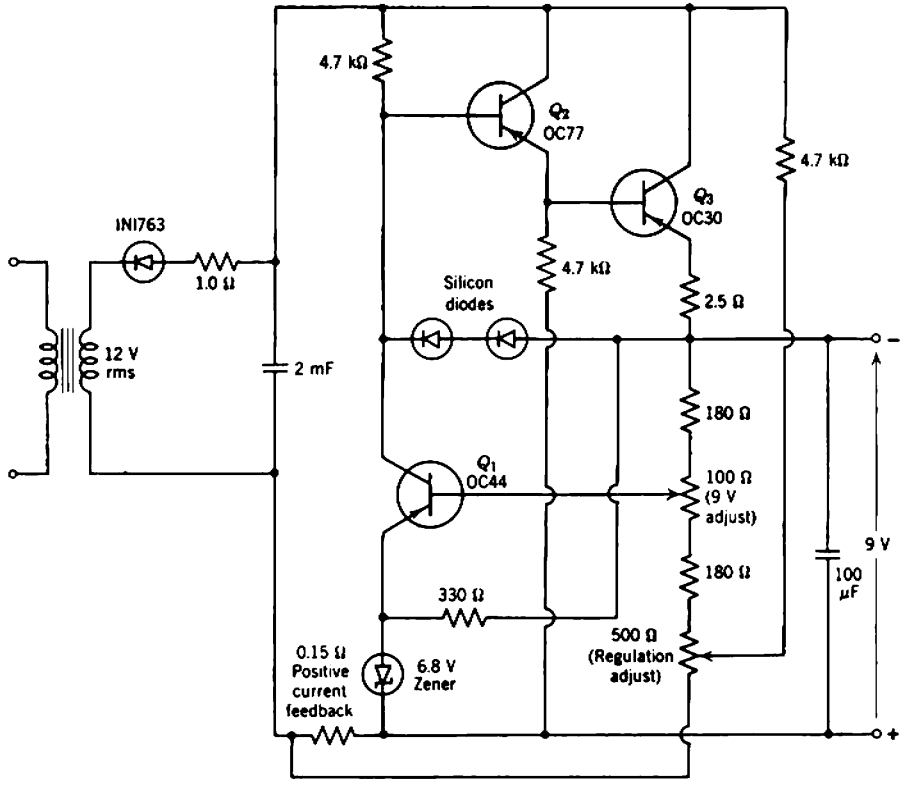


Fig. 16.47 Practical 9-V/250-mA common-collector regulator with drift compensation, positive current feedback, and overload protection.

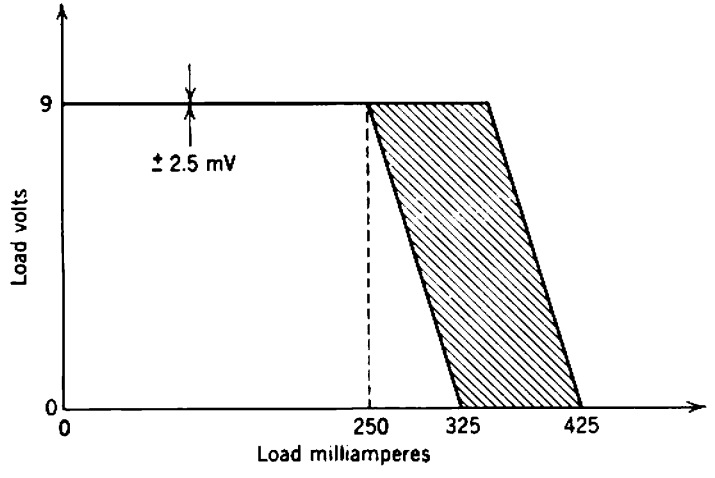


Fig. 16.48 Output characteristic for Fig. 16.47.

16.8.2.3 A Precision Regulator

Figure 16.49 is the circuit of a precision regulator that relies on high loop gain rather than compensation to achieve a low output resistance. The output is 20 V $\pm 0.05\%$ for load currents up to 500 mA and for a $\pm 10\%$ change in mains voltage. Electronic overload protection is provided.

The 4.7-V reference diode is fed from the regulated output. Because the diode resistance is 20 Ω, V_{ref} remains within $\pm 0.05\%$ of its nominal value unless V_L changes by almost $\pm 2\%$. Clearly, therefore, errors in V_L must arise predominantly from the amplifier rather than the reference voltage.

Changes in V_{unreg} have negligible effect on the collector currents of Q_1 and Q_2; Q_3 is used in the circuit of Fig. 16.41*b* to ensure that the compound output transistor (Q_4 and Q_5) is fed only with the differential output current from the comparator. Overload protection is provided by Q_6, which conducts and pulls down the base of Q_4 when the load current exceeds 600 mA. Notice that Q_3 is a germanium transistor which has a minimum collector current of a few microamperes. This residual current

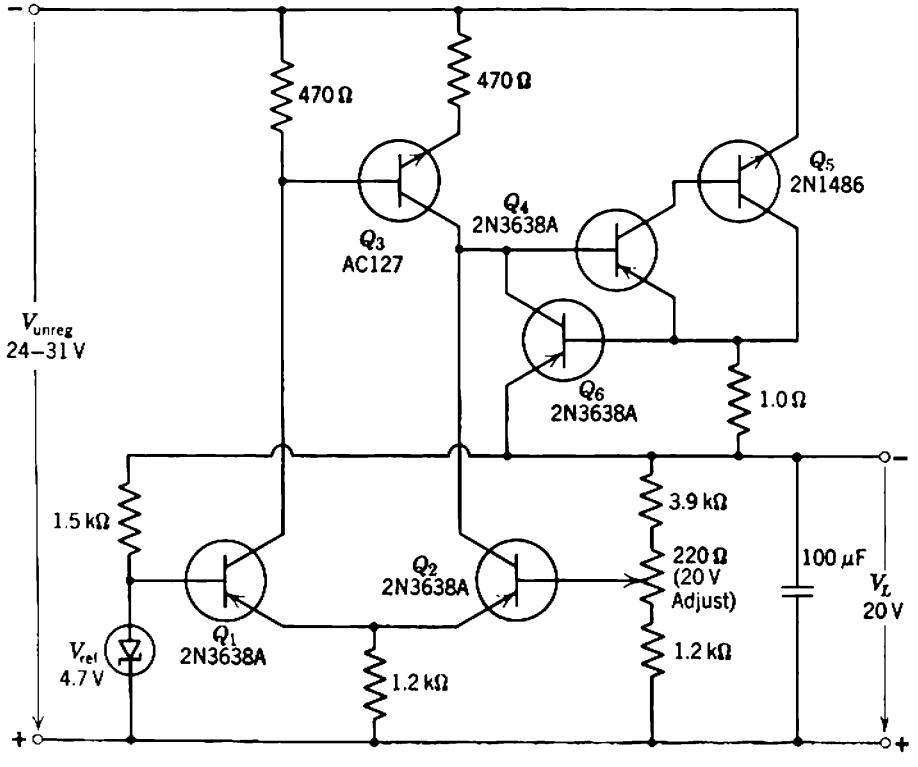


Fig. 16.49 A precision 20-V/500-mA regulator with overload protection.

ensures that the regulator always begins to operate when V_{unreg} is switched on; without this starting provision it is possible for Q_2, Q_3, and Q_4 to be locked in a mutually cutoff condition.

The loop gain of the regulator with a small load resistance R_L is

$$A_l = -G_T(Q_1, Q_2) \times 2 \times \beta_N(Q_4) \times \beta_N(Q_5) \times R_L \times \beta.$$

The output resistance is numerically equal to the load resistance for which the loop gain is unity; rearranging and substituting component values yields

$$R_o = 0.014\ \Omega.$$

As the load current is increased from zero to 500 mA, the output voltage should therefore fall by 7 mV; the measured fall is 8 mV, or 0.04% of the nominal output.

Chapter 17

Components, Construction, and Reliability

The foregoing treatment of low-pass amplifiers should convince the reader that amplifier design involves such a large number of variables that it approaches an art. It may be necessary to consider the following aspects of performance in any amplifier design: gain, gain stability, nonlinearity, bandwidth, transient response, dynamic range, and noise. Often the consideration given to some aspect needs only be trivial; for example, it is rarely necessary to worry about noise in power amplifiers. Nevertheless, the success of a practical amplifier depends on the ability of the designer to produce a realization which is a compatible solution of a number of different (and sometimes conflicting) problems. There is scarcely ever a unique solution to a design problem, so papers or articles which purport to present general, formalized design procedures should be viewed with suspicion.

The ideas developed in preceding chapters apply predominantly to the initial pen-and-paper stage of design; the subsequent transition from a final circuit diagram to a useful, reliable piece of hardware introduces a number of new variables and constraints, and compels the designer to consider questions such as the following:

(i) What physical form should the complete equipment take?
(ii) What type of mechanical construction should be used for sub-units?
(iii) How should devices and components be mounted?
(iv) What testing and maintenance facilities should be included?
(v) What is the expected (or required) reliability of the equipment?

Each of these questions leads to further questions, and correct answers are

certainly not unique. The ability to make correct decisions on such questions is not acquired readily, and comes only with experience. Thus, if pen-and-paper design is an art, then completion of the job to the hardware stage is an art of higher order. This is not to suggest that the pen-and-paper design is an abstract exercise carried out with no thought to construction, production, and maintenance, nor is it suggested that these latter problems are solvable only by witchcraft. Rather, the primary aim of this chapter is to help create among potential designers an awareness of these practical or technological problems.

Much emphasis is given in preceding chapters to the factors which limit the performance of amplifying devices; this is logical, because the inherent performance capability of an amplifier depends ultimately on the amplifying device properties. However, imperfections in passive components often degrade amplifier performance. A good example is flicker noise in carbon resistors (Chapter 8) which can degrade noise performance far below that expected from practical amplifying devices with ideal passive components. Equally, the inductance and capacitance associated with all practical resistors can degrade high-frequency performance. It is therefore desirable to consider briefly the advantages and limitations of the more common types of components; an outline is given in Section 17.1.

Amplifying devices and passive components can be united to form electronic equipment having a variety of constructional forms. A number of standardized mechanical arrangements have evolved, and some of these are presented for guidance in Section 17.2. The continuing search for greater operating speed, greater reliability, and minimum size has stimulated the development of quite new construction methods in which complete (or near-complete) circuits are manufactured as an integrated whole. In some instances devices, components, interconnections, and mechanical framework are all created in a single process. Such circuits are generally known as *microelectronic circuits* or *integrated circuits*. The two main production processes are based on diffusion and thin-film technologies; most practical microelectronic circuits use a combination of the two. Section 17.3 gives a very brief introduction.

Finally, some aspects of the important subject of reliability are introduced in Section 17.4. Reliability is a vast subject which embraces good design, the use of good components, sound construction and production techniques, and so on; in short, good reliability calls for good engineering in the widest sense. Some writers have been tempted to concentrate excessively on those aspects of reliability which can be given a mathematical significance, and reliability is often presented as a purely statistical study. Undoubtedly, some statistical ideas are required for the specification of reliability, but significant improvements come from careful investigations

of the physical reasons for failure, not from the mere use of more impressive statistical techniques.

17.1 PROPERTIES OF CONVENTIONAL COMPONENTS

Most conventionally constructed amplifiers employ many more passive components than amplifying devices, so a large demand has arisen for cheap, reliable components. This demand has been met very satisfactorily by the research and production facilities of component manufacturers. New components of smaller physical size, narrower tolerances, lower price, and higher reliability are being introduced continually, and the most satisfactory way of keeping abreast of these developments is to scan the periodical literature and manufacturers' data. The basic physical properties of matter used in some types of components are of great interest, but these and the equally interesting production processes are beyond the immediate scope of this book. The following paragraphs merely summarize the important features of a selection of readily available components.

17.1.1 Resistors*

Resistors are the most-used electronic components and it is essential that, of all components, these should be the most reliable. Commercially available resistors of all types exhibit satisfactory reliability for the majority of applications; greater reliability can be achieved only by increasing production costs. Resistors usually have a cylindrical geometry with lead wires emerging axially from opposite ends; in rarer examples the leads emerge radially from the ends, parallel to each other and at right angles to the axis of the cylinder. The resistance magnitude and tolerance are sometimes indicated by printed numbers, more usually by a series of colored bands which follow a standard color code. The two bands closest to the end of the resistor give the two significant digits of the resistance magnitude; the third band completes the specification of nominal resistance by defining the number of zeros that follow or precede the two digits, and a fourth band denotes the tolerance on this nominal value. The significance of the colors used is given in Table 17.1. Thus bands running orange, white, red, silver denote a 3.9-kΩ resistor with $\pm 10\%$ tolerance.

It is necessary to be careful in interpreting the manufacturers' quoted tolerances. Factors which can give rise to departures from the nominal resistance value include

(i) temperature changes,

* G. W. A. DUMMER, *Fixed Resistors* (Pitman, London, 1956).

(ii) initial production variations,
(iii) long-term variations (drift),
(iv) nonlinearity of volt-ampere characteristic in some high-valued types.

The effects of temperature change are often quoted separately as a temperature coefficient of resistance. This can be positive or negative, and is usually quoted as a change of so many parts per million per degree centigrade (ppm/°C). Some manufacturers provide information on the thermal properties of resistors so that temperature rise can be related directly to the power being dissipated. Resistors of normal size have thermal time constants of several seconds; this fact must be remembered if measurements of thermal resistance are attempted. The magnitude of initial production variations depends on the method of construction and the accuracy of the production process. Long-term variations occur due to imperfections in the manufacturing process; typical causes include ingress of moisture, gaseous contamination of resistor material, and chemical changes within the material.

Quoted tolerances are probably most convenient for the circuit designer when they refer to the total effect of all types of change in resistance. Such tolerances imply a limit to operating temperature (and hence power

Table 17.1 Resistor Color Code

	First and Second Bands	Decimal Multiplier (Third Band)	Tolerance (Fourth Band)
Black	0	1	—
Brown	1	10	±1%
Red	2	10^2	±2%
Orange	3	10^3	—
Yellow	4	10^4	—
Green	5	10^5	—
Blue	6	10^6	—
Violet	7	—	—
Grey	8	—	—
White	9	—	—
Gold	—	0.1	±5%
Silver	—	—	±10%
No color	—	—	±20%

dissipation) and a specified operating time (for example, 10,000 hours). However, this type of tolerancing covers both the short- and long-term stability of the component. Because short-term variations of resistance depend almost entirely on temperature changes, it is normal to specify the temperature coefficient (as stated above).

Readily available resistor values cover a range of about seven decades—from about one ohm to a few tens of megohms. The four most popular types of construction are:

1. *Wirewound resistors* consist of a coil of wire of accurately known resistivity and temperature coefficient wound on an insulating bobbin. Production tolerances smaller than 0.01% can be maintained, and temperature coefficients better than $\pm$ 100 ppm/°C can be achieved. Precision resistors of this type are expensive. The great disadvantage of wire-wound resistors is the relatively large series inductance which restricts operation to frequencies below a few megahertz.

2. *Composition resistors* are made from a mixture of carbon with other powdered materials. They have the advantage of cheapness and very low inductance, but they cannot be made with precision and they exhibit relatively large long-term variations; tolerances are rarely better than 5%. In addition, composition resistors generate flicker noise. Shunt capacitance restricts the application of high-valued composition resistors to frequencies below a few megahertz. Series inductance restricts the application of low-valued types to a few hundred megahertz. Composition resistors are used extensively in noncritical positions such as biasing resistors.

3. *Cracked carbon* or *pyrolitic resistors* are made by cracking hydrocarbons which have been deposited on ceramic rods, and then machining the cracked carbon into a spiral. The shunt capacitance is considerably smaller than that of composition types and the high-frequency limit occurs at a few hundred megahertz. Flicker noise is present, but less than in composition resistors. Typical tolerances lie between 1 and 10%.

4. *Metal film resistors* are made by depositing a film of metal on a glass support of *substrate*; they are something of a compromise between the cheapness of composition resistors and the precision of wire-wound resistors. The manufacturing process is flexible, it being possible to trade precision for cost to such an extent that metal film resistors of different types can be used throughout an amplifier: low-cost types for bias resistors, etc., and more expensive types of feedback resistors, etc. Tolerances smaller than $\pm$0.5% are possible in production batches. The useful frequency range for metal film resistors is up to a few hundred megahertz.

17.1.2 Capacitors*

In contrast to the relatively small number of resistor types in use, the range of available capacitor types is quite large and it may be difficult for the designer to select the best capacitor for a given requirement. Capacitors consist essentially of two conducting plates or electrodes separated by a dielectric. Capacitance is increased by increasing the area of electrodes, decreasing their spacing, and increasing the permittivity of the dielectric. Because high-permittivity materials have limitations in other properties such as leakage current or voltage rating, capacitors normally employ the lowest permittivity dielectric consistent with their capacitance value and physical size.

An initial simplified classification may be based on the approximate range of capacitance values achievable with each of six common types of capacitor (Fig. 17.1). It is important to note that all capacitors have some residual inductance, and the upper-frequency limit occurs when the inductive reactance becomes comparable with the capacitive reactance.

1. *Air dielectric capacitors* ($\varepsilon_r \approx 1$) are useful for achieving very precise low values of capacitance or, with movable vanes, may be used as variable capacitors. These components have high voltage ratings and are useful up to frequencies of several hundred megahertz.

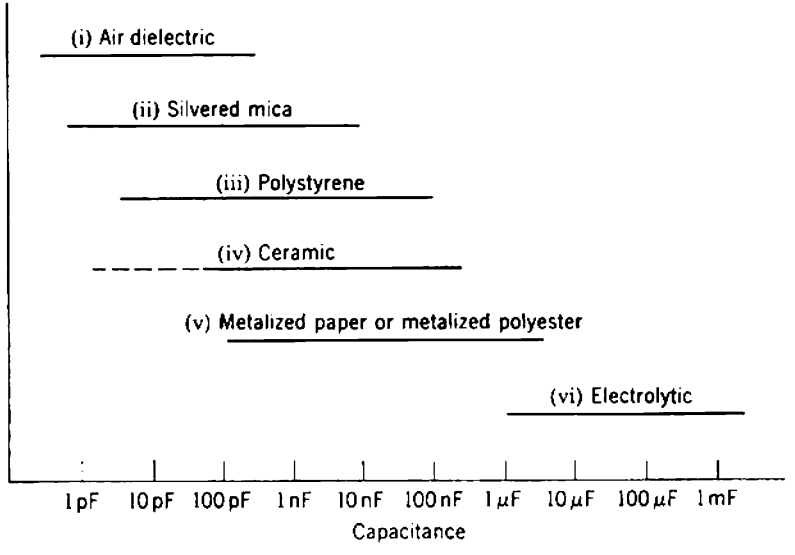


Fig. 17.1 Approximate capacitance range of commercially available units.

* G. W. A. DUMMER, *Fixed Capacitors* (Pitman, London, 1956).

2. *Silvered mica capacitors* ($\varepsilon_r \approx 5$) are produced by depositing thin films of silver on mica sheets. These are precise components, of moderate voltage ratings, and useful up to frequencies of several tens of megahertz. The tolerance may be as small as $\pm 1\%$.

3. *Polystyrene capacitors* ($\varepsilon_r \approx 3$) are somewhat similar to silvered mica capacitors but employ a synthetic dielectric which allows a higher maximum capacitance to be achieved in a given volume with a given voltage rating. The tolerance may approach $\pm 3\%$.

4. *Ceramic capacitors* ($\varepsilon_r \approx 1{,}000$) employ high-permittivity ceramic dielectrics. Their advantage is small physical size, but tolerances are much poorer than for silvered mica capacitors; a tolerance of $\pm 5\%$ is very good for these components. Ceramic materials of much lower permittivity but improved characteristics are used for small capacitors in the picofarad range.

5. *Metalized paper* or *metalized polyester capacitors* ($\varepsilon_r \approx 5$) are constructed by rolling up a sandwich of metal foil and a solid, pliable dielectric such as paper. They have moderate voltage ratings, and tolerances which may be as small as $\pm 10\%$.

6. *Electrolytic capacitors* (ε_r high and uncertain) are available in a variety of forms, most of which consist of the following essential elements:

(a) a metal positive electrode or anode,
(b) a dielectric which is a thin layer of oxide formed on the metal by electrolysis,
(c) an electrolyte,
(d) a metal negative electrode (the cathode) which makes contact with the electrolyte.

The electrodes are often made from aluminum or tantalum foil, and various wet and dry electrolytes are used. Unlike the other types, electrolytic capacitors require a unidirectional polarizing voltage to keep the oxide "formed." The oxide layer may require reforming after the capacitor has been stored or left out of operation for a period, and an excessive leakage current flows during this forming process. Electrolytic capacitors rarely have tolerances better than $\pm 10\%$, and are useful only at frequencies up to a few megahertz. As frequency increases, electrolytic capacitors exhibit first capacitive, then resistive, and finally inductive impedance.

Readers are urged to make a study of manufacturers' data to supplement this brief and inadequate discussion of capacitor types.

17.1.3 Inductors

Of the three basic passive components—resistors, capacitors, and inductors—the latter are the most difficult to realize accurately. Conventional

resistors and capacitors approximate to ideal elements except at quite high frequencies; inductors, however, have associated capacitance and resistance which must be considered even for operation over a small range of frequencies. Furthermore, inductors are expensive and are not available commercially to the same extent as other components. Inductors are wound on cores of air, laminated magnetic materials, powdered magnetic materials, and ferrites. Most circuit designers avoid the use of inductors in low-pass amplifier circuits.

17.1.4 Other Components

Other components used in amplifiers are transformers (discussed in Chapter 7), switches, plugs and sockets, dials, etc. It is quite impossible to discuss these and other components in a book of this size, but it should be emphasized again that considerable thought must be given to choosing every component used in a piece of equipment. It is a sobering thought that the availability of many sophisticated items of electronic equipment is impaired by inoperative pilot lamps!

17.2 STANDARD METHODS OF CONSTRUCTION

Complete electronic systems usually consist of a number of elemental sections such as amplifiers, waveform generators, logic circuit blocks, and power supplies. It is useful to differentiate between the methods used in constructing complete systems and the methods used in constructing the elemental sections. In the context of this book, the construction of elemental sections is more pertinent and is discussed at greater length.

17.2.1 Construction of Electronic Systems

Three main methods of construction in general use are:

(i) plug-in units,
(ii) rack-mounting units,
(iii) separate free-standing chassis.

These classifications are not mutually exclusive and it is often convenient to use combinations of construction methods.

Plug-in units are arranged to fit and lock into a special shelf unit in such a way that power supply connections (and sometimes signal connections) are made as the unit is installed. Often, the shelf accepts any one of a

number of alternative units (or modules); examples are the alternative plug-in amplifiers for most high-quality oscilloscopes. In other situations, the shelf accepts various combinations of plug-in units; this method is useful in flexible measurement systems, such as those used in nuclear research. An example of plug-in units of this type is shown in Fig. 17.2. This is a very useful construction method for professional equipment.

Rack-mounting units are positioned by fixing screws which hold a substantial front panel to a framework of vertical steel channels (the rack). The width of the front panel is often 19 inches, and various standardized panel depths may be used. The units can thus be stacked vertically on the rack as shown in Fig. 17.3. This method of construction is very convenient for vacuum-tube equipment, but for transistor equipment the minimum module size available is often inconveniently large. Aside from this difficulty, plug-in units and rack-mounted units share the essential advantages of flexibility (allowing systems to be changed readily) and easy maintenance (allowing inoperative units to be replaced with spare units while maintenance proceeds).

Separate free-standing chassis are used for many test instruments, signal generators, and other relatively small selfcontained equipment. The chassis is usually rectangular in shape, although a sloping front panel or top panel is sometimes used. A few free-standing chassis can be stacked without introducing too great a mechanical stability problem but, in general, free-standing units are an embarrassment in large systems. Some examples of free-standing units are shown in Fig. 17.4.

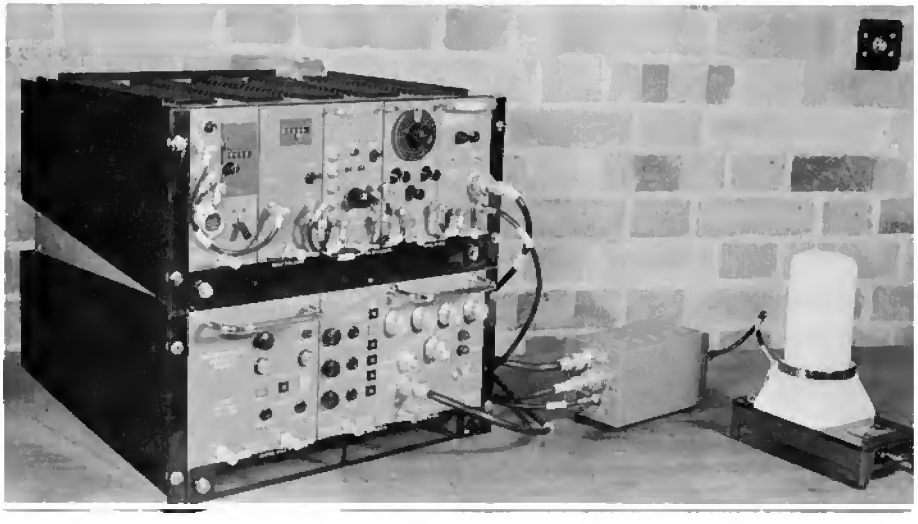


Fig. 17.2 Plug-in units and shelf. (Courtesy of United Kingdom Atomic Energy Authority.)

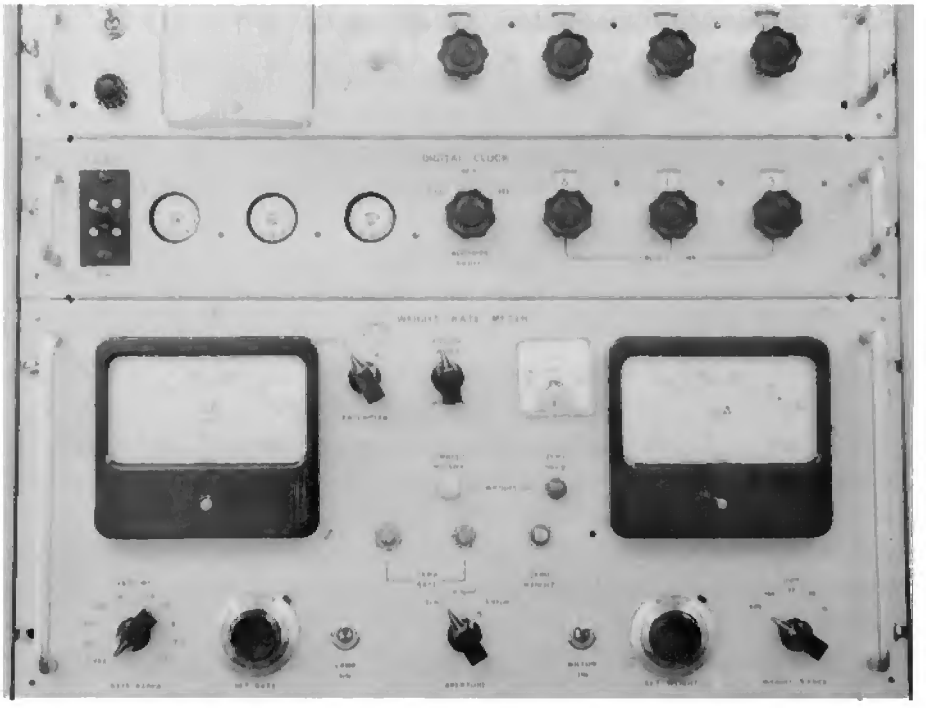


Fig. 17.3 Rack-mounting units.

Fig. 17.4 Typical laboratory use of free-standing equipment.

Irrespective of the construction method in use, the final electrical and mechanical design should consider such things as cooling requirements, mechanical fixing of devices and components, positioning of controls, and accessibility for maintenance. For a complete discussion of these problems the reader is referred to the literature.* Much can be learned from careful inspection of the products of reputable manufacturers.

17.2.2 Construction of Elemental Sections

The construction of vacuum-tube equipment has been based largely on the need to support relatively massive tubes in sockets which are recessed into a metal chassis or subchassis. The positioning of other components is dictated to a great extent by the disposition of these sockets. In amplifiers for use at high frequencies, it is usual to place the tube sockets as close together as is practicable and to wire all high-frequency components directly to and between socket pins; this is known as *point-to-point wiring* and is illustrated in Fig. 17.5. In a variation known as *cordwood wiring*, a number of components (perhaps all those for one stage) are connected together as a largely self-supporting module with very little extra bracing. In amplifiers for use at low frequencies, it is usual to mount most of the components on tagboards and to make connections to the sockets by means of relatively long leads; this is known as *tag-board wiring* and is illustrated in Fig. 17.6. It may be possible to combine point-to-point wiring with tag-board wiring by using the former method for signal paths and the latter for biasing components.

The disadvantages of these well-established constructional methods are that maintenance is difficult to perform, that production difficulties arise, and that the technique does not lend itself to the degree of miniaturization possible with semiconductor devices. Consequently, modern practice has swung toward the use of relatively small removable *boards* or *cards* onto which all the devices and components of an elemental section are placed. This technique is most useful for transistor equipment but may also be used for vacuum-tube equipment. The boards may be held in position by screws, or, alternatively, plugged into appropriate sockets. Several different types of board are in use; most of the boards in modern equipment have linear dimensions less than 6 inches.

* K. M. CARROLL, *Mechanical Design for Electronics Production* (McGraw-Hill, New York, 1956). Also, G. W. A. DUMMER, C. BRUNETTI, and L. K. LEE, *Electronic Equipment Design and Construction* (McGraw-Hill, New York, 1961).

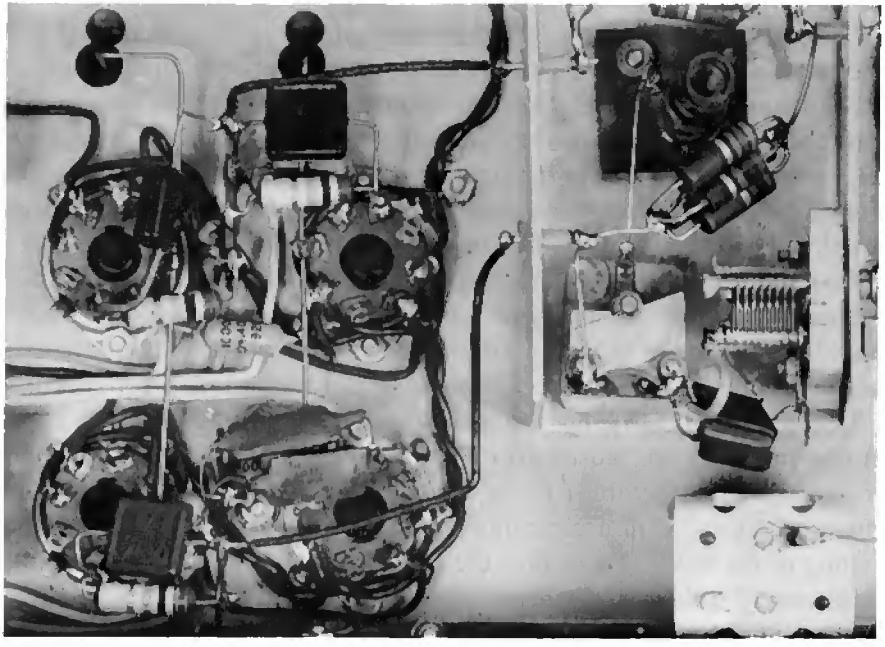
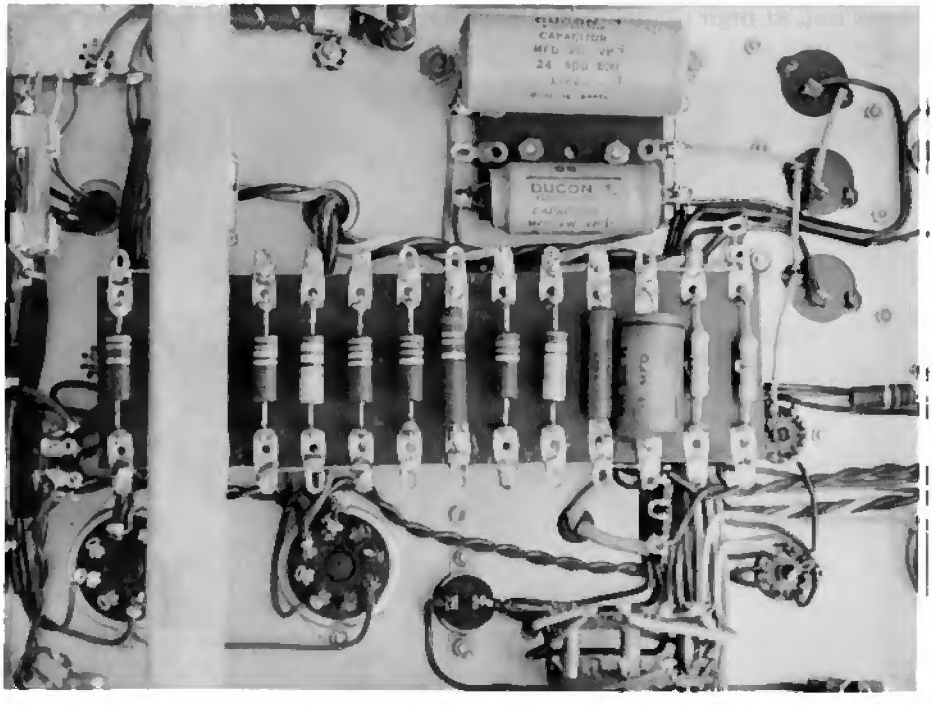


Fig. 17.5 Point-to-point wiring in vacuum-tube equipment.



1. *Perforated metal boards* are available from several manufacturers. The metal sheet approximates to an ideal ground plane which has negligible resistance and inductance between any points on its surface. Consequently, all ground connections can be made directly to any convenient point on the ground plane. This minimizes the residual inductance of individual ground leads and ensures that ground currents cannot produce undesirable signal voltages by flowing through common impedances. The use of multi-point ground plane techniques is almost mandatory in high-frequency circuits. Components are wired between insulated pins which thread selected holes in the board. Tube and transistor sockets are available, although it is usually preferable to solder transistors directly to pins.

2. *Copper-clad plastic boards* of the type shown in Fig. 17.7 with copper on one side only are most attractive for experimental and developmental work and are also suitable for mass production. Holes for pins are drilled, where necessary, with a clearance drill which removes the copper in the vicinity of the hole. Drive pins then grip firmly in the plastic and components can be soldered to them. The copper cladding forms an almost ideal ground plane, so this form of construction is very useful for high-frequency circuits.

3. *Plastic boards* without copper cladding are useful for low-frequency work but, at high frequencies, the ground plane provided by the copper is

Fig. 17.7 Copper-clad plastic board construction of a 300-MHz amplifier.

usually required to prevent excessive straying of signal fields. An example is shown in Fig. 17.8.

4. *Printed circuit boards* have one or both sides of the plastic copper-clad; sections of the copper are etched away to leave only that skeleton necessary for electrical connections (Fig. 17.9). Components and devices are soldered to islands in the printed wiring, or alternatively, plated eyelets are let into the board to receive components. Printed circuit boards are very convenient for mass production but are difficult to maintain. A further important disadvantage is given below.

A number of points must be considered in designing the wiring layout for a board. The more important of these can be summarized as follows:

1. Keep signal sections of the board as compact as possible. Long leads should be avoided if the circuit is to pass high-frequency signals. Note that the inductance of a cylindrical wire is given approximately by

$$L \approx 2l\left(\ln \frac{4l}{D} - 1\right) \quad \text{nH}, \tag{17.1}$$

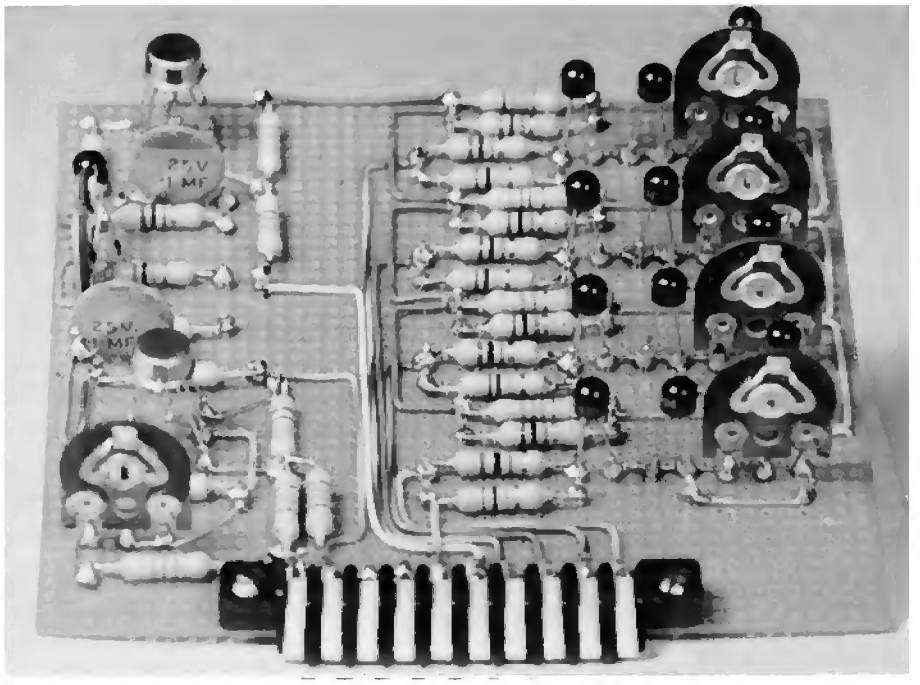


Fig. 17.8 Perforated plastic board construction of a low-frequency circuit.

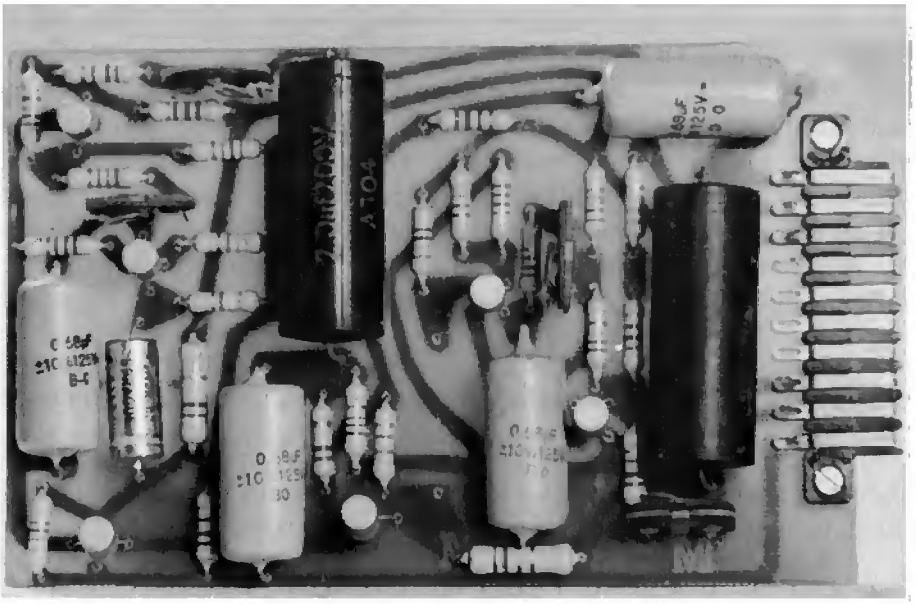


Fig. 17.9 Printed circuit board. The residual copper on the back of the board can be seen in silhouette.

where the diameter is D cm and the length is l cm. A similar calculation* for a flat conducting strip (as produced in printed wiring) of width w cm yields the result

$$L \approx 2l \ln \frac{2l}{w} \quad \text{nH.} \tag{17.2}$$

In either case the inductance is almost directly proportional to the length of the lead; the correction term for aspect ratio is not very significant.

2. Keep the capacitance between out-of-phase points (particularly input and output) low. Notice that quite high stray capacitances can exist between the strip conductors on printed circuit boards. For two parallel strips set distance d apart (Fig. 17.10), the capacitance is very nearly

$$C \approx \epsilon \frac{wl}{d}. \tag{17.3}$$

Lands on printed circuit boards for high-frequency amplifiers should be kept small to minimize the stray capacitance; some external wiring may be desirable also.

* The authors are indebted to M. O. Deighton for performing this calculation.

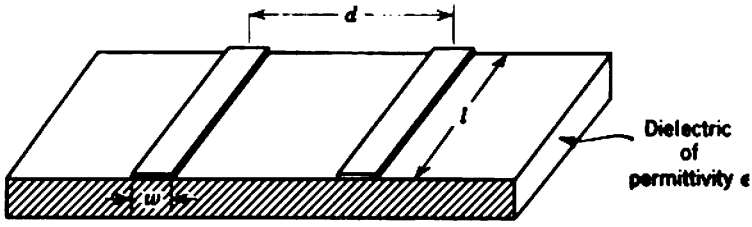


Fig. 17.10 Arrangement of printed conductors for calculation of stray capacitance.

3. Avoid excessive stray capacitance to ground at high-impedance points such as anodes or collectors. With small pins on copper-clad board, the stray capacitance to ground is about 0.7 pF per pin.

4. Use the smallest possible pins to avoid excessive heat storage during soldering. This results in a more reliable soldered joint, because good joints require rapid heating and cooling. Here it should be pointed out that solderless connections of the wrapped type have been used with success in very reliable equipment (see Table 17.2).

5. Ensure that components can be removed without difficulty during maintenance. This usually involves the use of separate pins for each component and not twisting component leads around pins.

6. Space pins adequately so that no excessive mechanical strains are developed in circuit elements due to mechanical stresses on leads or thermal stresses which develop during operation.

As in the previous section, the best advice which can be given is to suggest that the reader look critically at a number of commercially produced circuit boards.

17.3 INTEGRATED CIRCUITS

The packing density for components and devices on conventionally wired boards or printed circuit boards is considerably greater than the packing density on conventionally wired metal chassis, yet there is a demand for even greater packing densities—a demand that cannot be met economically by using still smaller components on smaller boards. As the size of discrete components is reduced, handling them becomes more difficult and production costs rise. Consequently, a major advance in this direction has been the development of production techniques which enable complete circuits to be manufactured without the use of discrete components and, in some cases, without the use of discrete devices. Since

all production processes are under the control of a single manufacturer, these *integrated circuits* yield substantial bonuses in the form of vastly improved reliability and lower cost. Often it is these rather than reduced size that justify the use of integrated circuit techniques. Although the subject is still in its infancy, many significant technological advances have been made.*

Present-day integrated circuits may be grouped broadly into two classes which are not mutually exclusive:

1. *Thin-film integrated circuits*† are produced by depositing thin films of material onto a glass or ceramic support or *substrate*. The normal method of deposition is by vacuum evaporation from a heated source, but other techniques such as sputtering, electroplating, and silk screening are also used. Conducting wires can be produced readily by depositing a suitable metal on the surface. Resistors are constructed from very thin films of tin oxide, tantalum, nichrome, or aluminum. Tin oxide adheres very well to clean glass substrates and resistance values from 50 Ω to 3 MΩ can be achieved; typical temperature coefficients lie in the range 0 to −1500 ppm/°C. The temperature coefficient of tantalum resistors is often smaller and lies between −50 and −100 ppm/°C. Tolerances better than 5% can be achieved with either type. Capacitors of up to about 5 nF with tolerances better than 5% can be produced by depositing a sandwich of metal, dielectric, and metal; silicon oxide and tantalum oxide are the most used dielectric materials. Useful values of inductance are difficult to achieve with simple thin-film techniques; this problem is of little concern in low-pass amplifiers but is a major hurdle in the realization of narrow-band amplifiers. Because films are essentially "all surface" the effective minority carrier lifetimes are extremely short and it does not appear that conventional bipolar transistors can be realized in thin films. Wiemer,‡ however, has produced thin-film insulated-gate field-effect transistors having gain-bandwidth products up to 30 MHz; these devices are not competitive with modern bipolar transistors in normal applications.

2. *Semiconductor integrated circuits* are essentially three-dimensional circuits fabricated in a single chip of semiconducting material (*single-chip* or *monolithic circuits*) or fabricated by combining a small number of chips (*multiple-chip circuits*). Each chip can contain several bipolar transistors, diodes, resistors, and capacitors. All of these individual elements exist in

---

* R. M. Warner and J. N. Fordemwalt, *Integrated Circuits* (McGraw-Hill, New York, 1965).

† L. Holland, ed., *Thin Film Microelectronics* (Wiley, New York, 1966).

‡ R. K. Wiemer *et al.*, "Integrated circuits incorporating thin-film active and passive elements," *Proc. Inst. Elec. Electronics Engrs.*, **52**, p. 1479, 1964.

localized regions or lands which are isolated from each other (and from the silicon substrate) by either a reverse biased *p-n* junction or a region of dielectric (usually SiO_2). For given geometrical dimensions, dielectric isolation results in smaller and constant parasitic capacitances to ground, at the expense of increased production complexity and cost. Most present-day integrated circuits employ *p-n* junction isolation and, because of the continually decreasing size of individual lands, parasitic capacitances have fallen below the values realized in carefully-constructed discrete-component circuits. The individual circuit elements and the isolating boundaries are produced by a relatively small number of compatible solid-state diffusion and epitaxial growth processes. Individual transistors having cutoff frequency f_T that approaches 1 GHz and current amplification factor β_N in excess of 50 can readily be produced. They are generally similar in characteristics to low-power high-frequency discrete transistor types. Thus present-day diffusion technology can realize far more useful amplifying devices than can present-day thin-film technology. Against this is the fact that diffused passive elements have a very restricted range of values, exhibit substantial nonlinearity, and have relatively wide tolerances. For example, it is difficult to realize diffused resistors and capacitors having nominal magnitudes in excess of 20 kΩ and 400 pF, respectively. Absolute tolerances on component values are seldom better than $\pm 20\%$, although ratios tend to be much better.

A newer class of semiconductor integrated circuit uses metal-oxide-semiconductor or unipolar transistors (m.o.s.t.s) instead of bipolar transistors. The individual m.o.s.t.s do not require isolation because the metallic gate electrode defines the control region. They can, therefore, be very small indeed, and it follows that the component density can be very large; several hundred m.o.s.t.s can be contained within a chip measuring 0.1 in. × 0.1 in.

At the time of writing (1967), semiconductor integrated circuits have been used more extensively in switching applications. However, with improvements in manufacturing technology and circuit design techniques appropriate to this technology, excellent integrated amplifiers are now available. Often these amplifiers are realized entirely within a single chip of silicon by utilizing the fact that parameters depending on resistivity and crystal temperature tend to vary in step with each other. In other instances the chip contains the transistors and noncritical components, while external elements (such as thin-film resistors and capacitors) are used for critical components like feedback resistors.*

* D. R. Breuer, "Integrated high-frequency dc amplifiers," *TRW Space Technology Laboratories*, Report No. 9990-6778-RU-000.

There are several differences in the design practices employed* for the construction of discrete component circuits and integrated circuits. The main difference arises from the economic factor that the total surface area of the lands controls production costs. Because transistors occupy smaller surface areas than capacitors and resistors, the total number of transistors is seldom a controlling factor and relatively large numbers of active elements can be used. Quite apart from this economic constraint, there are definite circuit reasons for using more transistors than is common in discrete circuits; for example, high-value capacitors cannot be realized and it becomes necessary to employ direct coupling between amplifier stages. This, in turn, necessitates the fairly general use of long-tailed pairs, even for amplifiers which are not required to respond to very low frequencies. Thus semiconductor integrated circuits employ relatively large numbers of transistors and relatively small numbers of resistors and capacitors. M.o.s.t. circuits seldom contain any passive components at all; transistor current sources are used as biasing resistors.

Despite these changes in some of the circuit design rules applicable to the new technology, the fundamental circuit design requirement is unchanged; that is, to design a circuit to perform its desired function predictably, reproducibly, and reliably at (ideally) the smallest possible cost. Circuit design in this area is a relatively new art and there is tremendous scope for creativity. The artificial barrier which has grown up between device and circuit engineers must be broken down, because the amplifier designer of the future will need to know even more about solid state physics and material technology than are required to understand the operation of present-day semiconductor devices.

17.4 RELIABILITY

The operating life of all electronic components is finite and failure will ultimately occur; usually the failure of any component renders equipment inoperative. Two essential design tasks are to ensure an adequate probability of successful operation over some specified time (i.e., to improve the reliability) and to ensure that the inoperative equipment can be maintained and returned to service as quickly as possible (that is, to improve the maintainability). These are engineering tasks and, as in most engineering

* P. Cooke, J. D. Evans, M. J. Gay, and J. S. Brothers, "Solid circuit design considerations," *Applications of Microelectronics* (I.E.E. conference publication No. 14, p. 11/1, September, 1965). Also R. J. Widlar, "Some circuit design techniques for linear integrated circuits," *Trans. Inst. Elec. Electronics Engrs.*, **CT-12**, 586, December 1965.

tasks, there are many factors to be considered and no simple universal rules are possible.

In order to put the assessment of reliability on a quantitative basis, it has become necessary to make use of statistical concepts. This has lead to the mistaken idea that statistical theory is all that is involved in reliability studies. In actual fact, the greatest difficulty of present-day reliability studies lies in the somewhat mundane task of establishing accurate and useful numerical data on the reliability of components, devices, and systems. These data are usually specified as a *failure rate*.

17.4.1 Failure Rate

The failure rate λ of equipment is obtained by dividing the fractional number of failures during a period of time Δt by the period of time. To be specific, suppose that N samples of some unit are being tested, and at any time t the total number of failures is N_F. The failure rate at time t is defined as

$$\lambda = \left(\frac{\Delta N_F}{N - N_F}\right) \Big/ \Delta t, \tag{17.4}$$

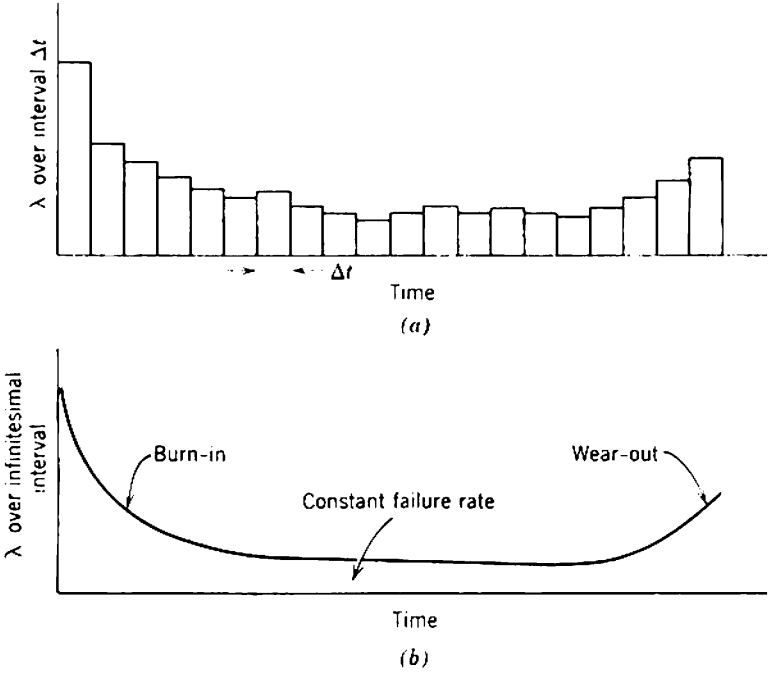


Fig. 17.11 Failure rate of typical equipment as a function of operating time: (*a*) form of experimental observations over finite time intervals; (*b*) theoretical idealization.

where ΔN_F is the number of failures during interval Δt, and $(N - N_F)$ is the number of operational units or *survivors* at time t. Numerical values of λ are often specified as a percentage failure rate per 1000 hours. Experimental data on failure rates are taken over finite time intervals Δt, and λ varies discontinuously as shown in Fig. 17.11*a*. In the limit, however, λ approaches a continuous function of time:

$$\lambda \to \frac{1}{N - N_F} \frac{dN_F}{dt}, \tag{17.5}$$

and the time variation has the general form shown in Fig. 17.11*b*. The high initial failure rate is known as the *burn-in period* which accounts for the early *infantile failures* due to design weaknesses or manufacturing imperfections. The burn-in period seldom lasts for longer than a few hours and should be exceeded in manufacturers' routine testing. Thereafter the whole of the equipment's useful life is characterized by a constant minimum failure rate λ. Although λ is small, it always has a finite value to account for occasional chance or *catastrophic failures* due to open-circuit resistors, short-circuit capacitors, breaking of bonds inside devices, etc. Eventually, the failure rate begins to rise again and *wear-out* is said to begin. Wear-out failures are a consequence of the wide-scale deterioration of aged materials. Immature designs often overstress some components, so that wear-out failure can be a serious factor. For more mature designs, however, such overstressing should not occur and wear-out failures should constitute only a small percentage of total failures. Moreover, if preventative maintenance is used to replace known short-life components such as microwave tubes, motor brushes, etc., wear-out is unlikely to be experienced before the equipment reaches obsolescence. Random catastrophic failures ought to account for more than 90% of the total failures in well-designed equipment.

The rate of catastrophic failures in a complete equipment depends on the number of individual elements and their failure rates, and on the number and character of the interconnections. If the equipment fails whenever any element fails, the individual elements contribute to the over-all failure rate as follows:

$$\lambda = \Sigma N_k \lambda_k \tag{17.6}$$

where N_k and λ_k are the numbers and failure rates of the various component types such as resistors, capacitors, transistors, and soldered joints. Thus, for the equipment failure rate to be low, particular attention should be paid to reducing the failure rates of the more numerous components.

In principle it is possible to determine the failure rates for all the elements used in any equipment and employ Eq. 17.6 to predict the over-all failure rate. In practice the difficulties of obtaining accurate and useful numerical data are all too obvious.

1. Large numbers of components must be life tested for long periods if reasonable confidence limits are to be achieved in the life-test data. This is expensive in power, time, and space and, very often, reliable data become available only when the unit is obsolete.

2. The failure rate of particular elements depends strongly on environment and any electrical, thermal, or mechanical stressing. For example, there may be a ten-fold increase in λ when a change is made from the ideal air-conditioned environment of a computer to normal laboratory conditions. A further tenfold deterioration may ensue when the environment changes to that of normal military conditions.

Published failure rates must be used with great caution, as they really only suggest the values of λ which can be achieved under certain conditions. It is essential that designers initiate schemes for feeding back information on the failure of operational equipment to see if the failure rates achieved for individual components are significantly worse than those claimed in published literature for supposedly similar environments. Table 17.2 lists failure rates which have been achieved under average laboratory conditions; rates for different environmental conditions are likely to vary by as much as two orders of magnitude.

Table 17.2 Failure Rates of Devices and Components, Given as a Percentage per 1000 Hours

Vacuum Tubes:		
Entertainment types		0.5
Professional types		0.02
Transistors:		
Entertainment germanium types		0.1
Professional silicon planar types		0.005
Capacitors:		
Mica		0.03
Paper		0.1
Electrolytic	aluminum foil	0.5
	tantalum foil	0.07
	tantalum solid	0.01
Resistors:		
Composition		0.03
Film		0.01
Wire-wound		0.2
Soldered Connections:		0.01
Wrapped Connections:		0.001

17.4.2 Exponential Law of Reliability

The standard definition of reliability R is the probability of nonfailure up to some time t; thus at time t

$$R(t) = \frac{N - N_F}{N} \tag{17.7}$$

and R always lies between 0 and 1. The variation of R with time may be obtained by integrating Eq. 17.5:

$$\int_0^t \lambda \, dt = \int_0^{N_F} \frac{dN_F}{N - N_F},$$

which yields

$$\frac{N - N_F}{N} = \exp(-\lambda t) \tag{17.8}$$

and, from Eq. 17.7,

$$R = \exp(-\lambda t) \approx 1 - \lambda t. \tag{17.9}$$

This exponential law of reliability emphasizes the great importance of failure rate. The time constant

$$m = \frac{1}{\lambda} \tag{17.10}$$

is known as the *mean time between failures* (sometimes abbreviated to m.t.b.f.), and for an operating time $t = m$ the magnitude of R drops from 1 to 0.37.

The reliability rating of a system must be based on experience. It is both theoretically and economically unsound to specify a reliability of 100%, because such a specification can be met only for a zero operating time. Usually the reliability is specified for a given operating period and this defines the failure rate required; for example, if an equipment is required to have a reliability of 90% for an operating period of 50 hours, the mean time between failure must be

$$m \approx \frac{t}{1 - R} = 500 \text{ hours.}$$

Moreover, if there are 1500 components and the failure of any one results in an equipment failure, the average component failure rate should not exceed

$$\lambda_k = \frac{\lambda}{N_k} = \frac{1}{mN_k} = 1.3 \times 10^{-6}/\text{hour}, = 0.13\%/1000 \text{ hours.}$$

Considerations such as these indicate the great difficulties involved in achieving high reliability in complex equipment over long periods. The

engineering approach to reliability improvement is to obtain information on which sections of the equipment have high failure rates and then to attend to such sections in the following ways:

1. Choose the most reliable components, devices, etc. from past experience, acceptance tests, or published failure rates.
2. Improve the electrical design by derating, using feedback, and minimizing the number of components by the use of simple proven circuits. Invariably, these steps conflict with each other and a compromise is necessary.
3. Improve the mechanical design to ensure adequate cooling, moisture protection, and shock protection.

It is important to realize that the aim of a reliability program is to *improve* reliability. Designers should not meekly preserve the status quo by assuming that accepted failure rates cannot be improved upon by more rigid inspection of all circuit elements. There is considerable evidence to suggest that accepted failure rates are dominated by the failure of a small proportion of the elements having constructional faults. If these can be separated from the majority of good quality elements, very substantial improvements in failure rate are possible. This again emphasizes the point that designers should look well beyond the incomplete picture presented by failure statistics.

If it is necessary to improve the reliability beyond that attained by careful inspection, selection of parts, derating, simplifying, and lowering environmental stresses, the designer must use *redundant* circuit techniques. In amplifier circuits, redundancy is achieved by making a number of alternative signal paths available. Almost all simple electronic circuits are nonredundant in the sense that the failure of any element causes the circuit to fail; under these circumstances, it follows from Eqs. 17.6 and 17.9 that the equipment reliability is the product of reliabilities for all elements, namely,

$$R = \Pi(R_k)^{N_k}. \tag{17.11}$$

In a redundant system it is the probabilities of failure (not the probabilities of nonfailure) which multiply. The probability of failure is $(1 - R)$ and for the system of Fig. 17.12 containing n identical redundant sections (parallel signal paths) of reliability R_n

$$1 - R = (1 - R_n)^n$$

or

$$R = 1 - (1 - R_n)^n. \tag{17.12}$$

Usually a redundant system has provision for switching the signal away

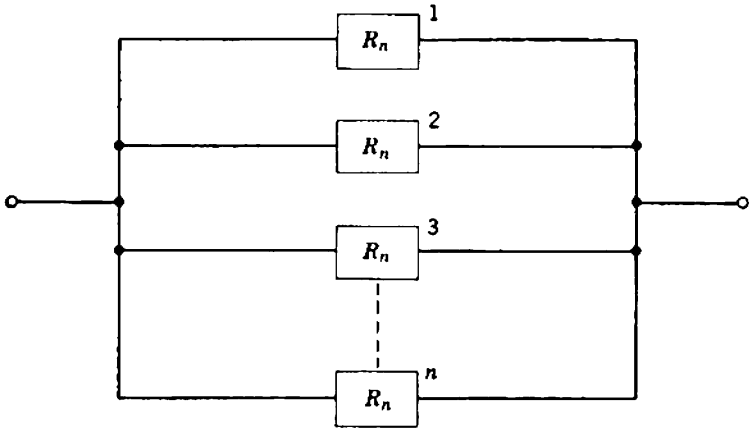


Fig. 17.12 Representation of a redundant (parallel-path) system.

from an inoperative section. The switching system itself has a finite reliability R_S which contributes to the over-all reliability so that

$$R = R_S[1 - (1 - R_n)^n]. \tag{17.13}$$

The realization of redundant systems of extreme reliability usually calls for much ingenuity from the designer.

17.4.3 Maintainability

Equipment such as that in missiles cannot be repaired or maintained and the probability of successful operation depends only on the actual mean time between failures and the mission time. In other instances, equipment can be maintained and an important consideration is the percentage time for which the equipment is available. A measure of availability is the *availability factor* A defined by

$$A = \frac{m}{m + m_R}, \tag{17.14}$$

where m_R is the *mean time to repair* the equipment. The availability factor can also be written as

$$A = \frac{\text{(up-time)}}{\text{(up-time)} + \text{(down-time)}}. \tag{17.15}$$

Numerical values of A are often expressed as percentages.

If a system is, on the basis of experience, required to have a 99.99% availability, with an estimated mean time to repair of 10 minutes, the mean time between failures needs to be 1667 hours. If this value of m cannot be

realized, the system can achieve the desired availability provided that steps are taken to reduce the mean time to repair. Such a reduction may be achieved by improving the design in some of the following ways:

(i) improve accessibility,
(ii) ease the removal of circuit sections and circuit elements,
(iii) label circuit elements clearly,
(iv) increase the number of monitoring points available,
(v) provide clear trouble-shooting schedules.

Reductions in the mean time to repair can be traded against the mean time between failures. Maintainability is thus a very important design consideration.

Practice Examples

These practice examples are designed to satisfy several different requirements and, consequently, the degree of difficulty or the time required for their solution varies considerably from one example to the next. Some of the examples are simple analyses designed to assist the reader in the initial understanding of theoretical developments. Others are aimed at filling gaps in the theoretical work either gaps in the detail of arguments outlined in the text or gaps due to the complete omission of some topics. The supreme example of complete omission is the important subject of measurements, and most chapters have examples designed to give insight into the more important techniques of measurement. A few examples are really literature research assignments which require the reader to seek additional information on such things as semiconductor device technology. Other examples serve the purpose of breaking down the necessarily arbitrary grouping of topics within chapters and so assist in broadening the reader's outlook to perceive the subject of amplifier design (or indeed electronic design) as a whole. Finally, and most important, many examples are straight-out design problems which, in the main, do not have unique solutions.

CHAPTER 1

1.1 Resistors and capacitors are available with nominal values which extend over the vast range of more than six decades. In the interests of standardization a range of *preferred nominal values* for each decade has been agreed upon. These preferred values are allotted so that there is no overlap due to the finite tolerances on each value. It is usual to assume a tolerance of $\pm 10\%$ about the nominal value and, to avoid overlap, adjacent preferred values M and N must be spaced to satisfy the requirement that

$$M + 0.1M = N - 0.1N.$$

Verify that this condition is achieved approximately when preferred values are taken as the first two significant figures of

$$10^{n/12}$$

where n is an integer.

Write out the $\pm 10\%$ preferred values for the decade from 10 to 100. Which of these preferred values should be used if component tolerances are relaxed to $\pm 20\%$?

The majority of circuit designs use components having the $\pm 10\%$ range of preferred values for nominal magnitudes, even if the actual component tolerances are much smaller (say, 1%). Nonpreferred values, in those rare cases when they are absolutely necessary, can be realized by connecting preferred-value components of appropriate tolerance in series or in parallel, although some of the more commonly used values (such as 50 and 75 Ω which are used to terminate coaxial cables) are available commercially.

Several other preferred-value series exist, and are appropriate when the tolerance is different from $\pm 10\%$. For example, there are series having 10, 20, 24, and 40 values per decade. What are the 24 values in the decade 10 to 100, appropriate for use with 5% tolerance components?

1.2 Sound reproducing systems often use two or more loudspeakers to cover different portions of the audio spectrum. Frequency-selective networks called *cross-over networks* are employed to route Fourier components of the signal to the appropriate loudspeaker. Usually, the cross-over networks comprise lossless LC combinations inserted between the amplifier and the loudspeakers. An alternative approach is to use RC cross-over networks in a low-level part of the system and to feed each loudspeaker from a separate amplifier.

To be specific, consider an audio system having two loudspeakers one each for the low- and high-frequency portions of the spectrum. Voltage signals v_{ol} and v_{oh} are to be applied to each loudspeaker in response to a low-level voltage signal v_i such that

$$v_{ol} = \frac{A_{vm}}{1 + s\tau} v_i$$

and

$$v_{oh} = \frac{A_{vm} s}{s + 1/\tau} v_i,$$

where A_{vm} is the mid-band voltage gain and $f_c = 1/2\pi\tau$ is the cross-over frequency.

(a) If the loudspeakers are of equal sensitivity, verify that aural addition of the two output signals results in a flat over-all frequency response.

(b) Show that the desired transfer functions v_{ol}/v_i and v_{oh}/v_i can be realized by combining the following elements:

a resistor R and capacitor C such that $RC = \tau$,
a voltage amplifier A_v of near-infinite input resistance,
a transfer resistance R_T of near-zero input resistance.

What constraints, if any, must be placed on the values of R, C, A_v, and R_T?

1.3 Figure E1.3 shows a representation of a nonideal amplifier having finite input and output resistances. Obtain expressions for the logarithmic power gain in

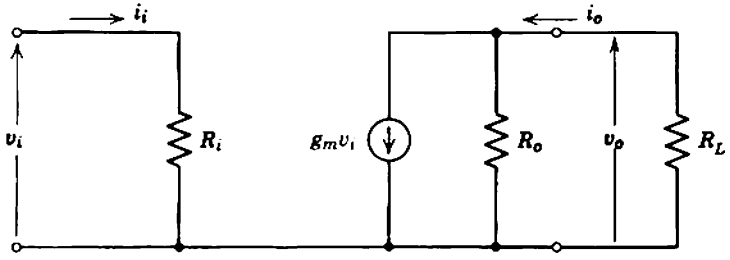


Fig. E1.3

terms of both the voltage gain v_o/v_i and the current gain i_o/i_i. Under what circumstances is the power gain equal to the square of the voltage or current gain? Is this special case ever important?

1.4 The frequency response of an amplifier can be determined directly by applying a test sinusoid to the input and measuring the magnitudes and phases of input and output signals. Usually, this simple test is limited in accuracy by differences in the calibration and/or frequency response of the measuring instruments. Furthermore, the input signal is often at a sufficiently low level for corrupting noise signals to intrude.

Such difficulties can be overcome by using an accurately calibrated attenuator to provide the magnitude information and a single measuring instrument to set two signal magnitudes to coincidence. If required, accurate phase measurements can be performed on the two signals of equal magnitude. A typical set-up for measuring the gain of a voltage amplifier is shown in Fig. E1.4. The attenu-

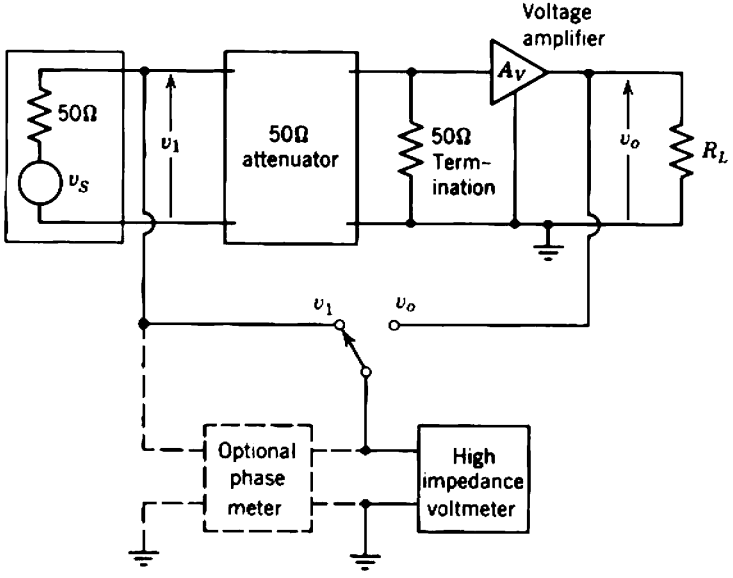


Fig. E1.4

ator is adjusted so that signal voltages v_i and v_o are equal and, since the attenuator is properly terminated, the attenuator reading is equal to the magnitude of voltage gain. (This assumes that the input resistance of the voltage amplifier is very much larger than 25 Ω.) It is essential to keep the signal levels within the extreme limits imposed by noise and distortion.

(a) If the attenuator reads 39.5 dB, what is the numerical value of A_V?
(b) Most attenuators have minimum steps of 1 dB or, more rarely, 0.5 dB. What is the percentage accuracy possible with the arrangement of Fig. E1.4? Suggest a method for improving this accuracy by using an additional calibrated decade resistor box or multiturn potentiometer.
(c) Outline three possible methods for measuring the phase angle between v_i and v_o and estimate the accuracy of each method.
(d) A transfer resistance amplifier has a nominal R_T of about 1 V/μA. The input resistance is expected to lie within the range of 60 to 200 Ω. Suggest a set-up for measuring the magnitude of R_T to an accuracy of 0.5%.

CHAPTER 2

2.1 A useful technique for experimental investigations of the performance of charge-controlled devices is to inject a current impulse (containing a known amount of charge) into the control electrode and observe the resulting output current. The current impulse can be obtained by applying a voltage step through the capacitor C as shown in Fig. E2.1*a*.

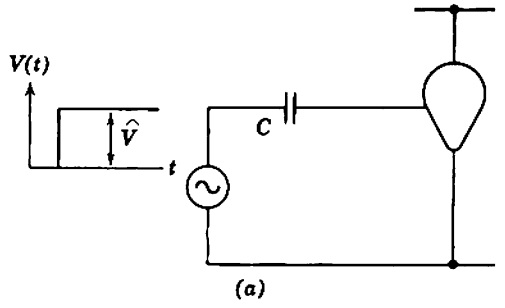
(a) Use any suitable argument to verify the possibility of ensuring that the injected charge is very nearly $C\hat{V}$. What restrictions, if any, must be placed on the magnitudes of C and $\hat{V}$?
(b) For a given amount of injected charge, the peak collector current pulse of a bipolar transistor varies with collector voltage in the manner shown in Fig. E2.1*b*. What is the functional variation of τ_1 with collector voltage? What evidence is there for assuming that τ_1 is independent of base voltage?
(c) It is found that the collector current transient in response to charge injection at $t = 0$ has the form

$$I_C(t) = I_C(0) \exp\left(-\frac{t}{\tau}\right).$$

If τ is 1 μsec, redraw Fig. E2.1*b* as a set of collector-current versus collector-voltage characteristics for various discrete values of base current.
(d) The input (base-to-emitter) voltage can be determined indirectly, but conveniently, by the scheme shown in Fig. E2.1*c*. The input charge is increased to twice that used in each of the tests leading to the results of Fig. E2.1*b* and the calibrated variable capacitor C_X is adjusted until the peak collector current is restored to its former value. In other words, C_X is adjusted until equal charges Q_C are injected into the base region of the transistor and the capacitor C_X. Then

$$V_B = \frac{Q_C}{C_X}.$$

If C_X is found to vary with Q_C as shown in Fig. E2.1*d*, determine and plot Q_C as a function of input voltage V_B.


V(t)
$\hat{V}$
t
C
(a)

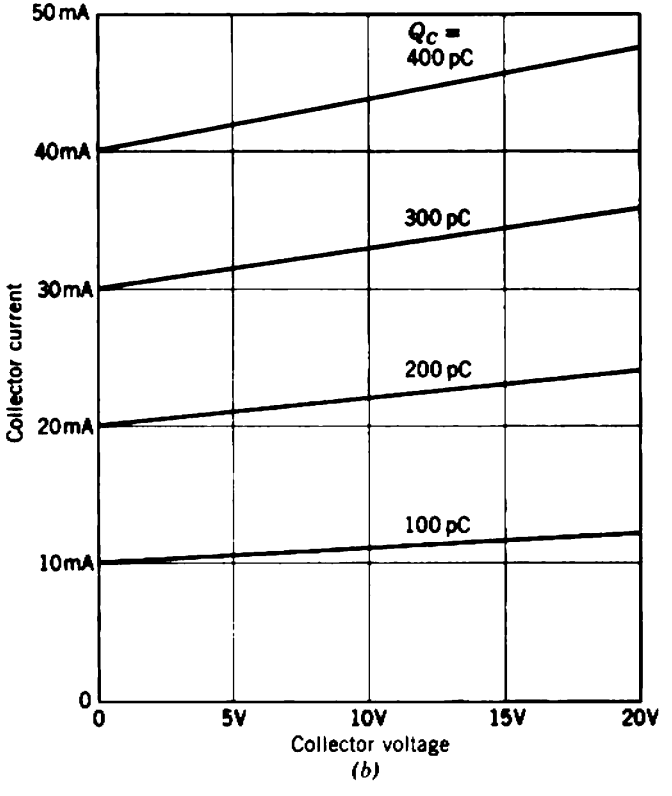
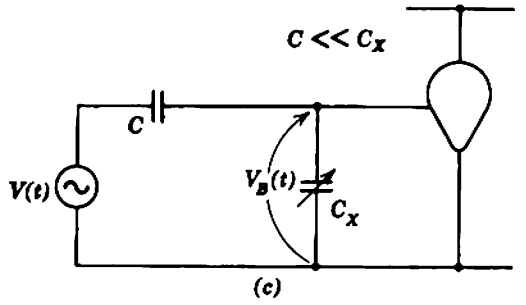
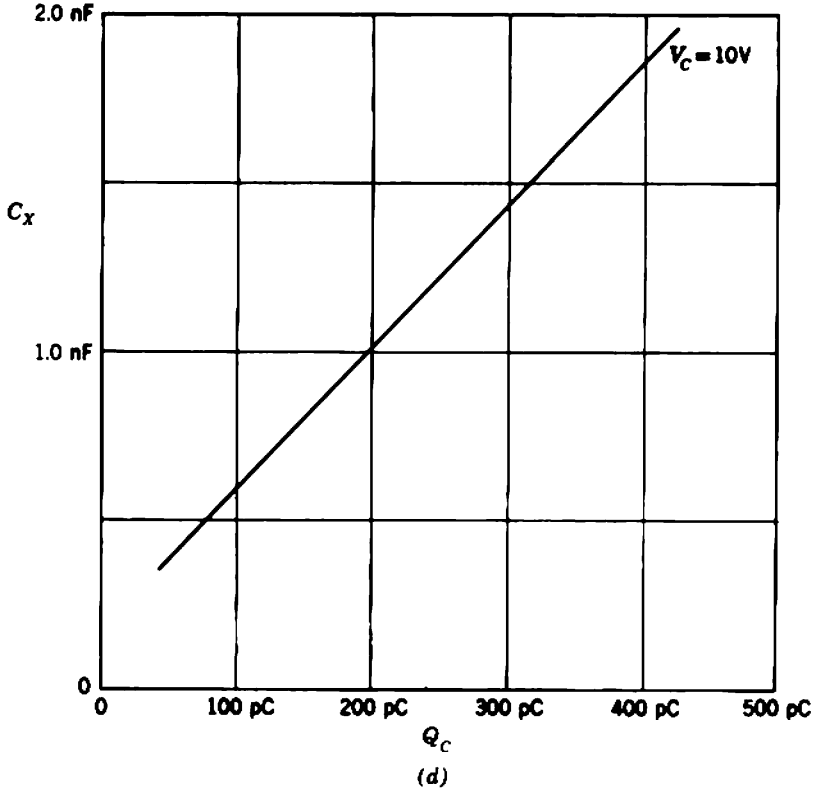

50 mA
40mA
30mA
20mA
10mA
0
Collector current
Q_C =
400 pC
300 pC
200 pC
100 pC
0
5V
10V
15V
20V
Collector voltage
(b)

Fig. E2.1



(c)



(d)

Fig. E2.1 (*continued*)

(e) Figure E2.1*d* corresponds to a collector voltage of 10 V; for a collector voltage of 20 V the $Q_C - V_B$ characteristic is shifted bodily to the right by 3 mV, whereas for V_C approaching zero the characteristic is shifted to the left by 3 mV. Use the preceding results to find the parameters of the small-signal equivalent circuit under the following quiescent conditions:

$$V_C = 10 \text{ V}, \qquad I_C = 10 \text{ mA}.$$

Note: The calculation of g_2 is difficult; you may *not* assume that

$$\left(\frac{\partial \tau_1}{\partial V_C}\right)_{V_B} = \left(\frac{\partial \tau_1}{\partial V_C}\right)_{Q_C}$$

2.2 Figure E2.2 shows a single-stage amplifier which uses a hypothetical charge-controlled device having negligible recombination. Current generator i_S and resistor R_S represent a signal source which is capacitively coupled to the control electrode; this form of coupling is necessary in order that the control electrode can be held steady at the quiescent voltage V_{11}.

The collecting-electrode supply voltage V_{22} is +200 V and V_{11} has a value such that the collecting-electrode current is 10 mA. What preferred value of R_L must be used for the quiescent collecting-electrode voltage to be about half V_{22}?

For the quiescent conditions specified, the device has the following small-signal parameters:

$$c_1 = 10 \text{ pF}, \qquad g_m = 5 \text{ mA/V}.$$

All other conductances and capacitances are negligible.

(a) If the bias resistor R_1 and source resistor R_S are each 100 kΩ and C_C is 1 μF, determine and plot the magnitude of $|v_o(j\omega)/i_S(j\omega)|$ over a range of 8 decades from 0.1Hz. What is the bandwidth of the amplifier? What is the mid-band transfer resistance v_{om}/i_{Sm}? Comment on the effect of C_C and c_1 on performance.

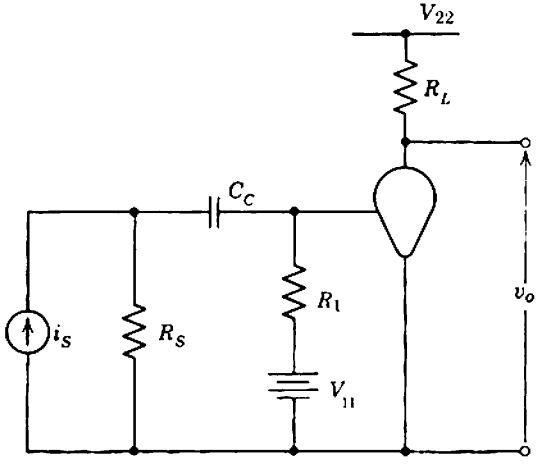


Fig. E2.2

(b) If the source signal is a small-amplitude step function at $t = 0$, use any suitable method to determine and plot the resultant waveform $v_o(t)$ from $t = 0$ to 5 μsec and from $t = 0$ to $t = 2$ sec.

(c) If the total resistance R in shunt with the input is the only significant source of noise, it follows from Chapter 8 that the mean-square output noise voltage is

$$v_{No}^2 \approx 4kTRA_{Vm}^2 \left(\frac{\pi}{2}\right) \mathscr{B},$$

where k ($= 1.37 \times 10^{-23}$ J/°K) is Boltzmann's constant,
T is absolute temperature,
A_{Vm} is the mid-band voltage gain,
$\mathscr{B}$ is the bandwidth in hertz.

What is the rms voltage at the output when no input signal is applied?

(d) From the previous results, estimate the approximate range of input signal magnitudes which can be amplified without excessive deterioration of the output waveform due to noise or nonlinearity.

(e) Is there any actual amplifying device which has characteristics like the hypothetical device assumed above?

2.3 A small germanium transistor has the following ratings:

maximum junction temperature = 75°C,
thermal resistance = 0.7°C/mW,
maximum collector voltage = 15 V.

At the low audio frequencies relevant to this problem, the junction temperature follows the time-varying power dissipated at the collector junction.

(a) Calculate the allowed peak collector power dissipation for an ambient temperature of 25°C.

(b) Plot the allowed operating region (bounded by the power dissipation hyperbola and collector voltage limit) on the I_C—V_C characteristics.

This transistor is used in the circuit of Fig. E2.3. The collector load impedance comprising an inductor and resistor in parallel ensures that the quiescent collector voltage is

$$V_{CQ} = V_{CC} = -7.5 \text{ V}.$$

Signal components of collector current having moderately high frequencies result in the development of a signal voltage V_L across the 1-kΩ load resistor. Assume that the transistor is perfectly linear and driven by a sinusoid so that the time-varying components of both I_C and V_L are sinusoidal; that is,

$$I_C = I_{CQ} + \hat{I}_{CS} \sin \omega t$$

and

$$V_L = \hat{V}_L \sin (\omega t + \phi)$$

where ϕ is the phase angle of the load impedance.

(c) Verify that instantaneous values of current and voltage are constrained to lie on an ellipse centered at the quiescent point on the I_C—V_C plot. This ellipse is called the *operating locus*.

(d) Calculate ϕ for a frequency of 1 kHz and then determine the value of I_{CQ} which gives maximum signal power to the load without the ellipse going

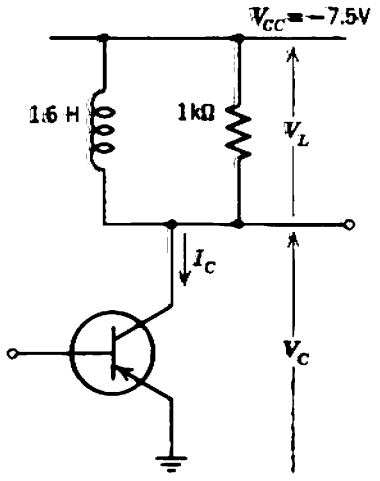


Fig. E2.3

beyond the allowed operating region of (b) above. What is the rms power delivered to the 1-kΩ load?

(e) If the transistor is biased to the same quiescent point as in (d) and has the maximum possible sinusoidal signal swing, superimpose the operating loci for frequencies of 100 and 10 Hz on the I_C—V_C plot. Comment on these results.

(f) What form does the operating locus take for frequencies well above 1 kHz?

2.4 In Section 2.5 it was stated that the forward transfer susceptance generator $j\omega c_{12}v_1$ in Fig. 2.7*a* is zero for vacuum triodes, bipolar, and field-effect transistors. This element was therefore omitted from the general, simplified equivalent circuit, Fig. 2.7*f*. Show that, if some new charge-controlled device is invented for which this generator is not zero, it must nevertheless be negligible compared with $g_m v_1$ over the frequency range for which the charge-control representation of the device is likely to be accurate. (*Hint*. What is the highest frequency at which charge-control theory is likely to apply accurately to a practical device?)

CHAPTER 3

3.1 Consult the manufacturer's data on the ECC82/12AU7 triode and plot the cathode current against the function $(V_A + \mu V_G)$. Use logarithmic scales, and hence determine the value of n in the expression

$$I_K = k(V_A + \mu V_G)^n.$$

The plot should cover the range of I_K between 1 and 60 mA.

3.2 In testing equipment for long-term performance, it is often useful to simulate a deterioration of tube characteristics with operating time by placing a small resistor R in series with the cathode, and to take the combination of "new" tube and resistor as representing an "aged" tube, Fig. E3.2.

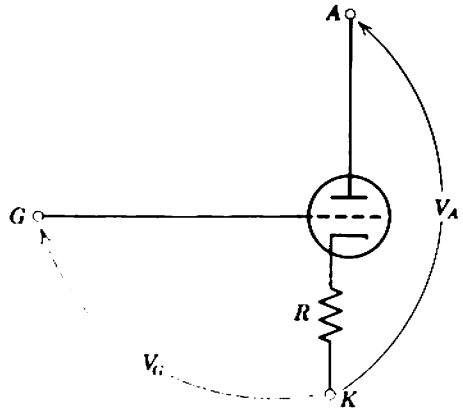


Fig. E3.2

(a) If the triode is half of type ECC82/12AU7 and $R = 470\ \Omega$, use suitable manufacturer's data for the "new" tube to determine the anode characteristics for the "aged" tube. Plot the two sets of characteristics on the same paper.
(b) Use these characteristics to determine g_m and r_A for "new" and "aged" tubes when V_A is 100 V and I_A is 5 mA.
(c) Is it possible to relate this method of simulating device deterioration with the physical processes actually involved?

3.3 Consult a manufacturer's data on the EF86/6BK8 pentode.

(a) Estimate a mean value for the ratio I_A/I_S and explain the method used in obtaining the result.
(b) Use the anode characteristics given for any one screen voltage as a basis to construct the anode characteristics for triode connection of the device ($V_S = V_A$, suppressor connected to cathode).
(c) Plot the manufacturer's quoted characteristics for triode connection on the same paper and compare with the results of (b) above.

3.4 The Miller bridge shown in Fig. E3.4 may be used for determining the voltage amplification factor of triodes. R_1 and R_2 are adjusted until no signal is heard in the headphones, and under these conditions it is possible to calculate μ from the known ratio between R_1 and R_2.

(a) Assuming that the oscillator signal is "small," redraw the above circuit with the tube equivalent circuit included and show how μ can be found.
(b) Describe bridge circuits for measuring r_A and g_m.

3.5 In discussing the necessity of adequately bypassing the screen grid of a pentode to the cathode, the *Radiotron Designers' Handbook* (4th ed., p. 498) states the following:

"An unbypassed screen resistor of $g_m r_S$ ohms gives the same degree of degeneration as a 1-ohm cathode resistor where

$$r_S = \text{dynamic screen resistance } \left(\frac{\partial V_s}{\partial I_s}\right)_{v_G}$$
$$\approx \text{triode anode resistance.}$$

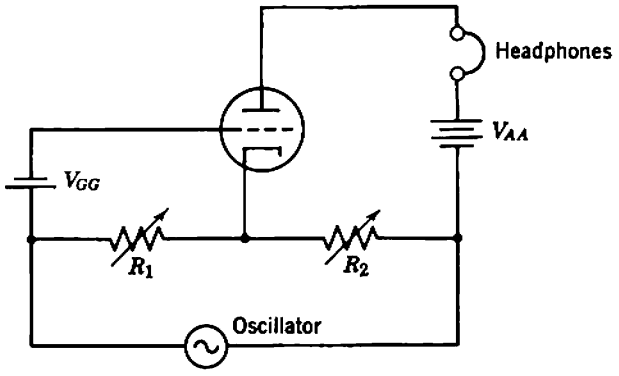


Fig. E3.4

For example, when $g_m = 2$ mA/V and $r_S = 10$ kΩ, an unbypassed screen resistor of 20 Ω will give the same degree of degeneration as a 1-Ω cathode resistor."

Discuss analytically the validity of this statement.

3.6 A change in the heater voltage of a vacuum tube changes the cathode temperature which, in turn, changes the position and potential of the barrier plane. Thus, changes in heater voltage change the anode current. It is convenient to represent the effect of a heater voltage change ΔV_H by including a voltage generator $k\,\Delta V_H$ in series with the cathode of an ideal tube as shown in Fig. E3.6*a*. The constant k is such that a 10% change in V_H produces a generator voltage of about 200 mV.

(a) Does the polarity of $k\,\Delta V_H$ shown correspond to an increase or a decrease in heater voltage?
(b) S. E. Miller has described a circuit (Fig. E3.6*b*) in which an extra auxiliary tube T_2 is used to compensate for heater voltage changes in the amplifying tube T_1. Replace the triodes by their small-signal equivalent circuits and then deduce the magnitudes required for resistors R_1 and R_2 in order that the output voltage V_o is independent of the heater voltage.
(c) What magnitudes are required for R_1 and R_2 when the triodes are identical in all respects?
(d) For the conditions of (c) above, determine the small-signal voltage gain v_o/v_i.
(e) Where and when was Miller's circuit first described?

3.7 Charge-control theory (Eq. 3.51) predicts a small positive anode-to-screen mutual conductance:

$$\left(\frac{\partial I_S}{\partial V_A}\right)_{V_A,V_S} = \frac{k}{r_A}.$$

In practice, a negative value is observed, as in Fig. 3.12. How does this error arise?

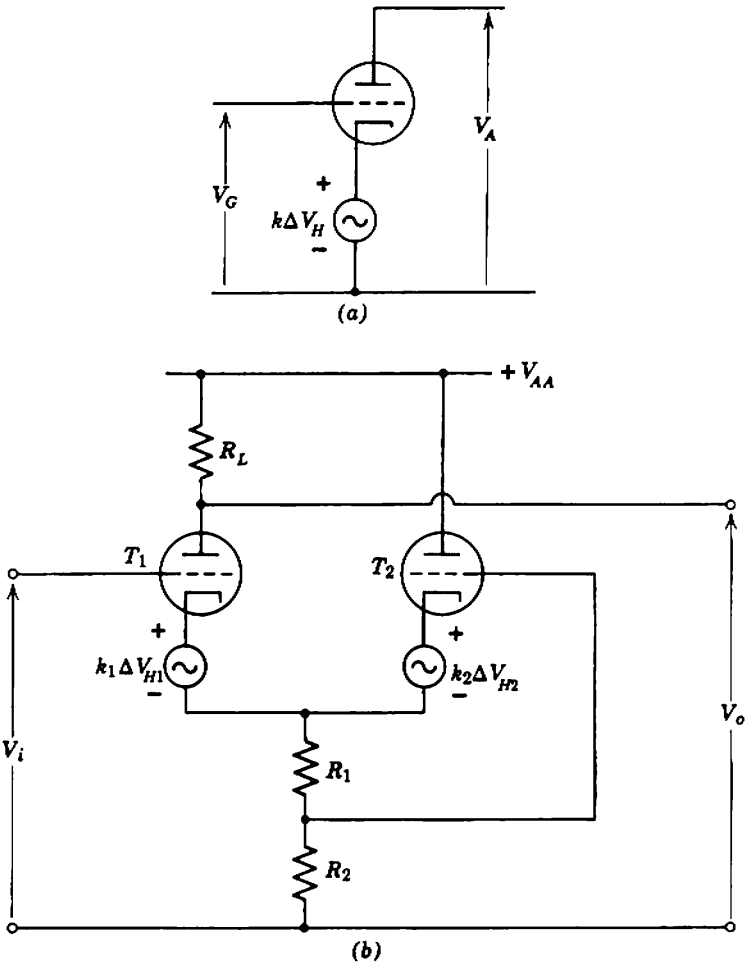


Fig. E3.6

CHAPTER 4

4.1 Verify that the equal and opposite donor and acceptor charges in the depletion layer of an abrupt *p-n* junction have magnitudes of

$$Q = q N_D A \frac{N_A/N_D}{1 + N_A/N_D} \left[2 \frac{\varepsilon}{q} \left(\frac{1}{N_A} + \frac{1}{N_D}\right)\right]^{1/2} (V_i - V)^{1/2},$$

where V is the applied voltage and A is the junction area. Thence obtain an expression for the incremental transition capacitance

$$c_t = \frac{\Delta Q}{\Delta V}$$

and compare with Eqs. 4.8 and 4.9.

4.2 A *punch-through* diode is a hypothetical device which could be used in voltage

regulator applications. At a particular reverse voltage, the depletion layer punches through the p region and the current increases to a value limited only by the external circuit.

(a) For the diode having the geometry shown in Fig. E4.2, determine the width, x, of the p region such that the diode regulates at -5 V. The p region resistivity is 10 Ω cm, whereas the n region resistivity is considerably less. Assume the following constants for the operating temperature of 25°C:

$$\varepsilon = 1.4 \times 10^{-12} \text{ farad/cm},$$
$$q = 1.6 \times 10^{-19} \text{ coulomb},$$
$$\mu_h = 1{,}800 \text{ cm}^2 \text{ V}^{-1} \text{ sec}^{-1},$$
$$V_i = 0.7 \text{ V}.$$

(b) Check that the maximum field is below the avalanche value, approximately 2×10^5 V/cm.

4.3 Transistors can be classified geometrically as being single-sided (with all leads connected to the same side of a supporting wafer) or double-sided (with emitter and collector leads connected to opposite sides of a supporting wafer). Several different production processes are used (or have been used) for realizing the geometrical types.

(a) Classify the following transistor types into single-sided or double-sided:

planar	diffused-base mesa
alloy	micro-alloy
rate grown	diffused-base, diffused-emitter mesa
epitaxial planar	micro-alloy diffused
grown diffused	double doped
surface barrier	

(b) Sketch the structure of each of the above types and comment on the accuracy of the representation assumed in Fig. 4.2. In particular, comment on the validity of using a fixed value for base resistance in the small-signal equivalent circuit of Fig. 4.9*b*. Which transistor types are likely to require the addition of further elements to this small-signal equivalent circuit?

(c) Outline the physical basis and major processing steps in the production of the above transistor types. Where possible, state the advantages and disadvantages of each process. Which processes yield the most advanced types of transistor?

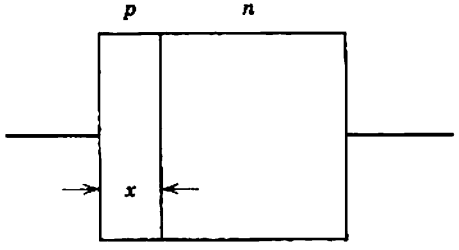


Fig. E4.2

4.4 Sketch the minority and majority carrier distributions in the base of a transistor under the following boundary conditions:

$V_{B'E}$ negative, $V_{CB'}$ negative;
$V_{B'E}$ positive, $V_{CB'}$ positive;
$V_{B'E}$ negative, $V_{CB'}$ positive;
$V_{B'E}$ positive, $V_{CB'}$ negative.

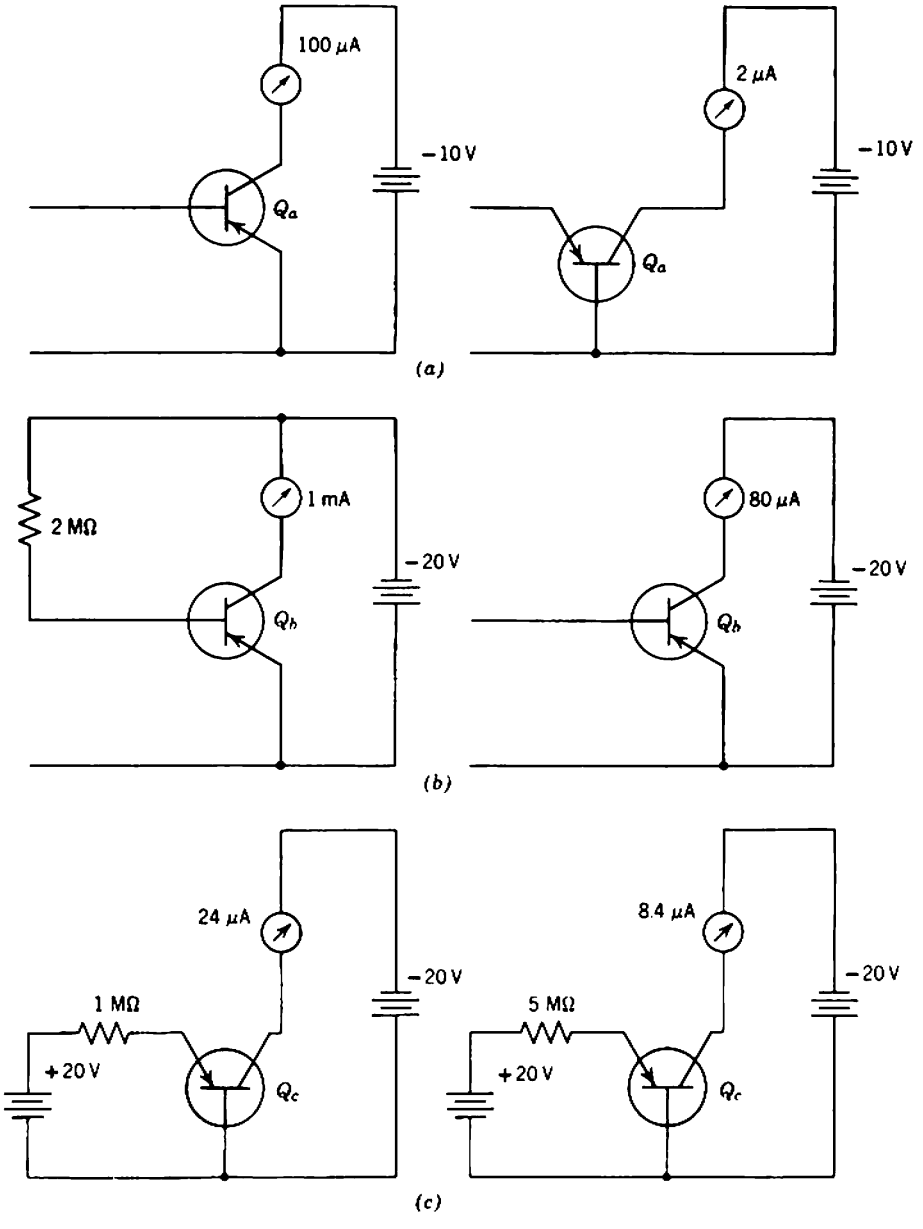


Fig. E4.5

For each condition, show the distributions for uniform-base and graded-base transistors, both *p-n-p* and *n-p-n*.

4.5 Figures E4.5*a*, *b*, and *c* indicate some possible dc tests which could be applied to *p-n-p* transistors Q_a, Q_b, and Q_c. In each case determine the parameters α_N, β_N, I_{CO}, and I_{CO}', noting any assumptions made. Indicate the actual directions of current flow.

4.6 The following information is extracted from a manufacturer's data on the type OC44 uniform-base germanium transistor. The operating conditions are 1 mA, 6 V, 25°C.

Parameter	Magnitude			Unit
	Minimum	Typical	Maximum	
V_{BE}	125	150	185	mV
I_{CO}	—	0.5	2.0	μA
r_{CE}	25	100	—	kΩ
$r_{B'C}$	2.0	—	—	MΩ
$r_{B'E}$	—	2.5	—	kΩ
$r_{BB'}$	40	110	250	Ω
g_m	—	39	—	mA/V
c_B	—	310	—	pF
c_{tE}	—	100	—	pF
c_{tC}	7.0	10.5	14	pF

(a) Are these values consistent with theoretical expectations?
(b) Estimate the values of all parameters for the following conditions:

(*i*) 1 mA, 6 V, 45°C,
(*ii*) 10 mA, 6 V, 25°C,
(*iii*) 1 mA, 3 V, 25°C,
(*iv*) 10 mA, 3 V, 45°C.

4.7 A 2N623 diffused-base transistor has a current gain which, according to the manufacturer, follows the empirical law

$$\frac{i_C}{i_E} = \frac{\alpha_N}{1 + j(2/3)(f/f_T)} \exp\left(-j\frac{0.3f}{f_T}\right).$$

Is this consistent with the quoted values of $f_T = 60$ MHz and $f_\alpha = 90$ MHz? Estimate the field parameter for this transistor.

Derive an expression for i_C/i_B from the above equation and plot the magnitude and phase as functions of frequency. Does $|i_C/i_B|$ obey the f_T/f law at high frequencies?

4.8 The base resistance of a transistor can be measured in a number of ways but the most useful methods are based on input impedance measurements or noise measurements.

(a) The input impedance of a common-emitter transistor (signal short-circuited at output) is measured with an impedance bridge to give the following results:

At very low frequency, $Z_i = (1730 + j0)\Omega$.
At $f = 200$ kHz, $Z_i = (900 - j830)\Omega$.

What is the value of r_B?

(b) If the transistor in (a) is operating at 2 mA, determine β_N and τ_T.

(c) A transistor is operated at a collector current of 0.5 mA so that

$$\beta_N r_E \gg r_B$$

and the base is signal short-circuited to the emitter. Show that the expected mean-square spot noise on the output current is

$$d(i_{No}^2) \approx 2qI_C\left(1 + 2\frac{r_B}{r_E}\right) df.$$

Assume that I_{CO} is negligible. If the mean-square output current noise (measured over a given bandwidth) is five times that expected from the collector current alone, determine r_B.

4.9 Figure E4.9 shows a simple test set for measuring f_T of transistors. The transistor is biased to desired values of emitter current and collector voltage and then excited by a small signal from a variable-frequency signal generator. An electronic millivoltmeter having an input capacitance (including strays) of 30 pF is connected as a detector. The signal frequency is varied until the millivoltmeter reading is the same for both positions of the single-pole switch.

(a) Verify that this signal frequency is simply related to f_T.

(b) What is the purpose of the 270-pF capacitor?

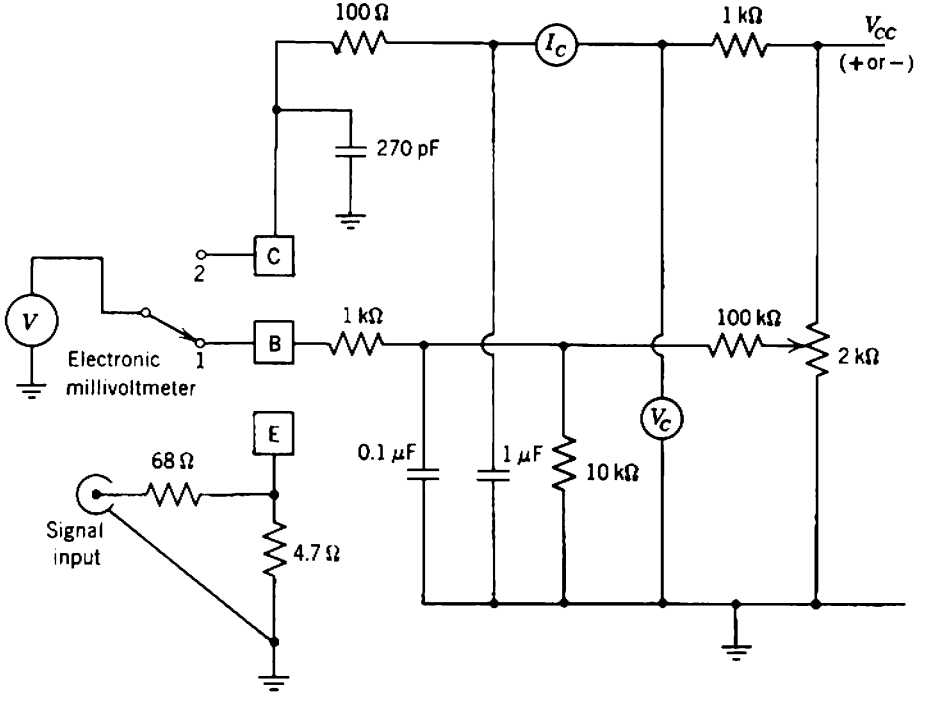


Fig. E4.9

(c) Estimate the approximate upper limit of f_T which can be measured with the basic unit shown, and suggest any modifications needed to accommodate 1 GHz devices.

(d) Show how a general purpose radio-frequency bridge can be used for measuring c_{tC}.

4.10 An insulated-gate f.e.t. is tested as shown in Fig. E4.10. The drain voltage is increased from zero to 10 V, at which $I_D = 1$ mA. Further increase in V_D does not change I_D appreciably.

(a) Deduce the I_D versus V_G transfer characteristic above pinch-off from the above information.

(b) What magnitude of drain voltage is required to obtain pinched-off operation at $V_G = +5$ V?

(c) What is the mutual conductance g_m under the conditions of (b) above?

4.11 Find the hybrid-π parameters of the following transistors at $I_E = 1$ mA, $V_{CE} = 6$ V, and $T = 16°C$.

OC71
T parameters at 3 mA/2 V/25°C

$$r_e = 6.5\ \Omega, \qquad r_b = 500\ \Omega,$$
$$r_c = 625\ \text{k}\Omega, \qquad \alpha = 0.979.$$

2N220
h parameters at 0.5 mA/4 V/16°C

$$h_{ie} = 3.57\ \text{k}\Omega, \qquad h_{re} = 9.44 \times 10^{-4},$$
$$h_{fe} = 65, \qquad h_{oe} = 25\ \mu\text{A/V}.$$

GET104
y parameters at 3 mA/10 V/16°C

$$y_{11} = 1340\ \mu\text{A/V}, \qquad y_{12} = 0.41\ \mu\text{A/V},$$
$$y_{21} = 79.2\ \text{mA/V}, \qquad y_{22} = 31\ \mu\text{A/V}.$$

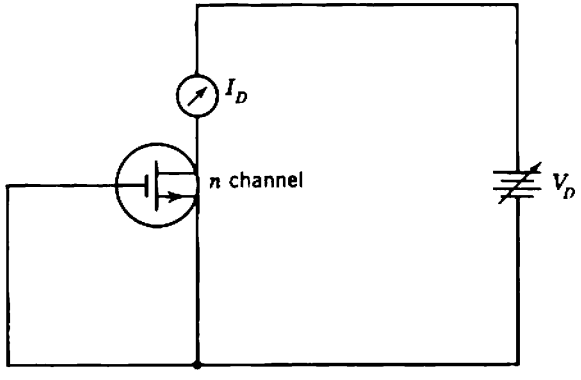


Fig. E4.10

CHAPTER 5

5.1 It is rewarding to have available data showing the dependence of gain precision on the nominal value of β_N, the tolerance on β_N, and the known value of r_E.

For the arrangement shown in Fig. E5.1*a*, Eq. 5.88 gives

$$\frac{i_{C2}}{i_{C1}} = \frac{\alpha_N R_P}{r_E}\left[\frac{1}{1 + (R_P + r_B)/\beta_N r_E}\right].$$

Obtain data to plot the variation of the correction term

$$\left[\frac{1}{1 + (R_P + r_B)/\beta_N r_E}\right]$$

normalized to

$$\left[\frac{1}{1 + (R_P + r_B)/\beta_{NA} r_E}\right]$$

as a function of β_N/β_{NA}. Plot the correction term vertically on a linear scale ranging from 0 to 200% and plot β_N/β_{NA} horizontally on a logarithmic scale ranging from 10 to 1000%. Show clearly the normally accepted limits of β_N/β_{NA} of 70 and 200%. Plot the relations for a number of realistic and usable values of the factor $\beta_{NA} r_E/(R_P + r_B)$. Make any pertinent comments on the form of the curves.

Are the foregoing results useful in estimating the precision of voltage gain v_o/v_S for the circuit shown in Fig. E5.1*b*?

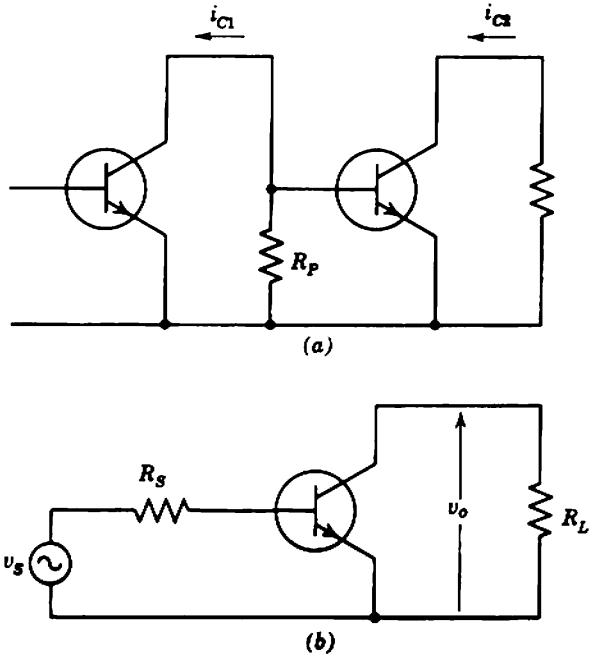


Fig. E5.1

5.2 An OC44 transistor (data given in Problem 4.6) is to be used as a small-signal voltage amplifier with a source of resistance 2 Ω as shown in elemental form in Fig. E5.2*a*.

(a) If the OC44 operates at 1 mA and 6 V, what is the maximum voltage gain v_o/v_S which can be achieved consistent with a gain stability of $\pm 20\%$? Use the complete equivalent circuit without approximation and assume that there is no dc-imposed upper limit to load resistance. (*Hint.* The upper value of R_L can be obtained by equating the ratio $(v_o/v_S)_{max} : (v_o/v_S)_{min}$ to 1.2 : 0.8.)

(b) Does alteration of the quiescent emitter current from $I_{E1} = 1$ mA to I_{E2} allow the achievement of greater nominal gain for the same gain stability? Follow (a) above but express the maximum allowed value of R_L as a function of I_{E2}/I_{E1}. Then plot the maximum and minimum limits of gain as a function of I_{E2}. What other factors influence the choice of I_{E2}?

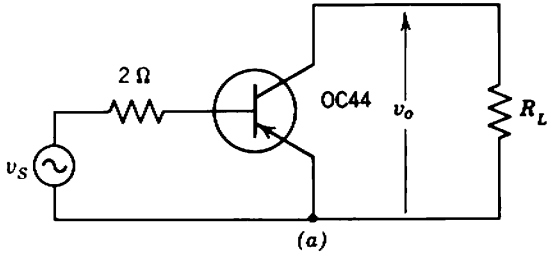
(c) If the source resistance is 2.2 kΩ instead of 2 Ω, what is the maximum gain v_o/v_S which can be achieved for a gain stability of $\pm 20\%$? What emitter current is used to achieve this gain?

(d) For the same source as in (c), what is the maximum stable gain v_o/v_S which can be achieved with two OC44's in cascade (Fig. E5.2*b*). Here the over-all (two-stage) gain stability should be $\pm 20\%$.

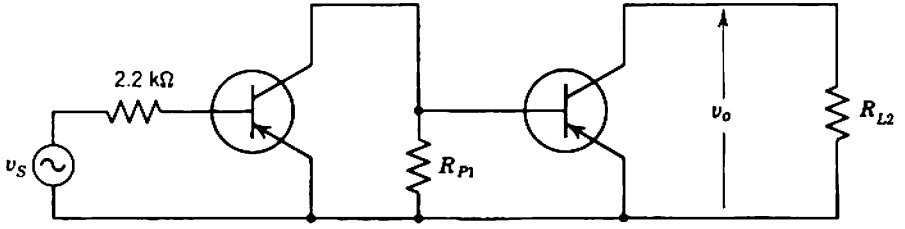
(e) Could the same order of gain stability be achieved if the source resistance is raised to 22 or 220 kΩ? Comment on the relative advantages of other types of device when the source resistance is large.

5.3 (a) Verify that the voltage gain of an emitter follower is given by

$$A_V = \frac{R_L}{(1 - \alpha_N)r_B + r_E + R_L}.$$



(*a*)



(*b*)

Fig. E5.2

(b) A two-stage amplifier requires an over-all gain stability of $\pm 20\%$. The source impedance and load resistance are both 1 kΩ, and collector supply resistors must be less than 10 kΩ to satisfy biasing requirements. If

$$\beta_{NA} = 100, \qquad 70 < \beta_N < 200,$$
$$r_{BA} = 100\ \Omega, \qquad 70\ \Omega < r_B < 200\ \Omega,$$

investigate the nominal voltage gain v_o/v_S which can be obtained with the circuits of Figs. E5.3*a* and *b*. (See also Problem 11.2.)

(c) Design a two-stage amplifier using the transistors of (b) above to meet the following specifications:

(i) $A_V = 300 \pm 20\%$.
(ii) Load resistance external to amplifier $R_{ext} = 3$ kΩ; no dc current should flow in R_{ext}.
(iii) $R_i \geq 3$ kΩ in worst case, including bias resistors.

5.4 Figure 5.4*a* shows a simple *p*-channel junction field-effect transistor stage, biased so that V_G is zero.

(a) Show that the mid-band voltage gain $v_o/v_i = -g_m/R_L$ has a maximum possible magnitude of

$$2\left(\frac{V_{DD}}{V_P} - 1\right)$$

which is independent of I_{DO}.

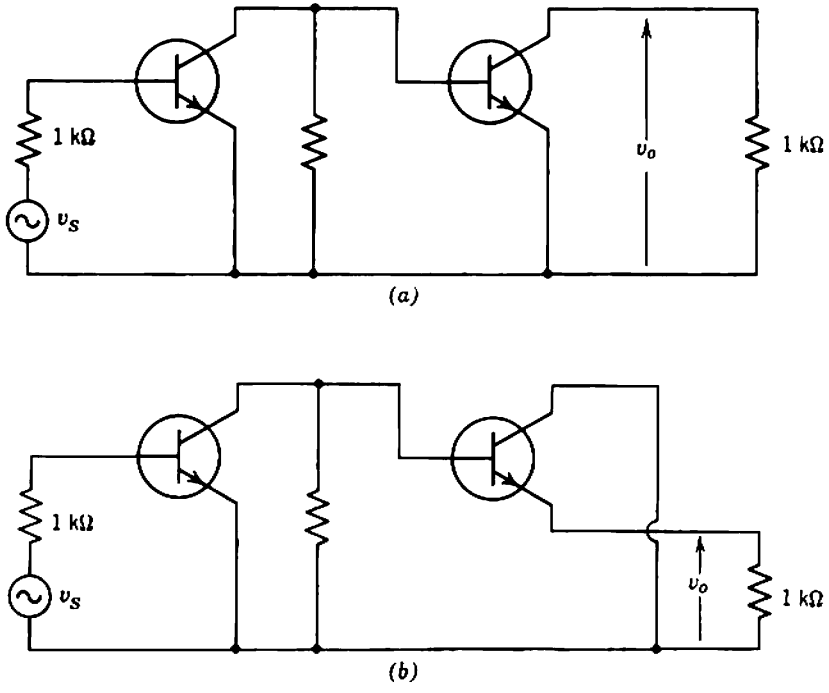


Fig. E5.3

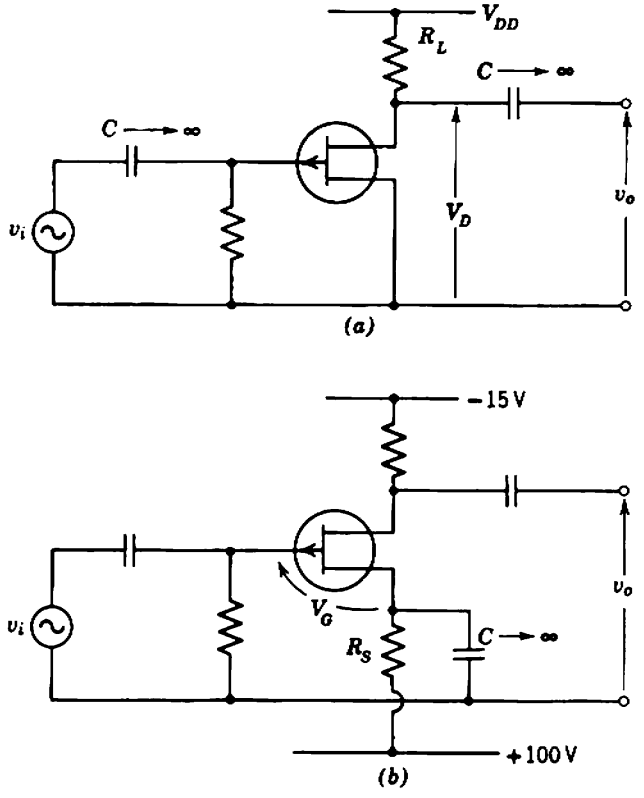


Fig. E5.4

(b) How does the value of I_{DO} affect the choice of R_L?
(c) The manufacturers of the type 2N2608 p-channel f.e.t. provide the following information:

	Minimum	Typical	Maximum	Units
Drain current at				
$V_D = 5$ V, $V_G = 0$	0.9	1.6	4.5	mA
Pinch-off voltage	1.5	2	4	V
Small-signal mutual conductance at				
$V_D = 5$ V, $V_G = 0$	1.2	1.6	—	mA/V

Estimate the maximum value of mutual conductance which can be expected.
(d) For this device, sketch the I_D versus V_D characteristic for the worst, typical, and best cases; show carefully the point at which pinch-off occurs. If the

device is used in the circuit of Fig. 5.4*a* with $V_{DD} = -15$ V, find the maximum preferred value of R_L which ensures that every device of this type number will operate in the pinch-off region. What are the minimum, typical, and maximum values of mid-band voltage gain?

(e) Compare these results with the maximum possible gains given by the expression obtained in (a). Is this expression of any practical importance?

(f) A considerable improvement can be obtained in both the magnitude and spread of gain by forcing the source current (hence the drain current) to a specified value, irrespective of the characteristics of the particular device in the circuit. One method of achieving this is shown in Fig. E5.4*b* where a high-voltage supply ensures that the source current remains almost constant. What is the magnitude required for R_S to keep I_D close to 0.9 mA? If I_D is held constant at 0.9 mA, find the expected spread in V_{GS}. Hence determine the expected spread in g_m.

(g) What is the largest preferred value for R_L which can be used? What is the magnitude and spread of voltage gain expected?

5.5 Design a single-stage amplifier using a pentode type EF86/6BK8 to have a nominal voltage gain of 100 and develop a signal of at least 56 V peak-to-peak across a 1 MΩ external load resistor. If the available supply is 200 V, investigate the effect of varying the screen voltage. What is the expected spread in voltage gain?

CHAPTER 6

6.1 The mutual characteristic of a pentode for $V_A = 250$ V and $V_S = 180$ V is specified by the following table:

I_A(mA)	V_G(V)
0	−6.0
1	−4.4
2	−3.6
3	−3.1
4	−2.6
5	−2.1
6	−1.7
7	−1.3
8	−1.0
9	−0.6
10	−0.3

On the average, the ratio of screen to anode current is 0.2, and μ for triode connection is 40.

(a) Determine and plot the family of mutual characteristics for other values of screen potential. Thence determine graphically the quiescent currents and voltages in the circuit of Fig. E6.1.

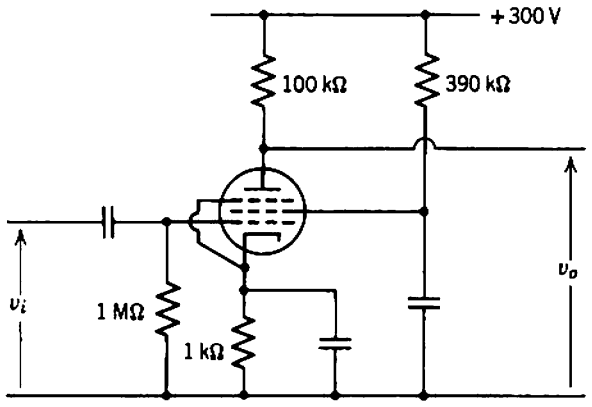


Fig. E6.1

(b) What is the mid-band voltage gain of this circuit?

6.2 A transistor has the following large-signal parameters:

$$V_{BE} = 0.2\text{ V} \quad \text{at} \quad I_E = 1\text{ mA},$$
$$\beta_N = 200,$$
$$I_{CO} = 10\,\mu\text{A}.$$

(a) Is it a germanium or a silicon device?
(b) Is it possible to self bias the device to an emitter current of 1 mA (Section 6.3.1.1)?

6.3 (a) Design a self biasing system to satisfy the following conditions:

$$V_{CC} = 12\text{ V},$$
$$I_E = 4\text{ mA nominal}, \pm 10\%,$$
$$R_C = 1.5\text{ k}\Omega,$$
$$\beta_N = 100,$$
$$I_{CO} = 2\,\mu\text{A},$$
$$V_{BE} = 0.2\text{ V}.$$

(b) Calculate the change in emitter current if
(i) β_N changes from 100 to 200, or
(ii) I_{CO} increases to 20 μA, or
(iii) V_{BE} increases to 0.3 V.

6.4 Design a self biasing circuit for a high-slope pentode, for example, ECF80/6BL8 or EF94/6AU6:

$$I_K = 10\text{ mA},$$
$$V_{SK} = 150\text{ V},$$
$$R_A = 4.7\text{ k}\Omega;$$

given

$$V_{AA} = V_{SS} = 250\text{ V}.$$

6.5 Calculate the quiescent currents and voltages in the two-stage combined emitter and collector feedback circuit shown in Fig. E6.5. The transistors have the following average parameters:

$$\begin{aligned} \beta_N &= 60, \\ I_{CO} &= 0.5\,\mu\text{A}, \\ V_{BE} &= 0.15\text{ V}. \end{aligned}$$

The maximum and minimum values of that transistor parameters (including production tolerances) over the temperature range 0 to 50°C are

$$\begin{aligned} 30 \leqslant \beta_N &\leqslant 200, \\ 0 \leqslant I_{CO} &\leqslant 10\,\mu\text{A}, \\ 0.05\text{ V} \leqslant V_{BE} &\leqslant 0.3\text{ V}. \end{aligned}$$

Comment on the stability of this circuit.

6.6 An interesting approach to transistor biasing circuit design is to arrange for the emitter current to be held constant by compensating for thermal changes in base-emitter voltage rather than fixing the emitter current with a high source resistance. The design problems associated with this technique can be investigated with the aid of a specific example, Fig. E6.6.

If the value of the emitter resistor can change with ambient temperature, determine its desired variation law over the range 15 to 60°C for constant $I_E = 1$ mA. The transistor has the following properties at 25°C:

$$\begin{aligned} \beta_N &= 50, \text{ proportional to } T^2, \\ I_{CO} &= 1\,\mu\text{A, doubles every } 8°\text{C}, \\ V_{B'E} &= 150\text{ mV, changes } -2\text{ mV every } 1°\text{C} \\ r_B &= 100\,\Omega, \text{ proportional to } T^2. \end{aligned}$$

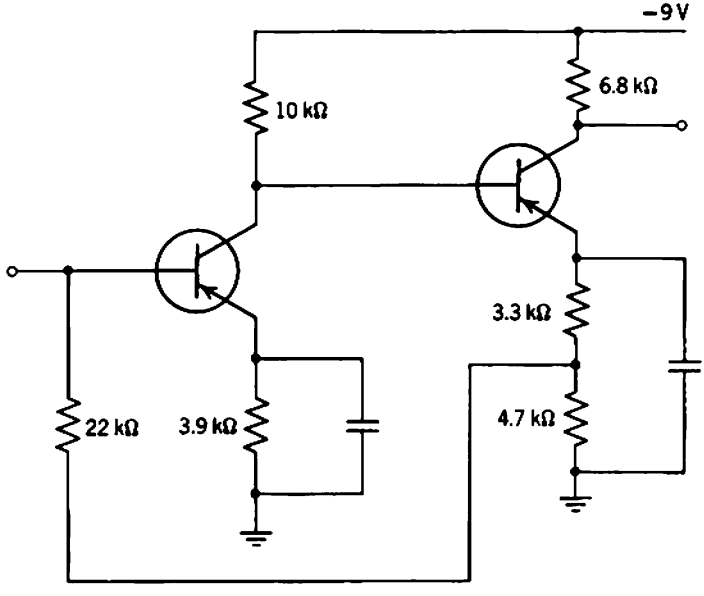


Fig. E6.5

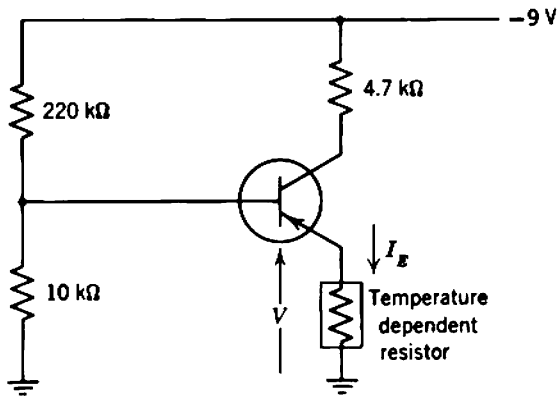


Fig. E6.6

Plot I_C as a function of temperature over the same range.

Suggest a physical realization for the temperature-dependent resistor. What is the advantage of this circuit over conventional emitter-feedback biasing? Suggest other ways of using temperature-dependent resistors in biasing circuits.

6.7 Calculate the quiescent conditions in the circuit of Fig. 14.43, given that the dc output from the rectifier is 350 V.

6.8 Design a collector-feedback biasing system to satisfy the following conditions:

$$\begin{aligned} V_{CC} &= -15\ \text{V}, \\ V_{BB} &= +10\ \text{V}, \\ I_E &\approx 4\ \text{mA nominal}, \\ R_C &= 2.2\ \text{k}\Omega, \end{aligned}$$

$$\frac{R_F R_D}{R_F + R_D} \geqslant 5\ \text{k}\Omega.$$

The nominal emitter current must be within $\pm 10\%$ of 4 mA and any additional change in emitter current (due to device tolerances and a temperature range of 0 to 50°C) should be less than $\pm 20\%$. Typical parameters are

$$\begin{aligned} \beta_N &= 70, & V_{BE} &= 0.3\ \text{V}, \\ I_{CO} &= 5\ \mu\text{A}, & r_B &= 100\ \Omega. \end{aligned}$$

Assume the tolerances given in Table 4.7.

6.9 In seeking to obtain the highest possible small-signal current gain from three transistors feeding a 1-kΩ load, a designer produced the circuit of Fig. E6.9. He claimed that, although the small-signal current gain was not stable (which did not matter in the application), the biasing circuit ensured that a stable quiescent current of 6 mA flowed in the output transistor and that this had been confirmed experimentally with several low-power transistors. Would the results claimed have been observed with germanium transistors and/or silicon transistors? What is the purpose of resistor R_3?

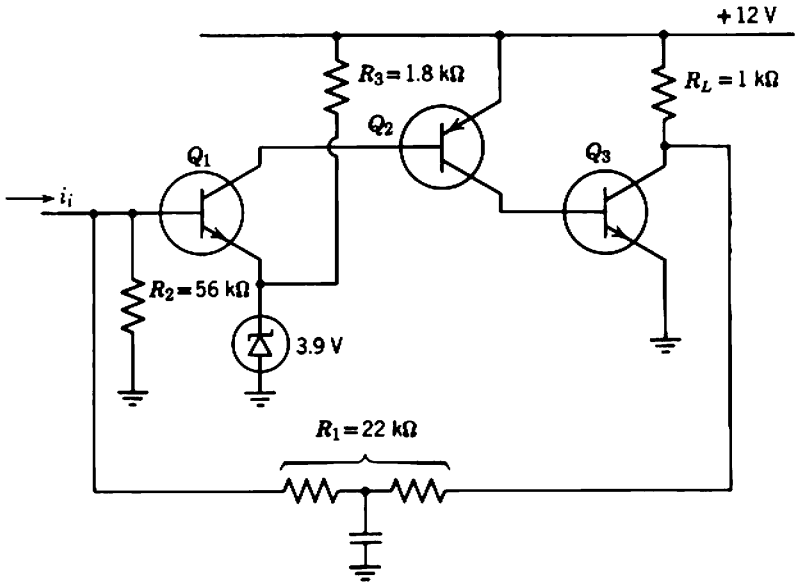


Fig. E6.9

CHAPTER 7

7.1 Figure E7.1 shows the coupling and bypass elements in a transistor amplifier stage. Data on the transistor are given in Problem 4.6.
(a) Obtain a literal expression for the low-frequency transfer function $v_o(s)/v_S(s)$.

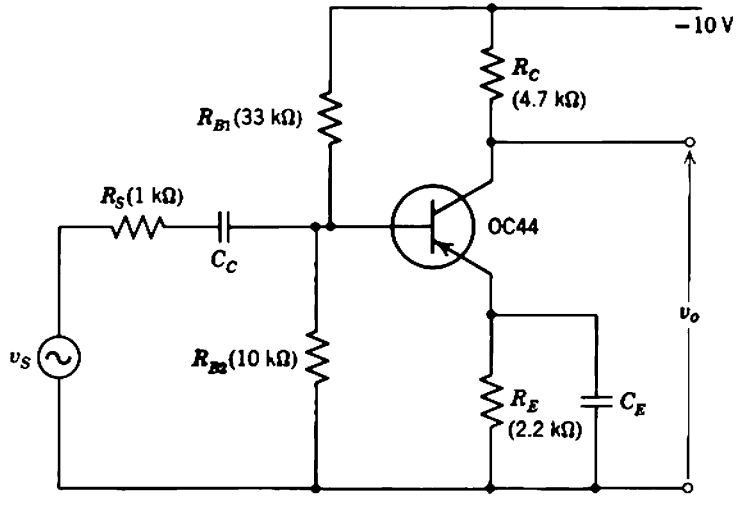


Fig. E7.1

Then comment on the validity of Eqs. 7.36, 7.37, and 7.38 for the three special cases discussed in Section 7.4.2.2.

(b) Is it possible for the poles to coincide?

(c) For the numerical values given in Fig. E7.1, find the values required for C_C and C_E to give a low-frequency half-power point at 50 Hz with minimum capacitor cost. It can be assumed that cost is proportional to the product (capacitance) × (voltage rating).

(d) If $v_S(t)$ is a 200-Hz square wave, determine the output waveform $v_o(t)$.

7.2 The base charging capacitance and depletion layer capacitances of some high-frequency transistors are so small that direct interlead capacitances become important. These are shown external to the dotted rectangle in Fig. E7.2.

(a) What is the cutoff frequency f_T of the transistor inside the dotted rectangle? Do the interlead capacitances change this for an actual transistor?

(b) If the transistor is fed from a source of resistance 1 kΩ and loaded with a 1-kΩ resistor, determine the mid-band voltage gain v_{om}/v_{Sm}.

(c) Compute and plot the magnitude of $v_o(j\omega)/v_S(j\omega)$ *exactly* up to the frequency at which the magnitude falls to 0.2 of the mid-band value.

(d) Assume that the effect of the two 0.5 pF feedback capacitors can be represented by a single 1.0 pF capacitor between either B and C, or between B' and C. For each case compute the magnitude of $v_o(j\omega)/v_S(j\omega)$ using the Miller capacitance approximation over the same frequency range as (c). Comment on the accuracy of the two approximations.

7.3 Calculate the nominal high-frequency singularities of the voltage gain $v_o(s)/v_S(s)$ for the circuit of Fig. E7.3 when OC44 transistors are used (data in Problem 4.6). What is the nominal mid-band gain and upper 3-dB frequency? How do these quantities vary with parameter tolerances?

7.4 Calculate the low-frequency singularities for the circuit of Fig. 12.17 and explain quantitatively how the low-frequency compensation is achieved.

7.5 Measurement of the pulse response of very-wide-band amplifiers is often difficult or impossible because of the restricted bandwidth of conventional

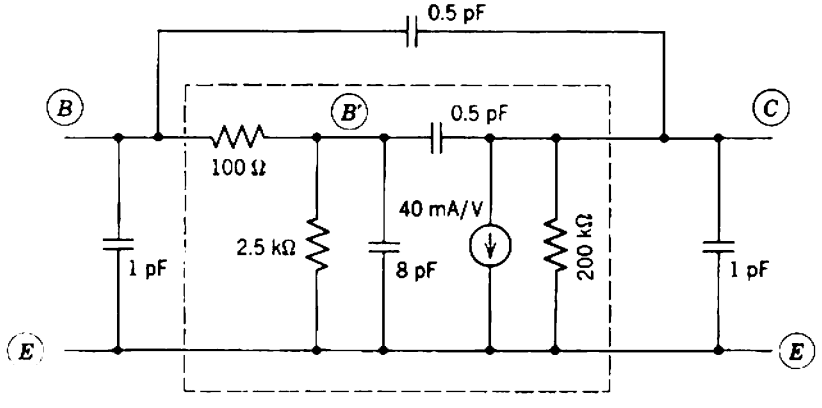


Fig. E7.2

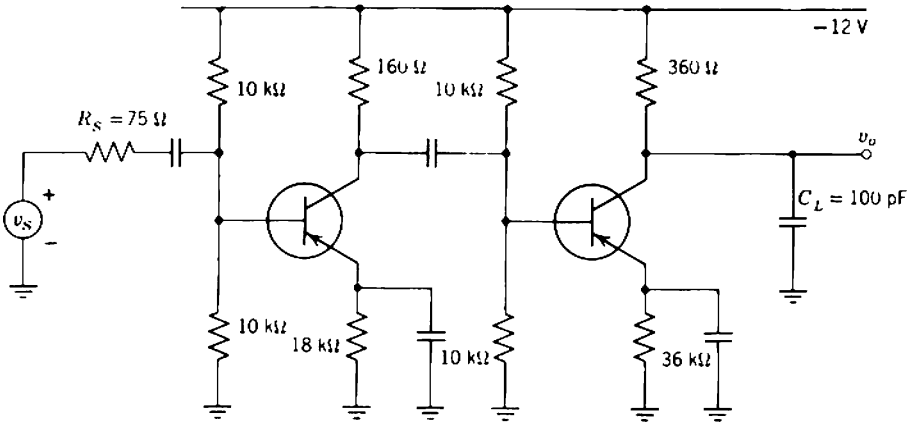


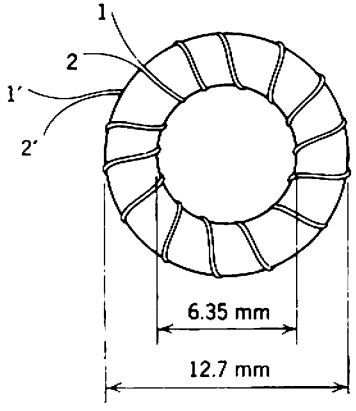
Fig. E7.3

oscilloscope amplifiers. This limitation can be overcome for the important special case of repetitive waveforms by using *sampling oscilloscopes.*

(a) Use block diagrams and typical waveforms to outline the mode of operation of a sampling oscilloscope and comment on the ultimate high-frequency limit.

(b) What is the maximum apparent bandwidth available in commercial sampling oscilloscopes and conventional (real time) oscilloscopes?

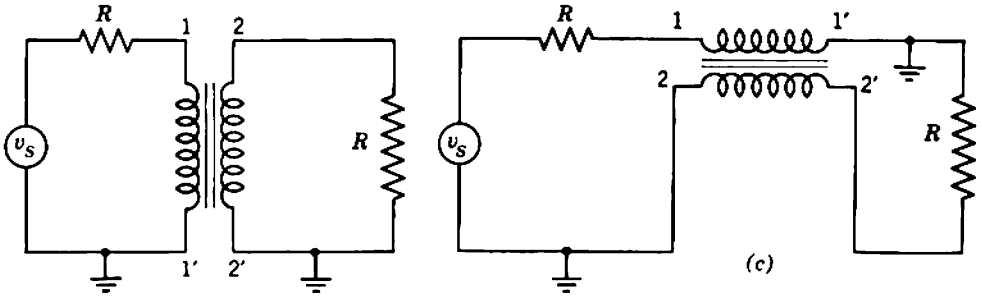
7.6 Figure E7.6*a* shows the winding detail of a special 1 : 1 transformer for use up to frequencies of hundreds of megahertz. The two windings 1–1′ and 2–2′ are bifilar-wound to ensure close coupling. The interwinding capacitance is 15.3 pF.

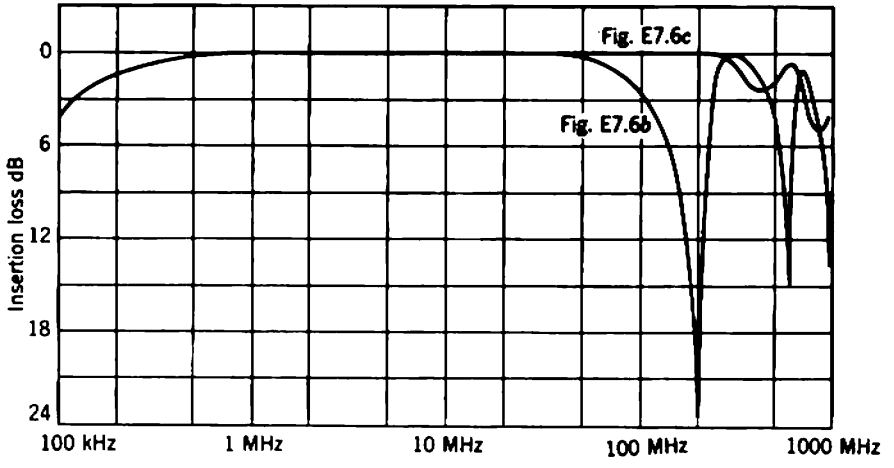


Thickness 3.17 mm

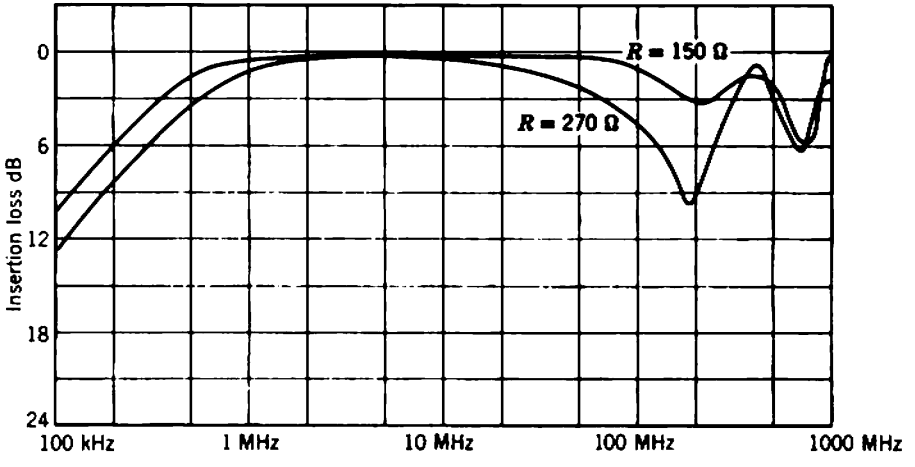
(a)

Fig. E7.6





(d)



(e)

Fig. E7.6 (*continued*)

When connected as shown in Fig. E7.6*b* (with terminating resistors R of 50 Ω) the insertion loss has the frequency variation given by the appropriate curve of Fig. E7.6*d*. Rearrangement of the two-port as shown in Fig. E7.6*c* results in an extended high-frequency response (Fig. E7.6*d*).

With the improved arrangement of Fig. E7.6*c*, variation of resistors R to 150 and 270 Ω results in the insertion-loss curves of Fig. E7.6*e*.

Explain the observed phenomena.

CHAPTER 8

8.1 In vacuum-tube amplifiers, two triodes are often connected as a *cascode* (Fig. E8.1) when the high voltage gain and low Miller capacitance of a pentode are desired but the partition noise of a pentode cannot be tolerated. Assuming identical triodes, show that the equivalent tube has the following parameters, approximately:

amplification factor = μ^2,
mutual conductance = g_m,
anode resistance = μr_A,
input capacitance = $c_{GK} + 2c_{AG}$,
spot noise voltage generator = $d(v_N^2)$,
spot noise current generator = $d(i_N^2)$.

8.2 A certain transducer has an output voltage which is directly proportional to the input excitation provided that the load resistance is not less than 50 kΩ. Thus, it is necessary to connect this transducer to an amplifier whose input resistance exceeds 50 kΩ; to be safe take the required input resistance as 100 kΩ.

Three possible input stages are shown in the elemental circuits of Fig. E8.2*a*; note that it is not necessary to connect one side of the source to ground. The transistors have the following parameters:

$r_B = 100\ \Omega$, $\quad I_{CO} = 0.5\ \mu A$,
$\beta_N = 100$, $\quad I_E = 0.5$ mA.

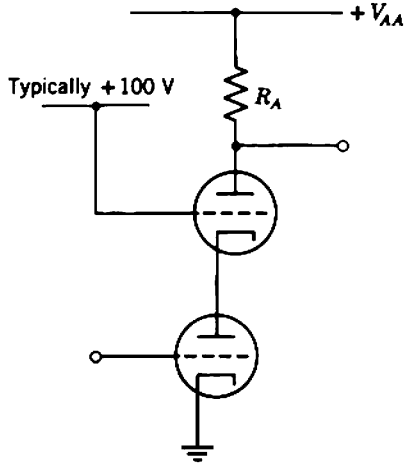


Fig. E8.1

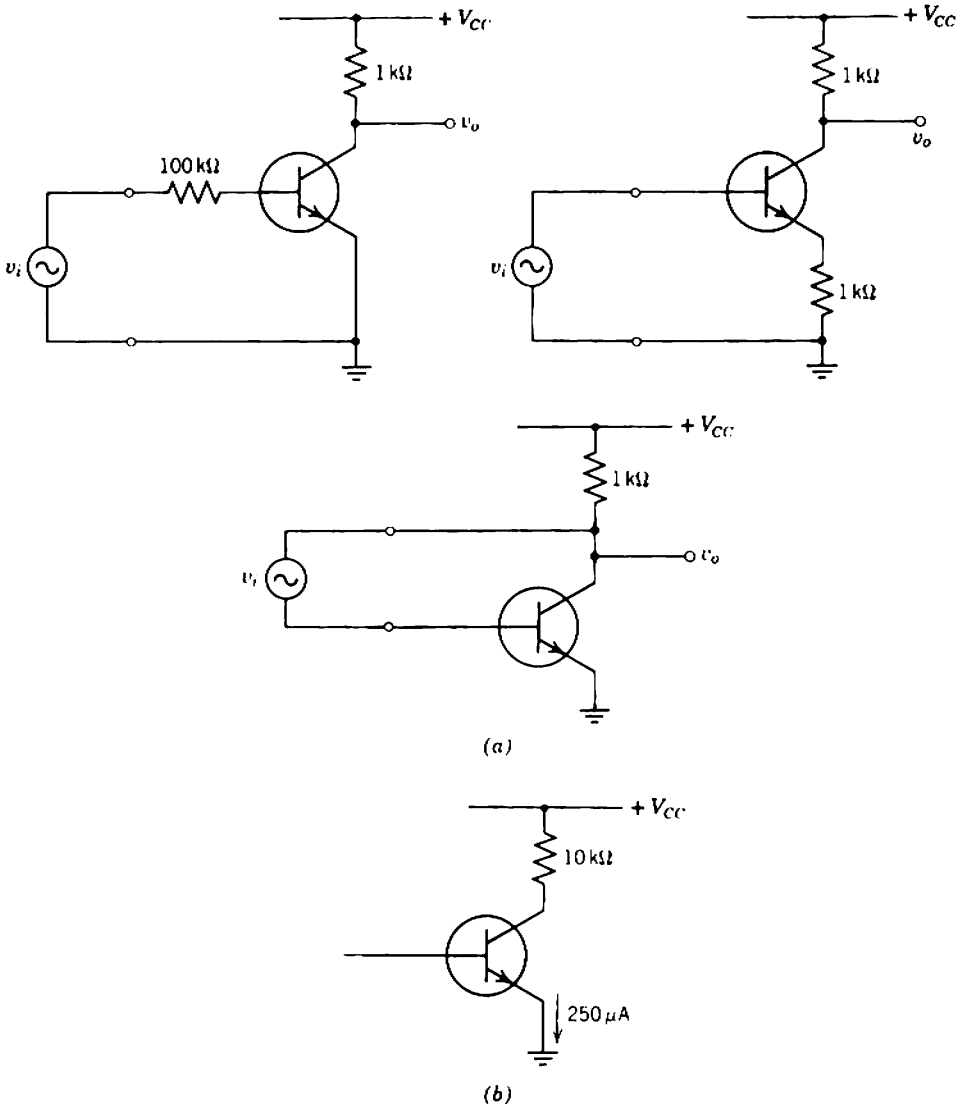


Fig. E8.2

(a) Verify that the input resistance is approximately 100 kΩ for each circuit. Comment on the effect of changes in β_N.

(b) Determine the mid-band voltage gain (v_o/v_i) for each circuit. Comment on the effect of changes in β_N.

(c) Superimpose all physical white noise generators (for the transistor and the resistors) on the equivalent circuit and calculate their contributions to the spot noise at the output of each circuit. Comment carefully on the dominant noise sources in each circuit and state any pertinent conclusions.

(d) If the first stage is followed by the voltage amplifier shown in Fig. E8.2*b*, would you expect any significant contribution to noise from this second stage?

8.3 A transistor has nominal values of β_N and r_B of 100 and 110 Ω, respectively, and is to be used as a low-noise input stage of an amplifier whose frequency response is such that neither flicker nor high-frequency noise is important. If the source resistance is 2 kΩ, determine the optimum emitter current required for minimum relative increase in noise. What is the optimum noise factor for this source resistance? Assume that all noise originates in the input stage, and that the amplifier has simple 20 dB/decade roll-offs at upper- and lower-break frequencies of 10 kHz sec and 200 Hz, respectively. If the mid-band input signal magnitude is 1 mV, determine the output signal-to-noise ratio.

8.4 Assume that an amplifier is operated as in Example 8.3, but the transistor has flicker noise resulting from depletion layer recombination so that an additional noise current generator exists of magnitude

$$\frac{d(i_N^2)}{df} = \frac{KI_C^2}{f},$$

where $K = 2.5 \times 10^{-11}$ for I_C in amperes. What is the new signal-to-noise ratio for an input signal of 1 mV?

8.5 The theory for optimizing the signal-to-noise ratio of a transistor stage fed from a resistive source (Section 8.5.1.1) is simplified by assuming that the small-signal and large-signal current amplification factors (β_N and $\bar{\beta}_N$, respectively) are identical. Obtain more general expressions for the relative increase in noise and the optimum emitter current that are valid when β_N and $\bar{\beta}_N$ differ substantially in magnitude.

8.6 Measurements of device spot noise generators involve the use of a high-gain, narrow-band amplifying system which has facilities for gain calibration and

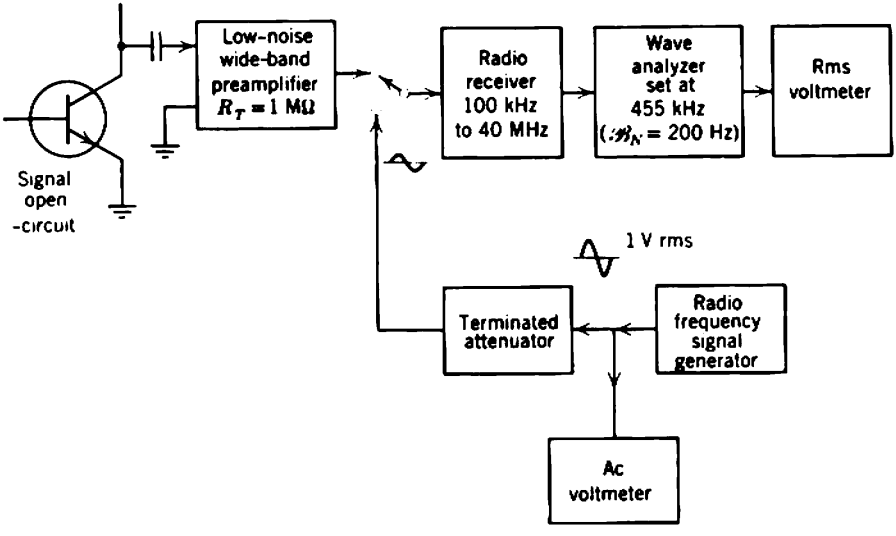


Fig. E8.6

tuning over a range of frequencies. A typical measuring system is shown in Fig. E8.6. The amplifying system consists of a wide-band preamplifier of transfer resistance 1 MΩ, followed by a radio receiver which is tunable over the range 100 kHz to 40 MHz. The system noise bandwidth is defined at 200 Hz by the wave analyzer. The gain of the receiver and analyzer is calibrated by feeding a sinusoid of known magnitude through the attenuator.

In tests performed on a silicon planar transistor with its base ac open-circuited, the signal generator output was 1 V rms and the attenuator was adjusted so that the rms output meter read the same for either position of the switch. The following results were obtained:

Frequency (MHz)	Attenuation (dB)
0.1	50.5
0.4	51
1.0	54
2.0	58
3.0	61
4.0	63.5
7.5	66
10	67
15	70
20	71
30	72
40	72

(a) Determine and plot the spectrum of the spot noise current output for the transistor.

(b) Is this consistent with the following measured operating conditions and parameter values?

$$I_C = 0.9\ \text{mA}, \quad \beta_N = 105,$$
$$I_B = 12\ \mu\text{A}, \quad f_T = 95\ \text{MHz}.$$

(c) What additional tests are required to characterize noise performance of the device completely?

8.7 Figure E8.7 shows the first stage of an amplifier fed from an inductive source. The complete system has a mid-band voltage gain of 10^4 and two coincident high-frequency poles at 20 kHz. The input transistor has a constant value of β_N of 100, negligible I_{CO}, r_B small in comparison with 1.2 kΩ, and an f_T of a few hundred MHz.

(a) If flicker noise is neglected, what is the optimum emitter current for the first stage?

(b) What are the rms values of output noise due to the transistor and the source resistor?

(c) How would the presence of flicker noise affect the optimum value of I_E?

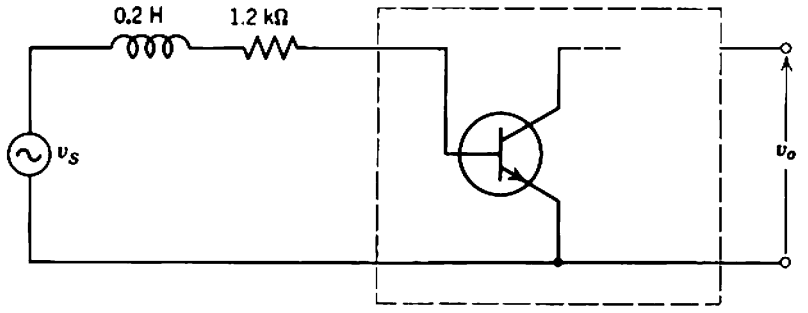


Fig. E8.7

8.8 The discussion of the nuclear pulse amplifier in Section 8.7 neglects flicker noise. Show that, if the tube has flicker noise on its anode current, then the optimum shaping time constant τ remains unchanged, but that the number of ion pairs noise is increased. Assume

$$d(i_N'')^2 = g_m^2\left(4kTR_N + \frac{A}{f}\right) df.$$

A selected type 6AK5 has the following parameters when connected as a triode:

$$\begin{aligned} I_K &= 10 \text{ mA}, \\ g_m &= 6.6 \text{ mA/V}, \\ R_N &= 380\ \Omega, \\ \Sigma\,(I_G) &= 7 \times 10^{-9} \text{ A}, \\ A &= 4 \times 10^{-13} \text{ V}^2. \end{aligned}$$

Calculate the optimum shaping time constant, and the degradation in noise performance due to flicker noise when $(C_S + C_i)$ is 50 pF.

8.9 Comment on the statement (in Section 8.4.3) that $d(v_{Ni}^2)$ and $d(i_{Ni}^2)$ are the spot noise voltage present and the spot noise current that actually flows in the input impedance of an amplifier.

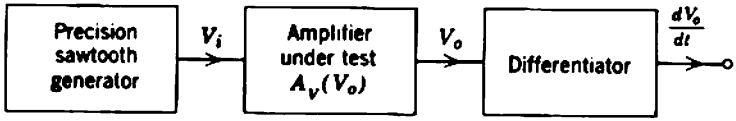
CHAPTER 9

9.1 The differential error of wide-band amplifiers can be measured conveniently with the arrangement shown in Fig. E9.1. For a voltage amplifier, the input voltage V_i is a linear sawtooth waveform, given between limits $-\hat{V}_i < V_i < \hat{V}_i$ by

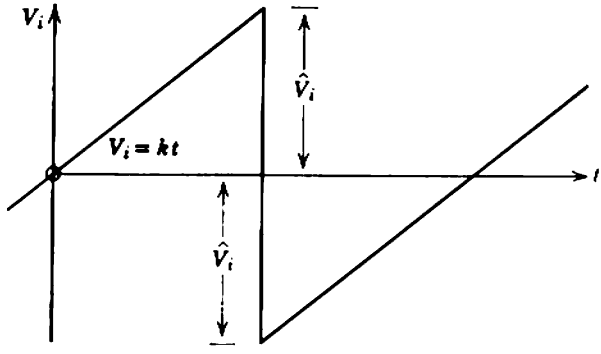
$$V_i = kt.$$

$\hat{V}_i$ usually corresponds to the mid-band dynamic range of the circuit. The factor k is ideally a constant and, in practice, its percentage variation with V_i needs to be at least an order smaller than the expected differential error.

(a) Show that an oscillographic display of dV_o/dt versus V_i or V_o can be scaled as a plot of differential error.



(a)



(b)

Fig. E9.1

(b) If the amplifier has a single high-frequency pole at radian frequency $-1/\tau_H$, determine and sketch the nonlinearity of the sawtooth arising from the finite frequency response. Hence show that this source of error is less than 1% for sweep times greater than 100 τ_H.

(c) Similarly determine and sketch the nonlinearity resulting from a low-frequency pole at radian frequency $-1/\tau_L$ and show that this source of error is less than 1% for sweep times of less than 0.02 τ_L.

(d) A simple *RC* circuit can usually be used for the differentiator. Investigate the effect of such a circuit on system performance.

9.2 (a) Occasionally, two or more stages in an amplifier contribute significantly to the nonlinearity. Verify that the total differential error at any point on the transfer characteristic is

$$\gamma_{tot} \approx \gamma_1 + \gamma_2 + \gamma_3 + \cdots.$$

(b) Figure E9.2 shows in elemental form the two-stage transistor amplifier described in Section 5.5.4. Use the normalized graph of Fig. 9.5 *twice* to determine the total differential error as a function of the instantaneous signal input voltage, and verify that nonlinearity in the second stage dominates:

(i) determine γ for G_T of Q_1,
(ii) determine γ for v_{C2}/i_{C1}.

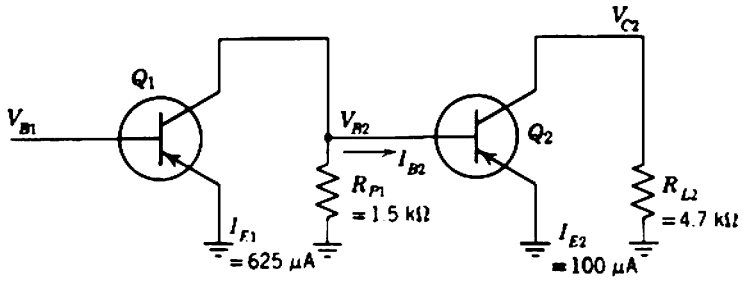


Fig. E9.2

(c) Hence calculate the rms value of the input sinusoid which produces 5% total harmonic distortion.

(d) Repeat (c) assuming that all nonlinearity originates in the transfer function v_{c2}/i_{c1}; use the approximate Eqs. 9.23, and the known small-signal voltage gain of the amplifier. What can you conclude about practical calculations of nonlinearity?

9.3 Figure E9.3 shows a form of two-stage amplifier which employs emitter coupling between stages. The quiescent input potential V_{IQ} is adjusted so that equal quiescent emitter currents flow in the two transistors.

(a) Use any suitable argument to estimate the clipping points in the $V_i - V_o$ characteristic. Then verify that

$$I_{E1} + I_{E2} \approx 1\ \text{mA}.$$

over the range of moderate nonlinearity.

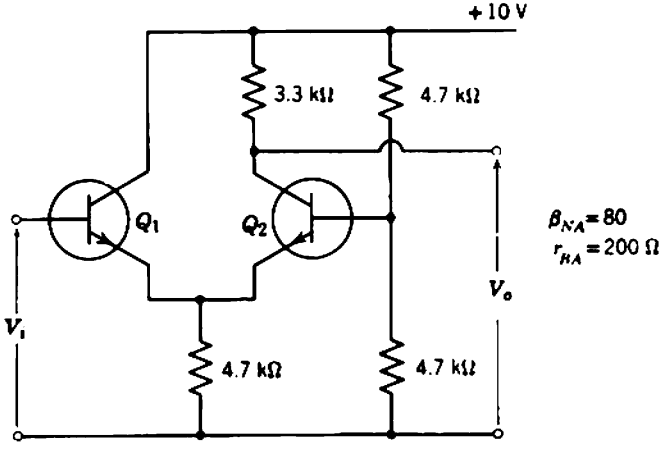


Fig. E9.3

(b) Determine and plot the differential error as a function of V_o. Estimate the peak magnitude of output signal for which the total harmonic distortion is 10%.

9.4 Compare Eqs. 9.17 with the approximate Eqs. 9.23 for the particular case of the ideal transistor amplifier in Section 9.4.1 for which Eq. 9.20 applies and

$$G_T = \frac{qI_C}{kT}.$$

(a) Calculate $\gamma(I_{CQ} + \frac{1}{2}I_{pp})$ and $\gamma(I_{CQ} - \frac{1}{2}I_{pp})$, and hence find the approximate values of D_2 and D_3 in terms of I_{pp}/I_{CQ}.

(b) Calculate the peak amplitude V_i of the sine wave input required to produce a peak-to-peak output I_{pp}:

$$V_i = \frac{\frac{1}{2}I_{pp}}{G_{TQ}}.$$

(c) Hence express the approximate values of D_2 and D_3 in terms of V_i, and compare with Eqs. 9.22.

Repeat the comparison for a practical transistor for which

$$G_T = \frac{\alpha_N}{r_B/\beta_N + r_E}.$$

Use Fig. 9.5 for the accurate predictions, and use the same values of $(r_B/\beta_N)/r_{EQ}$ as given in this graph. What can you conclude about the validity of Eqs. 9.23 for the useful range of signal amplitudes?

9.5 Calculate *exactly* the harmonic distortion in a f.e.t. amplifier as a function of input signal amplitude, assuming the ideal transfer characteristic (Eq. 4.141) applies. Use the results of Section 9.2.1 rather than Section 9.4.

9.6 Suggest a circuit, with two field-effect transistors as its essential part, whereby a sinusoid in the audio-frequency range can be doubled in frequency. Design such a circuit, using type 2N2497 *p*-channel junction f.e.t.s, for which the nominal parameters are:

$$V_P = 5\text{ V}, \qquad I_{DO} = 2\text{ mA}.$$

Specify the quiescent voltage V_{GQ}, and the range of signal amplitudes for which the circuit should be satisfactory.

(a) Include adjustments for trimming out unbalance between the f.e.t.s so that the input frequency is suppressed at all signal amplitudes.

(b) In view of the change in average current expected under signal conditions, what sort of biasing circuit is the most desirable?

CHAPTER 10

10.1 Readers familiar with flow-graph techniques can gain insight into the superposition process by drawing a flow-graph representation for the following set of equations:

$$v_1 = \frac{\Delta'_{11}}{\Delta'} i_1 + \frac{\Delta'_{k1}}{\Delta'} i_k,$$

$$v_2 = \frac{\Delta'_{12}}{\Delta'} i_1 + \frac{\Delta'_{k2}}{\Delta'} i_k,$$

$$v_j = \frac{\Delta'_{1j}}{\Delta'} i_1 + \frac{\Delta'_{kj}}{\Delta'} i_k,$$

$$i_k = g_m v_j.$$

Then use the rules of flow-graph reduction to check the following expressions for a_{12}, a_{1j}, z_{k2}, z_{kj}:

$$a_{12} = \frac{\Delta'_{12}}{\Delta'_{11}},$$

$$a_{1j} = \frac{\Delta'_{1j}}{\Delta'_{11}},$$

$$z_{k2} = \frac{\Delta'_{k2}}{\Delta'} - \frac{\Delta'_{12}\Delta'_{k1}}{\Delta'\Delta'_{11}},$$

$$z_{kj} = \frac{\Delta'_{kj}}{\Delta'} - \frac{\Delta'_{1j}\Delta'_{k1}}{\Delta'\Delta'_{11}}.$$

10.2 Figure E10.2 shows the circuit of a feedback pair which uses *p-n-p* and *n-p-n* transistors in combination. Use the emitter resistor R_{F1} as the reference variable to evaluate the mid-band closed-loop voltage gain. Hence determine the demanded gain and A_l. The results may be compared with those given in Section 14.4.1.1.

10.3 It is shown in Section 10.6.1 that it may be difficult to calculate the forward transfer function μ directly, because μ cannot be simply identified with the transfer function with the feedback network removed. However, any given β circuit defines a path through which the feedback signal flows; that is, β defines a section of the complete feedback loop of gain

$$A_l = \mu\beta.$$

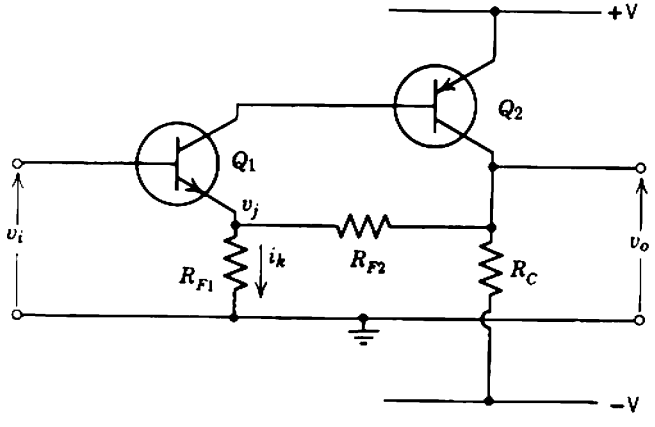


Fig. E10.2

Following Section 10.3.2, A_l may be determined directly by breaking the loop at any point where the signal is confined to a single path. The break can be in either the μ or β circuits. It is convenient, but not necessary, to break the loop at a point which has a large discontinuity in impedance on either side of the break, so that circuit conditions are not changed appreciably by making the break. For simple stages, useful points are the grid of a vacuum tube, the gate of an f.e.t., and the output electrode of a pentode, transistor, or f.e.t. When no such simplifications are apparent, the interaction of impedance on either side of the break must be taken into account as illustrated in the example of Fig. E10.3:

$$A_l = \frac{i_F}{i_S}.$$

Determine the mid-band loop gain for the feedback pair of Fig. E10.2 by breaking the loop between the collector of Q_1 and the base of Q_2.

10.4 (a) Which are the most appropriate points for breaking the loop to determine the mid-band loop gain of the voltage amplifier of Fig. E10.4. The transistors have nominal values of β_N of 200, r_B is negligibly small, and r_C can be assumed infinite. Q_1 operates at 93 μA, and both Q_2 and Q_3 at 1 mA. Calculate the nominal mid-band loop gain. What is the dominant source of variation in A_{lm}?

(b) What is the demanded closed-loop voltage gain? What is the actual closed-loop voltage gain?

(c) Determine the output resistance:

(i) by calculating the magnitude of additional external load resistance which reduces the actual closed-loop voltage gain to exactly one half of that in (b);

(ii) by means of Eq. 10.72.

(d) If a capacitive load exists across the output, investigate the character of the closed-loop dynamic response.

10.5 The White cathode follower of Fig. E10.5 has a voltage gain close to unity and a very low output resistance. Use superposition to obtain accurate algebraic expressions for A_V and R_o. Give reasons for the choice of reference variables.

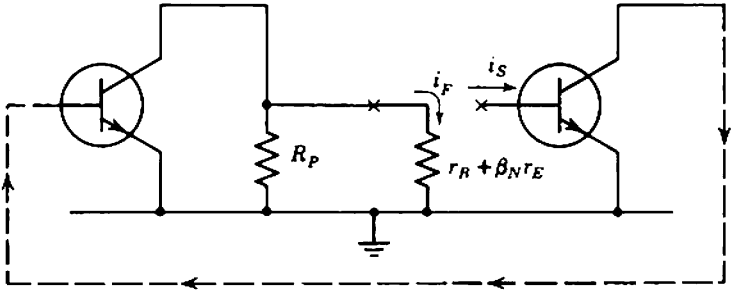


Fig. E10.3

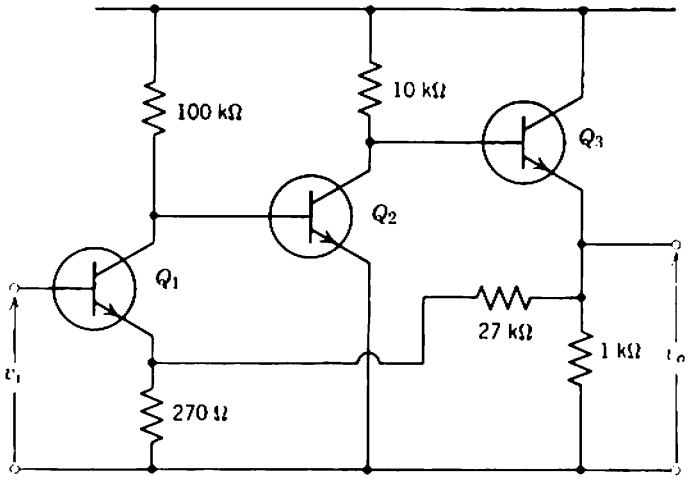


Fig. E10.4

10.6 The standard method of measuring loop gain is to break the feedback loop at a point of appropriate impedance discontinuity, feed in a sinusoidal voltage or current signal, and measure the magnitude and phase of the signal returned to the other side of the break. A typical arrangement for a voltage point is shown in Fig. E10.6*a*.

This method of measurement is often quite difficult to apply because a suitable point for breaking the feedback loop may not exist, and an alternative

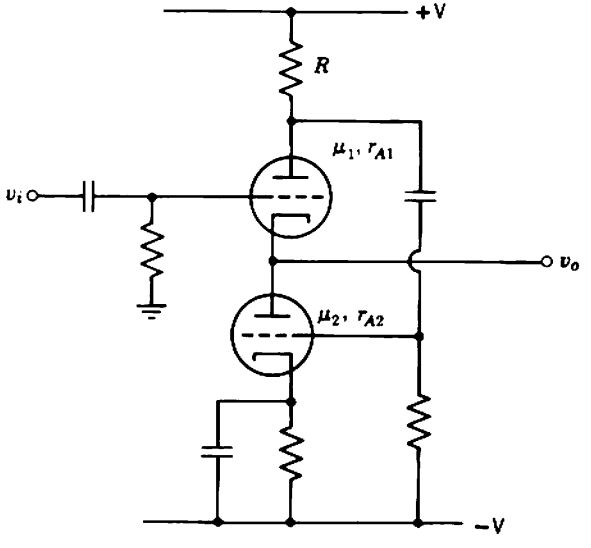


Fig. E10.5

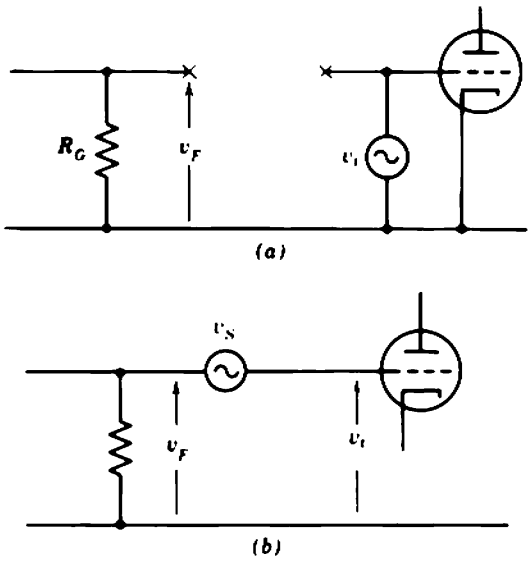


Fig. E10.6

approach which makes use of the properties of clip-on current probes has been suggested.* These probes are essentially small transformers which use a conducting wire in the circuit under test as the low-impedance winding; the series impedance induced into the wire is the shunt combination of about 10 mΩ and 1 μH. A clip-on probe is normally used to detect a current flowing in the low-impedance winding but it can also be operated in reverse to induce a voltage in a conductor.

(a) Verify that the magnitude and phase of loop gain can be determined by using a current probe for injecting a known test voltage v_S into the circuit, without breaking the feedback loop, and measuring the magnitudes only of v_i and v_F in Fig. E10.6*b*.

(b) Show how the loop gain can be determined by injecting a known signal current into a conductor from a high-impedance source and using a current probe to measure the magnitude of signal currents on either side of the injection point.

10.7 The circuit shown in Fig. E10.7 is a wide-band transistor voltage amplifier. Rise and fall times are approximately 2 nsec, corresponding to a bandwidth of approximately 200 MHz. Nominal parameters for the transistors used are

2N709	2N976
$\beta_N = 50$,	$\beta_N = 150$,
$r_{bb'} = 50\ \Omega$,	$r_{bb'} = 35\ \Omega$,
$f_T = 800$ MHz,	$f_T = 900$ MHz,
$I_{CO} = 5$ nA.	$I_{CO} = 3\ \mu$A.

* P. SPOHN, "A quick convenient method for measuring loop gain," *Hewlett-Packard Journal*, **14**, 5, January, 1963.

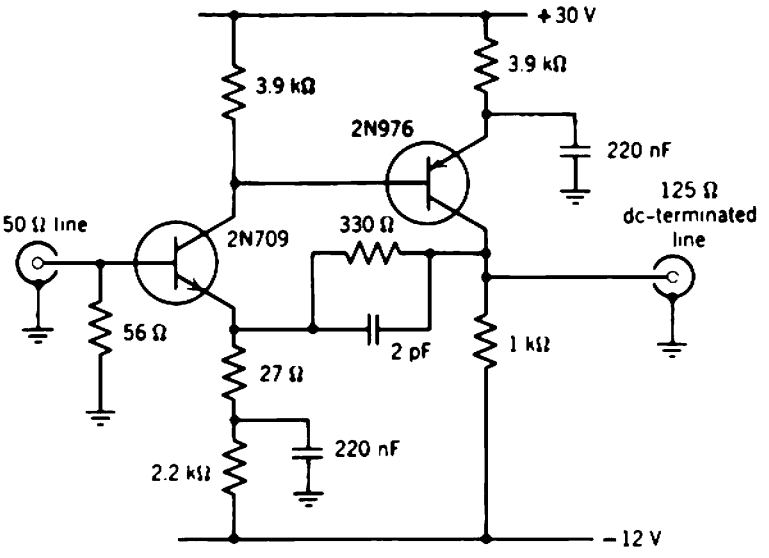


Fig. E10.7

(a) Calculate the spot noise current and voltage generators at the input of the amplifier.

(b) Assuming that the input is terminated by a 50-Ω line, calculate the equivalent noise referred to the input, assuming the amplifier has a "brick-wall" response with 200 MHz bandwidth. How accurate is this assumption?

(c) What is the minimum input signal for 20 dB signal-to-noise ratio at the output?

10.8 This problem illustrates an application of positive feedback.

Figure E10.8 shows a method for achieving zero output impedance. Feedback network β_1 derives normal negative voltage feedback from the output, whereas network β_2 derives positive current feedback.

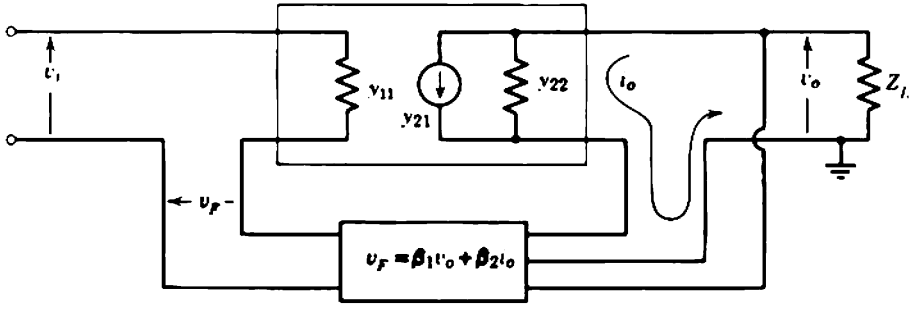


Fig. E10.8

(a) Show that the output resistance of the amplifier with combined positive and negative feedback is zero if

$$\beta_2 = \frac{1}{y_{21}}.$$

That is, the output impedance is zero if the β_2 loop gain into a short-circuit load is $+1$.

(b) Show that the total loop gain is

$$A_l = -\frac{y_{21}Z_L}{1 + y_{22}Z_L}\left(\beta_1 - \frac{\beta_2}{Z_L}\right).$$

Hence show that addition of a unity-gain positive current feedback loop to an existing negative feedback amplifier is unlikely to cause regeneration.

(c) Draw block diagrams illustrating the use of combined positive and negative feedback to obtain:

infinite input impedance,
zero input impedance,
infinite output impedance.

Suggest applications.

10.9 This problem illustrates another application of positive feedback.

Figure E10.9 shows a method for achieving very good linearity. Local positive feedback is applied to the first stage of an amplifier with over-all negative feedback. This raises the stage gain, and hence the gain around the over-all feedback loop. In the limit, the over-all loop gain becomes infinite when the first stage is on the point of oscillating.

(a) Show that increasing the positive feedback beyond the point of unity loop gain does not cause oscillation, or even change the system response appreciably as long as the first stage (without the positive feedback) contributes a dominant pole to the over-all feedback loop. (*Hint.* It is assumed in Sections 10.5.2 and 10.5.3.1 that $\mu(s)$ has no poles in the right half of the s plane.)

(b) Show that, as the positive loop gain approaches $+1$, the system non-linearity approaches that of the first stage alone with only the negative feedback present. Comment on the merit of applying positive feedback to the first stage, rather than the last stage of an amplifier.

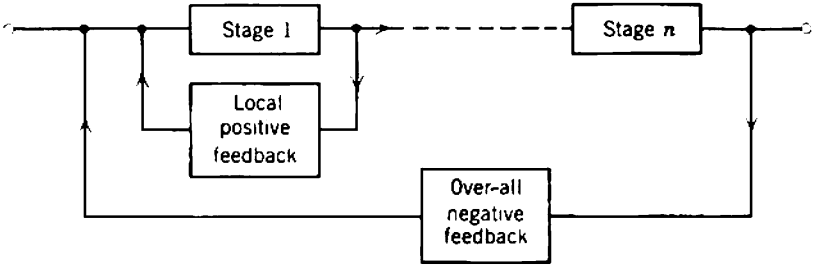


Fig. E10.9

(c) Comment on the proposal to use this system in order to achieve extremely high gain stability. At unity positive feedback, the over-all negative feedback becomes infinite and the stabiiity factor S_μ becomes zero.

10.10 The root-locus technique is not restricted to the investigation of feedback problems. An unusual application of root-locus ideas is illustrated by the double-tuned circuit of Fig. E10.10.

(a) Show that the transfer impedance of this passive circuit is given by

$$\frac{v_o(s)}{i_i(s)} = \frac{\pm k\omega_{01}\omega_{02}}{(1-k^2)C_1C_2} \times \frac{s}{\left(s^2 + \frac{\omega_{01}}{Q_1}s + \omega_{01}^2\right)\left(s^2 + \frac{\omega_{02}}{Q_2}s + \omega_{02}^2\right) - k^2\omega_{01}^2\omega_{02}^2}$$

$$= \frac{\pm k\omega_{01}\omega_{02}}{(1-k^2)C_1C_2} \frac{s}{(s-p_{1O})(s-p_{1O}^*)(s-p_{2O})(s-p_{2O}^*) - k^2\omega_{01}^2\omega_{02}^2}$$

$$= \frac{\pm k\omega_{01}\omega_{02}}{(1-k^2)C_1C_2} \frac{s}{(s-p_{1C})(s-p_{1C}^*)(s-p_{2C})(s-p_{2C}^*)}.$$

where ω_{01}, ω_{02} are the center frequencies of each tuned circuit with the other tuned circuit removed and Q_{01}, Q_{02} are the corresponding quality factors. The factor k is the coefficient of coupling.

(b) If the individual tuned-circuit poles p_{1O} and p_{2O} are given to be coincident at $-0.1\omega_0 + j\omega_0$, determine the loci of p_{1C} and p_{2C} over the range of realizable k.

(c) If it is desired that p_{1C} and p_{2C} have the same real part, say,

$$p_{1C} = -0.1\omega_0 + j1.1\omega_0,$$
$$p_{2C} = -0.1\omega_0 + j0.9\omega_0,$$

at a particular value of k, determine the loci of p_{1O} and p_{2O} for realizable tuned circuits.

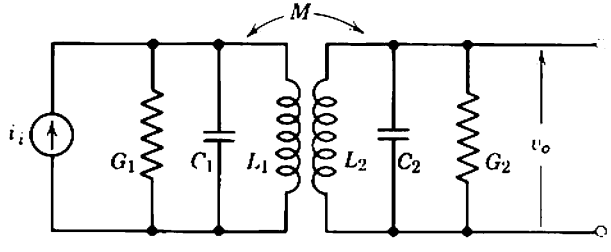


Fig. E10.10

CHAPTER 11

11.1 Sketch elemental circuit diagrams for single- or two-stage amplifiers to satisfy the following conditions. For each design, specify the following quantities if they are significant:

(i) emitter current,
(ii) average β_N,
(iii) minimum β_N.

(a) Transfer conductance = 25 mA/V;
Input resistance $\geqslant$ 1.0 kΩ.
(b) Transfer resistance = 50 V/mA for a load resistance = 5.0 kΩ;
Input resistance $\leqslant$ 250 Ω.
(c) Voltage gain = 1000 for a load resistance = 5.0 kΩ;
Average input resistance = 5.0 kΩ.

11.2 Refer to Problem 5.3 and redesign the two-stage amplifier to consist of a series stage feeding a shunt stage. Comment on the performance achieved in comparison with the other two realizations.

11.3 Design an audio amplifier for use with the photoelectric cell in a film projector. Use type OC44 transistors (data in Problem 4.6).

CELL CHARACTERISTICS.

(i) Maximum signal output current into 1-MΩ load: 1 μA rms.
(ii) Cell output resistance: 10 MΩ.

AMPLIFIER GAIN.

Require 1 V rms output into a 600-Ω load on signal peaks.

(a) Which type of transfer function should be used?
(b) What is its magnitude?
(c) Sketch an elemental circuit diagram showing the feedback resistors and emitter currents.
(d) Complete the mid-band design by finalizing bias circuit details and resistor values. Capacitor values need not be specified. (See also Problem 12.3.)

11.4 Design a 20-dB amplifier for use with 75-Ω coaxial cables. The output voltage is to be at least 1 V peak-to-peak. Use silicon *n-p-n* transistors which have the following average characteristics:

$$\beta_N = 75,$$
$$r_B = 100\ \Omega.$$

(a) Decide on a configuration of stages. (Look at Section 13.7.)
(b) Sketch an elemental circuit diagram, showing feedback resistors and emitter currents.
(c) Complete the mid-band design by finalizing biasing circuit details and resistor values. Capacitor values need not be specified.

11.5 Figure E11.5 shows a circuit having frequency-dependent feedback:
(a) What is the minimum loop gain?
(b) Determine the frequency response of the transfer impedance.

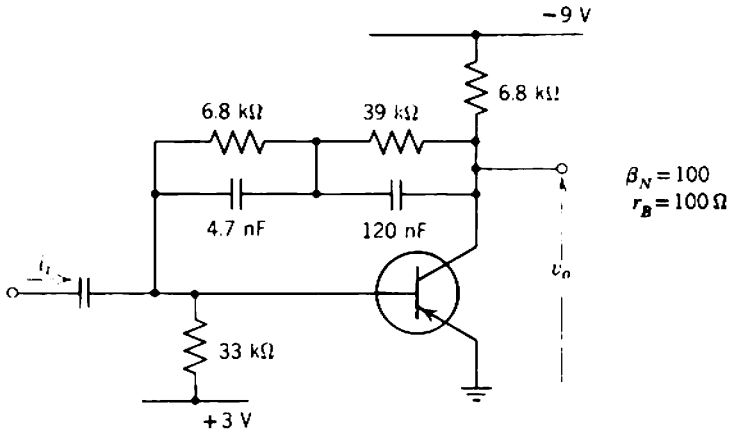


Fig. E11.5

CHAPTER 12

12.1 Investigate the decoupling required for an amplifier consisting of alternate series- and shunt-feedback stages.

(a) Find the equivalent voltage or current referred to the input of a series or shunt stage, due to a signal voltage on the dc supply rails.
(b) For the transistor amplifier shown in Fig. 12.18*b*, calculate:

(i) The 120 Hz hum voltage which must be present on either supply rail to give 10 μV equivalent hum referred to the input.
(ii) The internal resistance of the power supply necessary to give a loop gain of unity, for both mid-band and the worst possible frequencies. Is the mid-band phase relation such that the feedback is positive, that is, will the amplifier oscillate?

12.2 Figure E12.2 shows a proposed scheme for adjusting the low-frequency compensation of a transistor feedback amplifier. Variation of R changes the feedback blocking capacitor over the range C_F to 1.5 C_F, and moves the zero of the stage about half an octave.

(a) Find $\psi(s)$ of the stage, as a function of R.
(b) If the response of the rest of the amplifier can be represented by

$$\psi(s) = \frac{s}{s + 1/\tau},$$

where

$$\tau = 1.2\, R_F C_F,$$

plot the square wave response for various values of R. Assume the repetition frequency f_0 is $2.5/\tau$.

12.3 Complete the design of the audio amplifier of Problem 11.3 to provide a low-frequency response which is 3 dB down at 30 Hz. The solution should show:

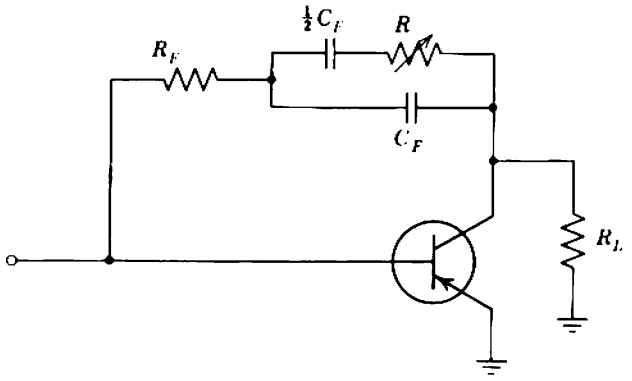


Fig. E12.2

(i) a complete circuit diagram,
(ii) the low-frequency singularity pattern for each capacitor,
(iii) the over-all low-frequency singularity pattern.

12.4 Figure E12.4 shows a circuit designed to provide a narrow band-pass transfer function centered at 10 Hz. Find the low-frequency singularities and hence determine the bandwidth of the transfer function $v_o(j\omega)/i_i(j\omega)$. Show how the circuit diagram can be simplified by representing the amplifier as a functional block.

12.5 This problem is the derivation of four important practical results for cross-coupling between the poles in a transistor amplifier—parts (e) through (h).

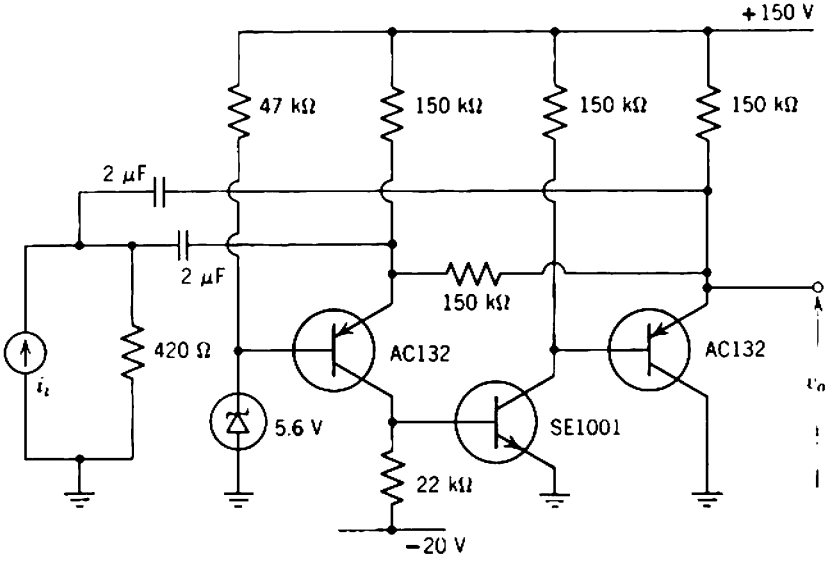


Fig. E12.4

(a) Show that the transfer admittance of Q_2 in Fig. E12.5 is

$$Y_T(s) = G_T\left(\frac{s - z}{s - p}\right),$$

where

$$p = -\frac{1 + G_T R_E}{R_E C_E}$$

$$z = -\frac{1}{R_E C_E}.$$

(b) Show that the input impedance of Q_2 is

$$Z_i(s) = R_i\left(\frac{s - p}{s - z}\right),$$

where

$$R_i = \frac{\beta_N}{G_T}.$$

(c) Show that the input impedance of the transistor in combination with R_B is

$$Z_i^*(s) = R_i^*\left(\frac{s - p}{s - z^*}\right),$$

where

$$R_i^* = R_i || R_B$$

$$|z^*| > |z|.$$

(d) Show that

$$\frac{i_{C2}(s)}{i_{C1}(s)} = -\frac{sR_i^* C_C(s - z)G_T R_C}{(1 + sR_C C_C)(s - z^*) + sR_i^* C_C(s - p)},$$

so that the total singularity pattern consists of

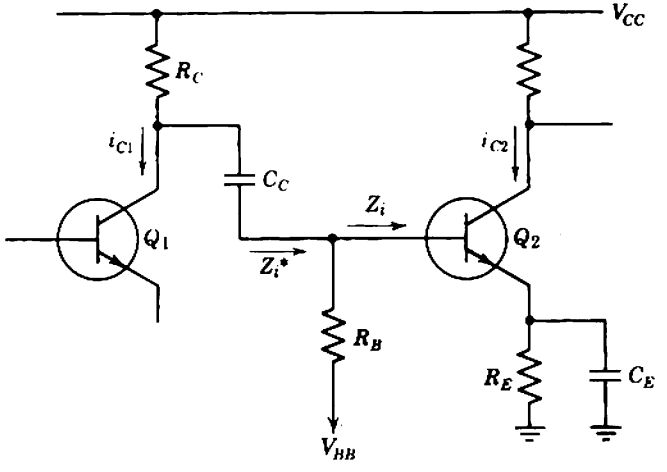


Fig. E12.5

(i) one zero at the origin (i.e., the zero due to C_C alone),
(ii) one zero at z (i.e., the zero due to C_E alone),
(iii) two poles, at the solutions of

$$\frac{s(s-p)}{(s+1/R_CC_C)(s-z^*)} = -\frac{R_C}{R_i^*}.$$

(e) By means of a technique similar to root locus plotting show that the dominant pole lies on the negative real axis between the limits p and $-1/R_CC_C$. Show that the pole approaches p as R_C/R_i^* approaches zero.
(f) Show that the nondominant pole lies on the negative real axis between z^* and the origin.
(g) Show that, in general, the dominant pole is at

$$s \approx -\frac{1/C_C - pR_i^*}{R_C + R_i^*}.$$

If $R_B \gg R_i$ and $G_TR_E \gg 1$, show that

$$s \approx -\left(\frac{C_E + \beta_N C_C}{C_E}\right)\left(\frac{1}{R_C + R_iC_C}\right).$$

(h) Comment on any practical advantage in designing a circuit with

$$p = -\frac{1 + G_TR_E}{R_iC_E} = -\frac{1}{R_CC_C}.$$

(i) Would application of series feedback to Q_2 affect any of these results?

CHAPTER 13

13.1 A pentode has equal input and output capacitances. If the total capacitance, is C, determine and compare the bandwidth reduction factors of the following coupling networks:

(i) simple resistance-capacitance,
(ii) shunt peaking (maximally flat),
(iii) series peaking adjusted to make as many derivatives as possible of $|\psi(j\omega)|^2$ zero at $\omega = 0$,
(iv) series peaking with additional "padding" capacitance added so as to give a true maximally-flat response.

For the same total capacitance C, what would be the effect of changing the ratio of input to output capacitance to values of 2, 3, and 4?

13.2 (a) A peaked shunt-feedback stage can be represented in the approximate manner shown in Fig. E13.2*a*. Determine the singularities of the loop gain with respect to the open-loop transfer function by breaking the feedback loop at the point shown. Hence, use root-locus plots to investigate the effect of various relative magnitudes of τ_β, τ_L, and τ_F on the closed-loop performance. Under what circumstances can a pair of complex conjugate poles be obtained in the closed-loop transfer impedance?
(b) A similar argument can be developed for the series stage by representing it

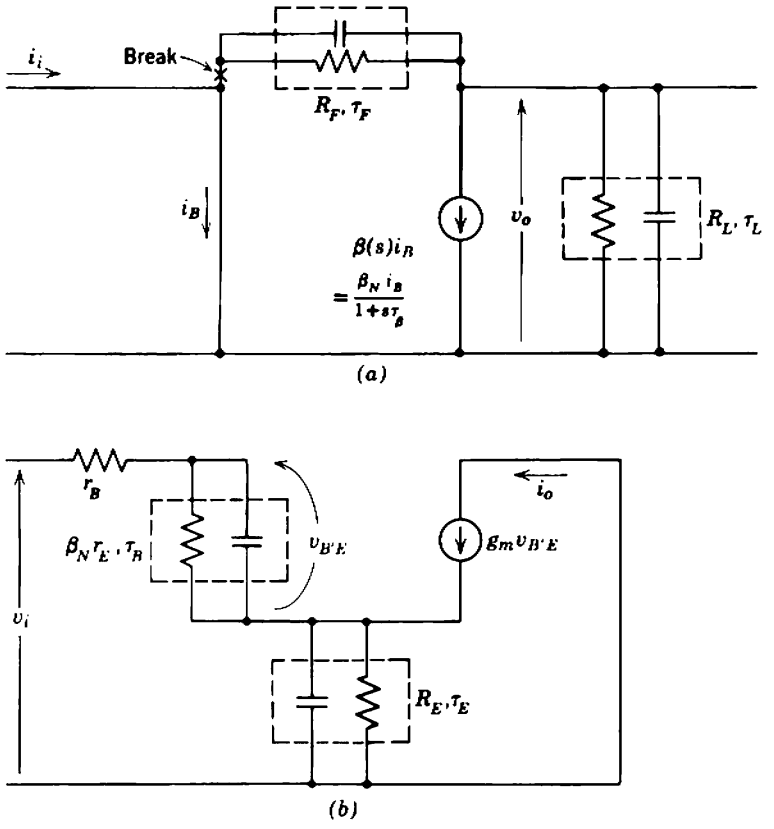


Fig. E13.2

in the approximate manner shown in Fig. E13.2*b*. Determine the loop-gain singularities directly and use root-locus reasoning to establish the criterion for a single-poled closed-loop transfer function.

Note. In all the following numerical problems, use type OC170/2N1516 transistors (*p-n-p* germanium) which have the following average parameters at 5 mA, 6 V, 16°C:

$$\beta_N = 150,$$
$$r_B = 50\ \Omega,$$
$$\tau_1 = 1.0 \text{ nsec},$$
$$c_{tE} = 50 \text{ pF},$$
$$c_{tC} = 2 \text{ pF}.$$

As a result of high-level injection, τ_1 rises to 1.4 nsec when I_E is increased to 20 mA.

13.3 Design a series-feedback stage whose transfer admittance into a small load to have a cutoff frequency of 100 MHz. Calculate its input capacitance and average input resistance.

13.4 Design a series-feedback stage to work from a 75-Ω source to a small load. The transfer function

$$\frac{\text{output current}}{\text{source voltage}}$$

is to have a cutoff at 100MHz.

What is the difference between Problems 13.3 and 13.4?

13.5 Design a shunt-feedback stage to give a 10 MHz, two-pole maximally-flat transfer impedance into 50 pF load capacitance.

13.6 Design a two-stage current amplifier to give a gain of 40 into a small load resistance. The bandwidth is to be at least 10 MHz, and the overshoot is to be less than 1%. Investigate the effect of feedback resistor magnitudes on bandwidth. What is the maximum bandwidth possible?

13.7 Design a two-stage amplifier to give 20 dB gain between 75-Ω terminations. The bandwidth is to be as large as possible consistent with 1% overshoot, and the tilt on a 50 Hz square wave is to be less than 5%. Solutions should show the following:

(i) the elemental circuit diagram, with emitter currents marked,
(ii) the complete circuit diagram,
(iii) the mid-band transfer function of each stage, and the interstage coupling efficiencies,
(iv) the high-frequency and low-frequency singularity patterns for each stage.

CHAPTER 14

14.1 In the discussion of maximum feedback amplifiers in Section 14.2, an optimum number of stages n_{opt} is derived for which $|A_{lm}|$ is maximum. Comment on the statement that a more realistic optimization should maximize $|A_{lm}|/n$. Derive this optimum number of stages for an amplifier without high-frequency peaking, and make a numerical comparison with Eq. 14.13.

14.2 Figure E14.2*a* is the elemental mid-band circuit diagram for a low-noise transfer-resistance amplifier designed by *X*. The forward path (enclosed in the broken rectangle is synthesized from series- and shunt-feedback stages. In describing this amplifier at a technical meeting *X* claimed the following characteristics:

	Calculated	Experimental
Transfer resistance	594 kΩ	585 kΩ
Input resistance	210 Ω	240 Ω
Equivalent noise current referred to the input	0.41 pA/$\sqrt{\text{Hz}}$	0.40 pA/$\sqrt{\text{Hz}}$

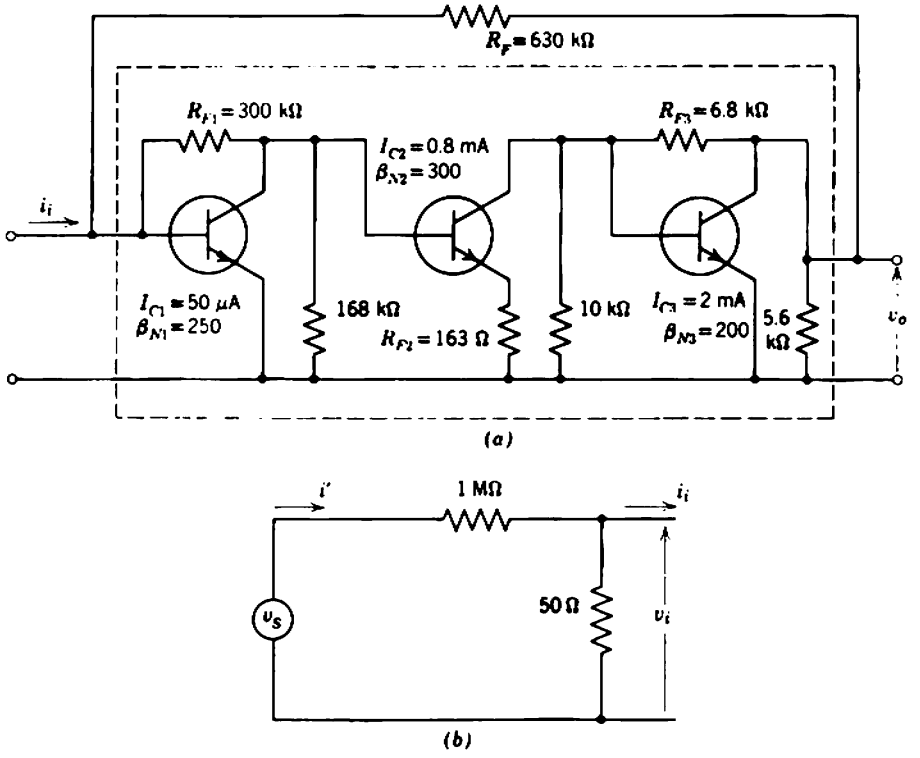


Fig. E14.2

A member of the audience, *Y*, queried these results and claimed that the noise current generator bridging any two nodes in a linear circuit can be identified as the thermal noise of the net resistance between nodes. *Y* then queried the measured value of input resistance and asked how it was obtained.

X explained that the measurement had been performed as shown in Fig. E14.2*b*. Since v_s and v_o are relatively large,

$$v_i = v_s \frac{50\ \Omega}{1\ M\Omega},$$

$$i_i = \frac{v_o}{R_T}.$$

Therefore

$$R_i = \frac{v_i}{i_i} = \frac{v_s}{v_o} \frac{50}{10^6} R_T.$$

This explanation caused *Y* to claim that the measurement was not valid, because the 50-Ω resistor across the input substantially short-circuits the feedback and the input resistance rises. *X* replied that the input impedance of any

circuit (even one including feedback) cannot be influenced by the method of measurement and that *Y*'s claims were groundless.

Investigate this circuit and comment carefully on the validity of the statements of *X* and *Y*.

14.3 Design an amplifier to meet the following specifications at mid-band frequencies:

$$A_V = 10^4, \pm 1\%,$$
$$R_L \text{ (external)} = 10\ \text{k}\Omega,$$
$$R_i \text{ (average)} > 100\ \text{k}\Omega,$$
$$v_o = 2\ \text{V rms}.$$

The design can be performed in the following steps:

(a) Design the mid-band circuit.
(b) Design the biasing system.
(c) Calculate the capacitor values for any frequency response that is thought reasonable.

> *Note.* In the following problems use transistors having the following parameters:
>
> $$\beta_N = 70,$$
> $$r_B = 100\ \Omega.$$

14.4 Design a current-feedback pair to give a mid-band current gain of 300. Investigate the effect of changing β_N down to 50 and up to 200.

14.5 Design a voltage-feedback pair to give a mid-band voltage gain of 200 into a 3-kΩ external load resistance. The input resistance (including biasing resistors) is to be 3 kΩ with average transistors.

CHAPTER 15

15.1 Figure E15.1 illustrates the use of a differential voltage amplifier as a detector

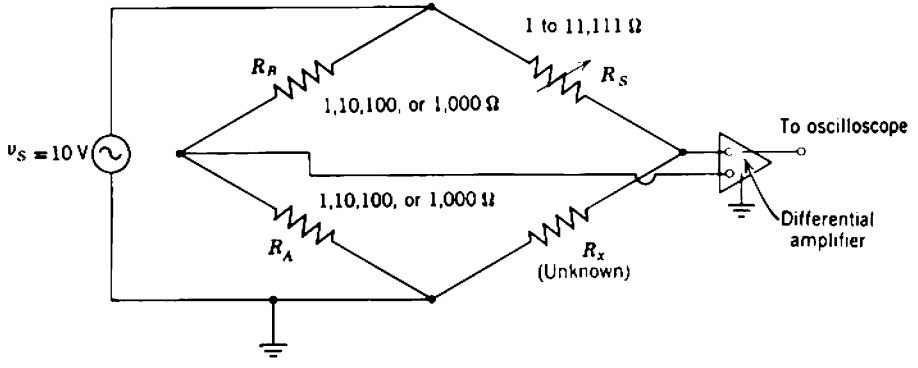


Fig. E15.1

for an ac bridge. The differential-mode voltage gain must be adequate to detect a 1-Ω change in the standard resistor R_S. Investigate the common-mode rejection required in the amplifier.

15.2 Design a wide-band differential amplifier using OC170 transistors (data in Problem 13.3) to satisfy the following requirements:

mid-band common-mode rejection $\approx 10^5$,
common-mode rejection at 100 kHz $\geqslant 10^4$.

15.3 Estimate the voltage and current drift generators per degree centigrade for transistor type OC44 (data in Problem 4.6), operating at 0.5 mA emitter current. Compare these with the values for transistor type 2N2484 (data in Fig. 15.20).

15.4 Design a dc amplifier to satisfy the following conditions, using type OC44 in the long-tailed pair input stage.

gain = 100
$\pm 5\%$ absolute value
$\pm 1\%$ short-term stability,
input resistance = 600 Ω $\pm$ 5%,
source resistance = 600 Ω,
external load resistance = 10 kΩ.

Estimate the separate drift contributions of the voltage and current generators and the total drift. With one additional potentiometer, devise a means for balancing the current drift. Would the availability of transistors of type 2N2484 ease the design problem?

15.5 Design an operational amplifier for use in the block of Fig. E15.5:

loop gain (dc to 100 Hz) $\geqslant 10^4$,
drift (referred to system input) $\leqslant$ 1 mV/°C.

The following design steps are suggested:

(a) What transfer function type is required for A, and what is its magnitude?
(b) Design a skeleton amplifier which will ensure high-frequency stability.
(c) Estimate the drift generators of the input stage.
(d) What drift generators are tolerable? Therefore, is a parallel-path amplifier necessary?
(e) Design the amplifier in detail.

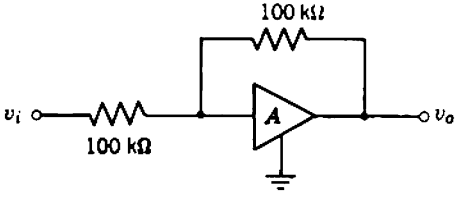


Fig. E15.5

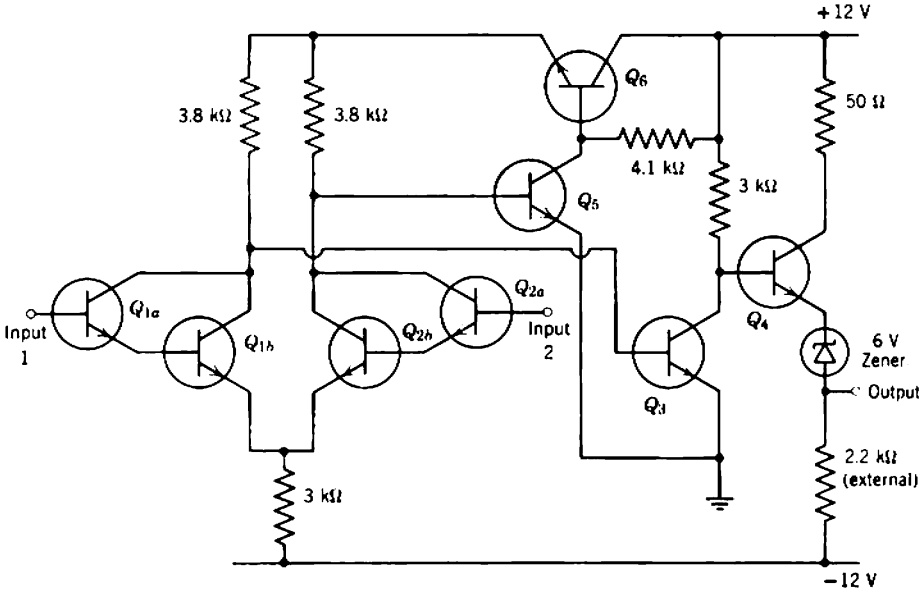


Fig. E15.6 (Courtesy Plessey Company, Ltd.)

15.6 Figure E15.6 shows the circuit of a commercially available integrated circuit operational amplifier (the SL701). The design makes use of the relatively narrow tolerances achieved in β_N and V_{BE} when all devices are manufactured in a single chip of silicon (Section 17.3). All transistors have the following parameters:

At 25°C, 5 mA, 5 V: $\beta_N = 60 \pm 5\%$, $V_{BE} = 0.78\text{ V} \pm 2\text{ mV}$.
At 120°C, 5 mA, 5 V: $\beta_N = 90$, $V_{BE} = 0.63\text{ V}$.

(a) Evaluate the quiescent voltages and currents in the circuit and then determine the voltage gain.
(b) Explain the action of the auxiliary circuit which controls the positive supply voltage to the long-tailed pair, and comment on the claimed common-mode rejection of 70 dB.

15.7 Suggest a method (in outline and detail) for measuring the common-mode rejection factor for the amplifier of Problem 15.2.

CHAPTER 16

16.1 The thermal resistance of transistors is measured by dissipating a known power until thermal equilibrium is established and then measuring the junction temperature. The "thermometer" used for measuring junction temperature is one of the temperature-dependent parameters of the transistor; examples are the leakage current of a reverse-biased junction and the voltage drop of a forward-biased junction. Once the heating power is disconnected from the

transistor under test, the junction temperature falls. Consequently, it is necessary to monitor the temperature-dependent parameters immediately after removal of the power. Figure E16.1*a* shows a typical test set-up which uses the emitter voltage V_E as the temperature-dependent parameter.

With the switch in position 1, a relatively large power

$$P_{C1} = V_{CC1}I_E$$

is dissipated in the device for a time sufficient to produce thermal equilibrium at which V_E becomes constant. When this occurs, the switch is moved to position 2 and a smaller power

$$P_{C2} = V_{CC2}I_E$$

is dissipated until thermal equilibrium again results. Figure E16.1*b* shows the power and emitter voltage waveforms; note the sudden discontinuity at the beginning of the cooling curve which results from the decrease in collector depletion layer width with collector voltage. The slow change ΔV_E is proportional to the fall in junction temperature due to the change in dissipated power:

$$\Delta P_C = (V_{CC1} - V_{CC2})I_E = (\Delta V_C)I_E.$$

(a) In a test on a power transistor in a heat sink,

$$\Delta P_C = 36 \text{ W}$$

and

$$\Delta V_E = 70 \text{ mV}.$$

If the temperature variation of V_E is such that

$$\frac{\partial V_E}{\partial T} = -2.0 \text{ mV/°C},$$

determine the thermal resistance between the junction temperature and ambient.

(b) Use any suitable argument to show that the two 2-terminal networks of Fig. E16.1*c* are equivalent provided that the time constants τ_1, τ_2, τ_3 are such that

$$\tau_1 \ll \tau_2 \ll \tau_3.$$

Then obtain an analytical expression for the temperature variation in a transistor due to a step in power dissipation.

(c) Figure E16.1*d* shows the detail of some experimentally obtained cooling curves of a good and a rejected sample of a power transistor. Subsequent microscopic examination disclosed that the rejected device had an incomplete bond between the junction assembly and the case. Show that the experimental curves can be represented very accurately by a 3-section thermal equivalent circuit involving thermal resistances from semiconductor to bond, bond to case, and case to ambient, together with thermal capacitances of the semiconductor, the bond, and the combination of case and heat sink. Comment on the accuracy of approximating the cooling curve with a single exponential.

16.2 The acceptable transistor of the previous problem is used in a single-ended Class *A* amplifier. If the maximum allowed junction temperature is 150°C,

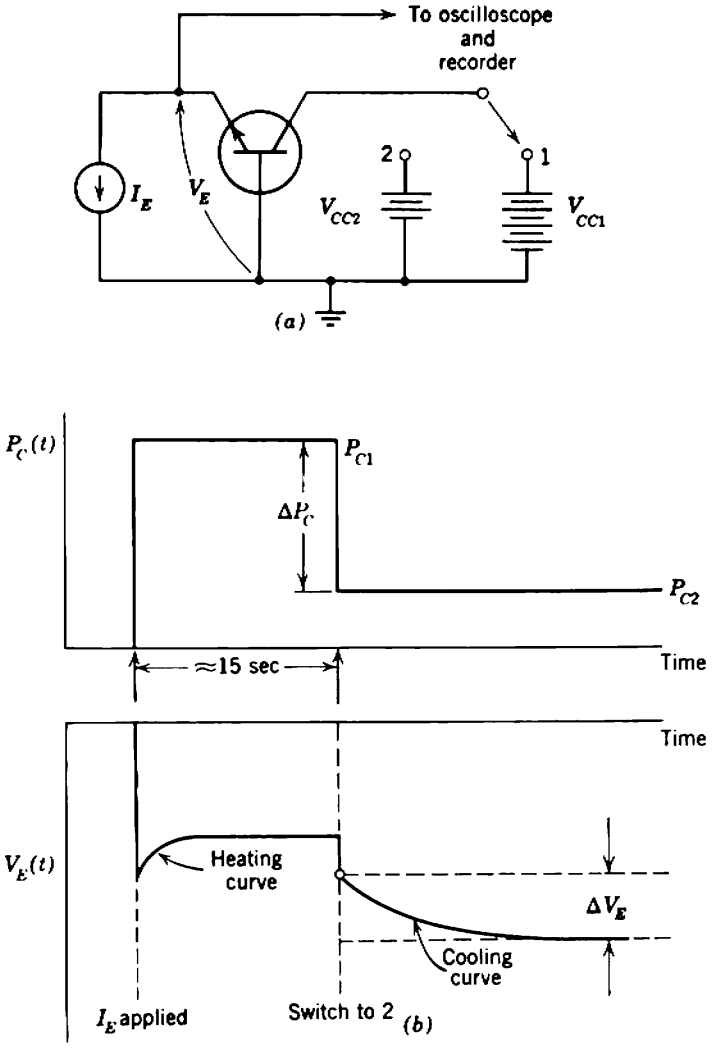
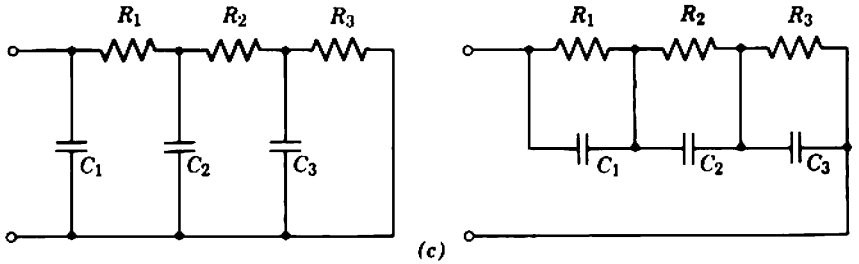
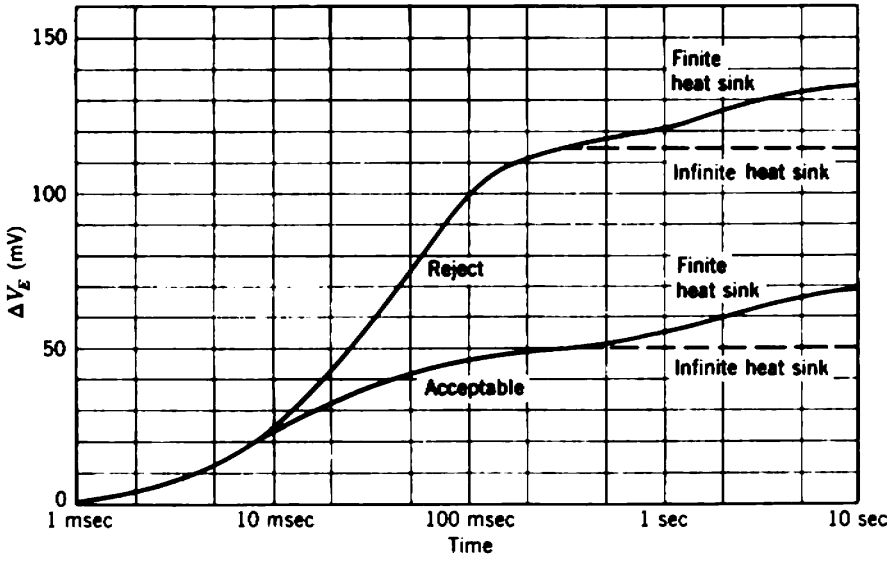

To oscilloscope
and
recorder
I_E
V_E
2
1
V_{CC2}
V_{CC1}
(a)
$P_C(t)$
P_{C1}
ΔP_C
P_{C2}
≈15 sec
Time
Time
$V_E(t)$
Heating
curve
ΔV_E
Cooling
curve
I_E applied
Switch to 2
(b)

Fig. E16.1



(c)



(d)

Fig. E16.1 (*continued*)

determine the peak power which can be dissipated when the time variation of power takes the following forms:

(a) Constant, independent of time.
(b) Sinusoid of frequency 100, 10, 1, 0.1, and 0.01 Hz.
(c) Square wave of frequency 100, 10, 1, 0.1, and 0.01 Hz.
(d) Rectangular pulse of width 1 msec at repetition rates of 100, 10, and 1 Hz.

16.3 The discussion of power relations in Class *B* amplifiers given in Section 16.2.3 assumes the presence of sinusoidal signals. The power rating of amplifiers usually is expressed as the maximum undistorted sine wave power $P_{LS(max)}$ that can be delivered to the load at mid-band frequencies. However, device dissipation ratings which are based on sinusoidal signal waveforms can be either optimistic or pessimistic. The following exercises throw further light on the topic. In each case assume that clipping points correspond to zero voltage or zero current.

(a) Determine and plot the average dissipation per device (as a percentage of $P_{LS(max)}$) in a resistance-loaded Class *B* amplifier against the percentage drive for sinusoidal, square-wave, and triangular-wave signals.
(b) White noise represents a better approximation to speech or music signals and readers strong in statistics should repeat (a) above for a clipped noise signal. Assume that the probability density distribution is Gaussian and express the percentage drive as the ratio of rms value of the noise to the peak swing allowed between clipping points.
(c) Figure E16.3 shows a type of Class *B* output stage having reactive coupling to the load resistor. Indicate how the I_L versus V_C characteristic can be constructed from a knowledge of the collector volt-ampere characteristics of the individual transistors. Notice that all diagrams in the text neglect the dependence of collecting-electrode current on collecting-electrode voltage. Superimpose the load line appropriate for frequencies such that

$$R_L \gg \frac{1}{\omega C_C}$$

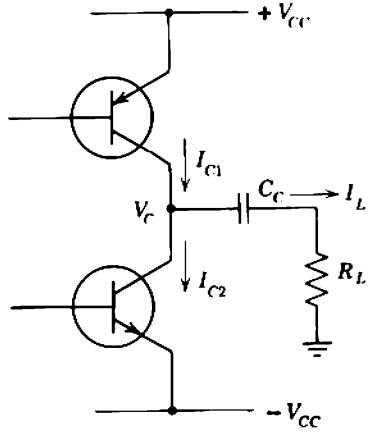


Fig. E16.3

and then determine the maximum peak power dissipation per device as a fraction of $P_{LS(max)}$.

(d) Show that the operating locus becomes an ellipse when a low-frequency sinusoid is applied to the amplifier and evaluate the peak power dissipation per device as a fraction of $P_{LS(max)}$ for a range of phase angles:

$$\phi = -\arctan \frac{1}{\omega C_C R_L}.$$

What is the worst case for this condition of drive?

(e) If any type of signal (including an overloading signal) is expected, what is the highest possible value of peak power dissipation per device?

(f) Comment on the criterion you would use for choosing transistors to deliver a specified load power $P_{LS(max)}$.

16.4 Figure E16.4*a* shows the elements of a biasing circuit arrangement suitable for each half of a Class *B* output stage.

When operated at 10 mA, the spread in diode voltage at 25°C junction temperature is given by

$$0.68 \text{ V} < V_D < 0.73 \text{ V}.$$

The spread in base-emitter characteristic of the low-power transistor at 25°C junction temperature is shown in Fig. E16.4*b*. It may be assumed that β_N exceeds 100 and that V_{BE} changes by -3 mV/°C.

(a) Determine the minimum value of R_E which ensures that the maximum emitter current (for junction temperature 25°C) does not exceed 5 mA. What is the corresponding minimum value of emitter current?

(b) Assume that the diode junction temperature remains at 25°C and that the transistor junction temperature is an independent variable that can alter over the range of 25°C to the maximum temperature of 125°C. Compute the maximum and minimum values of emitter current as a function of junction temperature. If V_{CC} is 20 V, plot power dissipation against junction temperature as the independent variable.

(c) Assume that power dissipation is the independent variable and plot the resulting junction temperature for an ambient temperature of 25°C and a thermal resistance from junction to ambient of 0.3°C/mW. Superimpose this plot on that of (b) above.

(d) From the results of (b) and (c) discuss the possibility of thermal runaway. If the junction-to-case thermal resistance of the transistor is 0.15°C/mW, would the addition of a small heat sink make runaway impossible? Is it possible to keep the maximum quiescent collector current below 10 mA by increasing the size of heat sink?

(e) Repeat the calculations of (b) and (c) when the diode is thermally bonded to the transistor case. Assume that diode voltage changes by -3 mV/°C. Comment on the improved thermal performance.

16.5 Class B output stages require one temperature-compensating diode for each base-emitter junction. Thus, at least two, and occasionally three, diodes are needed. It may be more economical in component and design costs if a transistor is used to produce the required bias voltage V as shown in Fig. E16.5. Verify that V can be set to some convenient value which is precisely related to the transistor V_{BE}. What type of transistor is most suited to this application?

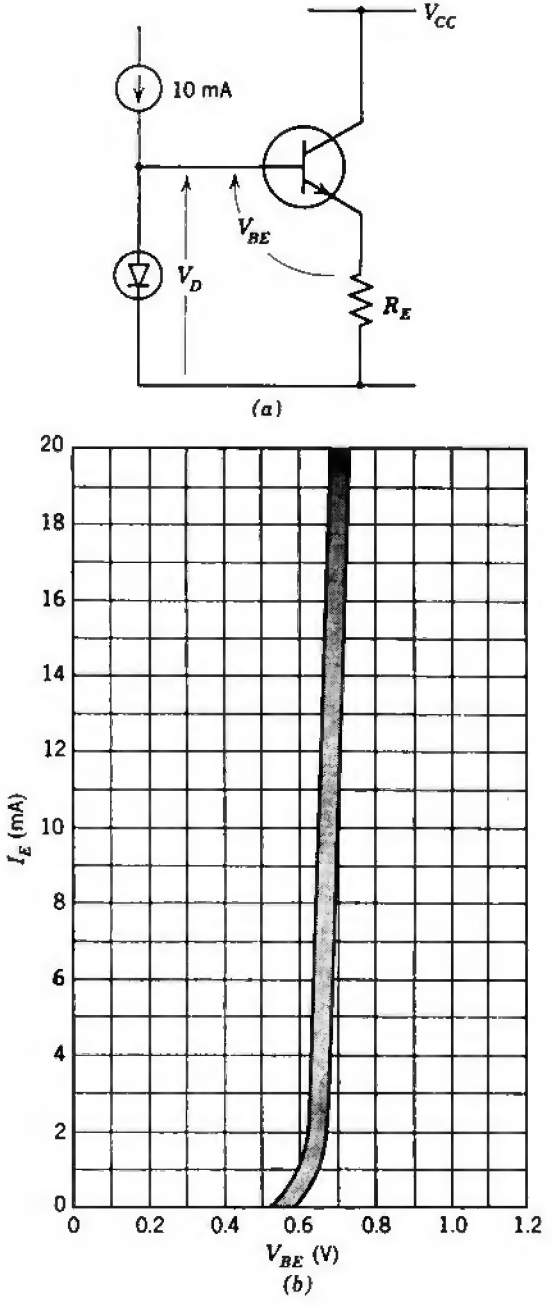

V_{CC}
10 mA
V_{BE}
V_D
R_E
(a)
20
18
16
14
12
10
8
6
4
2
0
I_E (mA)
0
0.2
0.4
0.6
0.8
1.0
1.2
V_{BE} (V)
(b)

Fig. E16.4

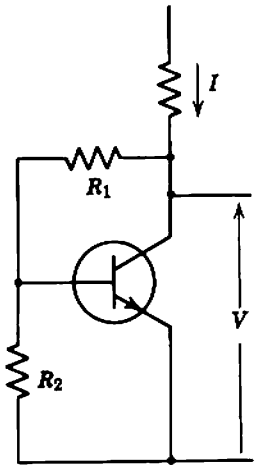


Fig. E16.5

16.6 The simple Class *A* design of Section 16.6.1 can be criticized on the following counts:

(i) The over-all signal feedback loop includes the finite impedance of the power supply.

(ii) The 6-V drop across the emitter resistor wastes excessive power.

(iii) The number of electrolytic capacitors is excessive.

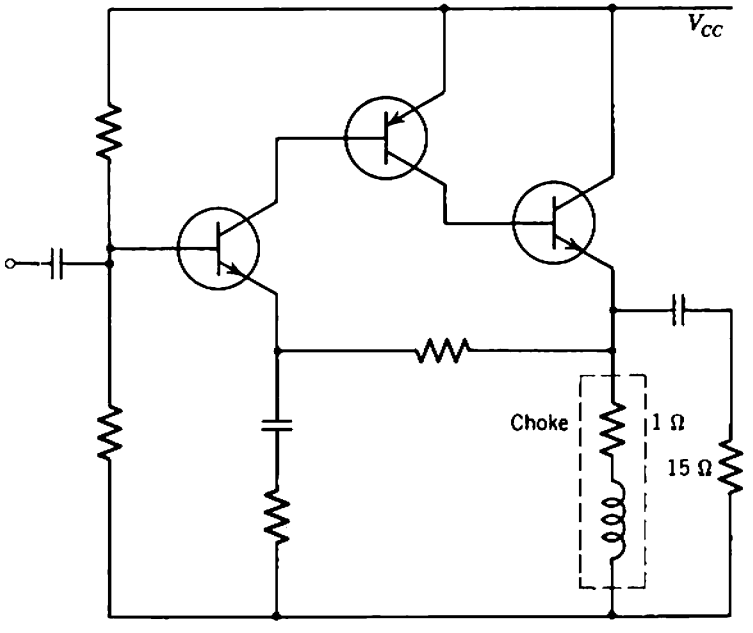
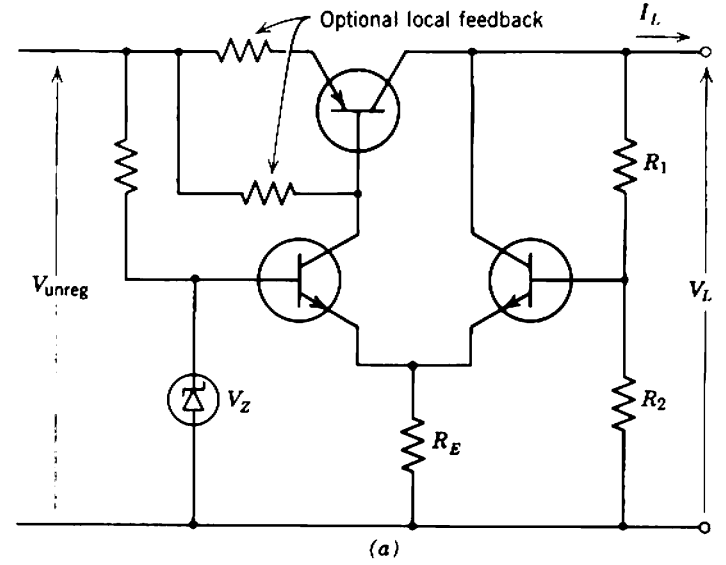
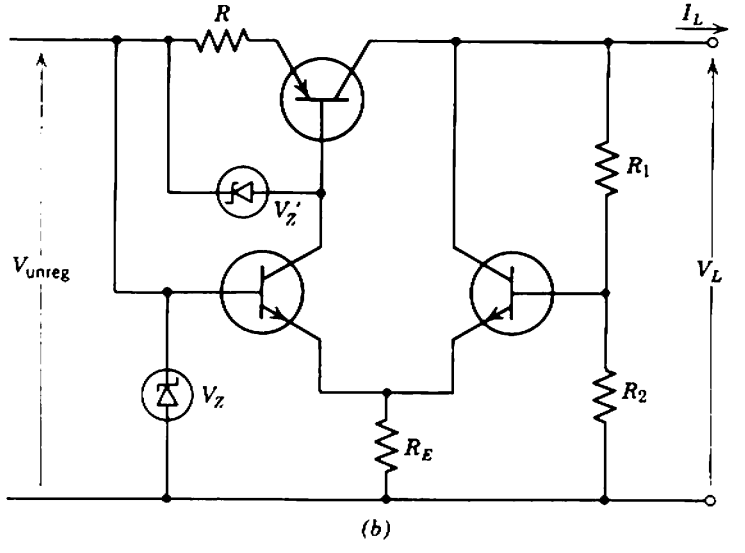


Fig. E16.6


Optional local feedback
I_L
R_1
V_{unreg}
V_L
V_Z
R_2
R_E

(a)


R
I_L
R_1
V_Z'
V_{unreg}
V_L
V_Z
R_2
R_E

(b)

Fig. E16.7

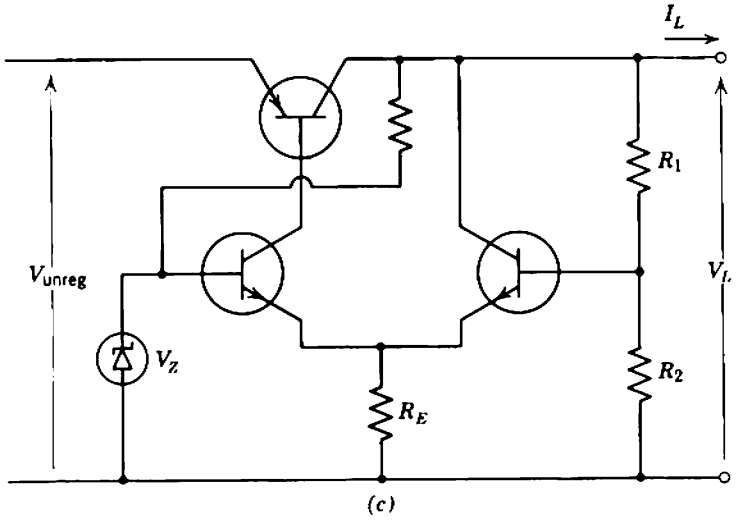


Fig. E16.7 (*continued*)

These limitations can be avoided (at the expense of extra design complication) by using the circuit of Fig. E16.6.

Complete the design of a 5-W audio amplifier to operate from a 14-V supply and to use the cheapest suitable transistors and components. Write a complete performance specification for the amplifier and, if possible, compare the cost with other competing designs.

16.7 The regulated power supply discussed in Section 16.8.2.3 uses an additional transistor to protect the output transistor when the load is short-circuited. In some instances it is possible to achieve protection without adding significantly to the basic regulation design. Figure E16.7 shows circuits which can be made self-protecting by appropriate choice of parameter values. Show the form of the load-voltage versus load-current characteristic for each circuit and comment critically on their relative merits. Investigate the feasibility of using one of these schemes in a 12-V, 100-mA regulated supply for general laboratory use. Use a capacitively filtered bridge rectifier for the unregulated supply, specify the voltage regulation with changes in load current and mains voltage, and specify the ripple content of the output voltage.

16.8 A constant current regulator is required to supply a load resistor which may vary from 0 Ω to 40 kΩ. The current should remain within 1% of its preset value for this change of load, and the preset value is to be continuously variable between 10 μA and 1 mA. If a regulated 50-V supply is available, design a suitable current regulator.

CHAPTER 17

17.1 The video amplifier of Fig. 13.35 is to be constructed on a board in such a way that electrostatic capacitances do not degrade performance. Thus, care must be taken to avoid excessive capacitance from any node to ground and the shunt capacitance of all resistors (typically 0.5 pF) must be considered.

(a) Design a practical layout in such a way that all components can be accommodated conveniently and, if necessary, removed and replaced readily. Show all pin spacings.

(b) Obtain failure-rate data from any pertinent source and estimate the mean time between failures of the complete amplifier. Make some estimate (give reasons) on the mean time to repair.

17.2 Compare the mean time between failures of the single-ended Class A power amplifiers shown in Fig. 16.49 and Fig. E16.6.

17.3 It is an interesting, if somewhat academic, exercise to investigate the effect of varying interconnections of elements on system reliability.

Find the system reliability as a function of element reliability for the situations shown in Fig. E17.3. Plot the results over the range $0.7 \leqslant p \leqslant 1.0$. For simplicity and explicitness assume that elements have identical reliabilities

$$R = p.$$

Assume also that series combinations of elements fail when any element fails, whereas parallel combinations fail only when all elements fail.

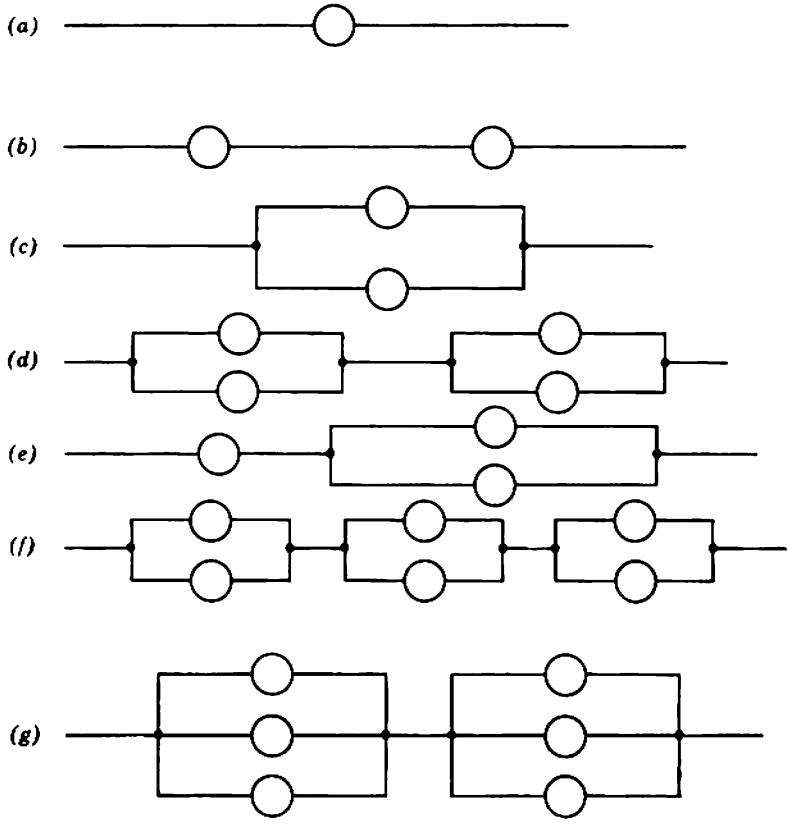


Fig. E17.3

17.4 Obtain failure-rate data on any commercial integrated-circuit operational-amplifier and compare this information with the predicted failure rate for comparable equipment using discrete devices and components.

17.5 If the dominant failure mechanism of electrolytic capacitors is a short circuit between electrodes, comment on the desirability of connecting two such capacitors in series.

Author Index

Subject Index

*There are too many entries to be listed.

*There are too many entries to be listed.

Suggested Readings . . . . .

INTRODUCTION TO FEEDBACK SYSTEMS

By L. DALE HARRIS, *University of Utah.*

"The central theme of the present text is the root locus technique and it is discussed in general and applied to the analysis of both control systems and feedback amplifiers. . . . The text has many problems and worked-out examples which enhance its expository value."—A. V. Balakrishnan in *Mathematical Reviews* *363 pages*

LINEAR NETWORK ANALYSIS

By SUNDARAM SESHU, *University of Illinois and* NORMAN BALBANIAN, *Syracuse University.*

"This presentation is written for the graduate student, preferably after he has had a course in functions of a complex variable and Laplace transforms. In spite of the extensive use of mathematics which the authors had to adopt to adequately cover the subject matter, the discussions covering each subject are exceedingly well written."—*Electrical Design News* *571 pages*

SEMICONDUCTOR ELECTRONICS EDUCATION COMMITTEE SERIES

Volume 1

INTRODUCTION TO SEMICONDUCTOR PHYSICS

By R. B. ADLER, A. C. SMITH, *both at the Massachusetts Institute of Technology and* R. L. LONGINI, *Carnegie Institute of Technology.*
247 pages

Volume 2

PHYSICAL ELECTRONICS AND CIRCUIT MODELS OF TRANSISTORS

By P. E. GRAY, *Massachusetts Institute of Technology;* D. DeWITT, *IBM Corporation;* A. R. BOOTHROYD, *Imperial College, England; and* J. F. GIBBONS, *Stanford University.* *262 pages*

Volume 3

ELEMENTARY CIRCUIT PROPERTIES OF TRANSISTORS

By C. L. SEARLE, *Massachusetts Institute of Technology;* A. R. BOOTHROYD; E. J. ANGELO, JR., *Polytechnic Institute of Technology;* P. E. GRAY *and* D. O. PEDERSON, *University of California.*
306 pages

JOHN WILEY & SONS, Inc.

605 Third Avenue **New York** **London** **Sydney**