

SOLID STATE RADIO **ENGINEERING**

Herbert L. Krauss
Charles W. Bostian
Frederick H. Raab

Solid State Radio Engineering

Solid

State

Radio

Engineering

Herbert L. Krauss
and
Charles W. Bostian

Department of Electrical Engineering
Virginia Polytechnic Institute and State University
Blacksburg, Virginia

Frederick H. Raab

Consultant
Green Mountain Radio Research Company
Burlington, Vermont

JOHN WILEY & SONS

New York • Chichester • Brisbane • Toronto • Singapore

Copyright © 1980, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

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

Library of Congress Cataloging in Publication Data

Krauss, Herbert L.

Solid state radio engineering.

Includes indexes.

1. Radio circuits. 2. Radio-Receivers and reception. 3. Radio-Transmitters and transmission. I. Bostian, Charles W., joint author. II. Raab, Frederick H., joint author. III. Title.

TK6553.K73

621.3841'2

78-27797

ISBN 0-471-03018-X

Printed in the United States of America

10 9 8 7 6 5 4

Preface

This book is about the analysis and design of the radio-frequency electronic circuits that are the building blocks of radio transmitters and receivers. It reflects the developments of the past decade, which have initiated an unprecedented growth in the use of analog radio systems for personal and business voice communications. Continuing advances in solid state technology have resulted in transmitters that are smaller, cheaper, and more reliable than ever before. Parallel developments have occurred in the home entertainment radio and television fields. As a result, radio engineers versed in the solid state art are in demand.

Because of the rapid changes in radio technology, teachers of courses in radio circuits have often had to rely on material that is scattered through many different textbooks, technical journals, and application notes. This volume meets the need for a comprehensive book on radio electronics.

Solid State Radio Engineering is unique because of its broad coverage of both receiver and transmitter circuits and its illustration of theoretical concepts with numerical examples from real circuits. Design that uses practical circuit elements instead of idealized mathematical models is emphasized. The letter symbols used for semiconductor device currents and voltages conform for the most part with IEEE Standard notation.

The last five chapters present for the first time in textbook form considerable information on RF power amplifiers. Currently, power amplifier design is often accomplished by using cut-and-try techniques and rules of thumb. Often, theoretical explanations of power amplifier operation are too complicated or require too much time for a designer to use. This book brings these principles to the student or practicing design engineer in a way that not only makes them understandable but also makes them useful for design. The discussions in Chapters 12 to 16 include not only accepted state-of-the-art technology, based on bipolar junction transistors, but also VMOS RF power FETs, high-efficiency techniques, envelope elimination and restoration, and other newly emerging technologies that are expected to play significant roles in radio engineering during the next decade.

This book is intended to be both a reference for the working engineer and a textbook for senior-level students in electrical engineering and electrical technology. A knowledge of complex algebra, Fourier series, and Fourier transforms will enable its reader to handle the mathematics in the book. As an

aid to self-study, practical design examples are included throughout. These are reinforced by homework problems that are keyed to the corresponding sections of the text.

The material presented is appropriate for either a two-semester or three-quarter course sequence. For shorter course offerings, some chapters may be omitted. For example, if receiving systems are of primary interest, Chapters 1 to 11 can be used. For transmitters, Chapters 1 to 8 and 12 to 16 (with the possible omission of Sections 14-3 to 14-6) are recommended. If the students have an adequate background in noise and modulation theory, Chapters 2 and 8 can be omitted. Prior knowledge of resonant impedance matching might permit skipping all of Chapter 3, with the exception of Section 3-6, which is used frequently in the following chapters.

A brief introductory chapter considers the concept of modulation and the functions performed in a typical transmitter and receiver. It is followed by a discussion of electrical noise because of its importance in the design of RF amplifiers and mixers in receivers. Chapters 3 to 7 include the component parts of receiver systems, and Chapter 8 provides the modulation theory necessary for an understanding of the operation of AM, SSB, FM, and TV receivers.

A thorough treatment of the design of narrowband, tapped resonant circuits for impedance matching, as well as the use of tapped mutual inductance circuits for both wideband and narrowband matching, is given in Chapter 3. The design of small-signal, tuned amplifiers for maximum gain with a specified degree of stability is considered next. This is followed by an analysis of sinusoidal oscillations in *LC* and crystal oscillator circuits; and a unique, laboratory-tested procedure is given for the design of a common-base Colpitts oscillator for specified output.

A phase-locked loop will soon be included in nearly every radio receiver, transmitter, and piece of test equipment. Hence, the simpler aspects of loop operation are outlined in Chapter 6, along with the characteristics of the basic loop components and some applications to communication equipment. This is followed by analysis of diode, BJT, and FET mixer circuits.

Chapters 9 to 11 are devoted to receivers. Because design techniques are changing constantly with the introduction of new integrated-circuit packages, the fundamental signal processing in each type of receiver is stressed without a detailed description of all possible circuits. Analysis of various types of AM and FM detectors is found in the corresponding chapters, along with information on ceramic, crystal, and surface-acoustic-wave IF filters. The basic principles of color picture transmission are given in Chapter 11, along with a block-diagram explanation of the receiver circuitry. Although not directly related to the subject matter of the book, this is the one piece of radio

equipment that the student owns and uses every day. Furthermore, it incorporates most of the principles studied in preceding chapters.

Chapters 12 to 16 are organized by practice rather than theory. Thus Class A and B amplification along with the broadband transformer and filter networks normally used in SSB transmitters are discussed first. Similarly, Class C and Class C mixed-mode power amplifiers and the discrete-element or transmission-line matching networks normally used with them, are presented together. Chapter 14 treats several types of high-efficiency power amplifiers (Classes D, E, F, and S). Chapter 15 includes CW, FM, and AM transmitters, since they have similar configurations. The last chapter examines single-sideband and multimode transmitters, envelope elimination and restoration, and other related techniques.

Herbert L. Krauss
Charles W. Bostian
Frederick H. Raab

Acknowledgments

This text originated with class notes written by H. L. Krauss; Chapters 1 to 11 were completed with the collaboration of C. W. Bostian. Chapters 12 to 16 were contributed by F. H. Raab.

Assistance in manuscript preparation was provided by members of the secretarial staff of the Department of Electrical Engineering at Virginia Polytechnic Institute and State University. We are indebted to Professors C. A. Holt and E. A. Manus for their helpful advice on various technical points, and our special thanks go to Professors R. O. Claus and H. R. Skutt, who class-tested the material and provided many worthwhile suggestions.

In the early stages of this project we received invaluable technical assistance from Roy C. Hejhall, of Motorola Semiconductor Products, and Ed Oxner, of Siliconix, Inc. For the material on power amplifiers, N. O. Sokal, of Design Automation, Inc., and A. D. Sokal, of the Physics Department, Princeton University, have been most helpful critics.

Finally, we thank our wives, Anne Krauss, Frieda Bostian, and Becky Raab, for their patience, devotion, and encouragement during the writing of this book.

H. L. K.

C. W. B.

F. H. R.

Contents

Chapter 1 Radio Communication Systems	1
1-1 Introduction	1
1-2 Elements of a Radio System	2
1-3 Modulation	5
1-4 Frequency and Time Multiplexing	7
1-5 Comparison of Modulation Systems	8
 Chapter 2 Electrical Noise	 9
2-1 Thermal Noise in Resistors and Networks	11
2-2 Noise in Receiving Antennas	14
2-3 Noise in Diodes, Transistors, and FETs	15
2-4 Definitions of Noise Terms	20
2-5 The Noise Figure	26
2-6 Amplifier Noise Considerations	30
Appendix 2-1 Choice of R_s to Minimize the Noise Figure	36
 Chapter 3 Resonant Circuits and Impedance Transformation	 38
3-1 Series Resonance	39
3-2 Parallel Resonance	42
3-3 Parallel Resonance with Series Load Resistance	43
3-4 Effects of Source and Coil Resistances	46
3-5 Parallel-to-Series Conversions for RC and RL Circuits	47
3-6 Tapped Resonant Circuits	48
3-7 Tapped Coil with Mutual Inductance	54
3-8 Single-Tuned Transformer	63
3-9 Double-Tuned Transformers	69
Appendix 3-1 Tables of Design Formulas	76

Chapter 4 Small-Signal High-Frequency Amplifiers	83
4-1 Definition of a Small-Signal Amplifier	83
4-2 Models for the Active Device	86
4-3 Amplifier Stability	89
4-4 Achieving Stability	93
4-5 Amplifier Power Gain	94
4-6 Design with Unconditionally Stable Device	96
4-7 Design with Potentially Unstable Device	99
4-8 Maximum Transducer Gain with Specified Stability	101
4-9 Alignability	102
4-10 Overall Design of the Tuned Amplifier Stage	102
Appendix 4-1 Relations Between Two-Port Parameters	110
Appendix 4-2 The y Parameters of the Hybrid-Pi Circuit	114
Appendix 4-3 Corresponding Quantities in Various Parameter Systems	115
Appendix 4-4 RF Transistor Data	117
Appendix 4-5 Self-Resonance Frequencies for Several Types of Capacitors and RF Chokes	125
 Chapter 5 Sinewave Oscillators	 128
5-1 The Criteria for Oscillation	128
5-2 Negative Resistance Oscillators	129
5-3 Feedback Oscillators	132
5-4 Oscillator Design Techniques	134
5-5 Colpitts Oscillator Analysis and Design	135
5-6 Other Oscillator Circuits	146
5-7 Maximum-Efficiency Oscillators	151
5-8 Crystal-Controlled Oscillators	151
5-9 Buffering	160
5-10 Frequency Stability	161
 Chapter 6 Phase-Locked Loops	 164
6-1 Simplified Explanation of PLL Operation	164
6-2 Linear Analysis of the PLL	165

6-3	Phase-Locked Loop Terminology	171
6-4	The Loop Oscillator	173
6-5	The Phase Detector	175
6-6	Phase-Locked Loop Applications	179
6-7	Example of PLL Design	184
Chapter 7 Mixers		188
7-1	Basic Mixer Theory and Spectral Analysis	188
7-2	Mixer Terminology	192
7-3	Balanced Diode Mixers	196
7-4	FET and BJT Mixers	201
7-5	Transistor (BJT) Mixers	202
7-6	FET Mixers	204
7-7	JFET Mixer Design	211
7-8	MOSFET Mixer Design	212
Chapter 8 Modulation		221
8-1	Amplitude Modulation	222
8-2	Double-Sideband and Single-Sideband Systems	228
8-3	Generation of Single-Sideband Signals	229
8-4	Angle Modulation	234
8-5	Spectra of Angle-Modulated Waves	240
8-6	Phasor Diagrams of Angle-Modulated Waves	245
8-7	Comparison of FM and PM	247
8-8	Pulse Modulation	250
8-9	Information Capacity of a Channel	258
	Appendix 8-1 Derivation of the Spectrum of Angle-Modulated Waves	263
Chapter 9 Amplitude Modulation Receivers		266
9-1	Receiver Performance Specifications	266
9-2	The RF Amplifier	268
9-3	The Mixer	269
9-4	The Local Oscillator	270
9-5	The IF Amplifier	271

9-6	Interstage Filters	272
9-7	Diode Envelope Detectors	280
9-8	Product Detectors	284
9-9	Automatic Gain Control	286
9-10	Squelch Circuit	288
9-11	The AM Receiver	289
Chapter 10 FM and PM Receivers		295
10-1	IF Amplifier Systems	295
10-2	FM Detector Characteristics	297
10-3	Practical Detectors	301
10-4	FM Stereo Reception	318
10-5	A Note on Quadraphonic Sound	322
10-6	Audio Considerations: Preemphasis and Deemphasis	322
10-7	An Example of a Complete FM Receiver	325
Chapter 11 Television Receivers		329
11-1	Monochrome Television	329
11-2	Bandwidth of the Composite Video Signal	332
11-3	Vestigial-Sideband Transmission	333
11-4	The Monochrome Receiver	334
11-5	Color Television	334
11-6	Transmission of Chrominance Information	340
11-7	The Color Receiver	342
Chapter 12 Linear Power Amplifiers		348
12-1	Class A Amplification	352
12-2	Class B Amplification	355
12-3	Practical Considerations	359
12-4	Intermodulation Distortion and Bias	361
12-5	Drive and RF Feedback	366
12-6	Wideband Transformers	371
12-7	Power Combiners and Splitters	379
12-8	Output Filters	382
12-9	Heatsink Design	384
	Appendix 12-1 RF Power Transistors	389

Chapter 13 Tuned Power Amplifiers	394
13-1 Current-Source Class C Amplifiers	394
13-2 Saturating Class C Amplifiers	399
13-3 Solid State Class C Mixed-Mode Amplifiers	402
13-4 Drive and Bias	408
13-5 Amplitude-Modulation Characteristics	410
13-6 Frequency Multipliers	412
13-7 Impedance Matching	417
Chapter 14 High-Efficiency Power Amplifiers	432
14-1 Class D Amplification—Idealized Operation	433
14-2 Class D Amplification—Practical Considerations	441
14-3 Class E Amplification	448
14-4 Class F Amplification	454
14-5 Class S Amplification	458
14-6 Other High-Efficiency Amplifiers	468
Appendix 14-1 Tabulation of PA Characteristics	472
Appendix 14-2 Performance of Class E PA	474
Appendix 14-3 Distortion in Pulse-Width Modulation	476
Chapter 15 CW, FM, and AM Transmitters	477
15-1 CW Transmitters	478
15-2 FM Transmitters	482
15-3 Amplitude-Modulated Transmitters	490
Chapter 16 Single-Sideband Transmitters	500
16-1 Transmitter Organization	501
16-2 Linear Amplifier Chains	505
16-3 Peak and Average Power	507
16-4 AGC and VSWR Protection	511
16-5 Envelope Elimination and Restoration	515
Index	525

Solid State Radio Engineering

1 Radio Communication Systems

1-1 Introduction

This book is devoted to the analysis and design of the electronic circuitry used in radio communication systems. It presumes that the reader is familiar with audio-frequency amplifiers; hence the primary emphasis will be on radio transmitter and receiver circuits. The transmitting and receiving antennas and the propagation path between them are important parts of an overall system, but a discussion of these elements is left to other texts.

Communication systems transmit information in the form of electrical signals that represent speech, music, television pictures, scientific and business data, and so forth. The waveforms of these signals are complex and continually changing, but the frequency spectrum of the signals is usually limited to a specified bandwidth either by the nature of the signal source or by filters in the transmitting equipment. Since many of these signals occupy a frequency band that extends downward to a few hertz, they cannot be transmitted in their original form over a common transmission path because it would not be possible to separate them at the receiving end. A separate transmission line or separate radio path for each signal would not be feasible from either an economic or a practical standpoint. Hence the overall communication system must provide a means for simultaneous transmission of a number of signals either by shifting them into different parts of the frequency spectrum or by sending samples of the signals on a time-shared basis.

The wavelength (λ) in meters of a radio wave is given by c/f , in which c is

the velocity of light (3×10^8 meters per second), and f is in hertz. (For RF calculations it is convenient to remember that f in megahertz $\times \lambda$ in meters = 300.) A radio antenna should have a physical size of one-half wavelength or more for reasonable efficiency. Hence, as the transmission frequency is increased, the physical size and cost of the antenna are reduced and its efficiency increases.

1-2 Elements of a radio system

The process whereby the original message is converted into a new form suitable for radio transmission is called modulation. The modulation process causes some property—such as the amplitude, frequency, or phase—of a high-frequency carrier¹ wave to be deviated from its unmodulated value by an amount proportional to the instantaneous value of the modulating (message) signal. Thus the content of the original message is shifted to a portion of the frequency spectrum in the vicinity of the carrier frequency. In the receiver this process is reversed in a detector that recovers the original signal.

Figure 1-1 shows a simplified block diagram of a radio transmitter and receiver in order to illustrate the signal processing that takes place. The function of each block is explained below.

1. The source of the message signal may be a microphone, phono pickup, television camera, or other device that transforms the desired information into an electrical signal.

2. The signal is amplified and often passed through a low-pass filter to limit the bandwidth.

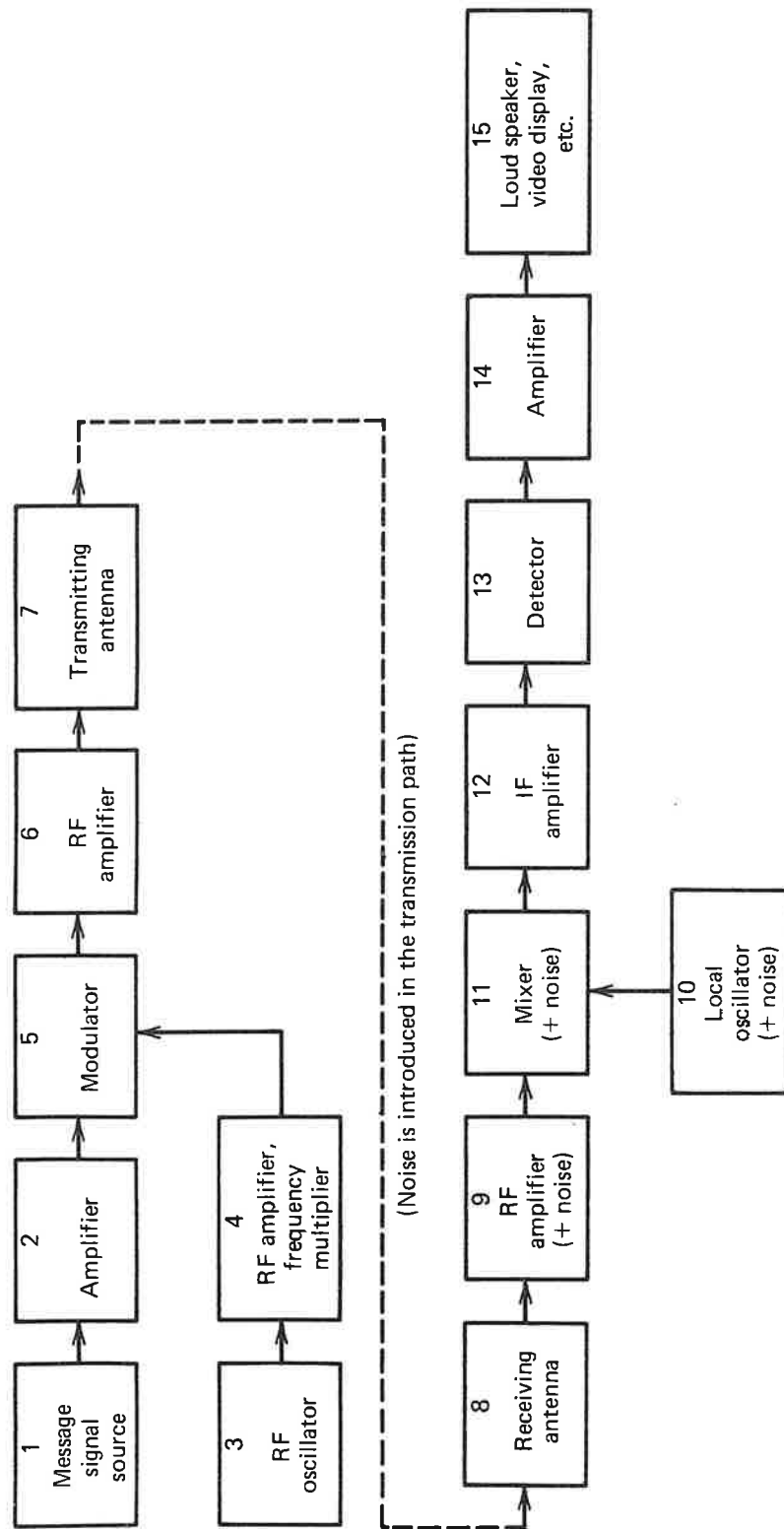
3. The RF oscillator establishes the carrier frequency or some submultiple of it. Since good frequency stability is required to keep the transmitter on its assigned frequency, the oscillator is often controlled by a quartz crystal.

4. One or more amplifier stages increase the power level of the signal from the oscillator to that needed for input to the modulator. Class C operation is used wherever possible to obtain high efficiency. Tuning of the output circuits to a harmonic of the input frequency results in “frequency multiplication” so that the final carrier frequency can be a multiple of the oscillator frequency.

5. The modulator combines the signal and carrier frequency components to produce one of the varieties of modulated waves discussed in Section 1-3. In the simplified system shown in Fig. 1-1 the output signal spectrum lies in the vicinity of the desired RF carrier frequency. In many transmitters a second oscillator and mixer (similar to blocks 10 and 11) are inserted between

¹The “carrier” may be a sinusoidal wave or a train of pulses.

Fig. 1-1 Radio transmitter and receiver block diagram.



blocks 5 and 6 in order to shift the modulated wave to a higher-frequency range.

6. Additional amplification may be required after modulation to bring the power level of the signal to the desired value for input to the antenna.

7. The transmitting antenna converts the RF energy into an electromagnetic wave of the desired polarization. If a single (fixed) receiver is to be reached, the antenna is designed to direct as much of the radiated energy as possible toward the receiving antenna.

8. The receiving antenna may be omnidirectional for general service or highly directional for point-to point communication. The wave propagated from the transmitter induces a small voltage in the receiving antenna. The range of amplitudes of the induced antenna voltage may be from tens of millivolts to less than 1 microvolt, depending upon a wide variety of conditions.

9. The RF amplifier stage increases the signal power to a level suitable for input to the mixer and it helps to isolate the local oscillator from the antenna. This stage does not have a high degree of frequency selectivity but does serve to reject signals at frequencies far removed from the desired channel. The increase in signal power level prior to mixing is desirable because of the noise that is inevitably introduced in the mixer stage.

10. The local oscillator in the receiver is tuned to produce a frequency f_{LO} that differs from the incoming signal frequency f_{RF} by the intermediate frequency f_{IF} ; that is, f_{LO} can be equal to $f_{RF} + f_{IF}$ or $f_{RF} - f_{IF}$.

11. The mixer is a nonlinear device that shifts the received signal at f_{RF} to the intermediate frequency f_{IF} . Modulation on the received carrier is also transformed to the intermediate frequency.

12. The IF amplifier increases the signal to a level suitable for detection and provides most of the frequency selectivity necessary to "pass" the desired signal and filter out the undesired signals that are found in the mixer output. Because the tuned circuits in blocks 11 and 12 always operate at a fixed frequency (f_{IF}), they can be designed to provide good selectivity. Ceramic or crystal filters are often used.

13. The detector recovers the original message signal from the modulated IF input.

14. The audio or video amplifier increases the power level of the detector output to a value suitable for driving a loudspeaker, a television tube, or other output device.

15. The output device converts the signal information back to its original form (sound waves, picture, etc.).

In addition to the desired signal that is processed by the receiver, electrical noise is added in the transmission path, and is generated within the RF

amplifier, local oscillator, mixer, and so forth. The block diagrams shown in Fig. 1-1 are for illustrative purposes only. In practice, so many variations in transmitter and receiver systems are encountered that no single block diagram could even be considered typical. The general layout of receivers or transmitters for particular applications will be discussed in detail in later chapters.

1-3 Modulation

To extend the concept of modulation introduced in the previous section, the basic definitions of commonly used types of modulation will be given here. Let the voltage of an unmodulated carrier wave be given by

$$v(t) = V_c \sin(\omega_c t + \phi) = V_c \sin \theta(t) \quad (1-1)$$

in which ω_c is the carrier (radian) frequency, V_c the amplitude, and ϕ an arbitrary phase angle.

Amplitude Modulation

In an amplitude-modulated (AM) wave, the deviation of the amplitude V_c from its unmodulated value is proportional to the instantaneous value of the modulating wave. In other words, if the modulating signal is $F(t)$, the carrier amplitude must vary in time according to the expression

$$V_c(t) = V_c[1 + m_a F(t)] \quad (1-2)$$

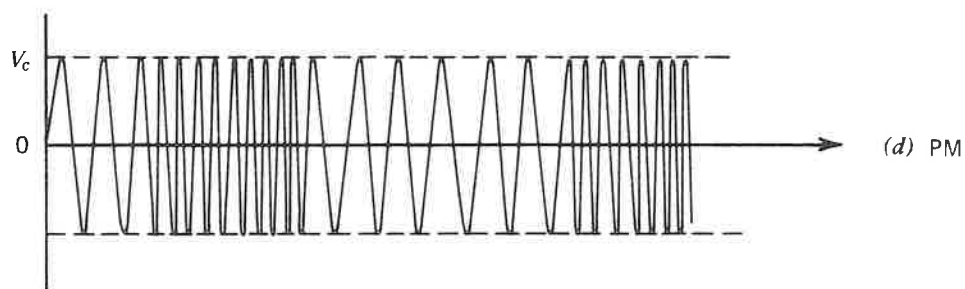
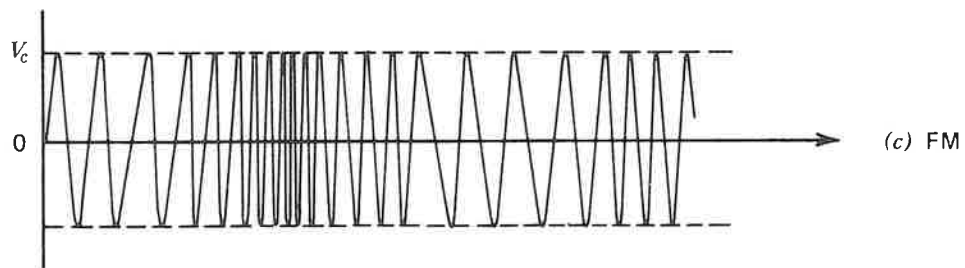
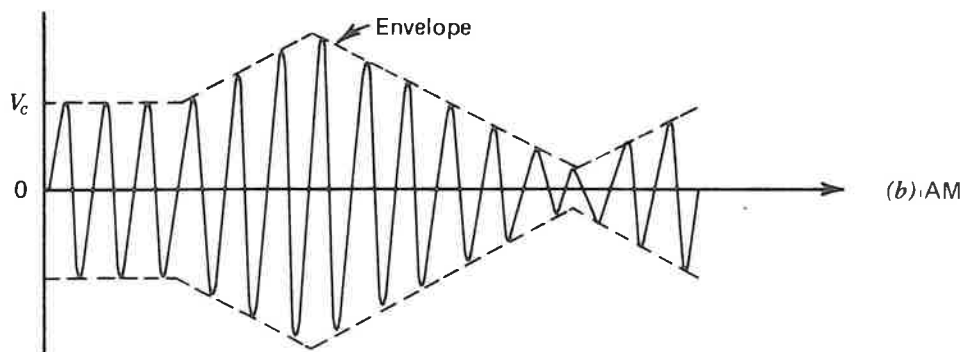
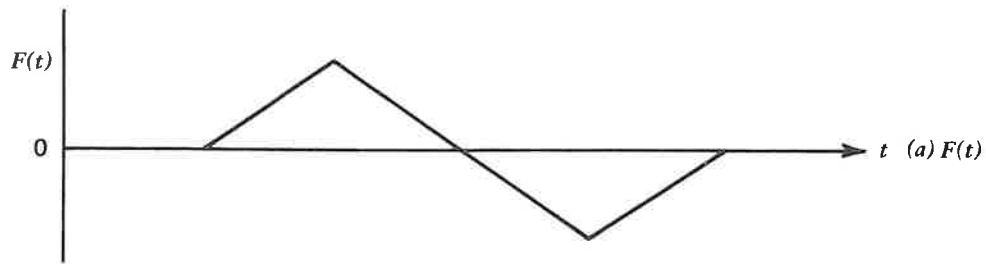
in which m_a is the *modulation factor*. The value of $m_a F(t)$ cannot exceed unity without introducing distortion. Figure 1-2a illustrates a modulating signal $F(t)$, and the corresponding amplitude-modulated wave is shown in Fig. 1-2b. Note that the *envelope* of the AM wave [given by (1-2)] has the same shape as the modulating signal.

Angle Modulation

In angle modulation, the angle $\theta(t)$ in (1-1), rather than the amplitude, is varied from its unmodulated value by the modulating signal. Phase and frequency modulation are particular forms of angle modulation. In *phase modulation* (PM), the angle $\theta(t)$ in (1-1) is made to vary proportionately to the modulating signal $F(t)$. In *frequency modulation* (FM) the instantaneous frequency,

$$\omega(t) = \frac{d\theta(t)}{dt}$$

Fig. 1-2 (a) A triangular modulating signal; and the resulting (b) AM, (c), FM, and (d) PM waves.



is caused to vary from its unmodulated value in a manner proportional to $F(t)$. In both cases the amplitude of the wave remains constant. Figure 1-2 illustrates AM, FM, and PM waves resulting from a triangular modulating signal.

Pulse Modulation

In pulse modulation systems, the “carrier” consists of a train of pulses that can be modulated in amplitude, repetition frequency, or spacing in a manner completely analagous to the AM, FM, and PM waves previously discussed. The sampling theorem shows that the continuous transmission of a message signal is unnecessary—it can be reconstructed completely from samples taken at a rate at least twice the highest frequency present in the signal. Thus, for a signal that is bandlimited to 4 kHz, a modulated pulse train with a mean repetition frequency of 8 kHz is sufficient and the pulses can be of arbitrarily short duration.

In pulse code modulation, each sample is represented by a set of seven or more pulses (or spaces) that represent a binary code for the sample amplitude. This system has greater immunity to noise at the expense of higher pulse repetition rates and greater bandwidth.

1-4 Frequency and Time Multiplexing

The modulated pulse train produced by any of the foregoing methods is still of relatively low frequency. For radio transmission, this pulse train forms the modulating signal that in turn modulates (by AM, FM, etc.) a high-frequency carrier.

In *frequency-division multiplexing* (FDM) the modulating process shifts the signal to a portion of the frequency spectrum in the vicinity of the carrier frequency. Since the portion of the spectrum to be utilized is determined by the carrier frequency, different signals can modulate carriers of different frequencies and all of them can be transmitted simultaneously. The receiver can choose the desired signal band by means of selective filters. This is done, for example, in AM, FM, and TV broadcasting, and in long-distance carrier-telephone systems.

Time-division multiplexing (TDM) is an alternate process in which a number of signals are transmitted over a common facility by letting the signals occupy the same frequency band on a time-sharing basis. This method is used with pulse modulation systems. The individual “samples” of a given signal are of sufficiently short duration that the time between samples can be used to transmit other signals. In such systems the transmitter is switched or

commutated to each signal channel sequentially. The receiving system must then be switched in synchronism with the transmitter to separate the various signals prior to final demodulation.

1-5 Comparison of Modulation Systems

Each of these modulation systems has its good and bad features. Amplitude modulation utilizes the simplest detectors and requires the least bandwidth (particularly if only one sideband is transmitted). However, it has the lowest immunity to noise and both transmitter and receiver circuitry becomes more complicated if single-sideband transmission is used.

Wideband frequency modulation employs simpler transmitter circuitry and provides much better rejection of noise and interfering signals than AM. It requires a bandwidth approximately five times that of a comparable AM signal, however. Pulse code modulation (PCM) provides even better transmission in the presence of noise but requires a further increase in circuit complexity and bandwidth. The choice of a particular technique depends upon the communication system requirements.

10 FM and PM Receivers

Receivers for FM and PM differ from their AM counterparts in that they have (1) larger IF bandwidths, (2) amplitude limiters, (3) frequency or phase detectors, and (4) stereo multiplex capability. Except for the increased bandwidth (200 kHz for broadcast FM, less for other services), the front-end design is the same as for AM receivers.

10-1 IF Amplifier Systems

The wide bandwidths required by angle-modulated signals and the necessity for amplitude-limiting the IF signal before detection (because most detectors respond both to amplitude and frequency or phase variations) once necessitated such specialized design techniques as stagger-tuned transformers and multiple-stage limiters in FM IF systems. The present availability of low-cost wideband integrated-circuit (IC) amplifiers and crystal or ceramic filters that outperform discrete-component IF strips has obviated the need for such techniques; hence the following discussion is restricted to IC system design.

The IF frequencies (or the last IF frequency in the case of multiple-conversion systems) are usually higher than in AM receivers; for example, the most common IF frequencies are 10.7 MHz for FM and 455 kHz for AM. The larger bandwidth required for FM dictates a higher IF frequency in order to ease the design requirements on the IF filter. This filter should have a sharp cutoff in order to reject adjacent channel signals, but at the same time to

avoid distortion it must exhibit phase linearity (vs. frequency) in the pass-band.

Phase distortion in an FM or PM system is the counterpart of amplitude distortion in an AM system. It arises when the relative phase relationships between the FM carrier and sidebands are altered and it hampers accurate recovery of the modulating waveform [1]. If the overall IF system (tuned circuits, filters, and amplifiers) has a transfer function whose phase decreases linearly with frequency, it will amplify FM signals without phase distortion. This may be shown mathematically [2]; intuitively, it may be argued that a network with a linearly decreasing phase characteristic simply introduces a frequency-independent time delay and thus passes the FM carrier and sidebands without changing their relative phases.

Unfortunately, there are no convenient formulas to relate phase non-linearity between the antenna and the detector to the distortion produced in an FM receiver. The designer seeks as linear a phase characteristic as can be justified economically and may achieve it by using components whose phase nonlinearities cancel each other. Phase distortion is obviously a more serious problem to the manufacturer of a high-quality home entertainment receiver than it is to the manufacturer of a taxicab radio.

A typical IF system (Fig. 10-1) consists of an *LC* tuned circuit, an integrated circuit amplifier, a crystal or ceramic IF filter, and either a detector or a second IF amplifier and then a detector. Hard limiting is achieved by driving the last amplifier stages into saturation and cutoff for all IF signals within the intended dynamic range of the receiver. The result is that the signal fed from the last IF amplifier into the detector is a frequency-modulated square wave.

Most commercial FM detectors are included in IC packages that also contain several stages of IF amplification. The detector specifications published by the manufacturer list the input voltages required for a variety of output characteristics such as decibels of quieting, SNR, recovered audio output, and limiting. The minimum input power required for satisfactory detector operation and the input power available from the antenna under the

Fig. 10-1 Block diagram of an IC IF system.

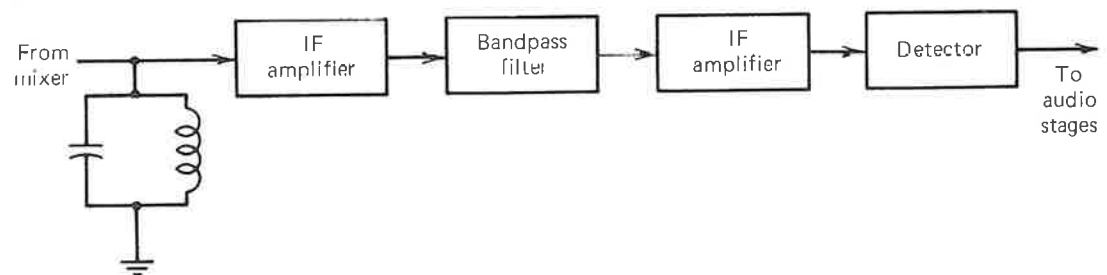
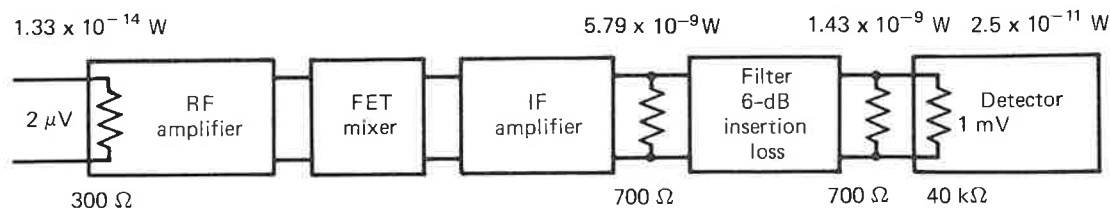


Fig. 10-2 Block diagram of an FM receiver.



desired minimum-signal conditions determine the overall gain required between the antenna terminals and the detector. The designer must apportion this gain between the RF amplifier, the mixer, and the IF system.

As an example of the calculations involved in the preliminary design of an FM receiver IF system consider the block diagram in Fig. 10-2. Assume that the minimum signal to be received will deliver $2\ \mu\text{V}$ into a 300-ohm RF amplifier input resistance and that a detector IC has been chosen which requires $1\ \text{mV}$ into $40\ \text{k}\Omega$ for full limiting. The detector is to be preceded by a ceramic filter requiring 700-ohm source and load resistances and having a 6-dB insertion loss. Under full limiting conditions the detector receives $2.5 \times 10^{-11}\ \text{W}$ and the filter-terminating resistor draws $1.429 \times 10^{-9}\ \text{W}$. The total power output from the filter is $1.454 \times 10^{-9}\ \text{W}$; if a 6-dB conversion loss is assumed, the IF stage must deliver $5.787 \times 10^{-9}\ \text{W}$. The power input to the RF amplifier is $1.333 \times 10^{-14}\ \text{W}$, and the overall antenna terminals to IF filter gain required is $56.4\ \text{dB}$. Since commercial IF amplifier ICs are capable of about 40-dB gain, a conservative 10 to 15-dB gain each may be allotted to the RF amplifier and to the mixer.

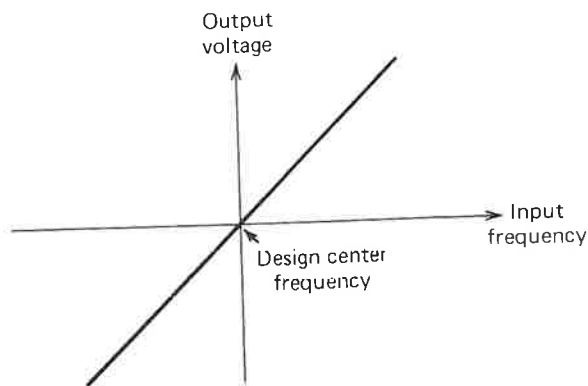
10-2 FM Detector Characteristics

Basically, an FM detector is a circuit whose output voltage is proportional to the difference between a reference frequency and the frequency of an input signal. Normally, the detector is adjusted so that the output voltage has the same magnitude, but opposite polarity, for input frequencies equally spaced above and below the reference. The transfer characteristic of an ideal FM detector is shown in Fig. 10-3. The characteristic of a real detector has finite frequency and voltage limits and is not perfectly linear.

Analysis of a Generalized FM Detector

Those FM detectors that include tuned transformers function as a frequency-to-amplitude converter driving a conventional AM envelope detector and a low-pass filter. Although this also at least partially describes the phase-locked

Fig. 10-3 Transfer characteristic of an ideal FM detector.



loop (PLL) detector, other circuits discussed in this chapter develop a train of pulses whose width is proportional to the input frequency and then integrate the pulse train to recover the modulation envelope. Nevertheless, all of these detectors have the same kind of transfer function, and the analysis to follow will illuminate the common characteristics of them all.

Following the approach of Taub and Schilling [3], and working in the frequency domain with the notation of Fig. 10-4, the frequency-to-amplitude transfer function $H(j\omega)$ is given by

$$H(j\omega) = j\sigma\omega = \frac{V_2(j\omega)}{V_1(j\omega)} \quad (10-1)$$

in which σ is a constant. In the time domain, (10-1) implies that

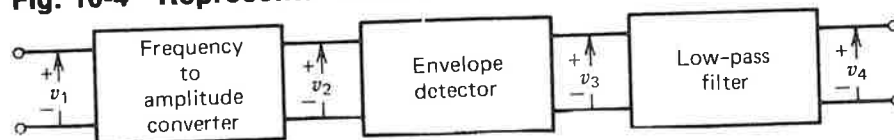
$$v_2(t) = \sigma \frac{dv_1(t)}{dt} \quad (10-2)$$

and the frequency-to-amplitude converter functions as a time-domain differentiator. The proportionality in (10-1) of output voltage to input frequency causes white noise in the upper part of the IF passband to produce correspondingly greater output than noise in the lower part. (See also the discussion in Section 8-7.)

If a signal with fixed amplitude A , carrier frequency ω_o , and variable phase angle $\theta(t)$ such that

$$v_1(t) = A \cos [\omega_o t + \theta(t)] \quad (10-3)$$

Fig. 10-4 Representation of an FM detector.



is applied to the converter, the output will be

$$v_2(t) = -\sigma A \left[\omega_o + \frac{d\theta}{dt} \right] \sin [\omega_o t + \theta(t)] \quad (10-4)$$

With input $v_2(t)$, the envelope detector output is

$$v_3(t) = \sigma A \omega_o + \sigma A \frac{d\theta}{dt} \quad (10-5)$$

The first term in (10-5) is a dc component that will be filtered out along with $\sigma A(d\theta/dt)$ components above the maximum modulating frequency. In an FM signal with modulating voltage $v_m(t)$ and modulation index m_f ,

$$\theta(t) = m_f \int v_m(t) dt \quad (10-6)$$

and the alternating component of $v_3(t)$ is equal to $\sigma A m_f v_m(t)$. Thus the modulating waveform $v_m(t)$ is recovered.

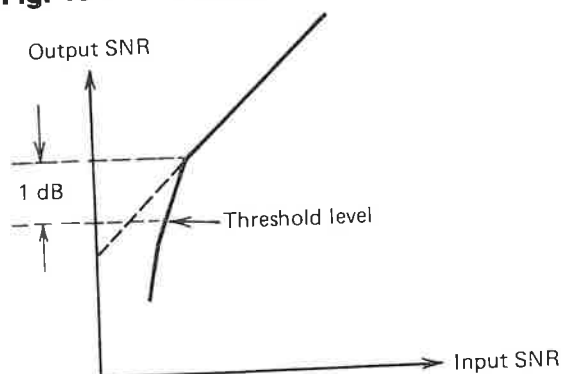
Response to Interference and Noise

When either noise or an interfering signal is present along with a wanted signal at an FM detector input, the SNR determines the nature of the detector output. For a given modulation index and detector type, a *threshold* SNR exists. So long as the input SNR is above this threshold, the output SNR will be significantly greater than the input SNR. Since the bandwidth of the IF signal at the detector input is larger than the bandwidth of the audio signal at the detector output, the SNR improvement with detection is commonly described as "trading bandwidth for SNR." It comes about because (1) signal power in the entire signal bandwidth contributes to the output signal power, but (2) only noise power within plus or minus the highest modulating frequency of the carrier contributes to the output noise power. Hence, *above threshold*, if the highest modulating frequency is fixed and the modulation index is increased, the output signal power will increase but the output noise power will not. However, raising the modulation index will raise the detector threshold and ultimately the input SNR will fall below threshold. Below threshold the output SNR improves with *decreasing* modulation index and IF bandwidth. This behavior resembles that of an AM detector. Figure 10-5 illustrates the general threshold behavior of a hypothetical FM detector.

The threshold sets the ultimate sensitivity limit of an FM detector, and considerable effort has been devoted to development of detectors with the lowest possible threshold. At present, one of the lowest thresholds is offered by the phase-locked loop [4].

The improvement of SNR with detection above threshold is responsible

Fig. 10-5 Threshold in an FM detector.



for the phenomenon called *capture*, in which only the stronger of two or more interfering signals is audible at the receiver output. If the ratio of desired to undesired signal is sufficiently high at the detector input, it will be so much higher at the output that the undesired signal effectively disappears. This behavior is in sharp contrast to AM systems, in which annoying interference can result even with a large ratio of desired to undesired signals. With today's crowded spectrum, the capture effect is one reason that mobile radio communications use FM.

Capture is measured by a quantity called the *capture ratio*. For practical purposes this is essentially the ratio in decibels of wanted to unwanted signal that will cause a 30-dB suppression of the unwanted signal. The formal (and exact) definition gives the capture ratio of an FM receiver as the result of a complicated measurement process in which the suppression of the signal from one signal generator by the signal from a second signal generator is observed at a series of standard input levels. The reader who needs to make capture ratio measurements should consult [5]. Typical values for good-quality home entertainment receivers lie between 1 and 2 dB.

Capture is responsible for an effect called *quieting*, where the audible noise output from an FM receiver decreases sharply or vanishes entirely with the appearance of a sufficiently strong input signal. What happens is that the large increase in postdetection SNR, which occurs above threshold, makes the output signal so much stronger than the noise that the noise is no longer heard.

Quieting may be described mathematically [1] by representing the noise input to the FM detector as a phasor with random amplitude and phase. Sudden phase and instantaneous frequency changes by this noise voltage produce audio noise at the detector output. When a signal is present, the input voltage to the detector can be represented as the phasor sum of the signal plus noise; when the signal amplitude is greater than the noise amplitude the phase and instantaneous frequency changes of the composite waveform are much

smaller than those of the noise alone and very little noise is present at the detector output.

The *quieting sensitivity* of an FM receiver is defined as the unmodulated carrier amplitude measured in microvolts rms that produces a specified reduction in output noise over the output noise present under no-signal conditions. A quieting sensitivity specification for a typical narrow-band FM receiver would be 0.5 microvolts for 20-dB quieting; a broadcast receiver might require 1.7 microvolts for 30-dB quieting.

10-3 Practical Detectors

FM detectors have existed for over 40 years, and many detector circuits have been manufactured that emphasize the technology available at a particular time. In vacuum-tube days, active devices were expensive, bulky, short-lived, and hot, and engineers designed FM receivers around circuits that required the fewest possible active components. Thus the earlier Foster-Seeley discriminator was later supplanted in many applications by the ratio detector, since the ratio detector was less sensitive to amplitude variations in the incoming signal and could tolerate IF amplifiers with fewer stages. With the advent of ICs, the economics of detector design were reversed and active devices became much cheaper to produce and package than tuned circuits; hence the design goal became one of eliminating inductors and transformers whenever possible. This effort was encouraged by the availability of ICs with hard-limiting IF stages, since the older tuned-transformer detectors could not demodulate a frequency-modulated square wave.

Which FM detector should be selected for a particular application? There seems to be no consensus among design engineers in industry at present. Some manufacturers are still using ratio detectors and discriminators; others have gone to fully digital detectors, and still others are using quadrature detectors or phase-locked loops. Although from an academic point of view it would seem that digital detectors are the simplest to install and align and that phase-locked loops offer the best weak-signal performance, economic arguments (particularly the cost of retraining service personnel) and personal and corporate feelings about achievable levels of AM rejection and distortion play important roles in today's market. For this reason this chapter will describe a variety of detectors, some of which may seem obsolete to academic readers and others of which may seem unproved to designers in industry.

Discriminators

The basic Foster-Seeley discriminator shown in Fig. 10-6 uses a frequency-sensitive phase shift to demodulate FM. To analyze its operation, consider first the voltage relationships in the coupled circuits $L_1C_1 - L_2C_2$ of Fig. 10-7.

If the impedance coupled into the primary circuit is negligible in comparison with the primary self impedance,¹ then the primary current I_1 will be

$$I_1 = \frac{V_1}{j\omega L_1} \quad (10-7)$$

The voltage induced in the secondary by I_1 is $j\omega MI_1$; at the secondary resonance frequency ω_c the secondary current will be

$$I_2 = \frac{j\omega_c MI_1}{R_2} = \frac{M}{L_1} \frac{V_1}{R_2} \quad (10-8)$$

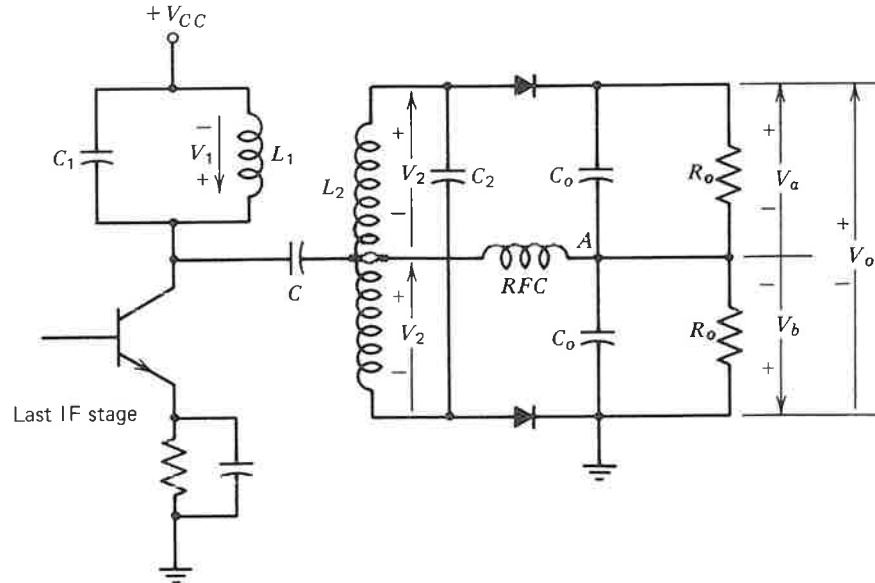
The voltage across the capacitor is

$$2V_2 = j\omega_c C_2 I_2 \quad (10-9)$$

therefore,

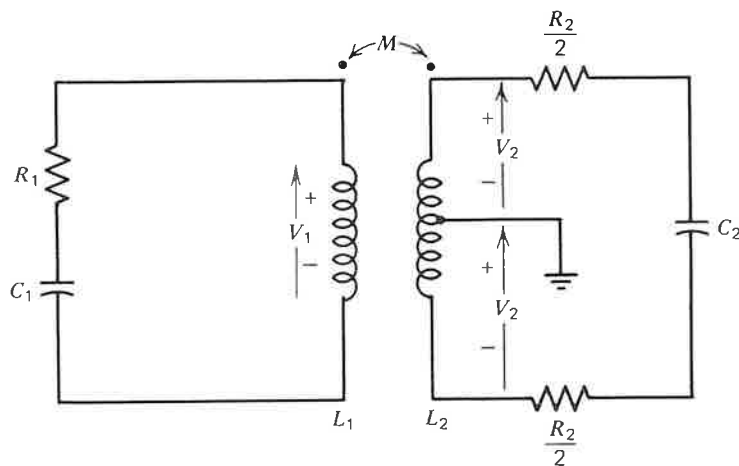
$$V_2 = \frac{j\omega_c C_2 M}{2L_1 R_2} V_1 \quad (10-10)$$

Fig. 10-6 The Foster-Seeley discriminator.



¹This condition will be met in practice; see [6, p. 607].

Fig. 10-7 Circuit used to illustrate phase relationships in the Foster-Seeley discriminator. The resistors shown represent the internal resistance of the coils.



and at resonance V_2 leads V_1 by 90° as in Fig. 10-8a. At frequency ω slightly different from ω_c ,

$$V_2 = \frac{j\omega C_2 M}{2L_1 R_2} \times \frac{V_1}{1 + (1/R_2)[j\omega L_2 + (1/j\omega C_2)]} \approx \frac{j\omega C_2 M V_1}{2L_1 R_2} \times \left\{ 1 - \frac{1}{R_2} \left[j\omega L_2 + \frac{1}{j\omega C_2} \right] \right\} \quad (10-11)$$

and the phase angle between V_1 and V_2 is given by

$$\text{Arg} \left(\frac{V_2}{V_1} \right) = \text{Arg} \left(j + \frac{\omega L_2}{R_2} - \frac{1}{\omega C_2 R_2} \right) = \cot^{-1} \left(\frac{\omega L_2}{R_2} - \frac{1}{\omega C_2 R_2} \right) \quad (10-12)$$

Then if

$$\omega = \omega_c + \Delta\omega \quad (10-13)$$

the phase angle formula reduces to

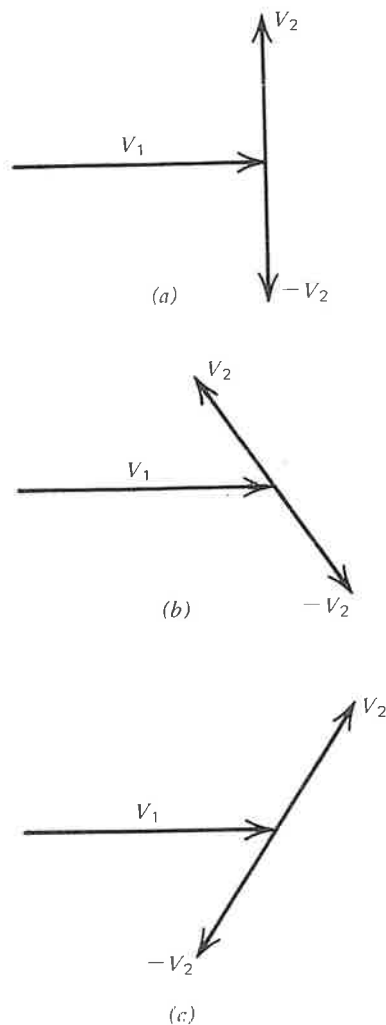
$$\text{Arg} \left(\frac{V_2}{V_1} \right) \approx \cot^{-1} \left(\frac{\Delta\omega L_2}{R_2} \right) = \cot^{-1} \left(\frac{Q\Delta\omega}{\omega_c} \right) \quad (10-14)$$

where Q is the Q of the secondary circuit. For $\Delta\omega > 0$ ($\omega > \omega_c$) the angle is less than 90° , and for $\Delta\omega < 0$ ($\omega < \omega_c$) the phase angle is greater than 90° . These two conditions are illustrated in Figs. 10-8a and b.

To carry this analysis over to the Foster-Seeley discriminator of Fig. 10-6, assume that C_o and C_c are RF short circuits and the RFC is an RF open circuit. These conditions place point A at RF ground, and the RF voltage V'_a applied to the top diode and the $R_o C_o$ network is

$$V'_a = V_1 + V_2 \quad (10-15)$$

Fig. 10-8 Phase relationships in the discriminator transformer. (a) Resonance: $\omega = \omega_c$; (b) below resonance: $\omega < \omega_c$; and (c) above resonance: $\omega > \omega_c$.



Under the same conditions the RF voltage V'_b applied to the bottom diode and filter network is

$$V'_b = V_1 - V_2 \quad (10-16)$$

The phase relationship between V_1 and V_2 is as shown in Fig. 10-8; Fig. 10-9 illustrates the variation with frequency of V'_a and V'_b . The basic relationships are

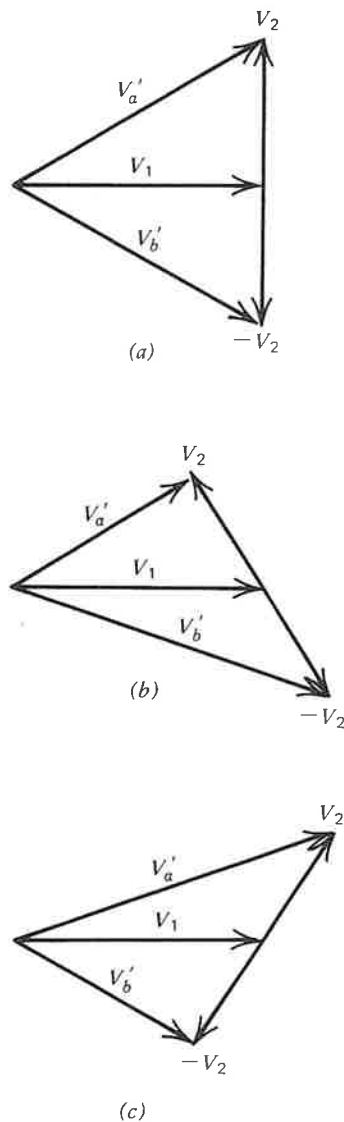
$$V'_a = V'_b \quad \text{when } \omega = \omega_c$$

$$|V'_a| < |V'_b| \quad \text{when } \omega < \omega_c$$

and

$$|V'_a| > |V'_b| \quad \text{when } \omega > \omega_c$$

Fig. 10-9 Phase relationships in the discriminator showing the diode voltages. (a) Resonance $\omega = \omega_c$; (b) below resonance: $\omega < \omega_c$; and (c) above resonance: $\omega > \omega_c$.

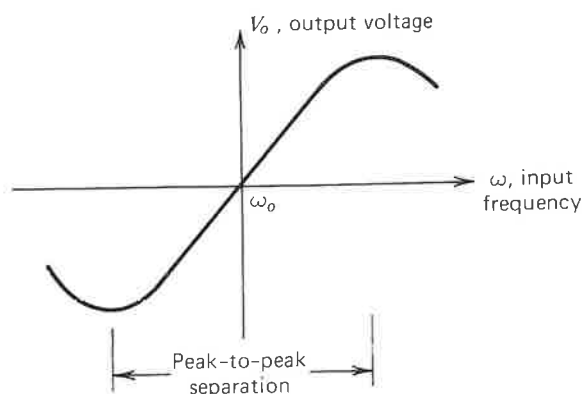


If the diode rectification efficiency is η , then the output voltage V_o is given by

$$V_o = V_a - V_b = \eta |V_a'| - \eta |V_b'| \quad (10-17)$$

A plot of V_o versus frequency is shown in Fig. 10-10. The principal shortcoming of the discriminator is evident from (10-17); any change in the input signal amplitude will be reproduced in V_a' , V_b' , and V_o and any AM on the incoming signal will be demodulated. A good limiter must precede the

Fig. 10-10 Foster-Seeley discriminator output characteristics.



discriminator for satisfactory operation. This requirement has eliminated the Foster-Seeley discriminator from almost all mass production entertainment receiver circuits in favor of the ratio detector, and many integrated circuits that are popularly called discriminators are actually ratio detectors.

The primary problem in discriminator design is the transformer; most designers avoid this by specifying off-the-shelf units. For a detailed discussion of the procedure and further references, see [7, pp. 300–311].

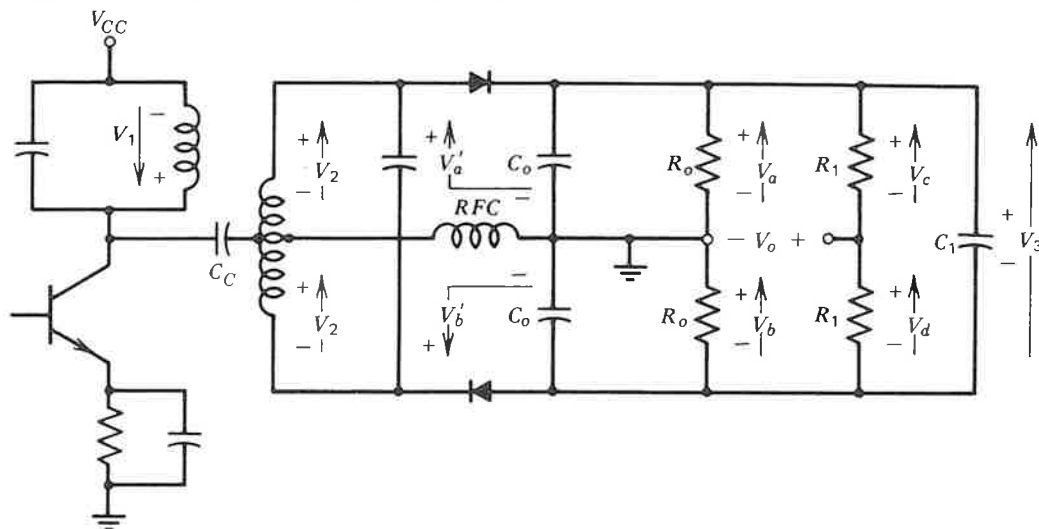
The Ratio Detector

The ratio detector offers better AM rejection² than the discriminator and will operate satisfactorily with poorer limiting in the preceding stages. For many years ratio detectors were standard in almost all FM receivers, and many manufacturers still feel that they offer superior audio performance.

The theory of ratio detector operation will be developed using Fig. 10-11. The ratio detector differs from the discriminator of Fig. 10-6 in three ways: the lower diode is reversed; an extra pair of resistors R_1 and a capacitor C_1 have been added; and the output voltage is taken between different points. From one viewpoint the circuit behaves as a half-wave rectifier with two diodes in series and an RC filter; the filtered output appears as V_3 . If the diode rectification efficiency is η , V_3 is proportional to $\eta|V_2|$, where V_2 is the phasor IF voltage on either side of the secondary center tap. Over the operating

²AM rejection is the ratio in decibels of the FM detector audio output power for a specified FM input to the audio output power for a specified AM input. The characteristics of the FM and AM signals vary with detector application and manufacturer; for example, RCA specifies that in measuring AM rejection in the CA 3013 IC the FM signal shall have an amplitude of 10 mV rms, a carrier frequency of 4.5 MHz, a modulating frequency of 1 kHz, and a ± 25 -kHz frequency deviation. The corresponding AM signal is required to have the same carrier frequency, carrier amplitude, and modulating frequency with 50 percent modulation.

Fig. 10-11 The basic ratio detector.



bandwidth of the detector $|V_2|$ and V_3 are essentially frequency independent. Since V_3 is divided across equal resistances (R_1) symmetric with respect to ground, the voltages V_c and V_d are of equal amplitude and

$$V_c = \frac{V_3}{2} \quad (10-18)$$

$$V_d = \frac{V_3}{2} \quad (10-19)$$

$$V_3 = V_c + V_d \quad (10-20)$$

In terms of V_2 and the phasor primary voltage V_1 , the phasor IF voltage V'_a between the anode of the upper diode and ground is

$$V'_a = V_1 + V_2 \quad (10-21)$$

and the phasor IF voltage V'_b between the cathode of the lower diode and ground is

$$V'_b = V_1 - V_2 \quad (10-22)$$

The corresponding rectified voltages are³

$$V_a = \eta |V_1 + V_2| \quad (10-23)$$

and

$$V_b = \eta |V_1 - V_2| \quad (10-24)$$

³The reader may note that, in Fig. 10-11, V_a and V'_a have the same polarity but V_b and V'_b have opposite polarities. This is because the top and bottom diodes also have opposite polarities.

By elementary circuit analysis

$$V_o = V_d - V_b \quad (10-25)$$

and

$$V_3 = V_a + V_b \quad (10-26)$$

From (10-19)

$$V_o = \frac{V_3}{2} - V_b \quad (10-27)$$

or

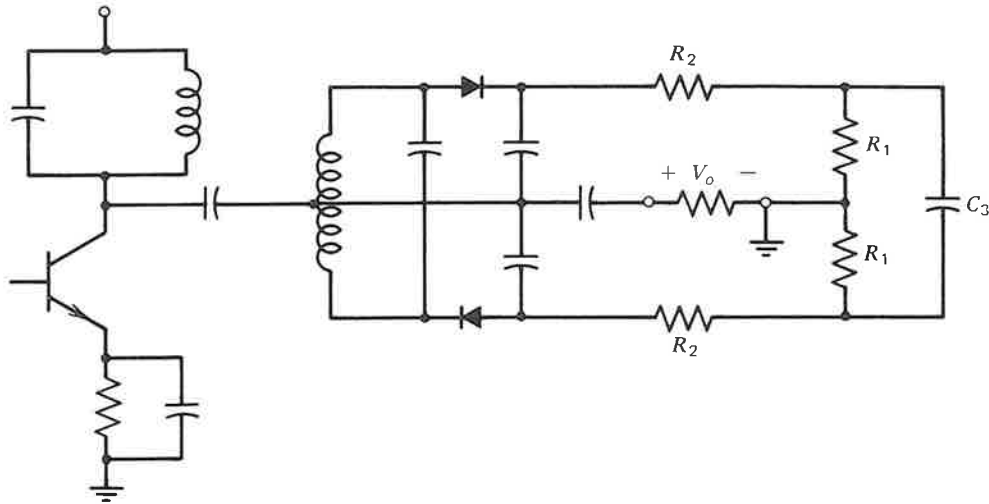
$$V_o = \frac{V_a}{2} + \frac{V_b}{2} - V_b = \frac{V_a - V_b}{2} \quad (10-28)$$

For the same circuit parameters the ratio detector is thus 6 dB less sensitive than the discriminator. The difficulties in obtaining optimum components (particularly transformers) for these circuits tend to reduce this difference, and in practice the sensitivity of the two circuits is about the same.

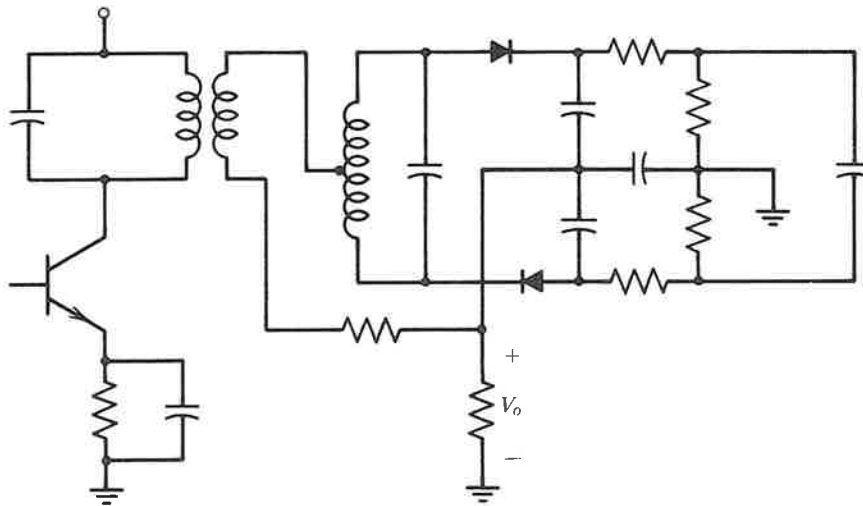
When the ratio detector is in steady-state operation, the diodes conduct during a small portion of each IF cycle to recharge the capacitors (i.e., to replace the charge supplied by the capacitors to the resistors and to the load). For the diodes to conduct, the instantaneous IF voltages V'_a and V'_b must be larger than the instantaneous values of V_a and V_b . If the input signal level drops, the peak values of V'_a and V'_b also drop and the diode conduction period is shortened. This reduces the secondary circuit dissipation and raises the secondary circuit Q because the diodes now connect the secondary windings to the resistors for a smaller fraction of each cycle. By (10-14) the phase angle between the primary and secondary voltages is a function of Q ; at a given frequency deviation $\Delta\omega$, an increase in Q makes $\text{Arg}(V_2/V_1)$ smaller and increases the ratio detector output voltage in the same way as an increase in $\Delta\omega$. This compensates for the decrease in output brought about by the decrease in IF amplitude that started the discussion. Similarly, an increase in input signal level would tend to raise the output voltage while at the same time causing a decrease in Q and a corresponding decrease in the detector output. The compensating effects of signal amplitude and circuit Q on the output voltage can be balanced by the insertion of additional resistors R_2 as in Fig. 10-11a and the ratio discriminator made relatively insensitive to amplitude variations whose period is short in comparison to the discriminator RC time constant. For details on this and other design considerations in maximizing ratio detector AM rejection, see [1, pp. 392-401].

The ratio detector can be modified in a number of ways to reduce costs or to suit particular applications. For example, the R_o resistors in Fig. 10-11 are unnecessary (R_1 and the load provide a suitable dc path) and, if the ground point is shifted as in Fig. 10-12a, the choke may be eliminated. Another variation (Fig. 10-12b) adds a third winding to the transformer and eliminates

Fig. 10-12 Ratio detector modifications.



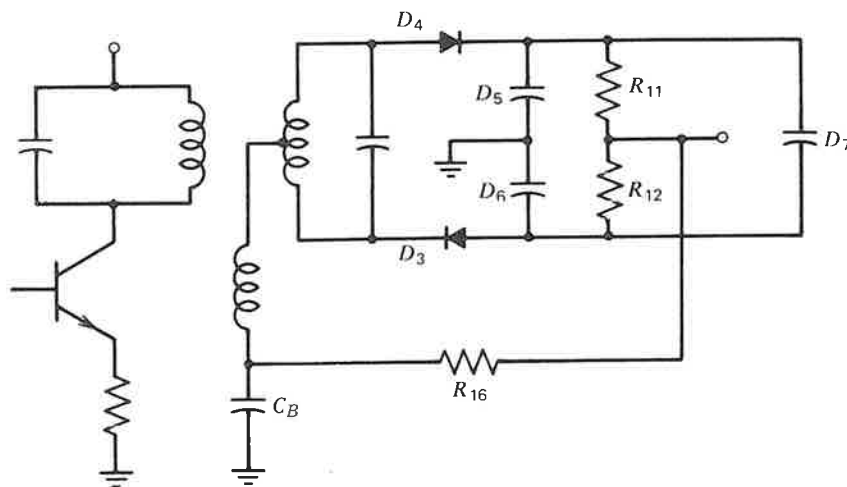
(a)



(b)

both the choke and the coupling capacitor. This technique is popular with IC users because almost all of the external components can be housed inside the transformer can. The third winding is nonresonant and tightly coupled to the primary so that its voltage is proportional to, and in phase with, the primary voltage [7]. It thus serves the same purpose as the coupling capacitor in Fig. 10-11: to connect the primary voltage in series with the secondary. The resistor in series with the extra winding improves AM rejection. Fig. 10-13 illustrates the ratio detector modification used in the RCA CA 3013 IC.

Fig. 10-13 Ratio detector used in RCA 3013 and 3014 ICs. All labeled components except C_B are on the IC chip itself; the capacitors with D labels are realized with diodes. (From "RCA Linear Integrated Circuits," SSD-201C, 1975. Courtesy of RCA Solid State Division.)



Coincidence or Quadrature Detectors

A quadrature detector is a circuit that (1) splits the IF signal into two parts, (2) passes one part through a network with a phase shift of 90° plus some constant times the IF deviation from center frequency, (3) multiplies the shifted and unshifted components together, and (4) selects the audio frequency portion of the multiplier output spectrum. Mathematically, this operation is easy to describe. For the IF center radian frequency ω_o and instantaneous radian frequency ω , let the phase shift $\Delta\theta$ in radians be given by

$$\Delta\theta = \frac{\pi}{2} - K(\omega - \omega_o) = \frac{\pi}{2} - K\Delta\omega \quad (10-29)$$

The inputs to the multiplier are $V_o \sin(\omega t)$ and

$$V_o \sin\left(\omega t + \frac{\pi}{2} - K\Delta\omega\right) = V_o \cos(\omega t - K\Delta\omega) \quad (10-30)$$

Multiplied together, these produce $V_o^2 \sin(\omega t) \cos(\omega t - K\Delta\omega)$, which has a low frequency component given by

$$\frac{-V_o^2}{2} \sin(k\Delta\omega)$$

For $K\Delta\omega < 0.25$ rad so that

$$\sin(K\Delta\omega) \approx K\Delta\omega \quad (10-31)$$

this term is linear in $\Delta\omega$ and reproduces the modulating waveform.

Figure 10-14 illustrates a quadrature detector broken down into functional blocks. The parallel RLC circuit resonates at ω_o and the phase shift $\Delta\theta$ between V_o and V_{in} is given by

$$\Delta\theta = \text{Arg} \frac{V_o}{V_{in}} \approx \frac{\pi}{2} - \tan^{-1}(Q_o\delta) \quad (10-32)$$

where Q_o is the resonant-circuit Q at ω_o and

$$\delta = \frac{\omega}{\omega_o} - \frac{\omega_o}{\omega} \quad (10-33)$$

For small frequency shifts, $\Delta\theta$ is sufficiently linear with frequency for the circuit to produce acceptable quality audio.

Analog multiplication and good IF limiting were difficult to achieve in vacuum tube circuits, and special purpose multigrid tubes like the 6DT6 were developed for service as quadrature detectors in TV receivers [1]. Here the lower quality required in TV audio as compared with FM broadcasting, the lack of a tuned transformer, and the relatively high audio output of the quadrature detector gave this circuit an economic advantage over the other FM detectors available at the time.

The TV quadrature detector depended upon special vacuum tubes that have no single-transistor counterpart. Solid-state quadrature detectors evolved with the analog multiplier replaced by a digital AND gate. This is a circuit, illustrated schematically in Fig. 10-15, that produces an output (+5 volts is a standard value) only if both of its inputs are positive. Since an AND gate may properly be called a coincidence circuit (because it produces an output when there are coincident signals at both inputs), the solid-state quadrature detector is frequently called a *coincidence detector*.

Fig. 10-14 Symbolic quadrature detector.

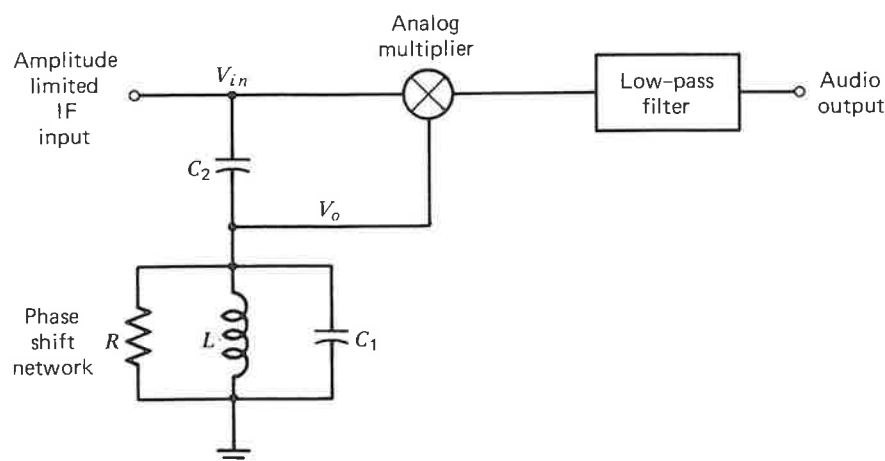
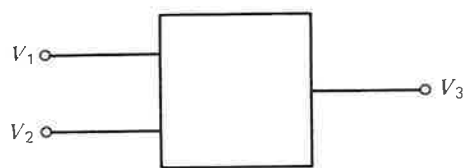


Fig. 10-15 Idealized AND gate.



Input-Output Relationships

V_1	V_2	V_3
> 0	> 0	+5
< 0	> 0	0
> 0	< 0	0
< 0	< 0	0

In the coincidence detector, the IF signal is hard limited to a square wave. The square wave is split; part of it goes directly to one input of the AND gate and the other goes through the phase shifter of Fig. 10-14 and then to the other AND input. Since the phase shifter contains a high- Q resonant circuit, the phase-shifted component will be essentially a sine wave at the IF center frequency. When the sine wave and the square wave are both positive, the AND gate output will be +5 volts. When one or both are negative, the output will be 0 volts. The circuit is sketched in Fig. 10-16.

The output of the AND gate is a train of pulses whose length is proportional to the phase difference between the sine and square waves. When the two are in phase, the pulse length is half the IF period, and when they are 180° out of phase the pulse length is zero (no output). Representative waveforms appear in Fig. 10-17. If the AND gate feeds an RC integrator, the integrator output voltage is proportional to the average value of the input pulse train and hence to the frequency deviation of the input signal. Thus integrator output reproduces the modulating waveform.

If the phase shift without modulation is not $\pm 90^\circ$, the coincidence detector is not truly a quadrature detector, but it will still demodulate FM. At $\pm 90^\circ$ phase shift the sine and square waves overlap for one-quarter of a period; at 0° they overlap for a half period and at 180° they do not overlap at all. At 180°

Fig. 10-16 Coincidence detector using an AND gate.

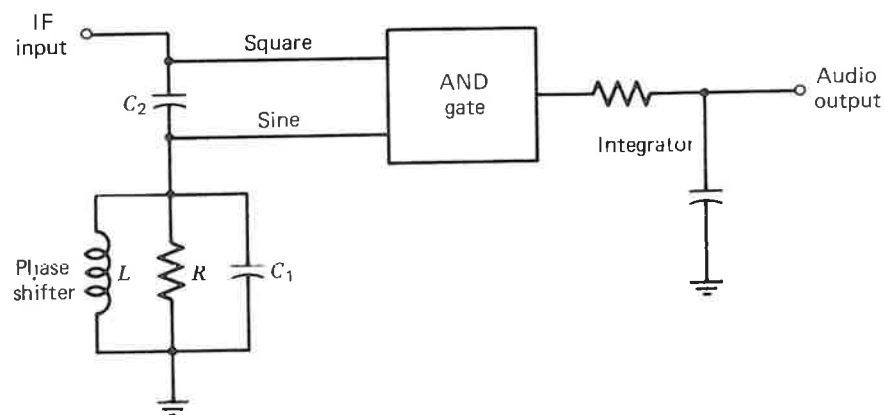


Fig. 10-17 Waveforms in the coincidence detector. (a) Output for a 45° phase shift ($\omega > \omega_o$); (b) output for a 90° phase shift (no modulation); and (c) output for a 135° phase shift ($\omega < \omega_o$).

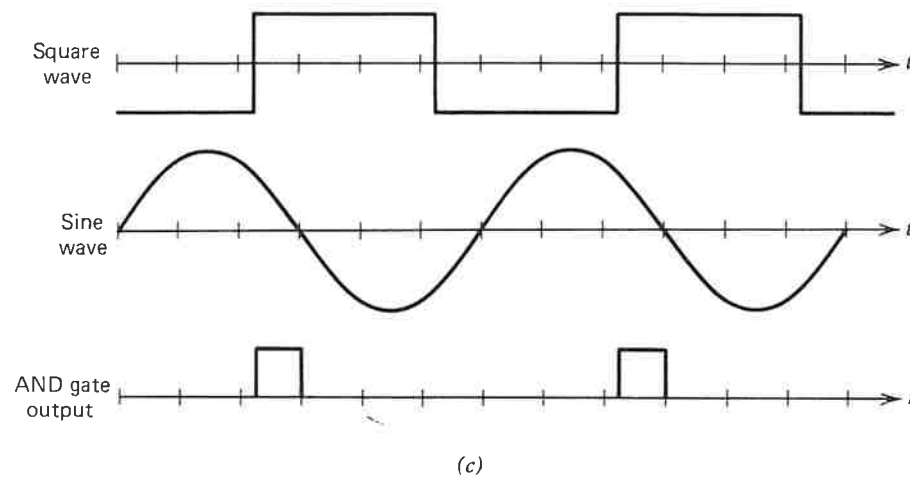
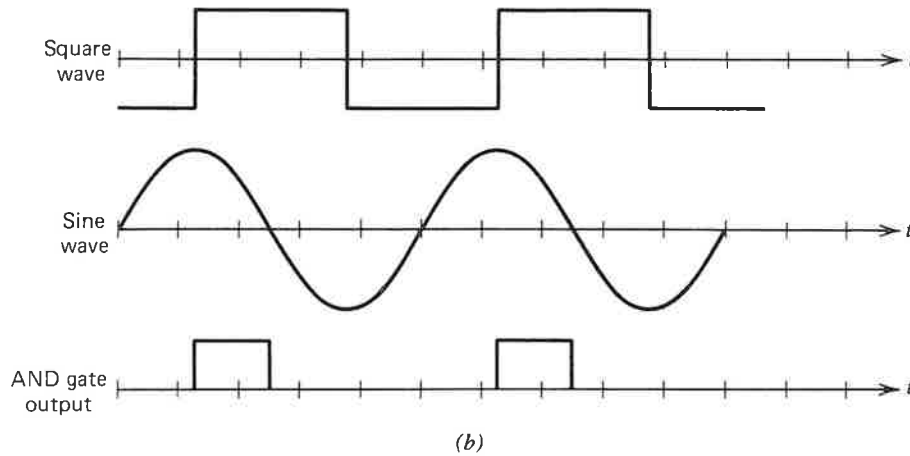
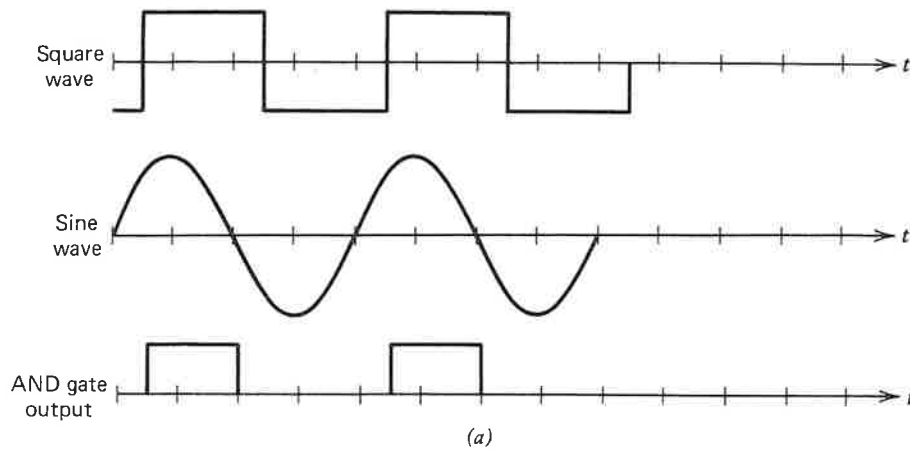
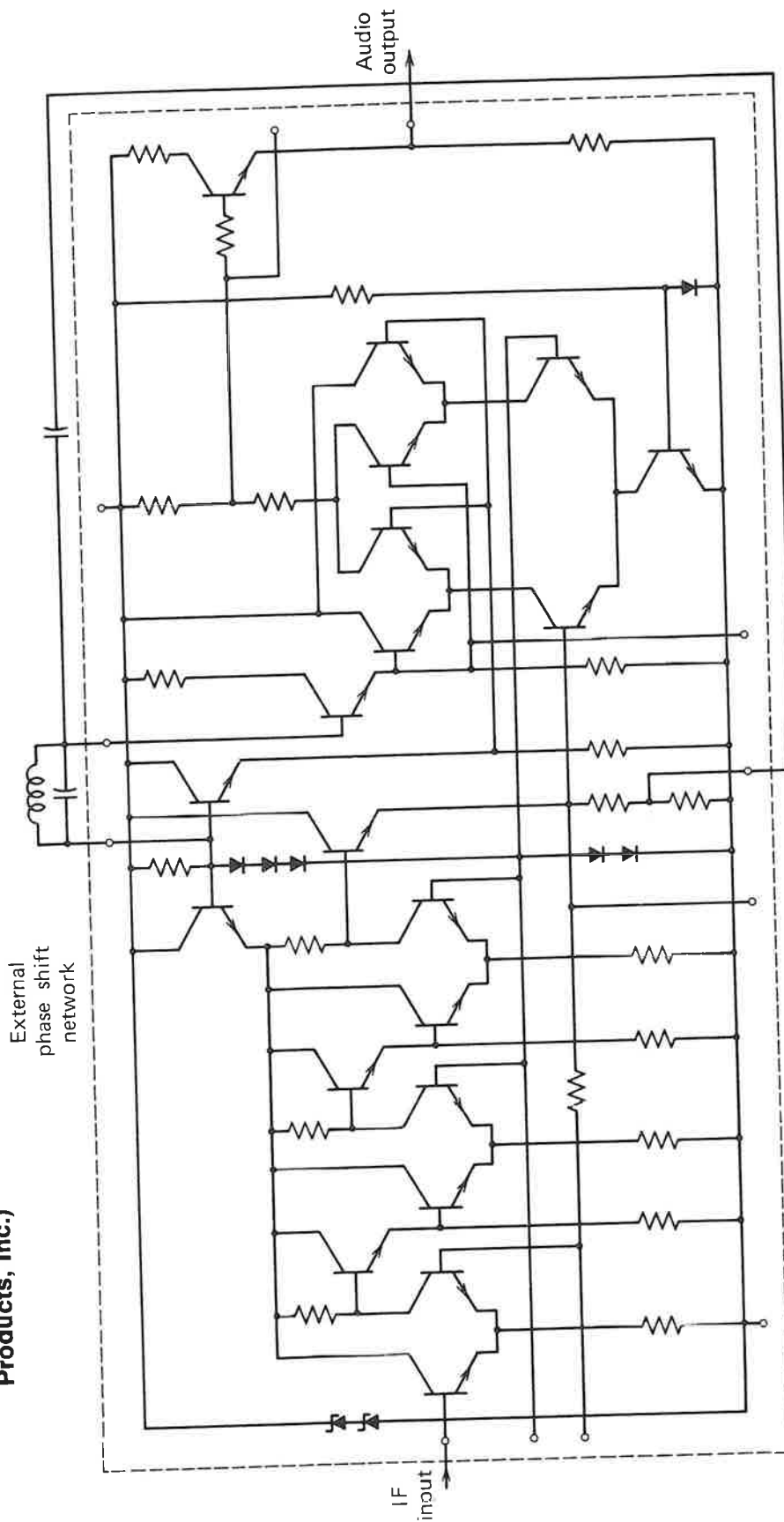


Fig. 10-18 Motorola MC1357 quadrature detector. (From "Data Sheet 9164 R1," 1970. Courtesy of Motorola Semiconductor Products, Inc.)



the AND gate has no output and at 0° it delivers its maximum output of 5 volts for half of an IF period. If the no-modulation phase shift is set at $\pm 90^\circ$, the detector can recover modulation corresponding to sine wave phase changes of $\pm 90^\circ$; at any other reference setting the operating range is reduced.

A reduced operating range is not a serious drawback in most applications, and a popular IC realization of the quadrature detector [8] shown in Fig. 10-18 is designed to operate with a 45° phase shift under conditions of no modulation. In the notation of Figs. 10-14 and 10-16, this means that at the IF center frequency ω_o , $R = 1/\omega_o C_2$. At instantaneous frequency ω the phase shift $\Delta\theta$ of (10-32) is given by

$$\Delta\theta = \tan^{-1} \left\{ \frac{\omega_o/\omega}{1 + Q_o(\omega_o/\omega)[(\omega/\omega_o) - (\omega_o/\omega)]} \right\} \quad (10-34)$$

For unambiguous detection $\Delta\theta$ must be in the first quadrant. This requires

$$\omega \geq \frac{\omega_o}{\sqrt{1 + 1/Q_o}}$$

and permits a maximum symmetric frequency swing with a modulation of

$$\pm \left(\omega_o - \frac{\omega_o}{\sqrt{1 + 1/Q_o}} \right)$$

around ω_o . See Problem 10-3.2.

Zero-Crossing Detectors

The basic problem in FM detection after a hard-limiting IF section is determining the instantaneous frequency of a frequency-modulated square wave. In principle this can be done by measuring the time interval τ between successive zero-crossings and doubling its reciprocal $1/\tau$ in a digital processor. Doing this digitally provides a frequency resolution far beyond that required in communications FM receivers; since it is also expensive, a combination of digital and analog techniques is used to convert the time interval between zeros to an analog voltage.

One of the earliest commercial applications of the zero-crossing detector was marketed in 1973 in the Heath Company's AJ-1510 digital FM tuner [9]. The circuit is sketched in Fig. 10-19 and the waveforms appear in Fig. 10-20.

The IF signal is applied to the input of an IC containing a trigger and a monostable multivibrator. The trigger fires the multivibrator whenever it senses a positive zero crossing in the input signal. When fired, the multivibrator puts a voltage (nominally +5 volts) on one of its outputs (called the Q output) and holds it for a fixed time interval equal to one-half cycle at the IF

Fig. 10-19 A zero-crossing detector.

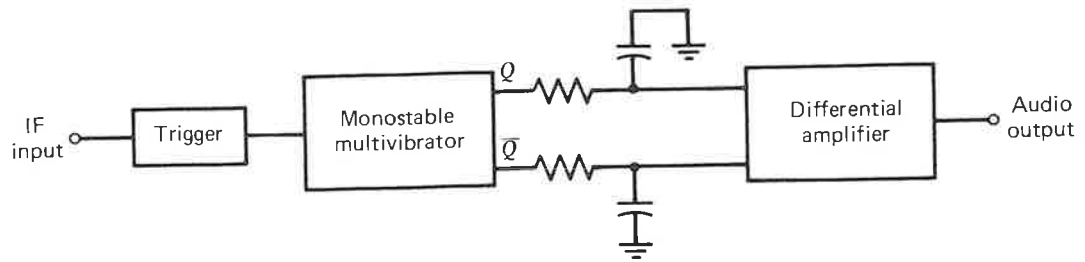
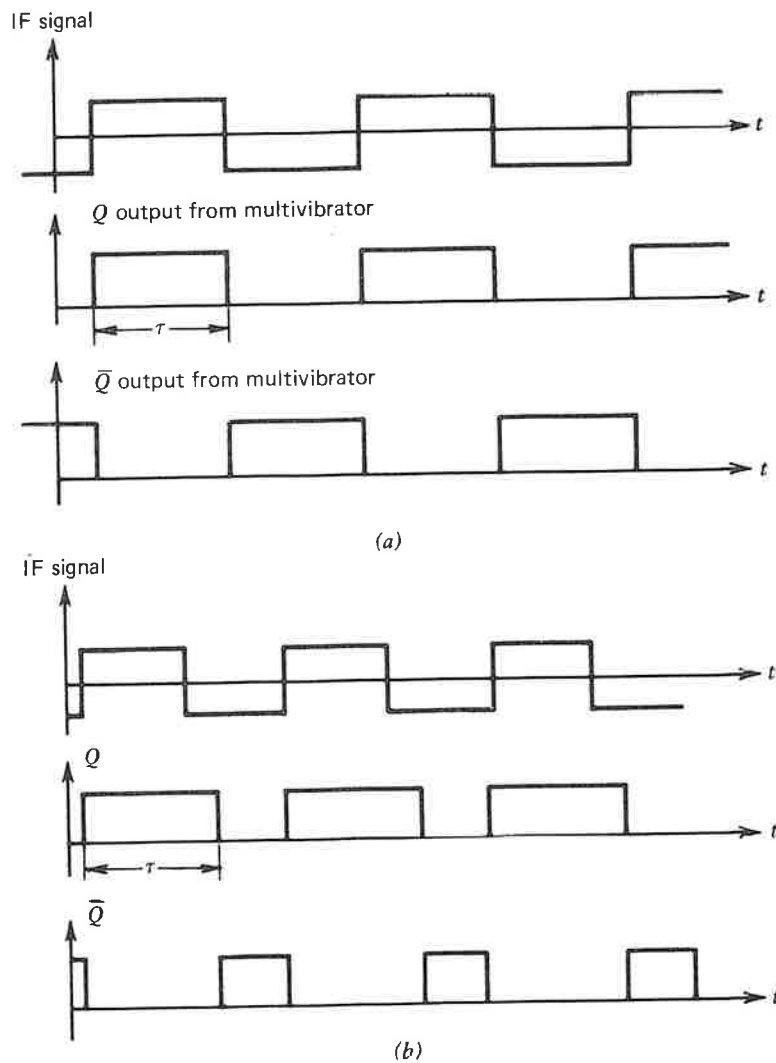


Fig. 10-20 Waveforms in the zero-crossing detector. (a) Signal frequency equal to IF center frequency; and (b) signal frequency above IF center frequency.



center frequency. When Q is at 5 volts, the other output $\bar{Q}$ (pronounced “not Q ”) is at 0 volts; when Q is at 0 volts, $\bar{Q}$ is at 5 volts. When an incoming unmodulated signal is tuned precisely to the IF center frequency, the Q and $\bar{Q}$ outputs are mirror images of each other, as shown in Fig. 10-20a.

The Q and $\bar{Q}$ outputs each go to RC integrators and then to the input terminals of a differential operational amplifier. The op-amp output is a constant times the difference between the average value of Q and $\bar{Q}$. If the signal frequency increases, Q pulses are initiated more frequently. The average voltage at Q increases and the average voltage at $\bar{Q}$ decreases, swinging the op-amp output positive. Similarly, a drop in signal frequency will make the op-amp output negative. The result is a linear transfer function relating op-amp output to IF frequency deviation.

Phase-Locked Loop Detectors

As described in Chapter 6 the phase-locked loop (PLL) is a feedback system that develops an error voltage proportional to the phase difference between an input signal and a signal produced by a voltage-controlled oscillator (VCO). The error voltage is filtered and applied to the VCO so as to make the VCO operate at the signal frequency. If the signal frequency increases, the error voltage will increase and raise the VCO frequency; the opposite will happen if the signal frequency decreases. The VCO control voltage thus follows the frequency modulation of the input signal and if the control voltage is accessible outside of the PLL chip, the PLL will serve as an FM detector.

Although phase-locked loops offer a low SNR improvement threshold and are a popular subject for mathematical analyses, very few receiver manufacturers have adopted them as FM detectors. Apparently, PLLs that are commercially available at the present time have relatively poor AM rejection and relatively high distortion. In addition, they require about the same number of external resistors and capacitors as other IC detectors and have no inherent cost advantage.

Detectors for Digital FM

The detectors described in the previous paragraphs are suitable for recovering the analog frequency or phase modulation used in voice communications systems and in commercial broadcasting. However, modern satellite and terrestrial microwave communications systems employ pulse code modulation and digital PM almost exclusively. Systems in current use generally employ either differentially coherent phase shift keying (DCPSK) or quadrature phase shift keying (QPSK), and the detectors for these angle-modulation schemes are quite unlike those described here.

10-4 FM Stereo Reception

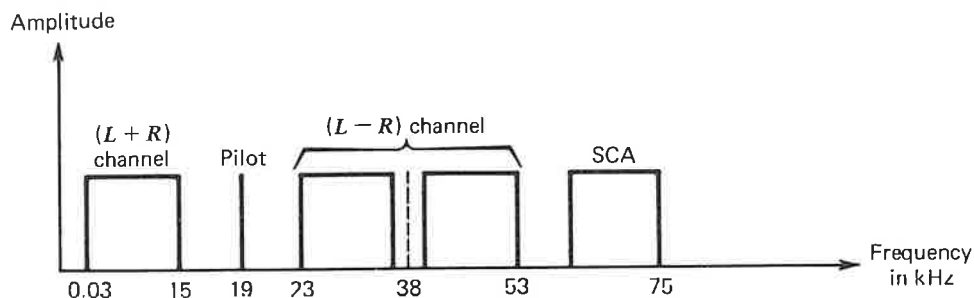
In stereophonic sound reproduction, the sounds captured by two independent microphones are replayed through two sets of speakers. This requires two independent audio channels. With reference to the microphone placement as seen by a listener facing the sound source, these are called the left channel (L) and the right channel (R).

After stereophonic disk recording became popular in the mid-1950s, there was a movement to extend this technique to FM broadcasting. Some stations holding dual AM and FM licenses conducted experiments in broadcasting one channel on their FM transmitter and the other on their AM transmitter, but obviously this approach was unsatisfactory. It was soon replaced by the present technique of broadcasting both channels simultaneously with a single FM transmitter.

In stereo broadcasting, the L and R signals are not themselves transmitted. Instead they are combined to form a sum ($L + R$) and a difference ($L - R$) channel. The sum channel is transmitted directly. The difference channel modulates a 38-kHz subcarrier, producing a DSB/SC signal. A 19-kHz "pilot" signal phase coherent with the 38-kHz subcarrier is transmitted to synchronize the subcarrier oscillator in the receiver with the subcarrier oscillator at the transmitter. The complete spectrum is sketched in Fig. 10-21.⁴ The composite waveform modulates an FM transmitter in the usual manner.

When stereo broadcasting was introduced, the FCC required that existing monaural receivers be capable of receiving stereo broadcasts as mono without

Fig. 10-21 Spectrum of the modulating waveform used in FM stereo broadcasting.



⁴The part of the spectrum labeled SCA is the *Subscription Carrier Authorization* band. It is used by some stations to transmit background music by subscription (i.e. for a monthly fee) to businesses.

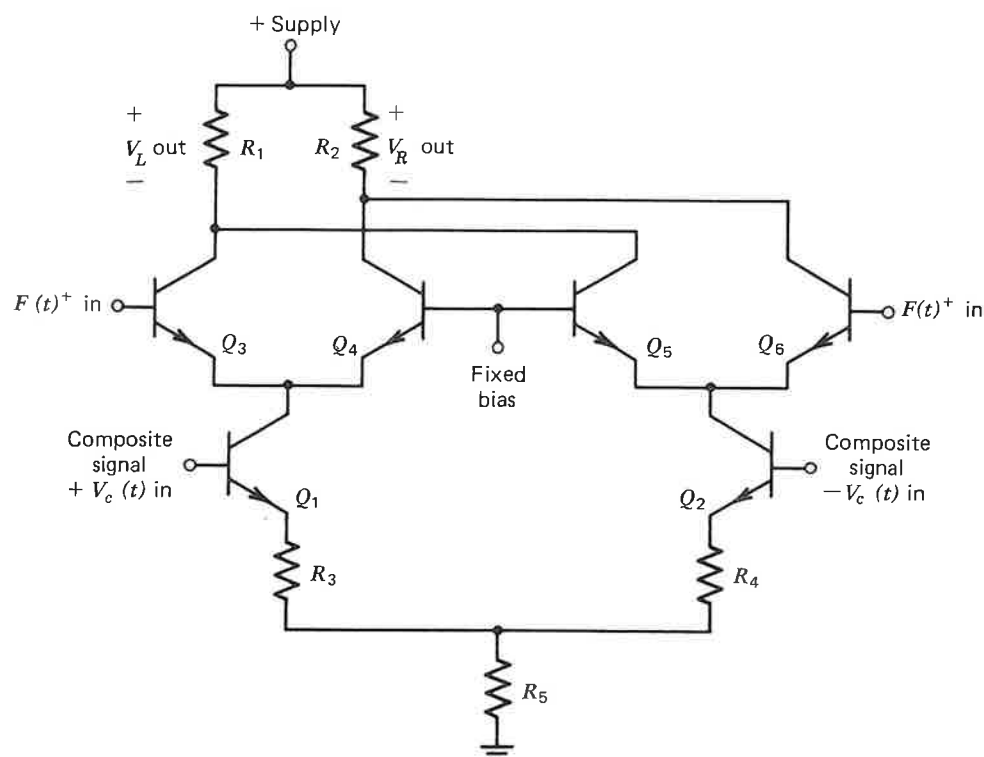
modification. This is the reason that the $L + R$ and $L - R$ signals are transmitted instead of L and R . The $L + R$ signal is identical to that broadcast by a monaural transmitter and this is all that a monaural receiver detects.

The pilot is transmitted rather than the subcarrier because 19 kHz falls into an empty portion of the composite spectrum. If the 38-kHz carrier were transmitted, it would have to be separated from the $L - R$ sidebands, which are only ± 30 Hz away. This would require expensive filtering.

Conceptually, an FM receiver demodulates the stereo signal and recovers the spectrum of Fig. 10-21. Since modulating frequencies above 15 kHz are not transmitted in ordinary monophonic broadcasting, the audio amplifier stages of a monaural receiver simply eliminate the pilot carrier and the receiver behaves as if it were receiving a monaural broadcast. On the other hand, a stereo receiver coherently demodulates the $L - R$ and $L + R$ signals and, by adding and subtracting the recovered $L + R$ and $L - R$ information, reproduces the original left and right channels.

A common circuit used to decode FM stereo broadcasting is shown in Fig. 10-22 [10]. It is similar to the phase detector of Section 6-5, and much of that discussion applies here. The composite signal is applied to the base of Q_1 . The Q_1 collector current is derived from the emitters of Q_3 and Q_4 . A 38-kHz

Fig. 10-22 FM stereo decoding circuit.



square-wave phase-locked to the 38-kHz suppressed carrier switches Q_3 and Q_4 on and off in turn. This will be called the switching waveform. Initially, ignore Q_2 , Q_5 , and Q_6 . The voltage V_L is then proportional to the product of the composite signal and the switching waveform $F(t)^+$, Fig. 10-23a, that drives Q_3 . Since Q_4 is off when Q_3 is on, and vice versa, the voltage V_R is proportional to the product of the composite waveform and a switching waveform $F(t)^-$, shown in Fig. 10-23b.

The switching waveforms are related by

$$F(t)^- = 1 - F(t)^+ \quad (10-35)$$

The function $F(t)^+$ is identical to the waveforms of Problem 7-3.1 with $\tau = T/2$, and its Fourier series representation is given by

$$F(t)^+ = \frac{1}{2} + \frac{2}{\pi} \cos \omega_o t - \frac{2}{3\pi} \cos 3\omega_o t - \dots \quad (10-36)$$

Here

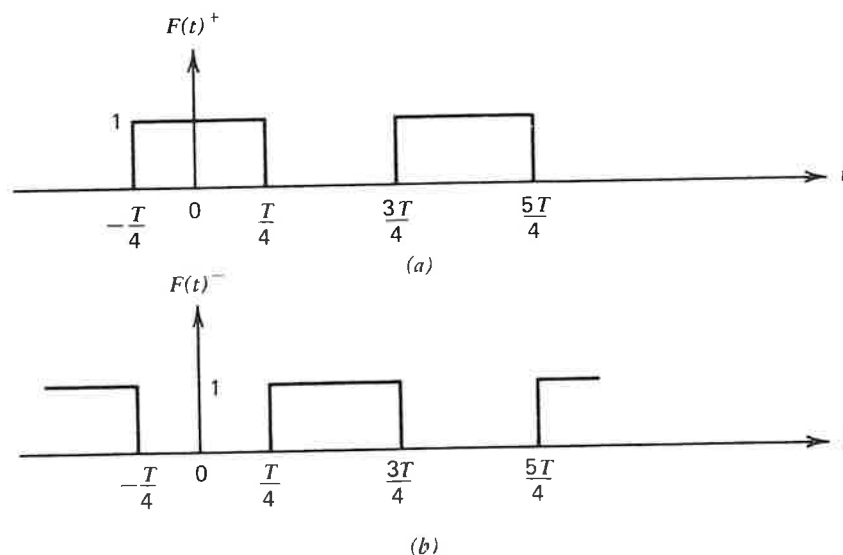
$$\omega_o = \frac{2\pi}{T} \quad (10-37)$$

is the radian switching frequency. Similarly, $F(t)^-$ can be expressed as

$$F(t)^- = \frac{1}{2} - \frac{2}{\pi} \cos \omega_o t + \frac{2}{3\pi} \cos 3\omega_o t - \dots \quad (10-38)$$

At any instant, the composite signal consists of an audio frequency component $(L + R)$, a pilot carrier $P_o \cos \omega_o t/2$, and the DSB/SC difference

Fig. 10-23 Switching waveforms used in the FM decoding circuit. (a) The $F(t)^+$ waveform; (b) the $F(t)^-$ waveform.



signal $(L - R) \cos \omega_o t$. The total composite spectrum can be written

$$V_c = (L + R) + P_o \cos \frac{\omega_o t}{2} + (L - R) \cos (\omega_o t) \quad (10-39)$$

Multiplying V_c with $F(t)^+$ and $F(t)^-$ to obtain V_L and V_R yields

$$V_L = K \left\{ \frac{L + R}{2} + \frac{1}{\pi} (L - R) + \text{higher-frequency terms} \right\} \quad (10-40)$$

where K is a constant of proportionality. Similarly,

$$V_R = K \left\{ \frac{L + R}{2} - \frac{1}{\pi} (L - R) + \text{higher-frequency terms} \right\} \quad (10-41)$$

The higher-frequency terms in (10-40) and (10-41) can be eliminated by a high-pass filter. We assume for convenience that K is unity; then combining terms yields

$$V_L = L \left(\frac{1}{2} + \frac{1}{\pi} \right) + R \left(\frac{1}{2} - \frac{1}{\pi} \right) = 0.82L + 0.18R \quad (10-42)$$

$$V_R = R \left(\frac{1}{2} + \frac{1}{\pi} \right) + L \left(\frac{1}{2} - \frac{1}{\pi} \right) = 0.82R + 0.18L \quad (10-43)$$

Hence V_L is approximately L and V_R is approximately R . The channel separation S for this arrangement is 13 dB, where S is given by

$$S = 20 \log_{10} \left| \frac{L \text{ component of } V_L}{R \text{ component of } V_L} \right| = 20 \log_{10} \left| \frac{R \text{ component of } V_R}{L \text{ component of } V_R} \right| \quad (10-44)$$

The channel separation can be increased to about 45 dB through small correction voltages supplied by Q_2 , Q_5 and Q_6 . The collector current of Q_2 is proportional to the negative of the composite signal: $-V_c(t)$. Since transistor Q_6 is driven by $F(t)^+$, its contribution to V_R is proportional to $-V_c(t)F(t)^+$. Call this contribution V_{R1} ; its low-frequency components are

$$V_{R1} = -K_1 \{0.82L + 0.18R\} \quad (10-45)$$

where K_1 is a constant of proportionality. The total value of V_R is then

$$V_R = (0.82 - 0.18K_1)R + (0.18 - 0.82K_1)L \quad (10-46)$$

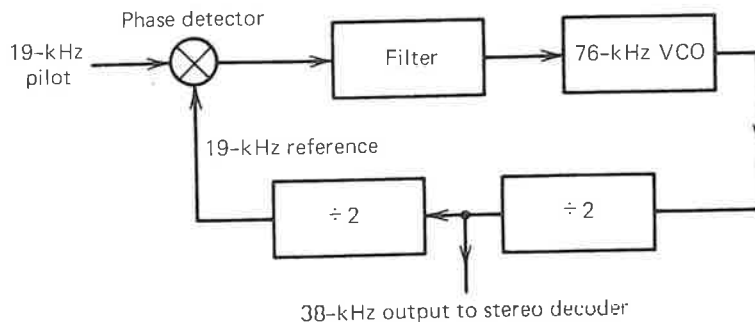
Similarly, for the left channel the total value of V_L is

$$V_L = (0.18 - 0.82K_1)R + (0.82 - 0.18K_1)L \quad (10-47)$$

If the circuit is adjusted so that $K_1 = 0.18/0.82 = 0.220$, theoretically infinite channel separation can be obtained. In practice 45 dB is adequate.

Thus far, nothing has been said concerning generation of the phase-

Fig. 10-24 Scheme for generating a coherent 38-kHz switching waveform.



coherent 38-kHz switching signal. This is done by a PLL system mounted on the same chip as the stereo decoder; the circuit [11] is sketched in Fig. 10-24. In it the filtered output of a 19-kHz phase detector drives a 76-kHz VCO. The VCO output is divided by 2 to provide a coherent 38-kHz signal for the stereo decoder and divided by 2 again to provide a 19-kHz comparison signal for the phase detector. The VCO is an *RC* relaxation oscillator and does not need to produce a symmetric waveform output. Symmetry is supplied by the divide-by-2 circuit.

10-5 A Note on Quadraphonic Sound

At the present time there is some interest on the part of audio enthusiasts in four-channel sound recording and reproduction. Called "quad," this system employs four independent microphone channels and a corresponding set of four speaker systems for reproduction. Quad is easiest to implement with four-channel tape equipment, but at least two mutually incompatible quad disk recording schemes are available. The situation with regard to broadcasting is confused; several systems have been proposed, but these are also mutually incompatible and have not been approved by the Federal Communications Commission for normal broadcasting use. For this reason, quadraphonic detectors and decoders will not be discussed here. The interested reader should consult the *Journal of the Audio Engineering Society* for detailed articles on the subject.

10-6 Audio Considerations: Preemphasis and Deemphasis

In deriving the characteristics of a generalized FM detector, it was pointed out that noise in the part of the receiver IF passband farthest away from the carrier is demodulated with greater audio output than noise in the part nearest

the carrier. This means that the noise level in the higher-frequency part of the demodulated audio spectrum will be greater than the noise level at the lower frequencies. To a listener the signal-to-noise ratio will seem higher for a low-frequency modulating signal than for a high-frequency modulating signal. If the modulating signal were processed before modulation by a filter whose transfer function was linearly proportional to frequency, the demodulated signal and the demodulated noise would have the same relative amplitudes at any modulating frequency. But this would introduce a further complication: distortion in the detected audio. This distortion could be removed by post-detection filtering with a filter whose transfer function varied inversely with frequency.

Such a procedure is followed in FM broadcasting. The filtering process is called preemphasis at the transmitter and deemphasis at the receiver. By FCC

Fig. 10-25 Preemphasis and deemphasis circuits and transfer functions. (a) Preemphasis; and (b) deemphasis. (From T. McMullen, *FM and Repeaters for the Radio Amateur*, The American Radio Relay League, Inc., © 1972.)

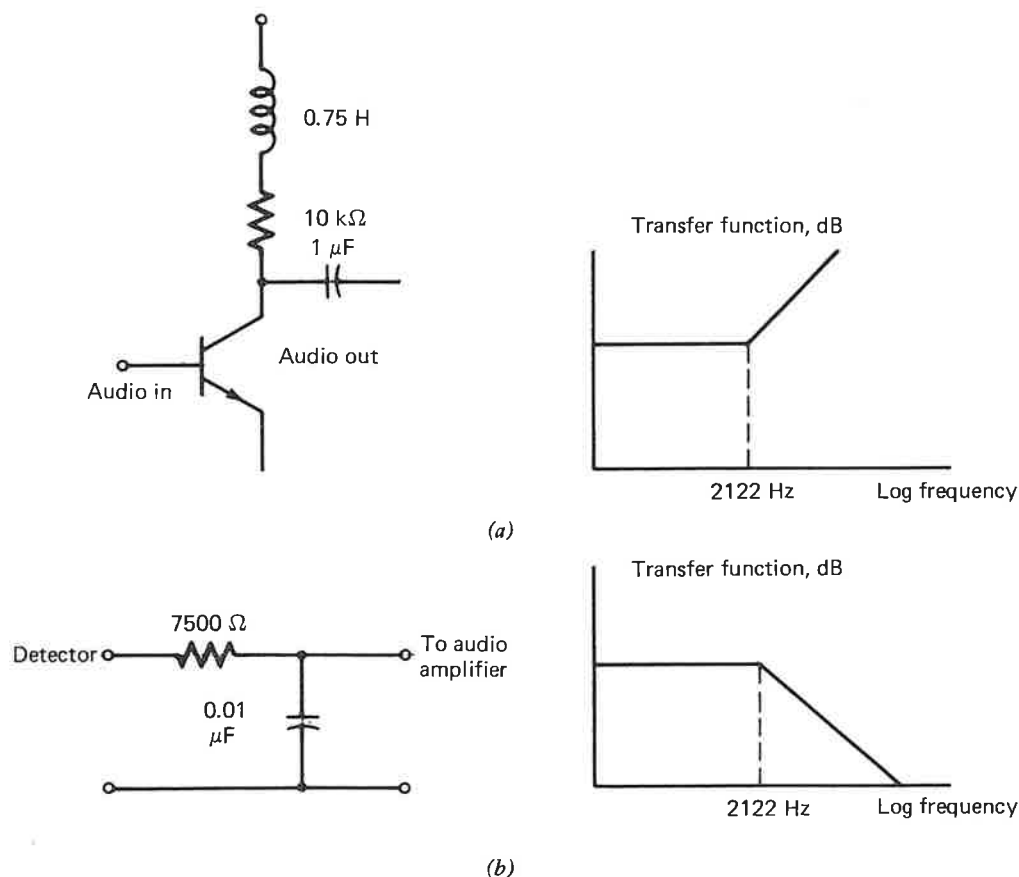
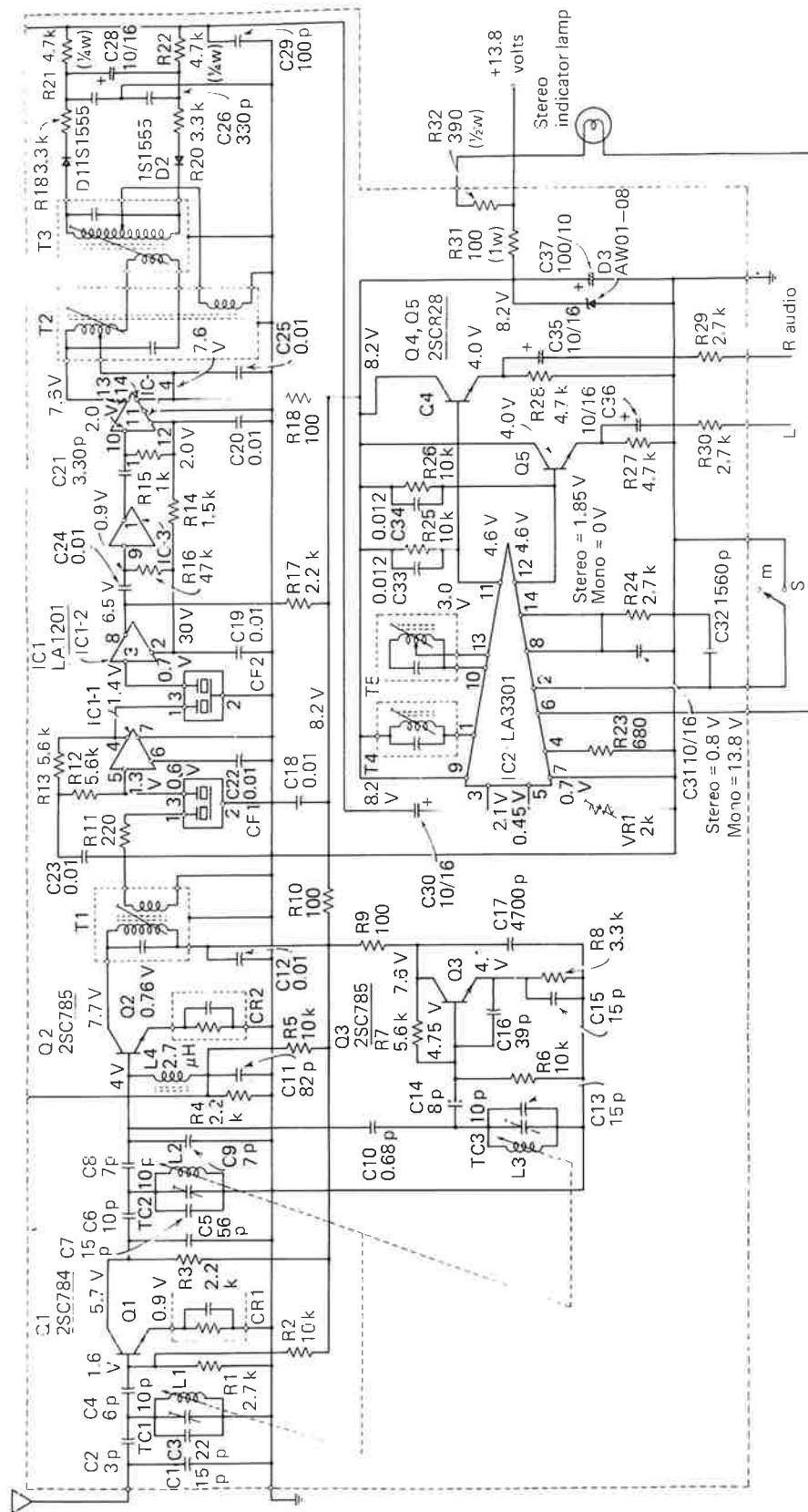


Fig. 10-26 A representative FM receiver. (From "Pioneer KP-250 Schematic." Courtesy of Pioneer Electronics of America).



Regulations the preemphasis filter for broadcast applications must be an RL network with a time constant of 75 microseconds and the deemphasis filter is usually an RC network with the same time constant. Typical preemphasis and deemphasis circuits are shown in Fig. 10-25 [12]. Electrically, the preemphasis network is a differentiator and the deemphasis network is an integrator. Therefore, the portion of the audio signal above 2122 Hz is differentiated before transmission and integrated after reception.

10-7 An Example of a Complete FM Receiver

The complete design of an FM receiver is a complicated process involving much of the material presented in all the chapters of this text. For obvious reasons it is not possible to go through such a design here. To illustrate a typical result, Fig. 10-26 reproduces the FM tuner portion of the Pioneer Electronic Corporation Model KP-250 automobile FM receiver and cassette tape player.

The KP-250 is typical of many commercial receivers in that it incorporates a blend of technologies to achieve a desired level of performance. The FM detector, for example, is a variation of the ratio detector realized with discrete components. On the other hand, the IF section employs IC amplifiers and ceramic filters.

REFERENCES

1. A. B. Cook and A. A. Liff, *Frequency Modulation Receivers*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1968.
2. K. K. Clarke and D. T. Hess, *Communication Circuits: Analysis and Design*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1971.
3. H. Taub and D. L. Schilling, *Principles of Communications Systems*, McGraw-Hill Book Company, New York, 1970.
4. J. Klapper and J. T. Frankle, *Phaselocked and Frequency-Feedback Systems*, Academic Press, Inc., New York, 1972.
5. *FM Broadcast Receivers. Standard Methods of Testing* (IEEE Standard 185-1975, IHF-T-200, 1975), available from IEEE, 345 E. 47th Street, New York, N.Y. 10017.
6. F. E. Terman, *Electronic and Radio Engineering*, 4th Ed., McGraw-Hill Book Company, New York, 1955.
7. C. E. Tibbs and G. G. Johnstone, *Frequency Modulation Engineering*, John Wiley & Sons, Inc., New York, 1956.
8. MC1357 Data Sheet, available from Motorola Semiconductor Products, Inc., Box 20912, Phoenix, Ariz. 85036.
9. D. M. Thomas, "Heath's Digital FM Tuner," *Radio Electronics*, **44**, 42-45, 50, 98, May 1973.

10. F. H. Hilbert, G. Cecchin, and J. Feit, "Stereo and Color Signal Demodulation with Integrated Circuit Techniques," *IEEE Trans. BC and TV Receivers*, BTR-14, 58, Oct. 1968.
11. L. Blaser and B. Cocke, "The μ A758, A Phase Locked Loop FM Stereo Multiplex Decoder," Application Note 319, Fairchild Semiconductor Corporation, Mountain View, Calif., June 1972.
12. T. McMullen, *FM and Repeaters for the Radio Amateur*, The American Radio Relay League, Inc., West Hartford, Conn., 1972.

PROBLEMS

10-1.1. The receiver in Fig. 10-2 is to be redesigned around a filter that requires 2000-ohm input and load resistances and has an insertion loss of 3.5 dB. Calculate the overall gain required between the RF amplifier input and the filter input.

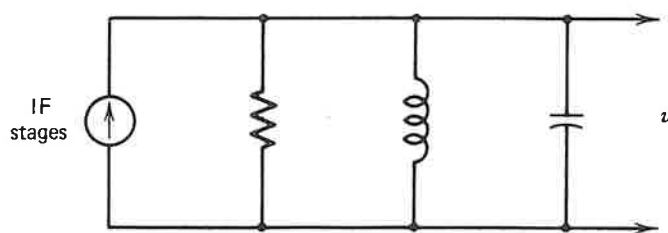
10-1.2. Show that a network whose transfer function has a constant amplitude and a phase angle that decreases linearly with frequency simply introduces a frequency-independent time delay and does not cause phase distortion.

10-1.3. The detector portion of the receiver in Fig. 10-2 is replaced by an IC detector requiring $160\ \mu\text{V}$ (rms) into a $9\text{-k}\Omega$ input resistance for full limiting. The detector must see a minimum 10-dB SNR at its input for satisfactory operation. The receiver is designed to receive broadcast FM (200-kHz bandwidth) and must reach full limiting for a $3\text{-}\mu\text{V}$ input signal at the RF amplifier input from an antenna with a 300-ohm source resistance and a noise temperature of 300 K. Calculate the overall gain and the minimum overall noise figure required for the stages ahead of the filter.

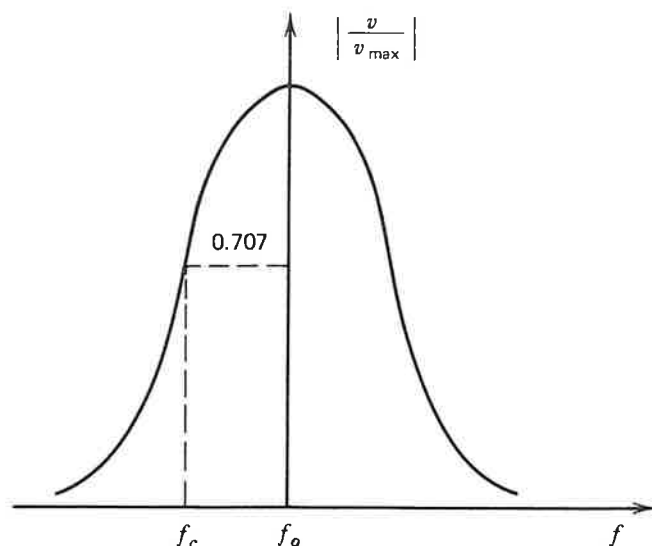
10-2.1. An AM receiver can be made to receive FM (after a fashion) by what is called slope detection. In it the receiver is tuned so that the carrier frequency (at IF) falls on the side of the IF response curve. To explore the slope detection process, imagine that the output of the receiver IF section can be modeled by a parallel resonant circuit with a $Q_t = 50$ and a resonance frequency f_o of 455 kHz driven by a current source as in Fig. P10-2.1a. The voltage v developed across the resonant circuit is applied to an envelope detector. The receiver is tuned so that the IF carrier frequency f_c lies at the frequency below resonance, where $|v/v_{\max}| = 0.707$, as in Fig. P10-2.1b.

- (a) Determine f_c .
- (b) Assume that the FM signal has a peak deviation of $\pm 5\text{ kHz}$. Plot the transfer characteristic of the detector and comment on its linearity.
- (c) Repeat part (b) for a peak deviation of $\pm 75\text{ kHz}$. This should tell you why slope detection does not work at all for broadcast FM.

Fig. P10-2.1 (a) Equivalent circuit of a slope detector; and (b) response curve indicating slope detector tuning.



(a)



(b)

10-3.1. Assuming that the integrated circuits are ideal, between what theoretical frequency limits will the zero-crossing detector of Fig. 10-19 operate linearly? Express your answer in terms of multiples of the IF center frequency.

10-3.2. Equation (10-34) gives the phase shift $\Delta\theta$ for a “quadrature” detector like the Motorola MC1357 in which the no-modulation phase shift is 45° . With reference to Fig. 10-14, this means that $\omega_o C_2 R = 1$, where ω_o is both the resonance frequency of the parallel combination of R , L , C_1 and the IF frequency without modulation. Remembering that $Q_o = \omega_o C_1 R$, derive equation (10-34).

10-3.3. For $Q_o = 50$ and $f_o = 10.7$ MHz, plot $\Delta\theta$ versus f over the frequency range

$$f_o \pm (f_o - \frac{f_o}{\sqrt{1 + 1/Q_o}})$$

10-3.4. For operation at $f_o = 10.7$ MHz the Motorola MC1357 data sheet specifies that $L = 2.0 \mu\text{H}$, $C_1 = 120$ pF, $R = 3900$ ohms, and $C_2 = 4.7$ pF. It states that the Q of R and L is 20. Assume that L is slug-tuned and can be adjusted to resonate with C_1 at 10.7 MHz, and then calculate the lower frequency limit of the detector in two ways:

(a) Assuming that $Q_o = 20$.

(b) Ignoring the data sheet value of 20 for Q_o and using $Q_o = \omega_o C_2 R$.

The measured value shown on the data sheet is 10.48 MHz.

10-4.1. Verify that multiplying V_c in equation (10-39) by $F(t)^+$ in (10-36) in an analog multiplier with gain K yields the expression for V_L given by equation (10-40).

12

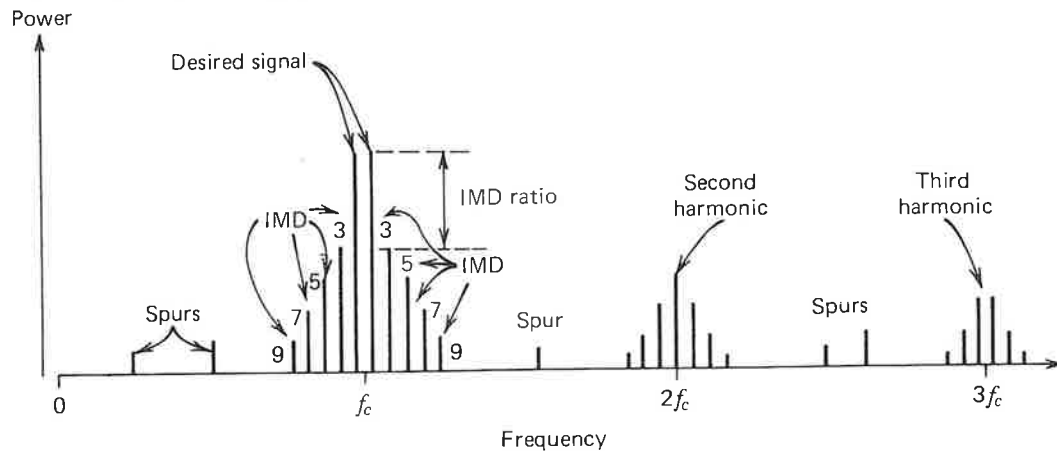
Linear Power Amplifiers

Power amplifiers (PAs) are used when the efficiency and output power of an amplifying circuit are the important considerations. The various types of PAs are identified by their classes of operation, that is, Classes A, B, C, D, E, F, G, H, and S. Except for Class A, all of these amplifier types are easily differentiated from small-signal amplifiers by their circuit configurations, methods of operation, or both. There is no sharp dividing line between small-signal and Class A power amplifiers; the choice of terms depends upon the intent of the designer.

Class A and B linear RF power amplifiers are considered in this chapter. These PAs have appreciable power gain, produce a magnified replica of the input signal voltage or current wave, and are typically used in SSB and multimode transmitters where accurate reproduction of both the envelope and the phase of the signal is required. (Linear RF amplification can also be accomplished with other types of PAs by the use of techniques described in Chapter 16). Tuned circuits or low-pass filters are not integral components of Class A or B amplifiers; however, they are often included to ensure adequate harmonic suppression.

Because the bandwidth of a typical RF signal is a small fraction of the carrier frequency (and of the bandwidth of the PA), the unwanted signals in the output of the power amplifier may be divided into three categories. Figure 12-1 depicts the relationship of these signals to the desired signal, which consists in this case of two tones of equal amplitudes at frequencies $f_c \pm f_m$. Amplifier nonlinearity produces two types of unwanted signals, called har-

Fig. 12-1 Distortion products in an RF power amplifier.



monics and intermodulation distortion (IMD) products. The IMD products (identified as third, fifth, seventh, and ninth orders in Fig. 12-1) are prominent near the carrier frequency. They cause distortion in the received signal and/or adjacent channel interference. Other unwanted signals include parasitic and subharmonic oscillations and mixer products; they are called spurious products or simply “spurs.” In an RF PA, the harmonics and some of the spurious products may be removed by filters; however, the IMD generated must be of an acceptably low level.

Bipolar junction transistors (BJTs) were used in most RF power applications prior to the publication of this text. The recently developed vertical MOSFET (VMOS) devices were being incorporated into many new designs because of their ease of biasing and immunity to second breakdown. Consequently, both BJTs and VMOS devices are used in the following chapters, with attention given to their relevant characteristics and the associated differences in design. (Similar design principles are generally applicable to high-power vacuum-tube equipment.) Construction of both bipolar and VMOS RF power transistors is discussed in Appendix 12-1.

To ensure PA reliability, the designer must pay careful attention to such transistor ratings as maximum base (gate) and collector (drain) voltages and currents, maximum junction temperature, and operating conditions that lead to second breakdown in BJTs. The meaning of peak voltage and current ratings is obvious, although the designer must note carefully how the given rating is related to other circuit conditions such as base bias. Junction temperature is discussed in Section 12-9. Second breakdown is caused by an excessive value of the instantaneous collector voltage-current product, resulting in hotspotting within the transistor.

Device specifications must be interpreted cautiously, since values given on data sheets are often related to the intended use of the device rather than its

maximum ratings. For example, a “28-V” linear RF power transistor¹ is intended for use in a Class B PA with a 28-V dc supply, and hence has a maximum voltage rating of at least 56 V. Similarly, the maximum dc supply current is sometimes given as a rating rather than the maximum instantaneous collector current.

Two parameters that are very useful for comparing different PA configurations and classes of operation are the efficiency and power output capability. *Collector power efficiency* (η), or simply *efficiency* is defined as the ratio of the RF output power to the collector circuit dc input power. The *normalized power output capability* ($P_{\max}$) is defined as the maximum RF output power that can be obtained without exceeding the peak collector voltage and current of 1 V and 1 A, respectively. The actual output power capability is obtained from $P_{\max}$ by multiplying it by the transistor collector voltage and current ratings. Values of $P_{\max}$, η , and other parameters for various types of amplifiers are tabulated in Appendix 14-1.

It will be convenient in this and the following chapters to express instantaneous voltages and currents in terms of *angular time*

$$\theta = \omega t = 2\pi f t \quad (12-1)$$

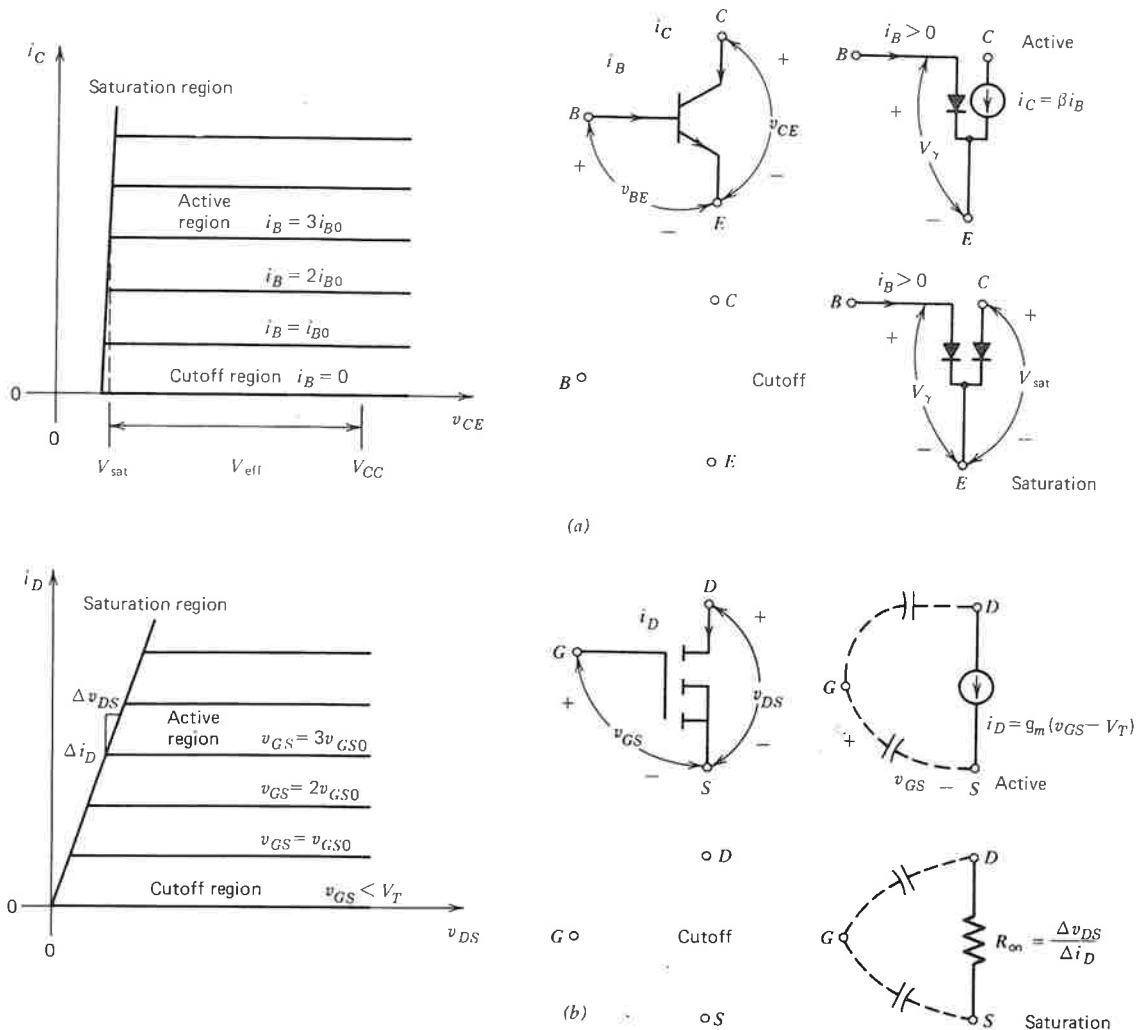
because most processes in power amplifiers are periodic and can therefore be described completely in terms of the behavior within a single RF cycle ($0 \leq \theta < 2\pi$).

Transistor characteristics have been simplified and high-frequency effects have largely been ignored to allow tractable equations and analytical modeling of PAs. Nonetheless, the designer will find the predicted results generally accurate to within a few percent and quite adequate for design purposes. Large signal transistor behavior is divided into three regions, designated *cutoff*, *active*, and *saturation* (Fig. 12-2). Note that the term *saturation* is used here with FETs in a manner analogous to its use with BJTs; many texts on semiconductor theory use *saturation* to refer to a limiting current.

An NPN BJT (Fig. 12-2a) is placed in the cutoff region by application of v_{BE} less than its cut-in voltage V_γ (≈ 0.7 V for silicon devices); in this region the BJT is essentially an open circuit. As v_{BE} is increased to just greater than V_γ , the transistor enters the active region. The base-emitter junction becomes a forward-biased diode and the collector-emitter junction becomes a current source whose current is linearly proportional to the base current. To operate in the active region, v_{CE} must be larger than the saturation voltage V_{sat} ($\approx V_{CE,\text{sat}} \approx 0.3$ V for a single BJT with a low collector current). If the load will

¹Common designations and their uses are 8 V (portable), 12 V (mobile and aircraft), 28 V (military), and 50 V (fixed).

Fig. 12-2 Simplified device characteristics and models. (a) Bipolar junction transistor (BJT); (b) VMOS field effect transistor (FET).



not allow $v_{CE} > V_{sat}$, the device enters the saturation region, and the collector-emitter junction voltage is roughly equivalent to a constant voltage V_{sat} .

An FET (Fig. 12-2b) is similarly placed in the cutoff region by application of v_{GS} less than threshold voltage V_T (≈ 2 to 3 V for VMOS power FETs); it is also essentially an open circuit under these conditions. As v_{GS} is raised above V_T , the FET enters the active region and the drain-source connection is roughly equivalent to a current source. The value of the drain current in a VMOS FET is roughly linearly proportional to $v_{GS} - V_T$ although in small-signal FETs it is proportional to $(v_{GS} - V_T)^2$. (See Appendix 12-1.) An FET is roughly equivalent to a resistance R_{on} when in the saturation region; it is placed in this region if the load would produce a drain voltage less than $i_D R_{on}$.

if the FET were in the active region. Vacuum tube characteristics are somewhat similar to FET characteristics; however, V_T is negative and a positive grid-cathode voltage will cause significant grid current to flow.

12-1 Class A Amplification

The circuit of a grounded-emitter Class A power amplifier (Fig. 12-3a) is similar to that of the analogous small-signal amplifier. In a PA application, the load resistance is usually low enough to make the effects of the device shunt resistance and reactance negligible. Since the Q (quiescent) point (i.e., I_{CQ}) is selected to keep the transistor in its active region at all times, the device is equivalent to a current source (Fig. 12-3b). The parallel-tuned circuit or equivalent filter is not a necessary part of a Class A amplifier. However, since no device is perfectly linear, a tuned circuit or filter is often included to prevent harmonic currents from reaching the load.

In the amplifier of Fig. 12-3a the driving signal and bias applied to the base are assumed to produce the collector current $i_C(\theta)$ shown in Fig. 12-3c.

$$i_C(\theta) = I_{CQ} - I_{cm} \sin \theta \quad (12-2)$$

(Methods of producing proper drive and bias are discussed later.) Harmonic components of $i_C(\theta)$ are omitted in (12-2) because they generally have little effect on the power or efficiency of a Class A amplifier. The dc and ac components of $i_C(\theta)$ must flow through the choke (RFC) and blocking capacitor (C_b), respectively, and therefore become the input current $I_{dc} = I_{CQ}$ and output current $i_o(\theta) = I_{cm} \sin \theta$, respectively. Harmonic currents produced by device nonlinearities are bypassed to ground by the parallel-tuned circuit. All of the carrier-frequency component of $i_C(\theta)$ flows through the load resistance R (assuming proper tuning of the parallel-tuned circuit and negligible effects from RFC and C_b) and produces the output voltage

$$v_o(\theta) = I_{cm}R \sin \theta = V_{om} \sin \theta \quad (12-3)$$

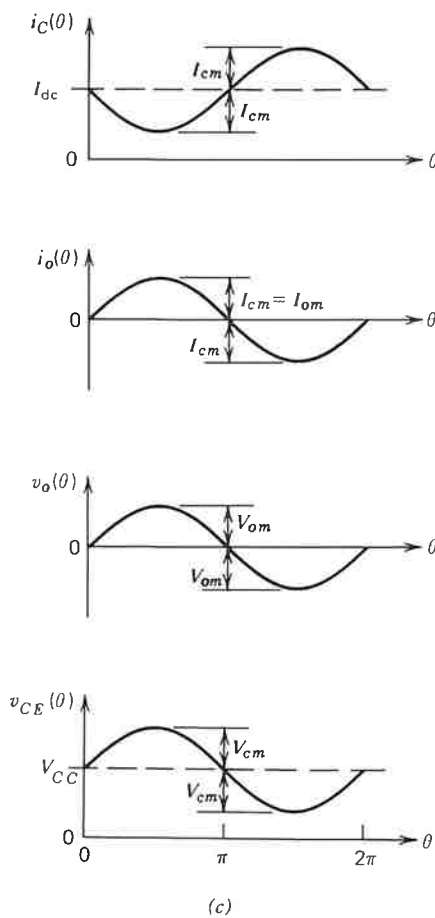
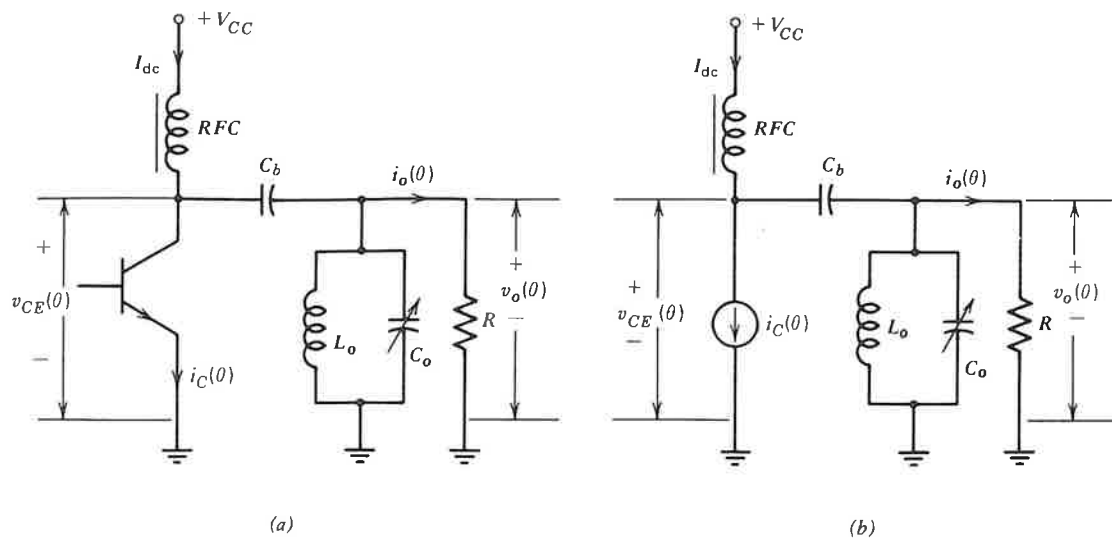
The collector voltage must have a dc component equal to the supply voltage and an ac component equal to the output voltage, thus

$$v_C(\theta) = V_{CC} + V_{om} \sin \theta = V_{CC} + V_{cm} \sin \theta \quad (12-4)$$

Since current-source operation of the device can be maintained only when $v_C(\theta)$ is positive, it is necessary to limit the output voltage V_{om} to values less than V_{CC} .² (Actually, the maximum V_{om} must be slightly less than V_{CC} due to

²An analogous requirement applies to VMOS FETs and vacuum tubes.

Fig. 12-3 Class A amplifier. (a) Circuit; (b) equivalent circuit; and (c) waveforms.



the saturation effects that will be discussed in Section 12-3.) Since the collector current $i_c(\theta)$ must similarly be positive, $I_{dc} = I_{CQ}$ is usually set equal to, or slightly larger than, the peak output current $I_{om} = V_{om}/R \leq V_{CC}/R$. Consequently, the input power is

$$P_i = V_{CC}I_{dc} = \frac{V_{CC}^2}{R} \quad (12-5)$$

the output power is

$$P_o = \frac{V_{om}^2}{2R} \leq \frac{V_{CC}^2}{2R} \quad (12-6)$$

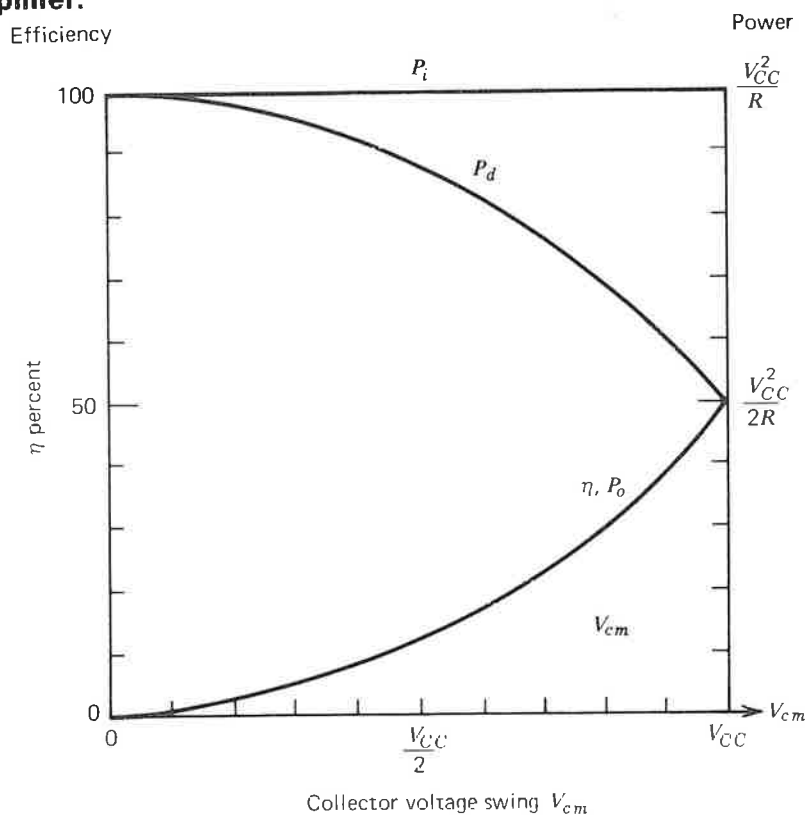
and the efficiency is

$$\eta = \frac{P_o}{P_i} = \frac{V_{om}^2}{2V_{CC}^2} \leq \frac{1}{2} \quad (12-7)$$

The power P_d dissipated in the transistor is the difference between P_i and P_o . The relationships of power and efficiency to the output voltage are shown in Fig. 12-4. A good design will assume the maximum value of $P_d = P_i$ to allow for no-signal or reduced-signal conditions.

Inspection of the waveforms in Fig. 12-3c shows that when the Class A PA is delivering its maximum output, the peak instantaneous collector voltage

Fig. 12-4 Efficiency and power versus collector voltage swing in a Class A amplifier.



and current are $v_{C,\max} = 2V_{CC}$ and $i_{C,\max} = 2I_{CQ} = 2V_{CC}/R$. The normalized power output capability is therefore

$$P_{\max} = \frac{P_{o,\max}}{v_{C,\max} i_{C,\max}} = \frac{V_{CC}^2/2R}{(2V_{CC})(2V_{CC}/R)} = \frac{1}{8} \quad (12-8)$$

If two devices operating as Class A amplifiers are combined in a complementary or a push-pull configuration (similar to that of Fig. 12-5a), $P_{\max} = \frac{1}{4}$. This arrangement cancels most of the even harmonic currents generated by device nonlinearity.

Class A RF power amplifiers most commonly are used as low-level driver amplifiers. In these applications, the power consumed by the Class A amplifier is a relatively small portion of the total transmitter power. Class A RF PAs are also used at microwave frequencies where it is difficult to employ other classes of amplification.

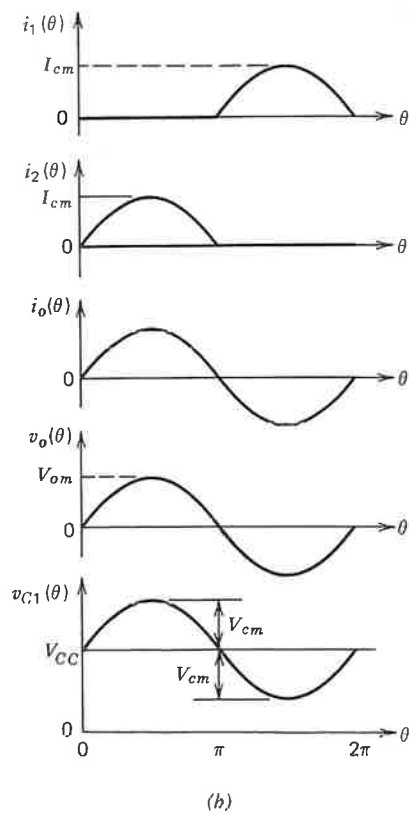
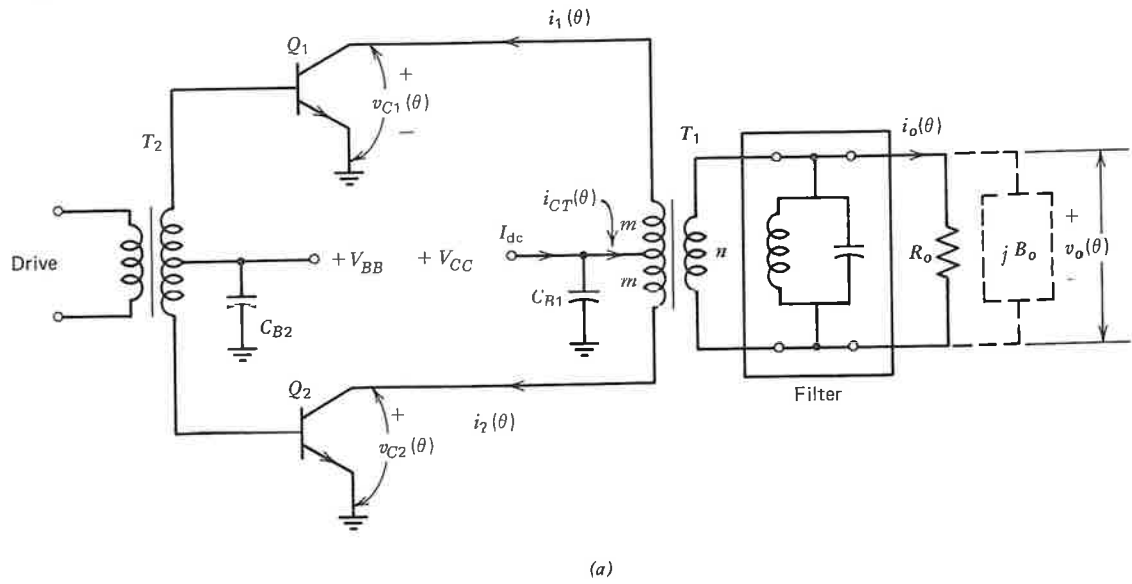
Example 12-1.1. Design (specify component values and ratings for) the output portion of a Class A PA that delivers 1 W to a 50-ohm load at $f = 10$ MHz with a supply voltage of +12 V: First, the peak RF output voltage from (12-6) is $V_{om} = (2RP_o)^{1/2} = 10$ V. The peak RF current is then $I_{cm} = V_{om}/R = 200$ mA; hence, $I_{dc} = 200$ mA, $P_i = 12(0.2) = 2.4$ W, and $\eta = 1/2.4 = 41.7$ percent. Peak device ratings are $v_{C,\max} = 12 + 10 = 22$ V, $i_{C,\max} = 200 + 200 = 400$ mA. During normal operation, $P_d = 2.4 - 1 = 1.4$ W; however, to allow safe operation with reduced signal amplitude, the heatsink must dissipate $P_d = P_i = 2.4$ W. For $Q = 5$, $X_{L_o} = X_{C_o} = \frac{50}{5} = 10$; hence $L_o = 0.159 \mu\text{H}$ and $C_o = 1592$ pF. The reactance of the RFC should be at least $10R$ to minimize its effect on the circuit, hence $RFC \geq 8 \mu\text{H}$. Similarly, for $X_{C_b} \leq R/10$, $C_b \geq 3200$ pF.

12-2 Class B Amplification

Class B operation is more efficient than Class A for linear RF amplification; hence Class B is frequently used in medium- and high-power linear PAs. The most common configuration is the push-pull transformer-coupled circuit of Fig. 12-5a, but complementary versions similar to those used for audio frequency (AF) amplification are also used [1]. A tuned, single-ended version can be used for narrowband linear amplification; it is similar to the Class C amplifiers discussed in Section 13-1.

The principles of operation of a Class B RF PA are similar to those of a Class B AF PA [2]. The two transistors are driven 180° out-of-phase so that each is active for half of the cycle and cut off during the other half of the cycle. When the transistors are in the active region, they are essentially

Fig. 12-5 Class B amplifier. (a) Circuit; and (b) waveforms.



current sources. The higher efficiency may be attributed to the zero collector current in the transistors when their collector voltages are highest. It is important to note that in Class B amplification, neither device by itself produces an amplified replica of the amplifier's input. The signal is split apart, amplified efficiently, and then reassembled.

Assume that the devices are perfectly linear and that each device is a source of a half-sinusoid current of peak amplitude I_{cm} (Fig. 12-5b). During a given half-cycle, only one-half of the primary winding of transformer $T1$ carries current. The transformation of the current in m turns of the primary to n turns of the secondary produces a sinusoidal output current

$$i_o(\theta) = \frac{m}{n} I_{cm} \sin \theta \quad (12-9)$$

which in turn produces the output voltage

$$v_o(\theta) = \frac{m}{n} I_{cm} R_o \sin \theta = V_{om} \sin \theta \quad (12-10)$$

Transformation of this voltage to the primary winding gives the collector voltage waveform

$$v_{C1}(\theta) = V_{CC} + V_{cm} \sin \theta \quad (12-11)$$

where the collector voltage swing is

$$V_{cm} = \frac{m}{n} V_{om} = \frac{m^2}{n^2} I_{cm} R_o = I_{cm} R \quad (12-12)$$

Above, R is the resistance seen across one-half of the primary winding with the other half open.

Maintenance of a nonnegative collector voltage requires that $V_{cm} \leq V_{CC}$, thus limiting the output power to

$$P_o = \frac{V_{cm}^2}{2R} \leq \frac{V_{CC}^2}{2R} = \frac{V_{CC}^2}{2(m^2/n^2)R_o} \quad (12-13)$$

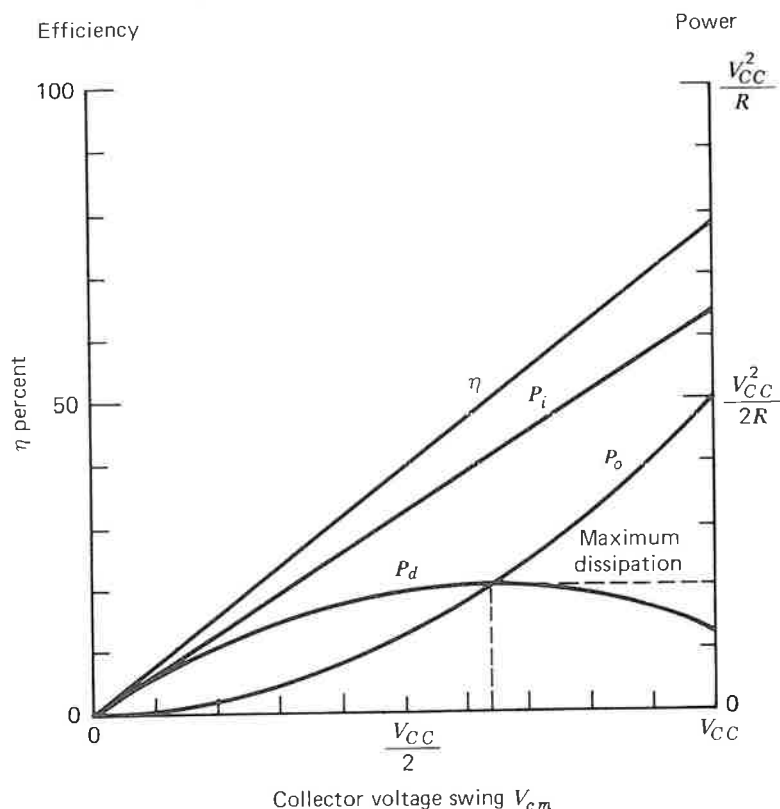
The center-tap current $i_{CT}(\theta) = I_{cm} |\sin \theta|$ is the sum of the two collector currents, and the input current I_{dc} is the dc component of the center-tap current, thus

$$I_{dc} = \frac{1}{2\pi} \int_0^{2\pi} i_{CT}(\theta) d\theta = \frac{2I_{cm}}{\pi} = \frac{2}{\pi} \frac{V_{cm}}{R} \quad (12-14)$$

The efficiency is then

$$\eta = \frac{P_o}{P_i} = \frac{\pi V_{cm}}{4 V_{CC}} \leq \frac{\pi}{4} \approx 0.785 \quad (12-15)$$

Fig. 12-6 Efficiency and power versus collector voltage swing in a Class B amplifier.



Since the peak collector voltage is $V_{CC} + V_{cm} \leq 2V_{CC}$ and the peak collector current is $I_{cm} \leq V_{CC}/R$,

$$P_{\max} = \frac{V_{CC}^2/2R}{(2V_{CC})(V_{CC}/R)} = \frac{1}{4} \quad (12-16)$$

The power dissipated in each device (P_{d1}) is half of the difference between P_i and P_o . The maximum dissipation in each transistor is

$$P_{d1,\max} = P_{d2,\max} = \frac{V_{CC}^2}{\pi^2 R} \quad (12-17)$$

and occurs when $V_{cm} = 2V_{CC}/\pi$ (see Fig. 12-6).

Example 12-2.1. Design a Class B amplifier to deliver 25 W to a 50-ohm load using a dc supply with $V_{CC} = 28$ V. Tuned matching networks are not to be used: From (12-13), $R \leq 28^2/(2 \times 25) = 15.7$ ohms. The most convenient value of R is 12.5 ohms, obtained with $m/n = \frac{1}{2}$. To obtain 25 W, $V_{cm} = \sqrt{2RP_o} = \sqrt{2 \times 12.5 \times 25} = 25$ V. Then $I_{cm} = V_{cm}/R = 2$ A, $I_{dc} = 2I_{cm}/\pi = 1.27$ A, and $P_i = 1.27 \times 28 = 35.6$ W. From these, $\eta = 25/35.6 = 70.1$ percent, and the maximum

collector dissipation in each device is $(28)^2/(\pi^2 \times 12.5) = 6.35 \text{ W}$ from (12-17). The transistors must withstand peak voltages of $28 + 25 = 53 \text{ V}$ and peak collector currents of 2 A .

12-3 Practical Considerations

The previous two sections have discussed idealized Class A and B power amplifiers. The performance of a real amplifier will not be quite as good as that predicted for the idealized model because real devices have saturation voltage and/or resistance and real loads can include reactance. At high frequencies, charge storage effects [3] can cause distortion of the collector current waveform, further reducing efficiency. In general, these effects may be neglected when $f < f_T/10$ but are quite significant when $f > f_T/2$.

Saturation Voltage

Bipolar junction transistors are characterized by a largely constant collector-to-emitter voltage during saturation (Fig. 12-2a). A BJT in the power amplifier circuits of Figs. 12-3 or 12-5 will enter saturation whenever the PA is driven hard enough so that the minimum instantaneous collector voltage is equal to the saturation voltage V_{sat} . To avoid saturation, the driving signal must be small enough to keep the collector voltage swing below

$$V_{\text{eff}} = V_{CC} - V_{\text{sat}} \quad (12-18)$$

The maximum output voltage, output power, output current, and input current are calculated using the *effective supply voltage* V_{eff} in place of V_{CC} . However, input power is still calculated using V_{CC} . This decreases the maximum efficiency by a factor of V_{eff}/V_{CC} ; for example, the maximum efficiency of Class A and B PAs is reduced to $\frac{1}{2}(V_{\text{eff}}/V_{CC})$ and $(\pi/4)(V_{\text{eff}}/V_{CC})$, respectively.

At low frequencies ($f \leq f_T/10$), V_{sat} is simply the specified $V_{CE(\text{sat})}$ for the device (typically 0.3 V for silicon transistors at low currents). If a Darlington arrangement is used, $V_{\text{sat}} \approx V_{CE(\text{sat})} + V_\gamma$, where V_γ is the forward-biased base-emitter voltage drop (typically, 0.7 V for silicon transistors). At higher frequencies ($f > f_T/10$), V_{sat} becomes larger than $V_{CE(\text{sat})}$ for steady-state operation. The increase in V_{sat} appears to be related to the time it takes for the entire transistor to become saturated. This “RF saturation voltage” may be as high as 1 or 2 V and is sometimes specified in device data sheets.

Saturation Resistance

The saturation (or "on") resistance (R_{on}) of FETs, like the saturation voltage of BJTs, limits the maximum output but does not affect operation in the active region. Saturation occurs when the minimum instantaneous drain voltage is equal to the product of drain current and saturation resistance; that is, for Class B, $V_{DD} - I_{dm}R = I_{dm}R_{on}$. Consequently, to avoid saturation, V_{dm} must be less than the effective supply voltage

$$V_{eff} = \frac{R}{R + R_{on}} V_{DD} \quad (12-19)$$

For Class A operation, or for a single-ended, tuned, Class B PA (see Section 13-1), the effects of R_{on} are doubled; that is, $V_{eff} = V_{DD}R/(R + 2R_{on})$. In either case, the maximum output voltage and power and input current are computed using V_{eff} , and input power is then computed using V_{DD} , as for the effects of saturation voltage.

Reactive Loads

Although most real-life loads include at least some reactance, most discussions of PA design techniques ignore load reactance. The causes of load reactance include mistuning, the variation of filter impedances with frequency, the transformer inductance, and antenna impedance variations. The effects of load reactance include reduced efficiency, increased dissipation, and increased second-breakdown stresses.

Standard circuit analysis techniques can be used to replace any passive linear circuit (at a single frequency) by a single conductance shunted by a single susceptance. These appear as $1/R_o$ and B_o in Fig. 12-5a and are transformed to an admittance $Y = 1/R + jB$ across one winding of the primary of T_1 . Since the transistors are operated as current sources, the output current $i_o(\theta)$ is still given by (12-9). However, the output voltage $v_o(\theta)$ now has a magnitude (see Problem 12-3.2)

$$V_{om} = \rho \left(\frac{m}{n} \right) I_{cm} R_o \quad (12-20)$$

where

$$\rho = \frac{1/R}{|Y|} = \frac{1}{\sqrt{1 + (BR)^2}} \quad (12-21)$$

and it is shifted in phase with respect to the output and collector currents. Power output and input are still computed from $V_{om}^2/2R$ and $V_{CC}I_{dc}$, respectively. However, to achieve the same output power, the dc supply current and input power must be increased by the factor $1/\rho$ from those required if the

reactance were not present. Consequently, the efficiency decreases by the factor ρ ; for example, a Class B PA with a reactive load has a maximum efficiency of $(\pi/4)\rho$.

A maximum power dissipation per device of

$$P_{d1,\max} = \frac{V_{CC}^2}{\pi^2 \rho^2 R} \quad (12-22)$$

is possible and occurs with a collector voltage of

$$V_{cm} = \frac{2V_{CC}}{\pi\rho} \quad (12-23)$$

(provided, of course, that this value is less than V_{CC}). In addition to the larger heatsinks required, the designer must also consider the instantaneous collector voltage-current product, since this leads to second breakdown. The *safe operating area* is commonly specified in the data sheets for low-frequency power transistors; if this is available, the designer can simply determine several instantaneous values of collector voltage and current and check them against the boundaries given in the data sheet. However, data sheets for RF power transistors more commonly give the maximum *voltage standing wave ratio* (VSWR) that the transistor can withstand. The VSWR can be determined from the load impedance using formulas given in Section 16-6.

Example 12-3.1. Determine the effects resulting from a susceptance of 20 mmhos across the load of the Class B amplifier of Example 12-2.1 if drive is increased to maintain the 25 W output: For a 25-W output, V_{cm} must remain at 25 V. The 20 mmhos at the load becomes 0.08 mhos at the collectors; hence, $|Y| = 0.113$ mhos and $I_{cm} = V_{cm}|Y| = 2.83$ A. Following Example 12-2.1, $I_{dc} = 1.80$ A, $P_i = 50.4$ W, $\eta = 49.6$ percent, and $P_{d1} = 12.7$ W. The voltage rating is the same as in Example 12-2.1, but the current rating is increased to 2.83 A, and the dissipation rating is more than doubled.

12-4 Intermodulation Distortion and Bias

Intermodulation distortion (IMD) consists of those distortion products occurring near the carrier frequency (Fig. 12-1) and results from an RF PA's inability to reproduce exactly the envelope and phase of the input signal. (Exact reproduction of the sinusoidal carrier is unnecessary, since harmonics are removed by the tuned output circuit or filter.) Two major causes of IMD are crossover effects (discussed subsequently) and gain reduction at high current (Section 12-5). Two other sources of IMD are device saturation and variation of collector capacitance with collector voltage. Saturation causes

both “flat-topping” of the envelope and unwanted phase modulation (due to storage-time effects) and is avoided by not overdriving the amplifier (i.e., by not exceeding the maximum output for the intended class of operation). The voltage-variable capacitance is inherent in semiconductor junctions, but its effect is usually small in most PA applications.

Intermodulation distortion is usually measured with a test signal composed of two sinusoids of equal amplitudes that are separated in frequency by 1 to 2 kHz. The resulting products (Fig. 12-1) are called “third-order,” “fifth-order,” and so forth, in the same manner as mixer intermodulation products (Section 7-5). The most commonly used measure of IMD [4] is the ratio of the largest intermodulation product to the amplitude of one of the two tones; -30 dB is usually the worst allowable IMD ratio for a transmitter. This ratio is easily measured on a spectrum analyzer but is too small to be measured accurately (or even detected) with an oscilloscope. It is interesting (and useful) to note that mirror-image upper and lower sideband intermodulation products are the result of either pure envelope distortion or pure phase distortion; unequal intermodulation products of the same order indicate simultaneous envelope and phase distortion [5].

A less commonly used test of linearity employs noise-like signals [6]. Consider a *band-limited white noise* signal, which has a power spectrum that is flat from $f_c - f_m$ to $f_c + f_m$ and zero at all frequencies outside of this band. If this signal is the input to an amplifier, any signals in the output that are immediately above $f_c + f_m$ or below $f_c - f_m$ are due to amplifier nonlinearity (this may be seen through use of the convolution integrals discussed in Section 7-1). Because it is difficult to produce a rectangular spectrum, it is convenient to use a noiselike signal with zeros in its power spectrum. Such a spectrum is provided by amplitude modulation of a carrier by a pseudorandom noise sequence of $+1$ s and -1 s; Zeros occur at $\pm f_m$, $\pm 3f_m$, $\pm 5f_m$, and so forth, from the carrier where f_m is the “chipping” (bit) rate. Energy present in the output at these frequencies is then a measure of the amplifier’s nonlinearity. The advantage of using a noiselike test signal is that it simulates more accurately the effects of a speech signal than does a two-tone test.

Crossover Distortion and Bias

Real devices (BJTs, FETs, or vacuum tubes) operated in Class B do not change abruptly from a cutoff mode to an active (linear current source) mode. The actual change is gradual and nonlinear, and usually involves an offset voltage. Consequently, an effect called “crossover distortion” alters the waveform as one device enters cutoff and the other device becomes active. These effects are most pronounced with signals of relatively small amplitudes (with respect to maximum PA output).

Crossover distortion is minimized by biasing of the bases or gates of the active devices to produce a small quiescent current in the collector or drain. Too little quiescent current results in no output when the input is less than a certain amplitude. Too much bias results in an excess gain for small-amplitude signals. The amount of quiescent current required for minimum distortion is most readily determined experimentally, since theoretical prediction is quite complicated. Typically, the minimum IMD ratio occurs with a quiescent current between 1 and 10 percent of the peak collector or drain current; the exact value is seldom critical. Because of this quiescent current, each device conducts current during slightly more than one-half of an RF cycle. This led to the use of the term "Class AB" for vacuum tube PAs and some early solid state PAs; however, many authors simply use "Class B" with the understanding that a quiescent current will be used.

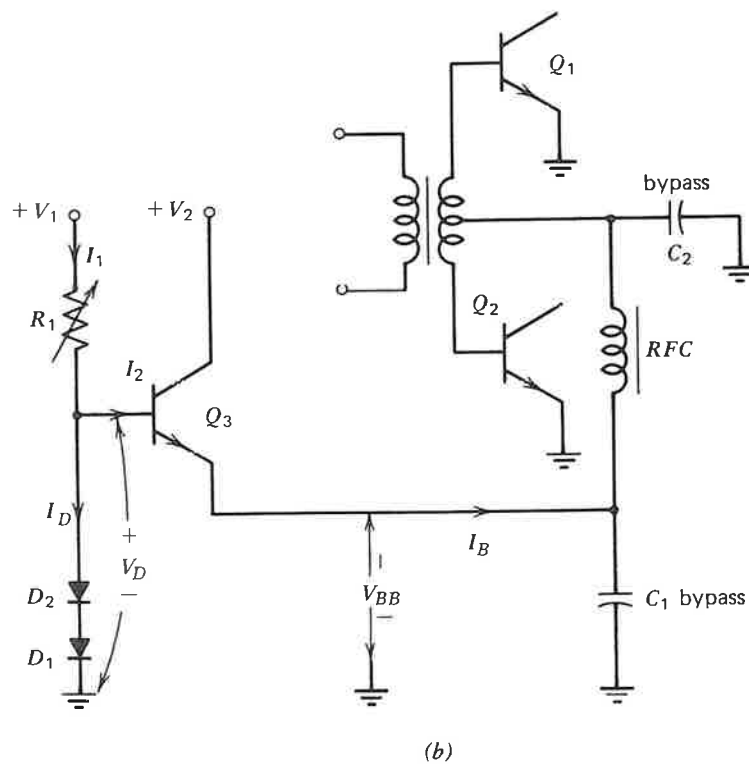
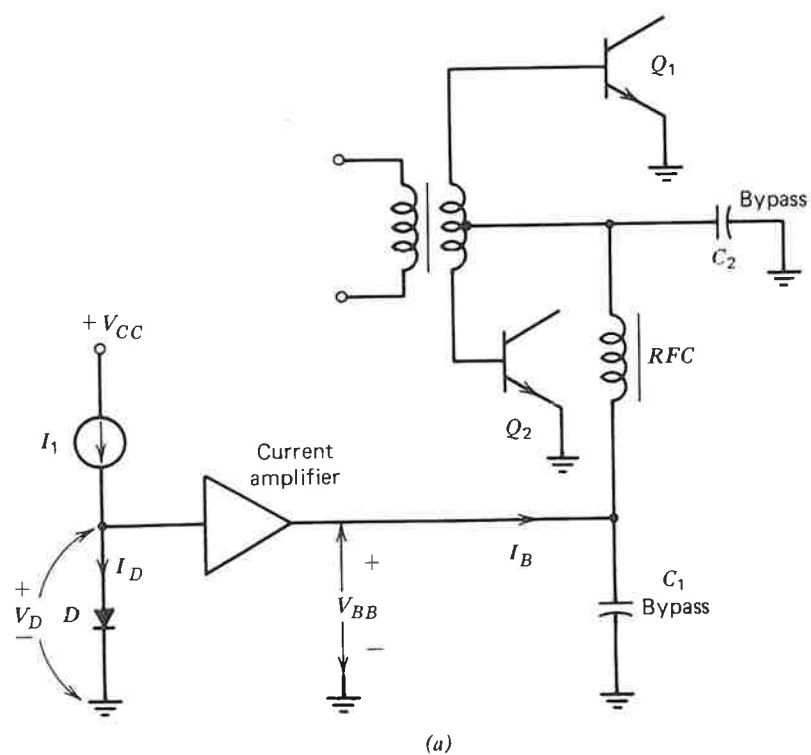
Bias Supplies

Suitable bias for FET PAs is easily obtained from a regulated supply voltage through an adjustable resistive voltage divider and an appropriate bypass network [5, 7]. However, a simple fixed voltage bias cannot be used with BJT PAs. The effect of increasing the temperature of a pn junction is approximately equivalent to a shift of the device characteristics by about $2 \text{ mV}/^\circ\text{C}$; that is, a 1°C increase in temperature produces approximately the same increase in current as would a 2-mV increase in junction voltage.³ Suppose that a bias voltage V_{BE} is applied to the base of a BJT and is quickly adjusted to produce the desired quiescent collector current. The quiescent current causes power to be dissipated in the transistor, raising the temperature of the junction. This in turn increases the collector current, which increases the dissipation. Since there is little dc resistance in the collector circuit, there is almost no limit to the collector current that can be drawn. This process is called thermal runaway and leads to device destruction.

Maintaining an essentially constant quiescent collector current then clearly requires a source of bias voltage that decreases with the junction temperature at approximately $2 \text{ mV}/^\circ\text{C}$. Such a voltage source can be obtained by mounting a diode of construction similar to that of the RF power transistor on the heatsink and biasing it with a constant current. The diode voltage controls a "current amplifier" that supplies bias current to the bases (Fig. 12-7a). The current amplifier may be an operational amplifier [9] or voltage regulator [10].

³This positive temperature coefficient is not obvious in (4-2), which does not show the temperature dependence of the reverse saturation current I_{ES} . The overall effect [8] is approximately equivalent to a change in V_{BE} of $1.7 \text{ mV}/^\circ\text{C}$ for germanium and $1.8 \text{ mV}/^\circ\text{C}$ for silicon.

Fig. 12-7 BJT bias circuits. (a) Using operational amplifier; and (b) using emitter follower.



An alternative current amplifier (Fig. 12-7b) uses a pair of diodes and an emitter (voltage) follower [11]. In all cases, current is supplied through the input transformer center-tap and a bypass network (C1-RFC-C2). Reference diodes are built into some RF power transistor packages, and special biasing components are available that match the temperature characteristics of RF power transistors and mount on the heatsink.

The current that the bias supply must furnish varies with the drive level. In Class B operation, each transistor base requires an approximately half-sinusoidal driving current of peak amplitude $(I_{cm} + I_C)/\beta$, where β is the current gain of the RF power transistor and I_C is the quiescent collector current. The center-tap current is then a full-wave rectified sine wave added to the bias current, and the maximum bias supply current (required at peak amplifier output) is

$$I_{BB,\max} = \frac{I_{cm}}{\pi\beta} + \frac{I_C}{\beta} \quad (12-24)$$

Bias supplies for Class A operation are similar to those for Class B operation. Although more base current ($I_{BB} \approx I_C/\beta$) is required, a slightly larger fractional variation of I_C with temperature is allowable, since its effect on IMD is small.

Real bias-current amplifiers require some input current (I_2 in Fig. 12-7b), and therefore tend to load the reference diode. Since any variation of the diode current degrades the ability of the bias amplifier to maintain constant quiescent current, the reference diode is typically biased to conduct 10 to 100 times the current required by the amplifier loading it. Exact analysis of bias circuits containing resistive voltage drops may require the use of graphical or numerical techniques. Effects of temperature differences can, however, usually be estimated through the 2 mV/°C equivalence.

Example 12-4.1. Design the bias supply circuit of Fig. 12-7b for the PA of Example 12-2.1, and determine the effects of 10°C temperature difference between the reference diode and the RF power transistor. Assume that $\beta = 20$ for all transistors and that a quiescent current of 25 mA per device gives the minimum IMD ratio: From Example 12-2.1, $I_{dc} = 1.27$ A without the quiescent current, and hence $I_{BB,\max} = (1.27 + 0.025)/20 = 66$ mA, and $I_2 = 0.066/20 = 3.3$ mA. A diode current of 10 times I_2 is selected so that $I_D = 33$ mA and $I_1 = 36.3$ mA. Since the 28 V supply is much larger than the approximate 1.4 V drop across the two diodes, a resistor $R_1 = (28 - 1.4)/0.0363 = 733$ ohms will behave like a current source. (R_1 would actually be a 2-k Ω variable resistor.) The effect of 10°C temperature difference will be about the same as that of a 20-mV change in V_{BE} ; the quiescent current will therefore increase by a factor of $\exp(0.02/0.026) = 2.16$ to a value of 53.9 mA per device.

12-5 Drive and RF Feedback

Large-signal operation of any device—BJT, FET, or vacuum tube—is quite different from small signal operation. Most noticeable are the increased effects of device nonlinearities at higher outputs. The discussion of distortion in small-signal amplifiers may have caused the reader to give up hope for linear power amplification. Linear large signal operation is possible, however, and is achieved by the careful design of drive and feedback circuits. Although design principles for output circuits are similar for all devices, the design principles relating to input circuits vary greatly from one device to another.

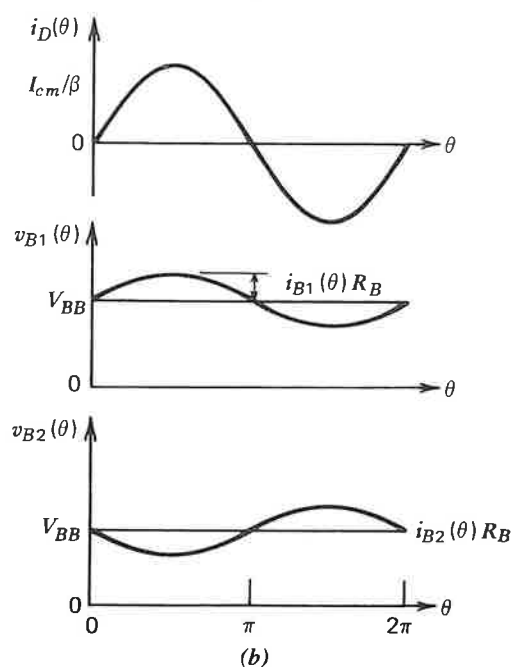
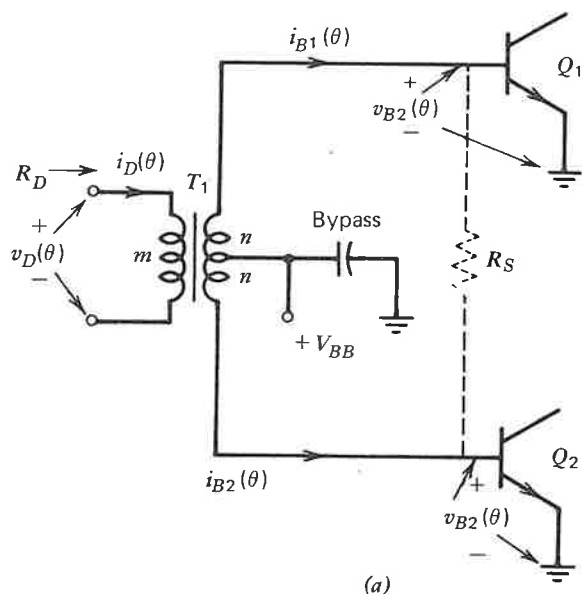
Drive

BJTs. The relationship of the collector current to the base-emitter voltage in a BJT is exponential, and is therefore inherently nonlinear for large-signal operation. For this reason, power amplifiers often use current drive rather than voltage drive. The collector current and the base current are nearly linearly related [12] because the base-emitter voltage is a logarithmic function of the current driven into the base. Ideally, the logarithmic and exponential functions cancel each other, but in practice, there is not perfect linearity. Most notable is the decrease in current gain β at high collector currents.

To understand how a Class B PA using BJTs is driven, suppose that a sinusoidal current is driven into the primary winding of transformer $T1$ in Fig. 12-8a. During the positive half-cycle of this current, there is a tendency for current in the secondary winding to flow from the base of $Q2$ toward the base of $Q1$. Since this direction of current flow is in the reverse direction for the base of $Q2$ and the forward direction for the base of $Q1$, current flows only from the center-tap to the base of $Q1$. The current flowing into the base of $Q1$ causes the voltage $v_{B1}(\theta)$ to rise above the bias voltage V_{BB} . The transformer reflects this as a decrease in the voltage $v_{B2}(\theta)$ on the base of $Q2$, as shown in Fig. 12-8b. During the negative half-cycle of the driving current, the base of $Q2$ is similarly forward-biased and the base of $Q1$ is similarly reverse-biased. If stray reactances and transistor capacitances are ignored, only one base is in the circuit at a given time.

If a constant current gain β is assumed, each base requires a half-sinusoidal driving current of peak amplitude I_{cm}/β . The sinusoidal driving current thus has a peak amplitude $(I_{cm}/\beta)(n/m)$. At any instant, the driving voltage $v_D(\theta)$ is formed by the transformation by $T1$ of the voltage waveform on the forward-biased base. Although the waveform $v_D(\theta)$ is nonsinusoidal, it is convenient to assume that it is sinusoidal. Its peak amplitude is then $(m/n)(I_{cm}/\beta)R_B$, where R_B is the large-signal equivalent base resistance. This resistance may be defined as $(m/n)^2$ times the ratio of the fundamental frequency

Fig. 12-8 Drive of a Class B BJT PA. (a) Circuit; and (b) waveforms.



component of the driving voltage to the amplitude of the driving current. Values of R_B (or impedance Z_B) are usually given in transistor data sheets; typically, values range from one to several ohms. This value may be used to determine the required driving power

$$P_{DR} = \left(\frac{I_{cm}}{\beta} \right)^2 \frac{R_B}{2} \quad (12-25)$$

and the driving-point resistance R_D , which is $(m/n)^2 R_B$. The power gain can be computed as the ratio of the output and driving powers.

Driver amplifiers are typically Class A or B PAs, and predrivers are typically Class A PAs or small-signal amplifiers (Chapter 16). Ideally, the driver should be a current source (with infinite shunt resistance), since the base voltage to collector current relationship is nonlinear. However, shunt resistances and reactances inherent in the transistors and coupling networks often make achievement of ideal current source operation difficult at low power levels. For this reason, negative feedback, a shunt resistance (R_S in Fig. 12-8a) or hybrid couplers (Section 12-7) are often used in the input circuitry of a PA.

FETs. In Class A operation, the ideal square-law relationship between the gate-source voltage and the drain current introduces no distortion in the fundamental frequency signal [13]. However, in Class B operation, the ideal square-law characteristic results in a square-law (i.e., nonlinear) relationship between the amplitudes of the fundamental frequency components of the input and output signals. Fortunately, the VMOS power FETs [14] have characteristics that are square-law at low levels and linear at medium and high levels. The effects of the square-law region appear as crossover distortion and are minimized by proper choice of the bias voltage.

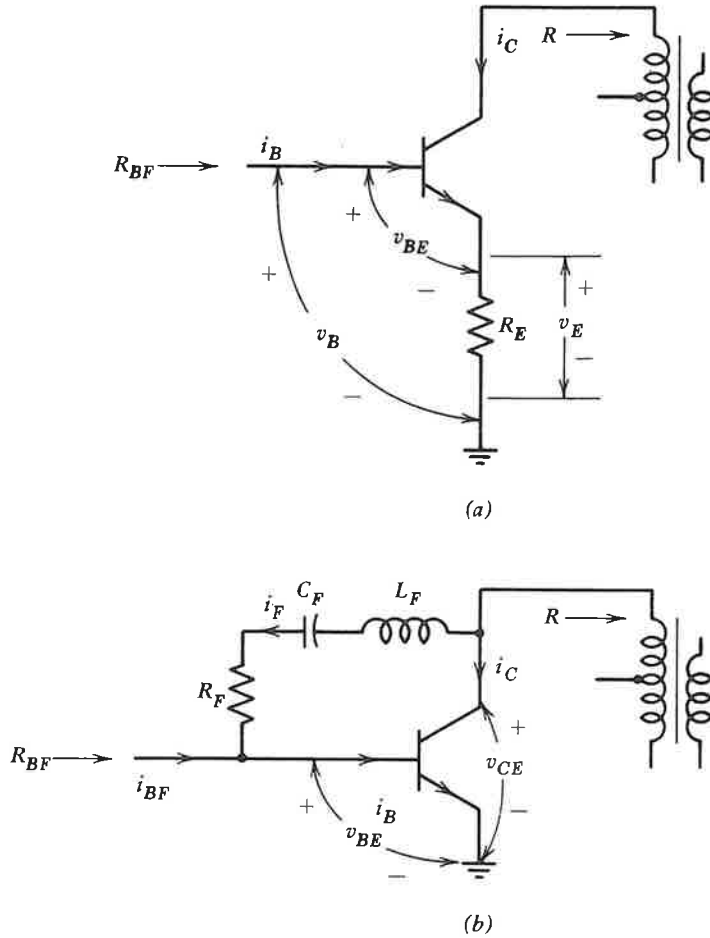
The drive of an FET PA must produce gate voltages corresponding to the desired drain currents. The input impedance of a MOSFET is primarily a large capacitive reactance with a small resistive component due to gate to drain to load coupling. Consequently, the input impedance of an FET PA is generally higher (and drive power is generally lower) than in a PA using comparable BJTs. It is important to note that current simultaneously flows in the gates of both the active and the cutoff FET. The input impedance of a single FET is therefore multiplied by $2(m/n)^2$ to obtain the driving impedance of a circuit similar to Fig. 12-8a. Because of the high, capacitive driving impedance, transformers used to couple the driving signal and bias voltage to the gates can be quite different [25] from those used in BJT PAs.

Emitter Feedback

An emitter resistor (Fig. 12-9a) can help to linearize the input voltage to output current transfer characteristic of a Class A or B power amplifier by suppressing the effects of variations in β with current, temperature, and frequency. Some emitter resistance ("ballast") is often included in RF power transistors to cause even division of current within the device (see Appendix 12-1) so that some emitter feedback is inherent in most PAs.

Insertion of resistance R_E in the emitter lead (Fig. 12-9a) produces an

Fig. 12-9 RF feedback. (a) Emitter feedback; and (b) collector feedback.



emitter voltage

$$v_E = (i_C + i_B)R_E = \left(1 + \frac{1}{\beta}\right)i_C R_E \approx i_C R_E \quad (12-26)$$

As long as β is large, it is apparent that the emitter voltage will be essentially proportional to the collector current, in spite of variations in β . By design, the variations in v_{BE} required to produce the collector current are small in comparison to v_E . Consequently, the base voltage v_B is simply $v_E + V_\gamma$ when the transistor is active. The driving voltage (applied to the base) and the collector current are thus forced to be proportional to each other.

Through a straightforward analysis, the driving impedance of the circuit of Fig. 12-9a can be shown (Problem 12-5.1) to be

$$R_{BF} = (\beta + 1)R_E + R_B \quad (12-27)$$

Similarly, the PA voltage gain can be shown to be R/R_E . These two expres-

sions are identical to those for small-signal amplification. The primary difference is that in a Class B PA, emitter feedback must act on the total instantaneous collector current rather than only its ac component. Shunting R_E with a capacitor can thus distort the current waveform.

Since R_{BF} depends directly upon β , it is often desirable to provide a fixed resistance shunt (R_S in Fig. 12-8a) for ac drive current. Also, it is important to note that emitter feedback has no effect if current (rather than voltage) drive is used. Source feedback in FET PAs is not generally practical because the large gate voltages produce large voltages on the source resistor and hence excessive power dissipation.

Collector Feedback

Another form of RF feedback that can be used with BJTs feeds a small portion of the collector current into the base. This is most easily accomplished with a simple resistor R_F (Fig. 12-9b) and a blocking capacitor C_F . If device gain falls off at higher frequencies, an inductor L_F may be added to decrease feedback action at those frequencies. (Effects of phase shifts due to L_F , the transistor, and the lead must be remembered in evaluating the PA's stability.) This form of feedback linearizes the amplifier's input current to output voltage transfer characteristic.

Since the base current is the sum of the drive and feedback currents, and since the feedback current is proportional to the output voltage and V_{be} is small in comparison with v_{CE}

$$\frac{i_C}{\beta} = i_B = i_{BF} + \frac{v_{CE}}{R} \quad (12-28)$$

For large values of β , the base current is very small, and the ac collector voltage is therefore forced to be proportional to the drive current. (Note that if the driver is a voltage source, collector feedback will not improve linearity.) For large values of β , the resultant equivalent current gain (i_C/i_{BF}) is approximately R_F/R , and the equivalent base input resistance is

$$R_{BF} = \frac{R_B}{1 + \beta R/R_F} \quad (12-29)$$

Other Forms of Feedback

The use of a transformer to couple a portion of the output current into the bases can be used to linearize a BJT amplifier's input current to output current transfer characteristic. Transformers may similarly be used with FET PAs to linearize the input voltage to output voltage characteristic. Envelope

feedback, discussed in Chapter 16, controls the driver (or a previous amplifier) to provide the drive amplitude that produces the desired output in spite of amplifier nonlinearities.

Example 12-5.1. Determine the drive requirements of the PA of Example 12-2.1 if $\beta = 20$ and $R_B = 4$ ohms. Then select a value of R_E that will result in an approximate voltage gain of 10. Further, determine the effects of the emitter resistance on amplifier efficiency and drive if β can vary from 20 to 40: First, with $\beta = 20$, each base requires a half sinewave of current whose peak amplitude is $2/20 = 0.1$ A. The drive power is then $P_{DR} = (0.1)^2 \times 4/2 = 20$ mW; hence the power gain is 31 dB. To get a voltage gain of 10, $R_E = R/10 = 1.25$ ohms. Efficiency is reduced according to the ratio of resistances; that is, now $\eta = 78.5 \times (12.5/13.75) = 71.4$ percent. The use of feedback increases the equivalent base resistance to $1.25(\beta + 1) + 4$, and hence values from 30.25 to 55.25 ohms are possible. The drive power required with emitter feedback is determined by first finding the current needed and then using the value of R_{BF} corresponding to a given current gain. The drive power varies from 151 to 69 mW, causing the power gain to vary from 22.2 to 25.6 dB.

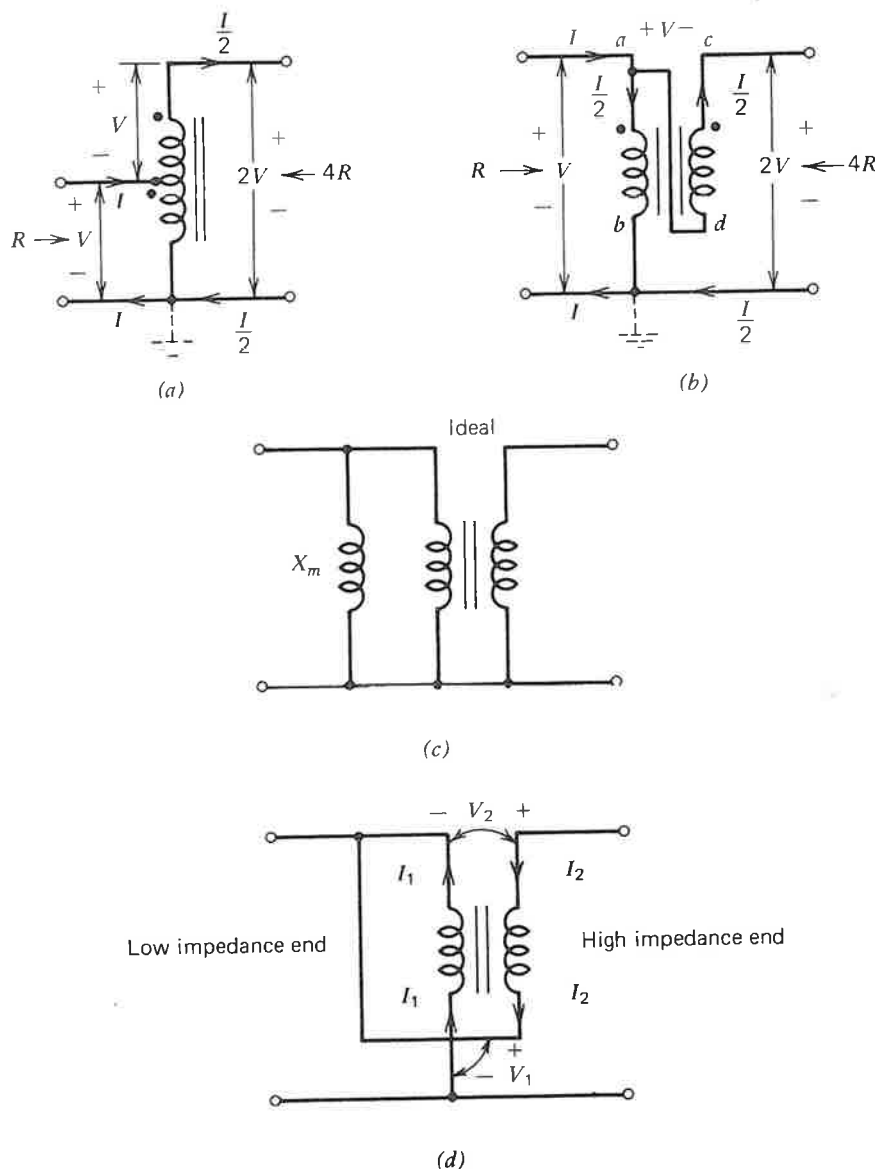
12-6 Wideband Transformers

Linear RF power amplifiers are often used in applications (such as HF SSB) that require operation over a large range of frequencies. Class B and D PAs generally require good transformer response to at least 3 times the output frequency, since the collector waveforms contain harmonics. Consequently, wideband transmission line transformers must be used rather than the conventional (colloquially called “wire-wound”) type of transformers, which have discrete windings.

The bandwidth limitations inherent in conventional transformers may be understood by considering the autotransformer of Fig. 12-10a, which has been tapped in the center to provide a 4:1 impedance transformation. Low-frequency performance is degraded by the shunt susceptance (magnetizing inductance) of the windings (Fig. 12-10c), while high-frequency performance is degraded by the series reactance (leakage inductance) of the windings [12, 16].

Transmission line transformers employ a simple but clever concept that allows improved high-frequency performance. Figure 12-10b shows the above-mentioned autotransformer rearranged into distinct primary and secondary windings. Since these remain connected, the 4:1 impedance transformation is preserved. Ideally, the current in each winding has the same magnitude, and the voltage across connections *a* and *c* is the same as the

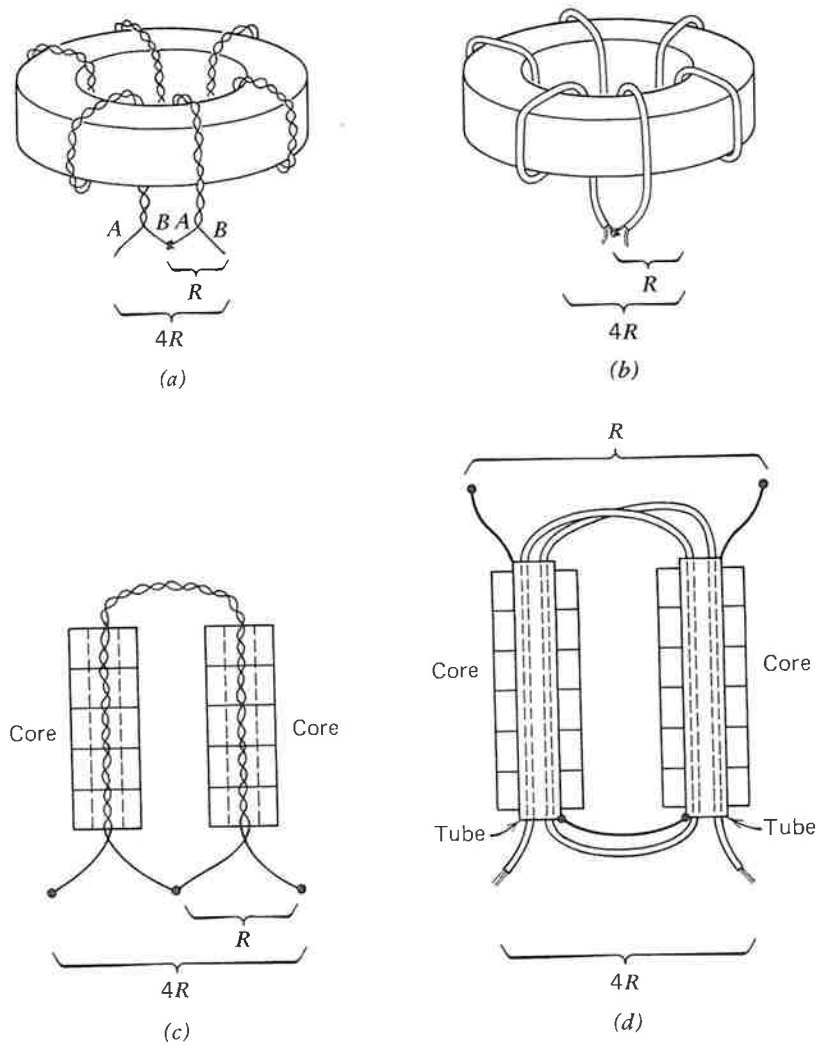
Fig. 12-10 Transmission-line transformer models. (a) Autotransformer; (b) transmission-line transformer; (c) low-frequency model; and (d) high-frequency model.



voltage across connections *b* and *d*. The two separate windings may therefore be replaced by the two conductors in a single transmission line as in Figs. 12-11a, b, or c.

At low frequencies, the shunt susceptance degrades performance, as with ordinary transformers. However, at high frequencies, the series reactance combines with the interwinding capacitance and the circuit behaves as a transmission line, greatly extending the high-frequency response.

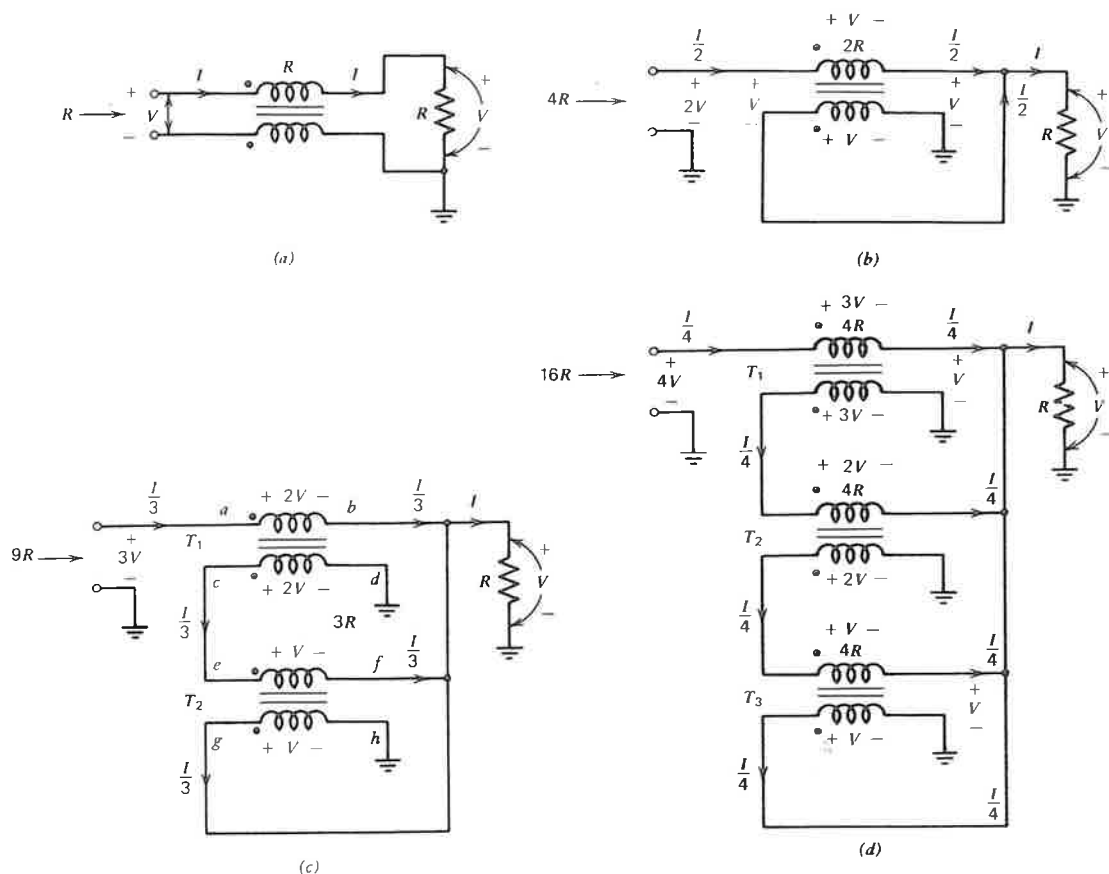
Fig. 12-11 Construction of transmission-line transformers. (a) Twisted wire; (b) coaxial cable; (c) stacked cores; and (d) brass tube.



Configurations

A variety of transmission line transformer configurations [17, 18] is depicted in Figs. 12-12 to 12-14. Easily recognized are the 1:1 (balun) and 4:1 configurations of Fig. 12-12a and b. Figure 12-12c and d shows how an $n^2:1$ impedance transformation can be obtained. A current $I/3$ driven into the upper winding of the upper transformer in Fig. 12-12c will cause a current of $I/3$ to flow in the lower winding of the upper transformer. This current then similarly produces a current $I/3$ in the lower winding of the lower transformer, resulting in a current I driven into the load R . A voltage V appearing on the load also appears across connections g and h of T_2 , causing it to appear across connections e and f as well. The voltage from e to f added to the

Fig. 12-12 Transmission-line transformer configurations. (a) Balun (1:1); (b) 4:1; (c) 9:1; and (d) 16:1.



voltage on the load produces 2 V across connections *c* and *d* of *T*₁. This similarly results in a voltage 3 V appearing at the input to the combination of *T*₁ and *T*₂. Figure 12-12*d* shows a similar arrangement that produces a 16:1 impedance transformation. Figure 12-13*a* shows an alternative to Fig. 12-12*d* that also produces a 16:1 impedance transformation. Note that in all cases, the characteristic impedance of the transmission line is specified by the voltage across the end of the line and the current flowing through the line.

Push-pull amplifiers such as Class B and Class D require slightly different arrangements. Figures 12-13*b* and *c* show how two or more transmission line transformers may be combined to enable push-pull operation (either for drive or output). As shown, the upper transistor (current source) is active, while the lower one is off and therefore forms an open circuit. In (*b*), a current *I* in the load results in a current *I* in the upper winding of *T*₂, and hence a current *I* in the lower winding of *T*₂ (whether or not the load is grounded). The current in the lower winding must also flow through the right-hand winding of *T*₁, which then produces a current *I* in the left-hand winding. The total current

Fig. 12-13 Transmission line transformer configurations. (a) Alternate 16:1; (b) push-pull 1:4; and (c) push-pull 1:9.

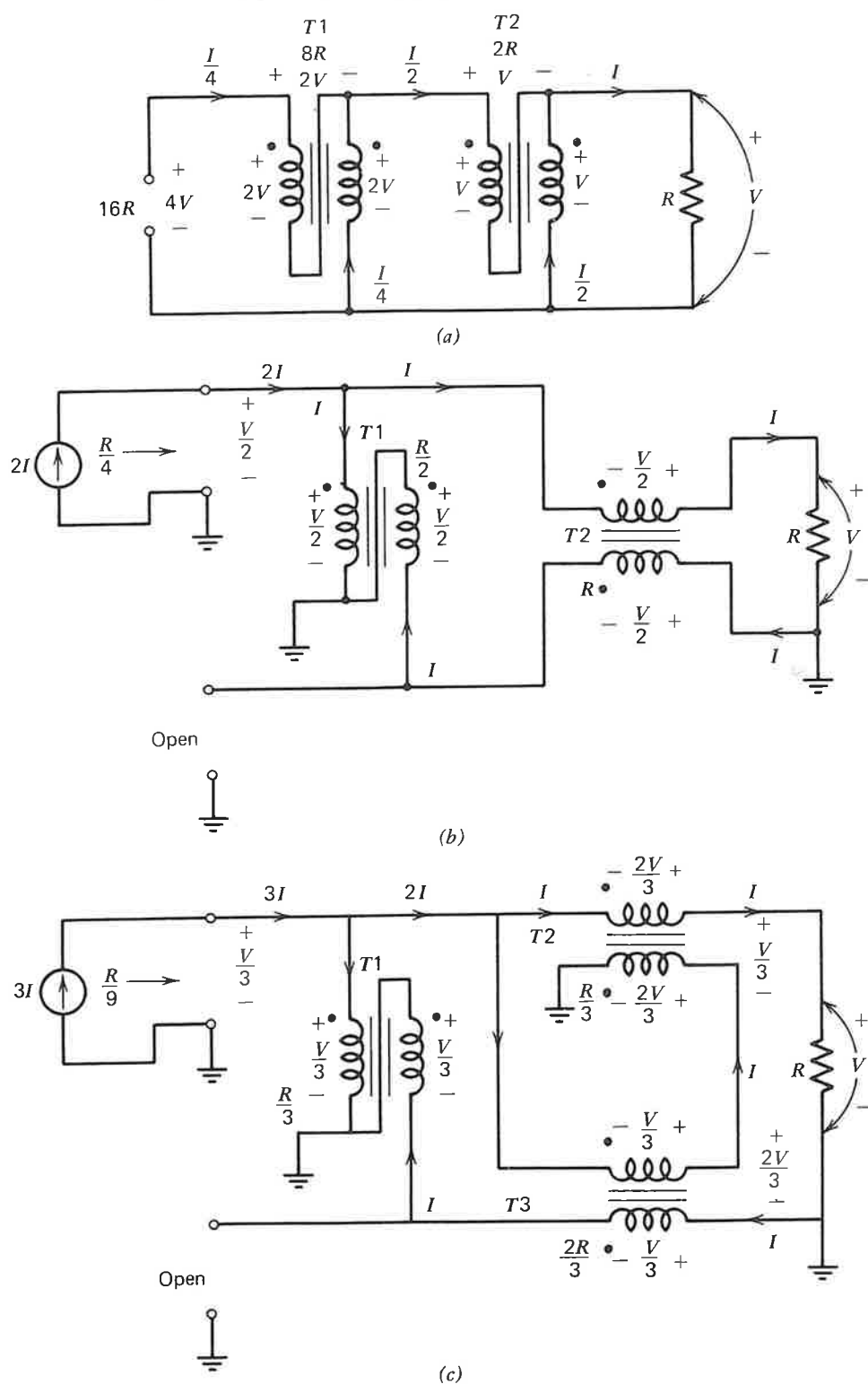
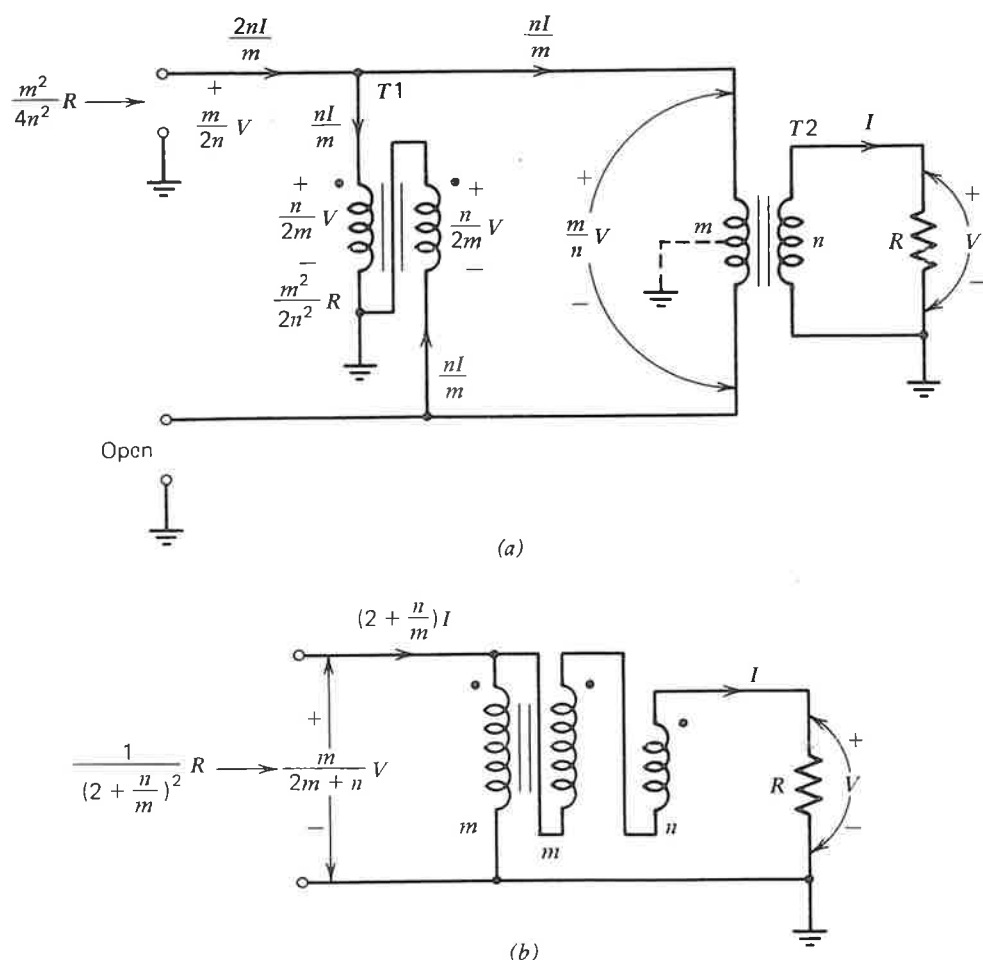


Fig. 12-14 Transmission line transformer configurations. (a) "Brass tube"; and (b) "Bootstrap."



supplied by the upper device is therefore $2I$. A voltage $V/2$ appearing on the device causes a voltage $V/2$ to appear across the right-hand winding of $T1$. This causes a voltage of $V/2$ to appear across $T2$, as shown, producing a voltage V on the load. Transformer $T1$ is commonly called a "hybrid," while $T2$ is commonly called a "balun." (Balun $T2$ could be connected as in Fig. 12-14a rather than as shown in Fig. 12-13b; however, the high-frequency response would be degraded.)

A commonly used version of the transmission line transformer employs a pair of brass tubes as a single-turn winding, as shown in Fig. 12-11d. Ferrite cores are stacked over the two tubes to provide the required minimum shunt reactance. The secondary winding is formed by wires running through the brass tubes. This arrangement may easily be visualized as a transmission line transformer when the secondary has only one turn. To visualize transmission line operation for other turns ratios, suppose that the secondary winding is

formed by winding and connecting two pieces of wire in parallel. The same voltage and current relationships are still present so that transmission line operation is preserved. Now suppose that the two pieces of wire are rearranged to form a two-turn secondary winding. The current flowing in one direction through a brass tube is equal to the sum of the currents flowing in the other direction in the two wires inside that tube. The voltage across each secondary winding is the same as in the 1:1 version. The transmission line mode of operation at high frequencies is therefore preserved, even though the secondary winding has been split into two (or more) turns. This type of transformer is often used in the arrangement shown in Fig. 12-14a, where an impedance transformation of $4n^2/m^2:1$ is obtained for push-pull operation. Note that either m or $n = 1$ if a “brass tube” transformer is used.

The transformer shown in Fig. 12-14b “piggybacks” a few extra turns onto a transmission line transformer to obtain a slightly different transformation ratio. Although high-frequency performance is not as good as with an ideal transmission line arrangement, this configuration can usually be used to provide up to a 6:1 transformation with only slight degradation [19,20]. The high-frequency response of a variety of other configurations may also be improved by forming some sort of transmission line with the windings.

Design Considerations

The first step in any design involving transmission line transformers is to select a configuration that gives the desired impedance transformation. The next step is to determine the characteristic impedances of the transmission lines needed (and to make sure that they are realistic). The performance of transformers in small-signal circuits is often given in terms of a 3-dB insertion loss [18]. However, since PAs act primarily as current (or voltage) sources, the insertion loss definition of bandwidth is not especially useful. More important to the design of PAs is the impedance seen by the PA (or devices) at different frequencies. As discussed in Section 12-3, any reactance appearing in the load of a Class A or B amplifier results in a decrease in efficiency and an increase in device stress.

Low-Frequency Response. As mentioned earlier, the low-frequency response of a transformer (transmission line or conventional) is degraded by a shunt susceptance that appears across the primary winding in Fig. 12-10c. This susceptance is due to the magnetizing inductance of the transformer [18], which is given by

$$L_m = \frac{X_m}{\omega} = 4\pi n^2 \mu \frac{A_e}{L_e} = n^2 A_L \quad (12-30)$$

where n is the number of turns in the primary winding, μ the core per-

meability, A_e the effective area of the core, and L_e the average magnetic path length. These effects are often combined in one quantity A_L , which is given (typically) in millihenrys per thousand turns squared.

High-Frequency Response. The designer may incorrectly assume that the upper-frequency limit of a transmission line transformer is determined by the quality of the transmission line used. However, the transformer does not employ the line as a transmission line, feeding power into one end and taking it from the other. The current is driven into only one (not both) of the conductors at a particular end of the transmission line, and the voltage is applied from end to end, rather than across one end. Compatibility with both ideal transformer and ideal transmission line operation is therefore only possible for frequencies at which the length of the line is a relatively small fraction of a wavelength (0.1 to 0.3 λ , maximum).

The voltages and currents (Fig. 12-10d) at the ends of a transmission line of characteristic impedance Z_o and length l are related by [21]

$$V_1 = V_2 \cos \gamma l + jI_2 Z_o \sin \gamma l \quad (12-31)$$

$$I_1 = I_2 \cos \gamma l + j \frac{V_2}{Z_o} \sin \gamma l \quad (12-32)$$

where γ is the propagation constant of the line ($\gamma = \omega/v$, where v is the phase velocity in the line). These equations can now be used to solve for the other voltages and currents in a circuit employing a transmission line transformer [22]. For the 4:1 configuration of Fig. 12-10d,

$$Z_{in}(\text{low end}) = Z_o \frac{Z_L \cos \gamma l + jZ_o \sin \gamma l}{2Z_o(1 + \cos \gamma l) + jZ_L \sin \gamma l} \quad (12-33)$$

with Z_L loading the high end and

$$Z_{in}(\text{high end}) = Z_o \frac{2Z_L(1 + \cos \gamma l) + jZ_o \sin \gamma l}{Z_o \cos \gamma l + jZ_L \sin \gamma l} \quad (12-34)$$

with Z_L loading the low end.

Transmission Line

Since deviations of the characteristic impedance of the transmission line from the desired value (or deviations of the load from its intended value) cause a degradation of the transformer's high-frequency response, careful selection of the line is important. Coaxial cable is desirable when it is available in the required impedance. Twisted wire [23] may be used to produce impedances between about 10 to 200 ohms. Thin wire and large spacing (insulation) produce a high characteristic impedance, while large wire, tight twisting, and

close spacing (enamel insulation) produce low characteristic impedances. It is also possible to connect two transmission lines in parallel to obtain a lower-impedance line. Empirical measurements made with the aid of a time-domain reflectometer or an RF impedance meter are most helpful in achieving the desired characteristics.

Flux Limits

The ferrite cores commonly used in RF transformers saturate at high flux densities. It is therefore necessary to design the transformers to keep the flux considerably below the saturation level. Failure to do so results in non-linearity, generation of harmonics, and inefficiency. At low frequencies the flux in the core is given by [23]

$$B = \frac{V_{\text{rms}} \times 10^8}{4.44fnA_e} \text{ gauss} \quad (12-35)$$

where V_{rms} is the rms voltage on the winding, n the number of turns in the primary, f the frequency, and A_e the effective area in square centimeters. At higher frequencies where the transmission line effects take over, the flux decreases below that given by (12-35). The maximum flux is thus determined by using the lowest frequency of operation in (12-35). The flux levels for core saturation and nonlinearity are usually given in the manufacturer's data sheets.

Example 12-6.1. Select broadband transformers for use with the PA of Example 12-2.1 and specify the characteristic impedance of the transmission lines needed. Match $R = 12.5$ ohms and $R_B = 4$ to 50 ohms: Since the output requires a 4:1 transformation, two transformers arranged as in Fig. 12-13b are used: $Z_o(T1) = 25$ ohms, and $Z_o(T2) = 50$ ohms. To transform 4 ohms to 50 ohms is awkward for a transmission-line configuration, but a 9:1 transformation will yield 36 ohms, which should be good enough. The configuration shown in Fig. 12-13c is therefore used to drive the bases, $Z_o(T1) = Z_o(T2) = 12$ ohms, and $Z_o(T3) = 24$ ohms.

12-7 Power Combiners and Splitters

When a larger output is needed than can be obtained from a single device or pair of devices, "hybrid coupler" [25] circuits can be used to combine two or more power amplifiers. Direct parallel operation of the devices is unsatisfactory because the currents do not divide equally among the devices.

Basic Hybrid Transformer

The basic hybrid transformer (Fig. 12-15a) can be used to combine the outputs of two PAs that are driven 180° out of phase with respect to each other. Shown here are two PAs operating as current sources and producing sinusoidal current outputs represented by I_1 and I_2 . Since the sum of the ampere-turns in an ideal transformer must be zero,

$$I_o = \frac{m}{n}(I_1 + I_2) \quad (12-36)$$

that is, the transformed sum of the two output currents is delivered to the load. The difference in the two output currents ($I_1 - I_2$) flows through the resistor R_h . (Note that the hybrid coupler technique is *not* used to combine the outputs of the two transistors in push-pull operation; to attempt this produces a different mode of amplification [26].)

When $I_1 = I_2$, no current flows in R_h , and $V_1 = V_2 = (m/n)V_o$, so it is apparent that both PAs deliver equal power to the load and each sees a load resistance of $R = (m/n)^2(2R_o)$. If $R_h = R/2$, both PAs see a constant load of value R even when $I_1 \neq I_2$ (see Problem 12-7.1), which is one advantage of the hybrid coupler.

The hybrid coupler is not, however, a cure-all for all possible problems encountered by a PA. If, for example, the load is shorted or opened, that effect is reflected to the collectors of each PA. Load reactances and variations are similarly transformed to the PAs.

Power Splitting

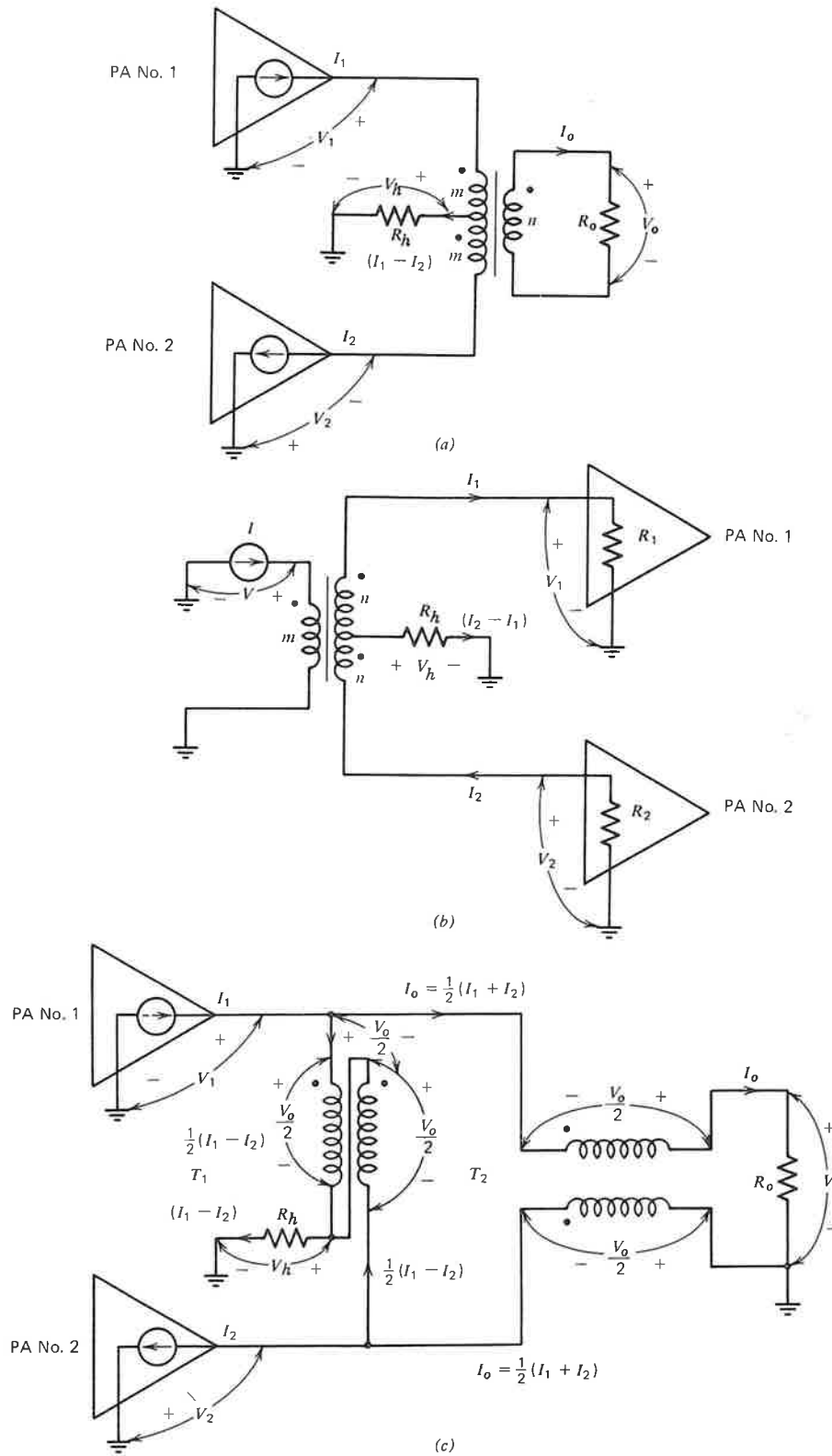
Figure 12-15b shows the basic hybrid transformer splitting the power from a current source driver into two loads, which might be the inputs of the PAs of Fig. 12-15a. If $R_1 = R_2$, the power and current are, of course, evenly divided. In general, however,

$$\frac{I_1}{I} = \frac{m}{n} \left(\frac{2R_h + R_2}{4R_h + R_1 + R_2} \right) \quad (12-37)$$

and the current is not equally split when $R_1 \neq R_2$, for any value of R_h .

To obtain equal drive currents in both PAs in spite of base resistance variations, the center-tap of the transformer may simply be omitted ($R_h = \infty$). Failure (open circuit input) of one PA will shut off the drive current to the other. This is not undesirable, since the resistor in the output coupler would have to dissipate half of the output of one of the PAs, and the net efficiency would be cut in half as well.

Fig. 12-15 Hybrid couplers. (a) Basic hybrid combiner; (b) hybrid splitter; and (c) broadband hybrid combiner.



Broadband Hybrids

The broadband transmission line transformer techniques discussed in the previous sections can also be used to make broadband hybrid couplers. For example, the previously discussed basic hybrid transformer is often made using brass tubes, as in Fig. 12-11*d*.

Figure 12-15*c* shows how a pair of transmission line transformers may be combined to make a hybrid coupler (commonly called a “180° hybrid”). The actual hybrid coupler is $T1$, and $T2$ is a balun. Using the same analysis as was used for the PA hybrid combiner, one can show that the load gets the sum of the driving currents (with $m = 2n$), while the difference in the driving currents flows through R_h . Now, however, $R = R_o/2$ (for $I_1 = I_2$), and $R_h = R_o/4$ to maintain a constant PA load resistance. Other aspects are similar to those of the basic hybrid coupler.

Other Hybrids

A variety of types of hybrid couplers exists. Reversal of the phase of I_2 in any of the previously discussed couplers will dump all of the power into R_h ; R_o then receives that power due to differences between I_1 and I_2 . Four PAs can be combined through a set of three splitters and three combiners. However, it is also possible [25] to make n -way hybrids directly. An example of the use of such a coupler is a 1-kW HF linear PA made by combining four smaller PAs [27].

Example 12-7.1. Design hybrid couplers to combine two of the PAs from Example 12-2.1 to get a total output of 50 W. Assume that $R_B = 4$ ohms and that I_1 and I_2 are within 10 percent of each other. Match (approximately) to 50-ohm input and output impedances: To get a 50-ohm load for the PAs requires that $n/m = \sqrt{2}$ in the circuit of Fig. 12-15*a*. A more practical choice is $m = 2$, $n = 3$, which gives $R = 44$ ohms. This will produce an 11-ohm load on the collectors, so slightly more current will be required to produce 25 W per PA. Since $R = 44$ ohms, $R_h = 22$ ohms. Assuming that the current difference is the maximum 10 percent, the power in R_h will be $(0.2)^2 \times 22/2 = 440$ mW. Since the PAs are current driven, the center-tap of the input hybrid will be omitted. The 4:1 transformer in the example gives a 64-ohm input to each PA. A splitter transformer with $m = 2$ and $n = 3$ will give a 56.9-ohm input, which is close to the desired 50 ohms.

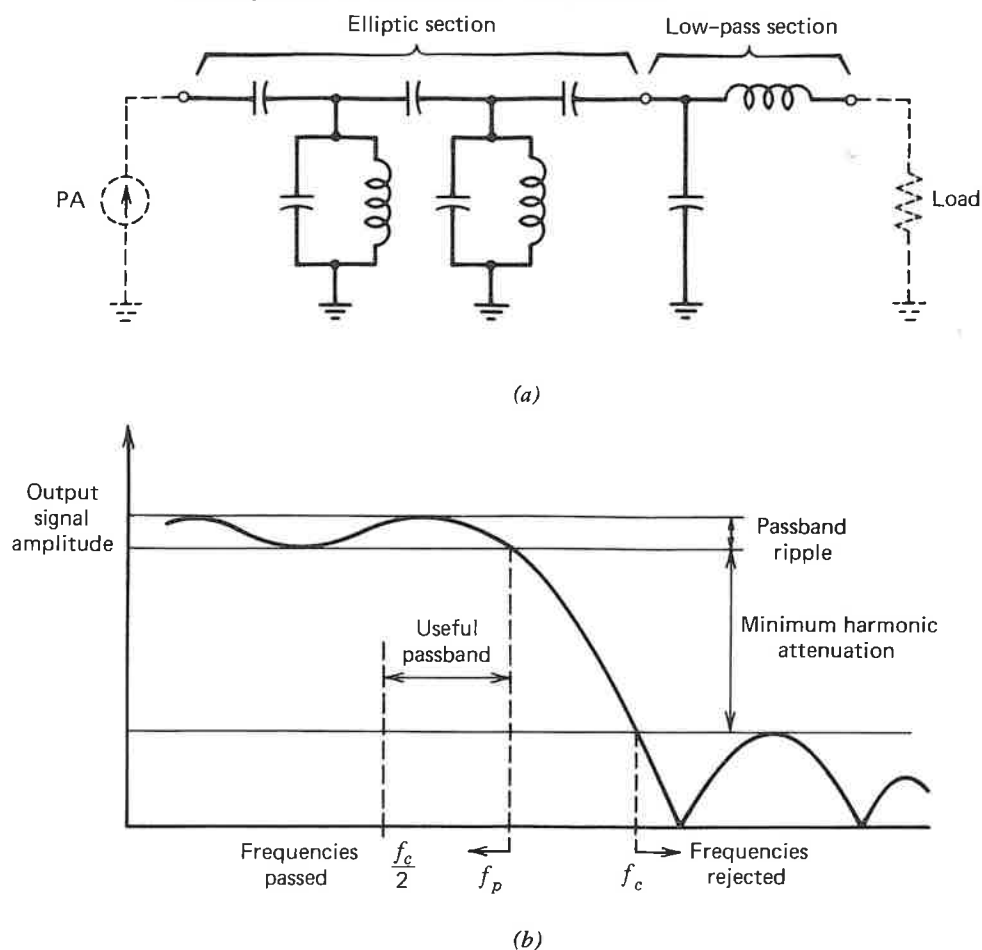
12-8 Output Filters

Real power amplifiers produce harmonic currents that must be filtered out if the PA is connected to an antenna. Crossover distortion is primarily respon-

sible for the odd harmonics, which are typically 15 to 20 dB below the signal level. Current-gain variations are responsible for the even harmonics, which are typically 20 to 40 dB below the signal. Typical specifications require the harmonics to be attenuated to at least 60 dB below the carrier.

Since this is not a text on filter theory, the discussion will be limited to those topics particularly applicable to modern power amplifiers. Along with the introduction of the broadband transformer came the use of a bank of filters rather than the tuned “tank” circuit that was common to vacuum-tube transmitters. The output filter must, for all frequencies for which it is to be used, pass the fundamental frequency while rejecting (bypassing) the harmonics; this limits the bandwidth of a given filter to a maximum of 1 octave (practical filters usually have a half-octave bandwidth). Since (assuming proper operation of the PA) there are no signals below the carrier frequency, low-pass filters are used instead of bandpass filters.

Fig. 12-16 Elliptic filter for a power amplifier. (a) Example of a circuit; and (b) example of a frequency response.



In principle, the filter response could be based on Butterworth (maximally flat) or Chebychev (equal ripple) characteristics. Because both a large passband and a large harmonic attenuation are normally desired, an elliptic (or Cauer or m -derived) section is normally used in conjunction with a low-pass section. The elliptic section of the filter of Fig. 12-16a has both the passband ripple (as in a Chebychev characteristic) and zeros in its response that give it a very rapid roll-off. An increasing attenuation at higher frequencies is ensured by the low-pass section.

Available references [29, 30] can be used to design a filter with specified characteristics. However, the designer must be aware [31] that the input to the filter is a high-impedance current source (if it is a Class A or B amplifier without feedback). The first element of the filter must therefore provide a path for the harmonic currents; a series inductor will have no effect on filter response (but will cause the PA to saturate at a lower output). Variations in load impedance can also alter filter response. Many filter design manuals and computer analysis programs are based on a noninfinite source impedance and purely resistive loads, and are therefore nearly useless for determining the response of the output filter of a PA.

Example 12-8.1. Specify the passbands (i.e., the cutoff frequency and the lowest usable frequency) for a set of output filters for 2 to 30 MHz operation. Assume that in the prototype design, the maximum acceptable passband attenuation occurs at f_p , which equals three-fourths the cutoff frequency f_c at which the minimum acceptable harmonic attenuation is obtained: Adequate harmonic rejection requires that no frequency lower than half the cutoff frequency be used. The ratio of the upper and lower passband frequencies is thus $0.75/0.5 = 1.5$. The number n of filters required then satisfies $2 \times (1.5)^n \geq 30$ MHz, and hence $n = 7$ is selected. The ratio r of passband frequencies then satisfies $2r^7 = 30$ so that $r = 1.47$. Filter passbands are then 2–3, 2.94–4.42, 4.34–6.5, 6.38–9.58, 9.4–14.1, 13.84–20.76, and 20.38–30.56 MHz. The lower frequencies were obtained by successive multiplication of 2 MHz by 1.47. The upper frequencies were obtained by multiplication of the lower frequencies by 1.5.

12.9 Heatsink Design

The reliability of transistors decreases with increasing junction temperatures. Since PAs in general dissipate a significant amount of power, it is important to provide them with a heatsink to dispose of the unwanted heat. In Class A operation with a resistive load, the maximum power dissipation is twice the maximum output power. In Class B operation with a resistive load, the

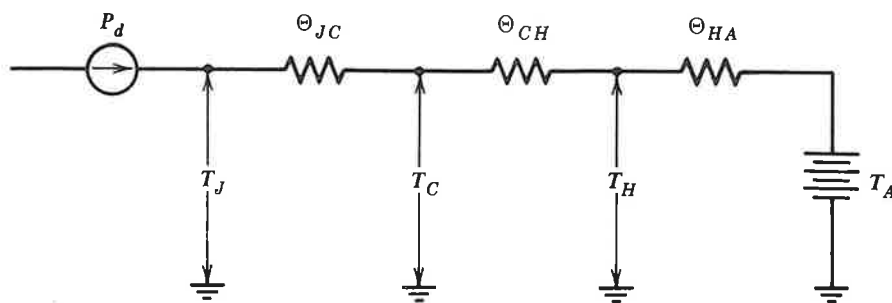
maximum total power dissipation is $2 V_{CC}^2 / \pi^2 R$, about 40.5 percent of the peak output power. In either Class A or Class B operation, the presence of reactance in the load will increase the maximum dissipation above these values, as discussed in Section 12-3.

Heat flow can be analyzed with a thermal circuit in the same way that current flow can be analyzed with an electrical circuit. In Fig. 12-17, the dissipated power is analogous to a current-source driver, temperatures are analogous to voltages, and each element in the thermal path has a thermal resistance Θ associated with it. Capacitors can also be added [9, 18] to model time constants due to heat storage.

The first element in the heat flow path is the junction-to-case thermal resistance Θ_{JC} , which is typically in the range of 1 to $10^\circ\text{C}/\text{W}$. The next element is the thermal resistance Θ_{CH} associated with mounting the device on the heatsink. This is typically a few tenths of a $^\circ\text{C}/\text{W}$, and is minimized by the application of a silicon thermal grease and the correct mounting torque [18]. (The case-to-air thermal resistance often quoted for small transistors is usually so high that it may be ignored when dealing with large power transistors.) The last element in the circuit shown is the heatsink-to-ambient thermal resistance Θ_{HA} . This resistance may include two parallel components corresponding to convection and radiation, as well as a series component corresponding to conduction from the mounting site to the surfaces. The convection component is usually nonlinear, but fortunately decreases with increasing temperature.

Heatsinks used in PAs are usually extruded-aluminum finned structures with a high area-to-volume ratio. The size of the heatsink is determined by comparing its required maximum thermal resistance with the curves given by the manufacturer. The maximum allowable thermal resistance is determined from the maximum allowable junction temperature (typically 200°C), the maximum ambient temperature, the amount of power dissipated, and the other thermal resistances in the circuit.

Fig. 12-17 Thermal circuit.



Example 12-9.1. Determine the maximum allowable thermal resistance of the heatsink for the PA of Example 12-2.1 if $P_d = 8 \text{ W}$, $\Theta_{JC} = 3^\circ\text{C/W}$, and $\Theta_{CH} = 0.3^\circ\text{C/W}$; T_J must be less than 200°C ; and the ambient temperature may be as high as 100°C . Assume a purely resistive load: The minimum junction-to-ambient temperature difference is 100°C , and 8 W must be dissipated; hence the total thermal resistance in the circuit can be no more than $(100)/8 = 12.5^\circ\text{C/W}$. Since the junction-to-case and case-to-heatsink resistances already total 3.3°C/W , the heatsink can have no more than 9.2°C/W .

REFERENCES

1. H. Granberg, "A Complementary Symmetry Amplifier for 2 to 30 MHz SSB Driver Applications," EB-32, Motorola Semiconductor Products, Inc., Phoenix, Ariz., Feb. 1975.
2. D. V. Jones and R. F. Shea, *Transistor Audio Amplifiers*, John Wiley & Sons, Inc., New York, 1968.
3. R. D. Peden, "Charge-Driven HF Transistor-Tuned Power Amplifier," *IEEE Jour. Solid State Circ.*, SC-5, pp. 55-63, Apr. 1970.
4. H. Granberg, "Measuring the Intermodulation Distortion of Linear Amplifiers," EB-38, Motorola Semiconductor Products, Inc., Phoenix, Ariz., Jan. 1975.
5. E. W. Pappenfus, W. B. Bruene, and E. O. Schoenike, *Single Sideband Principles and Circuits*, McGraw-Hill Book Company, New York, 1964.
6. "White Noise Test Evaluation of Power Tubes for Linear Amplifiers," Application Note 4530, RCA Electronic Components, Harrison, N.J., Jan. 1971.
7. L. Leighton, "Two-Meter Transverter Using Power FETs," *Ham Radio*, 9, pp. 10-15, Sept. 1976.
8. J. Millman and H. Taub, *Pulse, Digital, and Switching Waveforms*, McGraw-Hill Book Company, New York, 1965.
9. *Solid-State Power Circuits*, SP-52, RCA Solid-State Division, Sommerville, N.J., 1971.
10. H. Granberg, "Get 300 Watts PEP Linear Across 2 to 30 MHz from this Push-Pull Amplifier," EB-27, Motorola Semiconductor Products, Inc., Phoenix, Ariz., Sept. 1974.
11. B. Utter, "Z0-28: A Bias Device for High Power Amplifiers," Application Note 2.2.8.2, Communications Transistor Corporation, San Carlos, Calif., 1976.
12. P. E. Gray and C. L. Searle, *Electronic Principles: Physics, Models, and Circuits*, John Wiley & Sons, Inc., New York, 1969.
13. K. K. Clarke and D. T. Hess, *Communication Circuits: Analysis and Design*, Addison-Wesley, Reading, Mass., 1971.
14. A. D. Evans, D. Hoffman, E. S. Oxner, W. Heinzer, and L. Shaeffer, "Higher Power Ratings Extend V-MOS FETs' Dominion," *Electronics*, 51, 105-112, June 22, 1978.
15. F. H. Raab, "MOSFET Power Amplifier for Operation from 160 Through 6 Meters," *Ham Radio*, 11, 12-27, Nov. 1978.

16. W. L. Cassell, *Linear Electric Circuits*, John Wiley & Sons, Inc., New York, 1964.
17. H. Granberg, "Broadband Transformers and Power Combining Techniques for RF," Applications Note AN-479, Motorola Semiconductor Products, Inc., Phoenix, Ariz., June 1975.
18. J. Johnson, Editor, *Solid Circuits: Circuit Techniques*, Applications Book 2.2.8.0A, Communications Transistor Corporation, San Carlos, Calif., Mar. 1973.
19. H. L. Krauss and C. W. Allen, "Designing Toroidal Transformers to Optimize Wideband Performance," *Electronics*, **46**, 113–116, Aug. 16, 1973.
20. J. J. Nagle, "Use Wideband Autotransformers in RF Systems," *Electronic Design*, **24**, 64–70, Feb. 2, 1976.
21. R. G. Brown, R. A. Sharpe, W. L. Hughes, and R. E. Post, *Lines, Waves, and Antennas*, 2nd ed., The Ronald Press Company, New York, 1973.
22. C. L. Ruthroff, "Some Broad-Band Transformers," *Proceedings of the IRE*, **47**, pp. 1337–1342, Aug. 1959.
23. P. Lefferson, "Twisted Wire Transmission Line," *IEEE Trans. on Parts, Hybrids, and Packaging*, **PHP-7**, 148–154, Dec. 1971.
24. *Magnetic Design Manual*, Bulletin 550 1/71, Ferroxcube Corporation, Saugerties, N.Y., 1971.
25. E. F. Sartori, "Hybrid Transformers," *IEEE Trans. on Parts, Materials, and Packaging*, **PMP-4**, 59–66, Sept. 1968.
26. F. H. Raab, "The Class BD High-Efficiency RF Power Amplifier," *IEEE Jour. of Solid-State Circ.*, **SC-12**, 291–298, June 1977.
27. S. Chambers, "A 1000-W Solid-State Amplifier?," *Electronic Design*, **22**, 58–62, Apr. 1, 1974.
28. C. A. Holt, *Electronic Circuits*, John Wiley & Sons, Inc., New York, 1978.
29. P. R. Geffe, *Simplified Modern Filter Design*, John F. Rider, Inc., New York, 1963.
30. A. Zverev, *Handbook of Filter Synthesis*, John Wiley & Sons, Inc., New York, 1970.
31. V. B. Kozyrev, "Special Factors in the Design of Filters for the Output Stages of Radio Transmitters," *Telecommunications and Radio Engineering*, Part 2, **30/31**, 127–133, Feb. 1976.

PROBLEMS

12-1.1. Design a single-ended Class A PA to deliver 25 W to a 50-ohm load. Assume that $V_{CC} = 28$ V and $f = 2$ MHz. Specify the performance. *Hint.* Use a transformer to convert 50 ohms into a more suitable load resistance. Note that for RF power transformers, the number of turns in a given winding is a small integer (usually ≤ 5) so that an exact "match" may not be possible.

12-2.1. Design a Class B PA to deliver 100 W to a 50-ohm load. Assume that $V_{CC} = 28$ V and $f = 2$ MHz. Specify the performance.

12-2.2. Design a complementary version of the Class B PA of the previous problem. Select V_{CC} for maximum efficiency. Specify the ratings and performance.

12-2.3. Suppose that the load of a Class B amplifier is a parallel pair of diodes of opposite polarity (representing an unbiased pair of emitter-base junctions, the input to a Class D PA). The secondary of T_1 in Fig. 12-4a is replaced with $2n$ turns whose center-tap is grounded. The ends of this winding are connected to the anodes of the two diodes; the other ends of the diodes are grounded. Determine the power input, output, and dissipation for sinusoidal current drive, and the best choice for (m/n) . Assume that the diodes have an infinite reverse resistance and a zero forward resistance with voltage drop V_γ .

12-3.1. Assume that a battery voltage (V_{DD}) can vary from 24 to 28 V during its useful life. Design a 27 MHz Class B PA that will deliver 16 W PEP to a 50-ohm load over this entire supply voltage range. Use FETs with $R_{on} = 2.5$ ohm and specify the performance with both high and low values of V_{DD} .

12-3.2. Verify (12-21), (12-22), (12-23), and (12-24) for a Class B PA.

12-3.3. Consider the PA designed in Problem 12-2.1. Determine the minimum and maximum load resistance R_o (with no load reactance) for which the PA can deliver a full 100-W output. Also determine the minimum value of shunt reactance and maximum value of series reactance (with a 50-ohm resistance) for which the PA can deliver a 100-W output. Specify the performance in each of these four cases. Assume that $V_{sat} = 1$ V and that the transistors have 9-A peak current ratings.

12-4.1. Design a bias supply for the FET PA of Problem 12-3.1 if $V_{GS} = 2$ V gives the minimum IMD. Use a Zener diode with $V_Z = 3.3$ V and $I_Z = 10$ mA.

12-5.1. Derive (12-28) and the expression for the voltage gain of a PA with emitter feedback. Assume, when necessary, that $i_C = \beta(v_{BE} - V_\gamma) = i_B/R_B$, where β and R_B may vary.

12-5.2. Design an RF feedback system using transformers to accomplish current to current linearization, and describe its characteristics (gain, input impedance, etc.).

12-6.1. Design a network of transmission line transformers to allow operation of the PA of Problem 12-2.1 over the entire 2 to 30 MHz range. Assume that $R_B = 1.4$ ohms and that the driving impedance is to be approximately 50 ohms (unbalanced). Use the results of Problem 12-3.2 to establish performance criteria for the transformer nearest the collectors. Then use these criteria to design the transformer using a core with the following characteristics: OD = 3.6 cm, thickness = 1.5 cm, $A_L = 3639$ mH/(1000 turns)², $A_e = 0.977$ cm², and $B_{sat} = 3000$ gauss.

12-7.1. Show that the load presented to PA_1 in Fig. 12-15a is R for any value of I_2 when $R_h = R/2$. [Remember that $R = (m/n)^2(2R_o)$.]

12-7.2. Derive (12-38).

12-7.3. Design a hybrid power-splitting and combining network using transmission line transformers to combine four of the PAs of Problem 12-6.1 into an amplifier with a 400-W output. The transformers specified in Problem 12-6.1 may be modified if necessary.

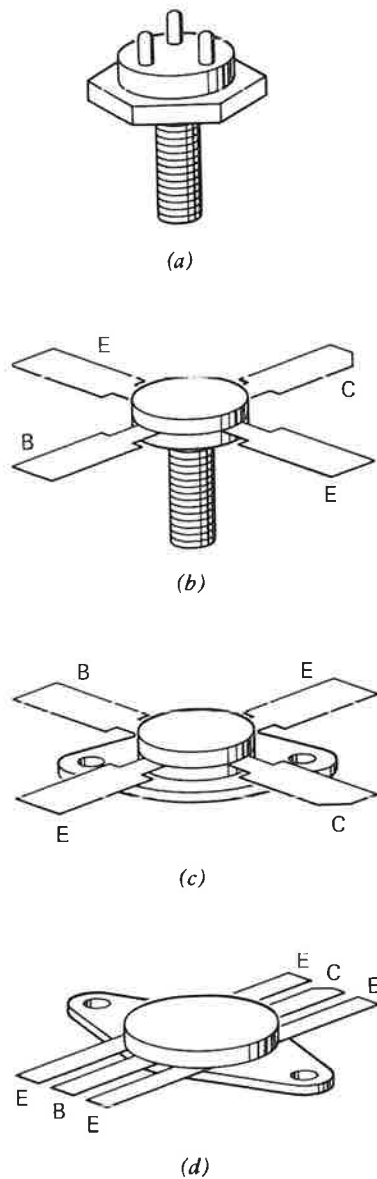
Appendix 12-1 RF Power Transistors

Detailed discussions of semiconductor theory and RF power transistor construction can be found in the literature [9]. Three aspects that are of special interest to the designer are discussed in this appendix.

Packaging

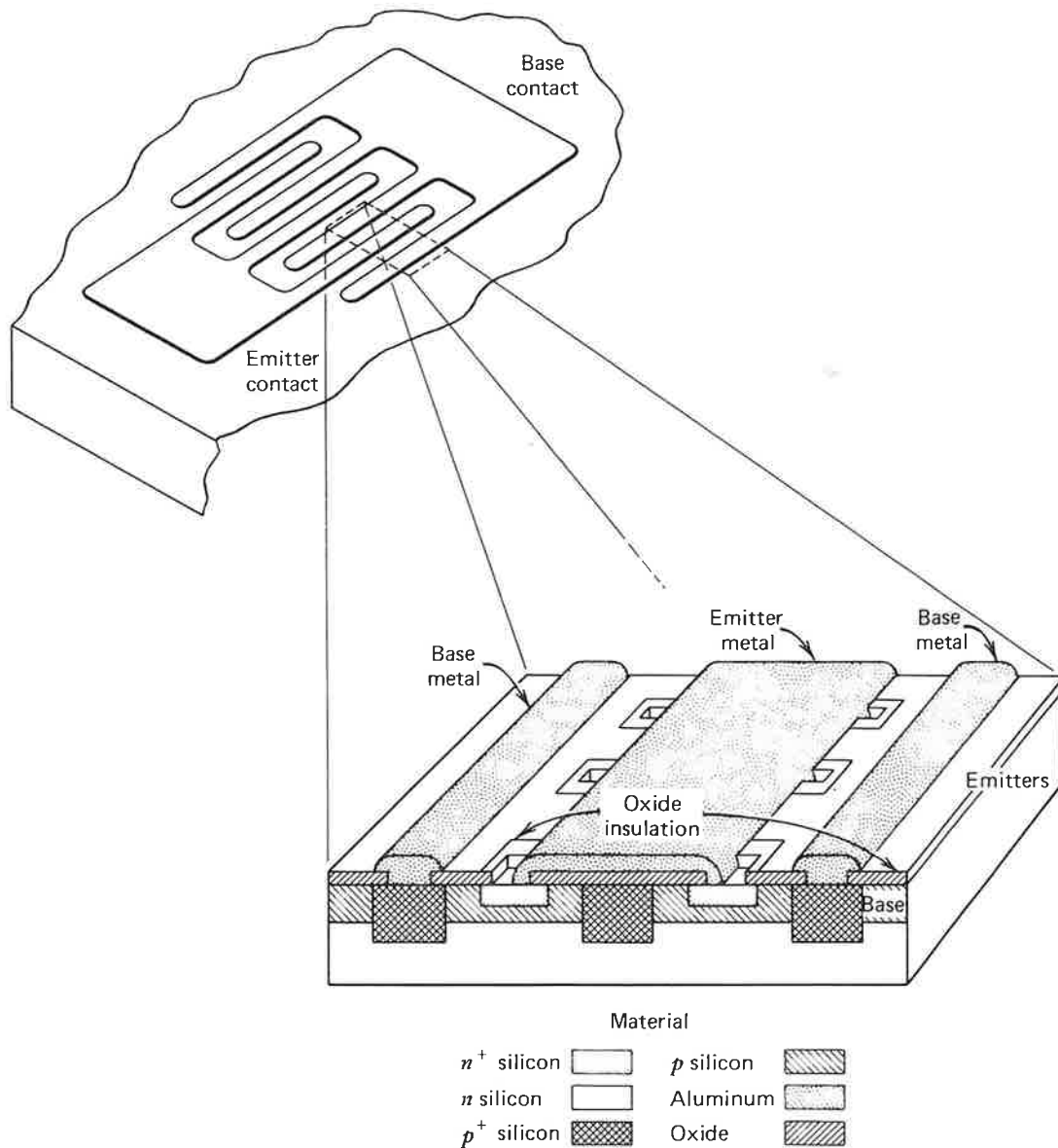
There are three basic case styles commonly used with RF power transistors, and many variations of each style. Figure 12-18a shows a simple stud-mounted package, commonly used on older RF power transistors. The stud provides for mounting with low thermal resistance.

Fig. 12-18 RF power transistor package types. (a) Stud-mounted; (b) Stud-mounted SOE; (c) flange-mounted SOE; and (d) flange-mounted improved stripline.



Figures 12-18*b* and 12-18*c* show “stripline opposed emitter” (SOE) packages with low-inductance “ribbon” leads that connect easily to PC board wiring. The emitter is sometimes connected to the case, allowing the two emitter leads to be omitted. A problem sometimes encountered with packages (a) and (b) is that excessive mounting torque can break the mounting stud, ruining the transistor. The flange-mounted package (c) overcomes this problem by providing two holes for screws and a large, flat surface for low thermal

FIG. 12-19 Construction of a bipolar RF power transistor. (From *Solid State Power Circuits*, SP-52, 1971. Courtesy of RCA Solid State Division, Somerville, N.J.).



Construction of Bipolar RF Power Transistors

The interdigitated construction allows for even distribution of the base current, a smaller distance from the base to the emitters, and lower contact resistances. The use of many small emitters makes possible local "hotspotting," which can lead to a thermal runaway chain reaction. To prevent this, resistance ("emitter ballasting") is often introduced in each emitter lead to force the current to divide evenly among the emitters.

The recently developed VMOS technique [14,28] has made possible FETs suitable for RF power applications. In a VMOS FET, the current flows vertically (hence the name *Vertical MOS*) rather than horizontally as in an

A cross-sectional diagram of a MOSFET structure. The diagram shows a central channel region labeled $N+$ (N+ body) flanked by P (body) regions. Above the channel is a SiO_2 gate layer. The source and drain regions are labeled P and $N+$ respectively. The top layer is labeled $Aluminum$. The substrate is labeled $N+ substrate$. The diagram illustrates the physical structure of the device, including the gate, source, drain, and substrate layers.

ordinary FET. Figure 12-20 depicts a cross section of a VMOS FET, showing the V-groove structure not present in conventional FETs. A positive gate (and body) potential with respect to the source creates an electric field that induces an *N*-type channel on both faces of the body adjacent to the gate. This completes the circuit from the *N*+ source through the *N*-channel and *N*-epi-layer, to the *N*+ substrate.

VMOS devices are characterized by high current density and a consequent low gate to drain capacitance. A low output conductivity results in nearly ideal current-source operation. At low current levels, the square-law gate voltage to drain current relationship holds. However, above a certain level, increased field strength does not increase the drift velocity. Because of this "velocity saturation" effect, the transfer characteristic for medium and high current levels is linear.

FETs for RF power applications had just been introduced when this was written. However, FETs offer several significant advantages over BJTs, including lower drive power and absence of storage time. Also, since there is a negative temperature characteristic, FETs are not subject to thermal runaway, current hogging, hotspotting, and second breakdown. In switching-mode operation (discussed in Chapter 14), VMOS FETs can pass drain current in either direction (i.e., they are *bilateral*); BJTs are, for the most part, *unilateral*.

This result is readily extended to any number of amplifiers, and its use is best shown by an example. See also Fig. 16-4.

Example 16-2.1. A transmitter is required to have an IMD ratio of -30 dB or better. The output of a mixer (assumed to be distortionless) will be amplified by three stages. Determine the IMD ratio required of each stage of amplification: An IMD ratio of -30 dB or better is equivalent to $d_{\text{chain}} \leq 1/31.62 \cong d_1 + d_2 + d_3$. If equal distortion in each stage is assumed, then $d_1 = d_2 = d_3 \leq d_{\text{chain}}/3$ is sufficient; hence, each stage must have an IMD ratio of -39.5 dB or better.

16-3 Peak and Average Power

The characteristics of the signal to be transmitted determine the transmitter peak-envelope power and average input and output requirements. The safest design allows for the worst-case power dissipation ($2.0 P_{oPEP}$ and $0.405 P_{oPEP}$ for Class A and Class B stages, respectively) and for input power corresponding to continuous power at the peak-envelope power level. However, accurate prediction of the input power requirements for portable and high-power transmitters requires some knowledge or assumptions about the signal characteristics. In addition, some low-cost transmitters reduce heatsink and power supply requirements to correspond to the expected usage rather than the worst-case-possible conditions.

Peak-to-Average Ratio

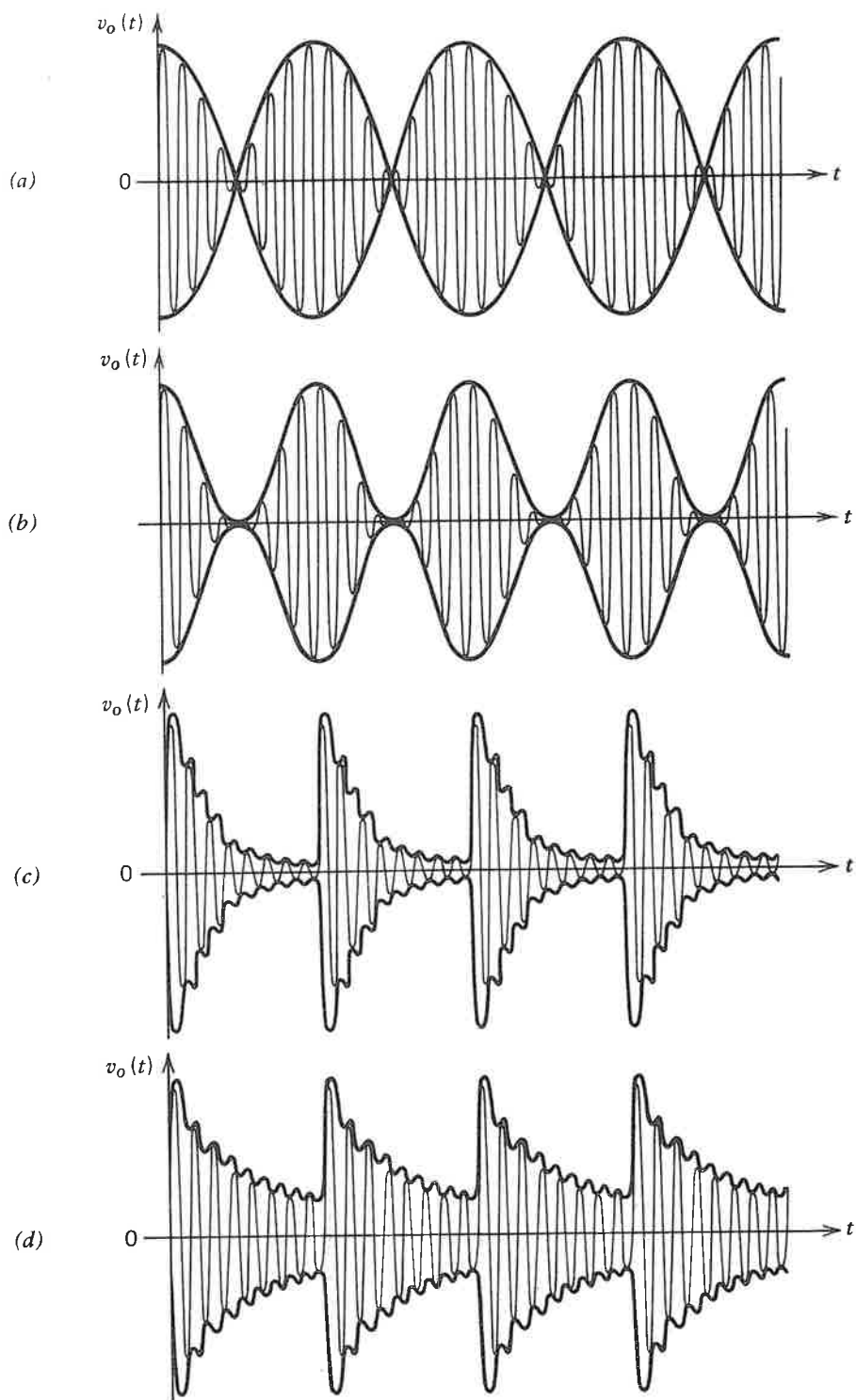
The commonly used term *peak-to-average* ratio is defined by

$$\xi = \frac{P_{oPEP}}{P_{oAVG}} \quad (16-3)$$

This is an important parameter of the signal being transmitted, since amplifiers must be designed to produce the peak-envelope-power output P_{oPEP} , but only deliver the average power output P_{oAVG} . For frequency-modulated signals, including frequency-shift-keyed radioteletype, $\xi = 1$. Since the carrier of a CW signal is intermittently switched on and off, the peak-to-average ratio depends upon the on-to-off ratio; $\xi = 2$ is often used for design purposes. The discussion in Section 15-3 determined that for single-tone, full-carrier amplitude modulation with $m_a = 1$ (Fig. 16-5b), $P_{oPEP} = 4P_{oCAR}$ and $P_{oAVG} = (\frac{2}{3})P_{oCAR}$; hence $\xi = \frac{8}{3}$ or 4.3 dB.

The *two-tone* signal commonly used to test SSB equipment consists of two tones of equal amplitudes and produces a spectrum equivalent to that of

Fig. 16-5 Signal envelopes, (a) DSB/SC one tone or SSB/SC two tone $\xi = 3$ dB; (b) AM one tone $\xi = 4.3$ dB; (c) SSB/SC speech $\xi \approx 16$ dB; and (d) SSB/SC speech $\xi \approx 10$ dB.



DSB/SC modulation of a carrier by a single tone [see (8-9) and (8-10)]. As shown in Figs. 8-9 and 16-5a, the envelope or amplitude-modulation apparent in the resultant signal is the absolute value of a cosine wave; that is,

$$A(t) = V_{oPEP} |\cos \omega_m t| \quad (16-4)$$

and the peak-envelope power output is therefore

$$P_{oPEP} = \frac{V_{oPEP}^2}{2R_o} \quad (16-5)$$

Inspection of (8-10) shows that each of the two components of the RF signal has an amplitude of $V_{oPEP}/2$. The average output power is then the sum of the power in each of the signal components or

$$P_{oAVG} = \frac{V_{oPEP}^2}{4R} \quad (16-6)$$

The peak-to-average ratio for the two-tone test signal is then 3 dB ($\xi = 2$).

The above results can be generalized for multitone SSB signals (which are used to transmit digital data) and multiple-carrier amplifiers (e.g., those used in repeaters). For any set of k sinusoids, all of which are of different frequencies,

$$P_{oAVG} = P_{o1AVG} + P_{o2AVG} + \cdots + P_{okAVG} \quad (16-7)$$

Unless the different tones or carriers are *coherent* (i.e., two or more have integer-ratio relationships among their frequencies), all of the tones or carriers will, at some point in time, simultaneously reach their positive maximum values, requiring a peak envelope output power of

$$P_{oPEP} = \frac{1}{2R} (V_{o1PEP} + V_{o2PEP} + \cdots + V_{okPEP})^2 \quad (16-8)$$

Thus, for multitone SSB signals or multicarrier signals, average powers are additive, as are peak-envelope voltages. If all k tones or carriers had the same amplitude (and were unmodulated), $P_{oAVG} = kP_{oCAR}$ and $P_{oPEP} = k^2P_{oCAR}$; hence the peak-to-average ratio is simply $\xi = k$.

Single-sideband signals arising from speech are characterized by somewhat larger peak-to-average ratios than are the two-tone SSB and single-tone AM test signals. Figure 16-5c illustrates a speech-derived SSB signal with a peak-to-average ratio of about 16 dB. Intelligibility of a speech signal in the presence of noise can be improved by either increasing the transmitter's peak envelope power or by *compressing* the amplitude of the speech signal or resultant SSB signal to reduce its peak-to-average ratio. Most methods of speech compression fall into two categories. In the first technique [7], the audio-frequency speech signal is applied to a nonlinear amplifier (e.g., one

with a logarithmic transfer characteristic) that amplifies signals with smaller amplitudes more than those with larger amplitudes. In the second technique [8], an RF clipper followed by an SSB filter is inserted immediately following the SSB modulator (and filter, if it uses the filter method). Figure 16-5d shows a speech-derived SSB signal with a peak-to-average ratio of about 10 dB, presumably resulting from the application of about "6 dB" of compression to the signal shown in Figure 16-5c.

Average Efficiency

The average efficiency of a power amplifier depends upon both the characteristics of the signal being amplified and upon the type of amplifier. Linear amplification in a Class A PA requires that $P_i = 2P_{oPEP}$ (because the bias current must be at least as large as the ac output current) so that

$$\eta_{AVG} = \frac{P_{oAVG}}{2P_{oPEP}} = \frac{1}{2\xi} \quad (16-9)$$

The efficiency of ideal Class A PAs is therefore only 25 percent for two-tone SSB test signals and only 18.25 percent for single-tone AM test signals.

The analysis for Class B amplifiers is more complicated, since the signal characteristics must be used. Consider first the amplification of a single-tone AM signal. In a Class B PA, the dc input current is proportional to the absolute value of the RF output current or voltage. The average input power for an AM signal is therefore the same as the input power for the carrier without any modulation, hence

$$\eta_{AVG} = \frac{P_{oAVG}}{P_{iAVG}} = \frac{(\frac{3}{8})P_{oPEP}}{(\frac{1}{2})(4/\pi)P_{oPEP}} = \frac{3\pi}{16} \approx 0.589 \quad (16-10)$$

The average efficiency of a Class B PA with the two-tone SSB test signal can be determined similarly. The dc input current, hence P_i , will vary with the output current, forming a full-wave rectified shape whose peak value is the current required by an amplifier delivering P_{oPEP} . The average dc input current is therefore $2/\pi$ times the peak dc input current and the average efficiency is

$$\eta_{AVG} = \frac{P_{oAVG}}{P_{iAVG}} = \frac{(\frac{1}{2})P_{oPEP}}{(2/\pi)(4/\pi)P_{oPEP}} = \frac{\pi^2}{16} \approx 0.617 \quad (16-11)$$

To determine the average efficiency of a Class B PA for multitone or speech signals, the average dc input current and the average envelope voltage must be found. To do this exactly requires knowledge of the statistical characteristics of the signal being amplified [10]. Useful and reasonably accurate estimates (within a few percent) can be made by assuming that the envelope has a Gaussian (normal) probability density function [11]. The

average output power is then σ^2/R , where σ is the rms voltage or standard deviation of the Gaussian random variable. The average absolute value of the output voltage is 0.675σ , which can be determined by examination of tables of Gaussian statistics. The peak-to-average ratio is roughly $\xi = (V_{oPEP}/\sigma)^2$, since there is very little likelihood of voltages larger than 2σ for $V_{oPEP} \geq 2\sigma$. The result is

$$\eta_{AVG} \approx \frac{\pi}{4} \frac{1.048}{\sqrt{\xi}} \quad (16-12)$$

All of the above discussions have determined average efficiencies for ideal amplifiers. The average efficiencies of real amplifiers are reduced by saturation voltage and resistance in the same manner as are the efficiencies at peak-envelope power. For example, if a Class B PA has an efficiency of 65 percent at peak-envelope power output, its efficiency when amplifying a two-tone test signal will be about $(0.617) \times (0.65) = 0.401$.

16-4 AGC and VSWR Protection

Automatic gain control is generally included in SSB transmitters to ensure that distortion does not occur because the final power amplifier was driven beyond its capability to reproduce the signal linearly. Similar circuitry can be used to reduce the drive automatically when the load deviates from its design value.

Automatic Gain Control

The amplitude of the speech signal delivered to the transmitter from the microphone is in general quite variable. Automatic gain control (Fig. 16-6a) is therefore used in most SSB transmitters [12] to prevent either gross under-drive (and hence low output power) or overdrive (and hence intermodulation distortion and clipping) of the power amplifiers. Automatic gain control in a transmitter is generally similar to AGC in a receiver.

In Fig. 16-6a, the RF output voltage is coupled through a blocking capacitor to an envelope detector. This detector has a fast attack (10 milliseconds or less) and a slow decay (0.1 to 1 second) characteristic so that overall amplifier gain is determined by the peak signal envelope rather than the average of the signal envelope. If the envelope detector output exceeds a preset or adjustable threshold voltage, the signal is applied to the controlled-gain stage to reduce its gain.

The controlled-gain stage usually operates in a low-power segment of the RF power amplifier chain. Any of several circuits can be used as the