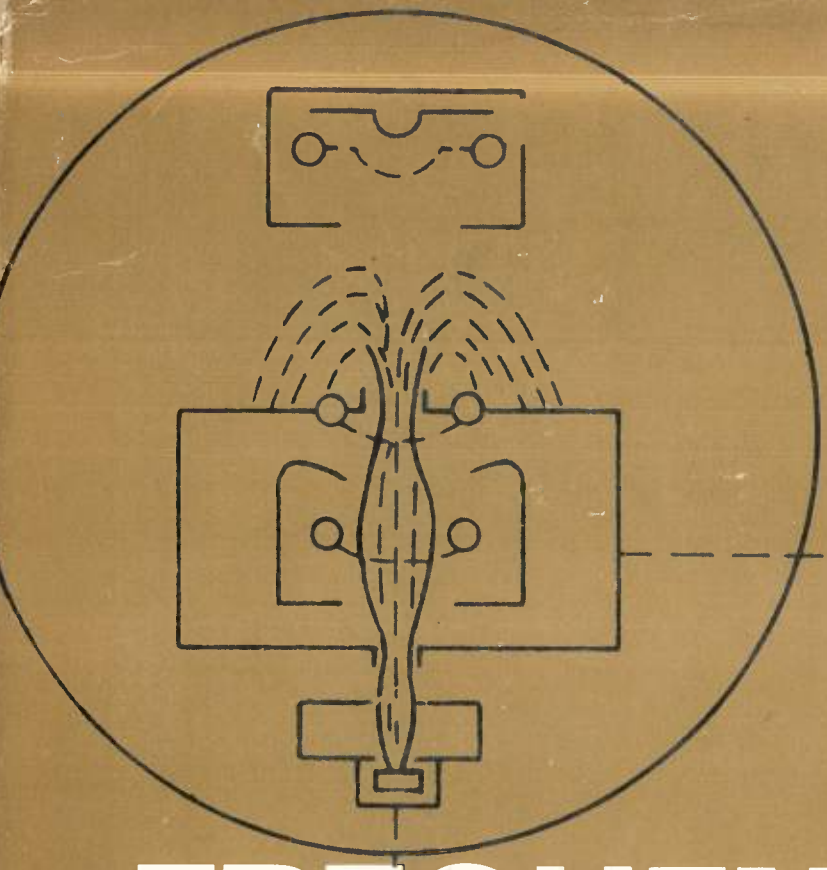


# FREQUENCY MODULATION RECEIVERS

A. B. Cook / A. A. Liff

The most comprehensive and up-to-date treatment available on the operational aspects of wide band FM receivers. Covers all circuits, auxiliary circuits, and other major aspects of the field. An ideal reference for in-service professionals.

# Other recommended books of interest . . .

## TRANSFORM CIRCUIT ANALYSIS FOR ENGINEERING AND TECHNOLOGY

by WILLIAM D. STANLEY, *Old Dominion College, Virginia*

Here is a book which details the fundamentals of transient circuit and systems analysis employing the LaPlace transform and pole-zero approach for analyzing and interpreting problems. Essential to engineers and technicians, this guide to the current transform methods is presented in a clearly illustrated step-by-step sequence, and is oriented to the engineering approach, thereby making only a minimum of advanced mathematics necessary to completely master the subject.

In its careful development of the material, the reader is carried from circuits through to basic systems, establishing a solid background for further work in electronics, control systems, and communications systems. Whether as a reference for working engineers and technicians, or as a self-instructional guide to the fundamentals of the subject, this book is a valuable working tool for anyone in electronic technology and engineering.

*Published 1967* *368 pages*

## ELECTRONIC TROUBLESHOOTING: A Self-Instructional Programed Manual

by Members of the Staff, *Philco Technical Institute*

This book gives the vital know-how to troubleshoot rapidly and accurately to minimize equipment "downtime"; it guides you to fast, accurate troubleshooting by showing you how to apply logical thinking and a systematic approach to each problem.

The author explains in the Preface, ". . . This is not an ordinary text; it is a self-instructional programed manual. Although the pages are numbered consecutively, you will not study them in that order. After studying a particular information block, you check your understanding by selecting one of the several choices offered as possible answers to a question pertaining to that information block. You then turn to the indicated page and are told whether or not you made the correct selection. If you have chosen the correct answer, you go on with the learning process. If you have selected an incorrect answer, you are told why and asked to review the information before going any further. Part I outlines troubleshooting principles and practice; Part 2 provides actual troubleshooting problems."

The major advance in this self-instructional programed manual is that you are never permitted to form erroneous ideas and/or arrive at the wrong conclusion. Following the procedure outlined in the Preface, you will acquire a real understanding of how to deal with any troubleshooting problem.

*Published 1966* *276 pages*

PRENTICE-HALL, Inc.
Englewood Cliffs, New Jersey

Providing one of the most comprehensive treatments on the subject of wide band FM receivers, this book offers an up-to-date reference to the entire subject of commercial Frequency Modulation broadcast receivers. It forms an ideal one-volume reference for in-service professionals—RF Engineers and Technicians, Communications Engineers, Broadcast Engineers, Service Technicians, as well as for those readers planning to enter any of these professions.

The book offers a complete discussion of the circuits which comprise an FM receiver, consequently covering all basic circuits such as the RF amplifier, mixer-oscillator, IF amplifiers, limiter, as well as all types of FM detectors. Major circuits are discussed in terms of vacuum tube, transistor, field effect transistors, and integrated circuit characteristics. Thus all areas of past and present practice are covered. Also discussed in detail are the various receiver auxiliary circuits such as tuning indicators, stereo multiplex, and squelch.

The two sections of this work covering FM detectors and tuning indica-

*(continued on back flap)*

*(continued from front flap)*

tors represent the most comprehensive treatments of both topics available anywhere in the literature of the field.

An important section of the book covers noise and VHF phenomena to account for the important factor of the very high frequencies at which commercial FM receivers operate and the weak signal reception which is a major characteristic of this receiver. Each section of the work ends with a detailed summary of the material covered, and is augmented by sample problems which can be directly applied to the reader's work.

Throughout, the work places the greatest stress on the operational aspects of FM receivers, and in its completeness and coverage of the most recent developments in the field, represents the most useful reference tool for any professional in FM technology.

---

A. B. COOK is Staff Engineer at Perkin-Elmer Corp.

ALVIN LIFF is Administrator of Special Projects at the RCA Institute, Inc. In addition to the present volume, he is the author of two widely used laboratory manuals in his field.

PRENTICE-HALL, Inc.
Englewood Cliffs, New Jersey

968 • Printed in U.S. of America

# FREQUENCY MODULATION RECEIVERS

A. B. COOK
*Perkin-Elmer Corp.*
A. A. LIFF
*RCA Institutes, Inc.*

PRENTICE-HALL, INC.
*Englewood Cliffs, N.J.*

To our wives and families without whose understanding and cooperation this book would not have been possible.



Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

Library of Congress Catalog Card Number: 68–20857
Printed in the United States of America

# FREQUENCY MODULATION RECEIVERS

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY. LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

PRENTICE-HALL SERIES IN ELECTRONIC TECHNOLOGY

DR. IRVING L. KOSOW, editor

CHARLES M. THOMSON AND JOSEPH J. GERSHON, consulting editors

ANDERSON, SANTANELLI AND KULIS Alternating Current Circuits and Measurements: A Self-Instructional Programed Manual
ANDERSON, SANTANELLI AND KULIS Direct Current Circuits and Measurements: A Self-Instructional Programed Manual
BABB Pulse Circuits: Switching and Shaping
BARRINGTON High Vacuum Engineering
BENEDICT AND WEINER Industrial Electronic Circuits and Applications
BRANSON Introduction to Electronics
COOK AND LIFF Frequency Modulation Receivers
CORNING Transistor Circuit Analysis and Design
COWLES Transistor Circuits and Applications
DOYLE Pulse Fundamentals
EADIE Introduction to the Basic Computer
FEDERAL ELECTRIC CORPORATION Mathematics for Electronics: A Self-Instructional Programed Manual
FEDERAL ELECTRIC CORPORATION Special Purpose Transistors: A Self-Instructional Programed Manual
FEDERAL ELECTRIC CORPORATION Transistors: A Self-Instructional Programed Manual
FLORES Computer Design
HEWLETT-PACKARD Microwave Theory and Measurements
JACKSON Introduction to Electric Circuits, 2nd ed.
JOHNSON Servomechanisms
KOSOW Electric Machinery and Control
LENK Data Book for Electronic Technicians and Engineers
LENK Handbook of Oscilloscopes: Theory and Application
LITTON INDUSTRIES Digital Computer Fundamentals
MALEY AND HEILWEIL Introduction to Digital Computers
MANDL Directory of Electronic Circuits with a Glossary of Terms
MANDL Fundamentals of Electric and Electronic Circuits
MANDL Fundamentals of Electronic Computers: Digital and Analog
MARCUS Electricity for Technicians
MARCUS Electronics for Technicians
MARTIN Technical Television
MEYERS Practical Semiconductor Experimentation
O'NEAL Electronic Data Processing Systems: A Self-Instructional Programed Manual
PHILCO TECHNOLOGICAL CENTER Electronic Precision Measurement Techniques and Experiments
PHILCO TECHNOLOGICAL CENTER Servomechanism Fundamentals and Experiments
POLLACK Applied Physics
POWER DC-AC Laboratory Manual
PRENSKY Electronic Instrumentation
RCA INSTITUTES Basic Electronics: Autotext—A Programed Course
RCA SERVICE COMPANY Fundamentals of Transistors: A Programed Text
RASKHODOFF Electronic Drafting and Design
REID AND KUBALA Experiments in Direct Current Circuits
SHIERS Design and Construction of Electronic Equipment
SHIERS Electronic Drafting
STANLEY Transform Circuit Analysis for Engineering and Technology
WHEELER Introduction to Microwaves
WHEELER Radar Fundamentals

# PREFACE

This text provides students of technical institutes, service technicians, practicing engineers and other interested parties with a comprehensive treatment of commercial Frequency Modulation (FM) broadcast receivers. Emphasis is placed on the wideband systems used for broadcasting in the 88 to 108 MHz bands.

Chapter 1 summarizes the FM receiver on a block diagram basis. The reader is given an integrated systems viewpoint first while subsequent chapters describe each "block" in detail. Transistor, field-effect transistor (FET), integrated circuit and vacuum tube aspects are covered.

Emphasis is placed on switching circuits which are found in AM-FM tuners as well as special controls (level, local-distant, afc on-off, etc.).

A summary and references have been included at the end of each chapter. This will make the book more useful to instructors in technical insitutes and junior colleges where it should find its greatest appeal. In order to achieve the desired simplicity, mathematics has been kept to a minimum while detailed explanations have been given for special topics which have not in the past, satisfied the average student. It is assumed that the reader has a background in basic electronic theory, fundamental algebra, and basic trigonometry. From experience gained by teaching, the authors have found that the phasor as a tool for analysis, leads the student to a deeper understanding of complex circuit conditions. Phasors are therefore used freely. Consequently, the student should be capable of handling all of the methods of phasor addition, etc. In order to minimize cluttering of text material with involved mathematics, a number of derivations have been included as appendices.

This book places the greatest stress on the operational aspects of FM receivers. The early chapters deal mainly with modulation concepts, inter-

ference and noise structure, and RF considerations. Since the frequencies employed for the FM broadcast band are in the 100 MHz region, a chapter has been included on very high frequency (VHF) effects as related to tubes, transistors and lumped and distributed properties.

The subject of FM detection is treated in great detail since these circuits represent the major difference between Amplitude Modulation (AM) and FM receivers. Also, characteristics such as fidelity, capture, and freedom from all spurious outputs are basically related to detector performance.

A chapter dealing with stereophonic multiplex broadcasting is included which treats the underlying theory of transmission. Completely integrated multiplex tuners are also discussed with emphasis on those circuits which are unique to the specialized needs of multiplex equipment.

With the widespread utilization of solid-state FM equipment, the need for a discussion of the circuits related to these components is clearly indicated. Therefore, each chapter (RF, IF, limiter, etc.) will contain an analysis of representative circuits.

The authors would like to thank their friends and colleagues on the faculty at R.C.A. Institutes for their cooperation during the preparation of this book. In particular, we would like to thank Benjamin Zeines, Herbert Friedman, Milton Liljestrand, James Hamilton, Waldemar Horizny, Matthew Mahoney and Mr. F. Mergner of the Fisher Radio Corporation for their aid and suggestions which we found helpful.

A. B. COOK
A. A. LIFF

# CONTENTS

# 1

# INTRODUCTION

The Second World War proved that frequency modulation was a practical and effective means of noise-free signaling. Frequency-modulation (FM) broadcasting remained relatively dormant in the years after the war because of public ignorance of the capabilities of FM and lack of interest in high-quality sound reproduction. With the introduction of the long-playing record in 1948, interest in high-fidelity sound increased and a better-informed public began to appreciate the noise-free properties and wide dynamic range of FM. The introduction of stereophonic records in 1957 (the Westrex 45/45 system) and wide-band FM receivers led to a fuller appreciation, by greater numbers of people, of quality sound reproduction.

In April of 1961 the Federal Communications Commission approved the transmission standards for stereophonic multiplex broadcasting. This was an important milestone in FM broadcasting, and it appears that FM as an entertainment medium has entered an era of unprecedented growth.

## 1-1. Advantages of FM over AM

FM transmitters are smaller and more efficient than their amplitude-modulation (AM) counterparts. The AM carrier is necessary for detection purposes and therefore must be transmitted at all times (for commercial AM). Since the AM carrier component conveys no information, it represents a waste of power. For the same power output, the FM transmitter requires a smaller input power than the AM transmitter and can thus be smaller.

For example, a 25-kw (average power, no modulation) AM transmitter must be capable of handling 100 kw of peak power when the modulation reaches 100 per cent. This is because the peak plate current in the last power amplifier doubles, and the power, being proportional to the square of the current, increases by a factor of four. In the case of the FM transmitter the total radiated power remains constant and independent of the degree of modulation. Increasing the degree of modulation simply results in a shift of power from the carrier component to the sidebands; the total power remains constant. An FM transmitter can transmit at full power at all times without provision for increases in the degree of modulation. Thus, in FM, the degree of modulation affects the bandwidth (as will be shown later) but not the power output, while in AM the degree of modulation affects the power output but not the bandwidth (assuming that the carrier is not over-modulated).

Noise, in general, tends to mask weak signal communications. AM is more susceptible to noise than FM, and therefore FM can satisfactorily operate with a lower signal-to-noise ratio ($S/N$). This is because the disturbance combines with the signal in such a way as to introduce a large degree of amplitude modulation of the carrier. Spurious frequency modulation is also produced, but to a lesser degree. The AM disturbance may be removed by noise-limiter circuits in the FM receiver. If that same disturbance should be introduced into an AM transmission, it could not be similarly removed without degrading the quality of the intelligence being transmitted. A conservative figure for the required $S/N$ ratio in FM is of the order of 2 to 1. This means that if the signal is approximately twice as strong as the interfering noise (or other interfering signal), the desired signal will suppress the disturbance. For similar suppression of the undesired noise or interference, the $S/N$ ratio for AM should be in the order of 100 to 1. FM also exhibits a channel-grabbing or "capture" effect that AM does not. This is the tendency of an FM receiver to ignore the weaker of two signals of nearly equal amplitude and equal or nearly equal frequency, permitting the stronger signal to take over or "capture" the FM detector. More will be said on this point in the next chapter.

Since the FM system (wide-band) enjoys a greater degree of freedom from interference than AM, much stronger interfering signals can be tolerated in FM than in AM. Therefore, FM stations can be placed much closer "geographically" than AM stations (on the same channel) and more can be accommodated in a given service srea.

## 1-2. Disadvantages of FM

(a) In order for FM to realize its noise-reduction advantages over AM, its bandwidth must be much wider than for AM. In other words, the FM system

sacrifices bandwidth occupancy in the transmission medium for superior noise performance. It will be shown later that a bandwidth of approximately 240 kHz is required in FM for a modulating frequency of 15 kHz, while the corresponding bandwidth in AM would be 30 kHz.

(b) FM transmitting and receiving equipment tends to be more complex than AM equipment, which means (as far as the consumer is concerned) that the average FM radio is more expensive than the average AM radio.

(c) The frequency band for FM is 88 to 108 MHz. At these frequencies, FM stations provide reliable reception that is limited approximately to "line-of-sight" distances.

(d) Tuning is somewhat more critical in FM compared to AM. This can be alleviated, however, by the use of some form of tuning indicator.

It should be stressed at this point that FM is *not* inherently a "better" system than AM. Some AM broadcast stations have modulated their carriers with 15-kHz audio signals, but this is relatively rare because of interference problems with adjacent stations. This is not the case with FM stations, however, and audio modulation up to 15 kHz is used regularly. Thus FM employs a greater frequency range than AM—*not* because of the system used but because of adjacent channel interference. If equal care is given to both AM and FM, both systems will provide excellent sound reproduction.

## 1-3. Functional systems description

Figure 1-1 is a block diagram of a typical FM tuner.† Each block will be discussed in detail in subsequent chapters; here we simply consider its general features.

All FM commercial broadcast receivers and tuners‡ are of the superheterodyne type. Thus, as shown in Fig. 1-1, the incoming FM carrier is first amplified by an RF amplifier. The major functions of the RF amplifier (regardless of the type used) are to increase the front-end $S/N$ ratio, minimize oscillator radiation, and provide a certain degree of selectivity against spurious signals. The amplified RF carrier is combined with a relatively large-amplitude signal from the local oscillator, whose function is to generate a constant-amplitude carrier wave. The combined signals are then applied to the input of a mixer or frequency-converting stage. The mixer heterodynes or beats the two signals, and the intermodulation products that appear in the output circuit contain a difference signal (among others) called the intermediate frequency or IF. The mixer must display a nonlinear characteristic if the beat component is to appear in the output. The resulting IF signal is

†A tuner contains all of the essential "blocks" except the audio voltage and power amplifiers and loudspeaker.

‡These tuners cover the broadcast range from 88 MHz to 108 MHz.

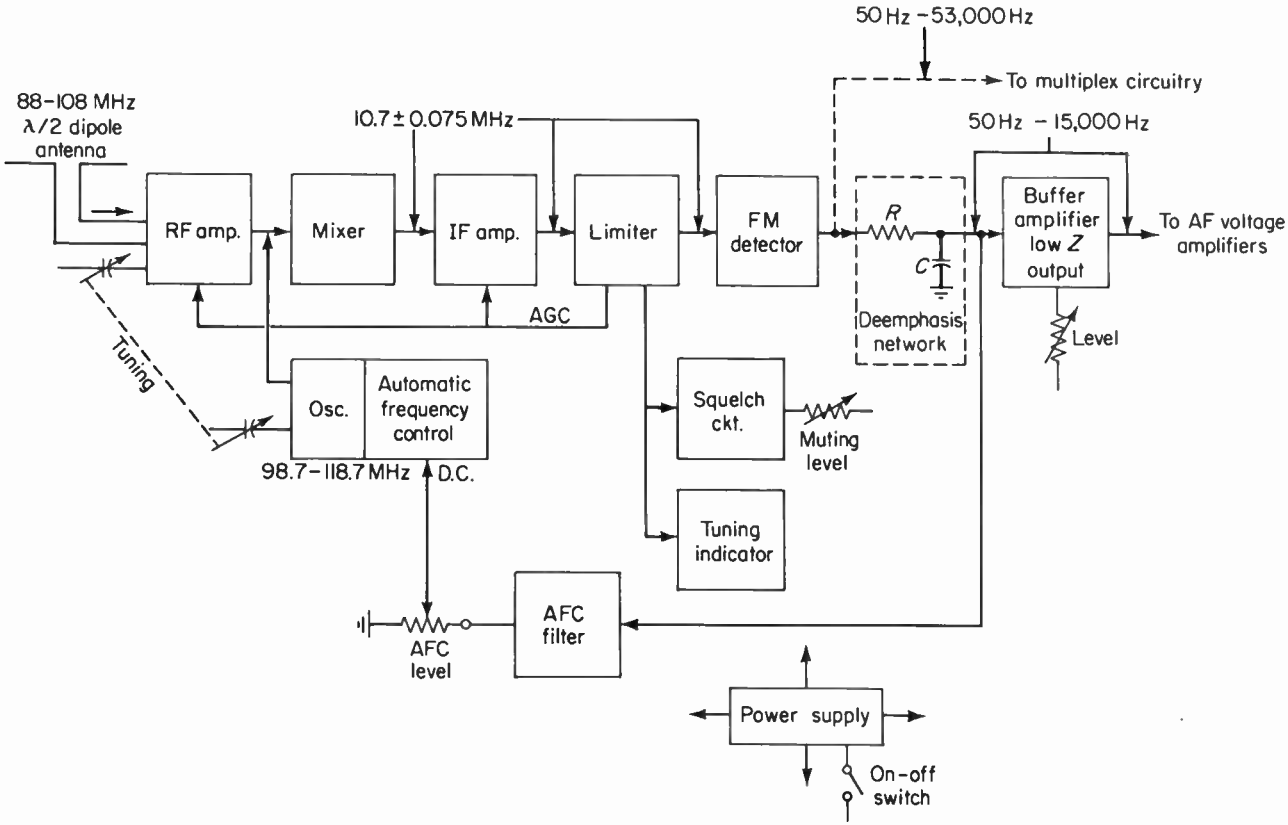



**Fig. 1-1.** The block diagram of a representative FM broadcast tuner including operating frequencies.

applied to one or more IF amplifiers for amplification. The main function of the IF amplifier is to provide high amplification (sensitivity) and selectivity for rejection of spurious signals. The output of the last IF amplifier is applied to the input of a stage that removes any amplitude modulation of the impressed wave by a process known as limiting. These amplitude limiters remove AM and retain the frequency deviations of the carrier. The amplitude-limited signal is applied to an FM demodulator (or detector). The FM detector converts frequency variations, in a linear manner, into audio-frequency voltage variations for application to the succeeding stages. The network following the FM detector reduces the amplitude of the higher-frequency audio components, since these components were accentuated at the tramsmitter for purposes of noise reduction. This network, termed a *deemphasis network*, is generally an *RC* network whose time constant, as prescribed by the FCC, is 75$\mu$sec. In addition to the foregoing basic building blocks, provision is sometimes made for: tuning indicators to assist in proper tuning; automatic gain control (AGC) to maintain constancy of audio output and minimize overloading of the RF and IF amplifiers; automatic frequency control (AFC) of the local oscillator to ensure frequency stability and further assist in proper tuning; an interstation squelch or muting circuit to quiet the receiver between stations; and a low-impedance emitter or cathode-follower output stage to permit long connecting cables between the tuner and the main amplifier chassis. The configuration shown in Fig. 1-1 is not the only possible arrangement, but it is representative of many good-quality FM broadcast tuners. The reader will appreciate that FM tuners are thus somewhat more complex than their AM equivalents. In the case of combination AM-FM tuners, various switching elements (not shown) are often provided.

## 1-4. Basic modulation considerations

*Modulation* of a high-frequency carrier is a process whereby some characteristic of the carrier signal is varied in accordance with the variations of another signal. This latter signal, in subsequent discussions, will be an audio-frequency voltage, which will be termed the *modulating signal.* The three basic properties of the high-frequency carrier wave are its amplitude, frequency, and phase angle. *Frequency* and *phase modulation* are termed *angle modulation* and, as will be shown, are always coexistent. Any one of the three basic properties of the wave can be varied in accordance with the modulating signal while the other two properties are held constant. For example, in *amplitude modulation* (AM) the frequency of the carrier is maintained constant (as well as the phase angle) while the amplitude varies in accordance with the amplitude and frequency of the audio modulating signal. In *frequency modulation* (FM) the carrier amplitude is maintained constant while the frequency of the carrier is caused to deviate from its average or unmodu-

lated value in accordance with the modulating signal. In *phase modulation* (PM) the amplitude of the carrier is again maintained constant while its phase angle (with respect to a reference) is caused to vary in accordance with the modulating signal.

The two principal methods, which will be studied in subsequent sections, are shown in Fig. 1-2. The unmodulated carrier for FM is shown having a frequency of $F_c$. The same sinewave at 1 kHz is used to modulate each wave. In the AM case the amplitude of the carrier is seen to vary about the unmodulated amplitude $E_c$. The *degree* to which the amplitude varies above and below $E_c$ is a function of the *amplitude* of the modulating signal $f_a$, while the *rate of change* of amplitude variation is a function of the *frequency* of the modulating signal. In the FM case the *degree* to which the frequency of the carrier varies above and below its unmodulated value is a function of the *amplitude* of the modulating signal, while the *rate of change* of frequency variation is a function of the *frequency* of the modulating signal.

In other words, at point $A$ in the modulation cycle the frequency of the FM carrier is at its unmodulated value of 100 MHz. Between points $A$ and $B$ the frequency of the FM carrier is deviating upward until it reaches a maximum of 100.075 MHz. Between points $B$ and $C$ the carrier frequency is deviating downward, until at $C$ it reaches its unmodulated value of 100 MHz. From $C$ to $D$ the carrier frequency is deviating downward (below its unmodulated value), until it reaches a minimum of 99.025 MHz at point $D$. From $D$ to $E$ the carrier frequency commences to deviate upward again until, at point $E$, it reaches its unmodulated value of 100 MHz. The change to *one* side of *center frequency* (100 MHz) is called the deviation, while the total deviation (given in this illustration as 150 kHz) is called the *frequency swing*. Thus, the frequency swing is equal to twice the deviation. We shall consider frequency deviation in much more detail later on.

## 1-5. Amplitude modulation

We have seen that in AM the amplitude of the carrier is varied in accordance with the modulating signal. The amplitude of the modulating signal determines the amount of amplitude change, while the frequency of the modulating signal determines the rate of change of amplitude variation. The *modulation depth* is a measure of how much of a change the carrier amplitude is undergoing; this is called the *modulation factor*, $m$. It is normally expressed as a percentage and should not exceed 100 per cent or unity. For our purposes, then, $m$ will be a value between 0 and 1. In order to discuss the *sideband spectrum* for AM, it is convenient to derive the composite wave equation for the modulated carrier. Only one trigonometric identity will be used, and the reader should have no difficulty in following the mathematical manipulations. Let the unmodulated carrier wave be represented by the expression:

$$e_c = E \cos \omega_c t. \tag{1-1}$$

$E$ is the *carrier amplitude*, which will vary at the audio rate about the average value of $E_c$ as shown in Fig. 1-2. The *angular velocity* of the carrier, which is the number of radians per second or $\omega_c$, is constant. $\omega_c$ is equal to $2\pi f_c$ and is sometimes referred to as the radian frequency. Thus, $E$ will consist of some average value (independent of modulation) in addition to a varying value, or

$$E = E_c + E_a \cos \omega_a t. \tag{1-2}$$

$E_a$ is the maximum value of the audio voltage, while its angular velocity is

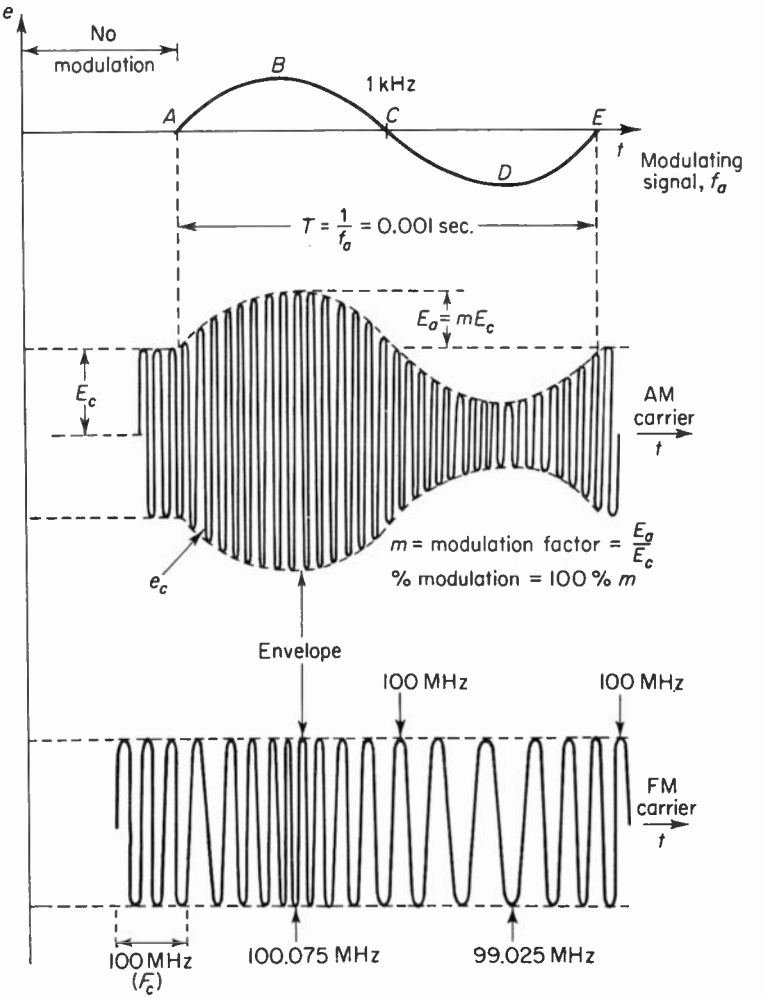



**Fig. 1-2.** Basic comparison between an amplitude modulated carrier and a frequency modulated carrier.

$\omega_a = 2\pi f_a$. If we multiply $E_a$ by $E_c/E_c$, then

$$E = E_c + \frac{E_c E_a}{E_c} \cos \omega_a t. \tag{1-3}$$

The ratio $E_a/E_c$ can be replaced by the modulation factor $m$ which is the percentage change of $E$ about the average value. Thus,

$$E = E_c + mE_c \cos \omega_a t. \tag{1-4}$$

Substitution of Eq. (1–4) in Eq. (1–1) gives

$$e_c = (E_c + mE_c \cos \omega_a t) \cos \omega_c t. \tag{1-5}$$

Expanding Eq. (1–5) gives

$$e_c = E_c \cos \omega_c t + mE_c \cos \omega_c t \cos \omega_a t. \tag{1-6}$$

An identity in trigonometry states that the product of the cosines of two angles is equal to the sum of one-half the cosine of the sum of the angles and one-half the cosine of the difference of the two angles or

$$\cos x \cos y = \tfrac{1}{2} \cos (x + y) + \tfrac{1}{2} \cos (x - y). \tag{1-7}$$

If we let $x = \omega_c t$ and $y = \omega_a t$, using Eq. (1–7) as a model, then

$$e_c = E_c \cos \omega_c t + \frac{mE_c}{2} \cos (\omega_c - \omega_a)t + \frac{mE_c}{2} \cos (\omega_c + \omega_a)t. \tag{1-8}$$

Equation (1–8) is the fundamental equation for the composite radiated AM wave. The physical interpretation of this equation is as follows: the right side of the equation contains three terms. The first term is exactly the same as that of Eq. (1–1) or the unmodulated carrier. The second term contains an angular velocity of $(\omega_c - \omega_a)$ or $2\pi(f_c - f_a)$ and is called the *lower sideband.* The third term contains an angular velocity of $(\omega_c + \omega_a)$ or $2\pi(f_c + f_a)$ and is called the *upper sideband.*

Two facts should be stressed at this point. First, all three terms are carriers, but of different frequencies. Second, nothing is removed from the original carrier in the modulation process; instead, energy is *added* in the form of the upper and lower sidebands. A further examination of the equation shows that if the modulation factor is unity (100 per cent modulation), the amplitude of each sideband is one-half the unmodulated carrier amplitude or $E_c/2$. If the modulation percentage does not exceed 100 per cent, *each* audio modulating signal gives rise to two sidebands, which are separated in frequency from the carrier by an amount equal to the modulating frequency $f_a$. Thus, in AM, the sideband separation is determined by the audio *frequency* while the sideband amplitude $(mE_c/2)$ is determined by the modulation factor, which is determined by audio *amplitude.* As an example, if a 1000-kHz carrier is amplitude-modulated by a single 1-kHz audio tone, the upper side-frequency is $f_c + f_a$ or 1001 kHz, while the lower side-frequency is $f_c - f_a$ or 999 kHz. Another point to be stressed is that the *intelligence* is contained in the side-

frequencies *only* and not in the unmodulated carrier. In fact, the intelligence is contained in *either* of the two side-frequencies, since both of the side-frequencies contain the modulation factor $m$. This factor is a measure of the amplitude of the audio signal, while the "$\omega_a$" contains the frequency of the audio signal. In double-sideband-plus-carrier transmission (conventional AM broadcasting) all three terms are "radiated." In double-sideband suppressed-carrier transmission the last two terms are "radiated" and the first term is suppressed. In single-sideband suppressed-carrier transmission, the first term is suppressed and *either* of the last two terms is "radiated."

If the modulation percentage is 100 ($m = 1$), the amplitude of the composite wave will rise to $2E_c$ and fall to zero. Equation (1–8) states that for each audio tone two side-carriers are generated and each is frequency-displaced from the carrier by an amount equal to the modulating frequency. This is true only if the modulation factor does not exceed unity. If $m > 1$, an infinite number of side-frequencies are generated per audio frequency. This is considered *overmodulation*. The modulation envelope is no longer sinusoidal and will consist of a series of harmonics. Therefore, a pair of side-frequencies will be generated for each harmonic. The modulation envelope will appear as shown in Fig. 1-3 with resultant distortion at the receiver.

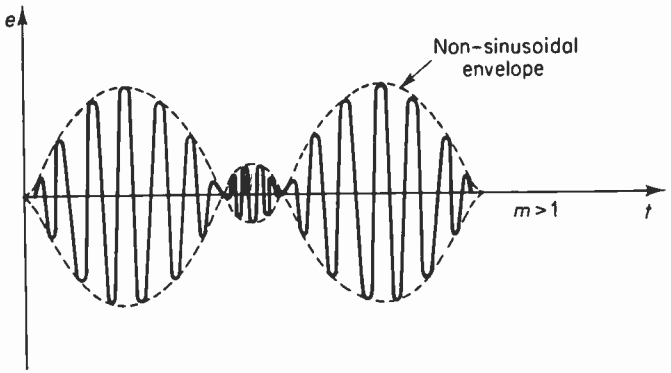



**Fig. 1-3.** The modulation envelope when the AM carrier is overmodulated.

We can show the sideband spectrum for a completely modulated AM carrier ($m = 1$) by considering that the amplitudes of each side-frequency are equal to $E_c/2$. Assuming *single-tone modulation* (a pure sinewave modulates the carrier), the audio frequency is $f_a$ and the carrier frequency is $f_c$. The spectrum is shown in Fig. 1-4. If the intelligence is to be recovered without distortion, neither sideband is to be attenuated in the transmission medium *or* in the receiver. This calls for a bandwidth of $2f_a$ or twice the *maximum* audio frequency. If 15 kHz is used to modulate the carrier, a bandwidth of 30 kHz would be required. If the AM channel separation is 10 kHz (this is the *minimum* separation between AM transmitters), it would appear that the

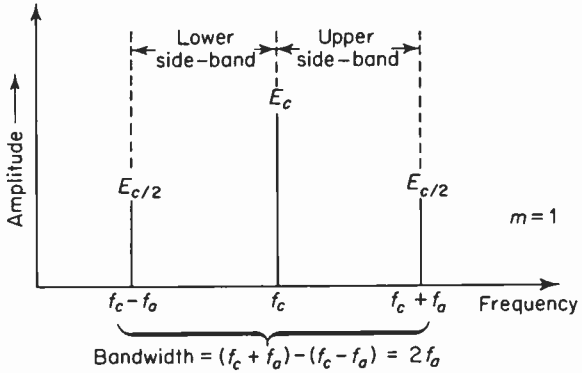



**Fig. 1-4.** The spectrum of an AM wave. A single audio tone will produce two side-frequencies which contain the complete specification of the modulating signal (*i.e.*, amplitude and frequency).

highest audio frequency permitted would be 5 kHz in order to prevent overlapping of sidebands on adjacent channels. However, the channel separation in a given service area (metropolitan New York City, metropolitan Los Angeles, and so on) is often 30 kHz, and as long as there is no interference between adjacent stations a transmitter *can* modulate up to 15 kHz and produce a bandwidth capable of transmitting the highest audible audio frequency.

If a transmitter should *overmodulate* ($m > 1$), then *many* sidebands are generated, which may interfere with nearby transmissions as a result of the sidebands "splattering."

Another important consideration in modulation concerns the energy in the modulated carrier. Since energy (or power) is proportional to the square of voltage, the power represented by the wave in Eq. (1–8) is proportional to

$$E_c^2 + \left(\frac{mE_c}{2}\right)^2 + \left(\frac{mE_c}{2}\right)^2 = E_c^2 + \frac{m^2E_c^2}{4} + \frac{m^2E_c^2}{4}. \qquad (1\text{–}9)$$

Equation (1–9) states that the *total* sideband power is 50 per cent of the total carrier power and that the portion contributed by the unmodulated carrier term ($E_c^2$) is constant or independent of the degree of modulation. Also, the sideband power is derived from the modulating source, for if the degree of modulation $m$ is zero, the sidebands will contribute zero power to the radiated wave. When the modulation factor is unity or less, the sideband power is proportional to the square of the modulation factor $m$. Since this involves a "square-law" relationship, the power contained in the information-bearing sidebands decreases very rapidly as the modulation factor is decreased. Generally, a large modulation factor is desired, but this can place severe design restrictions on the AM detector at the receiver. Thus the power contained in the first term (carrier component) in Eq. (1–9) is wasted as far as

intelligence is concerned. If the power contained in the carrier component can be placed in the sidebands, a more efficient transmission can be realized; this can be accomplished in single- and double-sideband suppressed-carrier techniques.

## 1-6. Phasor representation of an AM wave

A convenient way to illustrate the conditions in an AM wave is by means of a phasor diagram. This will help us understand not only AM but also the nature of noise as related to FM in a later chapter

The representation is based upon (a) a sinusoidal carrier and side-frequencies, (b) a modulation factor $m = 1$, and (c) a fixed modulating frequency $f_a$. Since the carrier and sideband voltages are sinusoidal, they can be represented by phasors that are rotating in a counterclockwise direction in the complex plane. Each is rotating at an angular velocity of $2\pi f$ or $\omega$. The carrier angular velocity is denoted by $\omega_c$, the upper-sideband angular velocity is denoted by $\omega_c + \omega_a$, and the lower sideband is denoted by $\omega_c - \omega_a$. Since we are concerned only with *relative* angular velocity, we can maintain the carrier phasor stationary and permit the sideband phasors to rotate about the reference carrier phasor. Since the angular velocity of the upper sideband is higher than that of the carrier, its phasor will rotate counterclockwise at an angular velocity that is the difference between the upper-sideband frequency and the carrier frequency (or $\omega_a$). The lower-sideband phasor will thus rotate at the same angular velocity, but since it rotates slower than the carrier phasor, we will rotate this phasor clockwise. The sideband phasors thus rotate in opposite directions, each about the tip of the carrier phasor at an angular velocity of $\omega_a$. This is shown in Fig. 1-5. The instantaneous resultant of the sidebands and the carrier gives the instantaneous amplitude

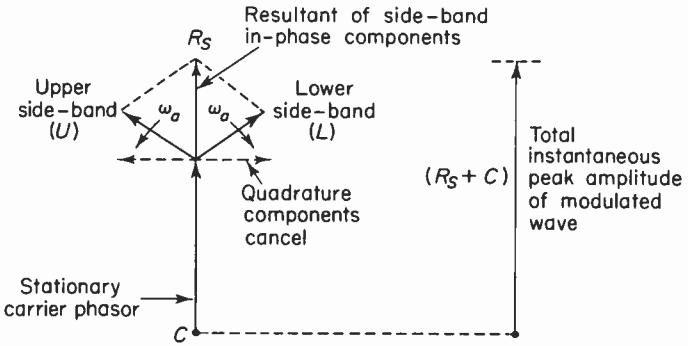



**Fig. 1-5.** The phasor representation of an amplitude modulated carrier for a modulation factor $m = 1$. The upper- and lower-sideband phasors rotate in opposite directions at the angular rate of $2\pi f_a$.

of the modulated wave. Each sideband phasor contains a component, at right angles to the carrier phasor, called the *quadrature* component. Each sideband phasor also contains an *in-phase* component, which is always "in line" with the carrier phasor. When two phasors are "in line," they are said to be collinear.

Note that the quadrature components always cancel, while the in-phase components add to form a sideband resultant that is always collinear with the carrier phasor. Hence we needn't show the quadrature components, since they contribute nothing to the variation in amplitude of the resultant signal.

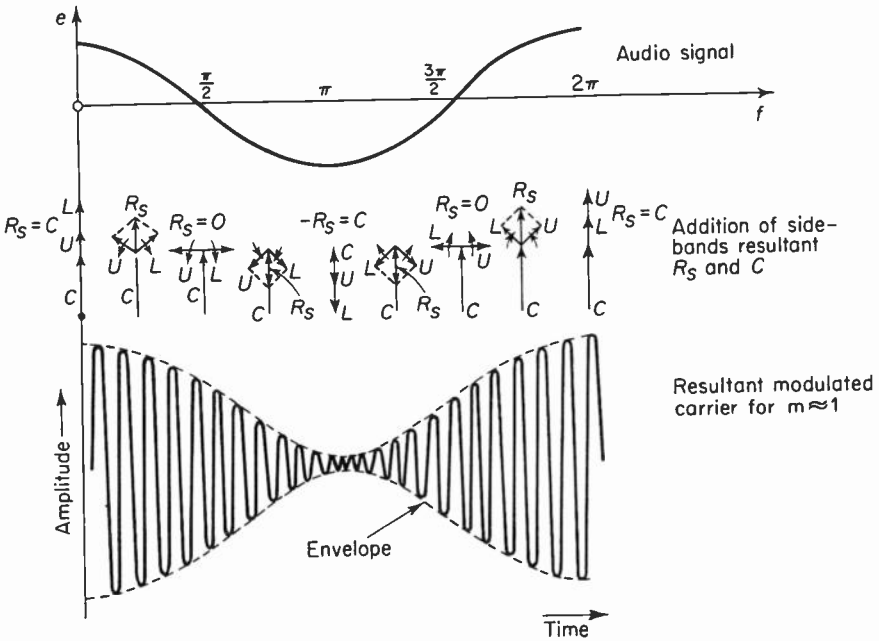



**Fig. 1-6.** The phasors similar to Fig. 1-5 for a number of selected points in the modulation cycle. The resultant modulation envelope is thus obtained by the phasor addition of the sidebands resultant $R_s$ and the carrier amplitude $C$.

Figure 1-6 shows how the modulation envelope is obtained graphically by adding the sideband resultant $R_s$ to the carrier to obtain the instantaneous amplitude of the modulated carrier. This is done for a number of time intervals in the modulation cycle for $m$ approximately equal to unity.

## 1-7. Double-sideband suppressed-carrier AM

It may seem inappropriate to discuss yet another form of AM in a book devoted to FM. However, with the advent of stereophonic FM multiplexing,†

†This topic is covered in Chapter 14.

*double-sideband suppressed-carrier* AM became important because this is precisely the method of modulation used for the stereophonic subchannel. The main channel in FM multiplexing utilizes frequency modulation (which will be discussed next), but the reader should have some idea of the broader aspects of the technique henceforth referred to as **DSSC**.

One possible (and certainly difficult) way to suppress the carrier component in a conventional AM wave is to feed all three voltages [see Eq. (1-8)] into a sharply tuned band-rejection filter that provides high attenuation of the carrier component. The reason for the difficulty is that the filter must provide an infinite attenuation for the carrier and zero attenuation for the sideband components. Conventional filters cannot provide this kind of performance and so we must resort to other methods—one being the balanced modulator. The theory and operation of balanced modulators are discussed in some of the references listed at the end of this chapter, but all we need be concerned with here, is that if a modulating signal and a carrier are applied to a balanced modulator the carrier term will be balanced out and only the sidebands will appear in the output.

A phasor representation of the **DSSC** signal can be developed by first considering the conventional AM phasor of Fig. 1-5 and simply omitting the carrier phasor $C$. The sidebands resultant $R_s$ is thus the phasor addition of the upper- and lower-sideband components. Figure 1-7 shows that the two sidebands are rotating about the point $P$ in opposite directions and at the audio rate. Also, part (b) of the same figure shows that the envelope of the sidebands resultant is nonsinusoidal and that the sidebands resultant shifts phase (by 180°) every half-cycle of the modulating signal. Part (c) of the same figure shows what the output of an AM envelope detector would look like if the carrier component were not reinserted at the receiving end. The rectified output would be twice the frequency of the original audio signal and would contain an infinite number of harmonics.

Thus, in order to recover the original sinewave modulation, it is necessary to reinsert the carrier component in the correct amplitude, frequency, and phase relationships. Amplitude is important because if the carrier amplitude is too small, it is the equivalent of a high modulation factor. If this factor exceeds unity, severe harmonic distortion could result. Frequency is important because the recovered audio will be related to the sum and difference frequencies ($\omega_c + \omega_a$ and $\omega_c - \omega_a$). The phase of carrier reinsertion at the receiver is important because any phase error between the original carrier (suppressed) and the reinserted carrier will produce severe harmonic distortion of the recovered audio. This is shown in part (d).

In multiplex stereophonic broadcasting the phase, frequency, and amplitude of the reinserted carrier are carefully controlled to maintain distortionless output.

Suppressed-carrier systems (such as are found in FM multiplexing)

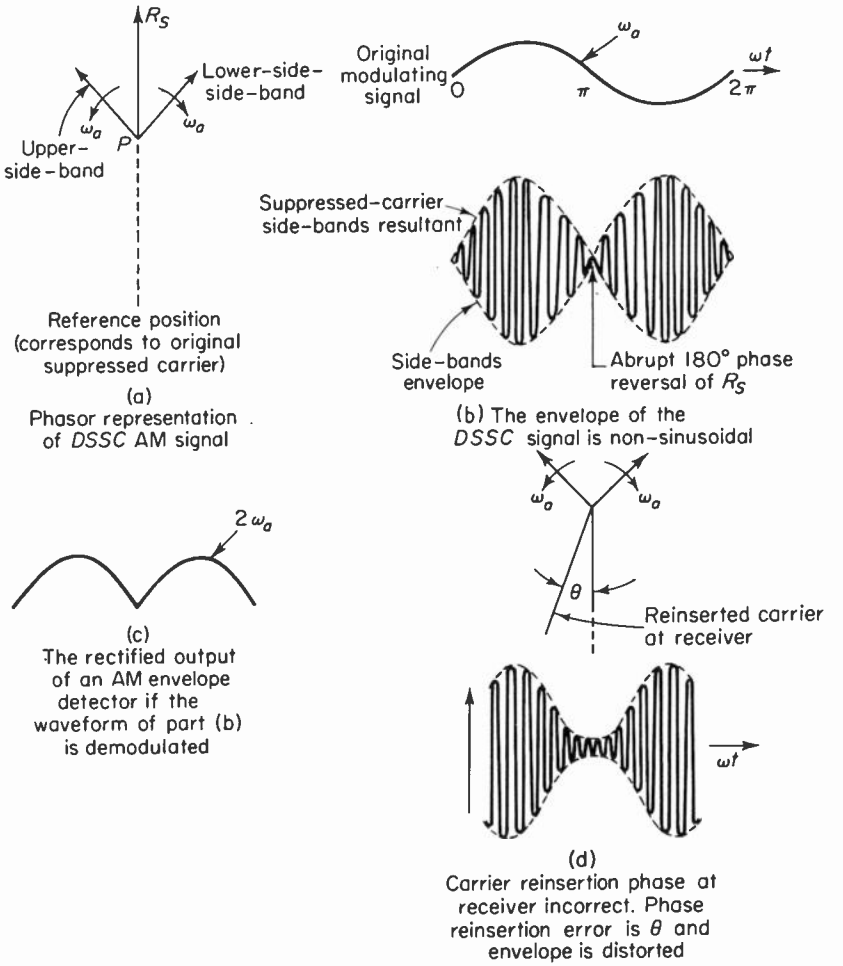



**Fig. 1-7.** Phasor and waveform conditions for double-sideband suppressed carrier transmission. When a locally generated carrier is added to the sidebands resultant of part (b), the waveform resultant of Fig. 1-6 is obtained and can properly be demodulated.

generally make use of a low-amplitude pilot carrier, which is "radiated" along with the main carrier. Its purpose is to control both the phase and frequency of the locally generated carrier so that the proper demodulation of the sidebands can take place. Two important reasons for suppressing the carrier in FM multiplexing are (1) more power can be placed in the sidebands and thus a larger effective modulation factor is realized and (2) the amplitude-handling capabilities of the transmitter can be used more effectively (see Sec. 14-2, "Interleaving or Nesting," for a more detailed discussion on this last point).

## 1-8. Frequency modulation

Frequency modulation (FM) is broadly defined as the process of varying the *frequency* of a carrier in accordance with the modulating signal and maintaining the *amplitude* constant. It will be shown that the *phase* of the carrier also varies with a change in frequency, but we need not concern ourselves with this notion at this point.

Assume that the carrier frequency is 100 MHz and that the audio modulating frequency is 2000 Hz. The *rate* at which the carrier frequency will vary or deviate above and below 100 MHz is then 2000 MHz. The rate of change of deviation of the carrier is thus the audio rate.

The *amount* of frequency deviation above and below the center or rest frequency (100 MHz in this case) is a function of the *amplitude* of the audio modulating voltage. The amount of change or deviation of the carrier frequency is thus a measure of the audio amplitude.

The reader should thoroughly understand the previous two paragraphs, since the distinction between the concepts of rate of change and amount of change is basic to the idea of modulation in general. Perhaps an examination of Fig. 1-8 will help to clarify the point. The audio voltage is assumed to have an amplitude of 1 volt and a frequency of 2000 Hz. The waveform below the audio voltage is the RF carrier and it is assumed to be (initially) 100 MHz. Between 0° and 90° of the audio cycle, the instantaneous frequency of the carrier deviates upward and reaches a maximum frequency of 100.050 MHz (100 MHz + 50 kHz). This will be referred to as a *positive deviation*. Between 90° and 180°, the instantaneous frequency of the carrier deviates in a downward direction until at 180° the carrier frequency is once again equal to the center frequency. Between 180° and 270° the carrier frequency deviates downward to its minimum value of 99.95 MHz. This will be referred to as a *negative deviation*. The total carrier *swing* is thus equal to twice the deviation.

The time it takes for the RF carrier to go through one complete swing is $\frac{1}{2000}$ second, and therefore the rate of change of deviation is 2000 times per second. If we assume that the deviation (amount of change) is linearly proportional to audio amplitude, then a 2-volt peak audio should produce a deviation of 100 kHz. Alternately, a 0.5-volt peak audio voltage will produce a deviation of 25 kHz.

A logical question at this point is: "How far above and below the center frequency is the RF carrier going to be allowed to deviate?" From a theoretical viewpoint, the carrier deviation could equal the center frequency and thus the instantaneous frequency limits would be between 0 Hz and 200 MHz for a 100-MHz carrier. To go beyond these limits would have no meaning, since, physically, there are no negative frequencies.

In commercial practice an upper limit to the frequency deviation is set by the Federal Communications Commission (FCC) and is a compromise

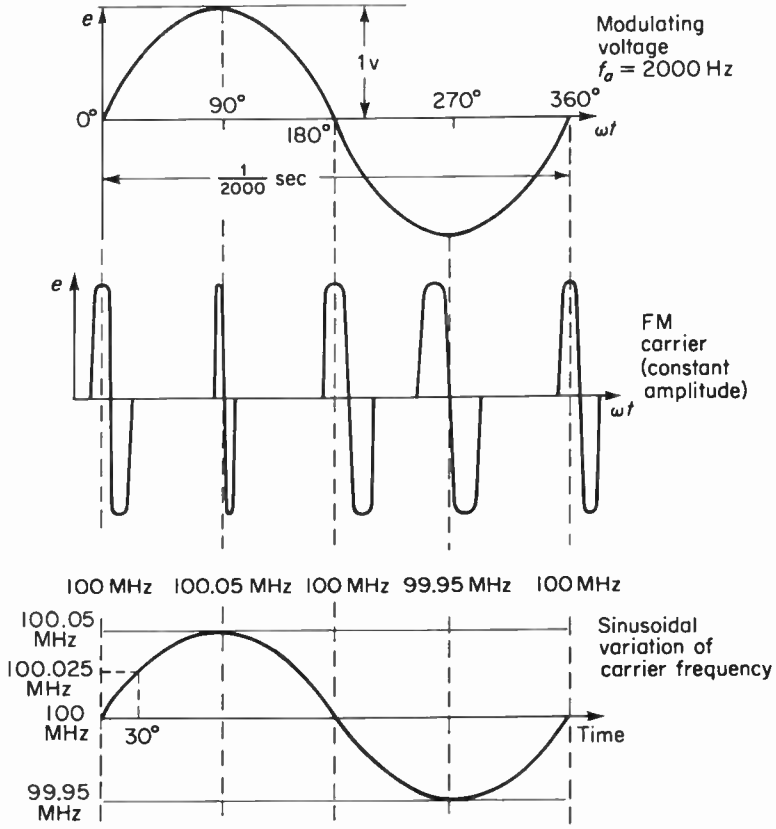



**Fig. 1-8.** Frequency deviation of a 100 MHz carrier.

between two important factors: signal-to-noise ratio and transmission-bandwidth requirements. The deviation should be as high as possible to obtain a high signal-to-noise ratio and as low as possible to reduce the transmission bandwidth. The value chosen for commercial broadcast service is 75 kHz, referred to as the "system deviation." A deviation of 75 kHz corresponds to 100 per cent modulation, while 33.333 per cent modulation corresponds to a carrier deviation of 25 kHz. Other systems use lower values (the system deviation for broadcast *television sound* is 25 kHz).

Since the carrier can deviate 75 kHz on either side of the center frequency, large bandwidths (compared to AM) can already be envisioned. Each station is allocated a bandwidth of 200 kHz, with 150 kHz employed for modulation and 50 kHz as a guard-band. The reason for providing a guard-band is to prevent audible "beats" between sidebands of two adjacent carriers, particularly if one of the two stations is slightly off-frequency. Bandwidth allocation is shown in Fig. 1-9.

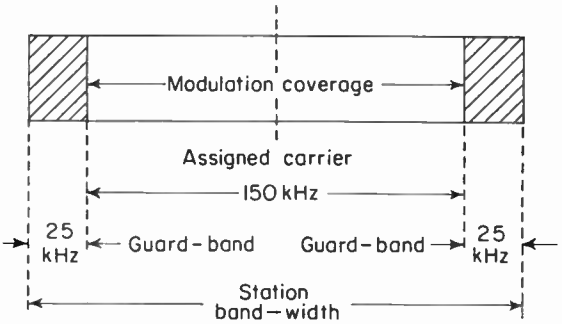
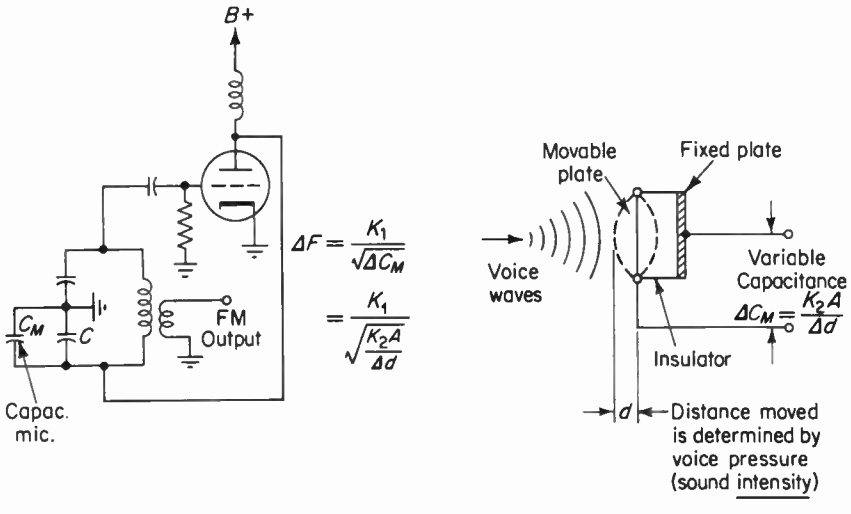



**Fig. 1-9.** FM station bandwidth allocation.

## 1-9. Simple method of generating an FM carrier

Figure 1-10 shows a conventional Colpitts oscillator circuit. A capacitor-type microphone is connected across one of the tuning capacitors *C*. This




(a)

Colpitts FM oscillator using capacitor microphone.

(b)

A capacitor microphone.

**Fig. 1-10.** A simple way to generate frequency modulation directly. The frequency of the oscillator output voltage varies directly with the intensity of the sound waves impressed upon the capacitor microphone.

is not the way a commercial FM transmitter generates an FM carrier,† but this fundamental approach will help one understand more complicated techniques. A capacitor microphone provides a change in capacitance proportional to voice pressure (sound intensity). The voice frequency determines the rate at which the capacitance varies.

As $C_m$ varies voice pressure, the total tuning capacitance of the oscillator changes, and thus the oscillator frequency varies above and below its center value (say 100 MHz). The amount of change on either side of 100 MHz is proportional to voice pressure, while the rate of change of frequency deviation is equal to the voice frequency. Figure 1-11 shows a plot of instan-

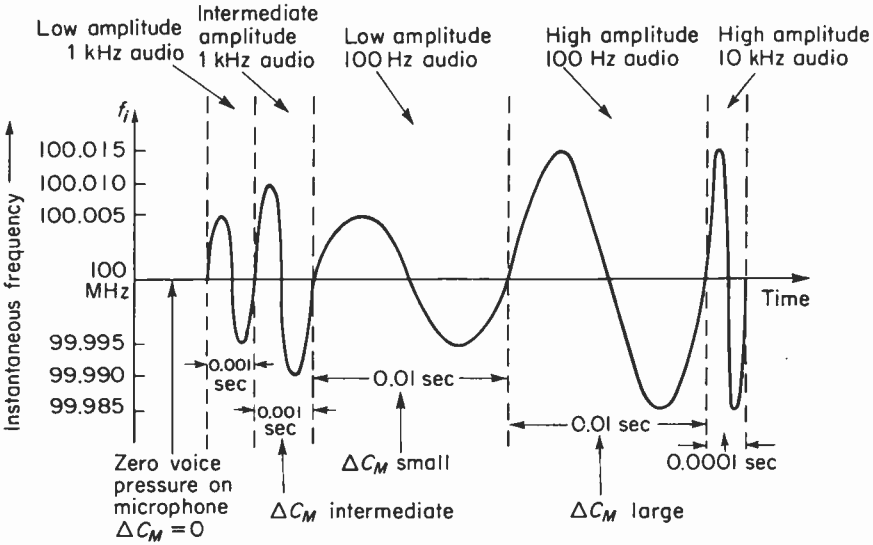



**Fig. 1-11.** The greater the sound intensity, the greater will be the change in microphone capacitance and the greater will be the carrier frequency deviation. The higher the audio frequency, the greater will be the rate of change of deviation.

taneous oscillator frequency versus time for different values and combinations of voice pressure and voice frequency. Small changes in microphone capacitance produce small carrier-frequency deviations, while large ones produce large swings. Low audio frequencies cause the capacitance to change slowly to produce small rates of change of deviation. As an example, a 2 per cent change in capacitance across the microphone produces a percentage change in oscillator frequency of

†Often, if not always, the microphone must be located a considerable distance from the transmitter. This would require long connecting cables, thus making the capacitor microphone arrangement untenable.

$$\text{per cent } \Delta F = \frac{1}{2\pi\sqrt{L(1.02)}}$$
$$= 1 \text{ per cent (approximate).}$$

Thus, the change in frequency is not absolutely linear with respect to changes in tuning capacitance (or voice pressure).

## 1-10. The reactance-tube modulator

An FM carrier may also be generated by means of a circuit that presents a variable reactance to an oscillator circuit. If this reactance is varied in accordance with the modulating signal, then once again the frequency of the oscillator will deviate above and below the center frequency. The basic circuit of the reactance tube is shown in Fig. 1-12. Assume that the oscillator is operating and that a voltage $e_o$ appears across the tank. That same voltage is applied between plate and cathode of the tube as well as across the $RC$ phase-shift network. This voltage produces a network current $i_n$. The network current produces two voltage drops that are at right angles to each other. Since $R \ll 1/j\omega C$, we can assume that $e_r \ll e_c$. The network current is in phase with $e_r$ and at right angles to $e_c$. The phasor addition of $e_r$ and $e_c$ produces $e_o$. The phasor addition of $i_n$ and $i_p$ produces $i_t$. Also, the plate current and grid voltage are in phase. The complete phasor diagram for the circuit is shown in part (c) of the same figure.

Notice that the impedance as "seen" by the oscillator tank is the ratio of $e_o$ to $i_t$. Since $e_o$ *lags* $i_t$ by almost 90°, the tuned circuit effectively sees a capacitive reactance. Thus a capacitor is "injected" across the oscillator tuned circuit. If this capacitance could be made to vary, the frequency of the oscillator would likewise change. For the time being, it can be said that the injected reactance is inversely proportional to the total current and that this total current is made up of two components ($i_p$ and $I_n$). If an audio modulating signal is applied to the grid of the reactance tube, then $i_p$ will change and this will change the equivalent capacitance injected across the tuned circuit, bringing about a change in oscillator frequency. This circuit requires compensation to produce a linear deviation, since the change in oscillator frequency is proportional to the inverse square root of the change in injected capacity.

One way to vary the plate current of the reactance-tube modulator is to vary the $g_m$, since the plate current in a pentode is $g_m e_g$. Since $R \ll 1/j\omega C$, the network current is determined almost entirely by $1/j\omega C$, or

$$i_n = e_o j\omega C.$$

The grid voltage is

$$e_g = i_n R = e_o j\omega RC.$$

The plate current is

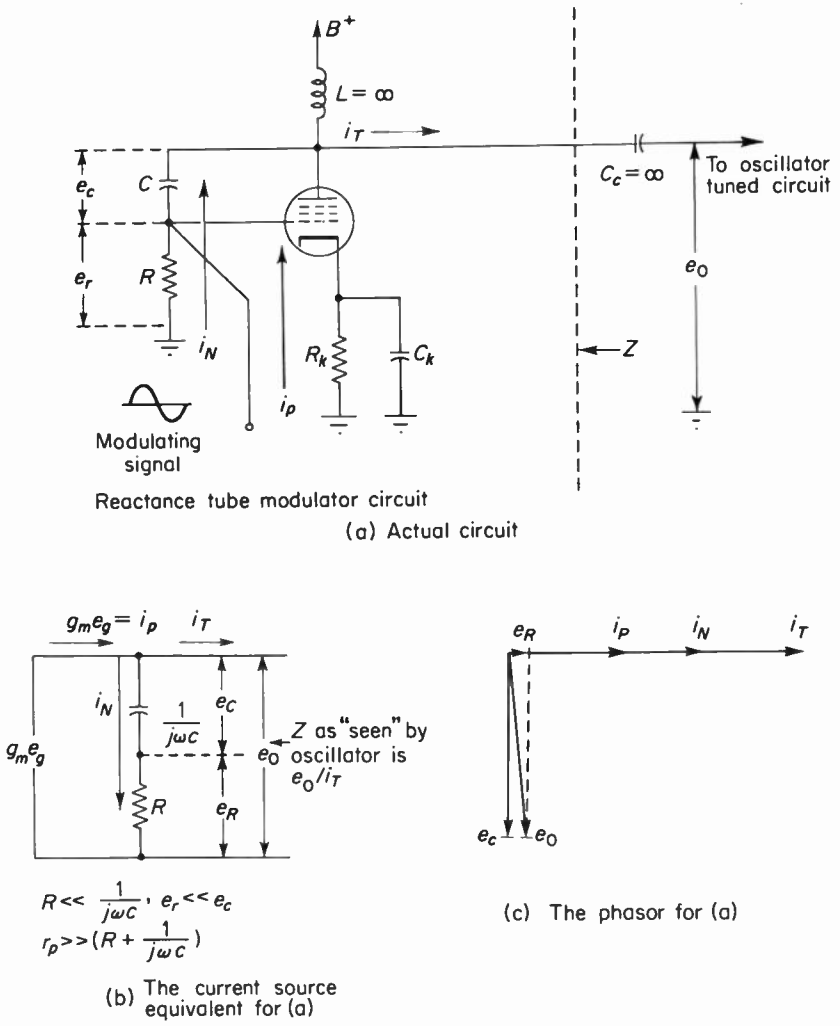



**Fig. 1-12.** The reactance tube modulator and corresponding phasor diagram. The phasor shows that a capacitance is "injected" across the oscillator tank since total voltage lags total current by approximately 90°.

$$i_p = g_m e_g = g_m e_o j\omega RC.$$

The impedance presented by the tube to the oscillator circuit is

$$Z_{\text{tube}} = \frac{e_o}{i_p} = \frac{e_o}{g_m e_o j\omega RC}$$

$$= \frac{1}{j\omega g_m RC}.$$

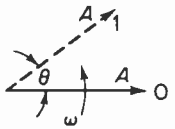
The term $g_mRC$ in the denominator is equivalent to a capacitor† whose value can vary as a function of the transconductance of the tube. Thus, the injected capacitance placed across the oscillator tank is $C_i = g_mRC$. In order to vary the $g_m$ of a tube, the tube must be operated about a point on its transfer characteristic where considerable curvature exists. The cathode-bias resistor $R_k$ shown in Fig. 1-12 provides sufficient cathode bias to operate the tube in the nonlinear region of its static transfer curve.

Thus, the amplitude of the modulating signal applied to the grid determines the amount of change of $g_m$, which is turn determines the amount of change of $C_i$. Since $C_i$ is effectively across the oscillator tuned circuit, the amount of deviation above and below center frequency is proportional to audio amplitude, while the audio frequency determines the rate of change of deviation.

## 1-11. Frequency modulation by means of phase modulation (indirect FM)

Before we discuss another method of generating an FM carrier, let us consider the concepts of phase and frequency and how they are related.

Consider first a signal phasor $A$ rotating in the complex plane at a constant




$$\omega = \frac{d\theta}{dt} = \text{constant}$$

**Fig. 1-13.** The angle $\theta$ locates the phasor $A$ in the complex plane. The angular velocity of the phasor remains constant so long as the phase angle changes at a constant rate.

angular velocity $\omega$, as shown in Fig. 1-13. The initial starting point is 0, and therefore the phase angle $\theta$ which locates the phasor in the plane is initially zero degrees. As the phasor rotates in a counterclockwise direction, it traces out an angle that changes at a constant rate. If the phasor, for example, sweeps out an angle of 360° or $2\pi$ radians in one second, then $\omega = 2\pi f = 2\pi$ rad/second and $f = (1/2\pi \text{ rad})(2\pi \text{ rad/second}) = 1$ Hz. Suppose the vector now sweeps out $1000\pi$ radians in one second. Then, $f = (1/2\pi \text{ rad})(1000\pi \text{ rad/second}) = 500$ Hz. Thus, in order to cause an increase in the angular velocity (and frequency) of the phasor, the time rate of change of phase angle must be increased. It can also be concluded that a decrease in the angular rate of change produces a decrease in frequency. Since the instantaneous

†The *tube* thus appears as a capacitive reactance. The magnitude of the capacitance is $g_mRC$.

frequency of the phasor is proportional to the rate of change of phase angle $\theta$, a fundamental expression can be written that relates angular rate of change and frequency or

$$f_i = \frac{1}{2\pi} \cdot \frac{d\theta}{dt} \cdot \dagger \tag{1-10}$$

If the phasor in Fig. 1-13 momentarily speeds up, then $\theta$ will change at a faster rate ($d\theta/dt$ is large) and the instantaneous frequency of the phasor increases. If the phasor momentarily slows down, then $\theta$ will change at a lower rate ($d\theta/dt$ is small) and the instantaneous frequency decreases. Therefore, a *change* in angular rate of change produces a change in instantaneous frequency. Also, a change in the instantaneous frequency produces a change in the instantaneous phase angle.

Assume now that the relative phase of a carrier is varied in accordance with a modulating signal. Instead of rotating the phasor, however, the complex plane (and axes) is rotated in the opposite direction at the same angular velocity.‡ This will permit a better examination of the modulation process. Figure 1-14 shows the phasor $A$ oscillating above and below point 0 in accordance with an audio signal. The audio *amplitude* determines the *amount* of phase deviation ($\Delta\theta$) on either side of the reference position (point 0) while the audio *frequency* determines the *rate of change* of phase angle or $d\theta/dt$. This process is called phase modulation.

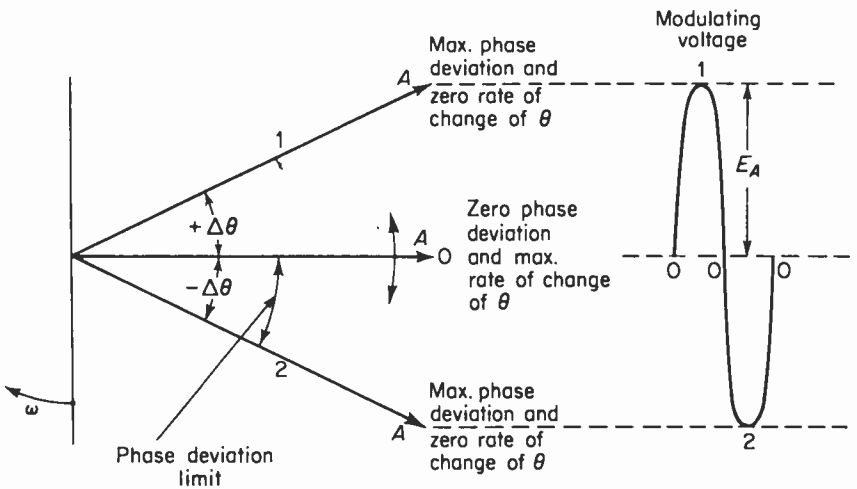



**Fig. 1-14.** A phase modulated carrier. The phase deviation limit (maximum angular excursion of the phasor) is proportional to $E_A$ while the rate of change of phase deviation is determined by the frequency of $E_A$.

†The *instantaneous radian frequency* is $\omega_i$, and $\omega_i = 2\pi f_i$. Radian frequency and angular velocity are synonymous.

‡The phasor thus appears to "remain stationary."

The phasor in Fig. 1-14 is therefore seen to alternately speed up and slow down in accordance with the phase deviations produced by the modulating signal. For example, at positions 1 and 2 the rate of change of phase angle is zero and the phase deviation is maximum. At position 0, the rate of change of phase angle is maximum and the phase deviation is zero. Since the angular velocity (or speed) of the phasor is related to the carrier frequency ($\omega = 2\pi f$), the carrier frequency must change during the phase-modulation process. Thus, a form of frequency modulation comes about when the phase angle of a carrier is varied. Since it came about indirectly, we call it "indirect FM." Frequency shift and phase shift are always coexistent, and the instantaneous value of frequency shift (above and below the *average* frequency) is proportional to the instantaneous rate of change of phase angle.

At points 1 and 2 of Fig. 1-14 the instantaneous rate of change of phase angle is zero. This produces a zero frequency deviation. At point 0 the instantaneous rate of change of phase angle is maximum and the instantaneous frequency deviation is maximum. Figure 1-15 shows that frequency modulation (FM) and phase modulation (PM) are at right angles to one another.

Figure 1-15 also shows that the instantaneous phase angle is a function of the amplitude of the modulating signal $e_M$. The instantaneous frequency is determined by the rate of change of amplitude of the modulating signal. The rate of change of $e_M$ depends upon modulation frequency ($1/T_a$) and modulation amplitude ($E_M$). If $E_M$ is increased, the phase deviation is increased and the phasor must move through a larger angle in a given time. This means a greater speed and a greater frequency deviation. If the modulation frequency is increased, the phasor must sweep out the same angle in less time (the phasor must speed up) resulting in an increase in frequency deviation. Also, a decrease in either amplitude or frequency of the modulating signal will produce less frequency deviation. The indirect FM produced as a result of PM can be calculated from the following equation:

$$\Delta F = \Delta\theta f_a \cos \omega_a t, \tag{1-11}$$

where

$\Delta F$ = equivalent FM as a result of PM,

$\Delta\theta$ = maximum phase deviation (determined by audio amplitude),

$f_a$ = modulating signal frequency,

$\cos \omega_a t$ = amplitude variations of the modulating signal at any time $t$.

When $\omega_a t$ is 0 or $\pi$ radians, signal amplitude is maximum at $+1$ and $-1$. If the maximum phase deviation is 60° or $\pi/3$ radians and a 2000-Hz tone phase modulates the carrier, then

$$\Delta F = \frac{\pi}{3}(2000)(\pm 1)$$

$$= \pm 2094 \text{ Hz}.$$

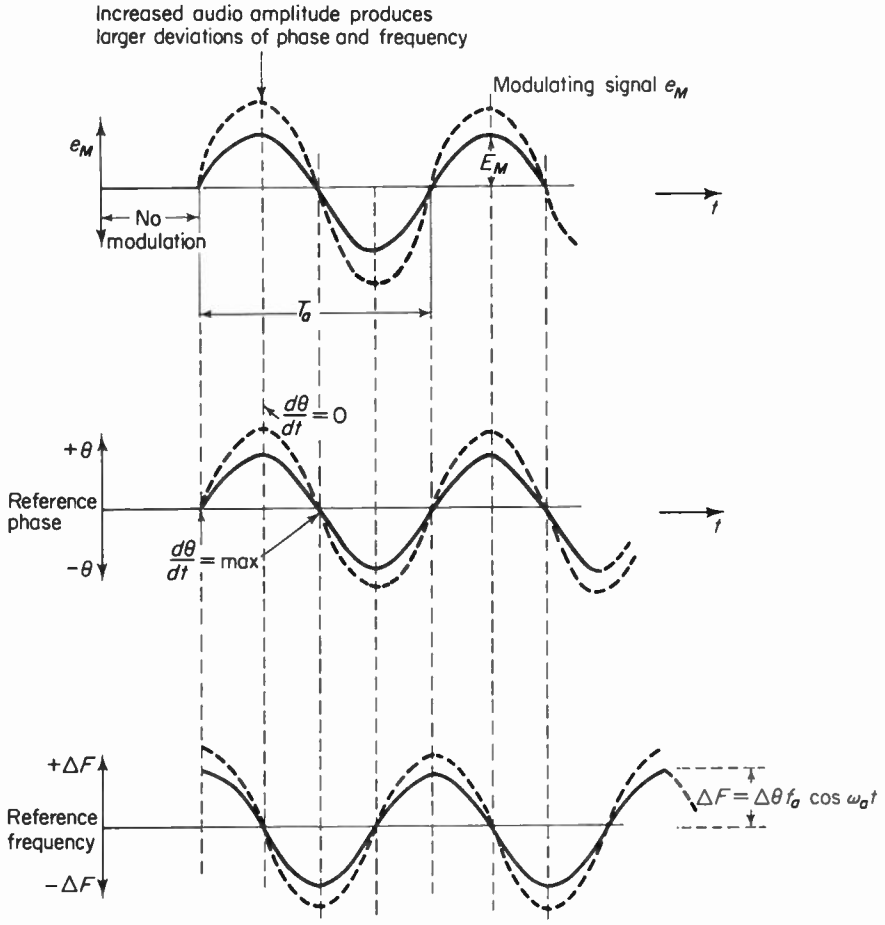



**Fig. 1-15.** The relationship between modulating signal amplitude, phase deviation and frequency deviation for a given modulating frequency of $1/T_a$.

For direct FM (the reactance-tube modulator, for example) the frequency deviation of the carrier was determined solely by the amplitude of the modulating signal. In PM we saw that the equivalent FM introduced by phase-modulating the carrier was determined not only by the amplitude but by the frequency of the modulating signal as well. In order to make PM identical to FM, we must make $\Delta F$ in Eq. (1–11) independent of the modulating frequency. In other words, if the modulating signal is fed into a low-pass filter whose output is inversely proportional to frequency, then the frequency deviations due to PM will be proportional to the phase deviation only. The system is now, in effect, a frequency-modulation system. This is shown in Fig. 1-16, with the audio correction network providing an output inversely

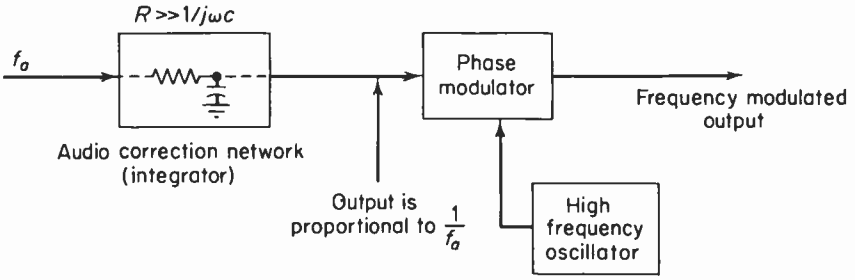



**Fig. 1-16.** An indirect FM system.

proportional to audio frequency. The output from the phase modulator is thus a true frequency-modulated output.

The simplest type of phase modulator is a crystal-controlled oscillator whose output is fed into a variable phase-shift network, such as shown in Fig. 1-17. The resistance is to be varied in accordance with a sinusoidal modulating signal; therefore the resistance variation is also sinusoidal. Part (b) of the same figure shows how the phase of the output voltage ($e_R$) varies between the phase deviation limits while the amplitude of the output also changes. This introduces undesirable AM into the output, but if we make $R \gg 1/j\omega C$, then the amount of AM introduced will be relatively small. Also, for a given size of $R$, the phase shift should not be too large or, again, the AM will be large. Distortion is held to within reasonable limits if the maximum phase deviation does not exceed about 25° or a little more than 0.5 radian.

Figure 1-18 shows how we might vary the resistance of the circuit shown in Fig. 1-17. We use a vacuum tube and vary its plate resistance by applying a modulating signal to the grid. The plate load resistor is made very large compared to $r_p$, so that the circuit to the right of point $P$ "sees" only $r_p$. This arrangement requires that the tube be operated at a point on its transfer characteristic where considerable curvature exists, in order to effect changes in $r_p$ with the applied grid voltage. A correction network is placed in the input circuit, so that the output is a frequency-modulated carrier as previously explained.

We may summarize the previous statements concerning phase and frequency modulation by stating that for FM, a phase shift is produced that is proportional to the instantaneous amplitude of the modulating signal. For PM, a phase shift is also produced that is proportional to the instantaneous *amplitude* of the modulating signal. However, the instantaneous phase shift as a result of frequency modulation is also inversely proportional to the *frequency* of the modulating signal. This is not the case for PM. This, then, constitutes the basic difference between PM and FM.

Thus, FM can be produced by "audio correcting" the modulating signal by means of a low-pass filter before application to a *phase* modulator (see

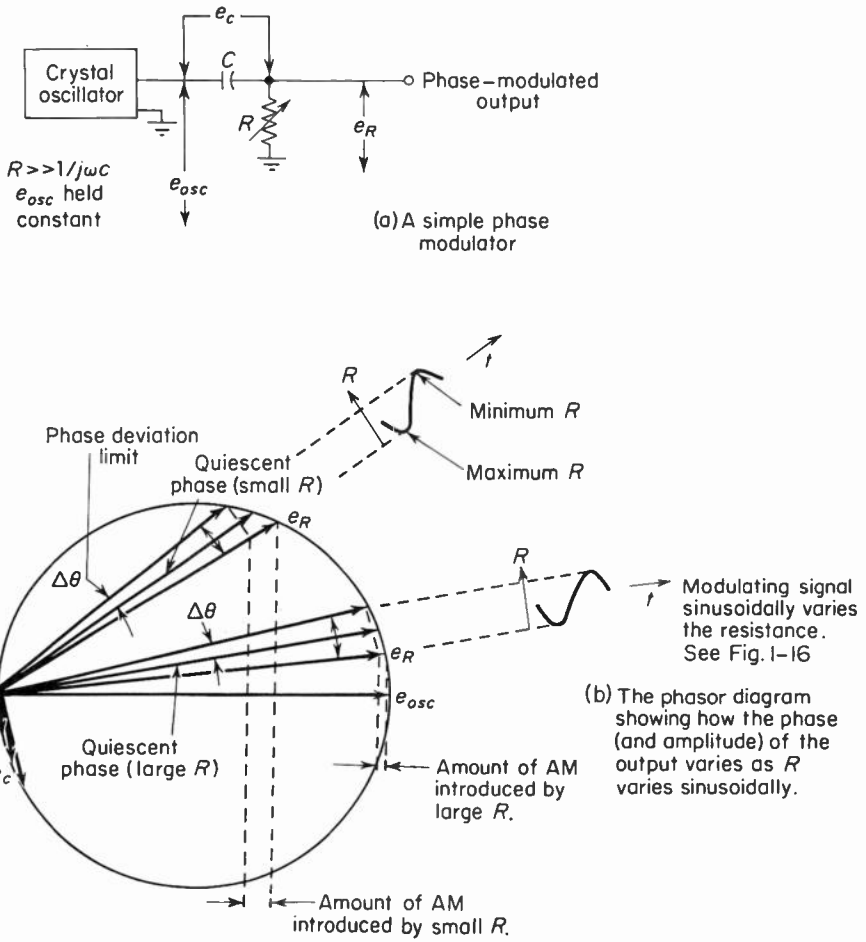



**Fig. 1-17.** A "simple" *R-C* network can produce phase modulation if we vary *R*.

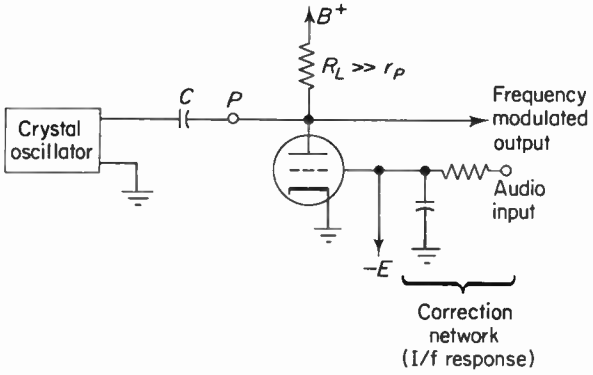



**Fig. 1-18.** A simple circuit for deriving indirect FM. The output is a narrow-band signal since the minimum phase deviation is limited to approximately .5 radian or 25°.

Fig. 1-18). On the other hand, PM can be produced by "audio-correcting" the modulating signal by means of a high-pass filter before application to a *frequency* modulator. This last provision is necessary in order to make the instantaneous frequency deviation proportional to the frequency as well as the modulating signal.

## 1-12. Frequency-modulation sideband structure

We saw in earlier discussions that when a carrier was amplitude-modulated, a pair of sidebands per audio modulating signal was produced. The sideband separation was determined by the modulating frequency, and this determined the system bandwidth. In FM (as well as PM), sidebands are also generated but they are much more complicated. It will be shown that for each sinusoidal modulating signal an infinite number of sidebands are produced. These sidebands are displaced from the carrier and are spaced at integral multiples of the modulating-signal frequency.

Now, if the intelligence conveyed by the modulated carrier is to be reproduced without distortion, the sidebands must be transmitted through the *entire* system so that their relative amplitudes, phases, and frequencies remain intact. The system bandwidth must be adequate so that all sidebands can be accommodated. This requires a knowledge of the sideband distribution, especially for frequency modulation because of the large number of sidebands as well as their complex nature.

We shall begin the discussion by setting down the equation for a frequency-modulated carrier as

$$e_c = E_c \sin\left(\omega_c t + \frac{\Delta F_c}{f_a} \sin \omega_a t\right), \tag{1-12}$$

where

$E_c$ = the peak amplitude of the radiated wave (this remains constant);

$\omega_c = 2\pi f_c$, where $f_c$ is the center frequency of the carrier;

$\Delta F_c$ = the frequency deviation to one side of the center frequency;

$\omega_a = 2\pi f_a$, where $f_a$ is the audio frequency that modulates the carrier and $f_a$ can be any frequency between 50 Hz and 15,000 Hz.

Using the trigonometric identity, $\sin(A + B) = \sin A \cos B + \cos A \cdot \sin B$ and letting $\omega_c t = A$, $\Delta F_c/f_a \sin \omega_a t = B$, and $\Delta F_c/f_a = M$, then Eq. (1–12) is

$$e_c = E_c[\sin \omega_c t \cos(M \sin \omega_a t) + \cos \omega_c t \sin(M \sin \omega_a t)]. \tag{1-13}$$

This may be written in the form (using Bessel functions of the first kind)†

†See A. Hund, *Frequency Modulation* (New York: McGraw-Hill Book Company, 1942), appendix, pp. 350–351.

$$e_c = E_c\{J_0(M)\sin\omega_c t + J_1(M)[\sin(\omega_c + \omega_a)t - \sin(\omega_c - \omega_a)t] + J_2(M)[\sin(\omega_c + 2\omega_a)t + \sin(\omega_c - 2\omega_a)t] + J_3(M)[\sin(\omega_c + 3\omega_a)t - \sin(\omega_c - 3\omega_a)t] + J_4(M)[\sin(\omega_c + 4\omega_a)t + \sin(\omega_c - 4\omega_a)t] + \ldots\}. \quad (1\text{–}14)$$

Equation (1–14) is a Bessel function of the first kind, order $n$, and argument $M$. The order $n$ refers to the subscript of the letter $J$ and corresponds to the number of the sideband pair. For example, $J_3$ in the *fourth* term refers to the *third* sideband pair. The argument $M$ requires that we define a new expression relating the frequency deviation ($\Delta F_c$) to the audio frequency ($f_a$) that produces that deviation, or

$$M = \frac{\Delta F_c}{f_a}. \quad (1\text{–}15)$$

$M$ is called the modulation index; it will be shown that this ratio determines the required transmission bandwidth of the system. The modulation index will range anywhere between zero ($\Delta F_c = 0$) and 1500 (75 kHz/50 Hz) and will, of course, continuously change when program material is modulating the carrier. Before we continue with the interpretation of Eq. (1–14), let us consider a very special case of the modulation index—the case where we divide the maximum permissible deviation (previously defined as the system deviation) by the maximum audio frequency. For wideband FM (commercial broadcasting) the system deviation is specified as 75 kHz and the maximum audio frequency as 15 kHz. Thus, the modulation index in this example is 75/15 or 5. This special case of the modulation index is called the "deviation ratio" and is defined as the ratio of the system deviation to the maximum audio frequency, or

$$D = \text{deviation ratio} = \frac{\Delta F_c\,(\text{maximum})}{f_a\,(\text{maximum})}. \quad (1\text{–}16)$$

For television sound, the system deviation is 25 kHz and the maximum audio frequency is 15 kHz. The deviation ratio for television sound transmission is therefore 25,000/15,000 or 1.665. In the next chapter we shall see that the noise improvement obtainable in FM (over AM) is directly proportional to the square of the deviation ratio, so that this is an important quantity.

We have just finished defining the modulation index $M$ for frequency modulation. Notice that this numeral depends upon audio-modulating amplitude as well as frequency. Phase modulation (PM) also contains a modulation index ($M_p$) and is defined as the maximum phase deviation of the carrier from its initial or quiescent phase position. This corresponds to the maximum angular excursion of the phasor $A$ shown in Fig. 1-14 or $\pm\Delta\theta$. This modulation index, being an angular quantity, has the unit of radians, and its magni-

tude is determined by the amplitude of the modulating signal. Thus, for a fixed-amplitude modulating signal fed to a phase modulator, the peak phase deviation (being independent of modulating frequency) will be the same at 100 Hz as at 10,000 Hz. Now, for a given modulating-signal amplitude and frequency, the values of modulation index are the same for both PM and FM and also have the same radian measure.† Thus, a frequency-modulated carrier of $M = 5$ (radians) produces a phase deviation of $\Delta\theta = \pm 286.5°$. For example, compute the maximum phase shift in degrees for an FM carrier modulated by a 70-Hz signal. The frequency deviation is 70,000 Hz. The modulation index $M = \Delta F_c/f_a = 70{,}000/70 = 1000$ radians. The peak phase shift in degrees is (57.3)(1000) = 57,000 degrees. Thus, many thousands of degrees of phase shift can occur in a "large-index" frequency-modulated signal.

We are now in a better position to interpret Eq. (1–14). An examination of this expression shows that a frequency-modulated carrier consists of an infinite number of sidebands located on both sides of the carrier component (represented by $J_0(M) \sin \omega_c t$) at integral multiples of the audio frequency. In other words, there are sidebands at $f_c \pm f_a$, $f_c \pm 2f_a$, $f_c \pm 3f_a$, and so on. The relative amplitude of these sidebands, as well as the carrier component, is determined by their respective Bessel coefficients.‡ These coefficients are obtained from published tables, and their values vary depending upon the modulation index chosen. It should also be noted that for sideband frequencies corresponding to odd multiples of the modulating frequency on either side of the carrier, the upper and lower sidebands are of opposite phase. That is, $J_1(M) \sin (\omega_c + \omega_a)t$ is the first-order (odd) *upper* sideband of relative amplitude $J_1(M)$ and frequency $f_c + f_a$. The first-order (odd) *lower* sideband of relative amplitude $J_1(M)$ and frequency $f_c - f_a$ is $-J_1(M) \sin (\omega_c - \omega_a)t$.

The power in each sideband is proportional to the square of its associated Bessel coefficient.* The total power in the radiated carrier is equal to the phasor sum of the carrier power and the individual sidebands power. During the modulation process, energy from the carrier is placed into the sidebands and energy from the sidebands is shifted back into the carrier. Since the envelope amplitude remains constant (with or without modulation), the average power of a frequency-modulated carrier remains constant with or without modulation. In amplitude modulation, on the other hand, the power associated with the carrier component remains fixed while the *total* radiated power increases with increase in the percentage modulation. Thus, for a given transmitter power and given power-output tube, the frequency-modulated transmitter will provide more output and is therefore more efficient. This

†One radian is equal to 57.3 degrees.

‡The Bessel coefficients in Eq. (1–14) are $J_0(M)$, $J_1(M)$, $J_2(M)$, and so on.

*The Bessel coefficient represents a voltage or current.

increased efficiency results from the reduction in energy content of the carrier component in frequency modulation (when the modulation is "on"), compared with the constant energy content of the carrier component in amplitude modulation.

In order to determine the carrier-component and sideband amplitudes for different modulaton indices, a series of curves (resembling damped sinusoids) is plotted for different indices on the abscissa versus carrier-component and sideband amplitudes on the ordinate. If we assume that the carrier amplitude [$E_c$ of Eq. (1–14)] is unity, then the relative carrier-component amplitude as well as the relative sideband amplitudes can be read off directly as a fractional part of the unit carrier. Figure 1-19 is a plot of Eq. (1–14) for modulation indices up to 15 and includes sidebands out to the twelfth order. This is sufficient for our purposes. In addition, only sidebands having amplitudes greater than 1 per cent of the unit carrier amplitude need be considered. These are designated as "significant," and we need include only these in making bandwidth calculations.

Let us now plot the frequency-modulation sideband spectra for a number of different modulation indices and modulating-signal frequencies. First, assume that the modulation index is zero. Figure 1-19 shows that for $M = 0$, all sideband-pair amplitudes are zero and the only term appearing is the unit carrier (that is, $J_0(M) = 1$). This is to be expected, since with $M = 0$, the carrier deviation is zero and no sidebands should be produced. This is shown in Fig. 1-20, with the carrier component shown as a unit quantity. Next, assume a modulation index of 5 and an audio frequency of 15,000 Hz. This corresponds to a frequency deviation of $Mf_a = (15{,}000)(5)$, or 75,000 Hz. The ordinate corresponding to $M = 5$ intersects the following sideband pairs: first, second, third, fourth, fifth, sixth, seventh, eighth, and ninth. The Bessel coefficient $J_9(M)$ for the ninth sideband pair is actually less than 0.01, so that this sideband pair can be considered as insignificant. The carrier component ($n = 0$) has an amplitude of 0.1776 and is negative for $M = 5$. Since we are concerned only with relative amplitudes, this polarity reversal can be disregarded when we plot the sideband spectrum. The relative amplitudes of the eight significant sideband pairs are:

| | |
|---|---|
| first: | 0.327 |
| second: | 0.046 |
| third: | 0.365 |
| fourth: | 0.391 |
| fifth: | 0.261 |
| sixth: | 0.131 |
| seventh: | 0.053 |
| eighth: | 0.018 |

These sidebands are shown plotted in Fig. 1-20(b). Now, Eq. (1–14) shows

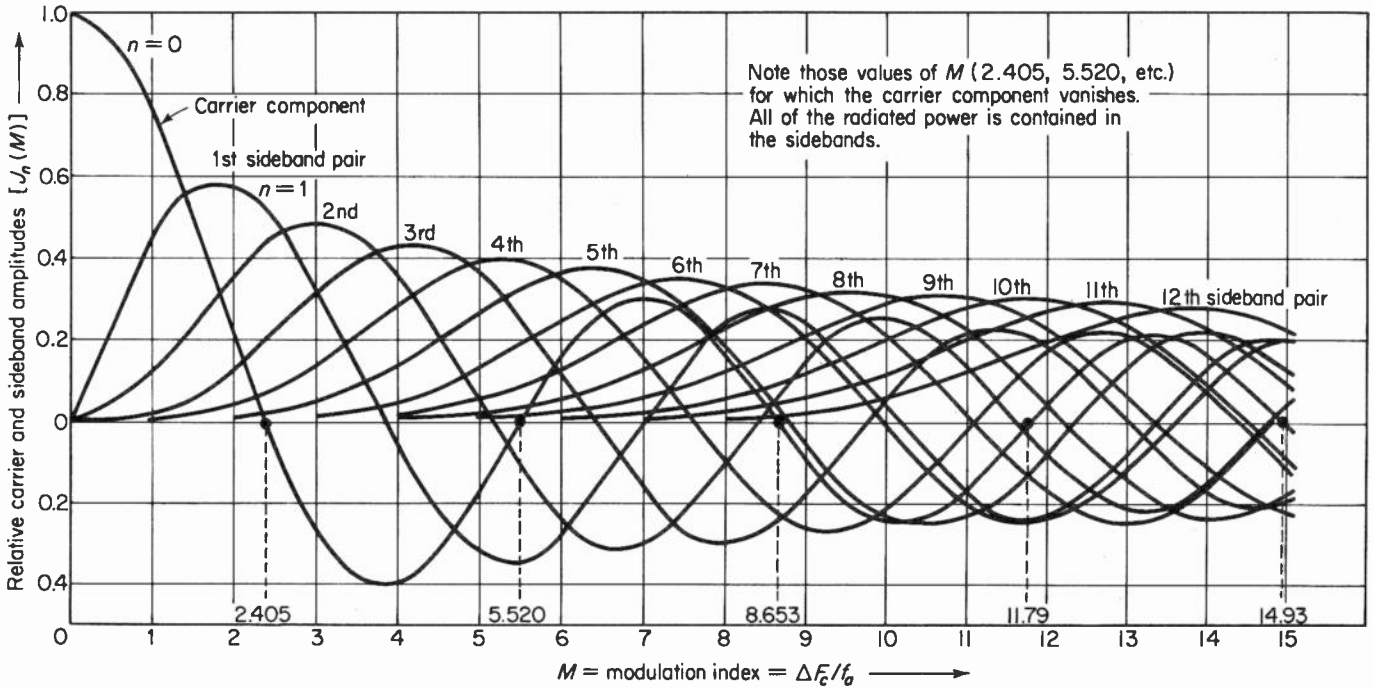



**Fig. 1-19.** The variation of carrier and sideband amplitude with modulation index. Only Bessel coefficients $[J_n(M)]$ greater than 0.01 are considered.

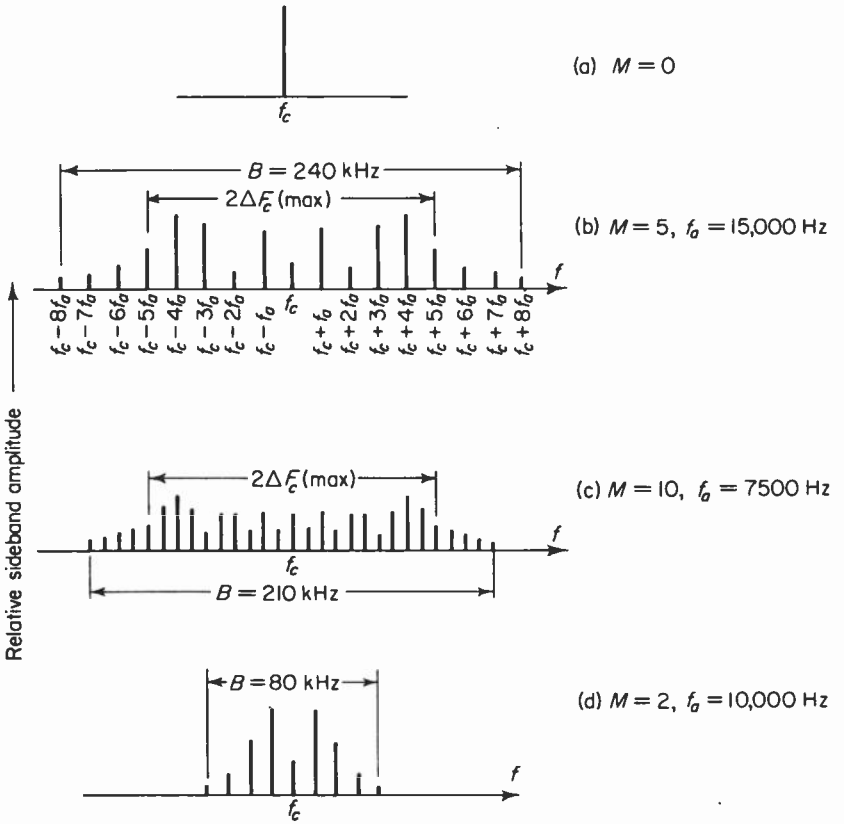



**Fig. 1-20.** The amplitude versus frequency spectra for four different conditions of modulation. Notice that for (b) and (c), the significant sidebands extend beyond the 150 kHz modulation limits.

that the sidebands are spaced at integral multiples of the modulating frequency removed from the carrier $f_c$. For a 15,000-Hz audio signal the first-order sidebands are at $f_c \pm 15{,}000$ Hz, the second-order sidebands are at $f_c \pm 30{,}000$ Hz, the third-order sidebands are at $f_c \pm 45{,}000$ Hz, and so on. Thus, the sideband *separation* is determined by the audio *frequency*, just as in amplitude modulation. The sideband amplitude is determined by the value of the modulation index, which is a function of both audio amplitude ($\Delta F_c$ is directly proportional to modulating-amplitude) and audio frequency.

It will be seen from Fig. 1-19 that the number of significant sidebands (*voltage* amplitudes greater than 1 per cent of the unmodulated carrier) *increases* with the modulation index. Since $M$ is proportional to audio *amplitude* and inversely proportional to audio *frequency*, it can be seen that for a constant audio-signal amplitude, the number of significant sidebands decreases with an increase in the audio frequency. The bandwidth occupied

by the significant sidebands in Fig. 1-20(b) is 240 kHz if all sixteen significant sidebands are to be accommodated by the transmission (and reception) medium. Thus, bandwidth occupancy for the significant sidebands in frequency modulation is equal to the total number of significant sidebands, $n_t$, multiplied by the audio frequency $f_a$, or

$$\text{bandwidth} = B = n_t f_a. \tag{1-17}$$

As the number of significant sidebands is reduced with increasing audio frequency, the bandwidth occupancy tends toward a constant value. The reason is that the separation increases but the modulation index decreases, and fewer significant sidebands are produced. Thus, there are simply fewer significant sidebands with greater separation between them.

The amplitude-frequency spectrum for two additional cases is shown in Fig. 1-20(c) and (d). Part (c) corresponds to a modulation index of 10 and an audio frequency of 7500 Hz. The curves of Fig. 1-19 indicate fourteen significant sideband pairs, each spaced at 7500-Hz intervals. The required bandwidth is thus

$$B = n_t f_a = (28)(7500) = 210 \text{ kHz}.$$

Figure 1-21 is a plot of bandwidth occupancy for a given deviation-versus-modulation index. Notice that the bandwidth occupied by the significant sidebands increases with increase in the modulating frequency ($f_a$). The reason is that for a given carrier-frequency deviation ($\Delta F_c$), the higher the modulating frequency, the greater will be the sideband spacing on both

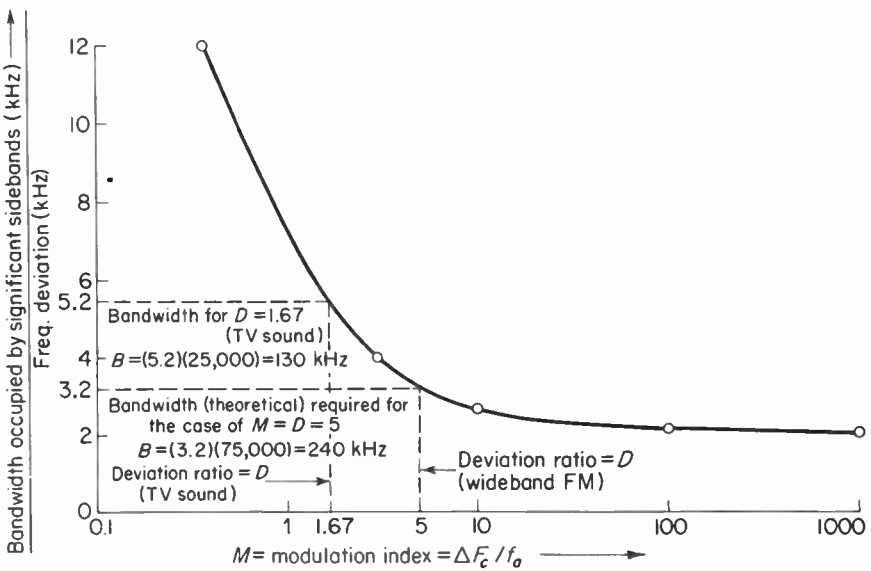



**Fig. 1-21.** Bandwidth occupancy of significant sidebands vs. modulation index.

sides of the carrier component. Thus, the *highest* modulating frequency determines the bandwidth. Notice, also, that for large modulation indices ($M = 10$), the bandwidth occupancy approaches $2\,\Delta F_c$ (max) or 150 kHz. Therefore, a large modulation index makes for more efficient use of the allotted spectrum. Notice that the deviation ratio for a given system sets the required bandwidth, since any other combination of deviation and modulating frequency will always produce a smaller bandwidth requirement—assuming that the maximum limits are not exceeded. For example, if we assume a carrier deviation of 75 kHz, then any audio frequency less than 15 kHz will result in a higher modulation index, and this will require a smaller bandwidth. This is because the sideband spacing on both sides of the carrier gets smaller, so that the sidebands begin to cluster near the carrier component. If, on the other hand, we assume a maximum modulating frequency of 15 kHz and reduce the carrier deviation below the maximum 75 kHz, then the number of significant sidebands diminishes and there are simply fewer sidebands spaced at 15 kHz on either side of the carrier component.

Suppose the modulation index is 4 and the carrier deviation is 50 kHz. To what bandwidth does this correspond? Using the curve shown in Fig. 1-21, the bandwidth occupied by the significant sidebands ($B$) divided by the 50-kHz carrier deviation is (from the curve) 3.5. The bandwidth ($B$) then is (50 kHz)(3.5) = 175 kHz.

Thus, typical IF bandwidths in commercial FM broadcast tuners (and radios) range from 150 to 200 kHz, the lower limit being 150 kHz. In addition to accepting the sidebands for minimum distortion, some tolerances should be provided to allow for oscillator drift and mistuning.

## 1-13. The phasor representation of a frequency-modulated carrier

Reexamination of Eq. (1–8) for the amplitude-modulated carrier shows that each sideband is preceded by a plus sign and that the phasor diagram for this wave (shown in Fig. 1-5) produced a sideband resultant ($R_s$) that was always collinear with the carrier phasor. If the same reasoning is used for the frequency-modulated equation (1–14), it is seen that the even-ordered sidebands [$J_2(M), J_4(M), \ldots$] all have the *same* sign (plus) while the odd-ordered sidebands have *different* signs; that is, the *upper* odd-ordered sidebands are preceded by plus signs while the *lower* odd-ordered sidebands are preceded by minus signs. The physical significance of this distinction is that the lower odd-ordered sideband phasors must be reversed from their "normal" position.† The first-order sideband phasors are shown in Fig. 1-22 together with the carrier phasor, which is labeled $J_0(M)$. The sideband phasors are labeled

†See Appendix 1-1.

$J_1(M)_L$ and $J_1(M)_u$ for lower and upper sidebands, respectively. Notice that this will produce a frequency-modulated signal (of the narrow-band variety, since we are including only two sidebands) in addition to some amplitude modulation of the resultant—virtually the situation shown in Fig. 1-14. The reason for the variation in carrier amplitude is that we are only considering one sideband pair. In wideband FM there will be many sideband pairs, all producing a resultant carrier whose amplitude is constant.

The foregoing situation can be summarized in a general way: all odd-ordered sidebands produce a resultant $e_{ro}$ that is in quadrature with or at right angles to the carrier component $e_c$, while the even-ordered sidebands produce a resultant $e_{re}$ that is always collinear with the carrier component.

Let us now construct the phasors diagrams for two different instants in time and for a modulation index of 3. From the curves given in Fig. 1-19 we see six significant sideband pairs. The Bessel coefficients are as follows: $J_0(M) = -0.26$ (carrier component), $J_1(M) = 0.339$ (first-order sideband), $J_2(M) = 0.486$ (second-order sideband), $J_3(M) = 0.309$ (third-order side-

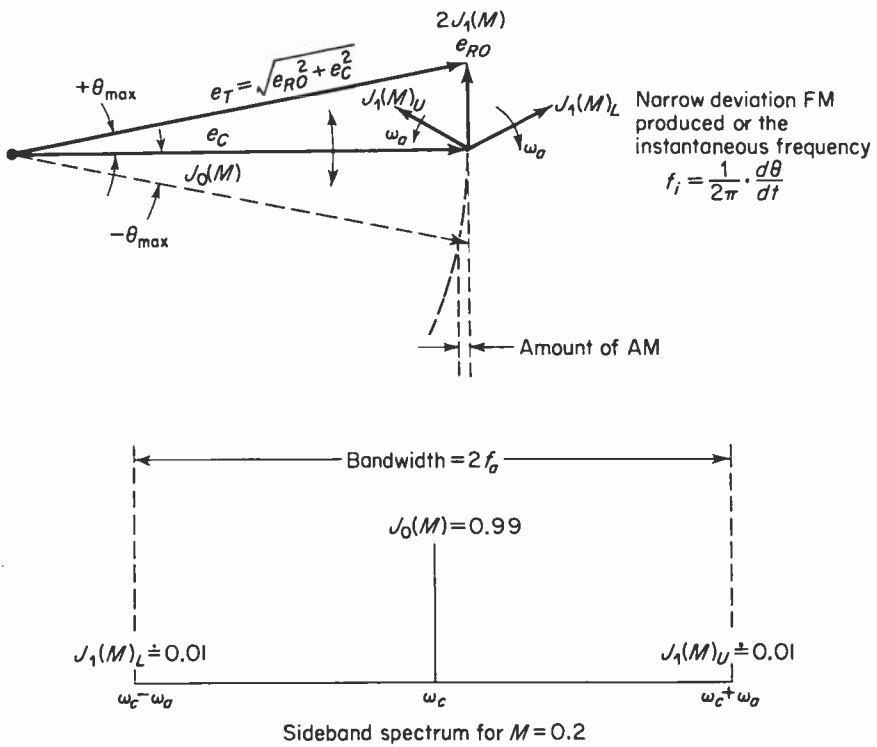



**Fig. 1-22.** A narrow-band FM signal ($M = 0.2$) is produced by letting the first-ordered sideband pair resultant ($e_{R0}$) phase modulate the resultant $e_T$. The sideband spectrum is also shown. How does the above situation differ from "conventional" AM?

band), $J_4(M) = 0.132$ (fourth-order sideband), $J_5(M) = 0.043$ (fith-order sideband) and $J_6(M) = 0.011$ (sixth-order sideband). If we assume that the carrier phasor $J_0(M)$ acts as a stationary reference, then all sidebands can be represented by phasors that rotate at angular speeds equal to the difference between the carrier velocity ($\omega_c$) and the corresponding sideband velocities $\omega_c \pm \omega_a$, $\omega_c \pm 2\omega_a$, $\omega_c \pm 3\omega_a$, . . .). For example, the first-ordered lower sideband $J_1(M)_L$ rotates in a clockwise direction at an angular velocity of $\omega_a$, the first-ordered upper sideband $J_1(M)_U$ rotates in a counterclockwise direction at a velocity of $\omega_a$, the second-order sideband (lower) $J_2(M)_L$ rotates in a clockwise direction at a velocity of $2\omega_a$, the second-ordered sideband (upper) $J_2(M)_U$ rotates in a counterclockwise direction at a velocity of $2\omega_a$ and so on.

The instant in time we will assume for the first vector is when all sideband vectors are as shown in Fig. 1-23 (a). Notice that all odd-ordered sideband vectors are in phase opposition, so that the vector resultant of $J_1(M)$, $J_3(M)$, and $J_5(M)$ is zero. All odd-ordered sideband pairs thus cancel.

The even-ordered sideband pairs add to give a resultant

$$\begin{aligned} e_{re} &= 2J_2(M) + 2J_4(M) + 2J_6(M) \\ &= 0.972 + 0.264 + 0.022 \\ &= 1.258. \end{aligned}$$

From this we must subtract the $J_0(M)$ carrier component, which has a Bessel coefficient of $-0.26$. The overall resultant $e_r$ is thus 0.998. The overall resultant should total unity, since we have been assuming a unit carrier; but since we have not included any sideband pairs higher than the sixth, a small amount of amplitude modulation is present, which we will neglect.

Let us now assume that the first-ordered sidebands pair has rotated through an angle of $\pi/2$ radians. Since the angular velocity of each pair is proportional to its sideband order $n$, an $n$th-ordered sideband should be rotated through $n\pi/2$ radians. For example, the second-ordered sidebands are rotated (in opposite directions) through an angle of $\pi$ radians, the third-ordered sidebands are rotated (in opposite directions) through an angle of $3\pi/2$ radians, and so on. Figure 1-23(b) shows the final positions of all six significant sideband pairs with their corresponding relative amplitudes (Bessel coefficients). We first obtain the sum of all odd-ordered sidebands $e_{ro}$ to obtain a resultant which is in quadrature with the carrier component $J_0(M)$. We then obtain the sum of all the even-ordered sidebands $e_{re}$, which is collinear with the carrier. $e_{ro}$, $e_{re}$, and $e_c$ are then combined into a right triangle to obtain the final overall resultant $e_t$, which should total unity. The slight difference is due to the omission of the higher-ordered sidebands. The final phasor is shown in Fig. 1-23(c). Notice that the odd-ordered sidebands produce phase modulation (and frequency modulation) and some amplitude modulation, while the even-ordered sidebands produce only ampli-

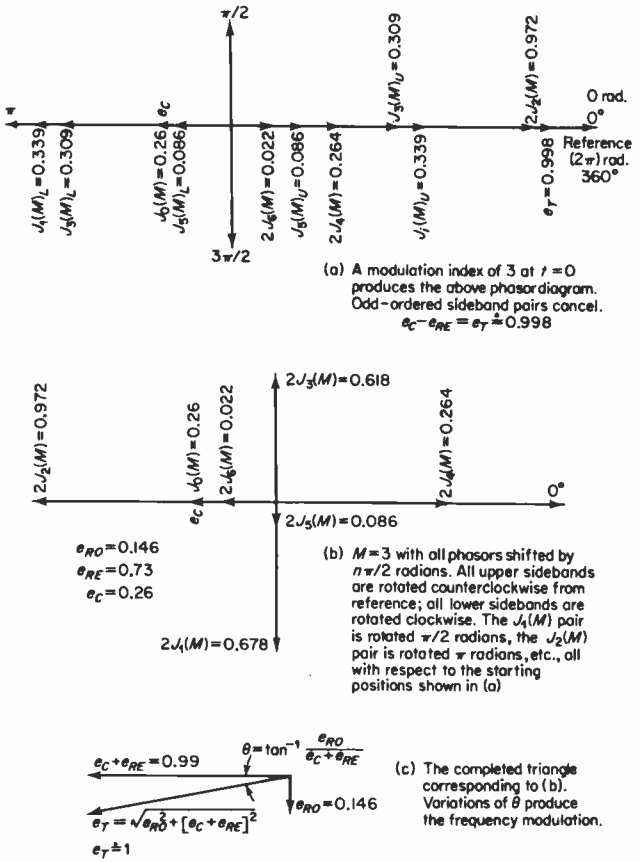



**Fig. 1-23.** The phasor construction for a frequency modulated carrier for $M = 3$.

tude modulation of the resultant. The variation of the phase angle $\theta$ produces the frequency modulation.

Since frequency modulation produces no change in carrier amplitude (assuming all sidebands are included), we can say that the root-mean-square value of the carrier component $J_0(M)$ and the sideband components $J_n(M)$ must total the root-mean-square value of the radiated wave, or

$$[J_0(M)]^2 + 2\sum_{n=1}^{\infty}[J_n(M)]^2 = 1. \tag{1-18}$$

## 1-14. Summary

We may summarize some of the more important aspects of FM developed in this chapter as follows:

**1.** In FM, sideband power is obtained from the carrier component. In AM, sideband power is obtained from the modulating source.

**2.** In FM, the *total* radiated power is independent of the degree of modulation. In AM, the *total* radiated power varies with the degree of modulation.

**3.** In FM, the degree of modulation determines the bandwidth. In AM, the audio frequency determines the bandwidth.

**4.** In FM, an infinite number of sidebands is generated per modulating frequency. In AM, two sidebands are generated per modulating frequency.

**5.** In FM, the sidebands produce both quadrature and collinear vector components. In AM, only collinear vector components are produced.

**6.** FM and phase modulation are always coexistent.

**7.** FM can satisfactorily operate with lower signal-to-noise ratios than AM but at the expense of bandwidth occupancy in the transmission medium.

**8.** FM can be generated by means of simple capacitor microphones, which directly vary the frequency of an oscillator, by reactance modulators, which also vary the oscillator frequency, or by phase modulation of the output of a crystal oscillator.

**9.** FM receivers are generally more complicated than their AM counterparts.

**10.** Commercial FM broadcasting uses carrier frequencies from 88 to 108 MHz. These are in the very-high-frequency band (VHF).

**11.** The maximum allowable carrier-frequency deviation for broadcast FM is $\pm 75$ kHz. For the television sound carrier it is $\pm 25$ kHz.

**12.** The FM broadcast-channel bandwidth is 200 kHz (including a 50-kHz guard-band).

**13.** Modulating frequencies for the FM carrier are from 50 Hz to 15,000 Hz.†

**14.** FM sidebands *can* extend beyond the maximum deviation limits. This does not violate the transmission standards.

**15.** A narrow-band FM signal ($M = 0.2$, for example) is similar to an AM signal except that the FM sidebands are in quadrature with the carrier, while in AM they are collinear with the carrier.

**16.** FM transmitters are more efficient than AM transmitters for the same input power.

**17.** More FM stations can be accommodated in a given service area than AM.

**18.** FM receivers differ from AM receivers in terms of IF bandwidth (200 kHz versus 5 to 10 kHz), limiters, the methods of detection, tuning techniques, and so on.

†This does not take account of stereophonic multiplex, which is covered in Chapter 14.

**19.** Double-sideband suppressed-carrier techniques are employed in stereophonic multiplex.

**20.** FM is proportional to the time rate of change of phase angle. Therefore, if an integrated modulating signal is used to phase-modulate a carrier, the result is FM.

## REFERENCES


1. Tibbs, E., and Johnstone, G. G.: *Frequency Modulation Engineering*. John Wiley & Sons, Inc., New York, 1956.
2. Arguimbua, L. B., and Stuart, R. D.: *Frequency Modulation*. John Wiley & Sons, Inc., New York, 1956.
3. Armstrong, E. H., "Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," *Proc. IRE*, **24** (May 1936).
4. Seeley, S. W., "Frequency Modulation," *RCA Review*, **5** (April 1941).
5. Black, H. S., *Modulation Theory*. D. Van Nostrand Co., Inc., Princeton, N.J., 1953.
6. Schwartz, M.: *Information Transmission, Modulation, and Noise*. McGraw-Hill Book Company, New York, 1959.
7. Goldman, S.: *Frequency Analysis, Modulation and Noise*. McGraw-Hill Book Company, New York, 1948.
8. Crosby, M. G., "Communication by Phase Modulation," *Proc. IRE*, February 1939.
9. Meewezen, W. D., "Interrelation and Combination of Various Types of Modulation," *Proc. IRE*, November 1960.
10. Jaffe, J. L., "Armstrong's Frequency Modulator," *Proc. IRE*, April 1938.
11. Carson, J. R., "Notes on the Theory of Modulation," *Proc. IRE*, **10,** 57 (1922), **17,** 187 (1929).
12. Pol, B. van der, "Frequency Modulation," *Proc. IRE*, **18,** 1194 (1930).
13. Sturley, K. R.: *Radio Receiver Design*, Part I: *Radio Frequency Amplification and Detection*, 2nd ed. John Wiley & Sons, Inc., New York, 1953.


## APPENDIX 1-1

This appendix discusses the phasor construction for an FM carrier. For clarity, only the first- and second-ordered sideband pairs are included. See Fig. 1-23 and Eq. (1–14). Assuming a unit carrier ($E_c = 1$), then

$$e = J_0(M) \sin \omega_c t + J_1(M)[\sin (\omega_c + \omega_a)t - \sin (\omega_c - \omega_a)t] + J_2(M)[\sin (\omega_c + 2\omega_a)t + \sin (\omega_c - 2\omega_a)t].$$

The significance of the minus sign preceding the first-ordered lower-sideband term is that its associated phasor $J_1(M)_L$ must be reversed with respect to the "normal"AM case [see Eq. (1–8) and Fig. 1-5]. Look at Fig. 1.

The in-phase components of the odd-ordered pairs cancel. The quadrature components add to provide a resultant $e_{RO}$ at right angles to the carrier component $e_c$. Each sideband provides a quadrature component of $J_1(M) \sin \omega_a t$. Since there are two such components per pair, $e_{RO} = 2J_1(M) \sin \omega_a t$.

Since all even-ordered pairs are preceded by a plus sign, their respective phasors must be shown as in the "normal" AM case. Look at Fig. 2.

The quadrature components of the even-ordered pairs cancel. The in-phase components add to provide a resultant $e_{RE}$ that is collinear with the carrier component $e_c$. Each sideband provides an in-phase component of $J_2(M) \cos 2\omega_a t$. Since there are two such components per pair, $e_{RE} = 2J_2(M) \cos 2\omega_a t$.

Combining phasors (1) and (2), we obtain Fig. 3.

The total radiated carrier is

$$e_T = \sqrt{[2J_1(M) \sin \omega_a t]^2 + [J_0(M) + 2J_2(M) \cos 2\omega_a t]^2}.$$

The phase angle of the resultant $e_T$ is

$$\theta = \tan^{-1} \frac{2J_1(M) \sin \omega_a t}{J_0(M) + 2J_2(M) \cos 2\omega_a t} = \frac{e_{RO}}{e_c + e_{RE}},$$

and variations of this angle produce frequency modulation. If *all* sidebands are included, then

$$\underset{\text{unit carrier}}{1} = \sqrt{e_{RO}^2 + (e_c + e_{RE})^2} \qquad \text{or} \qquad 1 = [J_0(M)]^2 + 2 \sum_{n=1}^{\infty} [J_n(M)]^2$$

and the resultant triangle as shown in Fig. 4.

Odd-ordered pairs produce FM (and some AM).

Even-ordered pairs produce AM.

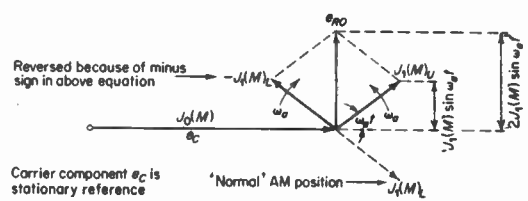



Fig. 1.

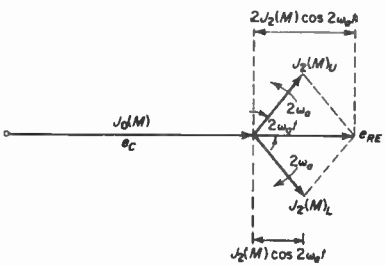


Fig. 2.

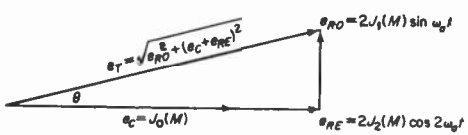


Fig. 3

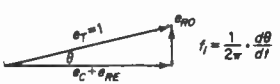


Fig. 4

# 2

# INTERFERENCE

## 2-1. Types of noise

The most important advantage that frequency modulation has over amplitude modulation is its ability to reduce the effects of all forms of interference and noise.† The improved noise performance stems from the use of changes in the instantaneous frequency of a carrier that are large compared to those occurring outside the transmitter. Noise and interference can originate from three sources:

1. *Impulse noise.* Static-type disturbances (lightning and other atmospheric conditions), automobile ignition systems, and switching transients generated by electrical machinery.

2. *Carrier interference.* Interfering radio-frequency carriers, which can be either unmodulated (CW) or modulated.

3. *Random noise.* Thermal agitation in conductors (wires, resistors, and so on), "shot" noise in vacuum tubes, and random fluctuations of minority-carrier current in transistors.

## 2-2. Impulse noise

The first source listed is impulse noise, with the impulses having very high peak voltages. The impulses are generally of very short duration compared to the time between pulses. These may be random or periodic, depend-

†Noise is defined as any undesirable disturbance that tends to obscure the transmitted signal.

ing upon their origin. If we perform a mathematical analysis of these pulses, we find that they comprise an infinite series of sinusoidal components. Those noise-frequency components (each is a harmonic of the fundamental pulse frequency) that appear at the receiver antenna terminals and fall within the RF acceptance band are amplified along with the desired carrier. The amplified impulses then shock-excite the tuned circuits in the IF section and produce a series of damped oscillations. The oscillations occur at the natural resonant frequency of the tuned circuits. Thus, the original noise pulse is modified by the selectivity (and sensitivity) characteristic of the IF amplifiers. If the IF amplifier is wide-band (as in broadcast FM), the peak amplitude of the noise envelope is increased while its time duration is reduced in proportion to the increase in bandwidth. If the IF amplifier is narrow-band (as in broadcast AM), the peak amplitude of the noise envelope is reduced while its time duration is accordingly increased in proportion to 1/bandwidth. In both instances the oscillation frequency is that of the center frequency of the IF tuned circuits, while the time duration of the main lobe of the wave train (see Fig. 2-1) is approximately equal to the reciprocal of the IF *half*-bandwidth. Thus, for broadcast FM using a 10.7-MHz IF and a 200-kHz bandwidth, the main lobe time duration is then approximately 1/100,000 or 10 microseconds.

Figure 2-1 shows the IF output waveforms due to an impulse for the narrow and wide-band cases. The increased *amplitude* in the wide-band case is relatively unimportant, since these amplitude fluctuations are removed by the limiters (see Fig. 1-1). However, we shall see shortly that this impulse produces *phase* modulation (and hence frequency modulation) of the resultant of the noise and signal voltages and that this is more difficult to contend with.

This form of noise produces an audible output not unlike that of a series of clicks *or* pops of a fairly high-pitched nature. Whether the noise output is a click or a pop depends upon such factors as the initial phase of the impulse (that is, when it occurs with respect to the signal), the relative amplitudes of the impulse and signal, the relative frequencies of the impulse and signal (if the receiver is slightly mistuned, their frequencies will not be the same), and other factors. We shall return to these forms of disturbances in a later paragraph.

Since the peak amplitude of a short-duration impulse is proportional to the IF bandwidth, the bandwidth should be no wider than necessary. An IF bandwidth wider than necessary serves only to provide the impulse with a better chance of exceeding the signal amplitude and suppressing the signal. The noise is thus said to "capture" the receiver. Since we are in effect limiting the bandwidth for more efficient impulse noise suppression, we must also limit the maximum signal deviation. This increases the effect of random-type noise (the third type listed above) and therefore makes the choice of a $\pm$75-kHz deviation a compromise.

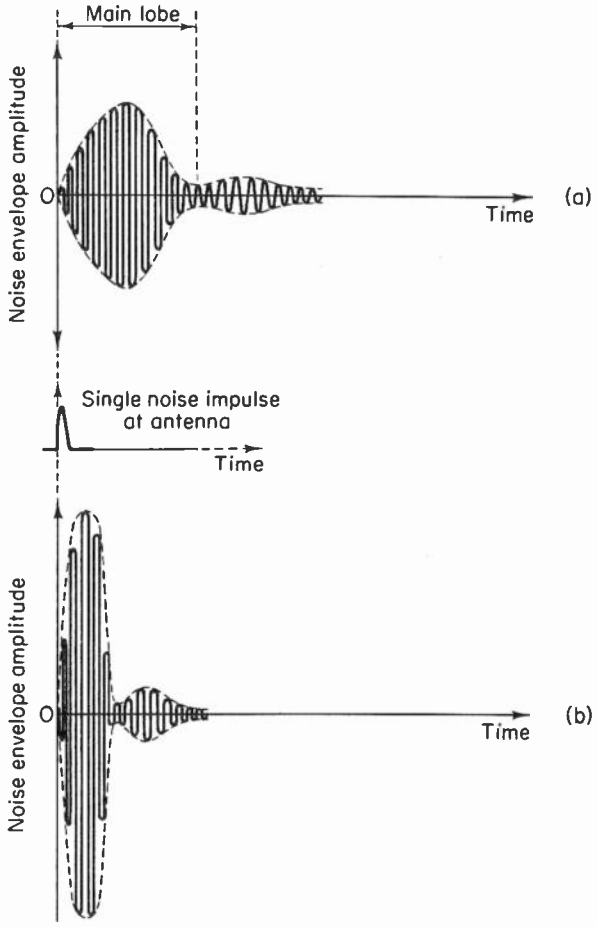
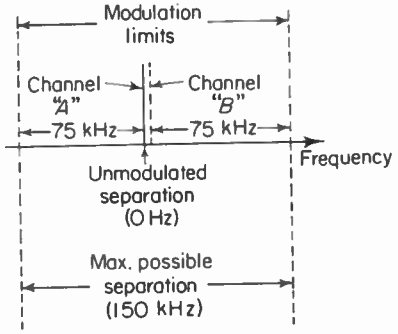


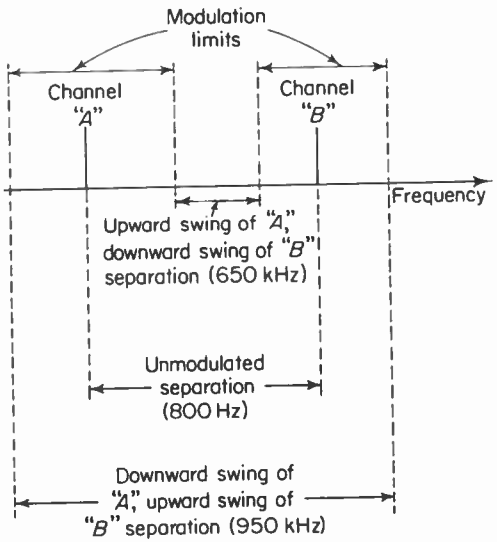

**Fig. 2-1.** IF output due to a single noise impulse. (a) narrow band IF. (b) wideband, flat top IF. The time duration of the main lobe is less for the wideband IF circuit because of larger losses (lower $Q$) in the tuned circuits.

## 2-3. Carrier interference

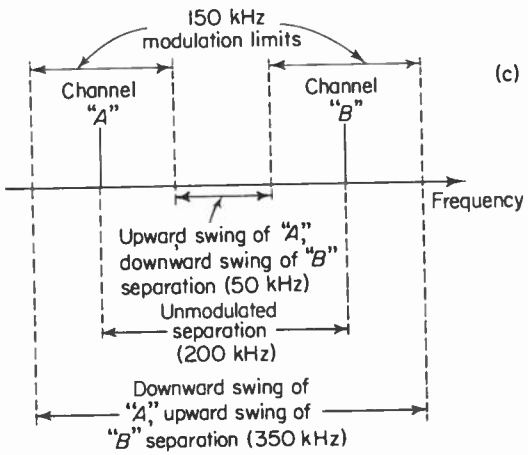
The second type of disturbance occurs when an interfering carrier is combined with the desired carrier at the antenna of the receiver. The undesired carrier can be either modulated or not. It can be co-channel or adjacent channel. If it is co-channel, it is of the same (or very nearly the same) frequency as the desired carrier. If it is an adjacent channel in the *same* service area, it can differ from the desired carrier by 800 kHz when neither carrier is modulated, by as little as 650 kHz when both are maximum modulated toward one another, or as much as 950 kHz when both are maximum modulated away from one another. If the undesired adjacent carrier is in an adjoining




(a) Co-channel, adjacent service areas. The degree of freedom from such interference is mainly a function of the relative carrier amplitudes and the "capture" abilities of the receiver.




(b) Adjacent channels, same service area. The degree of freedom from such interference is mainly a function of the IF and RF selectivity.




(c) Adjacent channel, adjacent service areas. Freedom from interference same as (b) above.

**Fig. 2-2.** Instantaneous channel frequency separations for three different conditions.

service area, it can differ from the desired carrier by 200 kHz when neither is modulated, by as little as 50 kHz when both are maximum modulated toward one another, or as much as 350 kHz when both are maximum modulated away from one another. These conditions are shown in Fig. 2-2. It should be noted at this point that co-channel interference is the most serious type, since it is entirely possible to produce audible beats as a result of the two carriers combining in the receiver. This is true even if neither carrier is modulated, if we keep in mind that only a small frequency difference exists between the carriers. We shall explore this last statement in greater detail shortly.

## 2-4. Random noise

The third form of noise listed is *random* or *fluctuation* noise. Its energy distribution is quite different from that of impulse, its frequency components cover a very wide range, and it produces a hissing sound in the output. Since random noise originates from many sources, its detailed analysis is reserved for a subsequent chapter.

## 2-5. Basic considerations

In order to see how frequency modulation provides superior noise (and interference) performance compared to amplitude modulation, we must see the effect that noise has on a carrier. We may simplify the discussion by taking a *single* noise-frequency component (voltage in this case), combining it with a desired carrier, and then observing the effect on the amplitude, phase, and frequency of the resultant. This single noise voltage may be a single impulse or a sinusoidal carrier. Since random noise consists of an infinite number of frequencies of almost equal amplitudes, we can even select one of these for our single noise-frequency component. We shall initially assume that the *peak* amplitude of the noise voltage ($E_N$) is smaller than the *peak* amplitude of the desired signal voltage ($E_S$). Thus the ratio of $E_N$ to $E_S$ is less than one and

$$\frac{E_N}{E_S} = a < 1. \tag{2–1}$$

The noise component can thus be treated as a sideband with respect to the signal carrier, and we shall assume that the noise frequency is greater than the signal frequency. If we treat both voltages as phasors, then we may let the noise phasor rotate counterclockwise at an angular frequency equal to the difference frequency between both phasors. We then maintain the signal phasor stationary. If the noise angular velocity is $\omega_N$ and the signal angular velocity is $\omega_S$, then the *relative* angular velocity is the *beat* frequency or the

difference between the two. Thus,

$$\omega_B = \omega_N - \omega_S, \tag{2-2}$$

where

$\omega_B = 2\pi f_B, f_B =$ beat frequency,

$\omega_N = 2\pi f_N, f_N =$ noise frequency,

$\omega_S = 2\pi f_S, f_S =$ signal frequency.

If the signal phasor is assumed to be one unit long, then by Eq. (2–1) the noise phasor is $a$ units long. This is indicated in Fig. 2-3 with the noise phasor shown initially at a zero relative phase angle. At this instant the amplitude of the resultant phasor $R$ is $(1 + a)$ and the phase angle $\phi$ of the resultant is zero degrees. As the noise phasor rotates about the tip of the unit signal, it traces out a circle of radius $a$. The tip of the resultant always touches this circle. Therefore, after 180° of rotation of the noise phasor (equivalent to the time of half a beat cycle), the resultant amplitude is $(1 - a)$ and the phase angle of the resultant is again zero. Thus, as the noise rotates about the signal phasor, the resultant of the two is both *amplitude* and *phase* modulated to the extent of $a$ for amplitude and $\phi$ for phase.† This is shown in the figure for $a = 0.5$.

The amplitude modulation of the resultant is removed by a limiter, so we need not concern ourselves with it for now. However, the FM detector responds to the frequency deviation of the resultant, since the time rate of change of $\phi$ represents changes in the carrier frequency about its average value. But what is actually taking place in Fig. 2-3 is not *frequency* modulation. This is because the *amount* of phase deviation (angle $\phi$) is independent

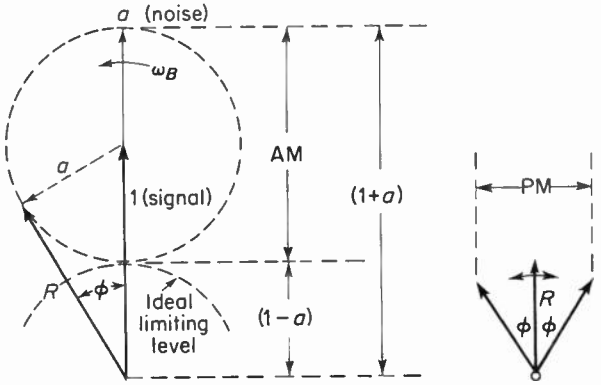



**Fig. 2-3.** The basic noise phasor. The phase modulation of the resultant is $\pm\phi$ and is proportional to $a$.

†By *extent* we mean an equivalent modulation factor for AM of $a$ and modulation index for PM of $\phi$.

of the beat frequency $f_B$ and depends only upon the amplitude of the noise phasor. Thus, the resultant is phase modulated by the noise. The equivalent FM produced as a result of this phase modulation will be explored after the amplitude and phase disturbances are evaluated.

## 2-6. Amplitude disturbance

Let us first obtain the expression for the amplitude disturbance of the resultant signal as a result of the superimposed noise. Again, the desired carrier is a unit phasor and the noise has an amplitude equal to $a$. Both voltages are assumed to be unmodulated, and the beat frequency $f_B$ is much smaller than the desired carrier frequency $f_s$. The noise phasor again rotates counterclockwise at $\omega_B = \omega_N - \omega_S$ and sweeps out an angle $\omega_B t$. The resultant $R$ sweeps out an angle $\pm\phi$. By the use of trigonometry (see Fig. 2-4)

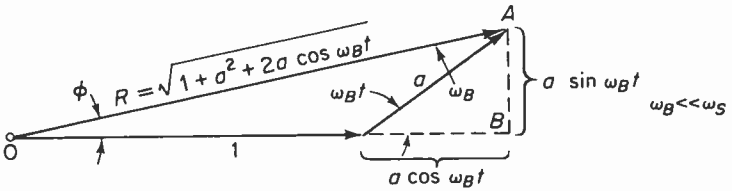



**Fig. 2-4.** We obtain the amplitude disturbance (the variation of $R$ over one beat cycle) by trigonometry.

the amplitude fluctuation (AM) of $R$ as a function of time is the square root of the sum of the squares of triangle $OAB$ or

$$A(t) = \sqrt{(1 + a \cos \omega_B t)^2 + (a \sin \omega_B t)^2}. \tag{2-3}$$

Expanding under the radical and grouping terms gives

$$A(t) = \sqrt{1 + 2a \cos \omega_B t + a^2(\sin^2 \omega_B t + \cos^2 \omega_B t)}. \tag{2-4}$$

Using the identity: $\sin^2 x + \cos^2 x = 1$, we have

$$A(t) = \sqrt{1 + 2a \cos \omega_B t + a^2}. \tag{2-5}$$

Now, if both phasors are in phase, $\omega_B t = 0$ and $\cos \omega_B t = +1$. Thus,

$$A(T)_{\max} = \sqrt{1 + 2a + a^2} = 1 + a.$$

If the phasors are in phase opposition, $\omega_B t = 180°$ and $\cos \omega_B t = -1$. Thus,

$$A(T)_{\min} = \sqrt{1 - 2a + a^2} = 1 - a.$$

The envelope of the resultant then varies between the limits of $1 + a$ and $1 - a$. A plot of Eq. (2–5) is shown in Fig. 2-5 for values of $a = 0.2$ and 0.9. Notice that when the noise amplitude is 0.9 of the signal voltage, the rectified envelope takes on a peaky appearance resulting in a large harmonic

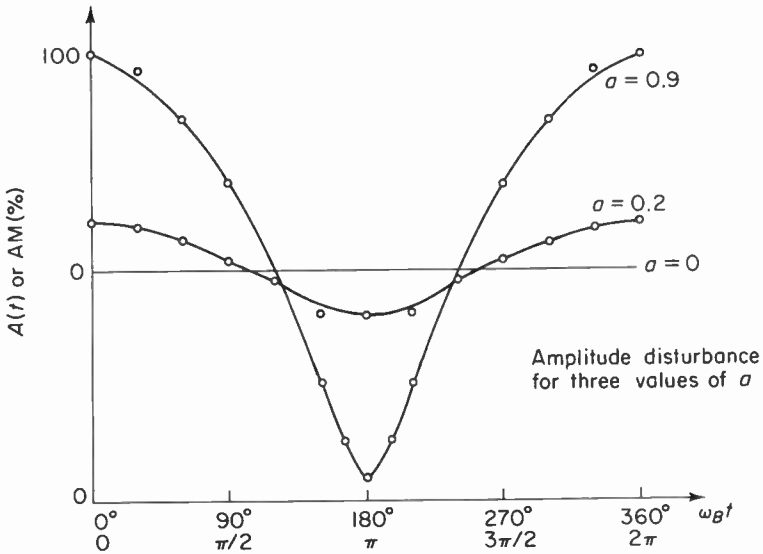



**Fig. 2-5.** The *rectified* heterodyne envelope as a result of a noise voltage of $a$ amplitude superimposed on a signal voltage of unit amplitude.

content. However, we are assuming that these amplitude disturbances are completely suppressed by efficient limiters.

## 2-7. Phase disturbance

Consider now the effect that the noise voltage has on the *phase* of the resultant. We must remember that any phase fluctuation of the resultant will produce an equivalent frequency fluctuation, since frequency is defined as the time derivative of phase [in accordance with Eq. (1–10)]. Now, it is precisely the *variation* in the angle $\phi$ of Fig. 2-4 that represents the phase modulation of $R$. First, let us obtain an expression for the phase angle as a function of time or $\phi(t)$. Once we have this, we may see what effect that noise has on the frequency of the resultant. We will then see how to contend with this frequency disturbance (if any) in order to prevent distortion in the final audio output.

With reference to Fig. 2-4,

$$\tan \phi = \frac{a \sin \omega_B t}{1 + a \cos \omega_B t} \tag{2–6}$$

or

$$\phi = \tan^{-1} \frac{a \sin \omega_B t}{1 + a \cos \omega_B t}. \tag{2–7}$$

Now, if the noise and signal phasors are in phase, $\omega_B t = 0$, $\sin \omega_B t = 0$

and $\cos \omega_B t = +1$. Thus $\phi(t)_{\min} = 0°$ when $\omega_B t = 0°$, 180°, 360°, and so on. The maximum value of phase shift depends upon the ratio of noise to signal but cannot exceed 90°. This maximum angular shift occurs when the noise phasor is at right angles to the resultant or

$$\phi_{\max} = \sin^{-1} \frac{a}{1} = \sin^{-1} a. \tag{2–8}$$

Thus, the maximum possible phase shift of the resultant (from its quiescent position, $\omega_B t = 0$) due to the noise pulse is, for a noise-to-signal ratio of unity, $\phi = \sin^{-1} 1 = 90°$ or $\pi$ radians. Thus, one way to minimize the effects of phase changes due to *noise* is to produce much larger phase changes due to *signal*. In other words, if we frequency-deviate a carrier 70 kHz by a 70-Hz audio signal, then the modulation index is 70,000/70 = 1000 radians. The peak phase deviation is (57.3)(1000) = 57,000 degrees. Certainly, a 90° phase shift due to noise will be swamped by a 57,000-degree phase shift due to signal. Thus, the larger the deviation ratio we choose, the more effective will be the noise suppression. It can be shown† that the signal-to-noise (voltage) ratio improvement of FM over AM is

$$\frac{\left(\frac{S}{N}\right)_{FM}}{\left(\frac{S}{N}\right)_{AM}} = \sqrt{3}\, D \tag{2–9}$$

where $D$ = deviation ratio.

Equation (2–9) is valid only when the peak signal exceeds the peak noise. If the peak noise should exceed the peak signal, the noise improvement is lost.

Since power is proportional to the square of voltage, we may state the improvement in signal-to-noise *power* as

$$\frac{\left(\frac{S}{N}\right)_{FM}}{\left(\frac{S}{N}\right)_{AM}} = 3D^2. \tag{2–10}$$

Equations (2–9) and (2–10) are for random-type noise. The improvement for impulse noise is of the same order of magnitude and slightly greater. For a deviation ratio of 5 (wide-band FM), the signal-to-noise power improvement is then $(3)(5^2) = 75$ or $10 \log 75 = 18.75$ db better than AM. For a deviation ratio of $D = 0.6$ (narrow-band FM), we have

$$\begin{aligned} \text{db improvement} &= 10 \log (3)(0.6)^2 \\ &\doteq 10 \log 1, \\ \text{db} &\doteq 0. \end{aligned}$$

†See M. G. Crosby, "Frequency Modulation Noise Characteristics," *Proc. IRE*, **25**, 472–514 (April 1937).

Thus, narrow-band FM provides no noise improvement over AM, since we are restricting the *signal* phase deviations with respect to the *noise* phase deviations.

Returning to Eq. (2–7), a plot of this function is shown in Fig. 2-6. Notice that for small noise-to-signal ratios ($a = 0.2$) the phase disturbance of $R$ is approximately sinusoidal and the *average* phase shift over one complete revolution of $a$ about the unit carrier is *zero*. For larger values of $a$ ($a = 0.9$) the phase variation is no longer sinusoidal but the *average* shift is still *zero*. The reason for the "peculiar" shape of $\phi(t)$ for $a = 0.9$ is as follows. Between 0° and 150° of rotation of the noise phasor about the signal phasor, the phase change is occurring at a fairly constant rate. Between 154° and 206° of rotation† the amplitude of the resultant is changing very rapidly. This is necessary since the noise phasor is rotating at a constant angular speed. In other words, $R$ sweeps out a maximum angle of $\phi_{max} = \sin^{-1} a = \sin^{-1} 0.9 = 64°$. This is equal to a maximum of $64°/57.3 \doteq 1.12$ radians swept out in the time of one beat cycle.

Now if the beat frequency is increased, the resultant must sweep out the same maximum angle (assuming $a = 0.9$) in less time. Thus the rate of change of phase is greater but the maximum angle swept out remains the same, since $\phi_{max}$ is a function of $a$ only.

## 2-8. Frequency disturbance

Finally, let us see the effect that phase modulation of the resultant (Fig. 2-6) has on the frequency of the resultant. We recall from Eq. (1–10) that frequency modulation is proportional to the rate of change of phase. In other words, the instantaneous frequency ($f_i$) of the resultant phasor is proportional to the *slope* of the curves of Fig. 2-6. To obtain an expression for the slope of Fig. 2-6 we must differentiate Eq. (2–7) by means of the calculus. When this is done,‡ the expression for the frequency *shift* due to the presence of the noise voltage is

$$f_N = \frac{1}{2\pi} \cdot \frac{d\phi}{dt} = \frac{a f_B (\cos \omega_B t + a)}{1 + a^2 + 2a \cos \omega_B t}. \tag{2–11}$$

This expression is exact and is valid for all values of $a$. A plot of Eq. (2–11) is shown in Fig. 2-7 for two values of $a$ (0.2 and 0.9).

The *total* instantaneous frequency of the resultant phasor of Fig. 2-4 is then equal to the frequency of the stronger signal (the unit phasor) plus a *shift* due to the disturbance. Thus

†If Eq. (2–7) is differentiated and set equal to zero, the instantaneous angles ($\omega_B t$) of the noise vector may be computed. These are the angles where the phase deviation of the resultant is a maximum.

‡See Appendix 2-1.

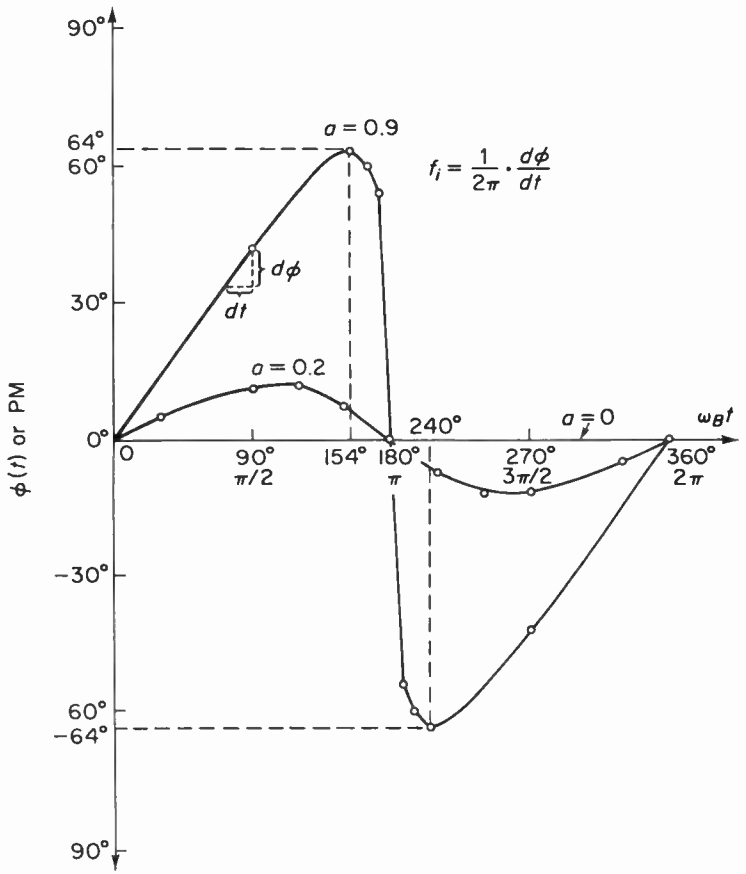



**Fig. 2-6.** The phase disturbance of the resultant of noise $a$ and a unit signal. The *average* value of phase shift over one beat cycle = 0.

$$f_i = f_c + \frac{1}{2\pi} \cdot \frac{d\phi}{dt} \tag{2-12}$$

$$= f_c + f_N \tag{2-13}$$

$$= f_c + \frac{a f_B(\cos \omega_B t + a)}{1 + a^2 + 2a \cos \omega_B t}. \tag{2-14}$$

Note that the shift ($f_N$) due to noise is directly proportional to the beat frequency $f_B$. For small values of $a$, $f_N$ is approximately cosinusoidal, producing a single beat-note in the detector output. If the beat frequency is 15 kHz or higher, it will be inaudible and we may ignore it. If the beat frequency is less than 15 kHz, then it will be audible. If the frequency of the desired carrier is the same as that of the noise, the beat frequency is zero. This produces no FM disturbance in the detector output, and only the desired carrier produces an output (provided it is being modulated).

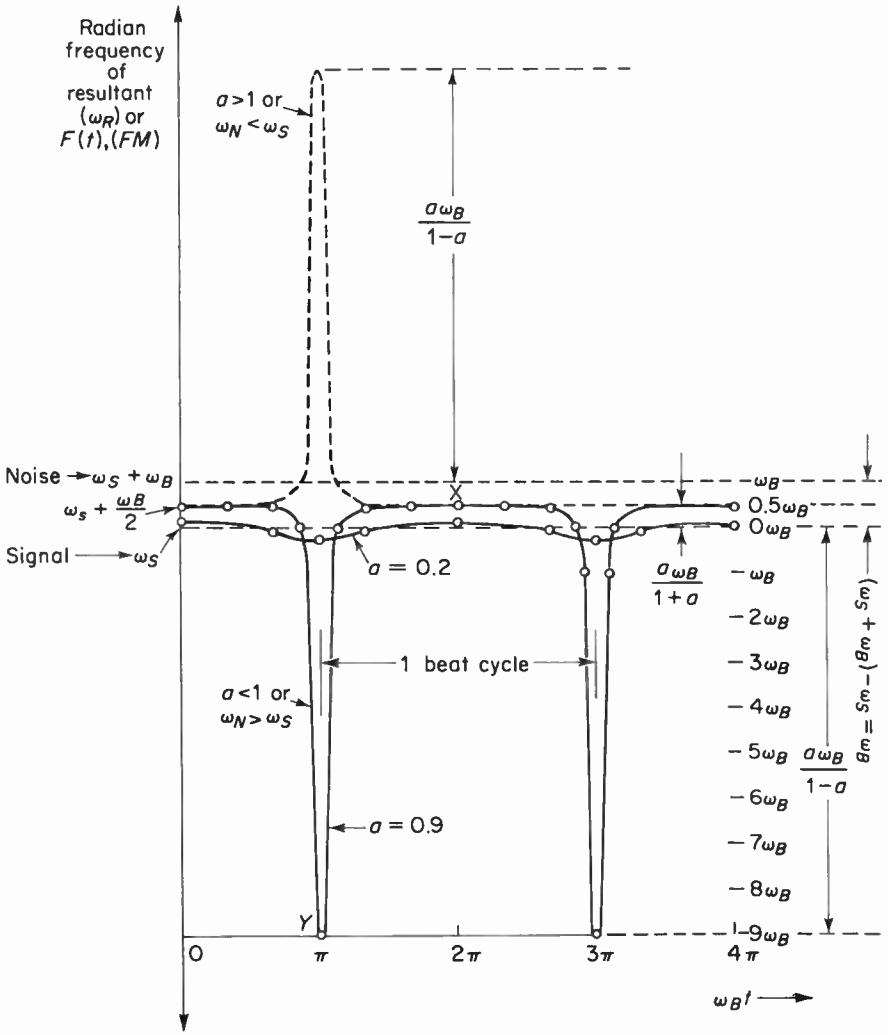



**Fig. 2-7.** The frequency disturbance due to a noise voltage of magnitude $a$ superimposed on a unit signal. The solid-line frequency spikes are shown for two complete rotations of the noise phasor about the signal phasor.

## 2-9. Capture effect

We may draw several conclusions from Fig. 2-7. As long as $a$ is less than unity, the average *phase* change of the resultant over one beat cycle is zero. Thus, the average frequency of the resultant is that of the stronger signal $\omega_S$. If one were to integrate Eq. (2–14) from zero to infinity, the result would be that of $f_c$, the frequency of the stronger carrier. The magnitude of frequency excursions of the resultant above and below $\omega_S$ increases with the

frequency difference ($\omega_B$) between the stronger and weaker carriers and with the magnitude of $a$. Should $a$ become larger than unity, the noise ($\omega_S + \omega_B$) "takes over" or captures the detector and suppresses the desired carrier. In other words, the average frequency of the resultant suddenly becomes that of the noise.

Then, if two co-channel IF carriers appear at the detector input, the average frequency detected is that of the stronger of the two signals. The

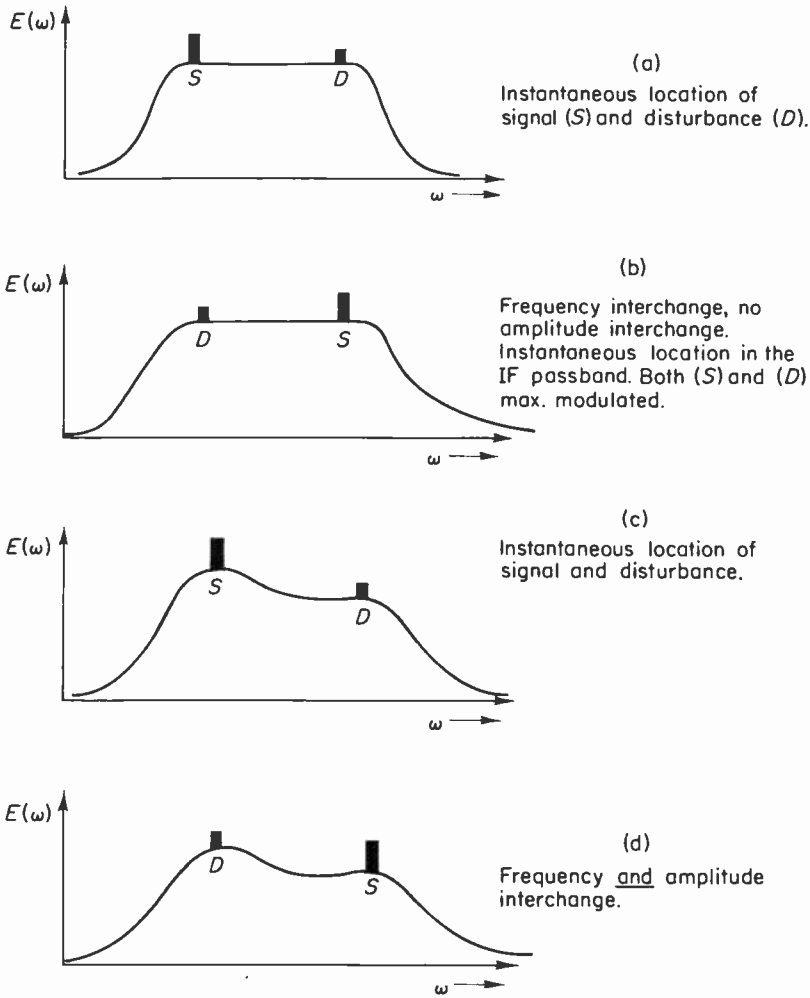



**Fig. 2-8.** The importance of a flat IF passband (careful design *and* alignment). Parts (a) and (b) illustrate the wanted signal ($S$) at the detector input greater than the disturbance ($D$) or $a$ 1. The disturbance is suppressed. Parts (c) and (d) illustrate how the wanted signal at the detector input is not maintained greater than that of the disturbance. The disturbance pattern of fig. 2-7 reverses and the disturbance captures the receiver.

disturbance pattern produced should not be "distorted" in any way by the detector. This is required so that the average frequency of the pattern does not change over one beat cycle. If the limiter-detector bandwidth is not able to accommodate the maximum excursion for a given $a$ and a given $\omega_B$, the pattern *is* distorted. The average frequency is shifted away from that of the stronger signal toward that of the undesired signal. This "frequency clipping" can be prevented by using a wide-band limiter-detector.†

It is also necessary that $a$ remain less than unity for low capture ratios.‡ Flat-topped and well-aligned IF stages will insure that during the frequency interchange there will be no amplitude interchange between the desired signal and the disturbance. In other words, so long as the disturbance amplitude $a$ is not allowed to exceed that of the wanted signal during the time of one beat cycle, the spikes will not reverse direction and the average frequency of the resultant remains that of the wanted signal. See Fig. 2-8.

We also see by Eq. (2–14) that the disturbance magnitude is proportional to the frequency difference between the two carriers. As the beat frequency increases, the disturbance will tend to get worse except that the deemphasis network, being a low-pass filter, will attenuate the higher-ordered beat notes.

The capture effect (ability of a wide-band FM system to suppress the weaker of two signals) provides FM systems with a distinct advantage over a "comparable" AM system. So long as adjacent service areas are kept reasonably apart, carriers in the respective areas may be co-channel, since the capture effect will prevail and a given receiver output will contain the intelligence of the stronger of the two carriers. Amplitude-modulation interference is directly proportional to relative signal strength, thus requiring much greater service-area separation for co-channel carriers. Frequency modulation, we conclude, is more efficient than amplitude modulation in terms of utilizing available allocations for large areas.

## 2-10. Pops and clicks as they relate to impulses

Let us now observe, in a graphical sense, the action of an impulse on the detected output of an FM receiver. As stated in Sec. 2-2, the noise pulse is modified by the IF selectivity and sensitivity characteristic of the IF amplifiers. The reason is that the impulse is short compared to the impulse period. It shock-excites the tuned circuits, and the tuned-circuit output is independent of the nature of the input pulse.

We may regard the impulse [see Fig. 2-9(a)] as a noise phasor $a$ of fre-

†See L. B. Arguimbau, *Vacuum Tube Circuits and Transistors* (New York: John Wiley & Sons, Inc., 1956), p. 531.

‡Capture ratio in db $= 20 \log_{10} 1/a$.

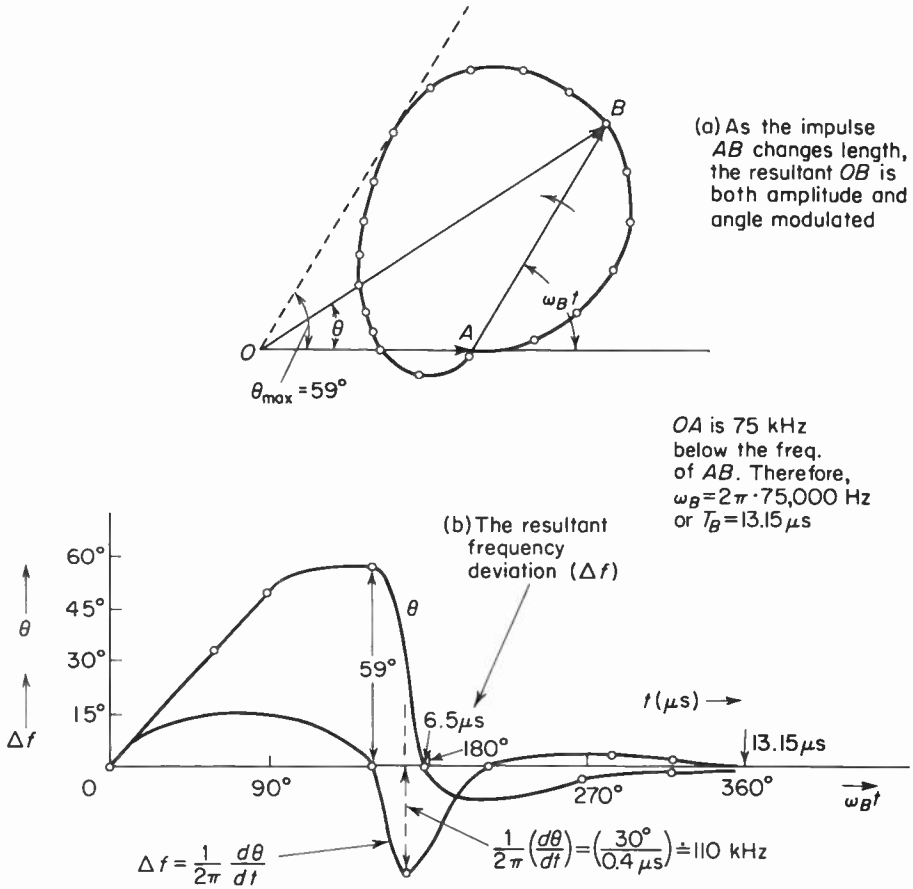



**Fig. 2-9.** The result of the impulse $AB$ is to produce a "click" in the audio output.

quency $\omega_B$ which grows to a maximum and then decreases to zero while rotating about the tip of $OA$. The initial relative phase angle between the signal and the noise is zero. As the noise phasor goes through successive changes in growth, the resultant ($OB$) is both amplitude and angle-modulated. As shown in Fig. 2-9(b), the net change in angle is zero; therefore the *net* area of the $\Delta f$ curve (obtained by $d\theta/dt$) is zero. The physical significance of the $\Delta f$ curve is that it lacks low-frequency content and produces a "click" (a relatively negligible disturbance).

On a purely random basis, the situation could be as depicted in Fig. 2-10(a). Here, the impulse starts "growing" at a later time compared to the start in Fig. 2-9(a). The result is that the origin is encircled. Whenever the origin is enclosed by the locus of the noise phasor, the resultant frequency deviation [as shown in Fig. 2-10(b)] has a nonzero net area. This produces

an audio output with appreciable low-frequency content or a "pop." We assume again that the noise is momentarily higher in frequency than the signal by 75,000 Hz. Thus, the beat period is 1/75,000 = 13.15 $\mu$sec. The maximum gradient of the phase plot is $(10.9 \times 10^6)$ degrees/sec, so that

$$\Delta f = \frac{1}{2\pi} \cdot (10.9)(10^6) = 1.74 \text{ MHz (max)}.$$

In the case of clicks, we see that the phase change is essentially momentary and less than the deemphasis time constant of 75 $\mu$sec. Experimentally, we determine that the audio output of the click relative to 100 per cent modulation (75 kHz) is approximately 26 db below the signal amplitude.

With regard to pops, we observe a sudden phase change, but thereafter, a relatively constant phase angle with respect to $\omega_B t$. This produces a sudden change in detector output, but thereafter the output remains constant.

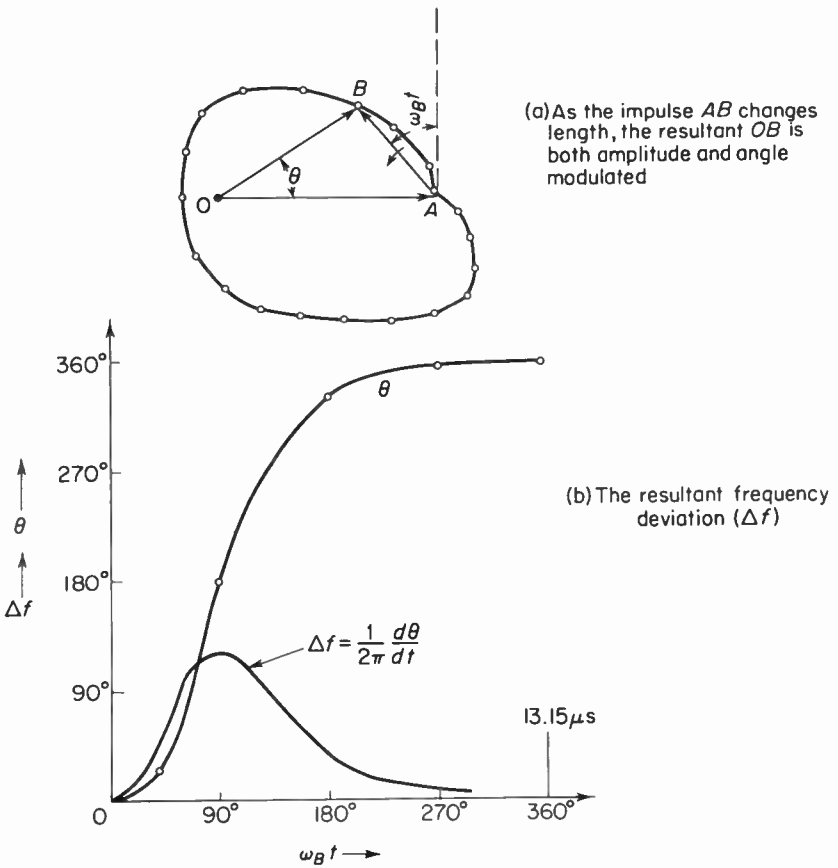



**Fig. 2-10.** The result of the impulse $AB$ is to produce a "pop" in the audio output.

Experimentally we find that the disturbance is only approximately 15.4 db below signal output relative to 100 per cent modulation (75 kHz).

In general, then, we see that pops produce a far more severe disturbance than clicks. The occurrence of either is, of course, purely random.

When an FM receiver is mistuned, the carrier frequency is inherently different from that of the impulse, and the pop intensity increases. Thus, the probability of occurrence of a pop increases with beat frequency. A pop, therefore, cannot occur if the receiver is precisely center-tuned and the signal is at center frequency. This points up the importance of properly tuning an FM receiver.

## 2-11. Interference suppression

It was stated in Sec. 2-1 that the improved noise performance of FM over AM is due mainly to the use of large changes in the instantaneous frequency of a carrier. This is borne out by an examination of Eq. (2–10). The improvement in $S/N$ power is proportional to the square of carrier deviation. Thus, the fundamental approach to interference suppression is to use *wide-band* FM—that is, use a large-index system. There is a limit to how large an index we can use to gain better suppression. Figure 1-21 tells us that bandwidth occupancy increases with increasing index and therefore we accept a greater amount of noise. In general, indices of five or greater constitute a wide-band system and give good interference suppression.

When the desired carrier amplitude is greater than that of the disturbance, the resulting effect for both AM and FM is to produce an output at the beat frequency ($\omega_B$). In AM the *intensity* of this output is independent of $\omega_B$, but in FM it is directly proportional to $\omega_B$. The interference effect in both systems is directly proportional to the amplitude of the disturbance. We may state this mathematically as follows:

$$E_N = \frac{N}{C} \qquad \text{for AM,} \tag{2–15}$$

$$E_N = \frac{\omega_B}{\Delta\omega_C} \cdot \frac{N}{C} \qquad \text{for FM,} \tag{2–16}$$

where

$E_N$ = amplitude of interference output referred to a percentage of maximum output,

$N$ = peak amplitude of disturbance,

$C$ = peak amplitude of desired signal,

$\omega_B$ = angular beat or difference frequency between the desired signal and the disturbance,

$\Delta\omega_C$ = angular frequency deviation of desired carrier $= 2\pi\Delta f_C$.

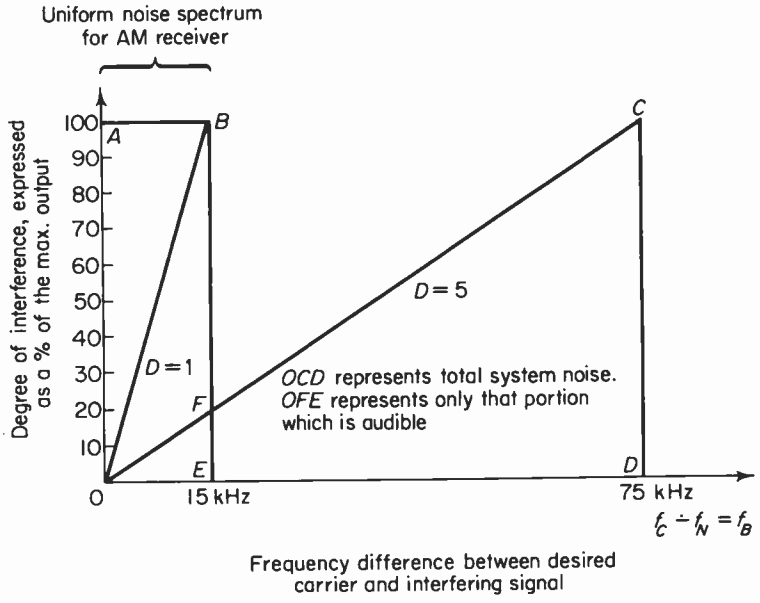



**Fig. 2-11.** The familiar noise spectrum. This spectrum is valid where the desired carrier amplitude is appreciably larger than the interfering signal, the desired carrier and interfering signal amplitudes remain constant and there is no preemphasis provided at the transmitter.

If we plot Eqs. (2–15) and (2–16), we obtain the familiar FM noise triangle (*OCD* for wide-band FM and *OBE* for narrow-band FM) and AM noise rectangle (*OABE*).

Figure 2-11 shows that the demodulated noise in AM follows a uniform spectrum for all beat frequencies up to 15 kHz (the normal limit of audibility). The noise spectrum for FM is entirely different. If an interfering signal is 15 kHz away from the desired signal, it will produce ten times the interference effect that would have occurred had the beat frequency been only 1500 Hz. Also, the area of the narrow-band FM system, being one-half that of the AM system, results in one-half the noise. If we use a wide-band system (*OCD*), then the $S/N$ ratio for FM is superior to a comparable AM system by a factor of $100/20 = 5$, or $20 \log_{10} 5 = 14$ db.

We note two important items at this point:

1. The basic improvement shown above is obtained when the undesired signal frequency is no more than 15 kHz away from the desired signal frequency or $\omega_B \leqq 15{,}000/2\pi$ Hz.

2. The triangular spectrum of an FM system (narrow or wide-band) results in a progressive increase in noise output with increasing frequency. Thus, the $S/N$ ratio at 15,000 Hz is lower than that at 1500 Hz. Also, the spectral density (watts/Hz) is less at 15,000 Hz, resulting in a smaller carrier deviation.

The obvious solution to this problem is to equalize the $S/N$ ratio across the entire band by *preemphasizing* the amplitudes of the higher-frequency tones at the transmitter. This will provide a measure of interference suppression if we furnish a complementary deemphasis network (low-pass filter) at the receiver. A typical configuration is shown in Fig. 2-12. This filter

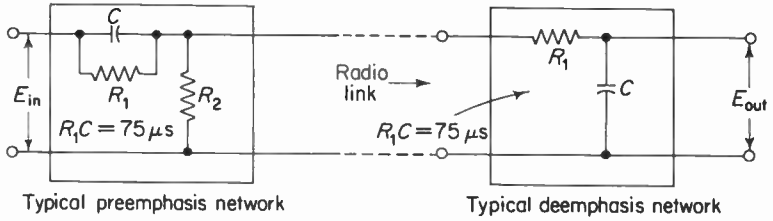



**Fig. 2-12.** Pre- and deemphasis networks. $E_{OUT}$ should be within a few db of $E_{IN}$, from 50 Hz to 15,000 Hz.

(nominal time constant of 75 $\mu$sec) attenuates the interference as well as the high-frequency audio notes, but since the high-frequency program content is restored to its original level, the interference is substantially reduced. The improvement in $S/N$ of FM over AM due to the use of preemphasis-deemphasis can be found by the use of the following equation:†

$$\text{improvement in db} = 10 \log_{10} \frac{1}{D} \tag{2-17}$$

where

$$D = 3\left(\frac{f_1}{f_m}\right)^3 \left(\frac{f_m}{f_1} - \tan^{-1} \frac{f_m}{f_1}\right), \tag{2-18}$$

$$f_1 = \frac{1}{2\pi R_1 C} \qquad \text{(see Fig. 2-12)},$$

$$f_m = 15{,}000 \text{ Hz}.$$

Substitution of appropriate numbers in Eq. (2–18) followed by the use of Eq. (2–17) will give a 20-fold or 13-db improvement (in voltage). This is based upon an FCC-specified preemphasis time constant ($R_1C$) of 75 $\mu$sec.

The effect of preemphasis at the transmitter is to "flatten" the triangular noise spectrum and make it essentially rectangular as shown in Fig. 2-13. Above about 2000 Hz the receiver audio passband drops linearly with frequency. This effect compensates for the tendency of the high-frequency noise components to deviate the transmitter excessively to produce a greater detector output. As a result, the noise output at the receiver detector remains essentially constant, and the improvement of an FM system over

†See M. Schwartz, *Information Transmission, Modulation and Noise* (New York: McGraw-Hill Book Company, 1959), p. 309.

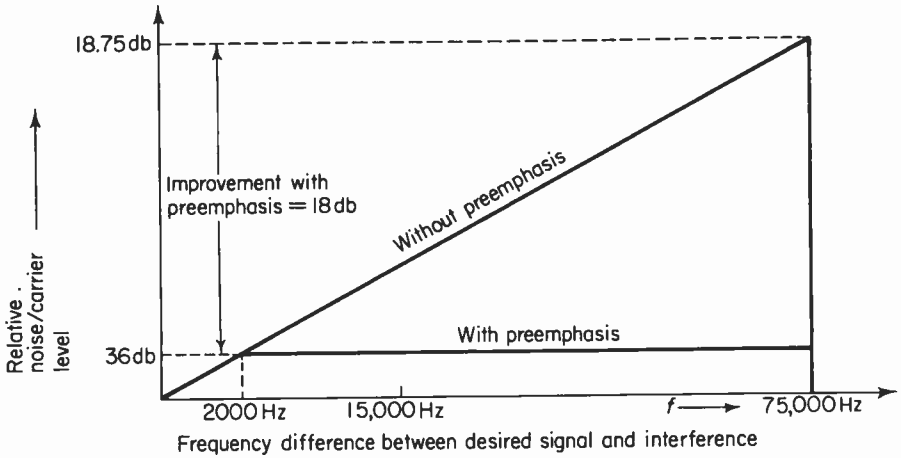



**Fig. 2-13.** Preemphasis effect is to "flatten" spectrum above about 2000 Hz. From the vantage point of the noise only, the system appears as narrow-band AM.

that of a comparable AM system is more like $10 \log_{10} 3(75{,}000/2000)^2 = 36$ db.

In summary then, the effective interference-suppression techniques are:

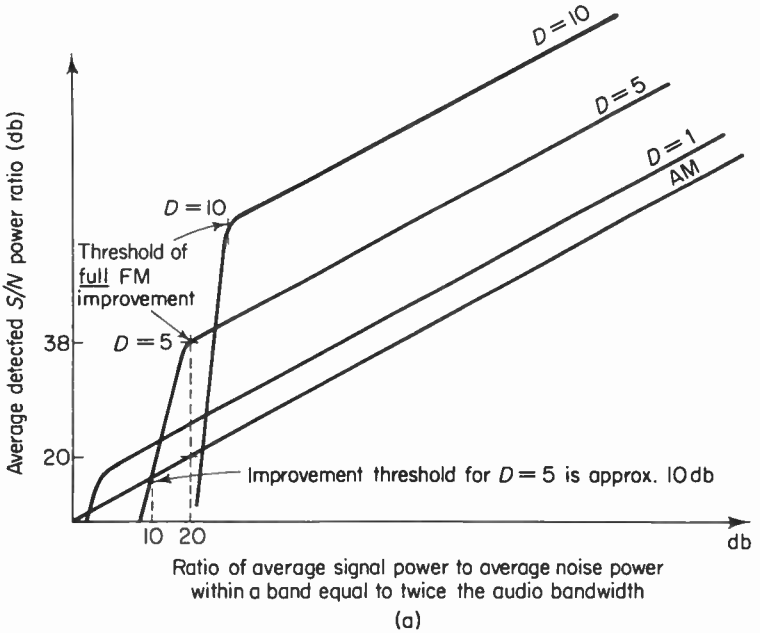
1. Use wide-band FM.
2. Use very wide-band limiters and detectors plus "flat," well-aligned IF stages.
3. Use preemphasis and deemphasis.

## 2-12. Threshold phenomenon

The previous discussions concerning the relative amplitudes of carrier to noise show a rather sharp transition between good and poor FM performance. In AM systems, the detected $S/N$ power ratio (in db) is linearly proportional to the $S/N$ power input to the receiver. This is not quite the case for FM systems.

There is a point below which the frequency-modulation improvement over an equivalent AM system is lost. This point is termed "improvement threshold." For input $S/N$ ratios above this point, the FM system "improves" at a greater rate than the AM system. As the input $S/N$ ratio continues to increase, a point is reached where maximum improvement of FM over AM is realized. This is designated as the threshold of *full* improvement. It is at this input $S/N$ level that Eq. (2–10) is valid. Beyond this point, the FM improvement over AM remains constant and is simply equal to $3D^2$. If greater FM-to-AM improvement is required, a higher deviation ratio is needed.

Figure 2-14 depicts these conditions, and from part (a) certain conclusions can be reached. Below the threshold of *full* FM improvement, the wide-band noise improvement is essentially independent of the deviation ratio ($D$) and is dependent upon the input $S/N$ ratio. Above the threshold of *full* FM improvement, the wide-band improvement is independent of the input $S/N$ ratio (the *slopes* of the FM and AM curves are practically identical) and is fixed for a given $D$.



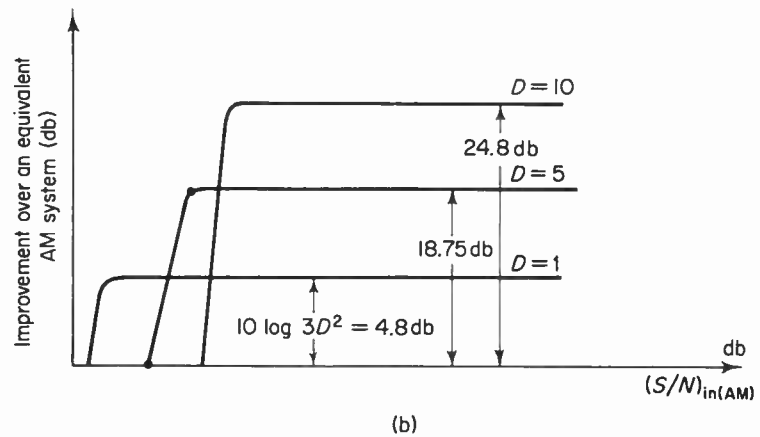




**Fig. 2-14.** FM improvement and threshold effects.

An additional conclusion, not so evident from Fig. 2-14, is that since detected noise power is dependent upon the receiver passband, the receiver with a wider passband is at a disadvantage. If the carrier power to two FM receivers of different bandwidths is reduced, the noise power will "reach" the carrier power in the wider-band receiver first. Therefore, the wide-band system ($D = 10$) requires higher carrier power before reaching the threshold than a narrowband system ($D = 1$). At low carrier power, the input $S/N$ power ratio may be above threshold for the narrow-band system and below it for the wide-band system. Therefore, under the above conditions, the narrow-band system will produce a *higher* average detected $S/N$ power ratio.

Figure 2-14(b) shows the improvement of several FM systems over an equivalent (same audio bandwidth) AM system. It should be noted that below the knees of the curves, the FM systems deteriorate rapidly.

## 2-13. Summary

**1.** Improved noise performance can be realized by large signal-induced frequency deviations.

**2.** The AM disturbance introduced into an FM receiver can be dealt with by means of amplitude limiting.

**3.** The phase and frequency disturbances introduced into an FM receiver can be dealt with by means of wide-band, linear, well-aligned equipment.

**4.** To minimize the possibility of the disturbance "capturing" the receiver from the signal, a flat-topped, accurately aligned and tuned receiver is required.

**5.** Phase and frequency modulation are mathematically related, but are not identical.

**6.** The signal-power to noise-power improvement in a wide-band FM system is proportional to the square of the deviation ratio.

**7.** A narrow-band FM system (deviation ratio = 0.6) provides essentially no noise improvement over an AM system of comparable IF bandwidth.

**8.** "Frequency clipping" (shifting the average value of the spike pattern of Fig. 2-7) can be prevented by using a wide-band detector-limiter.

**9.** FM systems are more efficient than AM in terms of utilizing available allocations.

**10.** The action of an impulse on the audible output of an FM receiver is to produce a series of "pops" and/or "clicks"—depending upon the initial phase of the impulse, the relative amplitudes of the impulse and signal, and so on.

**11.** Proper receiver tuning will minimize the occurrence of "pops."

**12.** The demodulated noise in AM follows a uniform spectrum for all beat frequencies, up to 15 kHz.

**13.** The demodulated noise in FM follows a triangular spectrum for all beat frequencies, up to the 75-kHz deviation limits.

**14.** Preemphasis provides $S/N$ improvement by means of equalizing transmitter deviations in the "noisy" higher audio-frequency range. This tends to flatten the triangular spectrum.

**15.** FM systems exhibit fairly sharp thresholds, below which the signal is completely submerged in noise.

**16.** Narrow-band FM systems can be superior to wide-band FM systems at low operating carrier levels (mobile, military, and the like).

## REFERENCES


1. Goldman, S.: *Frequency Analysis, Modulation and Noise.* McGraw-Hill Book Company, New York, 1948.
2. Lovering, W. F.: *Radio Communication.* David McKay Co., Inc., New York, 1958.
3. Cuccia, C. L.: *Harmonics, Sidebands and Transients in Communications Engineering.* McGraw-Hill Book Company, New York, 1952.
4. Johnstone, G. G.: "Limiters and Discriminators for FM Receivers," *Wireless World,* January 1957.
5. van Slooten, J.: "FM Reception under Conditions of Strong Interference," *Philips Technical Review,* July 1961.
6. Wheeler, H. A.: "Common-Channel Interference Between Two Frequency-Modulated Signals," *Proc. IRE,* January 1942.
7. Corrington, M. S.: "Frequency Modulation Distortion Caused by Common- and Adjacent-Channel Interference," *RCA Review,* 7 (December 1946).


## APPENDIX 2-1

The frequency fluctuation (FM) of the resultant $R$ above and below that of the stronger signal (see Figs. 2-3 and 2-4) is

$$f_N = \frac{1}{2\pi} \cdot \frac{d\phi}{dt} \tag{1}$$

where $d\phi/dt$ represents the FM disturbance.
If

$$\phi = \tan^{-1} \frac{a \sin \omega_B t}{1 + a \cos \omega_B t}, \tag{2}$$

then the instantaneous *radian* frequency of the disturbance is

$$\omega_N = \frac{d\phi}{dt} = \frac{d\left(\tan^{-1} \dfrac{a \sin \omega_B t}{1 + a \cos \omega_B t}\right)}{dt}. \tag{3}$$

Using the relationship

$$\frac{d(\tan^{-1} x)}{dt} = \frac{\dfrac{du}{dt}}{1 + u^2}$$

and letting

$$u = \frac{a \sin \omega_B t}{1 + a \cos \omega_B t},$$

then

$$\frac{d\phi}{dt} = \frac{d\left(\dfrac{a \sin \omega_B t}{1 + a \cos \omega_B t}\right)}{dt} \cdot \frac{(1 + a \cos \omega_B t)^2}{1 + a^2 + 2a \cos \omega_B t}. \tag{4}$$

Using the formula for the derivative of a quotient,

$$\frac{dy}{dt} = \frac{v\dfrac{du}{dt} - u\dfrac{dv}{dt}}{v^2},$$

and letting $v = 1 + a \cos \omega_B t$ and $u = a \sin \omega_B t$, proper substitution gives

$$\frac{d\phi}{dt} = \frac{a\omega_B(\cos \omega_B t + a)}{1 + a^2 + 2a \cos \omega_B t}. \tag{5}$$

Dividing through both sides of Eq. (5) by $2\pi$ gives

$$\frac{1}{2\pi} \cdot \frac{d\phi}{dt} = \frac{af_B(\cos \omega_B t + a)}{1 + a^2 + 2a \cos \omega_B t} = f_N. \tag{6}$$

When $\omega_B t = 0$, $\cos \omega_B t = +1$. This is when the unit carrier (desired signal) and the disturbance $a$ are *in-phase*. The rate of change of phase ($d\phi/dt$) is maximum and the instantaneous frequency is

$$f_N = \frac{af_B(1 + a)}{1 + a^2 + 2a} = \frac{af_B}{a + 1} \tag{7}$$

This frequency corresponds to point $X$ on Fig. 2-7.

When $\omega_B t = 180°$, $\cos \omega_B t = -1$. This is when the unit carrier and the disturbance $a$ are in *phase-opposition*. The rate of change of phase ($d\phi/dt$) is again a maximum and the instantaneous frequency is

$$f_N = \frac{af_B(a - 1)}{1 + a^2 - 2a} = -\frac{af_B}{1 - a}. \tag{8}$$

This frequency corresponds to point $Y$ on Fig. 2-7.

# 3

# NOISE

## 3-1. Introduction

The layman usually thinks of noise as sounds that are not harmonically related and are therefore not intelligible. In an electronic sense, voltage or current variations that tend to mask the desired signal are termed noise. At the output of the receiver, these noise currents produce rushing, hissing, or roaring sounds. Since their effect is to mask or interfere with the incoming signal, the maximum obtainable sensitivity of a receiver is limited by the electrical noise generated in the part of the receiver that obtains the greatest total amplification. This means that the RF amplifier and the mixer circuits are the most important circuits with regard to weak signal sensitivity.

Noise is contributed to the receiver by three main sources: (1) thermal agitation, (2) tube and transistor noises, (3) noise sources external to the receiver. In this chapter we shall consider the noise sources that are internal to the receiver, and what can be done to reduce their effect on the receiver.

## 3-2. Thermal agitation

All matter is made up of atoms or of groups of atoms called molecules. Conductors, in particular, at room temperature will release from their atoms vast numbers of free electrons. Owing to the energy imparted to them by the surrounding heat, all of these electrons, atoms, and molecules will be in violent motion. The greater the heat, the more violent the motion. Thus, the temperature of a body is the means by which the average motion (kinetic energy) of the component particles is measured. The temperature at which

all molecular motion ceases is minus 273.16 degrees centigrade (−459.68°F). This temperature is called absolute zero. The temperature scale that uses absolute zero as its reference is called the Kelvin or absolute scale. To convert from the centigrade scale to the Kelvin scale the following conversion formula can be used.

$$T = (\text{centigrade degrees}) + 273.16 \qquad (3\text{–}1)$$

where $T$ is in degrees Kelvin (K). This would mean, for example, that if an FM receiver's ambient temperature were 60°C (140°F), its absolute temperature would be 333.16°K.

In a conductor, millions upon millions of free electrons are in motion owing to the energy contributed to them by the surrounding heat. Each of these free electrons constitutes a minute current. These currents result in thermal noise—or, as it is sometimes called, thermal agitation noise. The direction and speed of the free electrons are random and of infinite variety. These currents, therefore, generate complex voltages whose waveform (over a period of time) can be resolved into a range of sinusoidal frequency components that appear across the conductor and are almost infinite in number. See Fig. 3-1. The amplitude of these components is seen to be constant up to a frequency of about $10^{13}$ Hz. Such noise (constant-amplitude noise-component spectrum) is often called fluctuation or white noise. Random noise is called white noise because of its similarity to white light (white light consists of equal-energy spectral components).

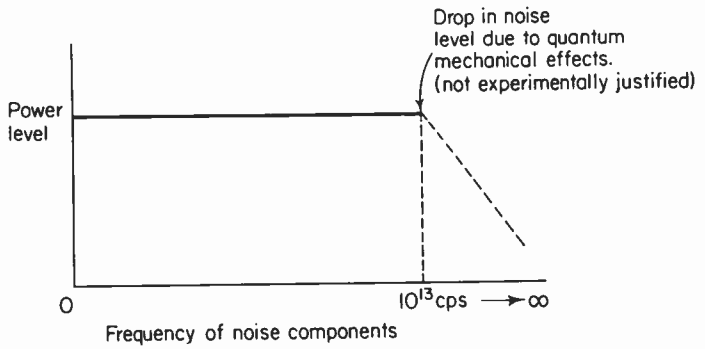



**Fig. 3-1.** Spectrum of sine wave components which random noise can be shown to be made up of.

White noise differs from impulse noise in that impulse noise can be resolved into an infinity of infinitely small sinusoidal frequency components whose amplitude decreases with frequency. See Fig. 3-2 and refer to Chapter 2 for more details regarding impulse noise.

White noise develops no DC component because there is just as good a possibility that electrons will move in one direction as another—therefore over a period of time the average (DC) value is zero.

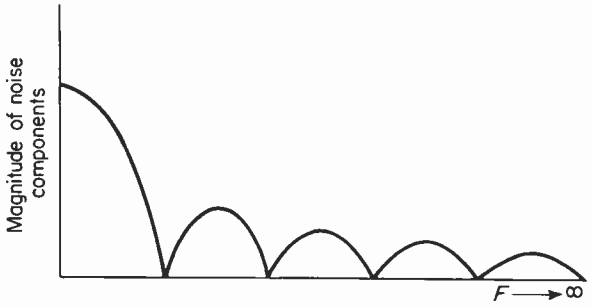



**Fig. 3-2.** Spectrum of impulse noise components.

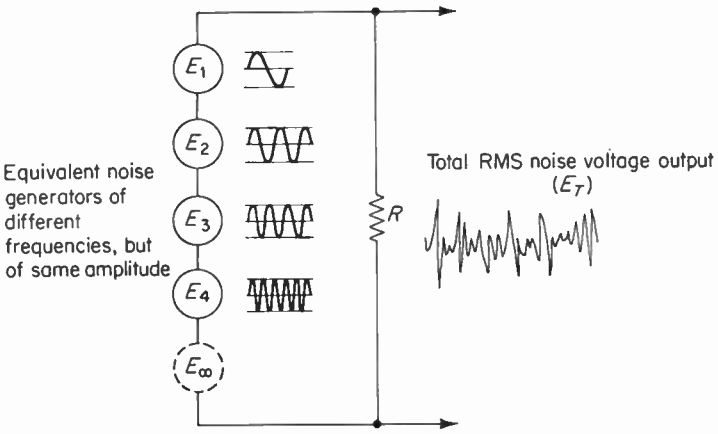
Noise components can be treated as different frequency generators which can be connected in series. Since all the generators are of different frequency and phase, the only way to evaluate such a combination of generators is on the basis of total power developed by the generators. As shown in Fig. 3-3, the voltage developed by such a combination of generators is an rms voltage whose magnitude is equal to the square root of the sum of the squares of all of the effective noise generators. Therefore, noise voltages and currents are always computed on an rms or effective value basis.

The magnitude of the noise voltage (rms) developed across a conductor depends upon three factors: temperature, resistance, and bandwidth. It is fairly obvious that an increase in temperature will, by definition, cause an increase in the velocity of free electrons in a conductor. This constitutes an increase in noise current through the conductor. As a result, an increased noise voltage appears across the conductor equal in magnitude to the noise current multiplied by the resistance of the conductor. Of course if the temperature should decrease, the noise voltage across the conductor would also decrease.

For a given temperature the average motion of the free electrons in a conductor is the same whether the resistance of the conductor is large or small. This means that the random-noise currents in a conductor of low resistance will generate a terminal voltage smaller than the voltage generated across a conductor whose resistance is large.

Bandwidth, *not the particular frequency of operation*, is the important consideration with regard to noise voltage generated by thermal agitation. This means that a wide-band amplifier will tend to generate more noise than will a narrow-band amplifier regardless of the frequency of operation, all other factors being equal. The reason is that the wider the bandpass of an amplifier the greater will be the *number* of noise components amplified (see Fig. 3-1). These amplified noise components will then add vectorially to one another, producing the noise output. Obviously, a narrow-band amplifier will pass fewer noise components than will a wide-band amplifier, and thus its noise output voltage will be smaller.

It is interesting to note that bandpass, in the discussion of noise, is not




The total power ($P_t$) dissipated across $R$ must be equal to the sum of the power delivered to $R$ by each noise generator. Thus;

$$P_t = P_1 + P_2 + P_3 + P_4 + \text{etc.}$$

$$\therefore E_t^2/R = E_1^2/R + E_2^2/R + E_3^2/R + E_4^2/R \ldots\ldots\ldots$$

Multiply both sides by $R$. Thus;

$$E_t^2 = E_1^2 + E_2^2 + E_3^2 + E_4^2 \ldots\ldots\ldots \text{(all voltages are rms)}$$

$$\therefore E_t = \sqrt{E_1^2 + E_2^2 + E_3^2 + E_4^2} \ldots\ldots\ldots \text{(rms voltages)}$$

**Fig. 3-3.** Derivation of an equation for finding the total rms voltage of a series circuit containing generators of different frequencies.

defined as the half-power points on the frequency-response curve of the amplifier or device in question. Mathematically, the amplifier's response is modified so as to avoid the necessity of dealing with noise voltages of different magnitudes (since the amplifier's bandpass is not ideal) when computing the total output noise voltage. The equivalent noise bandpass of the amplifier is rectangular (see Fig. 3-4) and its bandpass is found to enclose the same area as the actual amplifier's bandpass. It is also to be noted that the peak amplitudes of both curves are the same. The noise bandwidth of a single-stage high-$Q$ tuned amplifier will be found to be 1.57 times the half-power bandwidth of the stage. For most purposes where multiple tuned amplifiers are employed (good skirt response) the half-power bandwidth gives sufficient accuracy.

Granular-type carbon resistors will develop higher levels of noise than

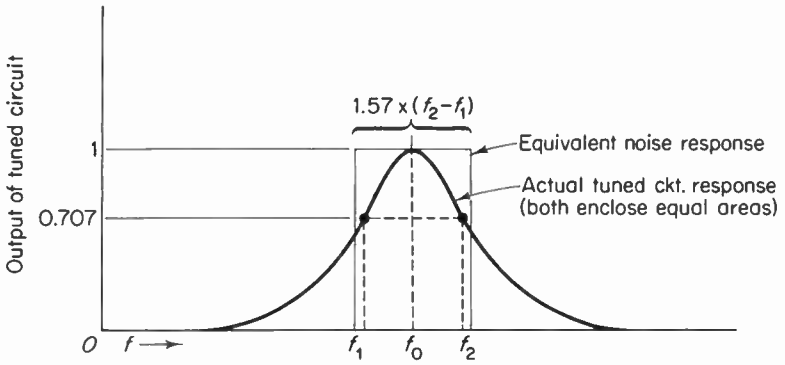



**Fig. 3-4.** High $Q$ tuned circuit response may be converted into an equivalent flat response in order to avoid the necessity of dealing with noise of different magnitudes.

will film or wirewound resistors, because contact effects between the grains of carbon introduce additional noise. Direct current flowing through a carbon resistor will also tend to increase the noise level above that generated by thermal agitation alone. This is due to internal arcing between the carbon grains which make up the resistor.

The basic work on thermal noise was done in 1928 by J. B. Johnson. This is why thermal noise is sometimes called Johnson noise. The equation Johnson developed for the rms value of noise is

$$e_{\text{rms}}^2 = 4KTRB \tag{3-2}$$

where

$T$ = the temperature in degrees absolute or Kelvin to which the resistance in question is raised.

$R$ = the resistance across which the thermal noise is developed. Its unit is ohms.

$B$ = the effective noise bandpass in cps.

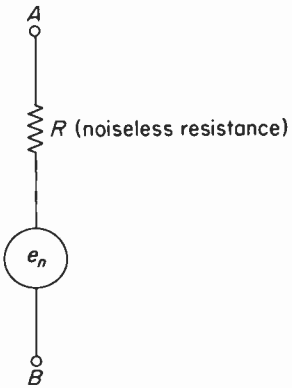
$K = 1.37 \times 10^{-23}$ joules per degree Kelvin. This is called Boltzmann's constant and represents the amount of energy contributed to the system by each degree rise in temperature.

This equation embodies the three factors $R$, $T$, and $B$, which determine the degree of thermal noise generated across a conductor. The magnitude of this voltage, for most conditions, is of the order of microvolts, so it can become important only when the overall amplification is very high.

For ease of calculation the noise voltage of Eq. (3-2) can be thought of as a noise generator in series with a perfect resistor that is incapable of

generating any noise. The magnitude of this perfect resistor is equal in value to that of the resistance which is actually producing the thermal noise. See Fig. 3-5(a). This circuit is a Thevenin constant-voltage equivalent circuit, which can be converted into a Norton constant-current equivalent circuit as shown in Fig. 3-5(b). Either equivalent circuit may be employed for thermal-noise calculations, using whichever method will produce the greatest mathematical simplifications.

A number of examples will clarify some of the problems that arise in the computation of thermal noise. Consider a 1000-ohm wirewound resis-




$$e_{n(\text{rms})} = \sqrt{4KTR \quad B}$$

(a) Thevenin equivalent noise generator.

To convert a Thevenin equivalent noise generator to a Norton equivalent noise generator the following procedure is employed;

Step 1. Short Thevenin equivalent circuit terminals $A$ to $B$.

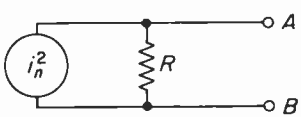
Step 2. Calculate current flowing through short circuit, this is the Norton equivalent current. Thus:

$$i_n = \frac{e_n}{R} = \frac{\sqrt{4KTRB}}{R}$$

$$i_n^2 = \frac{4KTRB}{R^2} = \frac{4KTB}{R}$$

Step 3. Use this current as the constant current generator in the Norton equivalent circuit.

Step 4. Place $R$ (Thevenin) across $i_n^2$ generator.




$$i_n^2 = \frac{4KT \quad B}{R}$$

(b) Norton equivalent circuit.

**Fig. 3-5.** (a) Thevenin constant voltage equivalent noise generator. (b) Conversion of Thevenin noise generator into an equivalent Norton constant current noise generator and its circuit diagram.

tor, at a temperature of 290°K (room temperature), which is placed at the input to an amplifier whose bandpass is 200 kHz. The noise voltage developed across the resistor due to thermal agitation will be

$$\begin{aligned} E_{n(\text{rms})} &= \sqrt{4KTRB} \\ &= \sqrt{4 \times 1.37 \times 10^{-23} \times 2.9 \times 10^{2} \times 10^{3} \times 2 \times 10^{5}} \\ &= \sqrt{3.18} \times 10^{-6} \\ &= 1.77 \text{ microvolts rms.} \end{aligned}$$

The peak or crest value of random noise has been found to be about four times larger than the rms value. Thus the peak value would be 7.1 microvolts.

Now consider the case of a parallel resonant circuit. As is well known, at resonance such a circuit is resistive and will, therefore, contribute thermal noise to an amplifier as would any other resistive element. At frequencies other than resonance the resistive component of the total impedance is reduced and thus contributes less noise. To compute the total rms noise introduced throughout the entire response of the tuned circuit would require that the noise level be calculated, by means of Eq. (3–2) [bandwidth $B$ equals 1 Hz], for an infinite number of adjacent frequencies. The total noise power generated throughout the entire response is the sum of all of the noise levels computed. This method is best accomplished by means of the calculus. For ease in calculation we can reach a good approximation of the noise generated in the passband by assuming that in a high-$Q$ tuned circuit the resistive component of impedance is constant. Thus, the noise generated in a parallel tuned circuit can be calculated by means of Eq. (3–2). Consequently, if it is assumed that the resistive component of the antiresonant circuit in Fig. 3-6 is constant throughout its bandpass, then the rms noise

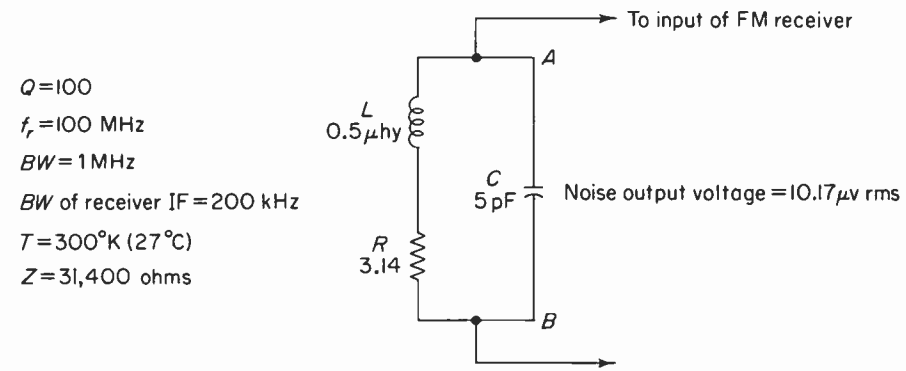



**Fig. 3-6.** An antiresonant circuit which produces a noise output equivalent to a 31,400 ohm resistor.

voltage developed may be calculated as follows:

$$\text{dynamic resistance at resonance} = QX_L$$
$$= 6.28 \times 10^8 \times 0.5 \times 10^{-6} \times 10^2$$
$$= 3.14 \times 10^4 \text{ ohms}$$

and therefore

$$e_n = \sqrt{4KTRB}$$
$$= \sqrt{4 \times 1.37 \times 10^{-23} \times 3 \times 10^2 \times 3.14 \times 10^4 \times 2 \times 10^5}$$
$$= 10.17 \text{ microvolts rms.}$$

Notice that in solving for the noise voltage the 200-kHz bandpass of the IF amplifier was employed rather than the tuned-circuit bandpass of $(f_r/Q) = 1$ MHz. This is because only those noise components which can pass through the bandpass of the amplifier will appear at the output of the amplifier.

It should be pointed out at this juncture that such circuit elements as inductance and capacitance do not generate thermal noise. Thermal noise is produced by the expenditure of energy [thus Eq. (3–2) is really a power equation] in the form of heat, whereas reactive elements can only store energy, not dissipate it. Consequently, these elements (disregarding such losses as distributed resistance, hysteresis losses, and so on) do not generate thermal noise. The reader may then wonder at the fact that in an antiresonant circuit, consisting of $L$, $C$, and $R$ components, the effective noise resistance at resonance is equal to $QX_L$, which is $Q^2$† times the series resistance of the inductive branch. This comes about because the series resistive losses in the circuit act as a noise generator that "sees" a *series resonant* circuit. Thus the noise

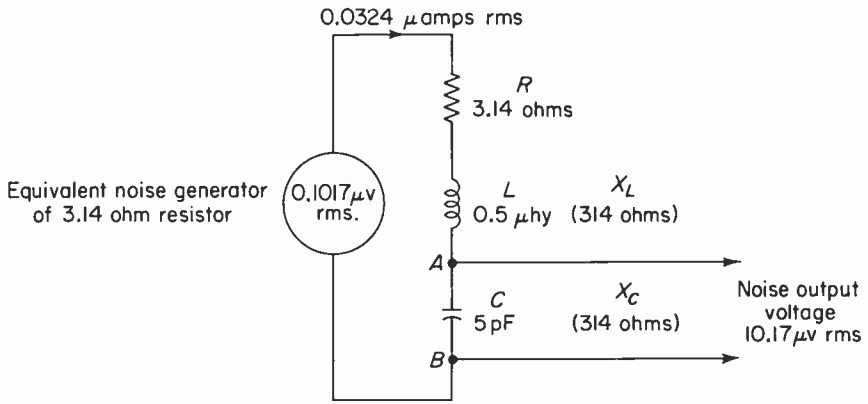



**Fig. 3-7.** The circuit of Fig. 3-6 is redrawn showing that the 3.14 ohm internal resistance and the effects of series resonance are responsible for the large noise output voltage.

†This is obtained as follows: $Q = X_L/R$; multiplying both sides by $Q$, $Q^2 = QX_L/R$, $Q^2R = QX_L$. Thus, $Q^2R = Z$ at antiresonance.

output across the tuning capacitor (terminals of the antiresonant circuit) is magnified $Q$ times the noise generated by the series resistance of the antiresonant circuit. The antiresonant circuit of Fig. 3-6 is redrawn in Fig. 3-7 showing the Thevenin equivalent noise generator of the 3.14-ohm series resistance driving a series resonant circuit. The noise generated across the 3.14-ohm resistance is

$$\begin{aligned} e_n &= \sqrt{4KTRB} \\ &= \sqrt{4 \times 1.37 \times 10^{-23} \times 3 \times 10^2 \times 2 \times 10^5 \times 3.14} \\ &= 0.1017 \text{ microvolts rms.} \end{aligned}$$

Thus the noise output across the tuning capacitor (terminals of the antiresonant circuit) is magnified $Q$ times the noise generated by the series resistance of the antiresonant circuit. Since the circuit $Q$ of the series resonant circuit is 100, the noise output voltage will be

$$\begin{aligned} e_{cn} &= e_n \times Q = 0.1017 \times 10^{-6} \times 10^2 \\ &= 10.17 \text{ microvolts rms,} \end{aligned}$$

which is exactly the same as the noise level found in Fig. 3-6.

## 3-3. Shot noise

An important contributor of noise in a high-amplification receiver is shot noise. This form of white noise is generated in the vacuum tube because of the random emission of electrons from the cathode and their random arrival at the plate.

In Fig. 3-8 a triode is arranged so that a DC plate current of 6 ma flows through the tube. In the same figure is a graph illustrating the flow of DC current versus time. A certain portion of curve is shown greatly enlarged. It can be seen that the plate current actually consists of two components, a DC component and a very small AC component. The AC component represents the variation in the number of electrons that arrive at the plate each instant. The magnitude of the AC component of plate current for a triode is found by means of the following equation:

$$i^2_{n(\text{rms})} = 4K(\theta T_c)\frac{g_m}{\sigma}B, \tag{3-3}$$

where

$\sigma = .88$,

$\theta = .644$,

$K =$ Boltzmann's constant, $1.37 \times 10^{-23}$ joules per degree Kelvin,

$T_c =$ cathode temperature in degrees Kelvin,

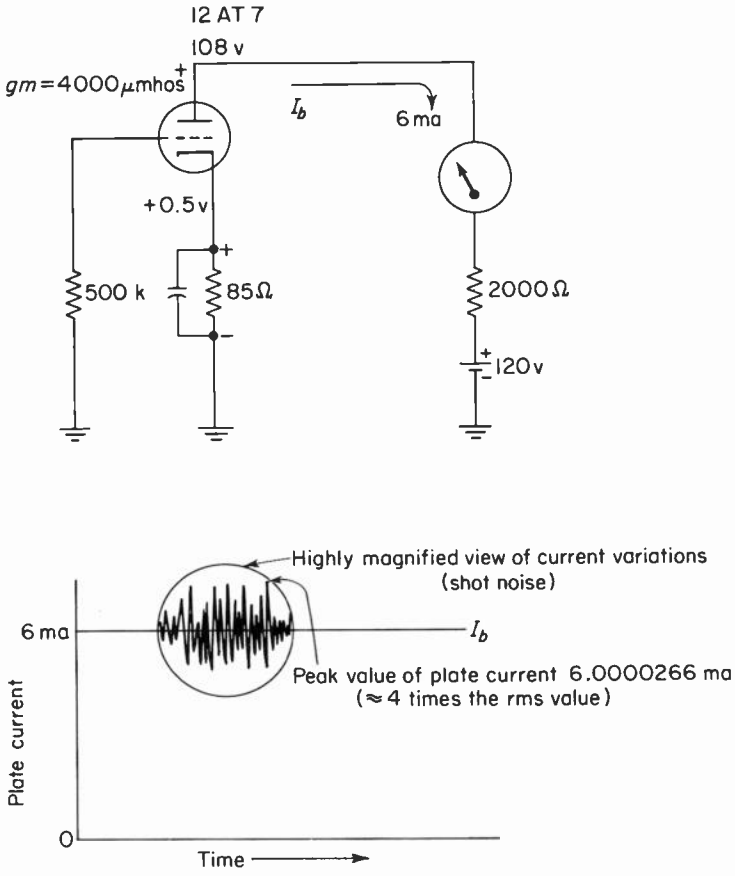



**Fig. 3-8.** Variations in the plate current of a triode amplifier due to shot noise.

$g_m$ = transconductance of a triode in micromhos,

$B$ = bandwidth of the system.

This equation assumes that the tube is not temperature-saturated (the condition where all emitted electrons are collected by the plate and no space charge is formed) but is space-charge limited. The noise level of the space-charge limited tube is lower than that of a temperature-limited tube. This is because the space charge, by means of its repelling action on newly emitted electrons, acts to make the distribution of electron velocities more uniform. Plate-current variations are, therefore, less than in the temperature-saturated case.

If it is assumed that the oxide coated cathode temperature of the triode in Fig. 3-8 is 1000°K, and if the bandpass is 200 kHz, the rms noise current

developed by shot noise is

$$i_n = \sqrt{4K(\theta T_c)\frac{g_m}{\sigma}B}$$
$$= \sqrt{4 \times 1.37 \times 10^{-23} \times 10^3 \times 4 \times 10^{-3} \times 2 \times 10^5 \times 7.3 \times 10^{-1}}$$
$$= \sqrt{32.4 \times 10^{-18}}$$
$$= 5.7 \times 10^{-9} \text{ amperes rms.}$$

The noise voltage developed across the load resistance (assumed to be noise-free) is simply

$$e_n = i_n \times R_L$$
$$= 5.7 \times 10^{-9} \times 2 \times 10^3$$
$$= 11.4 \times 10^{-6} \text{ volts or 11.4 microvolts rms.}$$

Shot noise actually appears in the plate circuit of an amplifier, but for ease of calculation it is generally referred to the grid where thermal and other noise sources are located. This can be done by placing a *fictitious* resistor into the grid circuit of the tube in question. The value of this resistor is chosen so that the amount of noise generated across the resistor at room temperature when amplified by a noiseless tube is equal to the shot noise ordinarily generated. This equivalent noise resistance is considered to act in series with the tube's grid. The magnitude of the equivalent noise resistance is found in the folllowing manner:

The AC component of plate current of a tube is equal to $g_m e_g$. Therefore, the rms noise current must be

$$i_n^2 = g_m^2 \times e_g^2. \tag{3-4}$$

Thus,

$$\frac{i_n^2}{g_m^2} = e_g^2. \tag{3-5}$$

Substituting Eq. (3-2) for $e_g^2$ and Eq. (3-3) for $i_n^2$,

$$\frac{4K(\theta T_c)\frac{g_m}{\sigma}B}{g_m^2} = (4KTR_{eq}B),$$

which is

$$\frac{\theta T_c}{\sigma g_m} = TR_{eq}. \tag{3-6}$$

Therefore

$$\frac{\theta T_c}{T\sigma g_m} = R_{eq}. \tag{3-7}$$

Since $T_c$ is assumed to be 1000°K and $T$ represents room temperature (290°K), $\theta T_c/\sigma T$ equals 2.5. Thus the equivalent noise resistance will be

$$R_{eq} = \frac{2.5}{g_m} \qquad \text{for a triode.} \tag{3-8}$$

Typical values of noise resistance for a triode are about 600 ohms.

If the equivalent noise resistance were to be calculated for the circuit of Fig. 3-8, it would be equal to 625 ohms. The thermal noise generated by this resistance at room temperature is $1.42 \times 10^{-6}$ volts rms. Since this voltage is applied to the grid of the tube, it will be amplified by eight times ($g_m R_L$). Thus the output noise developed by the tube will be 11.4 $\mu$v rms—which, of course, is the same as the noise level calculated by means of the noise-current equation.

## 3-4. Partition noise

Pentodes are about six times noisier than triodes because of an additional noise source called partition noise. In the pentode, although the space charge tends to smooth the total cathode current, the division or partition of current between the screen and the plate is random. Since the screen current is subtracted from the cathode current at a random rate, the randomness of the plate current is increased. As was the case with shot noise, the noise output of the pentode may be referred to the control grid of the tube. The equivalent noise resistance of a pentode is given by

$$R_{eq} = \frac{I_b}{I_b + I_{c2}}\left(\frac{2.5}{g_m} - \frac{20 I_{c2}}{g_m^2}\right) \tag{3-9}$$

where

$I_b$ = DC plate current of the tube,

$I_{c2}$ = DC screen current,

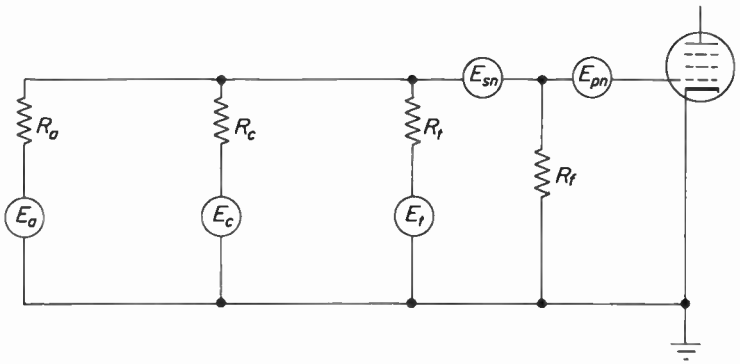
$g_m$ = transconductance of the pentode.

This expression is for the combined effects of shot noise and partition noise.

Values of equivalent noise resistance to be expected for a pentode range from 100 ohms (tube type 7788) to 10,000 ohms and higher.

The equivalent noise generator due to partition noise also acts in series with the grid of the tube as does the shot-noise equivalent generator. See Fig. 3-9. These generators and their associated thermal equivalent resistances are the only noise components that are connected in series with the grid. All other noise sources and their associated thermal resistances are connected in shunt with one another and are in turn connected in series with the shot and partition-noise generators. The reason is that shot noise and partition noise are generated in the tube and therefore act in series with noise sources that develop their noise voltages in circuits outside the tube.

It should be remembered that both partition-noise and shot-noise resistances are fictitious and, therefore, do not load the circuit that they are connected into. Consequently, selectivity, sensitivity, or operating conditions of the circuit will not be affected.




$E_a$ Thermal noise voltage due to $R_a$
$E_c$ Thermal noise voltage due to $R_c$
$E_t$ Thermal noise voltage due to $R_t$
$E_{sn}$ Thermal noise voltage generated by the equivalent shot noise resistance.
$E_{pn}$ Thermal noise voltage generated by the equivalent partition noise resistance
$R_a$ Antenna radiation resistance
$R_c$ Resistance developed by resonant tuned circuit
$R_t$ Input resistance due to transit time
$R_f$ Input resistance due feedback effects such as cathode lead inductance and Miller effect. Does not generate noise.

**Fig. 3-9.** Equivalent circuit of a pentode RF amplifier illustrating the various noise sources as referred to the grid.

## 3-5. Transit-time noise

Radio-frequency amplifiers that are to operate at frequencies where transit time of electron flight is an important part of one cycle of input signal suffer from additional noise called "induced grid noise." This internally generated tube noise is associated with resistive loading effects, between grid and cathode, which become important at high frequencies. The noise-generating resistance referred to here is not at all fictitious, as were the equivalent shot and partition-noise resistances, but is very real and exhibits loading effects as well as thermal noise that are both frequency-dependent. A real resistance in this instance does not refer to a physical object but rather to the voltage current relationships existing in the grid circuit and the phase angle between them. The magnitude of this transit-time resistance may be determined by means of the following expression:

$$R_{tt} = \frac{1}{Kg_m(2\pi f T_{kg})^2} \tag{3-10}$$

where

$K$ = a constant, typical value 0.05,

$g_m$ = transconductance of the tube,

$f$ = frequency of operation,

$T_{kg}$ = transit time between cathode and grid, typical value 0.001 to 0.003 $\mu$sec.

Notice that the loading resistance is inversely proportional to the square of the frequency. Thus the noise developed by this resistance will increase quite rapidly as the frequency of operation is decreased. However, it becomes more important at higher frequencies, because it shunts other noise sources and thereby determines the noise output.

From a physical point of view induced grid noise is due to the random motion of electrons as they pass the control grid on their way to the plate, electrostatically inducing minute noise currents into the control grid. The noise currents in the grid circuit set up noise voltages on the grid, which in turn vary the plate current at the noise rate, thereby contributing additional noise to the noise already existing at the output of the amplifier due to other sources.

The magnitude of the grid induced voltage may be calculated by

$$e_{R_{tt}(\text{rms})} = \sqrt{20KTR_{tt}B} \tag{3–11}$$

This formula is merely the expression for thermal noise given in Eq. (3–2). Certain differences exist in that the temperature ($T$) is fixed at room temperature, and the numerical coefficient is changed from 4 to 20. The increased coefficient can be attributed to the nearness of the cathode to the grid, thereby increasing the effective temperature five times above room temperature. It should be noted that since $R_{tt}$ is inversely proportional to the square of the frequency of operation, the noise level generated by this resistance will be frequency-dependent.

Other resistive components, which are introduced into the grid circuit by such feedback devices as Miller effect and cathode lead inductance (to be discussed more fully in a later section), do not introduce noise into the grid circuit, but do influence the signal-to-noise ratio of the receiver. This is because voltage division takes place between the circuit impedance and the feedback resistive components ($R_F$), so that less actual signal appears from grid to cathode of the RF amplifier than the signal developed at the input by the antenna. As can be seen from Fig. 3-9, the most desirable condition for developing maximum signal between grid and cathode occurs when the feedback loading resistance is large (less voltage division). Unhappily, these loading components become smaller as the frequency increases; therefore, less and less signal appears between grid and cathode of the tube. Of

course, this reduction of the effective input signal by means of $R_F$ adversely affects the signal-to-noise ratio.

## 3-6. Noise figure

The noise factor is the figure of merit for sensitivity of a receiver. It is defined as the signal-to-noise power ratio of an ideal receiver (noise generated by the antenna only) divided by the signal-to-noise power ratio of the actual receiver in question. The noise factor for vacuum tubes or field-effect transistors (FET) may be expressed either as a power ratio or a voltage ratio. It is most convenient for purposes of derivation to express it in terms of a voltage ratio. Thus

$$F = \frac{(S_i/N_i)^2}{(S_o/N_o)^2} = \frac{S_i^2 N_o^2}{S_o^2 N_i^2}, \tag{3-12}$$

where

$S_i$ = input signal voltage

$S_o$ = output signal voltage

$N_i$ = input noise voltage

$N_o$ = output noise voltage

The noise factor for a circuit such as that of Fig. 3-10, which assumes that the active device contributes no noise of its own, may be determined by evaluating the signal and the noise voltages in the circuit. Thus

$$N_i^2 = 4KTBR_s \tag{3-13}$$

and

$$N_o^2 = 4KTB\left(\frac{R_s R_{eq}}{R_s + R_{eq}}\right), \tag{3-14}$$

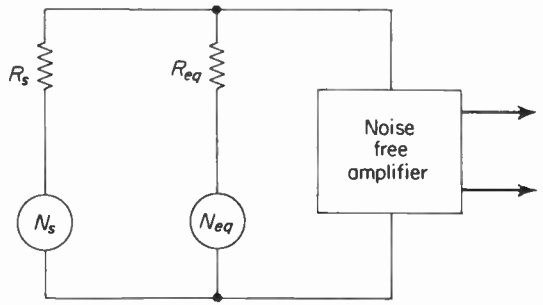



**Fig. 3-10.** The input circuit of a noise free amplifier. The noise voltage generated by $R_s$, the internal resistance of the source is $N_s$. The resistance designated Req. is the parallel equivalent of the circuit resistance ($R_c$) and the input resistance ($R_D$) of the amplifier. The noise voltage generated by Req. is Neq.

and since the output signal voltage is proportional to the input signal voltage and inversely to the total resistance of the voltage divider,

$$S_o^2 = \left[S_i\left(\frac{R_{eq}}{R_s + R_{eq}}\right)\right]^2 \tag{3-15}$$

where $R_s$ is the source resistance and $R_{eq}$ is the parallel equivalent of input resistance ($R_d$) of an assumed noiseless device (vacuum tube, transistor, or FET), and the shunt resistance ($R_c$) of any tuned circuit associated with the input. The noise factor may be found by substituting Eqs. (3–13), (3–14), and (3–15) into Eq. (3–12). Thus

$$F = \frac{S_i^2 N_o^2}{S_o^2 N_i^2} = \left(\frac{R_s + R_{eq}}{R_{eq}}\right)^2 \cdot \left(\frac{\frac{R_s R_{eq}}{R_s + R_{eq}}}{R_s}\right),$$

$$F = \left(\frac{R_s + R_{eq}}{R_{eq}}\right)^2 \cdot \frac{R_s R_{eq}}{(R_s + R_{eq})R_s} = \frac{R_s + R_{eq}}{R_{eq}},$$

$$F = 1 + \frac{R_s}{R_{eq}}. \tag{3-16}$$

or, expanding $R_{eq}$,

$$F = 1 + \frac{R_s}{R_c} + \frac{R_s}{R_d}.$$

When the noise factor is expressed in decibels it is referred to as the noise figure. Hence

$$NF = 10 \log F. \tag{3-17}$$

It is apparent from Eqs. (3–16) and (3–17) that under matched conditions, where $R_s = R_{eq}$, the noise figure is 3 db. It is also apparent that the smaller the ratio $R_s/R_{eq}$ becomes, the better the noise figure becomes. Thus, under ideal conditions the noise figure will be 0 db.

If the noise generated by the active device is to be included in the calculations for noise figure, the equivalent circuit appears as shown in Fig. 3-11. In this diagram $R_n$ represents the equivalent noise resistance of the device and is equal to $2.5/g_m$ for a triode and $2/3g_m$ for an FET. The noise factor of the circuit may be determined from the equivalent circuit as follows:

The noise input is

$$N_i^2 = 4KTR_s(BW). \tag{3-18}$$

The noise output is

$$N_o^2 = 4KT(BW)\left(\frac{R_s R_{eq}}{R_s + R_{eq}} + R_n\right). \tag{3-19}$$

The signal output is

$$S_o^2 = \left[S_i\left(\frac{R_{eq}}{R_{eq} + R_s}\right)\right]^2. \tag{3-20}$$

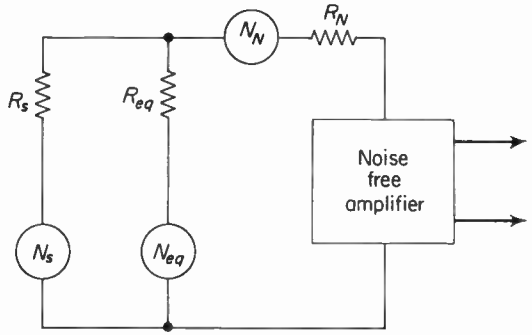



**Fig. 3-11.** The equivalent circuit of an amplifier where the noise generated by the amplifier is represented by an equivalent noise resistance designated $R_N$.

Combining Eqs. (3–18), (3–19) and (3–20) in Eq. (3–12) gives

$$F = \frac{(R_{eq} + R_s)^2}{R_{eq}^2} \cdot \frac{\dfrac{R_s R_{eq}}{R_s + R_{eq}} + R_n}{R_s}, \tag{3-21}$$

which may be simplified to

$$F = 1 + \frac{R_n}{R_s} + \frac{R_s}{R_{eq}} + \frac{2R_n}{R_{eq}} + \frac{R_s R_n}{R_{eq}^2}. \tag{3-22}$$

From this equation it can be seen that since $R_n$ is smaller for an FET than for a vacuum tube, an FET will have a lower noise figure than a vacuum tube.

Since $R_s$ may be varied by adjusting the turns ratio of a coupling transformer, it is important to know the value of $R_s$ that will provide the best noise figure. This may be done by taking the derivative of Eq. (3–22) and setting it equal to zero. Thus

$$\frac{dF}{dR_s} = \frac{-R_n}{R_s^2} + \frac{1}{R_{eq}} + \frac{R_n}{R_{eq}^2}, \tag{3-23}$$

$$\frac{-R_n}{R_s^2} + \frac{1}{R_{eq}} + \frac{R_n}{R_{eq}^2} = 0. \tag{3-24}$$

The optimum value of $R_s$ is

$$R_s = \sqrt{\frac{R_{eq} R_n}{1 + \dfrac{R_n}{R_{eq}}}}. \tag{3-25}$$

If the equivalent circuit resistance ($R_{eq}$) is very much larger than the noise resistance ($R_n$), as would normally be the case with vacuum tubes and FET's, then Eq. (3–25) reduces to

$$R_{s_{opt}} = \sqrt{R_{eq} R_n}. \tag{3-26}$$

Notice that the conditions for lowest noise factor ($R_s = R_{s_{opt}}$) do not coincide with the conditions for maximum power transfer. Thus a signal-power loss must be tolerated if the optimum noise figure is to be achieved.

## 3-7. Two-stage noise figure

The noise figure is determined primarily by the characteristics of the first RF amplifier and its associated input circuit. On occasion it is necessary to determine the composite noise factor (F) of an amplifier followed by a second stage. This may be done by first noting that Eq. (3–12) may be written

$$F = \frac{S_i^2/N_i^2}{S_o^2/N_o^2} = \frac{S_i^2/R_i/N_i^2/R_i}{S_o^2/R_o/N_o^2/R_o} = \frac{P_{si}/P_1}{P_{so}/P_2} = \frac{P_{si}/P_2}{P_{so}/P_1} = \frac{P_2}{GP_1} \tag{3-27}$$

where $P_{si}$ is the input signal power, $P_{so}$ is the output signal power, $P_1$ is the input noise power, $P_2$ is the output noise power, and $G$ is the power gain of the amplifier. Then, using Fig. 3-12, we may obtain the expression for the overall noise figure for two stages. Thus

$$F = \frac{P_2}{GP_1},$$

$$\therefore \quad F = \frac{(P_1 + P_3)G_1 + P_4}{G_1P_1}$$

$$= \frac{P_1G_1}{P_1G_1} + \frac{P_3G_1}{P_1G_1} + \frac{P_4}{P_1G_1}$$

$$= 1 + \frac{P_3}{P_1} + \frac{P_4}{P_1G_1}, \tag{3-28}$$

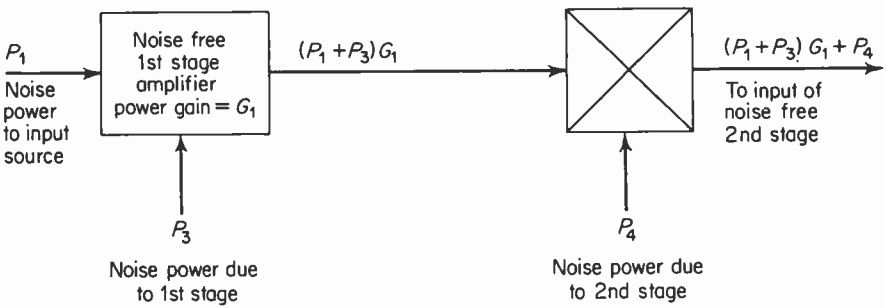



**Fig. 3-12.** The above equivalent circuit indicates the noise contributions of a 2 stage amplifier. This diagram is used to determine the effect upon the overall noise figure as the result of adding a second stage.

where $G_1$ is the power gain of the first stage, $P_3$ is the noise power contributed by the first amplifier, and $P_4$ is the noise power contributed by the second amplifier. Notice that the term $1 + P_3/P_1$ of Eq. (3–28) is another way of expressing Eq. (3–16), the noise factor of the first stage ($F_1$). Therefore, Eq. (3–28)

may be written

$$F = F_1 + \frac{P_4}{P_1 G_1}. \tag{3-29}$$

Again referring to Eq. (3–16) we can see that the ratio $P_4/P_1$ may be expressed in terms of the noise factor of the second stage ($F_2$). Thus

$$F_2 = 1 + \frac{P_4}{P_1} \quad \text{or} \quad \frac{P_4}{P_1} = F_2 - 1. \tag{3-30}$$

Substituting Eq. (3–30) into Eq. (3–29) yields

$$F = F_1 + \frac{F_2 - 1}{G_1}. \tag{3-31}$$

From this equation it can be seen that if the power gain of the first stage is very much greater than $F_2 - 1$, then the noise figure of the system is determined by the first stage.

The noise figure is sensitive to the frequency of operation. As the frequency is increased, the noise figure also increases—for two reasons. First, at low frequencies external noise (cosmic, lightning, and so on) is high compared to internal receiver noise. Therefore, it can be said that the antenna establishes the noise figure at frequencies below 50 MHz. Second, for vacuum tubes at high frequencies, transit time and cathode lead loading (in the case of a pentode RF amplifier) increase ($R$ decreases), which, owing to voltage division between $R_a$ and $R_t$ (see Fig. 3-9), causes a reduction of the component of noise, measured from grid to cathode, due to the antenna. At the same time, the noise developed from grid to cathode by $R_t$ increases at a faster rate than that lost by voltage division. The end result is a higher noise figure as the frequency of operation is increased. The same kind of reasoning will show that transistors and FET's are similar to vacuum tubes in this regard.

For triode RF amplifiers, operating at frequencies in the FM band, a noise figure of 3 to 5 db can be expected. Pentode RF amplifiers display typical noise figures of 8 to 14 db. Typical values of noise figure for transistors and FET's operating under the same conditions as the vacuum-tube amplifiers discussed above are 3.5 db and 2.5 db, respectively.

## 3-8. Transistor noise

With the advent of high-frequency transistors it was just a matter of time before transistorized FM receivers became commercially available. As with the vacuum-tube receiver, the main limiting factor governing maximum sensitivity of the transistor FM receiver is the degree of internally generated noise in the RF amplifier. Owing to the nature of transistor noise there is some possibility that transistor receivers may have lower noise figures than

vacuum-tube receivers. Noise figures of 3 db and less are common at frequencies of 100 MHz.

Transistor noise can be divided into four types:

1. Thermal agitation due to the resistance of the base lead. The magnitude of this voltage is obtained by means of the following expression:

$$e_{n(\text{rms})} = \sqrt{4KTr_bB}, \tag{3-32}$$

where $K$ is Boltzmann's constant, $T$ is the temperature in degrees Kelvin, $r_b$ is the base resistance, and $B$ is the effective bandwidth.

2(a). Shot noise due to input and output junctions of the transistor. The noise level generated by random variations in current flowing between emitter and base is

$$e_{n(\text{rms})} = \sqrt{2eI_er_e^2B}, \tag{3-33}$$

where $e$ is the charge on an electron ($1.6 \times 10^{-19}$ Coulombs), $I_e$ is the DC emitter current, $r_e$ is the emitter resistance, and $B$ is the bandwidth.

2(b). Shot noise due to random variations of current between the collector and the base junction. The noise voltage generated can be found by means of the expression

$$e_{n(\text{rms})} = \sqrt{2eI_{ce}r_c^2B}, \tag{3-34}$$

where $I_{ce}$ is the collector leakage current and $r_c$ is the collector resistance (AC).

3. Partition noise, which can be traced to the random division of hole current between base and collector. The noise current so developed can be evaluated by means of the expression

$$i_{n(\text{rms})} = \sqrt{2e\,\alpha I_e(1-\alpha)B}, \tag{3-35}$$

where $\alpha$ is the current-amplification.

None of the above-listed noise sources is frequency-sensitive.

4. Excess noise, often called semiconductor noise, flicker noise, and $1/f$ noise. This type of noise is also seen in vacuum tubes and FET's at very low frequencies, for reasons that are not well understood. Excess noise is called $1/f$ noise because of its characteristic of decreasing with an increase in frequency of operation. The magnitude of this voltage is

$$e_{n(\text{rms})} = \sqrt{\frac{K_cV^aR_c^b}{f^n}}, \tag{3-36}$$

where $K_c$ depends upon the material the transistor is made of, $V$ is the DC supply voltage, $R_c$ is the DC resistance of the collector, $a$ and $b$ are constants whose values are dependant upon the type material being used, and $n$ is equal to 1.1.

Observe Figure 3-13, which shows a plot of noise figure versus frequency for a junction transistor. There are three distinct regions in the curve; these are labeled $A$, $B$, and $C$. Section $A$ is attributed to the $1/f$ noise of the transis-

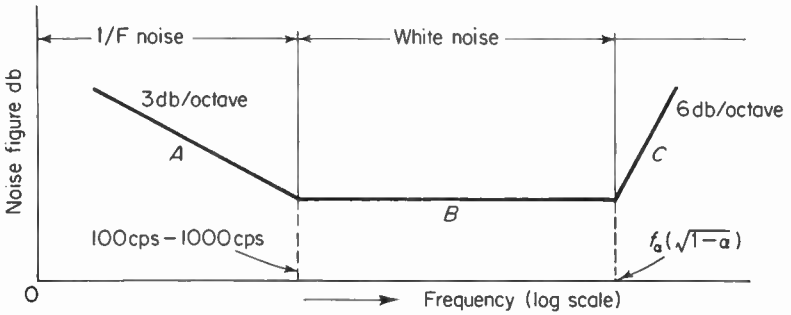



**Fig. 3-13.** Variation of noise figure versus frequency for a typical transistor.

tor, which is insensitive to temperature changes. Section *B* can be said to be due to the shot, thermal, and partition noise generated in the transistor. Of course, as can be seen from the curve, this noise is independent of frequency. Temperature does influence the magnitude of this component of noise. The poorer noise figure of section *C* is due not to an increase in noise from the sources mentioned but rather to a decrease in power gain of the transistor.

The simplified equivalent circuit for a common emitter, including the transistor noise generators, is shown in Fig. 3-14. This equivalent circuit

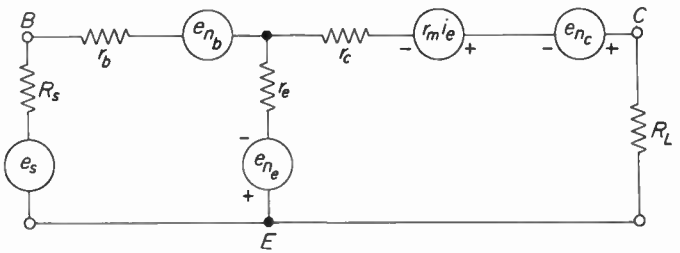



**Fig. 3-14.** Equivalent circuit of a transistor amplifier including the equivalent noise sources.

does not take into account the various shunt and distributed capacitances, which at high frequencies become important. Notice that the circuit contains two noise generators—one in the emitter arm of the equivalent circuit and the other in the collector arm. Typical values for the noise voltage generated in the collector circuit are about 5 $\mu$v rms, and for the emitter circuit about 0.05 $\mu$v rms. These values are obtained at a frequency of 1 kHz at a bandwidth of 1 Hz.

The noise factor of both the common-base and common-emitter amplifiers are identical. Using the equivalent circuit of Fig. 3-14 we can determine that the noise factor over the region of *B* and *C* of Fig. 3-13 is

$$F = 1 + \frac{r_b + \frac{r_e}{2}}{R_s} + \frac{(1-\alpha)\left[1 + \frac{1}{1-\alpha}\left(\frac{F}{F_\alpha}\right)^2\right](r_b + r_e + R_s)^2}{2\alpha r_e R_s} \qquad (3\text{–}37)$$

where $r_b$ is the base spreading resistance, $r_e$ is the emitter resistance, $R_s$ is the source resistance, $\alpha$ is the short-circuit current gain at low frequencies, $F_\alpha$ is the alpha cutoff frequency, and $F$ is the frequency of operation.

Equation (3–37) is based upon the assumptions that the collector resistance is very large and, therefore, can be ignored, and that the emitter current is very much larger than the leakage current $I_{co}$.

Minimum noise figure can be obtained by decreasing the base spreading resistance $r_b$, increasing $\alpha$, and selecting a transistor that has a high alpha cutoff frequency ($F_\alpha$). The noise factor may also be minimized by selecting the proper value of source resistance ($R_s$). The optimum value of $R_s$ is

$$R_{s(\text{opt})} = \sqrt{r_b^2 + \frac{2r_b r_e}{1 - \alpha}}. \tag{3–38}$$

Transistor noise also depends upon the operating voltages and currents of the transistor. In Fig. 3-15 is shown the variation in noise figure for a common-emitter circuit configuration. Note that the lowest noise figure can be attained by using low values of collector current and the proper value of source resistance. Notice, too, that the noise figure is independent of the value of $R_L$, the collector-to-emitter load resistance.

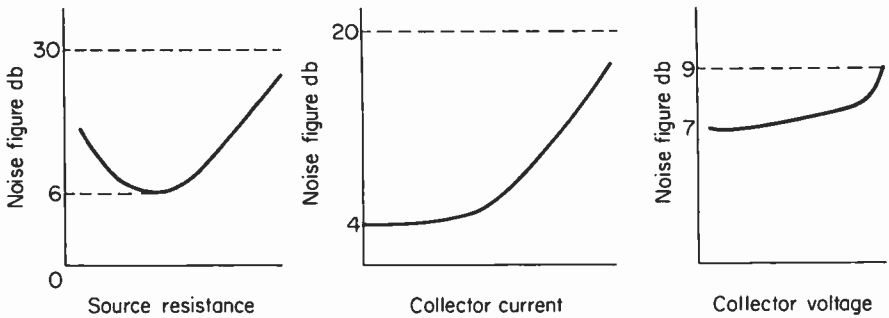



**Fig. 3-15.** Variation in noise figure with changes in operating conditions of a typical transistor.

## 3-9. Summary

**1.** Noise refers to any disturbance that produces undesired masking of signal intelligence.

**2.** Noise is introduced into a receiver by means of (a) thermal agitation, (b) tube noises, (c) sources external to the receiver.

**3.** Noise voltages due to thermal agitation are generated by the random motion of electrons within a conductor. The degree of motion depends upon the temperature of the conductor.

**4.** The magnitude of the noise voltage generated by thermal agitation

is determined by (a) the resistance of the conductor, (b) the temperature of the conductor, (c) the frequency bandwidth of the system.

**5.** Thermal noise (also called random, fluctuation, or white noise) is found to consist of an infinite number of sinewave components, all having the same amplitude.

**6.** Carbon granular resistors generate more noise than do wirewound resistors.

**7.** The total noise generated by a conductor is always measured as an rms value because of its infinite frequency spectrum.

**8.** Thermal-noise sources may be represented by means of Thevenin or Norton equivalent generators and their associated equivalent resistances.

**9.** Pure coils and capacitors are incapable of generating thermal noise.

**10.** Shot noise is a form of random noise that is generated in vacuum tubes by means of the random emission, at the cathode, and the random arrival of electrons at the plate of the tube.

**11.** A space-charge limited tube develops less noise than does a temperature-limited tube.

**12.** For ease of calculation shot noise is referred to the grid by means of an equivalent noise resistance.

**13.** Owing to partition noise, pentodes are six times noisier than triodes.

**14.** At high frequencies transit time of electrons introduces an important component of noise called grid induced noise.

**15.** Grid induced noise is frequency-dependent, in contrast to thermal noise, which is independent of frequency.

**16.** Shot and partition noise are considered to act in series with the grid of the tube, whereas grid induced noise, circuit noise, and noise due to the radiation resistance of the antenna act in parallel to one another and in series with the grid.

**17.** The effective noise voltage from grid to cathode of a tube is determined by the voltage division between all of the noise sources and their equivalent noise resistances.

**18.** Frequency changers (mixers) are from two to four times noisier than the same tube used as a class A voltage amplifier.

**19.** The noise figure of a receiver is a means of judging the weak-signal sensitivity of a receiver.

**20.** Noise figure is sensitive to the frequency of operation.

**21.** The best noise figure that a receiver can attain is zero db. Typical receiver noise figures range from 3 to 14 db, depending upon the design of the RF amplifier.

**22.** The noise figure of a transistor is frequency-sensitive. At low frequencies $1/f$ noise is dominant; at frequencies from approximately 1 kHz to $f_\alpha\sqrt{1-\alpha}$ the noise is random and is constant in level; above $f_\alpha\sqrt{1-\alpha}$

the power gain of the transistor drops, resulting in a poorer noise figure.

**23.** Transistor noise is dependent upon operating voltages and currents.

**24.** Noise reduction, in a triode, can be achieved by operating the tube with lower than normal cathode temperatures, high $g_m$, and high levels of grid bias.

## REFERENCES


1. Bennett, W. R.: "Characteristics and Origins of Noise," *Electronics*, March through July, 1956.
2. Johnson, J. B.: "Thermal Agitation of Electricity," *Phys. Rev.*, **32** (1928).
3. Nyquist, H.: "Thermal Agitation of Electric Charge in Conduction," *Phys. Rev.*, **32** (1928).
4. Thompson, B. J., North, D. O., and Harris, W. A.: "Fluctuation in Space Charge-Limited Currents at Moderately High Frequencies," *RCA Review*, January 1940; April 1940; July 1940; October 1940; January 1941; April 1941.
5. Friis, H. T.: "Noise Figures of Radio Receivers," *Proc. IRE*, **32** (July 1944).
6. van der Ziel, A.: *Noise*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1954.
7. Adler, R.: "Low Noise Devices," *IRE Student Quarterly*, May 1961.
8. Schwartz, Mischa: *Information Transmission, Modulation, and Noise*. McGraw-Hill Book Company, New York, 1959.
9. Metelmann, C.: "Noise Parameters in VHF-UHF Circuit Design," *Electronic Industries*, July 1959.
10. van der Ziel, A.: "Noise in Junction Transistors," *Proc. IRE*, **46** (June 1958).
11. Martens, N. H.: "Notes On Transistor Noise—What It Is and How It Is Measured," *Solid State Design*, May 1962.
12. Terman, F. E.: *Electronic and Radio Engineering*, 4th ed. McGraw-Hill Book Company, New York, 1955.
13. Sturley, K. R.: *Radio Receiver Design*, Part 1. John Wiley & Sons, Inc., New York, 1953.

# 4

# VHF EFFECTS

## 4-1. VHF amplifiers and input circuit loading

The operation of amplifiers used at very high frequencies (VHF) is determined, to a large extent, by the load presented by the amplifier to the input generator. The discussion below is centered around vacuum-tube and field-effect transistors (FET). Since these devices are usually considered to have high input-impedance, any loading of the input will tend to reduce the input impedance and can have important effects upon circuit operations. The transistor is also affected by input circuit loading, but since it is basically a low input-impedance device, most of the loading effects discussed below are "swamped out" by its low input impedance. In later chapters methods for determining the input admittance of high-frequency transistor amplifiers will be discussed.

At low frequencies (1 kHz) the input impedance of a vacuum tube or field-effect transistor is considered to be almost infinite. That is, a signal-voltage generator will draw zero AC grid or gate current when connected to the input of the device. However, as the frequency of operation is increased, the input generator is made to deliver current. The phase of this current with respect to the applied voltage is such that the input impedance will appear, to the input generator, as a parallel combination of a resistive component and a reactive component. The resistive component may be positive (dissipates electrical energy) or negative (acts as a generator), depending upon circuit conditions in the plate (drain) and cathode (source) circuits of the amplifier. Since the input generator usually contains significant internal impedance, it is not considered a constant-voltage generator. Assuming that

the resistive input loading component is positive, the AC current produces a voltage drop and a phase shift across the internal impedance of the input generator which subtracts from the source voltage and thus reduces the input voltage. If the load current becomes high enough (at very high and ultrahigh frequencies), the input voltage to the active element will become so small that the effective voltage amplification will be reduced to approximately zero. If the input generator contains a tuned circuit, the AC input current will dissipate energy that would have been stored in the electrostatic and magnetic fields of the capacitor and the inductance of the tuned circuit. This dissipation of energy means that the $Q$ of the circuit has been reduced. Thus, input loading can cause the input circuit to have a wider bandpass than the resonant circuit, taken by itself, would have had if there were no loading. Likewise the resistive component of loading may contribute additional noise to the already existing shot and partition noise present in all vacuum tubes and FET's. The capacitive component of input impedance must be taken into account when the resonant frequency of a tuned circuit, connected between grid (gate) and cathode (source), is considered.

If, on the other hand, owing to output circuit conditions the input impedance may contain a negative resistive component, the effect on the circuit will be somewhat different. In amplifiers that are tuned, a negative resistance, of the proper magnitude, may cause the amplifier to oscillate. Other values of negative resistance may cause regeneration with a subsequent increase in tuned-circuit $Q$ and narrow bandwidth.

The resistive component of the total input impedance is contributed to by at least four components:

1. Leakage resistance and dielectric hysteresis losses between the input circuit and the other elements of the device increase with frequency of operation. At the frequencies employed for FM broadcasting these losses are not of great importance.

2. Gas current, grid emission, and contact potential in a vacuum tube all contribute loading on the input generator because the input generator modulates these various currents and thus power is drawn from the input generator. The magnitude of these loading components is small ($R$ shunt is large) and thus is seldom included in calculations for total input resistance.

3. Feedback between the various elements of both the vacuum tube and the FET is responsible for two major components of input circuit resistive loading. In the case of the vacuum tube one of these components is the result of feedback via the interelectrode capacitance between the plate and the control grid. This feedback effect is commonly referred to as the Miller effect. The other resistive component is due to feedback between the cathode and the control grid via the interelectrode capacitance between them.

The FET is similar to the vacuum tube in that feedback between the drain

and the gate via the device's interelectrode capacitance results in a Miller effect phenomena. Likewise at ultra high frequencies (UHF), feedback between the source and the gate may also introduce a resistive component of input circuit loading. Owing to the relatively small effect which this type of loading has at very high frequencies (VHF) it will not be discussed in detail in this text.

4. The transit time of electrons from cathode to plate introduces a resistive loading effect from cathode to grid that depends upon the frequency of operation. Its value decreases very rapidly, being inversely proportional to the square of the frequency of operation. Transit-time loading will be examined in more detail in a later section of this chapter.

The capacitive (it is very rarely inductive) component of input loading is contributed to by three sources:

1. The physical size and spacing between the elements of the tube or FET will, by definition of a capacitor, provide capacitive loading on the input generator. This type of capacitance is the same whether the tube is in or out of the energized circuit.

2. Since the space charge of a vacuum tube is a conductive extension of the cathode itself, it is self-evident that the distance between grid and the effective cathode sleeve (space charge) is made smaller. Thus, the capacitance between grid and cathode is effectively increased above the "cold" capacitance $C_{gk}$ listed in the tube manual. The "hot" capacitance is 2 or 3 pF greater than the "cold" capacitance of the tube.

Space-charge capacity is the capacitive component associated with transit-time resistance and is considered to be in parallel with it.

3. As will be shown in a later section, feedback from the output to the input via interelectrode capacitance effectively introduces an important capacitive component across the input circuit; this is called the Miller effect. Another capacitive component of input loading that is introduced by feedback is provided by the cathode lead inductance and the grid-to-cathode capacitance $C_{gk}$. This effect becomes important at high frequencies when pentodes are employed as RF amplifiers.

In summary, therefore, it can be seen that there are at least five important shunt components that determine the input impedance of an amplifier operating at very high frequencies. Figure 4-1 illustrates how they are considered to appear in a typical grounded-source and grounded-cathode amplifier circuit. Since all of the components are in parallel, the total input $Z$, as seen by the generator, is often calculated by means of the addition of reciprocal values of $R$ and $X$. The reciprocal of $R$ is called conductance ($G = 1/R$), that of reactance is susceptance ($B = 1/X$), and the resultant impedance is called admittance ($Y = 1/Z$).

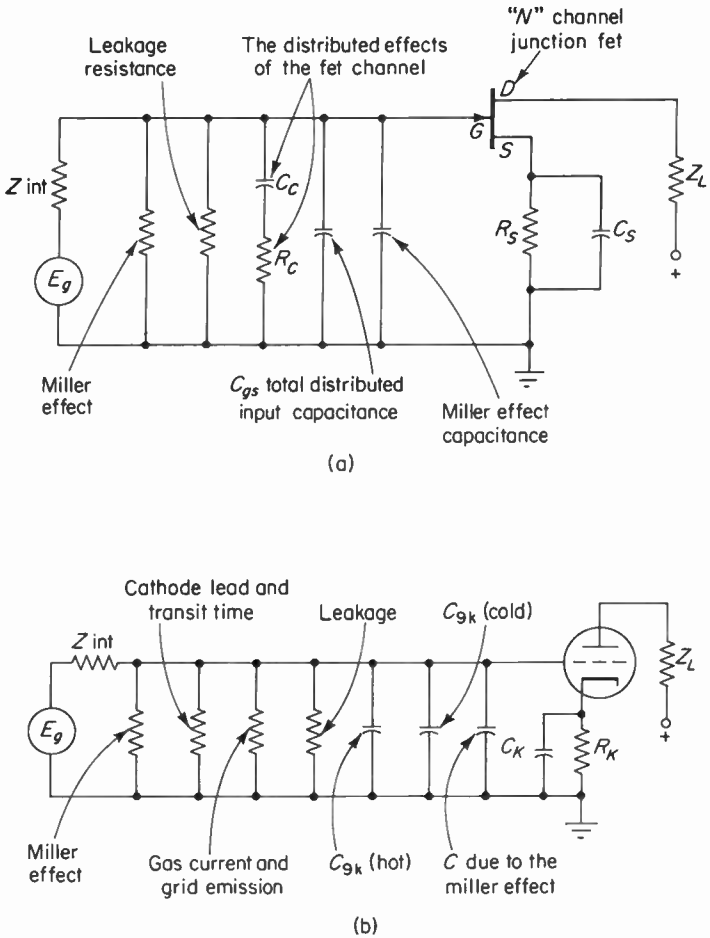



**Fig. 4-1.** (a) Indicates the components which represent the input impedance of a field effect transistor at low to VHF frequencies. (b) Illustrates the components which constitute the input impedance of a vacuum tube.

## 4-2. The Miller effect

The Miller effect is simply the effect of the output-circuit impedance upon the input-circuit impedance of a given active element. This is accomplished via coupling through the interelectrode capacitance of the device. Vacuum tubes, transistors, and FET's also display Miller effect, although the effect is less important to circuit operation in a transistor, owing to its small input impedance.

In order to understand how the output impedance can affect the input impedance we should review a number of basic ideas.

Remember that both resistance and reactance can be expressed as voltage-to-current ratios with certain phase relationships between $E$ and $I$. Thus, the *physical shape* of circuit components has nothing to do with whether a circuit is considered resistive, capacitive, or inductive. The sole determining factor is the phase between the applied voltage and the total current.

Figure 4-2 illustrates that in a resistor the voltage drop across it and the current through it are at a phase angle of zero degrees. It can be said, therefore, that anything (a tube, a resonant circuit, the gird circuit of a tube or a transistor) can be considered as a resistive device provided that $E$ and $I$ have a phase relationship of zero degrees between them.

If the voltage and current in a cirucit are related to each other by an angle of 90 degrees, and if the current leads the voltage, then the circuit *acts as though* a capacitor were connected across the generator. The point is that the device causing the leading current flow need not be a physical capacitor. Of course, if the current were to lag the applied voltage, then the circuit as "seen" by the generator would appear as an inductive reactance. See Fig. 4-2(b) and (c).

What is true for resistance and reactance is also true for impedance. Thus, in any circuit if $E$ and $I$ differ by an angle of more than zero degrees but less than 90 degrees [as in Fig. 4-2(d) and (e)], the circuit is an impedance whether or not any physical lumped components are connected across the input generator.

In summary, if in a given circuit an applied AC voltage causes an AC current to flow at a certain phase angle, a knowledge of the magnitude and phase relationships between the voltage and current will permit one to determine the type of load presented to the generator.

As is well known, phasors may be used to represent phase and magnitude relationships between sinusoidal voltages and currents [Fig. 4-2 (a), (b), (c), (d), and (e)]. They are convenient because they permit visual observation of mathematical relationships and thus provide a quick grasp of the phase and magnitude relationships between sinusoidal quantities.

Any phasor can be shown to be made up of two components at right angles to each other. Therefore, given the magnitude and phase of a voltage and current phasor, it is possible to determine the nature of the circuit both as to the configuration (series or parallel) and component size. Two examples will show how this idea is put to use. Note the two phasors $E$ and $I$ in Fig. 4-3(a). If the voltage is considered as the reference phasor, then the current may be resolved into two right-angle components. See Fig. 4-3(b). Electrically, this phasor diagram can be said to represent a circuit in which a capacitance is in shunt with a positive resistance. A positive resistance is any device that

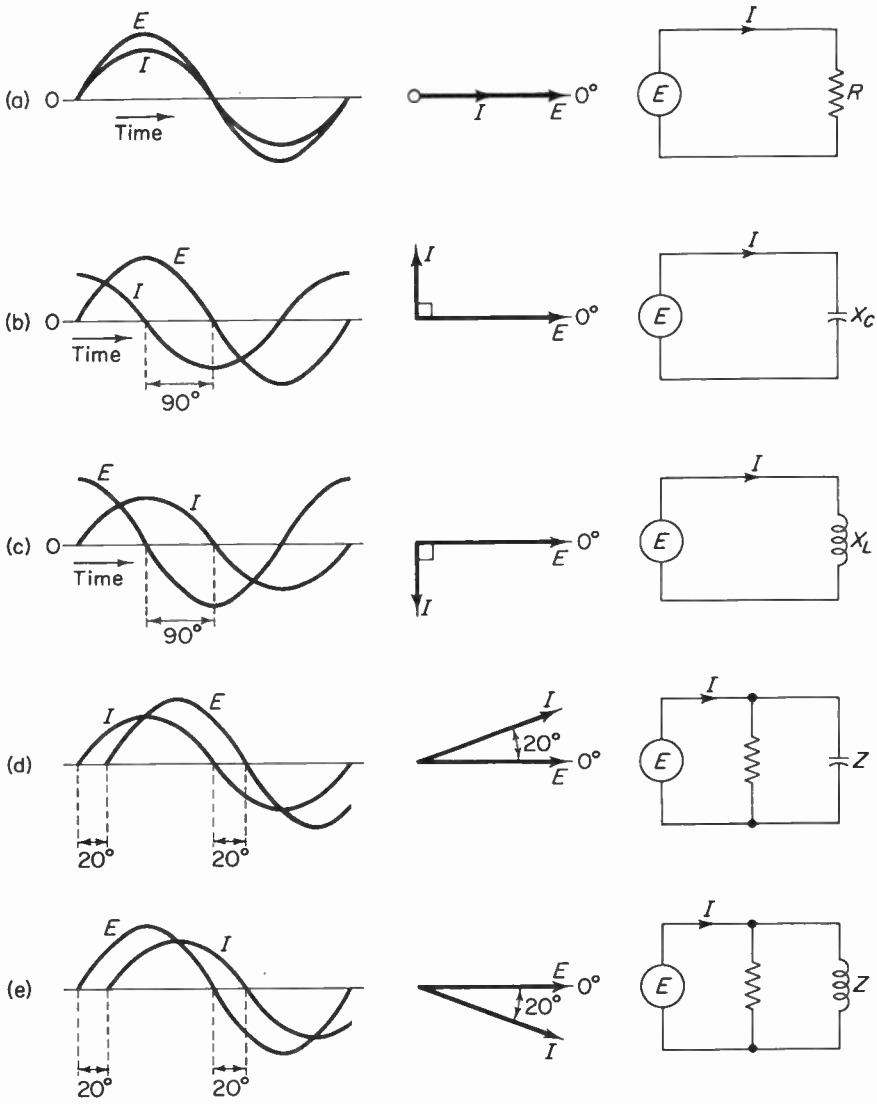



**Fig. 4-2.** The above figures illustrate the phase relationships which exist between the applied voltage and the total current for various types of shunt loads.

dissipates electrical energy. A second phasor diagram is shown in Fig. 4-4(a). In this case the current leads the applied voltage by more than 90 degrees. Again, if the voltage phasor is taken as reference, then the current phasor may be resolved into a capacitive reactive current and a negative resistive component of current. The parallel circuit this represents would therefore

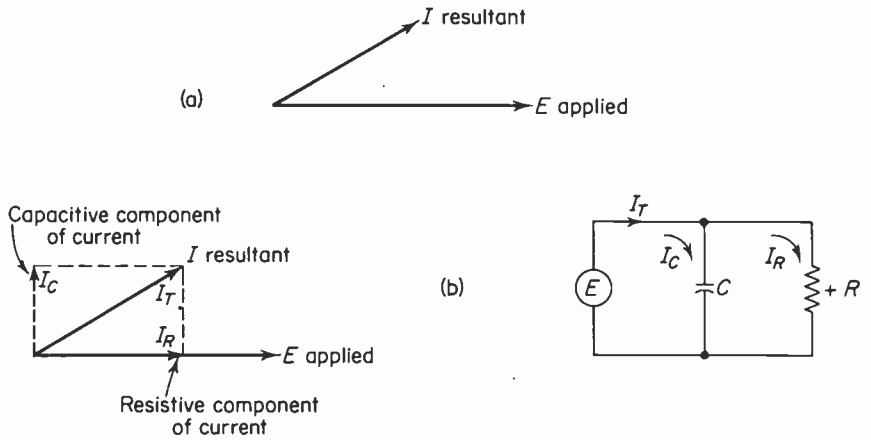



**Fig. 4-3.** Given a phasor diagram in (a) it is resolved into right angle components and the circuit which it represents in (b).

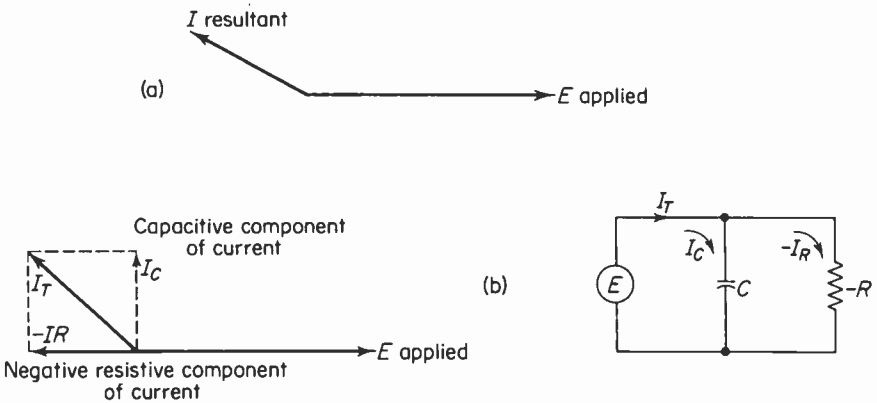
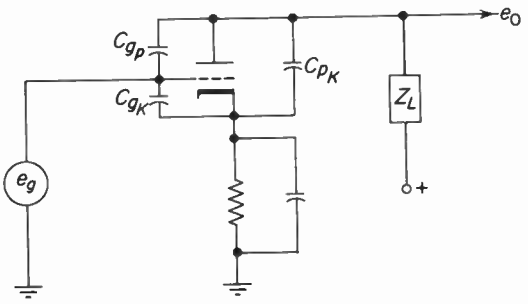


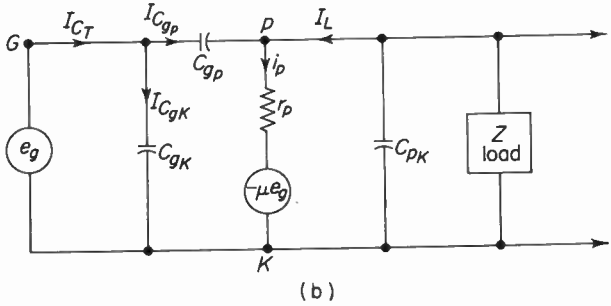

**Fig. 4-4.** In (a) the current leads the applied voltage by more than 90 degrees. In (b) the phasor is resolved into components of current which indicate that the circuit consists of a capacitance and a negative resistance in parallel.

consist of a capacitor and a negative resistance. See Fig. 4-4(b). A negative resistance can be thought of as a generator.

Of course, in each of the examples above, if the current phasor is taken as the reference phasor the equivalent circuit consists of a resistance and a reactance in series. In all of the following discussions the shunt equivalent circuits will be employed.

If the equivalent circuit of a vacuum-tube amplifier, employing a resistive load, is analyzed by means of a phasor diagram, as in Fig. 4-5, it will be seen that there are two components of grid current ($I_{c_R}$ and $I_{c_p}$), both of which lead the grid voltage ($e_g$) by 90 degrees. See Fig. 4-5(c). Thus, as far



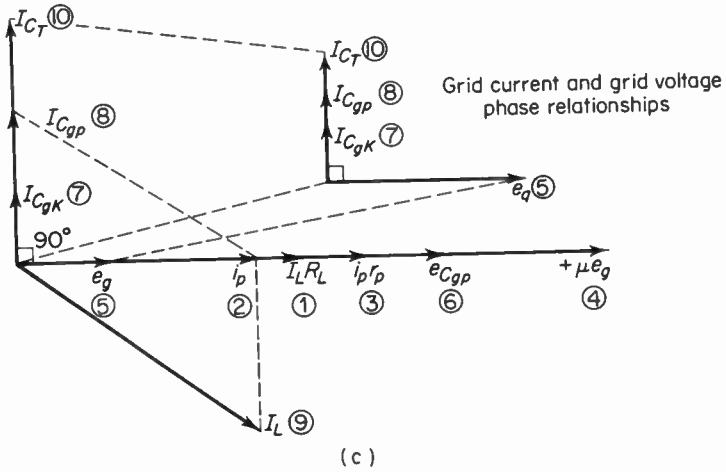




**Fig. 4-5.** (a) Triode amplifier showing various interelectrode capacitances. (b) Equivalent circuit of (a). (c) Phasor diagram of voltage and current relationships in (b). It is to be noted that the plate current (2) and the voltage across the load (1) are in phase whereas the load current (9) lags the load voltage. Thus the entire circuit from plate to cathode is resistive and load impedance is inductive.

as the input generator is concerned the total current drawn from the input generator is the phasor sum of these two currents. Since the currents are in phase, they may be added as a simple sum. The grid input generator is now supplying additional current ($I_{C_{gp}}$) above that demanded by the grid-to-cathode interelectrode capacitance ($C_{gk}$); therefore, the circuit behaves as though the capacitive reactance across the grid generator had been reduced. In other words, the input capacity of the tube has been effectively increased by the factor $C_{gp}(1 + A_v \cos \theta)$. The total input capacity of the tube† is then

$$C_t = C_{gk} + C_{gp}(1 + A_v \cos \theta), \tag{4–1}$$

$C_{gk}$ = grid-to-cathode interelectrode capacity,

$C_{gp}$ = grid-to-plate interelectrode capacity,

$A_v$ = voltage amplification of the stage,

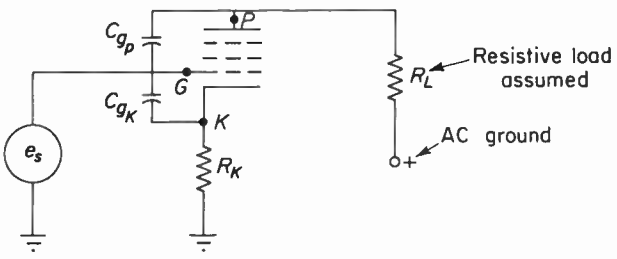
$\theta$ = angle between the grid voltage and the plate voltage, given by the equation below:

$$\theta = \tan^{-1} \frac{X_L}{R_L} - \tan^{-1} \frac{X_L}{r_p + R_L}. \tag{4–2}$$

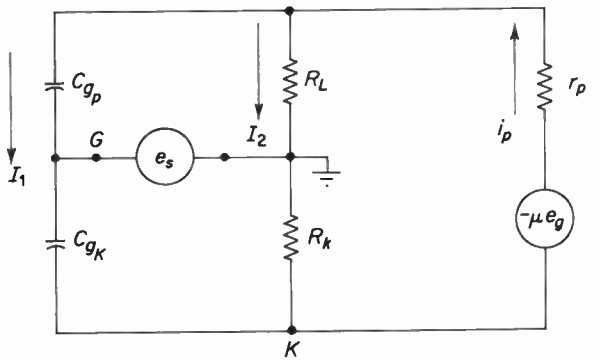
It can be seen from Eq. (4–1) that the magnitude of the input capacity depends upon the value of $C_{gp}$ and upon the stage gain. If a pentode is utilized as the amplifier, $C_{gp}$ is small compared to that of a triode (stray capacity from plate to grid due to wiring is also assumed to be negligible), and thus the input capacity due to feedback through $C_{gp}$ is not as great but is still a factor to be reckoned with.

Of course, the increase in input capacity for a triode amplifier must be taken into account whenever a resonant circuit is connected into the grid circuit. This is because the Miller-effect capacitance adds to the tuned-circuit capacity and will alter its resonant frequency. Furthermore, if AVC is employed in conjunction with a triode or pentode amplifier, any change in bias will cause a change in gain, which will then cause a change in input capacity, which in turn will detune the input resonant circuit. To be more exact, if the bias decreases, the voltage amplification of the amplifier (IF or RF) increases, since $g_m$ increases. Thus the input capacity due to the Miller effect increases. This then decreases the resonant frequency of the input tuned circuit. This change of input capacity, as well as the input capacity itself, can be neutralized by the simple expedient of leaving a cathode resistor of proper magnitude unbypassed. The value of this resistor is rather critical and is arrived at by balancing a bridge as shown in Fig. 4-6. When the bridge is balanced, the feedback current from plate to ground is zero.

†The method of construction of the phasor diagram and the derivation of Eqs. (4-1) and (4-2) will be found in Appendix 4-1.




(a)




$$\frac{I_2 R_L}{I_2 R_K} = \frac{I_1 X_{C_{g_p}}}{I_1 X_{C_{g_K}}} \quad \therefore \quad \frac{R_L}{R_K} = \frac{X_{C_{g_p}}}{X_{C_{g_K}}} \quad \text{thus} \quad \frac{\frac{1}{2\pi F C_{g_p}}}{\frac{1}{2\pi F C_{g_K}}} = \frac{R_L}{R_K}$$

$$\frac{2\pi f C_{g_K}}{2\pi f C_{g_p}} = \frac{R_L}{R_K} \quad \therefore \quad R_K \cong \frac{R_L C_{g_p}}{C_{g_K}}$$

(b)

**Fig. 4-6.** (a) Simplified diagram of a pentode RF which utilizes an unbypassed cathode resistor for neutralization of the Miller effect. (b) Equivalent circuit of Fig. 4-6(a) arranged to show its bridge structure. The approximate value of $R_k$ necessary to balance the bridge is found by the expression derived above.

Thus loading of the input generator is avoided. This method of neutralization will to some degree also compensate for changes in input capacity due to $C_{gk}$ (hot). Since $R_k$ introduces a small amount negative feedback, the amplifier is also less likely to oscillate.

If the plate load impedance is changed from resistive to capacitive, as would be the case for a resonant plate tank when the applied frequency is changed from resonance to a frequency above resonance, the phase relationships throughout the vacuum-tube circuit will be altered. The circuit and its constant-voltage equivalent as well as the phasor diagram for the circuit are shown in Fig. 4-7(a), (b), and (c). It can be seen from Fig. 4-7 (b) and (c) that the total grid current leads the grid voltage by less than 90 degrees. Since the grid voltage ($e_g$) is reference, it is apparent that the input generator "sees" a capacitance, generally larger than the interelectrode capacitance $C_{gk}$, which is in shunt with a positive resistance. The magnitude of the capacitive component is

$$C_t = C_{gk} + C_{gp}(1 + A_v \cos\theta)$$

which is the same as in the resistive plate-load case except that $\theta$ is now a negative angle less than 90 degrees. The magnitude of the shunt resistive component is

$$R = -\left(\frac{X_{C_{gp}}}{A_v \sin\theta}\right). \tag{4-3}$$

The resistive component is positive, when the plate load is capacitive, because $\theta$ (the angle between the grid voltage and the plate voltage) is negative—that is, the output plate voltage lags the grid reference voltage. In other words, this form of negative feedback results in the plate circuit taking power from the grid circuit.

Notice that the sign of $\theta$ depends upon the particular quadrant that $\theta$ falls into, and the trigonometric function of $\theta$ desired. For example, a capacitive plate load puts $\theta$ into the fourth quadrant. Therefore $\sin\theta$ is negative, resulting in an effective positive resistive component, and $\cos\theta$ is positive, resulting in an effective capacitive shunt component of input loading. As a matter of fact, for all conditions of plate load discussed in this text the cosine of $\theta$ is always positive. The sine of $\theta$, on the other hand, may be positive or negative, depending upon the type of plate impedance employed.

The practical effect of shunting the input circuit of an RF amplifier with a resistance and a capacitance is to broaden the bandpass of the tuned circuit and to change its resonant frequency. Of course, the resistive component of the Miller effect will also tend to influence the noise figure of the amplifier as well as the voltage amplification of the stage.

If we consider the condition where the plate-to-cathode load is inductive, as shown in Fig. 4-8, we find that the total grid current ($I_{C_{gp}}$ plus $I_{C_{gk}}$) leads the grid voltage by more than 90 degrees [Fig. 4-8(b)]. Thus, the input generator appears to be looking into a capacitance shunted by a negative resistance. The magnitude of these components may be obtained by means of Eqs. (4–1) and (4–3). Under these conditions both the sine and the cosine of $\theta$ will be positive.

The resistive component, being negative, will tend to cancel any positive

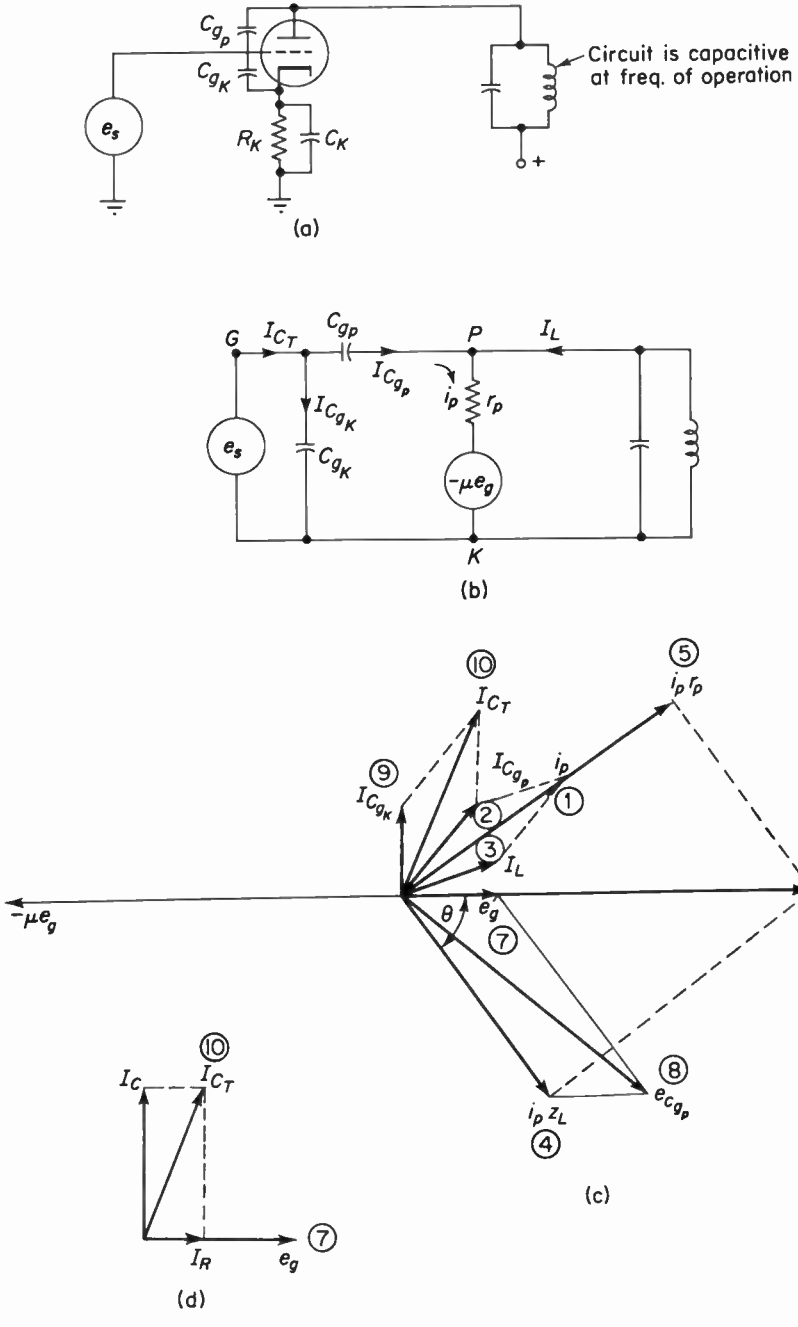



Fig. 4-7. (a) Represents the circuit and its interelectrode capacitance. (b) Is the constant voltage equivalent of (a). (c) and (d) are the phasor representations of the voltages and currents in the equivalent circuit.

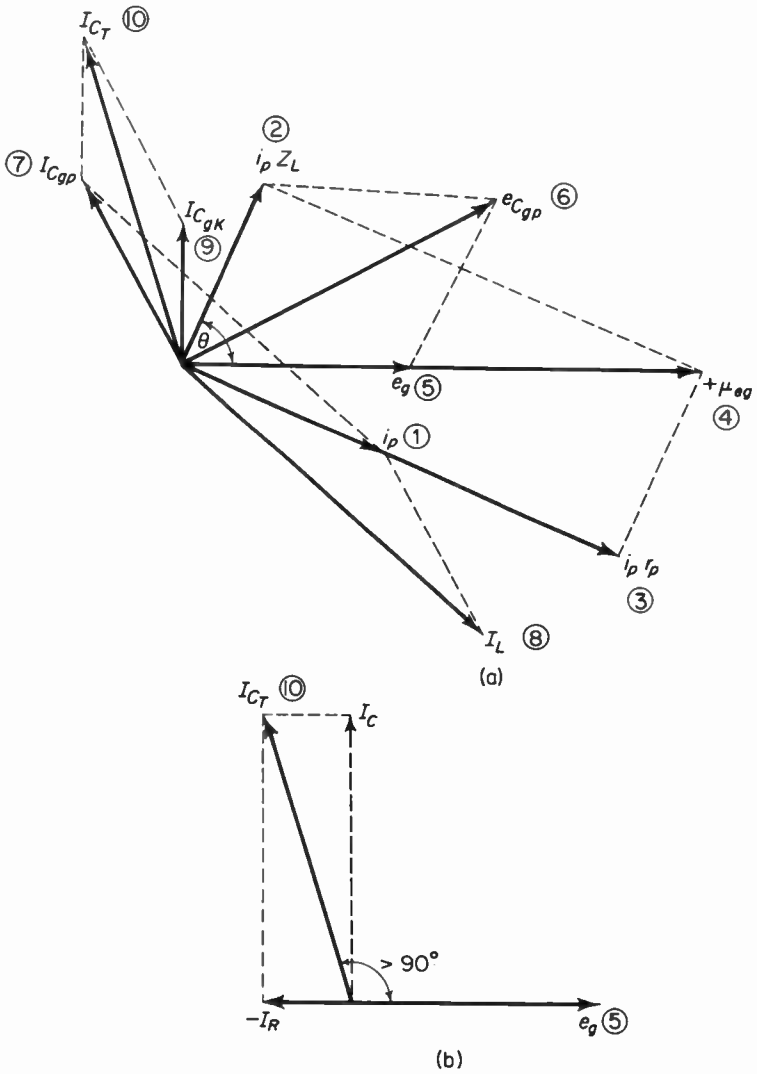
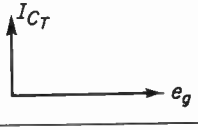
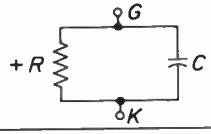
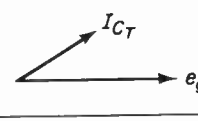
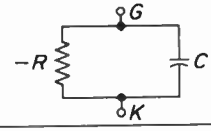



**Fig. 4-8.** (a) Illustrates a phasor diagram of an amplifier with an inductive circuit between plate and cathode. (b) Illustrates the phase relationships which exist between the grid voltage and grid current of the amplifier. See Fig. 4-7(a) and (b) for circuit and equivalent circuit upon which the phasors are based.

resistance in the grid circuit of the amplifier. Thus, the $Q$ of any tuned circuit connected between grid and cathode of the amplifier will be increased. If this is carried to an extreme, all losses in the circuit will be compensated for and the circuit will oscillate.

Table 4-1 summarizes the changes of impedance seen by the input generator that occur when the plate-load impedance is changed.

**Table 4-1.**

| Plate to cathode circuit impedance | Grid input impedance | Phasor relationship between grid voltage and grid current |
|---|---|---|
|   |   |   |
|   |   |   |
|   |   |   |

It can be seen from the expressions for $C_t$ and for $R_{in}$, that if $C_{gp}$ is made small ($X_{C_{gp}}$ is therefore large), the injected Miller-effect capacitance is reduced and the input resistance is increased. This reduction of the Miller effect can be best accomplished by the use of a pentode, with its inherent low value of $C_{gp}$, rather than a triode with its larger value of $C_{gp}$. If the screen grid of a pentode is not adequately bypassed, feedback may take place through the interelectrode capacity from screen to control grid. Miller-effect capacitance may be introduced into the control-grid circuit by this means.

It is possible to find the value of $C_{gp}$ that will bring an amplifier just to the point of instability. This value may be determined by the following expression:

$$C_{gp} = \frac{1}{f g_m Z^2}, \tag{4-4}$$

where

$C_{gp}$ = the grid-to-plate internal as well as the external interelectrode capacitance,

$g_m$ = transconductance of the tube,

$Z$ = input and output impedance of a single-tuned amplifier, assuming these to be equal.

The values of $C_{gp}$ thus obtained can be used as a guide for the selection of tubes used at high frequencies.

The highest voltage amplification possible for a single-tuned RF amplifier before it breaks into oscillation is found from the expression

$$\text{maximum stable gain} = \sqrt{\frac{g_m}{\pi C_{gp} f}}. \tag{4-5}$$

Of course, values of voltage amplification less than the maximum stable gain may cause important changes in bandpass and input loading.

The discussion above, although in terms of vacuum-tube symbols and terminology, is also applicable to the field-effect transistor. The characteristics of the FET are, aside from physical structure, almost identical to those of a high-mu triode. Thus, typical parameters of a field-effect transistor are $C_{dg}$, which is equivalent to $C_{gp}$, and is approximately 1 pF; $r_d$, which is equivalent to $r_p$ in a vacuum tube and is usually 50 to 100 k; and the $g_m$ of an FET, typically 1000 to 3000 $\mu$mho, which is of the same order as the $g_m$ of a tube. Owing to the large interelectrode capacitance $C_{dg}$, it is to be expected that the input impedance of an FET amplifier may be broken into components that also contain the Miller-effect resistance and capacitance. The equations for the Miller-effect capacitance and resistance of an FET are identical in form to those of a vacuum tube. The capacitance component is

$$C_m = C_{sg} + C_{dg}(1 + A_v \cos\theta) \tag{4-6}$$

and the resistive component is

$$R_m = -\left(\frac{X_{C_{dg}}}{A_v \sin\theta}\right). \tag{4-7}$$

## 4-3. Cathode lead inductance

As has been already been pointed out, pentodes are less susceptible to Miller-effect loading than are triodes. It might then appear that pentodes are free of input loading effects. Unfortunately this is not the case because of cathode lead inductance and transit-time loading, both of which become more important as the frequency of operation increases.

All conductors, no matter how short, have the property of inductance. At low frequencies this inductance is not important because of the low value of reactance developed, and generally it is not considered in any calculations. Although the inductance of a short piece of wire is small, it will exhibit appreciable reactance at high frequencies. For example, the cathode sleeve, connecting wires, and socket pins may have a total inductance of 0.002 $\mu$hy to 0.8 $\mu$hy. At a frequency of 100 MHz an inductance of 0.8 $\mu$hy will exhibit a reactance of 500 ohms. A reactance of this magnitude in the cathode circuit will, owing to the AC component of plate current, develop a voltage drop across it, which will combine with the input signal to produce a grid current flowing through $C_{gk}$, $L_k$, and the input generator. See Fig. 4-9(a).

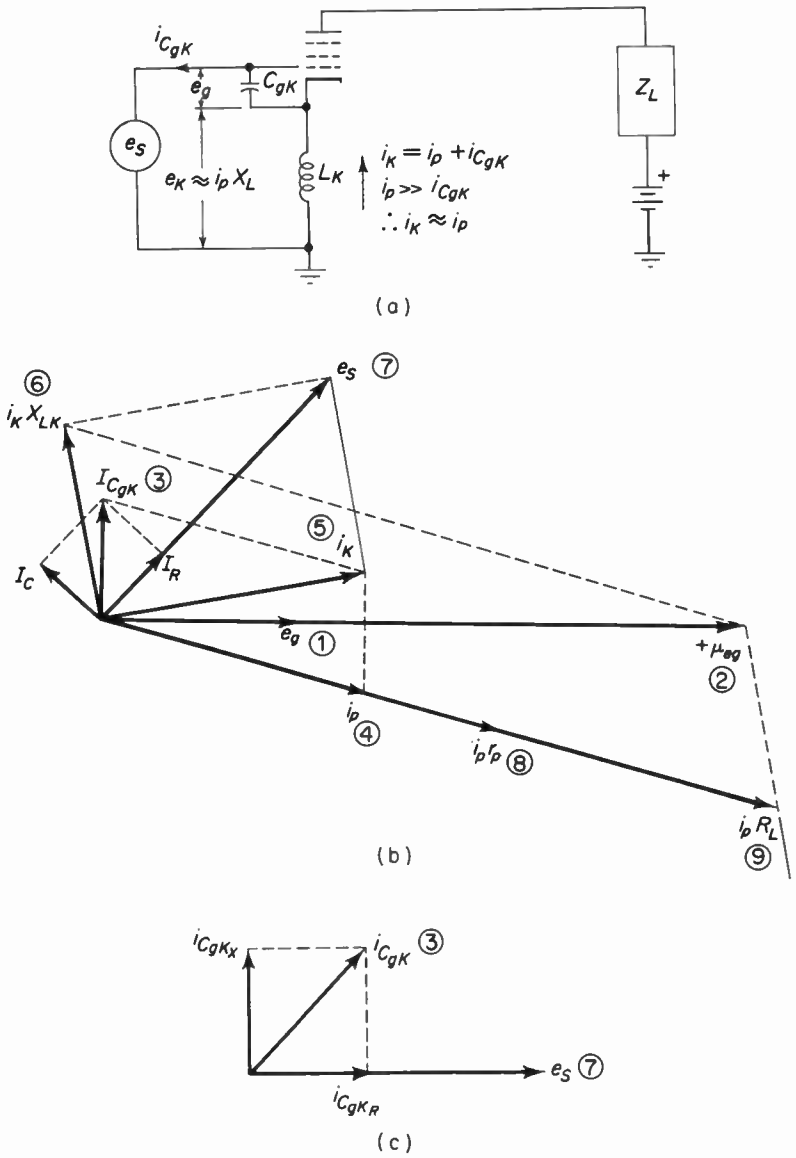



**Fig. 4-9.** (a) The input circuit of a pentode amplifier as seen by the input generator at high frequency. Screen and suppressor grids are not connected for purposes of simplicity. (b) Phasor diagram of a pentode amplifier operating at a frequency where the cathode lead inductance had become important. It is assumed that the plate load is resistive and that $i_p$ is much larger than $I_{Cgk}$. (c) Portion of figure (b) phasor diagram which illustrates that the input generator appears to be shunted by a resistance (the phasor component of current $I_{Cgkr}$) and a capacitance (the phasor component of current $I_{Cgkz}$).

The phase of this current with respect to the input voltage ($e_s$) is such as to appear to the input generator as a resistive and capacitive component shunting it. The nature of the loading can be determined from a phasor diagram of the circuit [Fig. 4-9(b)]. The magnitude of the capacitive component is approximately the same as the grid-to-cathode capacitance $C_{gk}$. The value of the resistive component may be found by means of the expression

$$R = \frac{1}{4\pi^2 f^2 g_m L_k C_{gk}}. \tag{4-8}$$

The method of construction of the phasor diagram and the derivation of Eq. (4–8) will be found in Appendix 4–2.

The equation for $R$ shows that as the frequency of operation increases, $R$ decreases inversely as the square of the frequency. This means that at high frequencies greater loading is placed on the input generator. For example, if the $g_m$ of a tube is 5000 micromhos, $C_{gk}$ is 2 pF, $L_k$ is 0.1 $\mu$hy, and the frequency of operation changes from 50 MHz to 100 MHz, it will be seen that the input loading resistance will vary from 101,000 ohms at 50 MHz, to one-fourth of that or 25,250 ohms at a frequency of 100 MHz.

Neutralization of cathode lead loading can be accomplished by at least five different methods:

1. Multiple cathode leads placed in parallel.
2. A bridge balancing coil placed in the plate circuit.
3. A bridge balancing coil placed in the screen circuit.
4. Division of the cathode inductance into two parts—one for the AC component of plate current and the other for the input grid current.
5. Series resonance between the cathode lead inductance and the stray capacity to ground.

The methods by which the twin cathode leads reduce cathode lead loading are shown in Fig. 4-10(a), (b), and (c). In Fig. 4-10(a) the cathode leads are

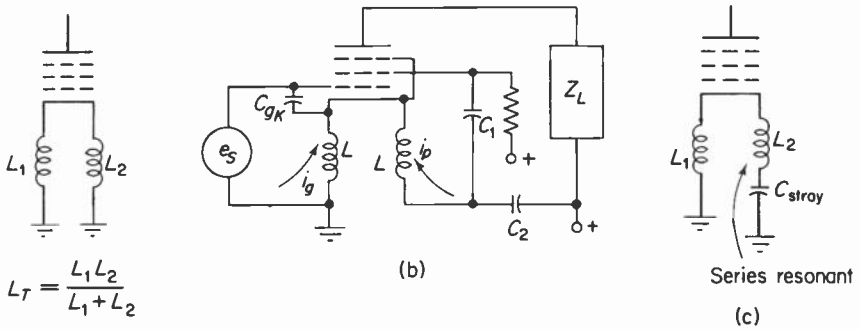



**Fig. 4-10.** Three methods of minimizing the effects of cathode lead inductance.

simply tied together, thus placing the two inductances in parallel. The total inductance is thereby reduced to one-half the inductance of either one.

The circuit in Fig. 4-11(b) uses the fact that the major cause of cathode lead loading is the voltage drop across $L_k$ produced by the AC component of plate current. This voltage drop shifts the phase between the signal voltage ($e_s$) and the grid-to-cathode voltage ($e_g$), and is, therefore, responsible for the phase difference between the signal voltage ($e_s$) and the grid current ($I_{C_{gk}}$). The circuit in Fig. 4-11(b) attempts to separate the plate current from the grid current by supplying separate paths for each. Note that the DC plate current returns to the cathode by means of the cathode lead $L_1$. The AC grid current due to the input generator also flows through $L_1$, whereas the AC component of plate current as well as the AC component of screen current flow through $L_2$. Thus input and output AC components of current are isolated from one another.

In Fig. 4-11(c), $L_2$ is made to resonate with the stray capacity to ground. Since this constitutes a series resonant circuit, the impedance from cathode to ground is very small. Thus the voltage developed from cathode to ground is reduced to a relatively insignificant level. Cathode lead loading is thereby reduced to very low levels.

Figure 4-11 shows a method for neutralization of cathode lead loading using the fact that a Wheatstone bridge exhibits infinite impedance in the

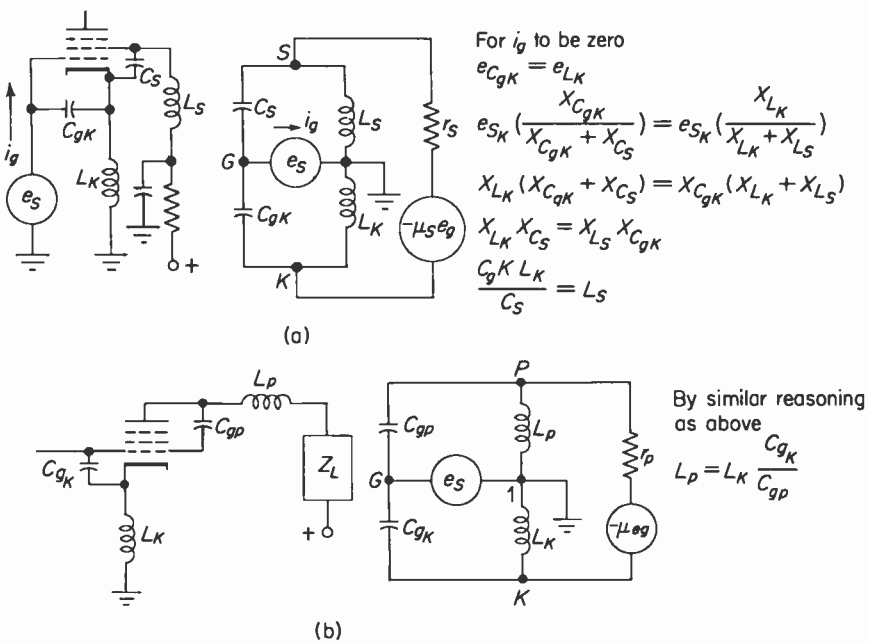



**Fig. 4-11.** Two methods for neutralizing cathode lead loading.

cross arm when the branch arms are in balance. To make a vacuum tube into a bridge, two methods are employed. The first method [Fig. 4-11(a)] is to place a coil in series with the screen of a pentode. The coil ($L_s$) feeds back energy to the grid circuit via the interelectrode capacity $C_{gs}$, which is opposite to that introduced by the cathode lead inductance. It might be said that the inductive screen is introducing a negative resistance into the control grid, since the Miller effect exists from screen to control grid if the screen is not completely bypassed. Of course, the negative resistance, so introduced, should equal and cancel the positive resistance developed by the cathode lead. The expression for determining the value of $L_s$ is developed in Fig. 4-11(a). The second method [Fig. 4-11(b)] is exactly the same as the first except that the neutralizing coil is placed in the plate circuit instead of the screen.

Of value in FM receiver oscillator circuits is the condition where the cathode circuit is made capacitive rather than inductive. The cathode circuit is made capacitive by placing a capacitor of large reactance in series with the lower-reactance cathode lead inductance. Upon evaluating the expressions for the input shunt components† we find that the capacitive component is approximately the same as $C_{gk}$, and the resistive component is negative and can be determined by the following expression:

$$R = -\frac{C_k}{g_m C_{gk}}, \tag{4–9}$$

where

$C_k$ = capacitance from cathode to ground,

$C_{gk}$ = capacitance from grid to cathode,

$g_m$ = transconductance of the tube.

This expression is valid only if $C_{gk}$ is larger than $C_k$.

A typical circuit, its constant-voltage equivalent circuit, and the associated phasor diagram are shown in Fig. 4-12. We can see that the input resistance is negative by noting, from the phasor diagram, that $I_{C_{gk}r}$, the resistive component of input grid current, and $e_s$, the generator voltage supplying this current, are 180 degrees out of phase with respect to one another.

An interesting difference exists between the input resistance due to cathode lead inductance and that due to a capacitive cathode impedance. It can be seen from the respective equations for the shunt component of resistance that the inductive cathode results in a frequency-dependent resistance, whereas the input resistance resulting from a capacitive cathode impedance is not frequency-dependent.The reason is that the voltage division of $e_s$ that takes place across $C_{gk}$ and $C_k$ is constant, since the reactances of both change in the same proportion with a change in frequency This is not so

†See the Appendix 4-3 for details.

in the inductive case, because for a given change in frequency that part of the input signal ($e_s$) across $C_{gk}$ changes inversely with frequency, whereas the voltage across $L_k$ changes directly with frequency. The end result is that the input resistance varies with frequency for an inductive cathode impedance.

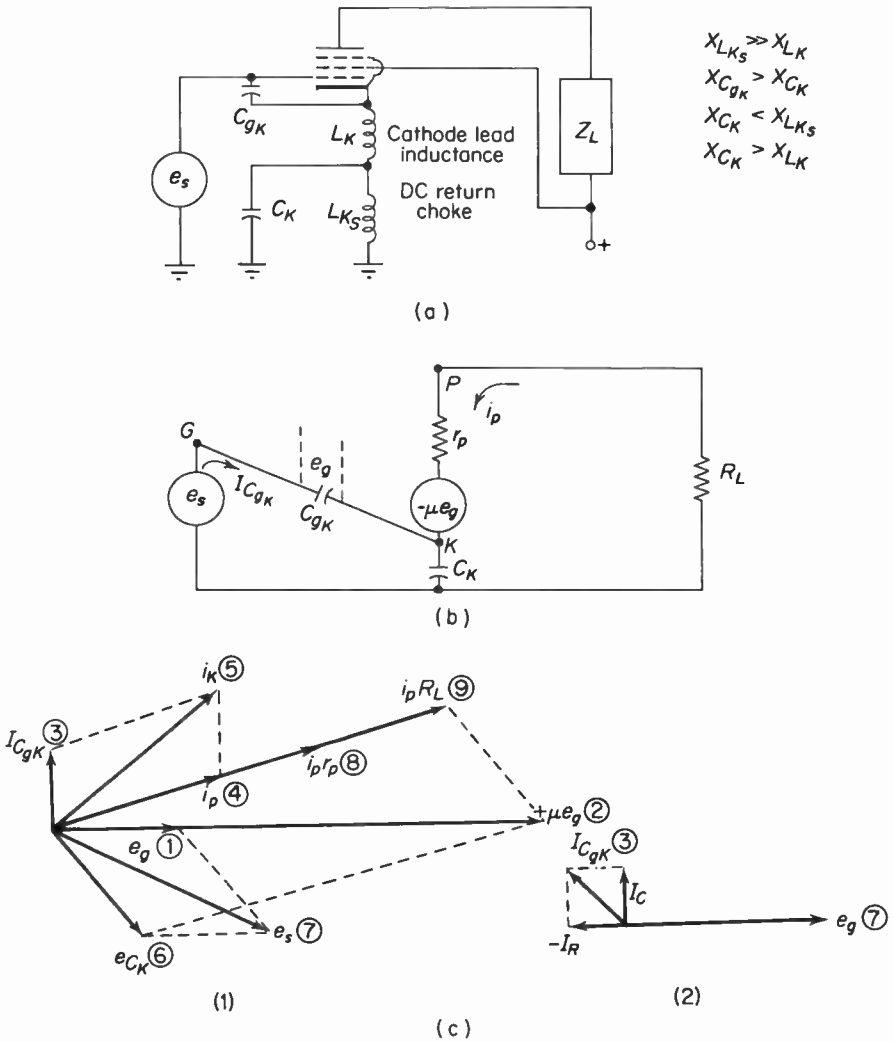
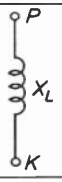
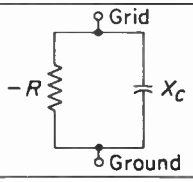
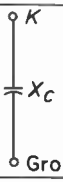
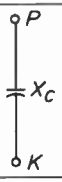
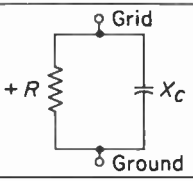
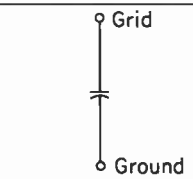



**Fig. 4-12.** (a) Schematic diagram of a pentode operating at very high frequency. The cathode circuit is capacitive. (b) Constant voltages equivalent circuit of a VHF amplifier that is the basis for the phasor diagram of (a). (c) Diagram number one is the complete phasor diagram of (b). Diagram number 2 is a portion of diagram number 1 which illustrates the input impedance seen by the input generator.

Table 4-2 points out an important relationship between the plate and cathode impedance and their effect on the input resistance of the tube. It can be seen that they have opposite effects. For example, if the plate load is capacitive, a positive resistive component is introduced (via the Miller effect) into the grid circuit. On the other hand, when the cathode impedance is capacitive ($i_k$ leads $e_k$), a negative resistance is introduced into the grid circuit. It has already been pointed out that when these two resistive components are equal, the total effective input resistance is infinite. Therefore, this concept may be employed as a means of neutralization of either effect.

**Table 4-2.**

| Impedance from plate to cathode | Impedance "seen" by input generator | Impedance from cathode to ground |
|---|---|---|
|   |   |   |
|   |   |   |
|   |   |   |

Note that the input capacity is larger than *Cgk* when the plate to cathode impedance is dominant, whereas the input capacity is less than *Cgk* when the cathode impedance is dominant.

## 4-4. Transit-time loading

The time of flight of electrons from cathode to plate in an ordinary vacuum tube is of the order of $1 \times 10^{-9}$ second, with most of the time spent between grid and cathode of the tube. At a frequency of 100 kHz this would represent the time necessary for the sinewave input signal to go through an amplitude variation corresponding to 0.036 degree out of 360 electrical

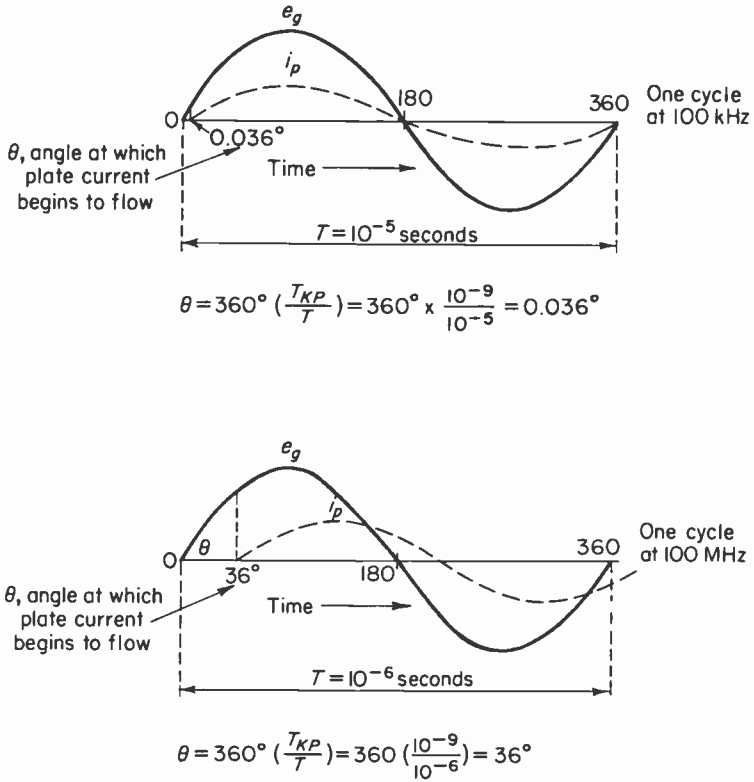



**Fig. 4-13.** Illustrates that as a result of transit time plate current lags the input voltage. At low frequencies (100 kHz) the delay is negligable whereas at high frequencies the delay can become very large.

degrees of input signal. Thus plate current would begin changing at approximately the same instant the grid voltage did. See Fig. 4-13. If the input signal is increased in frequency to 100 MHz, transit time will cause plate-current variations to occur 36 degrees after the grid voltage begins changing. Obviously, then, the plate current of a space-charge limited tube (the usual method of receiver tube operation) will lag behind the grid voltage by an angle that depends upon the transit time of electrons from cathode to plate. This plate-current phase shift is not very large at FM broadcasting frequencies, but it does play an important role in the design of UHF amplifiers and oscillators.

If we examine the interelectrode space from cathode to plate, we find that during one cycle of a 100-kHz signal the electron density throughout this space is uniform. This, of course, is due to the short transit time compared to the time of duration of one cycle of input signal. At 100 MHz the electron density from cathode to plate will not be uniform. Figure 4-14

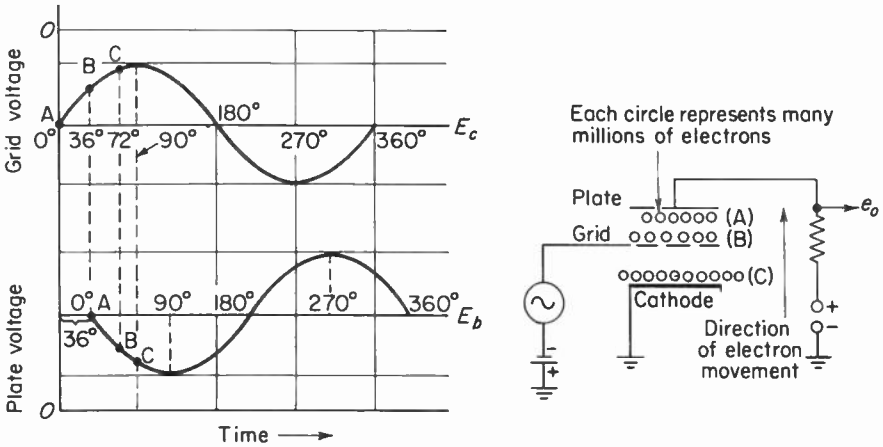



**Fig. 4-14.** The above diagram depicts the variation of electron density from cathode to plate when a high frequency signal is applied to the grid.

illustrates this fact. Note that when the grid-voltage magnitude corresponds to 72 degrees, the electrostatic field (which propagates at the speed of light†) of the grid is causing greater number of electrons to leave the space charge surrounding the cathode. At the same instant, the plate has just begun receiving electrons that left the cathode when the grid voltage was at 36 degrees. The importance of this fact is that the number of electrons approaching the grid is not the same as the number of electrons receding from the grid toward the plate. As a result, AC grid current flows even though the grid is biased negatively with respect to the cathode.‡ To understand this unusual effect, let us examine what happens when one electron travels from cathode to plate.

In Fig. 4-15(a) an electron is shown as it approaches the grid. Since the electron is negatively charged, it will repel a free electron in the grid wire,

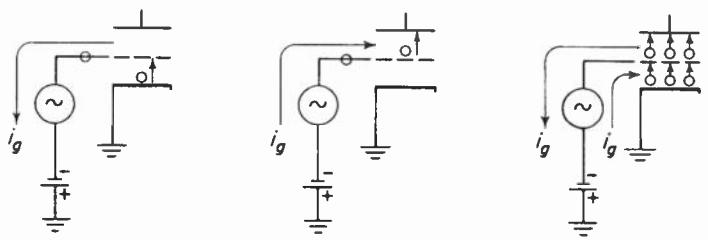



**Fig. 4-15.** (a) Shows the direction of the grid induced current as *an electron* approaches the grid. (b) Shows the direction of grid induced current when *an electron* has passed the grid on its way to the plate. (c) Indicates that the total grid current is zero when the number of electrons approaching the grid is the same as the number receding from it.

†If the space between grid and cathode were 0.3 cm, the grid would influence the space charge in $1 \times 10^{-11}$ second, or 100 times faster than transit time.

‡Of course there may be other causes for grid-current flow, notably the interelectrode capacitance $C_{gk(\text{cold})}$. For the sake of simplicity, the discussion from this point on will center around the *changes* that occur when electrons enter the space between grid and cathode. Thus, the capacitance $C_{gk(\text{cold})}$ between grid and cathode will be disregarded.

which in turn will constitute a grid current from grid to ground. When the electron has passed the grid plane [see figure 4-15(b)] on its way to the plate, its electrostatic field acting on the grid becomes weaker. Thus the electron that was repelled in a direction away from the grid now tends to return to its original position, thus constituting a current from ground to grid. Note, therefore, that as the electron approaches the grid, the grid-induced current is in one direction, whereas when the electron recedes from the grid, induced current is in the opposite direction.

The same principle can be applied, for normal low-frequency tube operation, where great numbers of electrons are moving from cathode to plate. That is, if the number of electrons approaching the grid is equal to the number receding from it, the net grid-induced current will be zero. See Fig. 4-15(c). This would be approximately the case during the peaks of the input signal at 100 kHz, where the electron density is the same from cathode to plate. On the other hand, at 100 MHz (as pointed out in Fig. 4-14) the number of electrons approaching the grid and the number receding are not equal; therefore, by means of electrostatic induction a net AC grid current will flow in the grid circuit. The magnitude of this current will depend on the various electrode potentials, the geometry of the tube, the velocity of the electrons (that is, transit time), and the frequency of the applied signal.

The phase angle between grid current and input grid voltage is determined by the same factors that determine the magnitude of the grid current. These relationships can be shown more clearly by a phasor diagram. See Fig. 4-16. If $e_g$ (the grid voltage) is used as reference, the electrons that pass the grid constitute a sinusoidally varying charge density ($q$) that will lag $e_g$ by an angle ($\theta$) that depends upon transit time. The magnitude of this angle must be a percentage of the period of one cycle of input RF. Thus $\theta$ must equal 360° ($T_{kg}/T$), where $T_{kg}$ is the transit time from cathode to grid, and $T$ is the period of one cycle of grid input voltage. Expressed in radians, 360° would be $2\pi$ and $1/T$ equals frequency; therefore, $\theta = 2\pi f T_{kg}$ radians.

It is known from basic considerations that $dq/dt$, the rate of change of

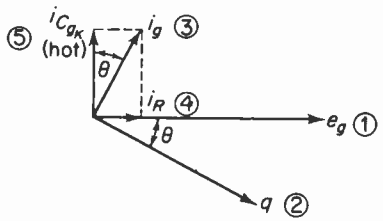



**Fig. 4-16.** Phasor diagram of the phase relationships between the voltages, currents, and charge magnitudes which exist between grid and cathode in a tube operating at a frequency where transit time affects the operation of the tube. The numbers refer to the order of construction of the phasor diagram.

charge, results in an induced current that leads $q$ by 90 degrees. That is to say, when the sinusoidal charge-density variation of $q$ reaches a maximum density level in the grid plane ($e_g$ was maximum an instant earlier), the number of electrons approaching and the number receding from the grid are equal. This means that the number of electrons repelled is maximum, but zero change in their number means that zero grid-induced current flows. When the rate of change of *charge* is maximum (90 degrees after $e_g$ was maximum) the number of electrons approaching and the number of electrons receding from the grid are at their greatest difference. Thus, the induced current is maximum. The conclusion reached, therefore, is that $q$ and $i_g$, the grid current, are 90 degrees out of phase, Since this action is electrostatic, $i_g$ leads $q$. This relationship is shown in Fig. 4-16.

From the phasor diagram (Fig. 4-16) it can be seen that the input generator delivers to the grid a current ($i_g$) made up of two right-angle components. The first component of current is leading $i_g$ by an angle of $\theta$, and is leading $e_g$ by 90 degrees. This is the capacitive component $I_{C_{gk(hot)}}$, which represents the current flowing through a shunt capacitor smaller than the interelectrode capacitance $C_{gk(cold)}$ listed in the tube manual. The tube manual lists the capacitance that exists between the tube elements without any cathode emission, which is therefore called $C_{gk(cold)}$. As already pointed out, when the cathode is heated, the space charge apparently moves closer to the grid, and thus interelectrode capacitance increases. An increase of 1 to 3 pF is typical and is referred to as $C_{gk(hot)}$. The shunt capacity introduced by transit time is $C_{gk(hot)}$. The total shunt capacity "seen" by the input generator is $C_{gk(hot)}$ plus $C_{gk(cold)}$ plus any other capacitance that may be introduced into the grid circuit.

The second component of $i_g$ is in phase with $e_g$ and thereby represents a resistive current being delivered by the generator to the grid. A resistive current delivered by the input generator represents a power loss and thus loads the input generator. The magnitude of the resistive load shunting the input generator may be derived from the phasor diagram (Fig. 4-16) as follows:

$$R = \frac{e_{g(max)}}{i_{r(max)}}$$

but by means of trigonometry we see that

$$i_{r(max)} = i_{g(max)} \sin \theta$$

$$\therefore \quad R = \frac{e_{g(max)}}{i_{g(max)} \sin \theta}.$$

The grid current is proportional to $i_p$ (the plate current); therefore $i_g$ can be said to be equal to some constant $K$ times $i_p$. The value of $K$ depends upon the geometry of the tube. A typical value of $K$ is 0.05.

$$R = \frac{e_{g(\max)}}{Ki_p \sin \theta}.$$

Since $i_p = g_m e_g$, or, in terms of maximum values,

$$i_p = g_m e_{g(\max)} \sin \theta,$$

$$\therefore \quad R = \frac{e_{g(\max)}}{Kg_m e_{g(\max)} \sin \theta \sin \theta}.$$

As was pointed out, $\theta = 2\pi f Tkg$ radians. It is a well-known fact that for *small angles*, $\theta$, expressed in radians, is approximately equal to $\sin \theta$.

$$\therefore \quad R = \frac{1}{Kg_m(2\pi f T_{kg})^2}. \tag{4–10}$$

As Eq. (4–10) shows, the degree of loading depends upon the frequency squared. Thus $R$ decreases very rapidly with an increase in frequency. At some high frequency the input resistance ($R$) of the tube will be close to zero, and thus the signal delivered between grid and cathode of the tube will be reduced to the point where useful amplification of the stage is not possible. The voltage from grid to cathode is reduced because of voltage division between the internal impedance of the input generator and the frequency-dependent transit-time resistive component. Transit-time loading also will reduce the $Q$ of any tuned circuit placed between grid and cathode of the tube. Thus, we must take transit-time loading into account when calculating the bandpass of input tuned circuits operating at high frequencies. Finally, we must consider transit-time loading when calculating total electrical noise generated by a tube.

## 4-5. Distributed properties of *L*, *C*, and *R*

At low frequencies (below 30 MHz) physically large circuit components of inductance, capacitance, and resistance (often called lumped components) are required to perform the sundry tasks assigned them in any electronic circuit. Unfortunately, the technician or student engineer often gets the idea that color-coded objects of cylindrical shape are called capacitors, while other shapes may be called resistors or coils. This approach is unfortunate because inductance, capacitance, and resistance are thought of as objects rather than as lumped electrical properties. These low-frequency concepts of inductance, capacitance, and resistance must be discarded when working at very high frequencies. The reason is that distributed components of inductance, capacitance, and resistance become more important as the frequency of operation is increased.

By distributed components we mean the inductance, capacitance, and resistance spread out along any length of an electrical device or conducting medium. To this must be added the distributed $L$, $C$, and $R$ existing between the electrical device and other objects as well as to ground. The shape of

the device will to some degree determine the amount and distribution of the distributed properties. Distributed properties exist along the leads and in the internal construction of all lumped circuit components. It is possible that at some very high frequency the distributed properties may become more important to the circuit operation than the lumped components of which the circuit is made. Depending upon application, this may or may not be a desired effect. Distributed inductance would, for example, introduce additional voltage drops and phase shifts in any circuit and thus could reduce the available signal delivered for amplification. Furthermore, it could combine with distributed or lumped $C$ to produce desired or undesired series or parallel resonance. Distributed $L$ may introduce coupling via mutual inductance, which may result in positive or negative feedback as well as crosstalk. Distributed $C$ can introduce into a circuit undesired voltage division, phase shift, resonance, and coupling. Distributed $R$, in addition to voltage division and undesired coupling, will also introduce a power loss. Such power losses make themselves felt as a reduction of tuned circuit $Q$ with subsequent increase in bandwidth.

We shall now examine distributed and lumped properties in more detail.

## 4-6. Distributed inductance

Inductance may be defined as the ability of a circuit element to generate a counter emf when a changing current is passed through it. This ability can by analogy be attributed to the inertia (mass) of the electrons in the current-carrying device. Thus all circuit elements possess the property of inductance.

The fundamental expression for inductance states that

$$L \text{ (in henrys)} = \frac{N\phi}{I} 10^{-8}, \tag{4–11}$$

where $N$ is the number of turns (unity in this discussion), $\phi$ is the flux linkages threading through the inductance, and $I$ is the current in amperes.

Any length of wire can be thought of as being made up of many parallel threads of wire. If we examine the magnetic fields surrounding each thread, we find, for a given current, that the center conductor is linked by the magnetic field of all of the other threads of wire, Thus, it has the greatest flux linkages of any component thread. By means of Eq. (4–11) we see that the center conductor will have a greater inductance than the outer ring of equivalent threads. This phenomenon gives rise to an increase in circuit resistance called skin effect. Not only does the inductance vary from center to outer edge, but the inductance of a wire also decreases by a measurable amount as the frequency of operation increases. This can be attributed to changes in magnetic flux density throughout the cross section of the wire as the frequency is increased.

The inductance of a straight wire operating at an infinite frequency and therefore its lowest value may be calculated by means of the following formula:

$$L = 0.000508\, l\left(2.303 \log_{10}\left(\frac{4l}{d}\right) - 1\right), \tag{4-12}$$

where $l$ is the length in inches and $d$ is the diameter of the wire in inches. $L$ is the inductance in $\mu$hy. From this equation we see that the inductance is directly proportional to the length of the wire and inversely proportional to its diameter. If the reactance of the distributed inductance is to be kept low at high frequencies, it becomes apparent that the length of all leads must be kept short, and the diameter of the wire must be as large as practical. From a physical point of view we may understand this by considering that an increase in length is equivalent to adding two or more lengths of wire (inductances) in series; thus their inductances add. An increase in diameter of a wire is equivalent to placing two or more lengths of wire in parallel; thus the total inductance is reduced because the inductances of the wires are in parallel.

Not only wires are important when we consider distributed inductance; any and all current-carrying bodies must be considered. For example, RF and IF bypass capacitors often make use of their distributed inductance (foil as well as lead length) to become series resonant circuits, thereby providing lower-impedance paths to ground. If the capacitance and its distributed inductance resonate below the frequency of operation, the bypass capacitor will appear to the circuit as an inductance. This condition would result in feedback and possible oscillation due to Miller effect from screen to control grid via $C_{sg}$, if the bypass capacitor were used in the screen circuit of a high-frequency IF amplifier.

Another problem associated with distributed inductance is that of mutual coupling. This occurs primarily where two or more leads, employed in circuit wiring, come in close proximity to one another. The various magnetic fields interact and provide stray coupling between one another. This stray coupling† may result in feedback, crosstalk, or other undesirable effects. To minimize stray inductive coupling, wires must be kept short, far apart from one another, and as far removed from the chassis or any tube or circuit shields as is possible. Where wires must cross one another they should do so at right angles so as to provide the smallest mutual inductance.

## 4-7. Distributed capacitance

Capacitance represents the storage of energy in the form of electrostatic lines of force. Since a potential difference can establish an electric field, it

†Distributed capacity also provides stray coupling.

becomes apparent that all energized circuit elements must exhibit distributed capacity across themselves as well as to one another. The capacitance between two parallel conductors of the same dimensions is given by the formula

$$C = 0.2249\left(\frac{KA}{d}\right), \tag{4–13}$$

where $C$ is in pF, $A$ is the area of one of the plates in square inches, $d$ is the distance between the plates in inches, and $K$ is the dielectric constant of the material between the plates.

Distributed capacitance may be undesirable because it can produce stray coupling, such as may occur from plate to control grid of an RF pentode amplifier, or undesired coupling between stages. This coupling is in addition to the interelectrode $C_{gp}$ and may be of such magnitude as to equal $C_{gp}$. Of course, distributed capacity may provide desired coupling. Some transformers, for example, use the distributed capacity between primary and secondary as a means of adjusting the coefficient of coupling between the two windings.

As the frequency is increased, the reactance of the distributed capacitance decreases to a very low value. If this capacity is in parallel with a load, fed by a generator having appreciable internal resistance, voltage division will take place and a reduced output will result. If the load is a resonant circuit, this voltage division may be reduced by making the distributed capacity part of the tuned circuit. In retuning the circuit to resonance it may be found necessary to reduce the size of the load inductance. The consequence is a reduced $LC$ ratio with a subsequent reduction in $Q$. On the other hand, if it is possible to resonate a coil with its own distributed capacity, an increased $Q$ may be obtained. This is as a result of a large $LC$ ratio at a given frequency.

The amount of distributed capacity depends upon the area of the conducting surfaces, the spacing between them, the nature of the material between the conductors, and to some extent the frequency of operation. In general, distributed capacitance can be kept to a minimum by the same procedures as outlined for minimizing distributed inductance.

## 4-8. Distributed resistance

Distributed resistance refers to the resistance a conductor offers a high-frequency current as compared to the resistance the same conductor offers a DC or low-frequency AC. The DC resistance depends upon the type of material used and is distributed uniformly in cross section as well as in length. The additional resistance that is apparently added at high frequencies, and is often referred to as the AC or RF resistance, occurs as a result of skin effect. As pointed out earlier, the inductance in the center of a wire, when high-frequency currents are flowing through it, is apparently greater than the

inductance of the wire on its surface. Thus, the reactance at the center of the wire is greater than the reactance of the wire along its surface. This means that high-frequency currents will tend to flow more readily along the surface of the wire than through its center. The higher the frequency, the thinner the "skin" through which the current flows—whereas at low frequencies or for DC the current is distributed evently throughout the cross section of the conductor.

Notice that with increasing frequency the cross-sectional area through which the current flows decreases, although there is no change in resistivity of the material. Thus, the increase in resistance at high frequencies is due to the reduction in cross-sectional area available for current flow.

The resistance of a straight copper wire at high frequencies due to skin effect may be determined from the following formula.

$$R = \frac{83.2\sqrt{f}}{d} \times 10^{-9}, \tag{4-14}$$

where $R$ is in ohms/cm, $f$ is the frequency in hertz, and $d$ is the outside diameter of the wire in centimeters. This equation assumes that the frequency is high enough so that the depth of penetration is small compared the radius of the wire.

The effect of the distributed resistance on circuit operation is to cause power losses, signal-voltage division, and reduction in tuned-circuit $Q$. These effects may be advantageous or disadvantageous depending upon application and the ingenuity of the circuit designer.

Reduction of the AC component of resistance, due to skin effect, is most often accomplished by silver-plating the conductor to a depth equal to the depth of penetration expected in the conductor at the frequency of operation. Since silver is a better conductor than copper, the effective resistance is reduced.

Also used for the reduction of AC resistance are large-diameter wire, which has greater surface area, and a multiple-strand enamel-covered wire, called Litz wire, which divides the current into many small currents and thereby tends to increase the total area of surface available for conduction. This last method is expensive and is seldom employed at very high frequencies.

Distributed resistance also includes power losses due to radiation and dielectric hysteresis loss.

Radiation losses are due to the generation of electromagnetic fields whenever high-frequency currents pass through a conductor. Energy is stored in these fields during a portion of the RF cycle. During the remainder of the RF cycle not all of the energy stored in the field returns to the wire. Thus, energy is supplied by the RF source to make up for the losses due to radiation. These losses become more important as the frequency of operation increases.

The power demand placed upon the activating generator by dielectric hysteresis loss results from the stressing and consequent distortion of the atomic structure of the insulating materials, which are acted upon by high-frequency electric fields. Certain materials, such as rubber, resist the distorting influence of the changing electrostatic field, thus placing a greater power demand on the generator than do materials, such as polystyrene or fused silica, that offer little or no opposition to the distortion of their structures. The power dissipated is directly related to both the type of insulating material and the frequency of operation; it is also inversely related to the square of the thickness of the material. Thus, we can reduce this type of loss by using low-loss insulating materials, and by making the thickness or separation between conductors as great as circuit conditions permit.

## 4-9. Characteristics of lumped components at very high frequencies

Lumped components are electronic circuit elements that exhibit concentrated amounts of inductance, capacitance, or resistance, usually with relatively little of either of the other properties present. Recalling what we have just discussed, we can see that at very high frequencies distributed properties change the effective characteristics of lumped components.

Thus, at frequencies where distributed properties become important, a resistor is not simply a resistive circuit element but appears as an impedance. Look at Fig. 4-17. This equivalent circuit is valid for carbon composition resistors. Wirewound resistors are seldom used at high frequencies because of distributed inductance and capacitance, which introduce undesired reactive and resonant effects. The capacitance across the resistor in Fig. 4-17 is due to the effects of distributed capacitance and the capacitance between the leads of the resistor. The size of the capacitance depends upon the physical shape and size of the resistance. In general, the larger the wattage rating the larger the total shunt capacitance. Typical values range from 0.1 pF to 1 pF for composition resistors. Also, as a general rule, the larger the ohmic value of the resistor the greater the change in the effective impedance with a change in frequency. The distributed inductance of a resistor is very small (0.0007 to 2 $\mu$hy) and is usually swamped by the distributed capacitance.

The complete equivalent circuit of a capacitor is shown in Fig. 4-18.

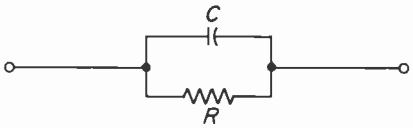



**Fig. 4-17.** The equivalent circuit of a carbon composition resistor, indicating its shunt distributed capacitance.

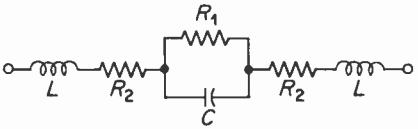



**Fig. 4-18.** The complete equivalent circuit of a lumped capacitance.

$C$ represents the actual capacitance of the capacitor, $R_1$ is the leakage resistance, $R_2$ represents the series equivalent resistive losses due to skin effect and dielectric hysteresis, and $L$ is the distributed inductance of the leads as well as the distributed inductance of the capacitor plates. The equivalent circuit illustrates the fact that the impedance of the capacitor varies with frequency. As a matter of fact all types of capacitors exhibit resonance at some frequency. It becomes apparent that at frequencies below resonance the circuit is capacitive, whereas above resonance the circuit behaves as an inductive impedance. See Fig. 4-19. Mylar capacitors resonate in a frequency range of 1 to 10 MHz, depending upon the lead length and the physical size of the capacitor. The physical size of the capacitor determines the total number of turns of foil needed to make the capacitor, and this in turn determines the distributed inductance. The dielectric material is generally of such quality that losses due to dielectric hysteresis (poor power factor) are large. These two factors combine to restrict paper capacitors to low-frequency applications.

Because of their smaller distributed inductance and losses molded mica capacitors are more useful at high frequencies than are Mylar capacitors. This type of capacitor tends to become self-resonant at frequencies ranging from 10 to 100 MHz.

Capacitors using as dielectric such materials as zirconium, titanite, and various salts of zirconium are called ceramic capacitors. Ceramic capacitors are by far the best suited for use at very high frequencies. They are economical, physically small, exhibit small dielectric hysteresis losses, and are self-resonant at frequencies of 400 to 500 MHz. The type of material and the sintering temperature determine the characteristics of the ceramic material. Thus, by the proper choice of materials and temperature, it is possible to obtain ceramic dielectric materials with dielectric constants ranging from 12 to about 10,000. Furthermore, the same technique can be used to obtain ceramic capacitors that have positive, negative, or zero temperature coefficients. As can be seen from Fig. 4-20, an increase in ambient temperature

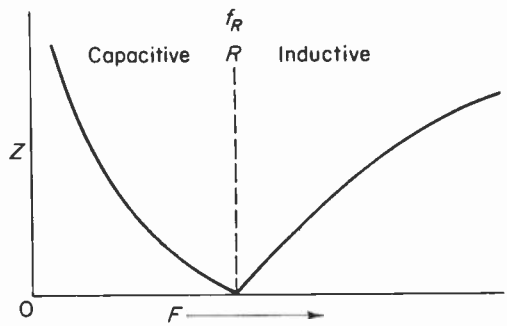



**Fig. 4-19.** The variation of impedance of a lumped capacitance at frequencies above and below series resonance.

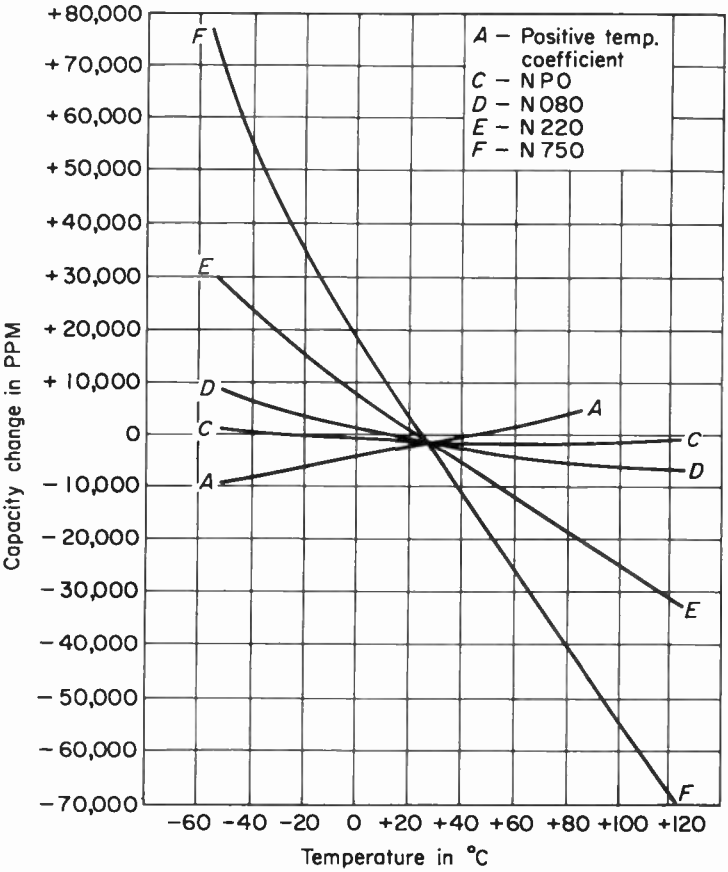



**Fig. 4-20.** The above graph indicates the changes in capacity that take place due to changes in temperature for 5 types of ceramic capacitors. The temperature coefficients range from "*P*" through NPO to N type capacitors.

will cause an increase in capacitance for a positive-coefficient capacitor, a decrease in capacitance for a negative-coefficient capacitor, and little or no change in capacitance for a zero-temperature-coefficient capacitor.

Positive-temperature-coefficient capacitors are designated by the letter "P" as in P100. P-type capacitors are seldom found in FM receivers.

The zero-temperature-coefficient ceramic capacitors designated NP0 (negative-positive-zero) as well as the N-type (negative-coefficient) capacitors are of great importance in sections of the receiver that operate at high frequencies and that may be temperature-sensitive. For instance, in order to reduce changes in critically tuned circuits, which might occur as a result of temperature changes, NP0-type capacitors may be employed as the resonating capacitors in the secondary winding of a Foster-Seeley discriminator

or ratio-detector transformer. They may also be used in the oscillator and RF stages to minimize variations in circuit conditions with changes in ambient temperature. NP0-type capacitors can be obtained in sizes from 0.5 to 300 pF. These capacitors are designed to maintain their rated capacity over a temperature range of $-20°$ to 85°C.

N-type capacitors are most often used in tuned circuits as a means of compensating for positive-temperature-coefficient inductance. In other words, an increase in temperature would ordinarily cause an increase in inductance, thereby lowering the resonant frequency of the tuned circuit, but if an N-type capacitor is placed in shunt with the coil, the decrease in capacitance will counterbalance the increase in inductance and the frequency of resonance will remain constant.

The rate at which an N-type capacitor changes its capacity with a change in temperature is obtained directly from the capacitor itself. Thus, an N750 capacitor is a capacitor in which a change of 1°C causes a change in capacity of 750 parts per million. For example, if a temperature increase of 40°C takes place in an N-type capacitor rated at 35 pF, the capacitance will decrease by a factor of

$$\left(\frac{750}{1{,}000{,}000}\right)(40) \quad \text{or} \quad 0.03.$$

The capacitance, therefore, decreases by

$$0.03(35 \text{ pF}) = 1.05 \text{ pF}$$

to a value of 33.95 pF. A formula that allows calculation of the final value of capacitance is

$$C_f = C_r(1 - N\,\Delta T \times 10^{-6}) \text{ pF}, \tag{4-15}$$

where $C_r$ is the rated capacitance in pF, $T$ is the change in temperature in degrees centigrade, and $N$ is the number of the N-type capacitor.

N-type capacitors can be obtained in sizes ranging from 0.5 pF to 500 pF, with negative-temperature-coefficient rates of change of from 30 parts per million to 5250 parts per million.

Lumped inductance is used throughout the FM receiver for such purposes as filtering (filament chokes), resonance (IF transformers), and tuned circuits (RF and oscillator). For the most part the frequency of operation dictates that the coils be physically small. The frequency of operation also dictates that the wire and coil diameter be large to reduce losses as well as to reduce changes in inductance due to temperature rise in the first half hour of receiver operation. Silver plating of the coils further reduces losses due to skin effect.

Because of the frequency of operation a high degree of tuned circuit selectivity is difficult to obtain. Also at these frequencies various losses conspire to reduce the $Q$ of the tuned circuits and further deteriorate the sharp-

ness of resonance and tuning. An increase in $Q$ can be affected by the use of high-permeability core materials since for the same inductance fewer turns of wire are required. At high frequencies three types of core materials are employed: powdered iron, ceramiclike ferrites, and brass or copper slugs.

Powdered-iron cores are made of high-permeability iron pulverized into particles as small as one micron in diameter. These particles are held together in any required shape by insulating binders. The particle size is reduced in an attempt to reduce eddy-current losses. Since eddy-current losses are proportional to the frequency squared and hysteresis loss is directly proportional to frequency, both losses become excessive at very high frequencies. To some extent hysteresis losses may be reduced by proper selection of magnetic materials. On the other hand, eddy-current losses can be reduced only by increasing the resistance or the resistivity of the material used for the core of a transformer or coil, thereby reducing the magnitude of the induced currents in the core. Thus, the very small particles that make up the powdered-iron core tend to reduce eddy-current losses because their individual resistances are increased as compared to a laminated or solid core of the same material. This approach is effective up to about 100 MHz, above which these losses again become excessive.

Ferrite cores are made of ceramiclike iron oxides, which exhibit high permeability and, unlike iron and other magnetic materials, high intrinsic resistivity. The high resistivity of ferrites is due to the ferrite crystal's being ionic. That is, the atoms of iron give up their valence electrons to the atoms of oxygen. This results in a situation where all electrons are tightly bound to their atoms and none are available for conduction. As a matter of fact, the resistivity of these cores can be as high as several megohms per centimeter, which is as much as a million billion times greater than that of metals. Thus, eddy currents are not an important cause of power loss in most ferrites.

As the frequency of operation is increased, a power loss due to magnetic resonance is introduced by the ferrite core. This loss is greatest when the signal frequency approaches the spin† frequency of the electrons, causing them to precess or wobble, and this results in a high degree of power loss. Ferrites may exhibit this loss at frequencies as low as one-tenth that of spin resonance.

†The concept of spin is a consequence of a quantum-mechanical approach to the structure of the atom. The electrons orbiting the nucleus of an atom spin upon their own axis, thereby generating a magnetic field. Most electrons are found in any atom to be grouped in pairs. Owing to the opposite spin of these electrons on their axis their magnetic fields cancel. On the basis of spin many magnetic effects can be explained. For example, iron, cobalt, and other ferromagnetic substances have an atomic structure that contains an unpaired electron. The uncanceled magnetic field arising from the spin of the unpaired electron produces the magnetic properties of these materials. In the case of a compound the electron spins of the elements "couple" with one another in such a way that each cancels the other's magnetic field. If one of the elements exhibits a stronger field than the other, the result is a ferrimagnet. Ferrites are ferrimagnets.

The resonant frequency depends upon the type of material employed and ranges from a few megahertz to many thousands of megahertz.

The expression for hysteresis loss (power loss $= KB^{1.6}f$) and the expression for eddy-current loss (power loss $= KB^2f^2$) reveal that when we are dealing with *low-permeability* materials the flux density ($B$) is a more important factor, with regard to power loss, than is the frequency of operation. This conclusion is evident in that both expressions raise $B$ to a power greater than unity, whereas $f$ (frequency) is raised to a power greater than one only in the case of eddy-current losses. Thus, for example, at a given frequency, if a magnetic material is used as a core, and its permeability is 1000, its power loss, due to eddy-currents alone, will be 1,000,000 times greater than a core whose permeability is approximately unity.

Solid brass and copper cores ($\mu < 1$) are occasionally employed to provide variable inductances at very high frequencies. This type of core material tends to reduce the coil's inductance because the eddy currents induced into the core develop a flux opposing the flux produced by the coil itself. The net flux threading the coil is reduced and thus the inductance is reduced. The change in inductance produced by brass or copper is approximately the same as that produced by an iron core. The sole advantage of nonferrous cores is that they are able to produce a variation in inductance while at the same time the losses are kept to a minimum. As already pointed out, the chief reason for the reduced losses is the low flux density. Another reason is that skin effect tends to keep the eddy-current depth of penetration very small and thus tends to reduce losses.

## 4-10. Summary

**1.** At low frequencies the grid-to-cathode circuit of a vacuum tube and the gate-to-drain circuit of an FET used as a Class $A_1$ voltage amplifier are considered to be open circuits that draw zero current from the input generator.

**2.** The apparent input circuit, as seen by the input generator, can be evaluated by noting the magnitudes of the input voltage, the grid current, and the phase angle between them.

**3.** The input impedance is generally considered to be a parallel combination of capacitance and resistance. The resistance may be either negative or positive, depending upon circuit conditions.

**4.** The resistive component of input loading can be divided into at least three major categories: (1) losses due to gas currents, grid emission, leakage resistance, dielectric losses; (2) feedback losses due to the Miller effect and cathode lead inductance; and (3) transit-time losses.

**5.** The input capacitance is contributed to by at least three sources: physical shape of the tube, the space charge, and feedback effects.

**6.** Owing to the Miller effect, triodes operated at high frequency exhibit severe input loading and a tendency to oscillate.

**7.** The input impedance of a triode depends upon the nature of the plate impedance. If the plate impedance is a pure resistance (as would be the case if the entire circuit were made resonant), the gird-to-cathode circuit impedance would appear to be a capacitance, larger than $C_{gk}$ by the additional factor of $C_{gp}(1 + A_v)$, in shunt with an infinite resistance. If the plate load is made capacitive, the input circuit will behave as though a positive resistance were placed in shunt with the total input capacity of the triode. Finally, if the impedance from plate to cathode is made inductive, the input-circuit impedance as seen by the input generator is apparently a negative resistance in shunt with the total input capacity of the tube.

**8.** At high frequencies, residual cathode inductance introduces feedback from cathode to grid, via $C_{gk}$, which draws power from the input generator.

**9.** An inductive cathode inpedance introduces a positive-resistive shunt component into the grid circuit, whereas a capacitive cathode impedance introduces a negative resistance into the grid circuit. In both of these cases a capacitance somewhat smaller than $C_{gk}$ is in shunt with the resistive components.

**10.** The effects of cathode impedance and of plate impedance on the input impedance of a tube are opposite; therefore, one may be used to neutralize the other.

**11.** At high frequencies transit time of electrons results in a power drain on the input generator.

**12.** Transit time of electrons also results in the plate current's lagging the input grid voltage.

**13.** Transit-time effects may be reduced by reducing space between cathode and grid and increasing plate-to-cathode voltages; and feedback can to some extent reduce transit-time loading ($R_{in}$ increases).

**14.** At high frequencies distributed properties rather than lumped components become important.

**15.** The inductance at the center of a wire is greater than the inductance at the edge of the wire. This gives rise to skin effect.

**16.** Owing to skin effect, high-frequency currents tend to flow near the surface of a wire. Thus, the effective resistance of the wire is increased as compared to its low-frequency value.

**17.** Litz wire, large-diameter wire, and silver plated conductors are used to reduce the increased resistance due to skin effect.

**18.** At high frequencies the presence of distributed properties of $L$, $C$,

and $R$ causes lumped components to appear to an applied generator as a complex impedance.

**19.** Ceramic capacitors are of great importance at high frequencies because of their low losses, high self-resonant frequency, and small size, and because they may be obtained with various temperature coefficients.

**20.** Owing to the high frequency of operation, the lumped inductances found in the front end of an FM receiver are generally made of large-diameter wire, silver plated, and may be tuned by means of cores made of powdered iron, ferrites, or brass.

**21.** Powdered iron reduces eddy-current losses by reducing the size of the particles that make up the core.

**22.** Ferrite cores reduce eddy-current losses because their resistivity is very high. At high frequencies they introduce a power loss due to magnetic resonance. This form of loss does not occur in cores of other types.

**23.** Since the permeability of brass and copper slugs is less than one, the magnetic flux density in these materials is low. Thus, hysteresis and eddy-current losses are kept to a minimum.

## REFERENCES

1. Ferris, W. R.: "Input Resistance of Vacuum Tubes as Ultra High-Frequency Amplifiers," *Proc. IRE*, January 1936.
2. Freeman, R. L.: "Input Conductance Neutralization," *Electronics*, October 1939.
3. Zepler, E. E.: "Valve Input Conductance at VHF," *Wireless Engineer*, February 1951.
4. Morecroft, J. H.: *Principles of Radio Communication.* New York: John Wiley & Sons, Inc., 1936.
5. *Aerovox Research Worker:* "Inductive and Reactive Effects In Straight Wires," April-May-June 1961.
6. Parker, B. E.: "Design Techniques for VHF and UHF," *Radio-Electronics*, May 1951.
7. Philpett, G. D.: "Temperature-Compensating Capacitors," *PF Reporter*, December 1959.
8. Gage, N. C.: "Ferrite Material and Components," *Electromechanical Design*, May 1961.
9. Harvey, R. L.: "Ferrites and Their Properties at Radio Frequencies," *Proc. of the Nat. Elect. Conference*, February 1954.
10. Miller, J. M.: "Dependence of Input Impedance of a Three-electrode Vacuum Tube upon the Load in the Plate Circuit," Sci. Paper 351, National Bureau of Standards.
11. Pierce, Robert: "When Is a Capacitor an Inductor?," *Electronic Products*, December 1966.

## APPENDIX 4-1 (THE MILLER EFFECT)

This appendix illustrates the manner of construction and the reasoning involved in the phasor constructions of Figs. 4-5, 4-7(c), and 4-8(a). These phasor constructions will be the basis for determining the expressions for the capacitive and resistive shunt components of input loading. In order to reduce unnecessary repetition, only phasor diagram 4-8(a) will be discussed in detail. From this construction, both expressions will be derived. It is assumed that the tube is a triode and that such other effects as cathode lead inductance and transit-time loading are negligible. The numbers associated with the phasor diagram of Fig. 4-8(a) indicate the sequence of construction employed for a tube and its inductive circuitry ($i_pZ_L$, the output voltage, leads the plate current $i_p$ by less than 90 degrees).

The first phasor component drawn is that of the plate current, $i_p$ (1). Since the plate circuit is inductive, the voltage developed from cathode to plate, $i_pZ_L$ (2), leads $i_p$ by less than 90 degrees. The voltage drop across $r_p$, $i_pr_p$ (3), is then drawn in phase with $i_p$. The phasor addition of $i_pZ_L$ and $i_pr_p$ will equal $\mu_{e_g}$ (4). By definition it is known that $e_g$ (5) and $\mu_{e_g}$ are in phase. The addition of $e_g$ and $i_pZ_L$ will provide the total voltage across $C_{gp}$. Thus, $e_{C_{gp}}$ (6) is drawn. The current through $C_{gp}$ $I_{C_{gp}}$ (7), is now found to be 90 degrees ahead of $e_{C_{gp}}$, since $C_{gp}$ is assumed to be a pure capacitance. The current through the load $I_L$ (8) lags the voltage across it ($i_pZ_L$) by an angle less than 90 degrees. The sum of $I_{C_{gp}}$ and $I_L$ equals $i_p$. The grid input voltage $e_g$ acting across $C_{gk}$ generates a current $I_{C_{gk}}$ (9), which leads $e_g$ by 90 degrees. The total grid current $I_{C_t}$ (10) is, therefore, the sum of $I_{C_{gk}}$ and $I_{C_{gp}}$. It is to be noted that $I_{C_t}$ leads $e_g$ by more than 90 degrees. Thus, the input generator ($e_g$) "sees" a capacitance and a negative resistance in shunt.

The phasor diagram of Fig. 4-8(a) may be used to derive the formula for the input resistance and capacitance.

The resistive component may be derived as follows:

$$R = \frac{e}{i} = \frac{e_g}{I_{C_{gp_r}}},$$

where $e_g$ is the input voltage and $I_{C_{gp_r}}$ is the resistive component of input current introduced by the plate into the grid circuit. Note, $e_g$ is at an angle of zero and $I_{C_{gp_r}}$ is at an angle of $-180$ degrees. But $I_C = E/X_C$, therefore

$$I_{C_{gp_r}} = \frac{e_{0_x}}{X_{C_{gp_r}}},$$

where $e_{0_x}$ is the quadrature component of plate voltage.

By means of trigonometry we see that $e_{0_x} = e_0 \sin\theta$, where $\theta$ is the angle between $e_g$ and $i_pZ_L$ ($e_0$). Thus

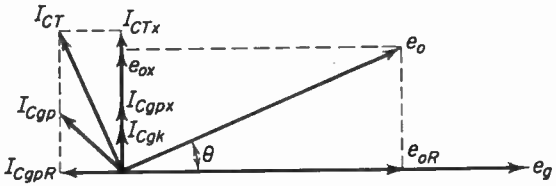



Fig. 1

$$R = \frac{e_g/0^\circ}{\dfrac{e_{0_x}/+90^\circ}{X_{C_{gp}}/-90^\circ}} = \frac{e_g X_{C_{gp}}/-180^\circ}{e_{0_x}} = -\frac{e_g X_{C_{gp}}}{e_0 \sin\theta}.$$

Dividing both numerator and denominator by $e_g$, we get

$$R = -\frac{X_{C_{gp}}}{A_v \sin\theta}.$$

By similar reasoning we can find the shunt component of capacitance:

$$X_{C_t} = \frac{e_g}{I_{C_{tx}}},$$

where $e_g$ is the applied grid voltage and $I_{C_{tx}}$ is the quadrature component of the total grid current.

From the accompanying phasor diagram it can be seen that

$$I_{C_{tx}} = I_{C_{gkx}} + I_{C_{gpx}},$$

where $I_{C_{gk}}$ is the AC current between grid and cathode via $C_{gk}$, and $I_{C_{gpx}}$ is the quadrature component of current flowing from plate to grid via $C_{gp}$.

The capacitive reactance equivalent to the total input capacity is

$$X_{C_t} = \frac{e_g}{I_{C_{gk}} + I_{C_{gpx}}} = \frac{e_g}{\dfrac{e_g}{I_{C_{gk}}} + \dfrac{e_{C_{gp}}}{X_{C_{gp}}}}.$$

But $e_{C_{gp}} = e_g + e_{0_r}$. Reference to the phasor diagram will show that $e_{0_r}$ is the in-phase (collinear with $e_g$) component of the output voltage. Therefore,

$$e_{0_r} = e_0 \cos\theta,$$

$$X_{C_t} = \frac{e_g}{\dfrac{e_g}{X_{C_{gk}}} + \dfrac{e_g + e_{0_r}}{X_{C_{gp}}}}$$

$$= \frac{e_g}{\dfrac{e_g}{X_{C_{gk}}} + \dfrac{e_g + e_0 \cos\theta}{X_{C_{gp}}}}$$

$$= \frac{e_g}{\dfrac{e_g}{X_{C_{gk}}} + \dfrac{e_g + e_g A_v \cos\theta}{X_{C_{gp}}}}.$$

By simple algebraic manipulation this reduces to

$$C_t = C_{gk} + C_{gp}(1 + A_v \cos\theta).$$

## APPENDIX 4-2 (CATHODE LEAD INDUCTANCE)

The construction of the phasor diagram of Fig. 4-9(b) and (c) is accomplished by the following procedure.

By definition $e_g$ (1) and $\mu_{e_g}$ (2) are drawn in phase. The grid-to-cathode voltage $e_g$ appears across the interelectrode capacitance $C_{gk}$; therefore, a quadrature current $i_{C_{gk}}$ (3) flows. Since $r_p$, $R_L$, and $X_{L_k}$ form an inductive circuit, the AC plate current $i_p$ (4) is drawn lagging $\mu_{e_g}$ by a very small angle [this is exaggerated in Fig. 4-9(b)]. From the circuit diagram [Fig. 4-9(a)] it is seen that $i_p$ and $i_{C_{gk}}$ add vectorially to produce the total cathode current $i_k$ (5). The voltage drop across the cathode lead inductance, $i_k X_{L_k}$, vectorially provides $e_s$ (7) ,the input signal. The vector sum of $i_k X_{L_k}$ plus the voltage drop across $r_p$, $i_p r_p$ (8), and the voltage drop across $R_L$, $i_p R_L$ (9), must equal $\mu_{e_g}$. From the final construction it can be seen that the input signal, $e_s$, "sees" a current $i_{C_{gk}}$ that leads $e_s$ by less than 90 degrees. Thus, the input generator apparently is looking into a parallel combination of a capacitor ($\approx C_{gk}$) and a resistance, the magnitude of which will be derived below.

This derivation is based upon the phasor construction of Fig. 4-9(b). From basic considerations it is known that the resistive component of input loading is

$$R = \frac{e_s}{I_r},$$

where $I_r$ is the component of input current that is in phase with $e_s$, $I_r$ is equal to $I_{C_{gk}} \cos\theta$, and $\theta$ is the angle between $I_{C_{gk}}$ and $e_s$.

$$R = \frac{e_s}{I_{C_{gk}} \cos\theta} = \frac{e_s}{\frac{e_g}{X_{C_{gk}}} \cos\theta} = \frac{e_s X_{C_{gk}}}{e_g \cos\theta}.$$

If $e_g \cos\theta$ is approximately equal to $i_p X_{L_k}$, then

$$R = \frac{e_s X_{C_{gk}}}{i_p X_{L_k}},$$

but $i_p = g_m e_g$, so

$$R = \frac{e_s X_{C_{gk}}}{g_m e_g X_{L_k}}.$$

Since $e_s$ and $e_g$ are assumed to be approximately equal,

$$R = \frac{X_{C_{gk}}}{g_m X_{L_k}}$$

$$= \frac{1}{\frac{2\pi f C_{gk}}{g_m 2\pi f L_k}}$$

$$= \frac{1}{(2\pi f)^2 L_k C_{gk} g_m}.$$

## APPENDIX 4-3

In this appendix we shall derive the formula for the resistive shunt component of input impedance when the cathode circuit is capacitive. See Fig. 4-12(a), (b), and (c) for the circuits and the phasor diagrams.

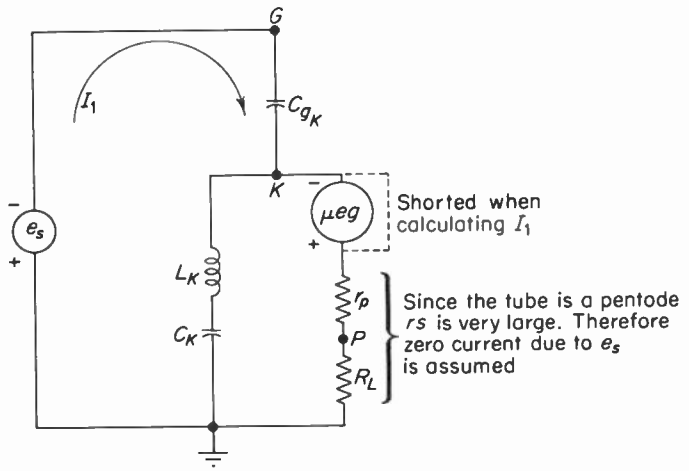



**Fig. 1.**

Using the principle of superposition, and referring to Fig. 1 of this appendix we find that the current passing through $C_{gk}$ due to $e_s$ alone is

$$I_1 = \frac{e_s}{j(X_{L_k} - (X_{C_k} + X_{C_{gk}}))}.$$

Using the second figure in this appendix it can be seen that the current passing through $C_{gk}$ due to the $\mu_{e_g}$ generator is

$$I_2 = \frac{e_{gk}}{-jX_{C_{gk}}}.$$

If $X_{L_k} \ll X_{C_k}$ and if $X_{C_k} \ll X_{C_{gk}}$, then

$$e_{gk} = i_p X_{C_k},$$

$$I_2 = \frac{i_p(-jX_{C_k})}{-jX_{C_{gk}}};$$

but $i_p$ equals $g_m e_g$, so

$$I_2 = \frac{g_m e_g X_{C_k} \angle 0°}{X_{C_{gk}}}.$$

As can be seen from the two accompanying figures, the current flowing through $C_{gk}$ due to $e_s$ is in the opposite direction from the current developed by $\mu_{e_g}$. Therefore, the total current through $e_s$ is

$$I_t = I_1 - I_2$$

$$= \frac{e_s}{Z_{in}} = \frac{e_s}{j(X_{L_k} - (X_{C_k} + X_{C_{gk}}))} - \frac{g_m e_g X_{C_k} \angle 0^\circ}{X_{C_{gk}}}.$$

If $e_s$ and $e_g$ are approximately equal, then

$$\frac{1}{Z_{in}} = \frac{1}{j[X_{L_k} - (X_{C_k} + X_{C_{gk}})]} - \frac{g_m X_{C_k}}{X_{C_{gk}}}.$$

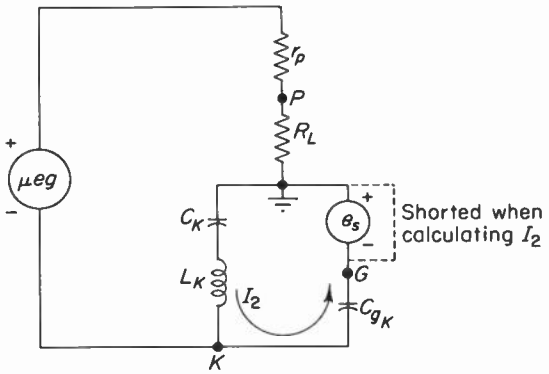



**Fig. 2.**

This first term of the equation is the magnitude of the reactive shunt component. The second term is the resistive shunt component of input admittance. Notice that it is a negative quantity.

The input resistance may also be expressed in terms of capacitance. Thus,

$$R_{in} = -\frac{C_k}{g_m C_{gk}}.$$

# 5

# RF AMPLIFIERS

## 5-1. Introduction

The RF amplifier used in FM receivers serves at least four important functions. (1) It provides a means for reducing local oscillator radiation. (2) When tuned, it reduces interference from spurious responses. (3) It increases overall sensitivity. (4) It improves the signal-to-noise ratio of the receiver. Its success in performing these functions depends on such factors as the type of mechanical tuning, cost, and the freedom from regenerative feedback. To a large degree, the practical consideration of cost determines the final compromise design for any receiver.

The sections that follow will examine in detail the means by which the RF amplifier accomplishes its tasks.

## 5-2. Local oscillator radiation

Local oscillator radiation in itself is not detrimental to the receiver from which it is emitted. It is of prime concern as a source of interference to other receivers and services operating over a range of frequencies extending from 98.7 to 118.7 MHz (assuming a 10.7-MHz IF). This band of frequencies (as well as their harmonics) includes such services as the FM broadcast band, some of the high-frequency TV channels, aviation radio navigation, and fixed and mobile services.

In an effort to reduce interference the Federal Communications Commission has set 50 microvolts per meter at a distance of 100 feet from the radiating device as the limit of radiation for any unlicensed device operat-

ing on a frequency range of 70 to 130 MHz. This, of course, includes FM receiver local oscillator radiation. The FCC rules prohibit the operation of receivers that exceed these limitations. Every commercial FM receiver manufactured after December 31, 1957, is required to carry a seal or label indicating that it complies with the radiation requirements of the FCC. Owners of receivers that exceed these requirements may be required to eliminate the interference or stop using the receiver.

The signals radiated from the FM-receiver local oscillator do so by direct radiation or by means of distributed coupling from the local oscillator to the antenna. Direct radiation can be kept to a minimum by proper shielding of the tubes and circuitry that make up the front end of the receiver. Receivers that do not employ RF amplifiers are very likely to have a great deal of distributed coupling between the antenna and the local oscillator, since the path between the oscillator and the antenna is very short and direct. As can be seen in Fig. 5-1, the coupling is accomplished via the interelectrode capacitance of the mixer tube and the distributed capacitance between the primary and secondary of the antenna input transformer.

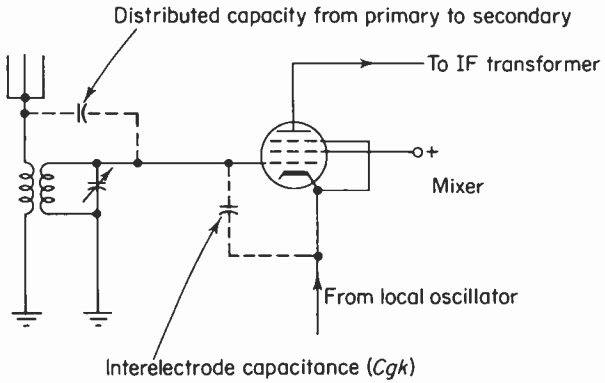



**Fig. 5-1.** In this circuit coupling from the local oscillator to the antenna is direct and is due to the distributed and interelectrode capacitance of the circuit.

The level of local oscillator signal that reaches the antenna can be materially reduced if an RF amplifier is placed between the antenna transformer and the mixer grid. The circuit illustrated in Fig. 5-2 shows that the parasitic capacitances between the oscillator and the antenna are effectively in series; thus, the total coupling capacity is reduced, and with it the oscillator radiation.

The advantages gained by the use of the RF amplifier may be lost if care is not taken with chassis layout and in the selection of the main tuning capacitor. It has been found that the shaft of the main tuning capacitor is a serious source of coupling between the local oscillator and the antenna circuit. Per-

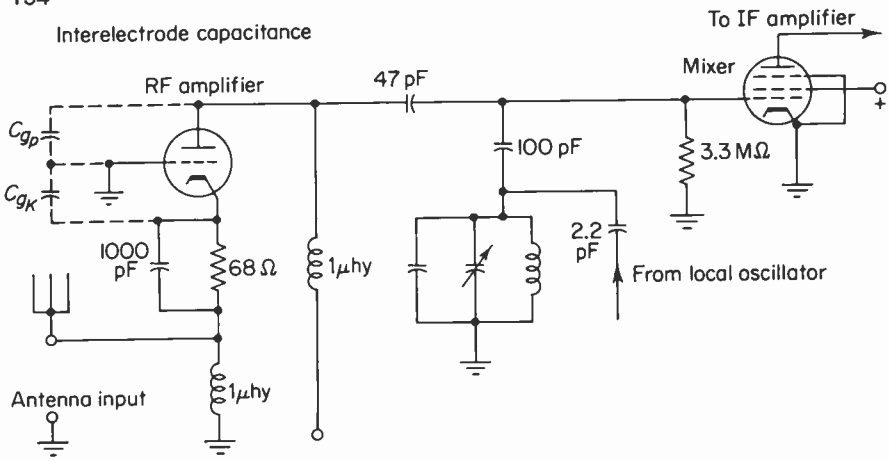



**Fig. 5-2.** Voltage division between the very low input impedance of the grounded grid RF amplifier and the relatively high reactance of the series connected interelectrode capacitance results in little local oscillator voltage reaching the antenna.

meability tuning (RF, mixer, and oscillator slugs are isolated from one another) has been suggested as a means of reducing this form of oscillator-to-antenna coupling.

## 5-3. Interference and spurious response

The superheterodyne receiver by its very nature tends to produce a multitude of spurious responses. That is, a superheterodyne receiver can deliver a signal output for a station that is operating at an assigned frequency other than the frequency to which the receiver is tuned. It is not unusual for the undesired spurious response to be very far removed in frequency from the dial setting of the receiver. This unique and unfortunate property of the superheterodyne is due to the ability of the mixer to generate sum and difference frequencies for any number of input signals that happen to be applied to the input of the mixer. It will be shown in Chapter 6 that the output of the mixer also contains higher-order detection products. These components can introduce serious forms of spurious response.

Spurious response can appear to the operator of an FM receiver as stations that are found at a number of places on the dial of the receiver. They may also fall on assigned frequencies of other stations and thus constitute a form of interference. This form of interference can be suppressed by the FM receiver, if the desired signal is stronger than the interfering signal by an amount at least equal to the capture ratio of the receiver. It is, of course, understood that if the spurious interfering signal is greater than the desired signal by the capture ratio, the desired signal will be suppressed. It is for

this reason that spurious response and other forms of interference must be attenuated or eliminated.

There are at least three important categories of interference and spurious response to which FM receivers as well as other types of superheterodyne receivers are subject: (1) adjacent and cochannel interference, (2) image response, (3) second-order intermodulation products.

### *Adjacent-channel and cochannel interference*

Adjacent-channel interference and cochannel interference affect not only the superheterodyne receiver but any type of receiver system (such as the TRF). Adjacent-channel interference is due to poor selectivity of both the RF and IF amplifiers. Cochannel interference, a condition where two channels operate on the same frequency at the same time, is, of course, a fault only of receiver location and the whim of propagation characteristics.

FM stations are assigned frequency allocations at 200-kHz intervals. The lowest assigned channel frequency is 88.1 MHz and the highest is 107.9

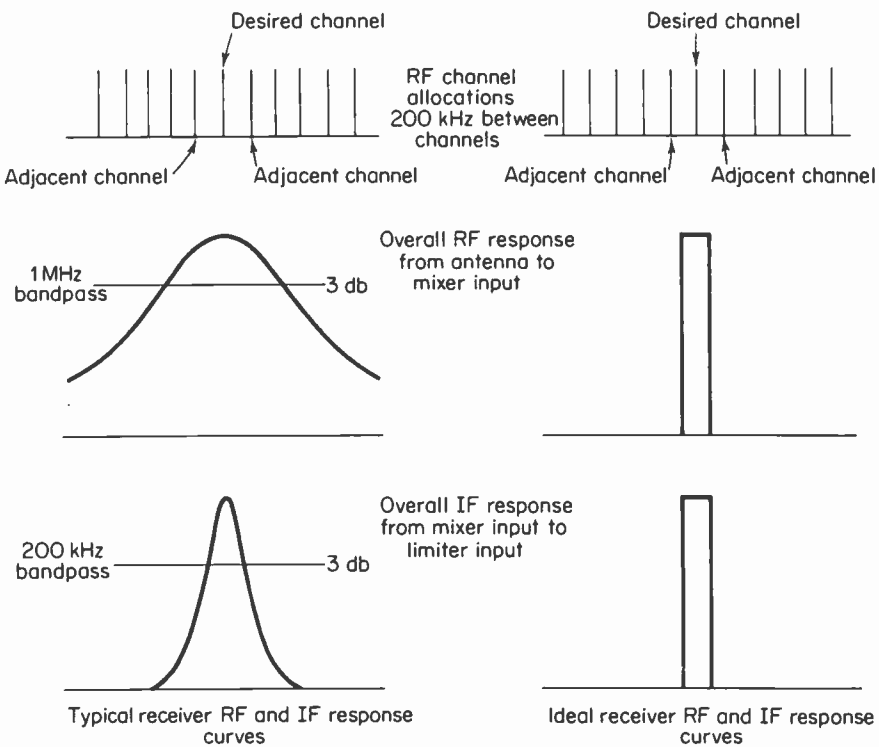



**Fig. 5-3.** The above comparison between the ideal and typical RF and IF response curves shows that in order to reduce adjacent channel interference both RF and IF response curves must have good skirt selectivity.

MHz. In an attempt to reduce adjacent-channel and cochannel interference the FCC maintains an 800-kHz separation between channels in any given locality. It is apparent that in a high-sensitivity receiver, adjacent-channel and cochannel interference may be a problem in areas between large population centers with their attendant multitude of FM stations.

Adjacent-channel interference can be reduced by improving the selectivity of the FM receiver. This requires good skirt selectivity in the IF amplifiers and if possible in the RF amplifier as well. See Fig. 5-3.

Cochannel interference can be minimized only by a low capture ratio, since the desired and the undesired signals are at the same frequency. Receivers with large capture ratios (most inexpensive receivers) will reproduce both the desired and the interfering signals at the same time.

### *Image interference*

For every position of the receiver tuning dial and thus every local oscillator frequency there are two RF signals that can produce an IF output of the mixer. One of these is higher than the oscillator frequency by an amount equal to the IF. The other is lower in frequency, and is also separated from the oscillator by the IF. Since it is current practice in FM receivers to set the local oscillator frequency above the RF, the lower-frequency response is the desired signal and the higher-frequency signal is called the image. It is also often referred to as a first-order image. This is illustrated in Fig. 5-4. From this illustration it can be seen that the image frequency is equal to

$$\text{image frequency} = \text{RF} + 2\text{IF}. \tag{5-1}$$

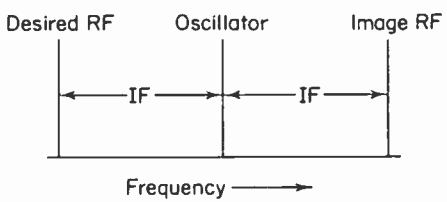



**Fig. 5-4.** Relative frequency relationships between the desired RF signal, the oscillator, and the image signal.

In receivers employing untuned (broad-band) RF amplifiers, such as the one shown in Fig. 5-5, image response can be a serious problem because both the desired and the undesired signals can reach the mixer grid. Thus, both can produce an IF output.

If the untuned RF amplifier is replaced with a tuned RF amplifier, such as the one in Fig. 5-6, image interference may be very much reduced. Reduction of 40 db is typical. This is because the image is made weaker than the desired signal by virtue of the position of the image on the response curve

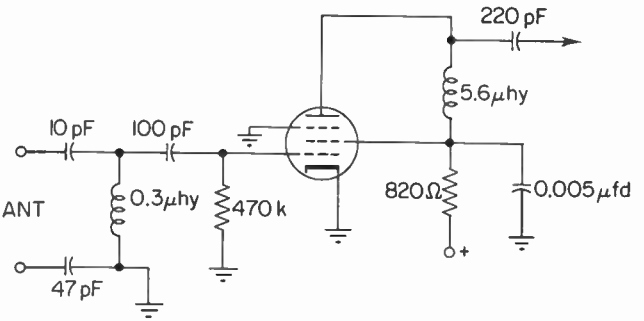



**Fig. 5-5.** An untuned (not containing a variable tuning element) pentode RF amplifier.

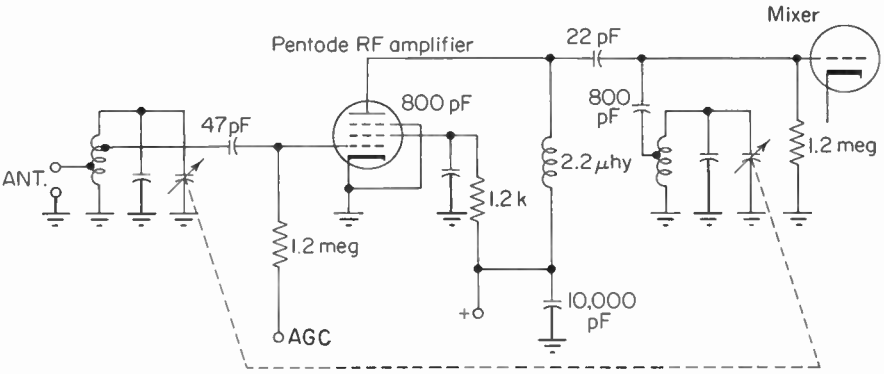



**Fig. 5-6.** Schematic diagram of a tuned RF amplifier.

of the tuned RF amplifier. A comparison between tuned and untuned RF amplifiers is made in Fig. 5-7(a) and (b).

It should not be overlooked that if the image interference is very much stronger than the desired signal, attenuation of the image by means of the RF response may not be sufficient to make the image weaker than the desired signal. Then the desired signal may be suppressed by the mechanism of capture and the resultant output would be the image.

Additional image rejection can be obtained by increasing the intermediate frequency. In this case the image frequency is further removed from the RF signal and thereby is further attenuated by the tuned response of the RF amplifier. See Fig. 5-7(b). If the IF is made greater than half the range of assigned frequencies (88 to 108 MHz, a range of 20 MHz), at no point on the receiver dial can an image appear that is also a channel on the FM band. For example, if the receiver is tuned to 88.1 MHz and if the intermediate frequency is 10.7 MHz, the oscillator frequency will be 98.8 MHz. The image frequency will be 10.7 MHz higher, or 109.5 MHz, which is outside the FM band. The likelihood of serious continuous interference is thus

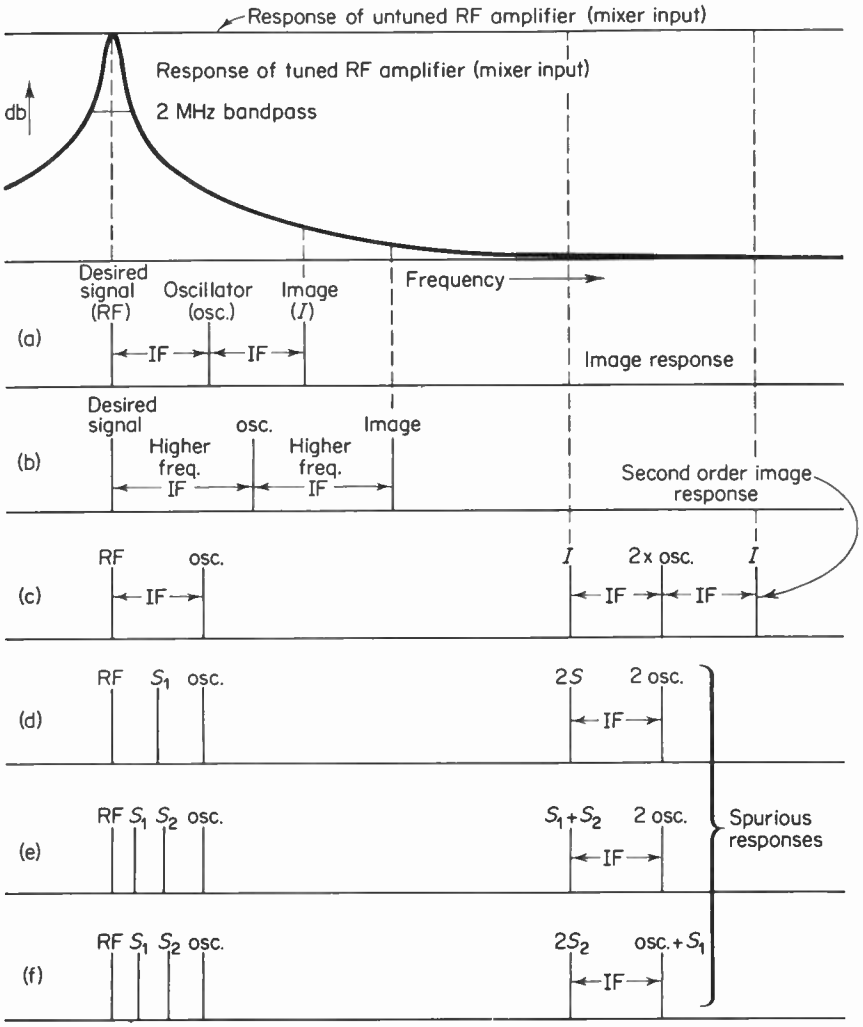



**Fig. 5-7.** The above diagram compares a tuned RF amplifier (both input and output circuits are tuned) with that of a hypothetical untuned RF amplifier (input and output circuits are untuned) and illustrates the attenuation which the undesired image and spurious responses suffer when a tuned RF amplifier is employed as the first stage of the receiver.

reduced, since the band of image frequencies (109.5 to 129.3 MHz) is used for aviation radio navigation and other aviation activities.

The situation would be somewhat different if the oscillator were lower in frequency than the desired RF signal. The band of image frequencies, assuming a 10.7-MHz IF, would then range from 66.8 to 86.5 MHz. These frequencies fall into the assigned frequencies of TV channels 3, 4, 5, and 6

(60–66 MHz, 66–72 MHz, 76–82 MHz, and 82 to 88 MHz, respectively). It is primarily for this reason that the local oscillator frequency is placed above the RF signal frequency in all modern commercial FM receivers.

### *Second-order image interference*

Owing to the nonlinearity of the oscillator, the output of the oscillator is rich in harmonics. These oscillator harmonics, when combined with the RF signal in the mixer, generate sum and difference frequencies at the output of the mixer in the same manner as the oscillator fundamental and RF signals generate an IF. If the RF input to the mixer contains signal components that differ from the oscillator harmonics by an amount equal to the IF, these signals will appear at the output of the mixer as an intermediate-frequency signal and will be amplified without discrimination by the IF amplifiers. This is termed second-order image response when the second harmonic of the oscillator is involved in the production of such interference. The frequency range of the local oscillator's second harmonics extends from 197.4 MHz to 237.4 MHz. Thus, the second-order image frequencies range from 186.7 MHz to 248.1 MHz. Included in this band of frequencies are TV channels 9 (starts at 186 MHz), 10, 11, 12, and 13 (ends at 216 MHz). Since these stations may radiate power measured in the hundreds of kilowatts, they can constitute a strong source of second-order image interference. When this form of interference is heard on an FM receiver it appears as 60-Hz buzz (vertical sync), or TV sound may be heard on the FM band.

It was pointed out in the discussion of first-order images that a tuned RF amplifier tends to discourage the production of spurious response or interference. This is also true in the case of second-order images. Figure 5-7(c) compares input to a mixer, fed by a tuned and an untuned RF amplifier, when the input to the RF amplifier is a second-order image.

Third- and fourth-order images are also possible but are rarely a source of interference in practice. The RF signals necessary to produce such interference are very far removed, in frequency, from the FM band; thus, they are very highly attenuated by the RF tuned circuits.

### *Second-order intermodulation products*

In the normal mixing process the harmonics of the oscillator can beat with the mixer-generated harmonics, sum frequency, or any other frequency combination of one or more FM stations and produce new sum and difference products. Furthermore, the FM signal themselves can beat against one another (in the RF amplifier or mixer) to produce new harmonics and sum and difference signals. If any of these intermodulation products are separated from each other by an amount equal to the IF, a spurious response will be apparent at the output of the receiver.

This form of spurious response is very common and is characterized by reception of the same signal on a number of places on the dial.

Three important sources of spurious response are illustrated in Fig. 5-8. In Fig. 5-8(a) a second harmonic of the oscillator ($2(O)$) beats with the second harmonic of a signal ($2(S)$), and since they are separated from each other by the IF a spurious response results. This can be stated in an equation:

$$\text{IF} = 2(O) - 2(S)$$

or

$$\text{IF} = 2(S) - 2(O). \tag{5-2}$$

Another source of interference can occur in the following manner. Two FM signals ($S_1$ and $S_2$) can combine in the mixer to produce a sum frequency ($S_1 + S_2$), which beats with the second harmonic of the oscillator ($2(O)$). Again a spurious response will be produced if the difference frequency is the intermediate frequency of the receiver. This is shown in Fig. 5-8(b). This is written

$$\text{IF} = 2(O) - (S_1 + S_2). \tag{5-3}$$

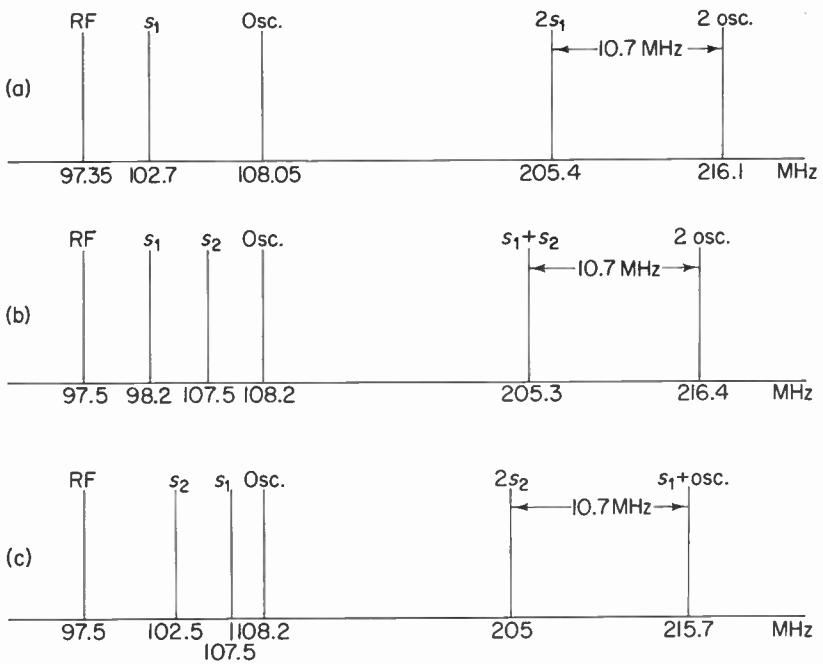



**Fig. 5-8.** Three mechanisms by which spurious responses can be generated are illustrated in the above diagram. In each case the receiver is tuned to the RF signal, thus if the RF signal is present the spurious response will appear as possible interference. Of course, if the RF signal (dial setting) is not being radiated the spurious response will appear at the output of the receiver as a signal which occurs at a number of places on the receiver dial.

An interesting combination resulting in a spurious response is shown in Fig. 5-8(c). The equation in this instance is

$$\text{IF} = (O + S_1) - 2S_2. \tag{5-4}$$

In other words, the second harmonic of an FM signal ($S_2$) beats with the sum frequency of the local oscillator and another signal ($S_1$), thereby producing a difference signal that is equal to the IF. For comparison purposes the above three sources of spurious response are also shown in Fig. 5-7(d), (e), and (f).

It should be evident that in each case if the undesired signals are prevented from reaching the mixer no spurious responses can be generated. In order to accomplish this, tuned RF amplifiers are employed between the antenna and the mixer grid.

## 5-4. Sensitivity and Types of RF Amplifiers

FM receivers use four types of vacuum-tube RF amplifiers: the grounded-cathode pentode, the neutralized grounded-cathode triode, the grounded-grid triode, and the cascode RF amplifier. Transistor RF amplifiers are either of the common-base or the common-emitter variety. Typical examples of these six amplifiers are shown in Fig. 5-9.

At first thought, voltage amplification might be considered the most important factor in the choice of an RF amplifier. Upon investigation we find that voltage amplification by itself is useless unless it can be accomplished with a low noise. Thus, the single most important factor in the selection of a weak-signal high-frequency RF amplifier is a low noise figure.

Noise in itself does not limit the maximum obtainable voltage amplification of the RF amplifier. Noise simply places a limit on the lowest level of RF signal that can produce a usable output. Such factors as the need to match the antenna to the low input impedance of the tube, the low $L/C$ ratio of high-frequency operation, and the need for stable operation combine to keep the voltage amplification of the RF amplifier low.

At low frequencies the input matching transformer provides substantial gain (step-up). At high frequencies a step-up of 5 or less is the maximum that can be obtained, owing to the need for impedance matching between the radiation resistance of the antenna and the low input impedance of the RF amplifier. Thus, if a practical stage gain of 10 is assumed (grid of RF amplifier to grid of mixer), an overall gain (as measured from the antenna terminals to the grid of the mixer) of approximately 50 is possible. The low stage gain can be accounted for in the manner outlined below.

The voltage amplification of a tuned pentode RF amplifier is equal to $g_m Z_L$. Since the parallel resonant plate-circuit impedance ($Z_L$) is equal to $L/CR$, it becomes apparent that the larger $L/C$ can be made, the greater

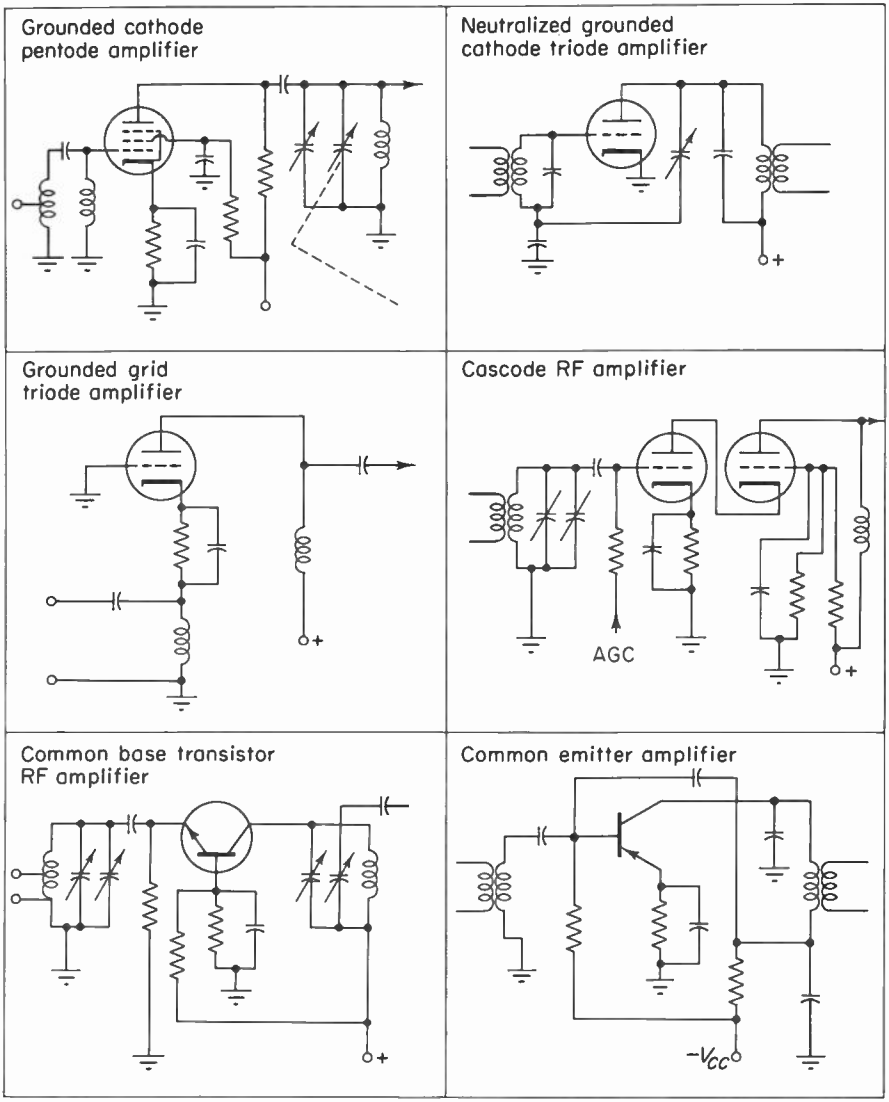



**Fig. 5-9.** RF amplifiers found in FM receivers.

$Z_L$ will become, and with it the gain of the stage will increase. Unfortunately, high gain is difficult to obtain at high frequencies because the distributed capacitance fixes the highest $L/C$ ratio that can be obtained in any given circuit at any given frequency.

For example, a pentode RF amplifier whose $g_m$ is 5000 micromhos is to operate at a frequency of 100 MHz. The resonating plate tuned-circuit

inductance is assumed to be 1.59$\mu$hy and the circuit $Q$ is 100. Under these conditions the voltage amplification would be

$$\begin{aligned} A_v &= g_m Z_L \\ &= 5 \times 10^{-3} \times 10^5 \\ &= 500, \end{aligned}$$

and the resonating value or $C$ required is

$$\begin{aligned} C &= \frac{L}{Z_L R} \\ &= \frac{1.59 \times 10^{-6}}{10^5 \times 10^1} \\ &= 1.59 \quad \text{pF}. \end{aligned}$$

Such a low value of $C$ is clearly impossible to attain in practice, since the output capacitance $C_{pk}$ of the RF stage by itself may be greater than 1.59 pF. In addition to $C_{pk}$, the wiring capacitance and the grid-to-cathode ($C_{gk}$) capacitance of the mixer combine to provide a total distributed capacity of as much as 20 pF.

If this practical value of $C$ is used in the example above, it is a simple matter to determine that the resonating inductance is 0.12 $\mu$hy. The plate load impedance is therefore

$$\begin{aligned} Z_L &= \frac{L}{CR} \qquad \left(\text{where } R = \frac{X_L}{Q}\right) \\ &= \frac{0.12 \times 10^{-6}}{20 \times 10^{-12} \times 0.75} \\ &= 8000 \quad \text{ohms} \end{aligned}$$

and the voltage amplification becomes

$$\begin{aligned} A_v &= g_m Z_L \\ &= 5 \times 10^{-3} \times 8 \times 10^3 \\ &= 40. \end{aligned}$$

In practice the voltage amplification would be lower than the value calculated, owing to the loading introduced by the mixer and the need for stability.

### *Gain-bandwidth product*

An interesting and important relationship can be obtained when the stage gain of an amplifier is multiplied by the half-power bandwidth of this amplifier. The resultant expression is independent of frequency and is, therefore, considered a constant. This fact is important, because if the type of amplifier circuit is known, and if the desired bandwidth of the stage is also known, we can determine the maximum voltage amplification that can be

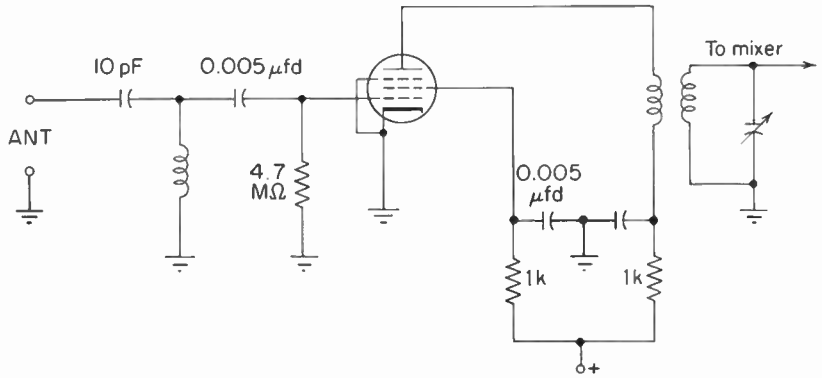



**Fig. 5-10.** A single tuned pentode RF amplifier.

obtained from this stage. Thus, for a single-tuned RF amplifier such as the one shown in Fig. 5-10, the gain-bandwidth product is

$$A_v B = g_m Q X_L \left(\frac{f_r}{Q}\right)$$
$$= g_m X_L f_r.$$

But $f_r = 1/2\pi C X_c$, so

$$A_v B = \frac{g_m X_L}{2\pi C X_c} \quad \text{Hz}$$

At resonance $X_L = X_c$ and

$$A_v B = \frac{g_m}{2\pi C} \quad \text{Hz} \tag{5-5}$$

In terms of the previous example the gain-bandwidth product becomes

$$\frac{g_m}{2\pi C} = \frac{5 \times 10^{-3}}{2\pi 2 \times 10^{-11}} = 40{,}000{,}000 \quad \text{Hz}$$

Since the frequency of operation was assumed to be 100 MHz and the circuit $Q$ was 100, the circuit bandwidth is

$$\frac{f_r}{Q} = \frac{100 \text{ MHz}}{100} = 1{,}000{,}000 \quad \text{Hz}$$

Dividing the gain-bandwidth product by 1,000,000 Hz results in the previously calculated stage gain of 40.

It is plain, therefore, that the greater the gain-bandwidth product, the greater will be the available gain per required bandpass. To increase the gain-bandwidth product the $g_m$ of the RF amplifier should be as high as possible and the capacity added to the circuit by the tube should be as small as possible.

The ratio between the $g_m$ of a tube and the interelectrode capacity is often referred to as the figure of merit of the tube. Needless to say, the

higher the figure of merit of a tube, the better suited that tube is for use at high frequencies.

### *Automatic gain control (AGC)*

Automatic gain control can serve two important functions in an FM receiver. First, it tends to keep the signal input to the limiter or ratio detector constant, thereby materially aiding the process of limiting This is especially important for receivers using the ratio detector, which has limited effectiveness in suppressing low-frequency amplitude variations (fading) .

Second, AGC can be applied to the RF and IF amplifiers to prevent possible overload, which could result in nonlinearity and change in the bandpass characteristics of these stages. Changes of bandpass or nonlinearity tend to introduce distortion and to degrade the capture ratio, thereby opening the way to various forms of interference. Cross-modulation, which is the transfer of undesired amplitude modulation from one carrier to another, is not a source of interference in FM receivers as it is in AM receivers because the limiter circuits of the FM receiver can easily remove such amplitude modulation as may be introduced.

### *The effects of overloading RF and IF amplifiers*

Overloading of the RF and mixer circuits due to strong RF input signals (100,000 $\mu$v or more) can result in reduced tuned-circuit $Q$ with its attendant poor selectivity. As has already been pointed out, poor selectivity encourages the production of spurious response.

Transistor RF and mixer stages are from 10 to 100 times less effective in preventing the generation of spurious response than are similar vacuum-tube or FET stages. This is primarily because the low imput impedance of the transistors shunts their associated tuned circuits, thereby reducing the selectivity of these circuits. Some improvement in selectivity can be obtained by the use of tapped transformers which provide proper impedance matching. Nevertheless, the selectivity obtained by the use of vacuum-tube and FET amplifiers is superior to that obtained by the use of transistor RF amplifiers. Thus, any situation that tends to deteriorate the poor selectivity characteristics of a tuned transistor amplifier will also tend to increase the generation of spurious responses.

Such deterioration of a receiver's selectivity occurs when a transistor RF amplifier is overdriven—that is, when the transistor input signal current is sufficiently great to drive the transistor into the region of either collector-current saturation or collector-current cutoff. Look at Fig. 5-11(a). Operation of a transistor in the region of saturation, where the slope of the output characteristics is almost vertical, indicates that the output tuned circuit (input to mixer) is shunted by a very low collector-to-emitter (CE) or col-

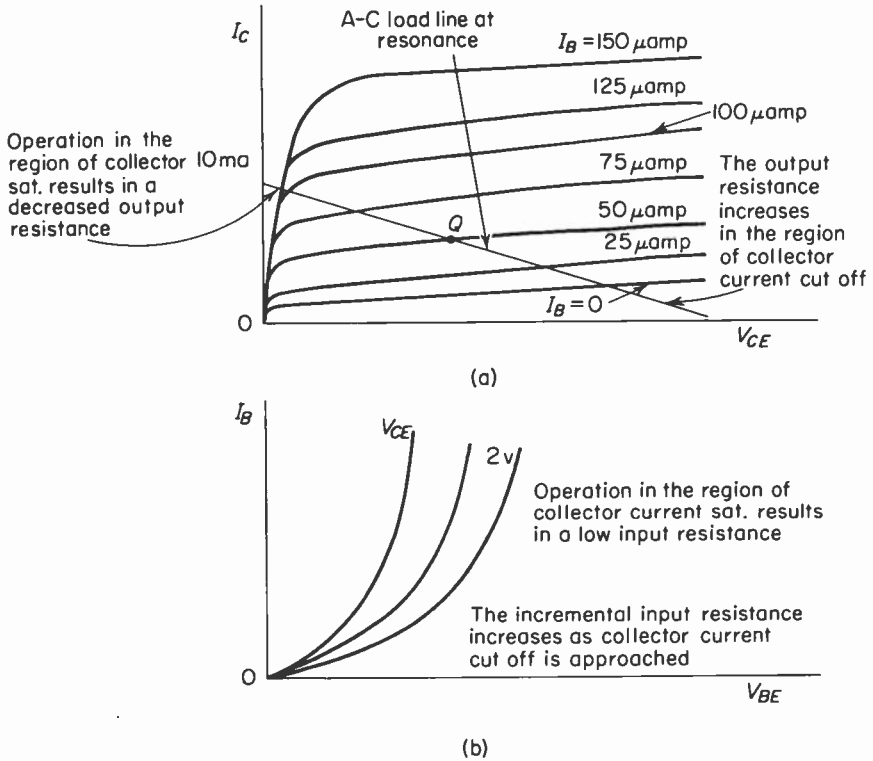



**Fig. 5-11.** The above diagram shows the effect on the incremental output resistance ($r_0 = \Delta V_{CE}/\Delta I_C$) and the incremental input resistance ($r_i = \Delta V_{BE}/\Delta V_{BO}$ for conditions of operation in regions of non-linearity.

lector-to-base (CB) resistance. At the same time that the transistor is driven into saturation, the increasing base current causes the incremental resistance from base to emitter to become smaller. See Fig. 5-11(b). Associated with these changes in input and output resistance are the changes in interelectrode capacitances. Capacitance variations can be accounted for by the changes in the thickness of the depletion areas that accompany changes in input or output operating points. Thus the bandpass, the bandpass linearity, and the center frequency of the input and output tuned circuits change during overload conditions.

On altermate half-cycles of the input signal the overdriven transistor may be driven into the region of collector-current cutoff. In this region of operation the transistor incremental input and output resistances increase (Fig. 5-11) and the interelectrode capacitances decrease. Again the net effect is changes in the input and output tuned-circuit characteristics.

Overdriving a transistor amplifier not only causes deterioration of selectivity but also may result in modulation distortion (harmonic and intermodulation distortion). In FM receivers, modulation distortion is not introduced by nonlinearity of tube or transistor operation, as is the case in AM receivers. This form of distortion results in a range of frequency components that are outside the passband of the amplifier's tuned circuits and, therefore, are bypassed without influencing the receiver output distortion levels. Modulation distortion is introduced in an overloaded FM receiver because of nonlinear phase characteristics associated with detuning of the input and output tuned circuits of the RF and IF amplifiers. Thus a change in frequency can introduce spurious phase modulation, which appears at the output of the receiver as distortion.

### *Methods for automatic gain control*

Both vacuum-tube and transistor automatic gain control vary the stage gain in inverse proportion to signal strength by changing the quiescent operating point of the controlled stages.

In the case of the automatic gain-controlled vacuum tube the operating point is normally changed by a shift in grid bias, which causes a change in the transconductance of the tube. Since the voltage amplification of a stage is directly proportional to the tube's $g_m$ ($A_v = g_m Z_L$), any variation in $g_m$ will result in a change in stage gain. The slope of the static grid transfer curve is defined as the $g_m$ of the tube. If this slope is linear, there can be no change in $g_m$ as the result of a change in grid bias. This is shown in Fig. 5-12. Nonlinearity of the grid transfer curve can be used to provide the necessary $g_m$ variation. For this purpose we can employ remote-cutoff triodes and pentodes, which are designed to exhibit a nonlinear grid transfer curve.

Another method of obtaining a remote-cutoff characteristic, and therefore a means for controlling gain, is to employ a sharp-cutoff pentode whose screen-voltage regulation is poor. Poor screen-voltage regulation results in a remote-cutoff characteristic because the effect on plate current due to a variation in control-grid voltage can be offset by a change in screen voltage. For example, if the RF signal level should increase, the AGC voltage, obtained from the limiter grid-leak bias, will also increase. This increase in the control grid-to-cathode voltage of the controlled stages will cause the plate and screen currents to decrease and the operating point of the tube to move toward cutoff. As a result of reduced screen current the voltage drop across the screen-dropping resistor will become smaller and the screen-to-cathode voltage will increase. Increased screen voltage will tend to move the operating point of the tube away from the plate-current cutoff, thus defeating the initial effects of an increasing negative control-grid bias. Since a "sliding screen" tends to prevent a decrease in plate current, a remote-cutoff characteristic is established.

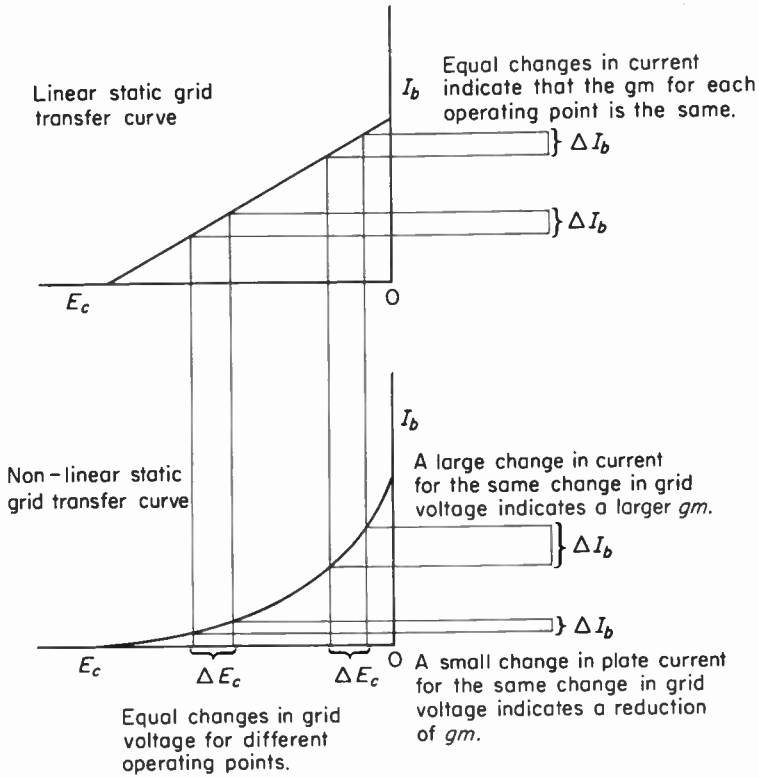



**Fig. 5-12.** The above diagram shows that a vacuum tube which exhibits a linear grid transfer curve does not produce a change in *gm* ($gm = I_b/E_c$) with a change in operating point. In contrast a tube with a non-linear grid transfer curve does exhibit a change in gm when a change in operating point occurs.

### *Transistor AGC methods*

The power gain of a transistor RF or IF amplifier may be controlled by either of two AGC systems. Both of these schemes vary gain by changing the operating conditions of the transistor. If the emitter current is used as the controlling factor, the AGC system is referred to as reverse AGC. If the collector-to-emitter voltage is used as the gain-controlling agent, the system is referred to as forward AGC.

These methods are similar in that the base is the transistor element to which the AGC voltage is applied. They differ in that an increase in RF level will cause a decrease in base-to-emitter voltage for reverse AGC, which in turn causes a decrease in $I_e$. For the same increase in RF level, forward AGC results in an increase in base-to-emitter voltage. An increase in the input forward bias results in an increase in collector current and a decrease in collector-to-emitter voltage. A reduction of collector-to-emitter voltage

will only occur provided that a large bypassed resistor (from 2 to 10 kilohms) is placed in series with the collector tuned circuit. It should be noted that in effect the collector supply voltage is being changed, since the bypass capacitor ties the collector tuned circuit to AC ground. A typical schematic diagram of both types is shown in Fig. 5-13.

Reverse AGC under increasing-signal conditions tends to shift the operating point of the transistor toward collector current cutoff. Forward AGC, under the same conditions, tends to shift the operating point of the transistor amplifier toward regions of collector-current saturation. The result of these changes upon transistor current gain (alpha or beta depending upon circuit configuration) and upon the incremental input and output

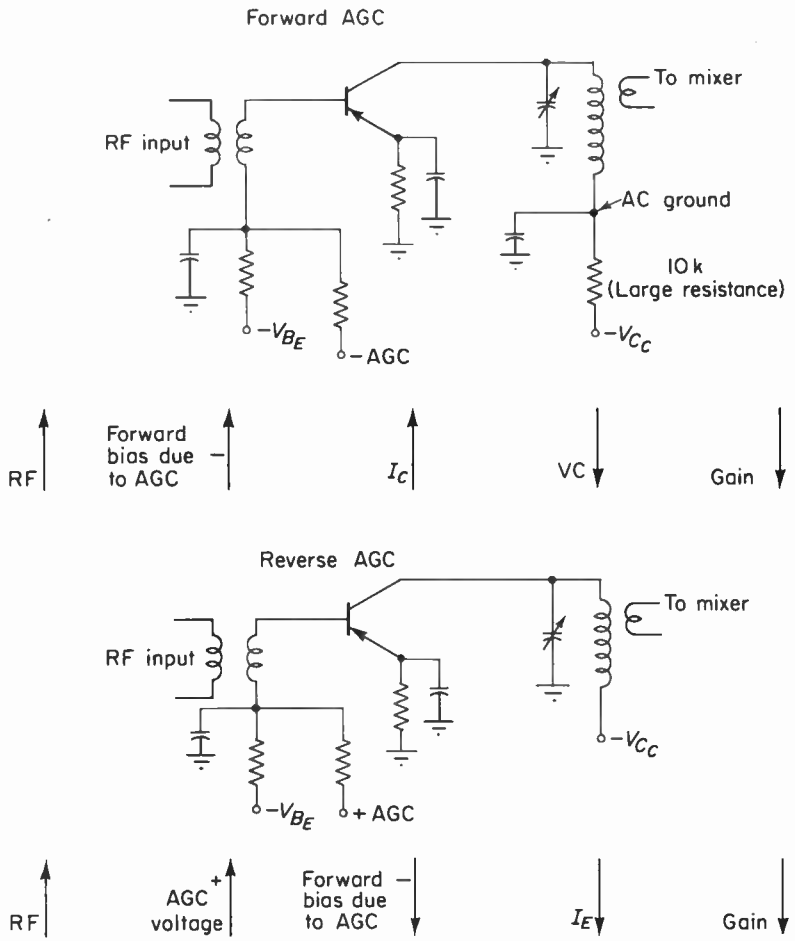



**Fig. 5-13.** Schematic diagram of forward and reverse AVC. The arrows indicate the changes in voltage, current, and gain that takes place as the result of an increase in RF input levels.

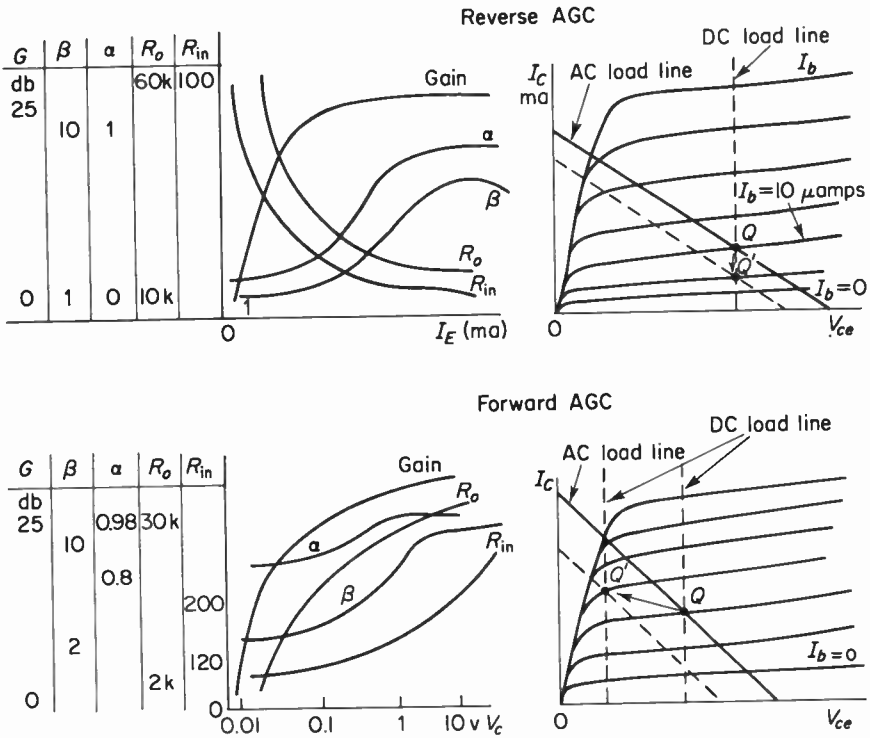



**Fig. 5-14.** The collector characteristic curves on the right show the shift in operating point for both reverse and forward AGC due to an AGC controlled change in base current. In both cases the base current change is due to an increase in RF level. The curves on the left illustrate the variations in gain, alpha, beta, $R_{in}$, and $R_o$ for a common emitter configuration. Similar characteristics will be obtained if the circuit uses a common base configuration.

impedances is seen in Fig. 5-14. In the case of reverse AGC, moving the operating point closer to cutoff causes both the input and output resistance to rise and the current gain to fall. Forward AGC, on the other hand, tends to decrease both the incremental input and output resistances as well as to cause a decrease in current gain. These effects cause a decreasing power gain (RF signal increasing), since the power gain of the transistor depends upon both the current gain and the degree of impedance match between the transistor and its associated input and output circuits.

### *AGC control voltage and filter time constant*

The AGC controlled vacuum tubes of an FM receiver generally obtain the AGC control voltage either from the grid of the last limiter stage or across the ratio-detector stabilizing capacitor. Typical circuits illustrating

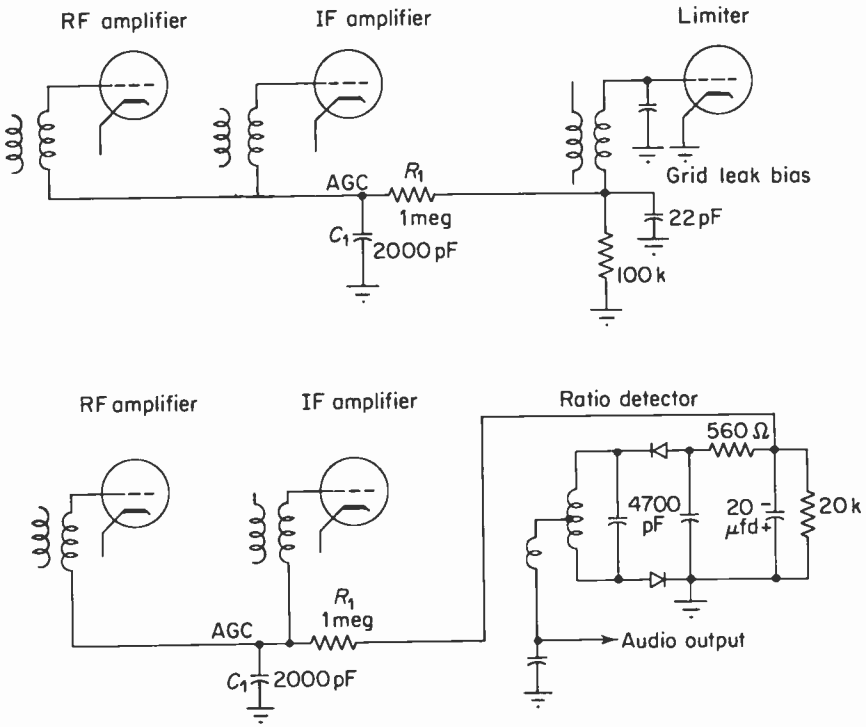



**Fig. 5-15.** Two methods of AGC take-off. Another method by which the AGC control voltage may be delivered to the grid of the RF amplifier is shown in Fig. 5-6.

the methods of AGC feed are shown in Fig. 5-15. The means by which these circuits develop a DC voltage proportional to the signal level is fully treated in Chapters 8 and 10.

As is the case in AM receivers, the AGC control voltage must be filtered before it is applied to the RF or IF stages. This is done to eliminate any signal feedback, which if left unattenuated would lead to oscillation. At the same time DC voltage feedback, from the limiter to the RF and IF amplifiers, must be unattenuated. Thus a low-pass filter such as $C_1$ and $R_1$ in Fig. 5-15 is generally used. The time constant of this low-pass filter will be found to range from 0.005 to 0.1 sec and is not very critical, since rapid or slow amplitude changes have little effect upon the audio output if the limiter stages are effective during these amplitude excursions.

Transistor FM receivers develop voltages suitable for AGC from the limiter and ratio detector in a manner similar to that of vacuum-tube circuits. However, an important difference does exist between the needs of transistor AGC circuits and vacuum-tube AGC circuits. The reason is that transistor AGC controlled amplifiers constitute a substantial power demand on the AGC

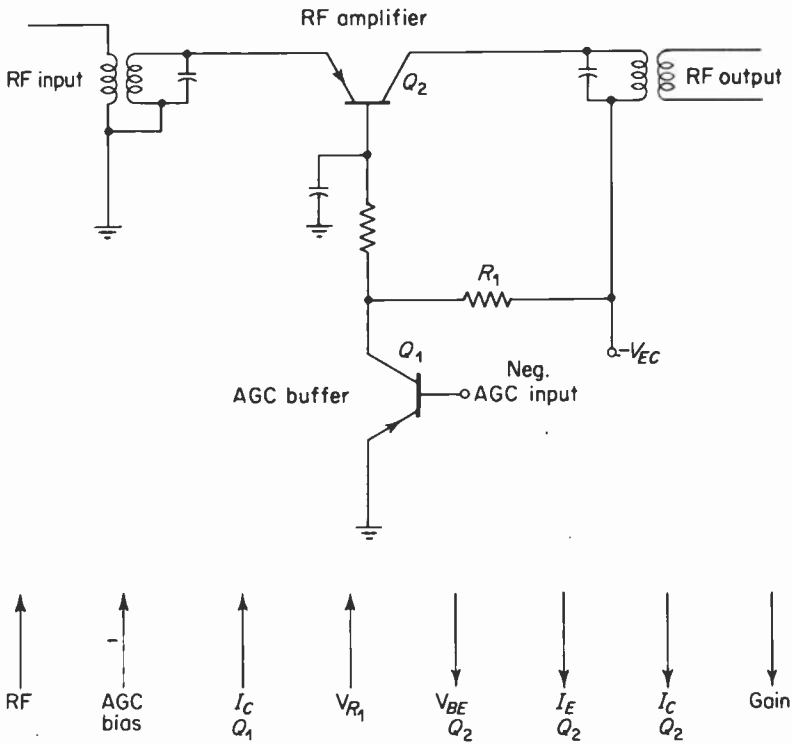



**Fig. 5-16.** An AGC buffer amplifier. The arrows indicate the changes in voltage, current, and gain that takes place as the result of an increase in RF input level.

sources. To avoid serious loading effects on the limiter or ratio-detector stages a DC buffer amplifier is sometimes employed between the AGC source and the controlled stages. Such a circuit is shown in Fig. 5-16.

*Input impedance variations due to AGC*

A major drawback both in vacuum-tube and transistor AGC controlled high-frequency amplifiers is the change in input and output impedance that occurs during changes in AGC voltage. In the case of the transistor amplifier these changes are inherent in its AGC action, and thus changes in bandpass, bandpass linearity, and center frequency of the RF and IF amplifiers can be expected. Such changes tend to increase spurious responses due to a widened bandpass, harmonic distortion due to nonlinear phase relationships of the detuned coupling circuits, and, in the case of distorted tuned circuits, rapid capture interchange. The last-mentioned effect occurs as two interfering signals deviate over a tuned response that has peaks and valleys: the stronger signal at times becomes weaker and the weaker signal may become stronger, thus by means of capture suppressing the originally

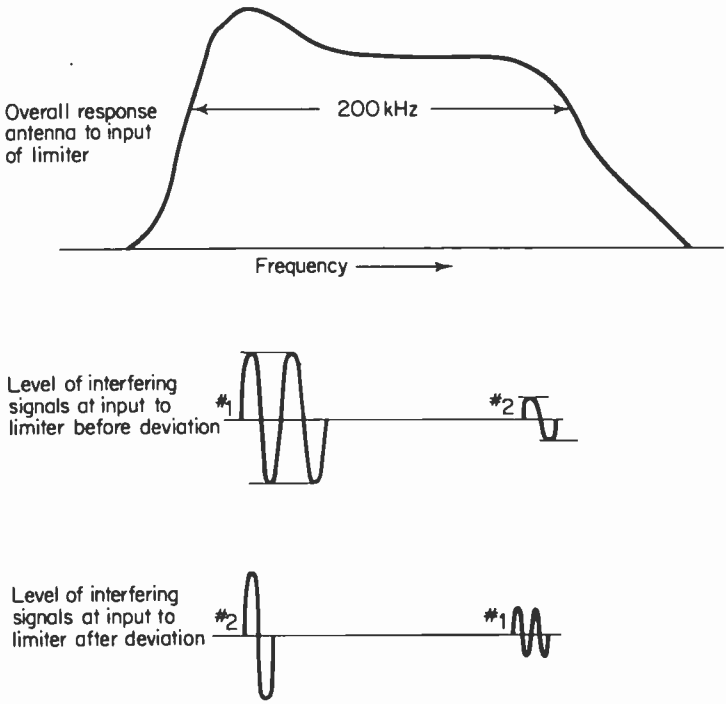



**Fig. 5-17.** The above diagram shows how a distorted receiver bandpass can result in capture interchange.

stronger signal. See Fig. 5-17. Rapid capture interchange may thus appear as a form of cross-modulation.

Vacuum-tube amplifiers also exhibit changes in input and output impedance when their operating point is shifted by AGC-derived voltages. As a result of these changes of impedance, vacuum-tube amplifiers can exhibit the same undesired effects outlined above. The vacuum tube does not usually exhibit as large a variation in its impedance for a given change in input signal as a transistor amplifier does.

These effects can be neutralized in the vacuum-tube amplifier by the simple expedient of leaving the proper size cathode-bias resistor unbypassed, and in the transistor amplifier by leaving the emitter stabilizing resistor unbypassed.

### *Types of AGC systems*

When the AGC voltage is obtained directly from a source that develops a control voltage for all levels of input signal, this system is often referred to as simple AGC. The circuit illustrated in Fig. 5-15 is an example.

Simple AGC tends to reduce the signal-to-noise ($S/N$) ratio of the receiver, because weak signals develop an AGC voltage that tends to reduce the amplitude of the already low-level signal. Since the noise level of the mixer

is left unchanged, a reduction of RF signal delivered to the mixer results in a poorer $S/N$ ratio.

A form of AGC where low-level signals generate no AGC voltage and thus operate the receiver at maximum $S/N$ ratio is called delayed AGC. "Delay" in this case does not refer to time but to amplitude. That is, delayed AGC does not develop any AGC voltage until the *level* of the input signal exceeds a certain predetermined magnitude. Thus for weak signals the RF amplifier provides maximum gain and, therefore, maximum $S/N$ ratio. High-level signals that exceed the predetermined level provide normal AGC action.

A circuit incorporating delayed AGC is shown in Fig. 5-18. In this circuit the AGC takeoff is obtained from the grid-leak bias of the limiter stage. Since the cathode of this stage returns to ground through a cathode-bias resistor, no appreciable grid current will flow until the input signal exceeds this bias voltage. Therefore, no AGC voltage will be generated for weak input signals.

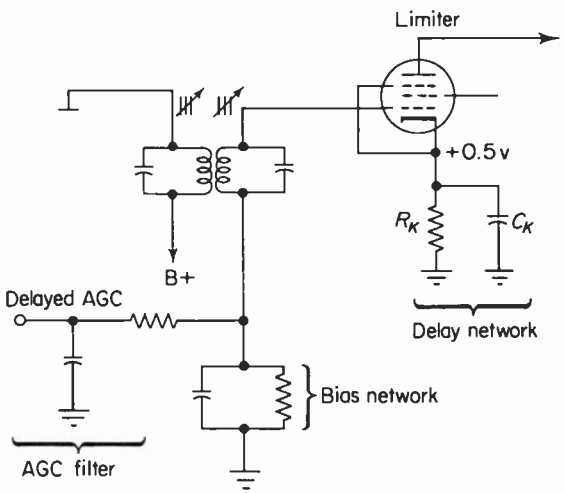



**Fig. 5-18.** Circuit arrangement for delaying the AGC voltage until the peak carrier exceeds 0.5 volt. This circuit will not properly drive a tuning indicator.

## 5-5. Characteristics of VHF amplifiers

### *The tuned grounded-cathode pentode RF amplifier*

A typical schematic diagram of a grounded-cathode pentode RF amplifier can be seen in Fig. 5-9(a).

The pentode class A RF amplifier is characterized by relatively good

stability, high gain, moderately high input impedance, and a high internally generated noise level. The pentode itself may be either sharp or remote-cutoff, depending upon whether or not AGC is employed.

Good stability is achieved because the shielding action of the screen and suppressor grids reduces (but does not eliminate) feedback from plate to control grid to very low levels. Thus we can obtain a reasonably high maximum stable gain from a single-stage VHF RF amplifier. Overall voltage amplification from the antenna to the grid of the mixer can be expected to be as high as 50.

From the constant-current equivalent circuit of Fig. 5-19(a) and (b) it can be seen that if the input impedance of the tube and the generator source impedance are ignored, the stage gain is equal to $g_m Z_L$. On the other hand, when the input shunt components (due to cathode lead inductance, transit time, leakage resistance, and Miller effect) are taken into account, the gain in terms of the input signal is $A_v = e_o/e_a$, where $e_o$ is the AC output voltage of the amplifier delivered to the grid of the next stage, and $e_a$ is the voltage delivered by the antenna to the input circuit of the amplifier. To determine the voltage amplification of a pentode in terms of its input resistance ($R_{in}$) and the Thevenin impedance of the input circuit [this includes all the lumped and distributed components of the circuit between grid (G) and cathode (K) of the amplifier except $R_{in}$], the following procedure may be used.

The total voltage amplification of the circuit (Fig. 5-19) is the product of the voltage amplification of the pentode and its load ($A_{v_p}$) and the voltage step-up of the tuned input circuit ($A_{v_t}$). Thus, $A_v = A_{v_p} \times A_{v_t}$—or, in terms of the circuit voltages, $(e_o/e_g)(e_g/e_a)$. From Fig. 5-19(b)

$$A_{v_p} = \frac{e_o}{e_g} = \frac{g_m e_g Z_L}{e_g} = g_m Z_L.$$

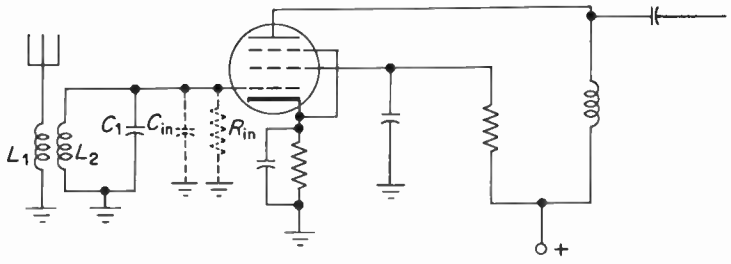
One method for determining $A_{v_t}$ is to convert the input circuit into a Thevenin equivalent circuit. See Fig. 5-19(b) and (c). The Thevenin voltage ($e_{th}$) is the voltage between gird and cathode of Fig. 5-19(b) with $R_{in}$ disconnected. Thus,

$$e_{th} = \frac{e_a \omega M Q_s}{Z_p},$$

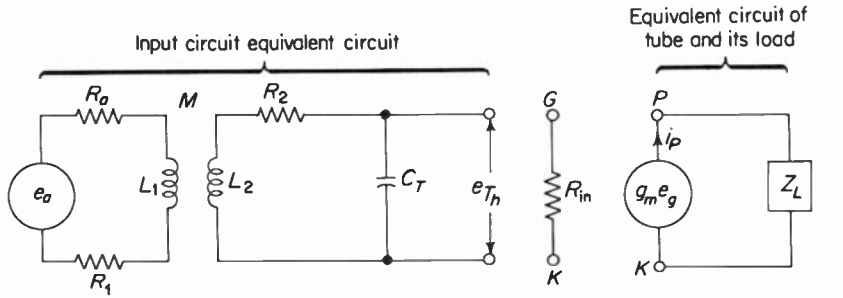
where $Q_s$ is the unloaded secondary $Q$, $Z_p$ is the primary impedance, and $M$ is the mutual inductance, The Thevenin impedance ($Z_{th}$) is the impedance between the grid and cathode of the resonant tuned circuit [Fig. 5-19(b) and (c)] with the source ($e_a$) replaced by its internal resistance ($r_a$). Thus,

$$Z_{th} = -j\left(\frac{1}{\omega C_t}\right)\frac{(R_2 + j\omega M L_2 + Z'_R)}{R_2 + Z'_R}.$$

Refer to Fig. 5-19(b) and (c) for circuit terminology.




(a)




$e_g$ — antenna voltage
$R_g$ — antenna resistance
$R_1$ — primary resistance

$$Z_R = \frac{(\omega M)^2}{Z_S}$$

$$Z_S = R_2$$

$R_2$ — secondary resistance
$R_{in}$ — input resistance of tube
$M$ — mutual inductance
$Z_L$ — plate load impedance

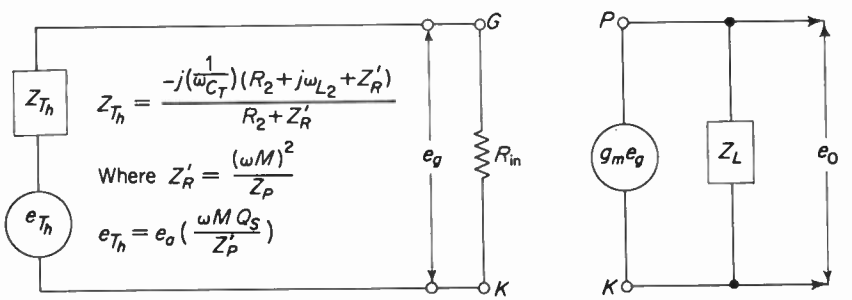
$$Z_P' = R_g + R_1 + j\omega_L + Z_R$$

$$Z_P = R_g + R_1 + j\omega_L$$

$L_1$ = primary inductance
$L_2$ = secondary inductance
$C_T = C_{in} + C_1$

$$Z_L = \frac{r_P Z}{r_P + Z}$$

(b)




$$Z_{Th} = \frac{-j(\frac{1}{\omega C_T})(R_2 + j\omega_{L_2} + Z_R')}{R_2 + Z_R'}$$

Where $Z_R' = \frac{(\omega M)^2}{Z_P}$

$$e_{Th} = e_g\left(\frac{\omega M Q_S}{Z_P'}\right)$$

(c)

**Fig. 5-19.** The above figures show the development of the equivalent circuit of a pentode amplifier. (a) Is the pentode amplifier. (b) Is the equivalent circuit of both the input circuit and the vacuum tube amplifier and (c) shows a Thevenin equivalent of the input circuit which can be used to derive an expression for voltage amplification in terms of the input circuit.

We can now determine $A_{v_t}$. From Fig. 5-19(c) it is evident that

$$e_g = \frac{e_a \omega M Q_s}{Z'_p}\left(\frac{R_{in}}{R_{in} + Z_{th}}\right),$$

$$A_{v_t} = \frac{e_g}{e_a} = \left(\frac{\omega M Q_s}{\frac{Z'_p Z_{th}}{R_{in}} + Z'_p}\right)$$

The total stage voltage amplification is

$$A_v = A_{v_p} \times A_{v_t} = g_m Z_L \left(\frac{\omega M Q_s}{\frac{Z'_p Z_{th}}{R_{in}} + Z'_p}\right). \tag{5–6}$$

Pentodes are six times noisier than triodes, mainly because of partition noise. The higher noise levels associated with the pentode are troublesome only when reception of weak or fringe-area signals is attempted. In strong-signal areas, such as might be found in metropolitan areas, this disadvantage does not reduce the pentode RF amplifier's usefulness, and it has been widely employed in many receivers.

### *Neutralized triode amplifier*

Since triodes generate less internal noise than do pentodes, triodes should be used as RF amplifiers in receivers that are expected to produce usable outputs for very weak input signals. However, triode amplifiers are less stable than pentode amplifiers and unless they are neutralized they may oscillate.

A typical schematic diagram of a neutralized triode RF amplifier is shown in Fig. 5-9(b). The triode is operated class $A_1$ in order to minimize the possible introduction of distortion and spurious responses.

In a triode amplifier that does not employ neutralization the effect upon the input impedance due to transit time and cathode lead inductance is completely swamped by the component of input impedance due to the Miller effect. The nature of this input impedance depends upon the plate load and the frequency to which both input and output tuned circuits are tuned. For example, if both input and output circuits are tuned to the same frequency, and if the applied signal is at a frequency lower than the frequency to which both circuits are tuned, the Miller-effect impedance will appear to shunt the input tuned circuit with a negative resistance [see Eq. (4–24)]. This may result in undesired instability and oscillation.

The shunt impedance that the grid source "sees" due to the Miller effect may be effectively eliminated by reducing the AC current drawn from the plate circuit by the grid input generator to zero. The most effective way of doing this is to make the input generator, connected between grid and cathode

of the amplifier, the galvanometer of a Wheatstone bridge. Obviously when the bridge is balanced the feedback current passing through the input circuit will be zero, and neutralization has been accomplished.

### *Neutralization methods*

Many methods of neutralization exist. Some are based upon capacitor bridge balance; others make use of coils. Rice and Hazeltine neutralization require tapped coils and thus have not found widespread use in receiver

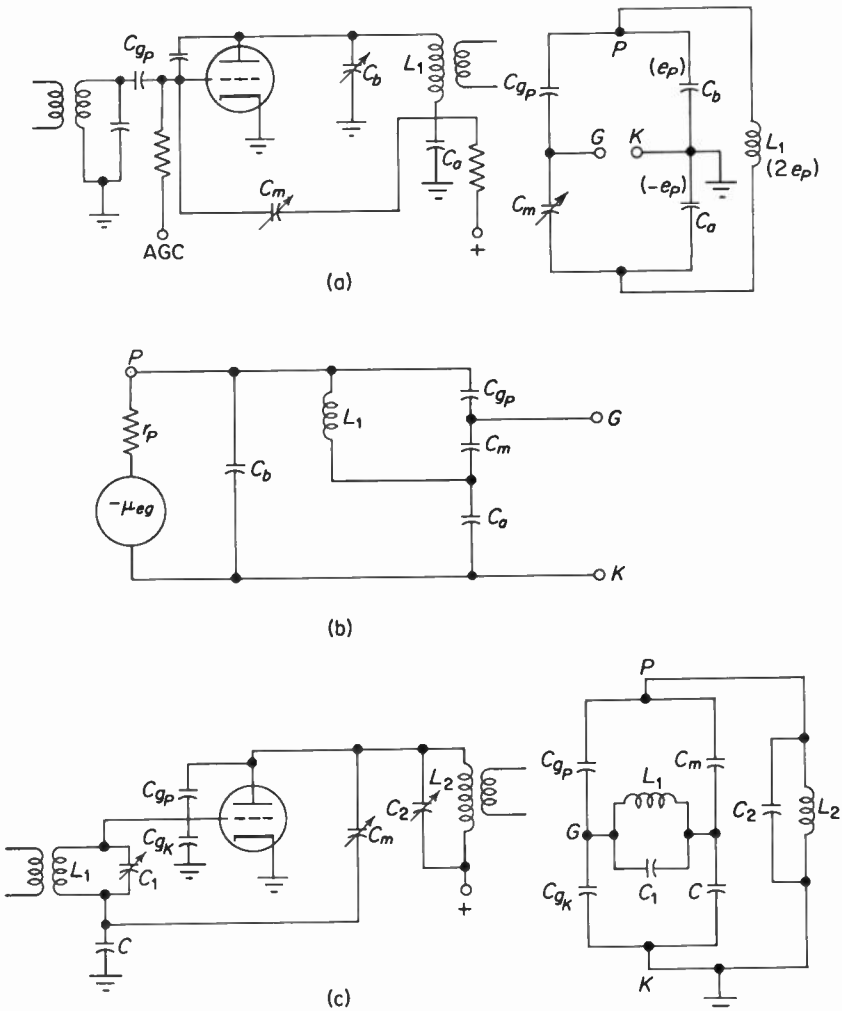
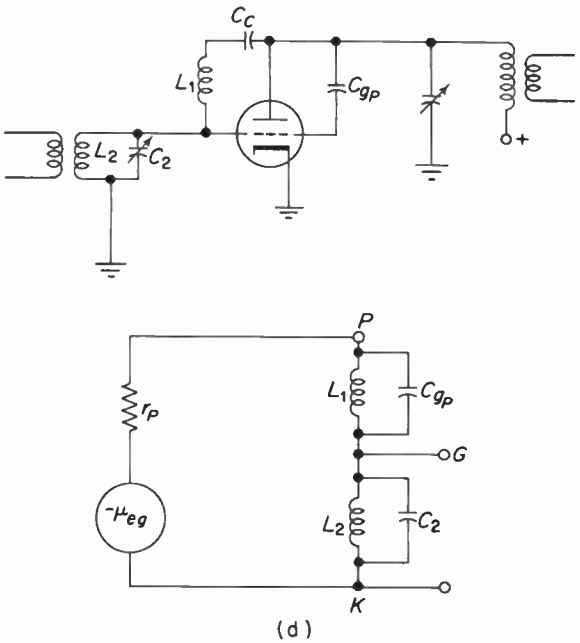



**Figure 5-20(a), (b), (c) and (d).** Three circuits and their equivalents that are commonly employed for neutralizing triode RF amplifiers.




(d)

**Fig. 5-20 (*cont'd*)**

design. Since these methods have been treated in detail elsewhere, they will not be treated here.

Three methods for neutralizing a triode are shown in Fig. 5-20. Wherever possible, accompanying each method is the bridge equivalent of each circuit.

In Fig. 5-20(a) neutralization is obtained by capacitively coupling, by means of $C_n$, part of the plate tank voltage to the grid of the triode. Of course, this feedback voltage must be 180 degrees out of phase with the feedback voltage developed in the grid circuit by $C_{gp}$. To obtain the proper phase shift the plate tank circuit behaves as a phase-shift network and provides the phase reversal required for neutralization.

The equivalent circuit of Fig. 5-20(a) is drawn in Fig. 5-20(b). It is assumed that lead inductance, transit-time effects, and other parasitic circuit conditions are negligible. It is further assumed that $X_L = 2X_{C_a}$ and that $X_{C_{gp}} + X_{C_n} \gg X_L$. $X_L$ is made twice as large as $X_{C_a}$ to provide the necessary amplitude and proper phase for the feedback voltage. Since $X_{C_{gp}} + X_{C_n}$ is in parallel with $L$, their total reactance must be made large in order to avoid changing the effective value of $L$.

It can be seen from Fig. 5-20(b) that to reduce the feedback voltage between grid and cathode to zero, the sum of $e_{C_n}$ and $e_{C_a}$ must equal zero. This can be shown as follows:

$$e_{gk} = e_{C_n} + e_{C_a}.$$

By proportion

$$e_{C_a} = e_p\left[-\frac{jX_{C_a}}{j(X_L - X_{C_a})}\right]$$
$$= -e_p\left(\frac{X_{C_a}}{X_L - X_{C_a}}\right),$$

but if $X_L = 2X_{C_a}$,

$$e_{C_a} = -e_p\left(\frac{X_{C_a}}{2X_{C_a} - X_{C_a}}\right)$$
$$= -e_p.$$

Thus the voltage across $C_a$ has been shifted with respect to $e_p$ by 180 degrees.

Similarly, by proportion,

$$e_{C_n} = e_L\left[\frac{-jX_{C_n}}{-j(X_{C_{gp}} + X_{C_n})}\right].$$

See Fig. 5-20(b). If it is assumed that $X_{C_n} = X_{C_{gp}}$,

$$e_{C_n} = e_L\left(\frac{X_{C_n}}{2X_{C_n}}\right),$$
$$= e_L\left(\frac{1}{2}\right).$$

Furthermore,

$$e_L = e_p\left[\frac{jX_L}{j(X_L - X_{C_a})}\right],$$

but if $X_L = 2X_{C_a}$,

$$e_L = e_p\left(\frac{2X_{C_a}}{2X_{C_a} - X_{C_a}}\right),$$
$$= 2e_p.$$

Therefore the voltage across the tank is twice the plate-to-cathode voltage of the tube. Thus,

$$e_{C_n} = (2e_p)\left(\frac{1}{2}\right)$$
$$= e_p.$$

Finally,

$$e_{gk} = e_{C_n} + e_{C_a}$$
$$= e_p + (-e_p)$$
$$= 0.$$

The phasor diagram of these relationships is shown in Fig. 5-21. It is obvious from the phasor diagram that since $I_L = I_{C_b}$, the circuit is anti-resonant. Furthermore, it is clearly seen that $e_{C_n}$ and $e_{C_a}$ are equal and of opposite phase; hence the feedback voltage will be zero.

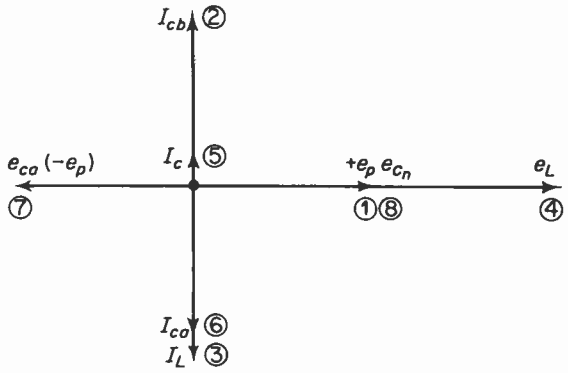



**Fig. 5-21.** Phasor diagram illustrating the phase relationships between the voltages and currents of the neutralizing circuit shown in Fig. 5-20(b).

The bridge equivalent of this circuit is shown in Fig. 5-20(c). Unfortunately, the bridge equivalent configuration obscures the phase-shift nature of this method of neutralization. It may also lead to some confusion, since the source voltage of the bridge appears to be that of the load inductance rather than the plate-to-cathode voltage of the tube. However, the analysis above shows that by proper choice of circuit components it is possible to make the voltage across each of the capacitors equal to $e_p$, and thus the circuit can be thought of as a bridge. The bridge will be balanced when the following relationship is satisfied:

$$\frac{C_{gp}}{C_n} = \frac{C_b}{C_a} \tag{5-7}$$

Balance of the bridge does not mean that the input generator sees an infinite impedance. The reason is that bridge balance eliminates only the components of grid current that originate in the plate circuit. The input generator still sees the various capacitive components that comprise the bridge. Thus the input capacity of the circuit is the parallel combination of $C_{gp}$ in series with $C_b$, and $C_n$ in series with $C_a$.

The circuit in Fig. 5-20(c) is another type of bridge neutralization, which employs a capacitor voltage divider ($C_n$ and $C$) to provide the neutralizing voltage. In this case there are two feedback paths: the first is $C_{gp}$ and $C_{gk}$; the second is purposely introduced to prevent instability and consists of $C_n$ and $C$. The voltage drop across $C_{gp}$ and $C_n$ must be equal for purposes of neutralization. As can be seen from the bridge equivalent of this circuit, the input generator is the cross arm of the bridge, and thus under conditions of balance the plate circuit can have no effect upon the input impedance as seen by the input generator.

It is to be noted that the resultant voltage drops across $C_{gk}$ and $C$ are due to the voltage division of the plate-to-cathode voltage added to the

voltage drop due to voltage division of the input signal voltage, When these voltages are combined across $C_{gk}$, it is apparent that the voltage from grid to cathode is the result of two oppositely phased voltages, and thus negative feedback is a byproduct of this scheme of neutralization.

Bridge balance in this case occurs when

$$\frac{C_{gp}}{C_{gk}} = \frac{C_n}{C}. \qquad (5\text{–}8)$$

As in the previous system of neutralization, the input impedance seen by the input generator is not infinite under conditions of balance. This is because the input source delivers current to the series combination of $C_n$ and $C$ in one branch and $C_{gp}$ and $C_{gk}$ in the other. Thus, neglecting other loading effects (transit time, leakage resistance, and so on), the input impedance is the parallel combination of each series branch.

If the loading effects listed in the last paragraph are ignored, it is apparent that neutralization is accomplished by means of capacitor voltage dividers. Thus the ratio of voltage division is frequency-independent, and consequently neutralization should be relatively unaffected by frequency variations. This is not completely true in practice because of resistive loading components that in fact do influence bridge balance.

In Fig. 5-20(d) is a circuit that is not truly a form of neutralization but is nevertheless employed in conjunction with triode RF amplifiers as a means of providing circuit stability. In this scheme the impedance from plate to grid is increased by making the interelectrode capacitance $C_{gp}$ and an external "neutralizing" inductance form an antiresonant circuit. It is obvious that the impedances between plate and grid and between grid and cathode are acting as a voltage divider. If the impedance from grid to cathode is very much smaller than the impedance from plate to grid (this will be so if $L$ is a high-$Q$ coil), very little plate-circuit signal voltage will be developed between grid and cathode of the tube.

If both circuits are at resonance, the feedback voltage will be in phase opposition to the applied signal, and negative feedback will result. Amplifier stability in this case is a result of plate-to-grid feedback reduction and gain reduction due to negative feedback.

Since the impedance from plate to grid and the impedance from grid to cathode do not vary in direct proportion to frequency, this type of neutralization is useful only over a limited frequency range.

If the plate load impedance of a triode RF amplifier is small (such as in the first stage of a cascode RF amplifier), the tendency for the circuit to be unstable is reduced and, therefore, the need for neutralization is not apparent. However, a triode amplifier that is not neutralized will exhibit considerable Miller-effect loading. The major effect of this loading is to change the optimum source resistance required to obtain a minimum noise figure for the RF amplifier. The lack of neutralization also results in less voltage amplification,

since the tuned input transformer will be loaded by the Miller-effect resistive component. The voltage amplification of the neutralized triode RF amplifier can in fact be equal to or greater than that of a pentode RF amplifier of equal $g_m$ and equal plate load impedance. This is shown by Eq. (5–6).

For a pentode, $R_{in}$ consists of the parallel combination of resistive loading components due to cathode lead inductance, transit time, leakage resistance, and *Miller effect*. The Miller effect must still be considered in a pentode because $C_{gp}$ is not zero and in fact may be twice as great as given by the published data, owing to the effects of distributed circuit capacity. In a *neutralized triode*, on the other hand, the Miller-effect resistance is infinite, and thus $R_{in}$ can be two to three times larger than that found in a pentode. Thus in practice the gain of the neutralized triode may be the same or somewhat greater than that of the pentode, owing to the voltage step-up contributed by the tuned input transformer.

### *Oscillator radiation*

The neutralized triode RF amplifier provides high gain and good stability, but in doing so provides an additional path (via $C_n$) for the local oscillator signal to reach the antenna. Unless care is taken, local oscillator radiation may exceed the limits set for it by the FCC. To minimize this a Faraday shield may be placed between the primary and secondary of the input transformer. Another possibility is the placing of low-pass filters between the grid of the amplifier and the antenna.

### *The grounded-grid RF amplifier*

A triode circuit configuration that exhibits somewhat less gain than a pentode, low noise, and good stability is the grounded-grid amplifier. Figures 5-2 and 5-9(c) illustrate typical schematic diagrams of the grounded-grid amplifier.

In this circuit, just as in the grounded-cathode amplifier, the input signal is applied between the grid and cathode of the tube. However, in this case the grid is at AC ground potential and, therefore, acts as a Faraday shield minimizing any capacitive feedback from plate to input generator (cathode). Thus neutralization is not required.

As a result of the manner of circuit connection (see the AC equivalent circuit in Fig. 5-22), the AC plate current must flow through the input generator. Therefore, the input impedance,† assuming $X_{C_{k_2}}$ is very large, as seen by the input generator is

$$Z_{in} = \frac{e_s}{i_p},$$

$$= \frac{r_p + r_s(1 + \mu) + R_L}{(1 + \mu)}. \qquad (5\text{–}9)$$

†See Appendix 5-1 for the derivation of this equation.

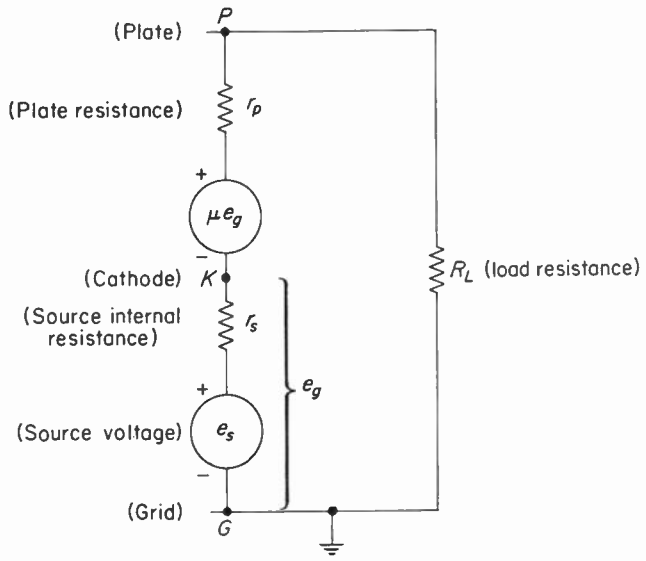



**Fig. 5-22.** The equivalent circuit of the grounded grid amplifier.

If $\mu$ is much greater than 1 and if $R_L$ is approximately equal to $r_p$, as may be the case with triodes,

$$Z_{in} = \left(\frac{2}{g_m}\right) + r_s. \tag{5-10}$$

If the $g_m$ of the tube is assumed to be 5000 $\mu$mhos and if the source resistance ($r_s$) is 100 ohms, $Z_{in}$ will be found to be 500 ohms.

In shunt with the input impedance, $(2/g_m) + r_s$, is the transit-time resistance. Transit-time loading is unaffected by circuit configuration and, therefore, will have the same magnitude as in a grounded-cathode amplifier.

Any residual Miller effect (due to $C_{kp}$) will result in a slightly reduced input capacitance, since the plate voltage and grid voltage are in phase as measured to ground. This means that the voltage across $C_{kp}$ will be the difference between $e_p$ and $e_{gk}$. Thus, the current through the feedback capacitance must be small compared to the situation where $e_p$ and $e_{gk}$ are 180 degrees out of phase (measured to ground) as in the grounded-cathode amplifier. The value of $C_{in}$ for a grounded-grid amplifier can be shown to be

$$C_{in} = C_{gk} + C_{gp}(1 - A \cos \theta), \tag{5-11}$$

where

$C_{in}$ = effective capacitance between cathode and grid,
$C_{gk}$ = interelectrode capacitance between grid and cathode,
$A$ = voltage amplification of the stage, and
$\theta$ = phase angle between input and output voltages.

The voltage amplification of the grounded-grid amplifier can be determined from the equivalent circuit of the tube. Thus,

$$A_v = \frac{e_o}{e_s},$$

$$e_o = i_p R_L,$$

$$i_p = \frac{e_s + \mu e_g}{r_p + r_s + R_L};$$

but

$$e_g = e_s + i_p r_s,$$

and

$$\therefore i_p = \frac{e_s(\mu + 1)}{r_p + r_s(\mu + 1) + R_L},$$

$$e_o = \frac{e_s(\mu + 1)R_L}{r_p + r_s(\mu + 1) + R_L},$$

$$A_v = \frac{(\mu + 1)R_L}{r_p + r_s(\mu + 1) + R_L}. \tag{5-12}$$

Since the plate current is common to both input and output circuits, it can also be seen that

$$A_v = \frac{e_o}{e_{gk}} = \frac{i_p R_L}{i_p R_{in}} = \frac{R_L}{R_{in}}. \tag{5-13}$$

Furthermore, since $i_p^2 R_L = P_o$ and since $i_p^2 R_{in} = P_{in}$, the ratio of power output to power input (power gain) must be identical to that of voltage amplification. Hence

$$A_v = PG = \frac{R_L}{R_{in}}. \tag{5-13a}$$

The noise figures of the grounded-grid triode amplifier and the grounded-cathode triode amplifier are about the same, assuming identical values of circuit parameters. In both cases minimum noise figure is obtained when the effective value of input source resistance is

$$R_{s_{opt}} = \sqrt{\frac{R_1 R_{tt} R_n}{R_1 + R_{tt}}}$$

$$= \sqrt{R_L R_n}, \tag{5-14}$$

where

$R_{s_{opt}}$ = optimum source resistance,
$R_1$ = shunt loading components due to circuit conditions other than transit time,
$R_{tt}$ = transit-time resistance,
$R_n$ = equivalent shot-noise resistance,
$R_L$ = parallel combination of $R_1$ and $R_{tt}$.

If $R_{tt} \gg$ than $1/g_m$ and if $\mu \gg 1$, the expression for noise figure of these amplifiers, assuming the source resistance is optimized, is

$$NF_{opt} = 1 + 2\sqrt{\left(\frac{1}{R_1} + \frac{1}{R_{tt}}\right)R_n}. \tag{5-15}$$

The disadvantages of the grounded-grid amplifier are its low input impedance, which makes it difficult to use variable tuning in the input circuit, lower available amplification than is provided by the neutralized grounded-cathode amplifier, and extreme variation of input impedance when AGC is employed.

### *The cascode amplifier*

If an ideal RF amplifier could be constructed, it would have the noise figure of a triode and the gain of a pentode. This seeming contradiction can be resolved, provided that a two-stage RF amplifier is employed in place of a single RF stage. To provide the lowest noise figure possible both stages must be triodes. This in itself provides a problem, in that unless care is taken the circuit will be unstable.

Wallman has determined that of the nine possible combinations of two-stage circuit configuration (GK-GG, GK-GK, GK-GP, and so on) the grounded cathode feeding the grounded grid (see the block diagram in Fig. 5-23) provides the best noise figure, stability, and gain. This circuit is called the cascode amplifier.

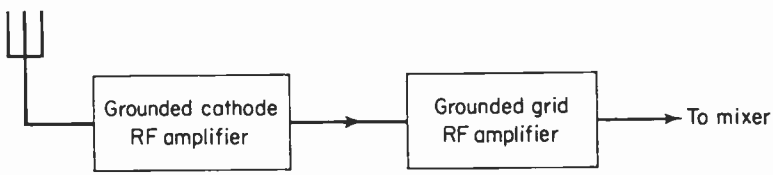



**Fig. 5-23.** A block diagram of a cascode RF amplifier.

Figure 5-24 shows two methods of coupling that may be employed between the grounded-cathode stage and the grounded-grid stage of the amplifier. In Fig. 5-24(a) a coupling capacitor is used to feed the signal from the plate of $V_1$ to the cathode of $V_2$. Since the input impedance of $V_2$ is very low ($2/g_m$), the interstage circuit is broadly tuned over the FM band by means of $L_x$ and $C_{dis}$.

### *Stacked B plus and AGC*

The circuit shown in Fig. 5-24(b) is sometimes referred to as a stacked B arrangement. This is because the two tubes are connected as a series voltage divider across the B supply. If both tubes are identical and have the same

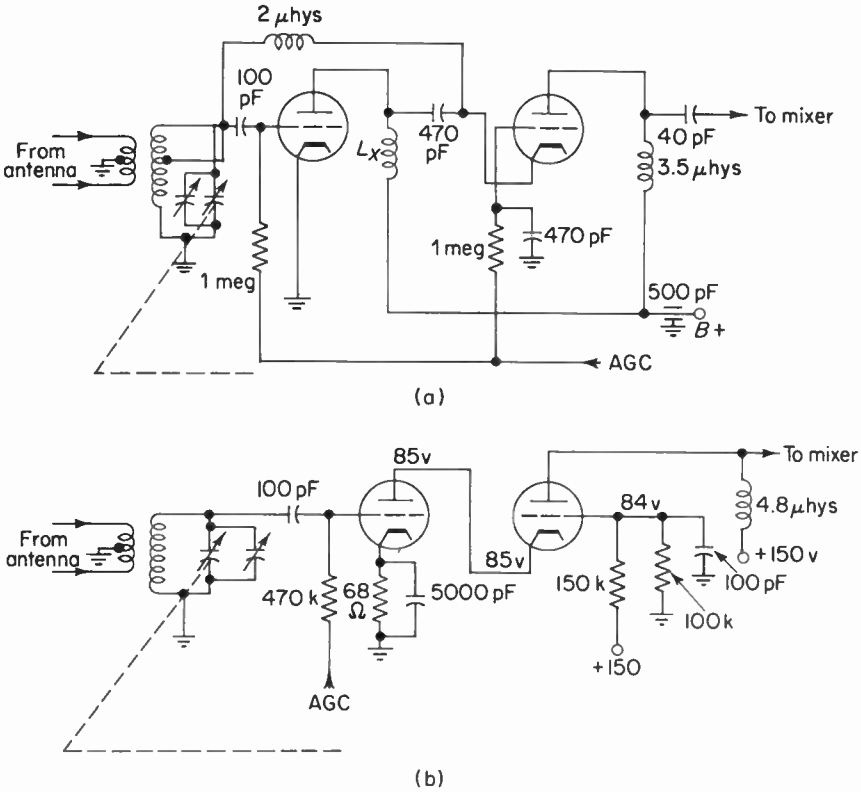



**Fig. 5-24.** Two methods of coupling the grounded cathode amplifier to the grounded grid amplifier in a cascode arrangement.

operating conditions, the DC voltage across $V_1$ will be the same as the voltage across $V_2$. Simplicity of circuit connection is the obvious advantage of this system. This circuit also provides a remote-cutoff characteristic when AGC is employed. This comes about because any change in bias of the first stage causes the plate voltage to change in a direction that opposes the change in current. That is, an increase in bias in the first stage will cause a decrease in plate current that will tend, owing to voltage division between tubes, to increase the plate voltage of the first tube. Since the first stage is a triode, an increase in plate voltage tends to increase plate current, thus increasing the magnitude of the grid voltage required to cut off the tube. In effect the tube has become semiremote cutoff.

The bias developed between grid and cathode of the second stage is provided by returning the grid to a potential somewhat less positive than the cathode. The grid is held at AC ground by means of an RF bypass capacitor.

*Stage gain*

Since the second stage of the cascode amplifier is a grounded-grid amplifier, it can be said that under normal operating conditions this stage is stable. The second-stage voltage amplification is†

$$A_{v_2} = \frac{(\mu + 1)R_L}{r_p + R_L}. \tag{5-16}$$

Of the two stages this stage provides the greatest voltage amplification. However it is interesting to note that in terms of power amplification the second stage delivers less power gain than does the first stage. This can be seen from the expression for power gain in terms of voltage amplification. Thus,

$$PG = \frac{P_o}{P_{in}} = \frac{e_o^2}{R_L} \times \frac{R_{in}}{e_{in}^2} = A_v^2 \frac{R_{in}}{R_L}. \tag{5-17}$$

In the case of a grounded-grid amplifier it has been shown that $R_{in}/R_L = 1/A_v$. Therefore, Eq. (5–17) when applied to the second stage resolves itself into $PG = A_v$. On the other hand, the first stage (grounded-cathade) may have a $R_{in}/R_L$ ratio of 100 to 1 or more, which when multiplied by $A_v^2$ becomes a power gain considerably greater than the second-stage power gain. The large $R_{in}/R_L$ ratio is obtained because $R_L$ of the first stage is the input impedance of the second stage ($2/g_m$), which may be quite small.

A suitable tube for the second stage should have a large $g_m$ in order to provide a low input impedance and thus increase the power gain of the first stage. Of course, this also increases the power gain of the second stage. The tube used for the second stage should also have a low cathode-to-plate interelectrode capacitance in order to insure stability. If the tube is to be used in a stacked B arrangement, the maximum heater-to-cathode voltage rating should be high enough to prevent heater-to-cathode breakdown.

The cascode amplifier can be considered a special case of grounded-grid amplifier. This is clearly seen when a comparison is made between the equivalent circuits of Figs. 5-22 and 5-25. It is apparent that in the case

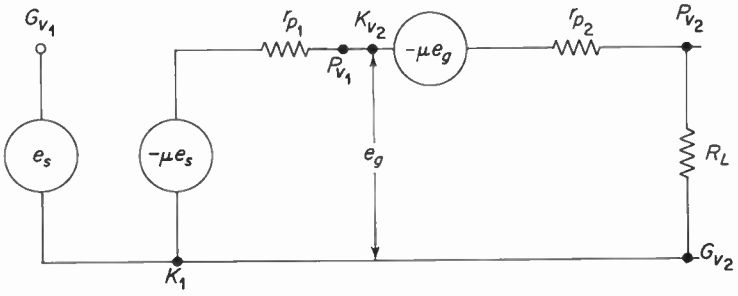



**Fig. 5-25.** The equivalent circuit of a cascode amplifier.

†See Appendix 5-2 for the derivation of equation 5-16.

of the cascode amplifier the driving source of the second stage is the grounded-cathode amplifier, whereas in the grounded-grid amplifier of Fig. 5-22 the driving source is the input signal. In both cases the driving source is loaded with the input impedance of the grounded-grid amplifier. This impedance, assuming $R_L$ is one half of $r_p$, is therefore approximately equal to $1.5/g_m$ [see Eq. (5–9)] and constitutes the plate load of the first stage of the cascode amplifier. This impedance is chiefly responsible for the low-voltage amplification developed by the first stage. The voltage amplification developed by the grounded-cathode first stage is

$$A_{v_1} \approx g_m\left(\frac{1.5}{g_m}\right) \approx 1.5. \tag{5–18}$$

The exact expression for the voltage amplification of the first stage is†

$$A_{v_1} = \frac{\mu}{\left(\dfrac{1}{1 + R_L/r_p}\right)\left(\dfrac{1}{1 - \{\mu/(\mu + 1)\}}\right) + 1}. \tag{5–19}$$

This expression assumes that both tubes have identical parameters.

Since the two stages are connected in cascade, the product of Eq. (5–16) and Eq. (5–19) will lead to an expression for the overall voltage amplification. Thus

$$\begin{aligned} A_{v_t} &= A_{v_1} \times A_{v_2} \\ &= \frac{R_L\mu(\mu + 1)}{r_p(2 + \mu) + R_L}. \end{aligned} \tag{5–20}$$

For example, if both tubes are identical and if $\mu$ is 50, $r_p$ is equal to 10,000 ohms, $R_L$ the load on the second stage is 4000 ohms, and if Eq. (5–19) is solved on the basis of the data above, it will be found that $A_{v_1}$ will be equal to 1.37. The application of the same data to Eq. (5–16) leads to a voltage amplification for $V_2$ of 14.5. The product of $A_{v_1}$ and $A_{v_2}$ is 19.8. If a pentode were employed as an RF amplifier and its $g_m$ was the same as that of the triodes used above, the voltage amplification achieved by this single stage alone would be

$$A_v = g_m R_L = 5000 \times 10^{-6} \times 4 \times 10^3 = 20.$$

Thus the overall voltage amplification of the cascode amplifier is almost the same as that of a pentode RF amplifier of equal $g_m$. It should be noted that the voltage amplification of the cascode amplifier is always less than that of pentode RF amplifier (assuming $r_p$ is very much greater than $R_L$). This can be seen by taking the ratio of the voltage amplification of the pentode to that of the cascode amplifier. This ratio is therefore

$$\frac{\text{voltage amplification of pentode}}{\text{voltage amplification of cascode}} = \frac{R_L}{r_p(1 + \mu)} + \frac{2 + \mu}{1 + \mu}. \tag{5–21}$$

†See Appendix 5-3 for the derivation of equation (5-19).

It can be seen from Eq. (5–21) that the second term is greater than unity and therefore the first term of the equation simply makes the ratio larger.

*Stability and noise*

The stability of the first stage of the cascode amplifier is provided by its inherently low voltage amplification. That is, the proportion of output voltage fed back by the interelectrode capacitance $C_{gp}$ is not changed, but since the ouptut signal voltage is low, owing to the reduced voltage amplification, the feedback voltage is insufficient to sustain oscillation. Neutralization is nevertheless employed (see Fig. 5-24) as a means of establishing a low noise figure for the amplifier. Furthermore, neutralization reduces the changes in input impedance, and therefore minimizes bandpass variations that accompany the application of AVC to the first stage.

Since both tubes of the cascode amplifier are triodes, partition noise is not a source of noise in the cascode amplifier. Thus the cascode amplifier will be less noisy than an equivalent pentode RF amplifier. Furthermore, owing to the low input impedance of the second stage, the noise contribution of the second stage to that of the first is negligible. As a matter of fact a number of authorities have pointed out that the second stage has a source conductance almost equal to the optimum source conductance. Since the second-stage noise contribution is negligible, the first stage determines the noise figure of the system. Thus, the expression for optimum noise figure of the cascode amplifier is the same as Eq. (5–15), which is the noise figure of a neutralized triode.

## 5-6. Transistor RF amplifiers

Small-signal transistor RF amplifiers employed at very high frequencies use class A operated common-base or common-emitter circuit configurations such as those illustrated in Fig. 5-26. Common-collector configurations are seldom used because of their inherently low power gain. In some respects these circuits are very similar to their low-frequency counterparts. It is apparent that regardless of the frequency of operation a temperature-stable operating point must be selected and maintained. Thus resistors $R_1$, $R_2$, and $R_3$ of Fig. 5-26 perform the function of setting the operating point and maintaining temperature stability. This may be somewhat more important for weak-signal VHF amplifiers,† since any small change in input and output impedance (due to a change in operating point as a result of a change in temperature) will be felt in the form of detuning, and thus loss of gain as well as harmonic distortion may be introduced. Of course any change in operating point will also affect the noise figure of the transistor.

†Typical value for stability factor at low frequency is 5, and that commonly found in VHF amplifiers is 2.5.

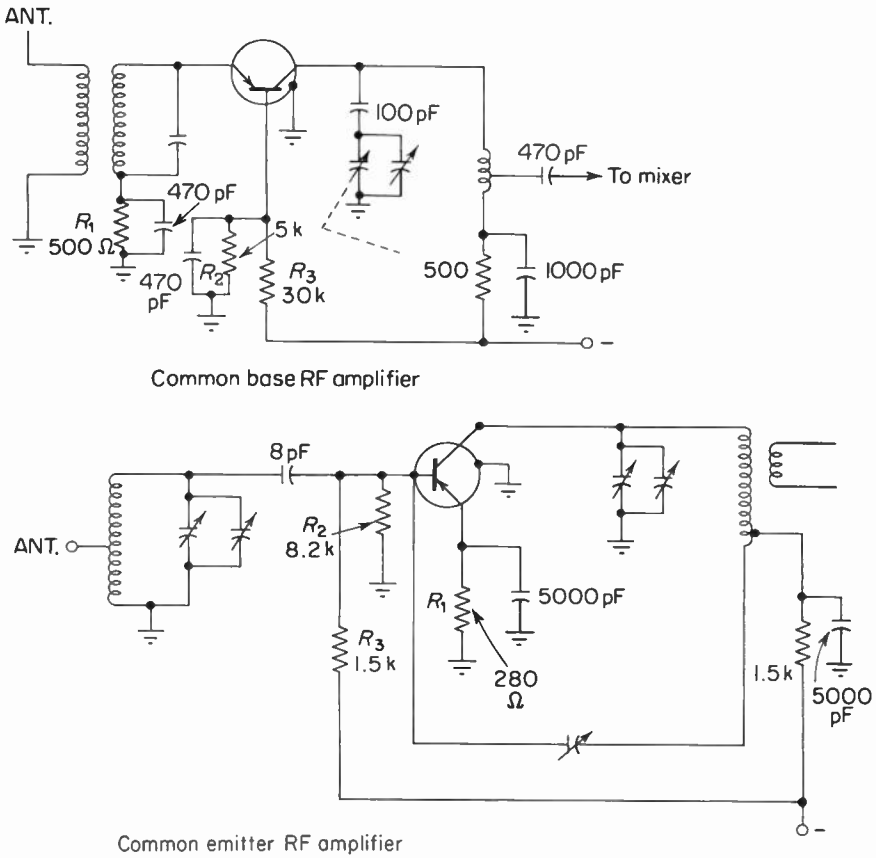



**Fig. 5-26.** Common base and common emitter RF transistor amplifiers.

Except for the physically small size of the input and output tuned circuits, the tuned-circuit considerations at high frequencies are about the same as at low frequency. These circuits are utilized to provide the required selectivity and proper impedance match or deliberate mismatch (for stability) between stages.

### *High-frequency requirements*

At frequencies of the order of 100 MHz, transistors and the circuits in which they are employed must meet certain requirements if optimum performance is to be obtained. First and foremost the transistor must be able to provide usable power amplification at very high frequencies. Thus, in order to be effective at FM broadcast frequencies the transistor must posess a gain-bandwidth product ($f_t$)† of 100 MHz or more. To meet this

†The gain-bandwidth product ($f_t$) is defined as the frequency at which $h_{fe}$ decreases to unity.

requirement the transistor must be fabricated in such a manner as to reduce the transit time of the current carriers passing through the base. In terms of the hybrid pi equivalent circuit this is the same as reducing the value of $C_{b'e}$. See Fig. 5-27. It is also apparent from the hybrid pi equivalent circuit that in order to obtain maximum power gain, the input resistance $r_{bb'}$ (due to the resistivity of the base) must be made as small as possible. To provide stable operation at very high frequencies, $C_{b'c}$, the interelectrode feedback capacitance between collector and base must also be made as small as possible.

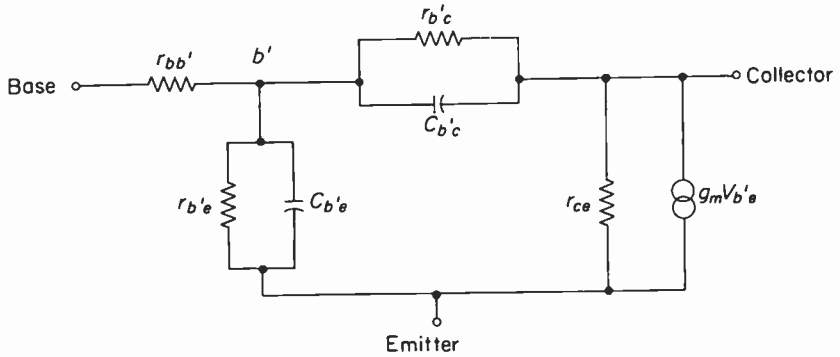



**Fig. 5-27.** Hybrid $p_i$ equivalent circuit for the common emitter high frequency transistor amplifier.

### *Types of transistors*

The history of transistor development is a history of the development of transistors that can be made to operate effectively at high frequencies. All of the seemingly endless variety of transistor fabrication processes that have been devised are basically methods by which transistors can be made to operate at higher and still higher frequencies.

One of the first attempts by transistor manufacturers to reduce emitter-to-collector transit time was by means of base thickness reduction. This approach resulted in such transistor types as the surface-barrier and the microalloy transistors. Figure 5-28(a) and (b) illustrates the construction of these types. These transistors can operate at frequencies as high as 50 MHz.

Further development resulted in a technique whereby an accelerating electrostatic field is set up in the base. This has the effect of reducing transit time by increasing the speed of transit of the current carriers. The accelerating drift field is obtained by doping the base so that it has a graded resistivity between emitter and collector. As is shown in Fig. 5-28(c), the resistivity of the base in the region of the emitter is low as the result of an increased impurity concentration. At the collector the base resistivity is maximum because of the scarcity of the impurity atoms. When sufficient voltage (a

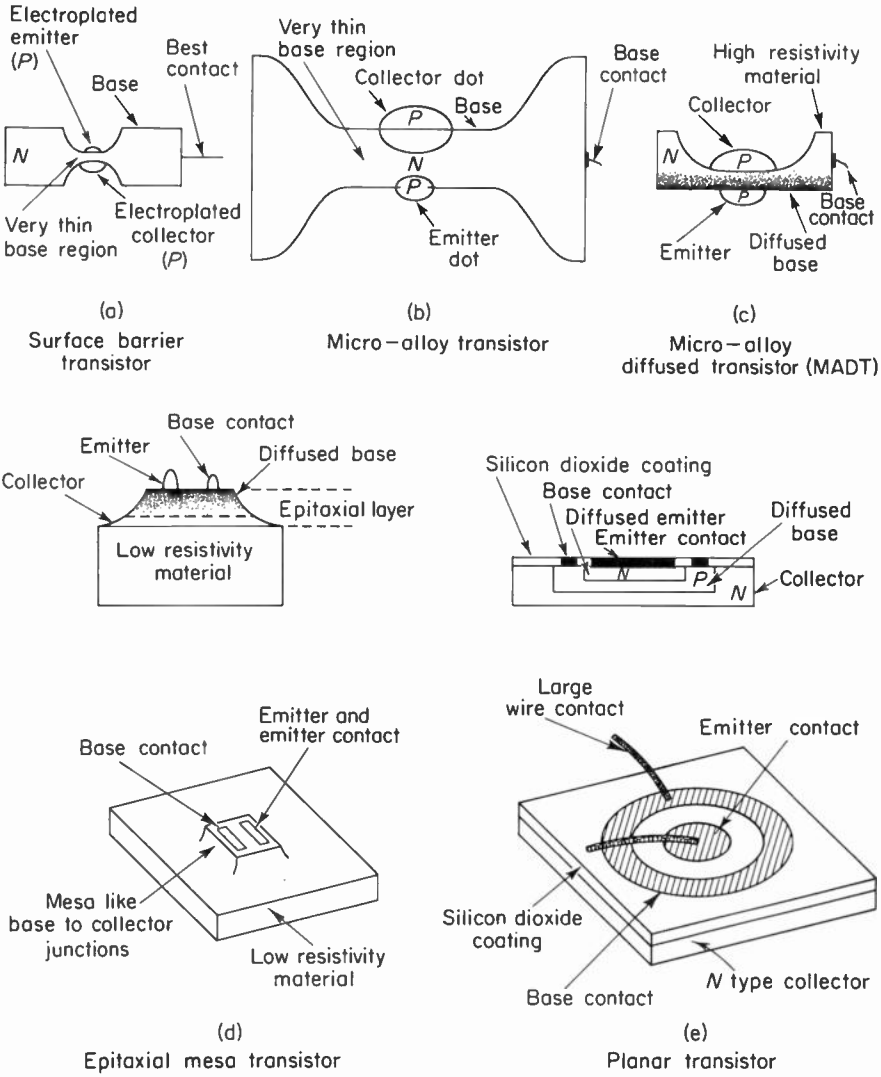



**Fig. 5-28.** The structure of five types of high frequency transistors.

few volts) is applied to the collector of a drift transistor, the transit time of the hole-current carriers is reduced to about one-fourth the time that would have been necessary if hole current were due to diffusion alone. In the micro-alloy diffused transistor (MADT), etching (to produce a thin base) and a drift field are combined to produce a quality VHF transistor. These transistors can be made to operate at frequencies as high as 200 MHz.

Other types of very-high-frequency transistors make use of etching, diffusion, and other specialized construction techniques in order to obtain

the desired high-frequency characteristics. See Fig. 5-28(d) and (e). These types, given such names as mesa, epitaxial, and planar transistors, show certain differences as well as similarities. Planar transistors differ from the microalloy and the microalloy diffused transistors in that the base and the emitter are deposited, by evaporation, on the collector, which is used as supporting structure for these elements. In the case of the MADT and other similar types the base is the supporting structure for the emitter and the collector. Both the mesa and the planar transistors employ diffused base construction in order to obtain a drift field. As the figures show, mesa and planar transistors differ in physical appearance. Needless to say, it is by these differences that they obtain their names.

Mesa as well as planar transistors may employ epitaxial layers in their construction. Epitaxial films are structures in which the film atoms have been aligned with those of the crystal so as to preserve the uninterrupted lattice structure of the crystal. These films can be controlled precisely and thus transistor performance can be improved, since greater control of transistor parameters can be achieved. Mesa and planar transistors exhibit gain-bandwidth products in excess of 200 MHz.

### *The field-effect transistor*

The field-effect transistor (FET) is a physically small, rugged semiconductor device whose power consumption is low. The FET is formed by the same techniques used in constructing planar transistors and integrated circuits. Like the triode vacuum tube, it is a three-element device: the cathode, grid, and plate of the vacuum tube correspond functionally to the source, gate, and drain of the FET. Schematic symbols of both the vacuum tube and the FET are shown in Fig. 5-29.

Under DC and low-frequency conditions, the FET has exceptionally high input impedance, of the order of $10^{15}$ ohms, and fairly high output resistance, of the order of 100,000 ohms or more. Thus, the FET can be expected to behave in a manner similar to a high-mu triode or pentode vacuum tube. This similarity is reinforced in that the interelectrode capaci-

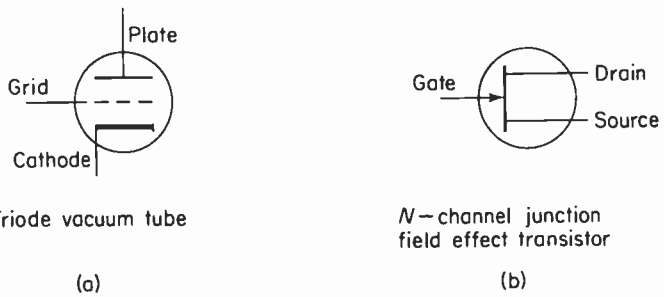



**Fig. 5-29.** Schematic symbols of a vacuum tube triode and a junction field effect transistor.

tance between input and output of the FET and the triode are of the same order of magnitude. At low frequencies the similarity between the vacuum tube and the FET becomes apparent, since both devices are voltage-activated and do not require power from the input signal source. It is, therefore, more meaningful to speak of voltage amplification ($A_v$) than power gain ($G$) when discussing the FET. The vacuum tube and the FET are also similar at high frequencies, since both exhibit a decrease in input impedance due to the Miller effect and other input loading phenomena. Thus, at high frequencies it becomes more meaningful to speak of power gain since both devices take power from the input signal source.

Other than the obvious physical differences between vacuum tubes and FET's, there are important electrical differences also. For example, the FET does not require filament power, and the drain-to-source voltage is of the order of 20 volts as compared to the usual 100 volts between plate and cathode of a vacuum tube.

The FET is found in two basic forms which can be made, by fabrication to operate in one of two modes. The two basic forms are the junction field-effect transistor (JFET) and the insulated-gate field-effect transistor (IGFET)—or, as it is often called, the metal-oxide semiconductor field-effect transistor (MOSFET). See Fig. 5-30 for cross-sectional and schematic representations. Notice that in both JFET and MOSFET the medium

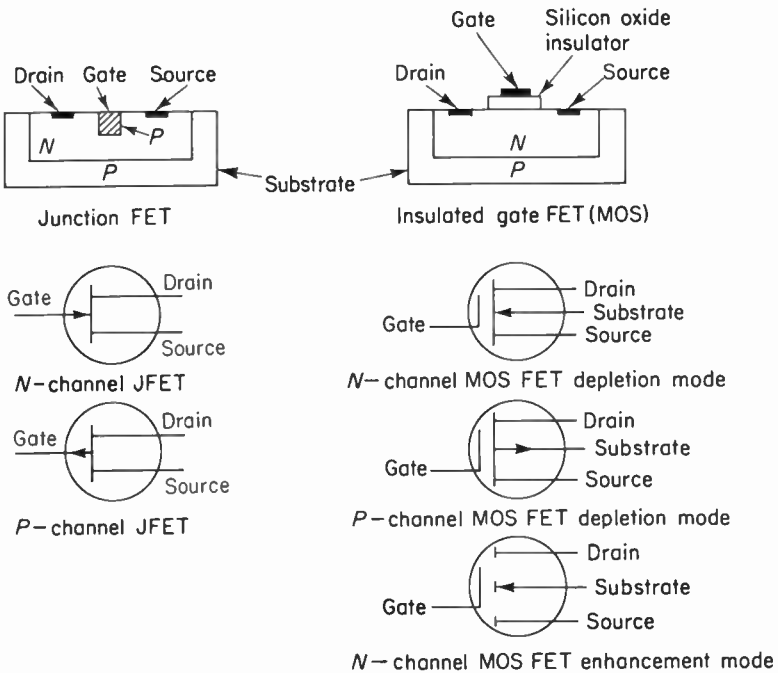



**Fig. 5-30.** Cross-sectional and schematic representations of the junction and MOS field effect transistors.

between source and drain is either N-type or P-type semiconductor material. Thus, an FET is usually referred to as either an N-channel or P-channel FET. In the following discussion N-channel FET's will be assumed, since this type if most analogous to the vacuum tube and since its higher ratio of transconductance to stray capacitance makes for a better high-frequency transistor.

The junction field-effect transistor, like the vacuum tube, requires a reverse bias between its input terminals. This maintains a high input impedance and thus prevents input circuit loading. The MOSFET is structurally different from the JFET, and is so arranged that the gate is completely insulated from the source-to-drain channel. Therefore, the input circuit of the device is not polarity-sensitive. Thus, bias is used only to set the operating conditions of the MOSFET and may be either positive or negative with respect to the source, depending solely upon the desired operating conditions. For the MOSFET bias is not required to maintain a high input impedance.

The two basic modes of FET operation are called depletion and enhancement. The depletion mode of operation refers to the condition where an increase in input-circuit reverse bias causes a decrease in output current. The output volt-ampere characteristics of an FET that can be used in the depletion mode is shown in Fig. 5-31. Notice that the drain current is maximum with zero gate voltage and decreases with an increase in reverse bias. Notice also that the appearance of the characteristic curve is similar to that of a pentode vacuum tube.

The enhancement type of FET is structurally different from the depletion FET and is cut off (drain current is approximately zero) at zero bias. Drain current in an enhancement FET does not increase until the bias increases in a positive direction above zero. Typical output characteristics are shown in Fig. 5-31(b). These characteristic curves are similar to that of a pentode vacuum tube except that cutoff occurs at zero gate voltage.

The small-signal equivalent circuit of a MOSFET is shown in Fig. 5-32. This simple field equivalent circuit is valid up to about 100 MHz, above which distributed inductance of the leads and of the case would have to be included. The significance of the equivalent circuit elements is as follows:

$r_{gs}$ and $r_{gd}$ represent the leakage resistance between gate and source, and gate and drain, respectively. Typical value for the parallel combination of these resistances is $10^{15}$ ohms.

$C_c$ and $r_c$ represent the distributed network between the gate structure and the channel. Typical values are 3 pF and 100 ohms, respectively. At high frequencies the transconductance of the FET is controlled by the voltage $e_c$ which appears across $C_c$. As the frequency of operation is increased the reactance of $C_c$ decreases with the result that the voltage $e_c$ also decreases. Thus the time constant $r_cC_c$ controls the gain at high frequencies. At low frequencies $C_c$ becomes an open circuit and the voltage $e_c$ becomes $e_{gs}$.

$C_{gs}$ is the distributed case and interlead capacitance between the gate and source. Typical value for this capacitance is about 2 pF.

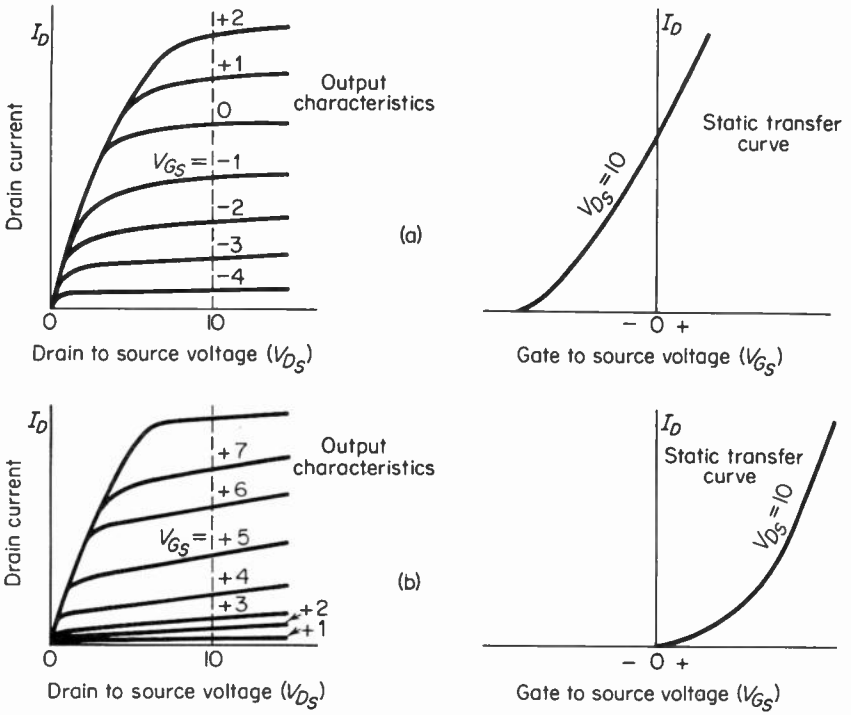



**Fig. 5-31.** Output characteristics and their static transfer curves for (a) depletion mode (b) enhancement mode of operation.

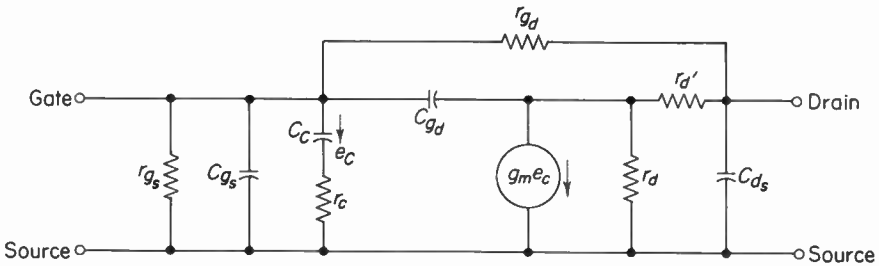



**Fig. 5-32.** The simplified equivalent circuit of a field effect transistor.

$C_{gd}$ is the distributed feedback capacitance from drain to gate. A representative value for this capacitance is 1 pF.

$r_d$ is the output resistance of the FET; its value is determined by the slope of the output characteristics of the FET and the operating point of the device. Depending upon the operating point of the FET, the value of $r_d$ may range from a few hundred ohms to several hundred megohms.

$g_m$ is the transconductance of the FET. For low frequencies it is determined from the slope of the $I_d/E_g$ transfer characteristic. Typical value of $g_m$ ranges from 3000 to 10,000 micromhos.

In summary, it is evident that for FM receiver applications the microalloy transistor is suitable for IF and limiter stages. The MADT, mesa, planar, and FET types are more useful as mixer and RF amplifiers.

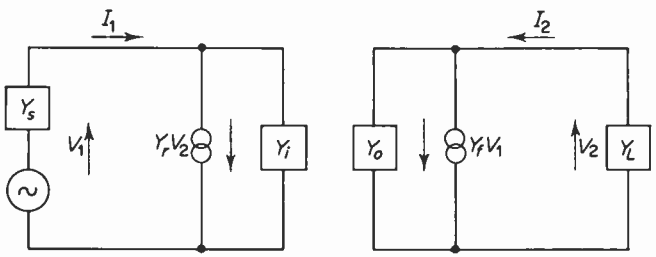
Since the semiconductor revolution is far from over, no one type has become the universal favorite. No doubt in time new semiconductor contrivances will supplant the present formidable arsenal.

### *The "Y"-parameter equivalent circuit*

Prediction of transistor small-signal behavior is not easily obtained by means of conventional T or hybrid $\pi$ equivalent circuits when the transistor is operated at frequencies approaching alpha cutoff or above. The reason is primarily that the measurements needed for usable equivalent circuits are very difficult to make. Furthermore, since almost all of the parameters are complex and since the equivalent circuits themselves are complicated, calculations become very involved. The use of these parameters (T or $\pi$), however, does yield some insight into the physical processes of the transistor operated at high frequencies.

An equivalent circuit, useful over a relatively narrow bandpass, that simplifies parameter measurements as well as circuit calculations is one based on admittance parameters. It is called a *Y*-parameter equivalent circuit or a short-circuit parameter equivalent circuit. Figure 5-33 illustrates the *Y*-parameter equivalent circuit of a transistor amplifier. Of course, each of the admittances shown can be resolved into component conductances ($G$) and susceptances ($B$). Thus, for example, $Y_o = G_o + jB_o$. Note that a plus $jB$ represents a capacitive reactance and a minus $jB$ represents an inductive reactance.

The equivalent circuit is applicable to any circuit configuration. Parameter symbols intended for use with the common-base amplifier would be followed by a subscript $b$; those intended for the common emitter would




$Y_s$ source admittance
$Y_i$ input admittance ($I_1/V_1$) with output short circuited
$Y_r$ reverse transfer admittance ($I_1/V_2$) with input short circuited
$Y_f$ forward transfer admittance ($I_2/V_1$) with output short circuited
$Y_o$ output admittance ($I_2/V_2$) with input short circuited
$Y_L$ load admittance

**Fig. 5-33.** Transistor equivalent circuit based on *Y* parameters.

be followed by a subscript *e*. Thus, for example, $g_{te}$ would represent the input conductance of the common-emitter amplifier.

The parameters are determined by the quiescent operating conditions ($I_e$ and $V_c$) and the frequency at which the transistor is to function. Consequently, for any design it is necessary that measurements or conversion calculations be performed in order to obtain the appropriate data at the intended circuit conditions. Curves illustrating *Y*-parameter variations due to changes in $I_c$ can be obtained from the transistor manufacturer. Figure 5-34 shows typical curves for a transistor used in the common-base and common-emitter modes of operation.

The design of an RF amplifier must be based upon the needs of the

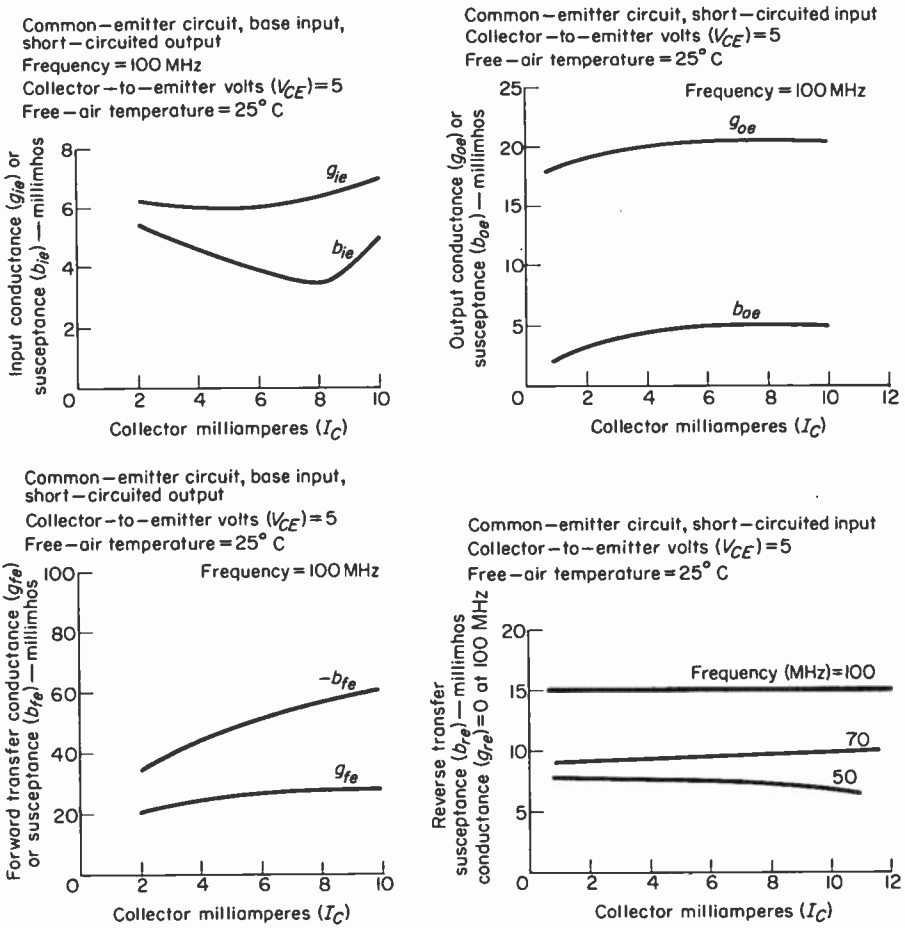



**Fig. 5-34.** Variation of the small signal *Y* parameters of a transistor used as a common emitter amplifier.

designer and the characteristics and limitations imposed by the amplifying device. Thus, for example, to provide maximum power gain the source admittance and the input admittance of the next stage must be conjugate-matched† to the respective input and output admittance of the amplifier. To provide such a match the designer must know both the input and output admittance of the transistor. But since these depend upon the quiescent operating point, and since noise is a function of the $Q$ point, it is also necessary to select the operating point for low-noise operation. On the basis of the data above, and the other data required for the construction of a $Y$-parameter equivalent circuit, the power gain as well as the stability of the small-signal VHF amplifier can then be determined.

### *Internal feedback, and stability*

Under most conditions of VHF operation the internal feedback associated with the common-base amplifier is regenerative, whereas the common emitter displays inherent degenerative feedback.‡ Therefore, the common-base amplifier may be subject to instability, and uniformity of bandwidth, power gain, input and output impedances, and so on may in general be difficult to maintain if for any reason the transistor is replaced or if for any reason the circuit is to be duplicated. If the operating point and the load and source impedances are maintained for both types of circuit configurations it is possible, as a result of the internal feedback, for the common-base to develop greater power gain than can the common-emitter amplifier. As a result the common-base amplifier can often employ transistors of lower $f_t$ rating than would be possible if a common-emitter circuit configuration were used.

### *Neutralization and unilateralization*

Stability of the amplifier depends upon the internal feedback impedance associated with a given circuit configuration, the frequency of operation, the overall power gain, the nature of the load impedance, and the quiescent point of operation. To eliminate or reduce the tendency for the circuit to oscillate, one of three methods may be employed: neutralization, unilateralization, or circuit mismatch. Neutralization is a general term used to describe any method whereby an undesired effect is canceled. In terms of stability and input circuit loading, due to feedback, neutralization is usually designed to cancel the negative-resistance component of input resistance. Unilateralization refers to a method whereby a device that is bilateral (operates in two directions) is converted into a one-way device (unilateral). Thus, unilateralization is able to cancel both the resistive and reactive effects of

†The condition where the admittance of a generator and its load are equal but opposite in phase. Thus the reactances are tuned out and maximum power is transferred.

‡According to R. F. Shea [*Transistor Circuit Engineering* (New York: John Wiley & Sons, Inc., 1957), p. 190] the common emitter is potentially unstable only up to a fraction of alpha cutoff. At *higher frequencies* the common emitter is *unconditionally* stable.

feedback. In the case of an amplifier this means that none of the output signal is returned to the input.

*Neutralization methods*

Transistor amplifiers that require neutralization use essentially the same methods that are used to neutralize vacuum-tube amplifiers. An important difference is that the internal feedback in transistors is through a network consisting of both resistive and reactive components ($r_{b'c}$ and $C_{b'c}$ of Fig. 5-27), whereas the internal-feedback network associated with the vacuum tube is purely reactive ($C_{gp}$). Thus for complete neutralization (unilateralization) the transistor external neutralizing network may be more complicated than that required by a comparable vacuum-tube amplifier. For example, if complete neutralization is required of a common-emitter amplifier, the circuit may appear as shown in Fig. 5-35. Compare this with a similar

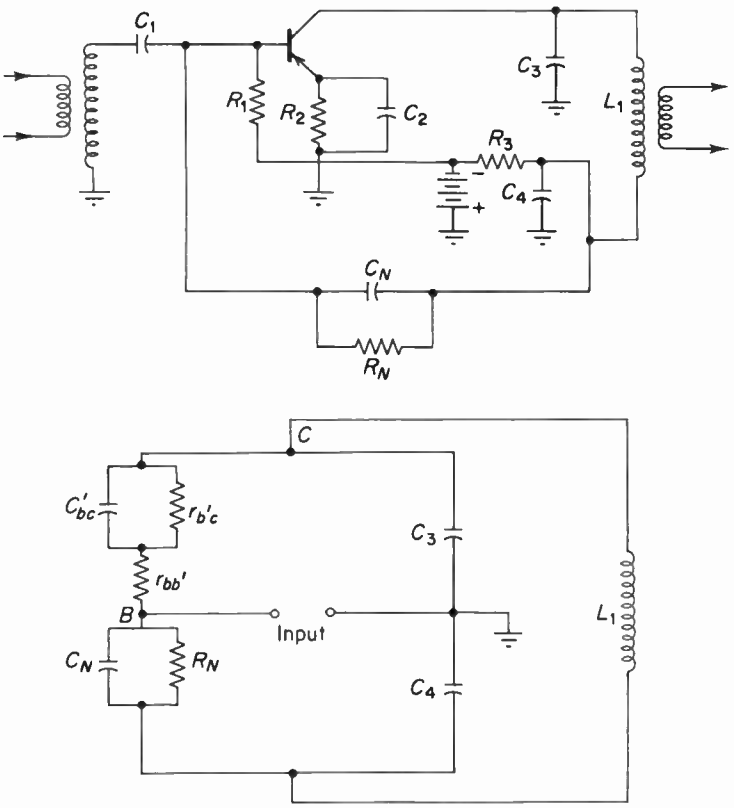



**Fig. 5-35.** Neutralized common emitter amplifier and its equivalent bridge circuit.

circuit used to neutralize a vacuum-tube amplifier [Fig. 5-20(a)]. Both circuits use bridge neutralization, but in the transistor case the neutralization capacitor ($C_n$) is shunted by a resistor ($R_n$), which is necessary to achieve perfect bridge balance. Since $r_{b'c}$ is usually very large (measured in megohms), $R_n$ is often not used in balancing the bridge. Of course, technically speaking this means that the amplifier is not unilateralized. The balance equations for both the transistor and vacuum-tube bridge neutralization will be found to be similar to Eq. (5–7). Thus,

$$\frac{C_{b'c}}{C_n} = \frac{C_3}{C_4}.$$

Common-base RF amplifiers are also neutralized by means of bridge-balancing techniques. A typical example is shown in Fig. 5-36(a). In order to show the bridge structure of this method of neutralization, we resort to the T equivalent of the transistor. Since the circuit shown in Fig. 5-36 is obviously a form of Wheatstone bridge, the condition of balance occurs when

$$C_n R_n = C_c r_{bb'}. \tag{5–22}$$

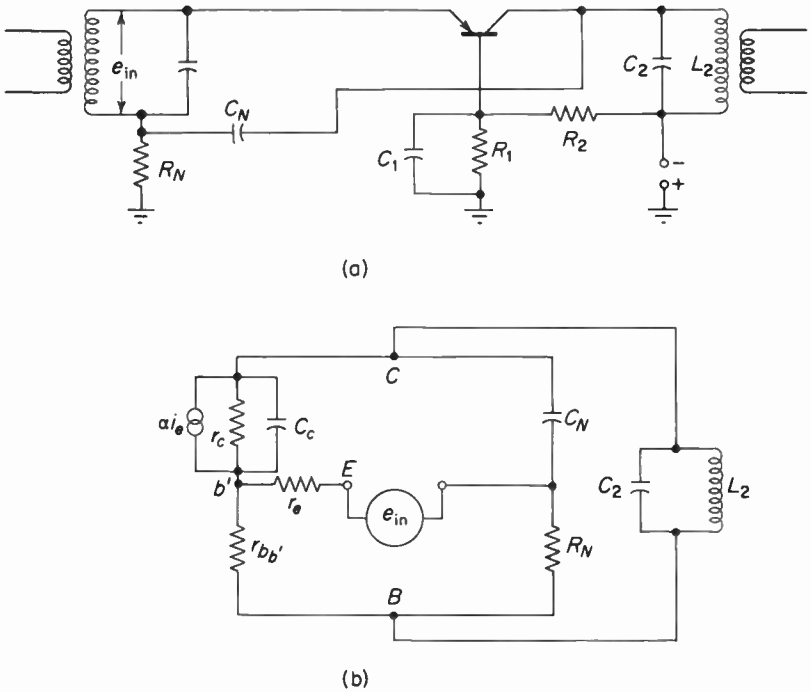



**Fig. 5-36.** Neutralized common base RF amplifier and its equivalent bridge circuit.

To completely unilateralize this circuit $C_n$ may be shunted with a resistor of proper magnitude.

### *Maximum stable gain and mismatch*

As for vacuum-tube amplifiers, there exists for transistor amplifiers a condition of maximum stable gain. That is, if the gain of a transistor amplifier exceeds this value, the amplifier will tend to oscillate. According to Hunter,† the maximum stable transducer power gain for an unneutralized common-emitter amplifier must be less than

$$\frac{P_L}{P_s} < \frac{g_m}{1.25(\omega)C_{ob}}, \tag{5-23}$$

where $\omega$ is $2\pi f$, $P_L$ is the power in the load, $P_s$ is the power delivered by the source, $g_m$ is the ratio $i_c/V_{b'e}$, and $C_{ob}$ is the sum of the collector-to-base interelectrode capacitance and the distributed lead capacitance from collector to base.

It is obvious from the last expression that if the power gain of a transistor amplifier is sufficiently reduced, the circuit will not oscillate. One of the simplest methods of reducing the power gain of an amplifier is to provide a mismatch between source and load. This method of insuring amplifier stability will also result in a greater independence of the input impedance from the output impedance.

### *Y-parameter circuit design*

To design a high-frequency RF amplifier we must know such circuit conditions as input admittance, output admittance, power gain, and circuit stability. The *Y*-parameter equivalent circuit can be used to calculate all of these conditions.

The input admittance of a transistor amplifier is the admittance seen by an input generator excluding its own internal source admittance. This quantity must be known if optimum noise figure is to be obtained, or for that matter if proper matching is to be secured between input source and the input circuit of the transistor.

In the following derivations a common-base amplifier will be assumed, although the equations may also be used for the common-emitter amplifier. The designations $Y_{ib}$, $Y_{rb}$, $Y_{fb}$, $Y_{ob}$ are the transistor parameters usually supplied by the manufacturer and are defined in Fig. 5-33. These parameters may be expanded into their real and complex terms. Thus, the complex terms are designated $B_{ib}$, $B_{rb}$, $B_{fb}$, $B_{ob}$, and the real terms are designated $G_{ib}$, $G_{rb}$, $G_{fb}$, and $G_{ob}$, respectively. $Y_{\text{in}}$ and $Y_{\text{out}}$ refer to the input and output

†L. P. Hunter, *Handbook of Semiconductor Electronics*, 2nd ed. (New York: McGraw-Hill Book Company, 1956), pp. 12–48.

admittances "seen" by the source and load, respectively. $V_1$ and $V_2$ designate the input and output terminal voltages of the amplifier, and the $I_1$ and $I_2$ represent the input and output currents of the amplifier. The source admittance is referred to as $Y_s$, and the load admittance is designated $Y_L$. From Fig. 5-33 it can be seen that the input admittance of the common-base amplifier with the source removed must be

$$Y_{in} = \frac{I_1}{V_1},$$

$$I_1 = Y_{ib}V_1 + Y_{rb}V_2,$$

and

$$V_2 = -\frac{Y_{fb}V_1}{Y_{ob} + Y_L}.$$

Substituting,

$$Y_{in} = \frac{I_1}{V_1} = \frac{V_1 Y_{ib} + Y_{rb}\left(-\dfrac{Y_{fb}V_1}{Y_{ob} + Y_L}\right)}{V_1}$$

$$= Y_{ib} - \frac{Y_{rb}Y_{fb}}{Y_{ob} + Y_L}. \tag{5-24}$$

Equation (5–24) indicates that the input impedance depends upon the output load conditions. It also indicates that under conditions where the first term is smaller than the second term the input admittance becomes negative, which may result in oscillation.

The output admittance of a transistor amplifier is the admittance seen by the load (but not including the load) looking back into the circuit of Fig. 5-33. Thus,

$$Y_{out} = \frac{I_2}{V_2},$$

$$I_2 = Y_{ob}V_2 + Y_{fb}V_1,$$

$$V_1 = -\frac{Y_{rb}V_2}{Y_{ib} + Y_s},$$

$$Y_{out} = \frac{Y_{ob}V_2 + Y_{fb}V_1}{V_2} = \frac{Y_{ob}V_2 + Y_{fb}\left(-\dfrac{Y_{rb}V_2}{Y_s + Y_{ib}}\right)}{V_2}$$

$$= Y_{ob} - \frac{Y_{fb}Y_{rb}}{Y_s + Y_{ib}}. \tag{5-25}$$

The output admittance of a single-stage amplifier must be known if matching (or mismatching) between two cascaded stages is to be obtained. If maximum power is to be transferred (and thereby maximum available power gain obtained), $Y_L$ must be conjugate-matched to $Y_{ob}$. The admittance $Y_L$ is made up of the input admittance of the second stage and the tuned circuit associated with the first stage.

If both input and output circuits are tuned to resonance, both circuits become pure conductances. Thus Eq. (5–24) reduces to

$$Y_{in} = Y_{ib} - \frac{Y_{rb}Y_{fb}}{(G_{ob} + G_L) + j(B_{ob} + B_L)}$$

$$= Y_{ib} - \frac{Y_{fb}Y_{rb}}{G_{ob} + G_L} \tag{5–26}$$

and by similar reasoning Eq. (5–25) becomes

$$Y_{out} = Y_{ob} - \frac{Y_{fb}Y_{rb}}{G_s + G_{ib}}. \tag{5–27}$$

The power gain of a tuned RF amplifier at resonance may be found as follows.

$$PG = \frac{P_o}{P_{in}} = \frac{V_2^2 G_L}{V_1^2 G_{in}} = A_v^2 \frac{G_L}{G_{in}}.$$

But from Eq. (5–24)

$$V_2 = -\frac{Y_{fb}V_1}{Y_{ob} + Y_L},$$

therefore,

$$\frac{V_2}{V_1} = -\frac{Y_{fb}V_1}{Y_{ob} + Y_L} \times \frac{1}{V_1}.$$

The voltage amplification $V_2/V_1$ is equal to $-Y_{fb}/(Y_{ob} + Y_L)$, but since the power amplification is $(A_v)^2 G_L/G_{in}$, it may be rewritten

$$PG = \left(\frac{|Y_{fb}|^2}{|Y_{ob} + Y_L|^2}\right)\frac{G_L}{G_{in}}. \tag{5–28}$$

In a tuned amplifier this reduces to

$$PG = \left(\frac{|Y_{fb}|^2}{|G_{ob} + G_L|^2}\right)\frac{G_L}{G_{in}}.$$

According to Linvill and Gibbons, the criterion of stability of a two-port network in terms of $Y$ parameters of a common-base amplifier is given by the following expression:

$$C = \frac{|Y_{fb}Y_{rb}|}{2\,\mathrm{Re}\,(Y_{ib})\,\mathrm{Re}\,(Y_{ob}) - \mathrm{Re}\,(Y_{fb}Y_{rb})} \tag{5–29}$$

where Re means "the real part of." The device is stable if $C < 1$. That is, under these conditions all input and output terminations will not introduce oscillation. This is a test for stability with both input and output terminals of the transistor open-circuited; thus it represents a worst-case condition. If $C$ is greater than $|1|$, the circuit is potentially unstable—that is, under certain load and source admittances the circuit will oscillate.

It is often desirable to be able to determine whether an actual circuit is stable. This may be done by employing the stability criterion established

by Stern, which takes into account actual source and load admittances connected to the transistor. This stability factor ($K$) in terms of a common-base amplifier may be written

$$K = \frac{|Y_{fb}Y_{rb}| + \text{Re}(Y_{fb}Y_{rb})}{2(G_{ib} + G_s)(G_{ob} + G_L)}, \tag{5-29a}$$

where Re means "the real part of." If $K$ is less than $|1|$, the circuit will be stable; if $K$ is larger than $|1|$, the circuit will be unstable. The advantage of $K$ is that it provides a stability evaluation for a specific circuit. The stability factor $C$ is able to predict the potential stability of a transistor with an open-circuited source and load.

Of course, $Y$ parameters may be used for vacuum tubes and FET's as well as transistors for analysis and circuit design.

A numerical example will illustrate how the $Y$ parameters may be used for amplifier design. Assume that a transistor operating at 100 MHz and at a given $I_c$ has the following $Y$ parameters for common-base and common-emitter configurations. All values are in millimhos.

$$\begin{aligned} Y_{ib} &= 23.7 - j13.05, & Y_{ie} &= 9.2 + j11, \\ Y_{rb} &= -0.5 - j1.85, & Y_{re} &= -0.035 - j0.84, \\ Y_{fb} &= -14.4 + j23.3, & Y_{fe} &= 14 - j21, \\ Y_{ob} &= 0.5 + j2.71, & Y_{oe} &= 0.5 + j2.71. \end{aligned}$$

If it is also assumed that the source admittance of a common-base amplifier is

$$Y_s = 80 + j15.85$$

and that the load admittance is

$$Y_L = 4.5 - j2.57,$$

then, using Eq. (5-24), the input admittance of the amplifier will be

$$\begin{aligned} Y_{\text{in}} &= (23.7 - j13.05) - \frac{(27.4\angle 121.7)(1.92\angle 255.3)}{0.5 + 4.5} \\ &= 13.7 - j15.85. \end{aligned}$$

The output admittance of the amplifier, determined from Eq. (5–25), is

$$\begin{aligned} Y_{\text{out}} &= (0.5 + j2.71) - \frac{(27.4\angle 121.7)(1.92\angle 255.3)}{80 + 23.7} \\ &= 0.01 + j2.57. \end{aligned}$$

Linvill's critical stability factor is found from Eq. (5–29) to be

$$C = \frac{|(27.4)(1.92)|}{2(23.7)(0.5) - (50)} = |1.93|.$$

Thus, the transistor is potentially unstable, since $C > 1$. If we calculate Stern's stability factor to see the effect of the generator and load conductances, we find that

$$K = \frac{|(27.4)(1.92)| + 50}{2(23.7 + 80)(0.5 + 4.5)} = |0.099|,$$

which indicates that the amplifier is stable under loaded conditions. If the power gain of the amplifier is now calculated, it will be found to be

$$PG = \frac{|27.4|^2}{|0.5 + 4.5|^2} \times \frac{4.5}{13.7} = 9.85 \text{ or } 9.9 \text{ db}.$$

Repeating the above calculations for a common-emitter amplifier, we obtain the following results for a source and a load admittance of

$$Y_s = 27 - j16.58 \quad \text{and} \quad Y_L = 1.51 - j3.036.$$

The input admittance becomes

$$Y_{in} = (9.2 + j11) - \frac{(25.2/56.3)(0.84/269.8)}{0.5 + 1.51} = 0.9 + j16.58.$$

The output admittance is

$$Y_{out} = (0.5 + j2.71) - \frac{(25.2/56.3)(0.84/269.8)}{27 + 9.2} = 0.014 + j3.03$$

and

$$C = \frac{|(25.2)(0.84)|}{2(9.2)(0.5) - (17.8)} = |2.46|.$$

Thus, the transistor in this circuit configuration is also potentially unstable. When $K$ is calculated, it becomes

$$K = \frac{|(25.2)(0.84)| + 17.8}{2(9.2 + 27)(0.5 + 1.51)} = |0.266|.$$

Thus, the common-emitter amplifier is stable under loaded conditions. The power gain of the amplifier will be

$$PG = \frac{|25.2|^2}{|0.5 + 1.51|^2} \times \frac{1.51}{0.9} = 263 \text{ or } 24.2 \text{ db}.$$

In each of the amplifiers discussed above, stability was the result of a mismatch between the input and output circuits and the transistor.

To provide the necessary bandpass for the amplifier, taking into consideration the input and output admittances as well as the source and load admittances, the input and output tuned circuits must act as impedance-matching transformers. An example of how this is done will be discussed in a later section.

## 5-7. Input circuit considerations

RF amplifier input circuit has at least five important functions:

1. It provides a match between the antenna transmission line and the input impedance of the amplifier to minimize line reflections and therefore reduce signal loss.

2. It converts the transmission-line resistance to the optimum source resistance if optimum noise figure is desired.

3. It provides minimum insertion loss.

4. The tuned input circuit of an RF amplifier is designed to provide the required bandpass.

5. The input circuit may be designed to convert balanced-circuit conditions into unbalanced-circuit conditions as a means of reducing transmission-line impulse-noise pickup.

The first two of these functions may not be compatible, since in a given circuit the conditions for match and the conditions for minimum noise figure may not be the same. In general, the difference is small and is not of much moment, except in the case of a specifically designed weak-signal receiver. Thus, in most cases the input circuit will be designed to achieve a compromise between the conditions of maximum power transfer and best noise figure. A discussion of optimum source resistance will be found in Sec. 3-6.

Insertion loss is defined as the power loss introduced by a network placed between the source and a load. In decibels this is

$$\text{db} = 10 \log \frac{P_o}{P_L}, \tag{5-30}$$

where $P_L$ is the power delivered to the load with the network, and $P_o$ is the power delivered to the load without the network.

This loss that may be contributed by the transmission line feeding the RF amplifier as well as the input transformer tends to degrade the $S/N$ ratio. This occurs because the loss tends to reduce the signal and, at the same time, contributes noise to the system. Insertion loss due to the transmission line is generally disregarded as negligible except in the case of low-noise and weak-signal applications.

The insertion loss introduced by the input circuit is most easily determined as a function of unloaded and loaded tuned-circuit $Q$. Thus,

$$L_i = \left(\frac{Q_u}{Q_u - Q_L}\right)^2, \tag{5-31}$$

where $L_i$ is the insertion loss [the power ratio of Eq. (5-30)], $Q_u$ is the unloaded $Q$ of the tuned circuit, and $Q_L$ is the loaded $Q$ of the tuned circuit. It can be seen from Eq. (5-31) that the insertion loss can be kept to a minimum by making the ratio $Q_u/Q_L$ as large as possible.

### *Input circuits*

At very high frequencies the input impedance of vacuum-tube and transistor amplifiers is very low. If a tuned circuit is placed across the input of these devices, the operating $Q$ of the tuned circuit will be greatly reduced. This, in turn, will result in increased bandpass and poor selectivity. To

maintain adequate selectivity and to maintain an impedance match, tapped transformers such as those in Fig. 5-6 are used. The tapped transformers are autotransformers, which transform the low input impedance of the amplifier into a larger effective impedance, equivalently shunting the entire tuned circuit. By means of the impedance-changing properties of the transformer the selectivity and bandpass can be made adequate for proper operation of the circuit.

The following example assumes that the unloaded tuned circuit has no losses (infinite $Q$) and that the coupling between turns is perfect ($k = 1$). If the tuned circuit is at resonance and is shunted by a 5000-ohm load ($R_1$) (due to the input resistance of the amplifier), and if the reactance of the coil is 50 ohms, the loaded $Q$ of the circuit will be

$$Q_L = \frac{R_1}{X_L} = \frac{5000}{50} = 100.$$

See Fig. 5-37(a). At the resonant frequency of 100 MHz the bandpass ($B$) is

$$B = \frac{f_r}{Q_L} = \frac{100}{100} = 1 \text{ MHz}.$$

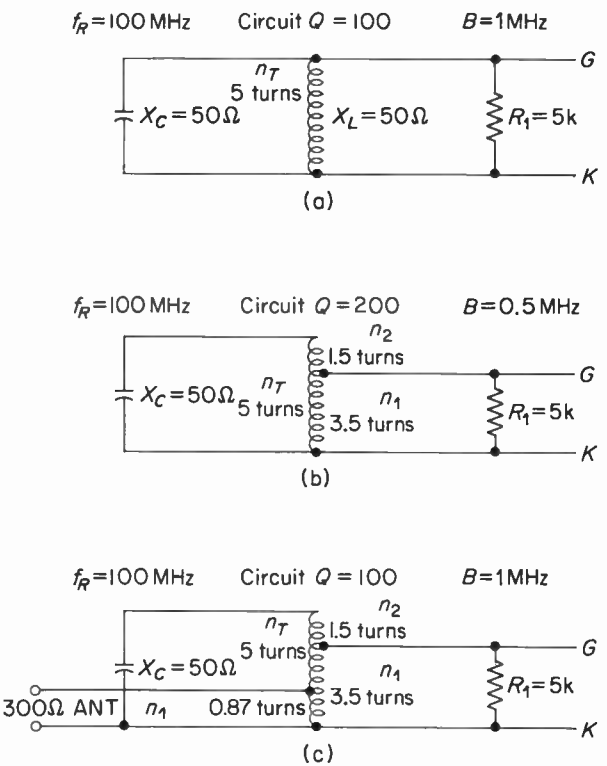



**Fig. 5-37.** The tapped transformer as a means for providing a power match for a given bandpass.

If this circuit is used as an antenna input transformer, its $Q_L$ will be reduced to one-half (50) when the antenna is connected (matched conditions assumed). This means that the bandpass will increase to 2 MHz.

If, as often happens in tunable RF amplifiers, the antenna connected bandpass is to be 1 MHz, the unloaded (only the antenna disconnected) $Q_L$ must be 200. See Fig. 5-37(b). To achieve a loaded $Q_L$ (both antenna and $R_1$ connected) of 100, we tap the resonant tuned circuit, converting it into an autotransformer. This results in a higher effective shunt resistance across the coil. The magnitude of the effective *shunt* resistance is found by the following expression:

$$R_{\text{eff}} = n^2 R_{\text{load}}, \tag{5-32}$$

where $n$ is the turns ratio $(n_1 + n_2)/n_1$. If the turns ratio of the circuit shown in Fig. 5-33(b) is 1.41, the effective shunt resistance due to the input resistance of the amplifier and appearing across the terminals of the transformer will be $(1.41)^2 \times 5000 = 10{,}000$ ohms. Since, 10,000 ohms is effectively shunting the tuned circuit, the $Q_{\text{eff}}$ of the circuit without the antenna connected will be 200. To find the proper tap in terms of the required $Q_{\text{eff}}$, coil reactance, and load resistance, the following formula may be used:

$$n_1 = n_t \sqrt{\frac{R_1}{Q_{\text{eff}} X_L}} = \frac{n_t}{n} \tag{5-33}$$

where $Q_{\text{eff}}$ is the $Q$ of the circuit due to $R_{\text{eff}}$, $n_t$ represents the total number of turns, and $n_1$ is the number of turns across which $R_1$ is placed. Assuming $n_t$ is five turns and solving Eq. (5–33),

$$n_1 = 5\sqrt{\frac{5\text{ k}}{200 \times 50}}$$

$$= 5\sqrt{0.5} = 3.5 \text{ turns},$$

$n_1$ is found to be 3.5 turns from the bottom of the coil.

When the antenna (a 300-ohm folded dipole) is connected as shown in Fig. 5-37(c), it must also be connected to a tap if a proper match is to be obtained. Under conditions of match the effective shunt resistance across the tuned circuit due to $R_1$ must equal the effective shunt resistance across the tuned circuit introduced by the antenna resistance. Therefore, the total effective shunt resistance across the entire coil is 5000 ohms. Under these conditions the loaded $Q$ with the antenna connected will be

$$Q_t = \frac{R_{\text{eff}}}{X_L} = \frac{5000}{50} = 100.$$

Consequently, the bandpass of the antenna-connected tuned circuit becomes

$$B = \frac{f_r}{Q_t} = \frac{100}{100} = 1 \text{ MHz},$$

which is the desired bandpass for a tunable RF amplifier.

The proper tapping point may be found by Eq. (5–33), provided that in place of $R_1$ the antenna resistance is used The $Q_{eff}$ is assumed to be 200 since it is determined with $R_1$ disconnected. For this example this calls for a tap 0.865 turns from the bottom of the coil.

Since the example above is based on the assumption that the coil is lossless, the results are to be considered as first steps in determining the correct tapping point.

### *Capacitive matching*

As the frequency of operation increases, the size of the resonating inductance of a tuned circuit decreases. At frequencies of 100 MHz values of inductance as low as 0.5 microhenry or less are required to produce resonance. In practice, tapping of the inductance to provide power match becomes very difficult when the tap is a fractional part of a turn. Furthermore, the input capacitance of the transistor or tube (typical value for a transistor, common-emitter mode, is 3 to 100 pF, for a tube 5 to 10 pF) shunts the tuned-circuit inductance and for a given frequency of operation will make the resonating inductance smaller.

To provide proper match and to increase the required value of the tuned-circuit inductance, capacitive coupling between the input impedance of the amplifier and the tuned circuit may be used. See Fig. 5-38. The coupling

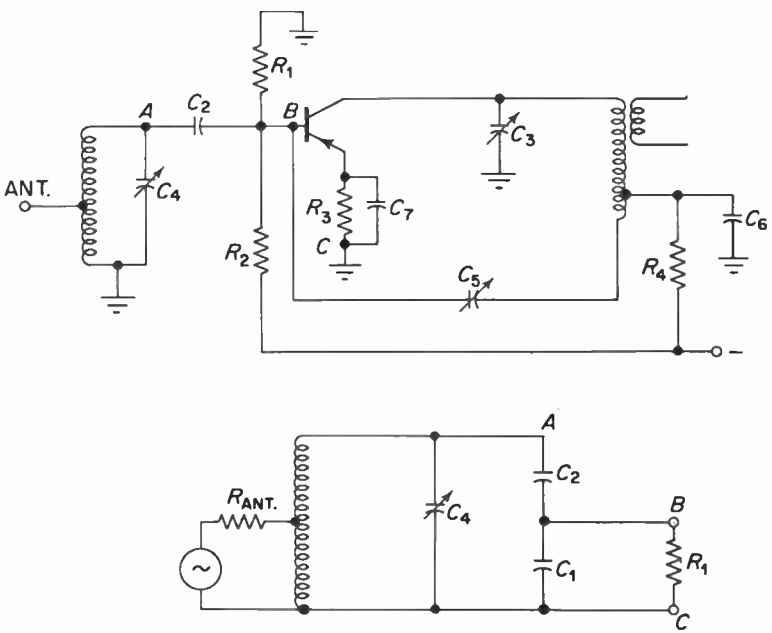



**Fig. 5-38.** A neutralized common emitter amplifier employing capacitive matching and its input circuit equivalent.

capacitor is generally much smaller than the input capacity of the amplifier and therefore isolates the amplifier input capacity from the tuned circuit. The ratio between the input capacity of the amplifier $C_1$ and the coupling capacitor $C_2$, determines the impedance transformation. If the input resistance of the amplifier is larger than the reactance of $C_1$, the effective turns ratio of the circuit becomes

$$n = \frac{C_1 + C_2}{C_2}. \tag{5-34}$$

This effective turns ratio may be used like the turns ratio of an inductive circuit. Thus,

$$R_p = R_1 n^2, \tag{5-35}$$

where $R_1$ is the input resistance of the amplifier and $R_p$ is the effective resistance across the tuned circuit. From the above two equations the value of $C_2$ necessary to provide a match is found to be

$$C_2 = \frac{C_1}{\left(\frac{R_p}{R_1}\right)^{1/2} - 1}. \tag{5-36}$$

The total capacitance shunting the tuning inducatnce, due to $C_1$ and $C_2$, is

$$C_t = \frac{C_1 C_2}{C_1 + C_2}. \tag{5-37}$$

In many circuits the matching capacitors are shunted by the variable tuning and trimmer capacitors; these must be added to the value of $C_t$ determined above. The resonant frequency of the tuned circuit is determined by value of $L$ and the total capacitance across it. The inductance may also be tapped to permit a proper match between the antenna and the input impedance of the amplifier.

Consider the following example. A folded dipole ($Z = 300$ ohms) operating at 100 MHz is to match a transistor input impedance consisting of a 40-pF capacitor in parallel with a 250-ohm resistance. The tuned-circuit inductance is to be used as an autotransformer and is considered lossless. The loaded bandpass of the tuned circuit is to be 1 MHz. The inductive reactance of the coil is assumed to be 100 ohms.

In order for the bandpass to be 1 MHz, the effective shunt resistance across the tuned circuit under fully loaded conditions must be $R_{\text{shunt}} = QX_L$. Since $Q = f_r/B$ or 100 MHz/1 MHz = 100, $R_{\text{shunt}} = 10{,}000$ ohms. If a power match is desired, then this 10,000-ohm effective resistance must be considered to be made of two equal shunt components of 20,000 ohms each. One of these shunt components is due to the transformed antenna resistance, the other to the transformed transistor input resistance. Using Eq. (5-32), the 300-ohm antenna resistance is transformed into 20,000 ohms, and the necessary turns ratio may be determined. Thus, $n = \sqrt{20{,}000/300}$

= 8.16. If the tuned-circuit inductance consists of 6 turns, then the proper tapping point would be found by Eq. (5–33) to be 0.73 turns from the bottom of the coil.

Using Eq. (5–36) to determine the matching capacitor, $C_2$ is found to be 5.04 pF. This matches the 250-ohm transistor resistance to the 20,000-ohm effective antenna resistance shunting the tuned circuit. Thus, the total effective shunt resistance across the tuned circuit is 10,000 ohms. Since the $Q$ required for a 1-MHz bandpass is 100 ($B = f_r/Q$), and since the effective shunt resistance is 10,000 ohms, the required capacitive reactance is 100 ohms ($Q = R/X$). At 100 MHz a 15.9-pF capacitor has a reactance of 100 ohms. Solving Eq. (5–37), $C_1$ and $C_2$ are found to contribute 4.45 pF; therefore, the main tuning capacitor and the trimmer must provide the remaining 11.44 pF. To resonate the circuit an inductance of 0.159 microhenry is required.

### *Tuned and untuned input circuits*

RF amplifiers may be broadly classfied into two types: those whose inputs are tuned by means of a variable capacitor or inductor, and those whose inputs are fixed-tuned.

The fixed-tuned inputs are commonly associated with the grounded-grid amplifier (see Fig. 5-2). In such amplifiers the input resistance of the vacuum tubes is low and, therefore, loads the input tuned circuit. This loading will result in an input tuned-circuit bandpass of 20 MHz or more, which is required in order to permit reception of the commercial FM band. A wide bandpass will tend to degrade the interference-rejecting qualities of the receiver. On the other hand the variable-tuned input circuit can be designed to provide bandwidths of 1 MHz, thus providing reception that is relatively free of interference.

Most receivers that use variable tuning employ capacitor tuning rather than variable-inductance tuning—chiefly because of mechanical difficulties associated with the latter. These difficulties are related to smoothness of operation, reproducibility of tuning, and the complexity of the devices required to move the ganged cores or slugs in and out of their associated coils. Variable-inductor tuning may also require specially shaped slugs to provide tracking between the oscillator and the RF amplifier tuning. This does not mean that variable-capacitor tuning is faultless. Variable capacitors may introduce undesired coupling between stages via the common metal shaft of the ganged capacitor and the distributed inductance of the metal plates that make up the capacitor. Also variable-capacitor tuning is more prone to microphonics and wiper contact noise than is variable-inductance tuning.

Variable-capacitor tuning and variable-inductance tuning have different effects on gain and bandpass. Variable-capacitor tuning will tend to maintain

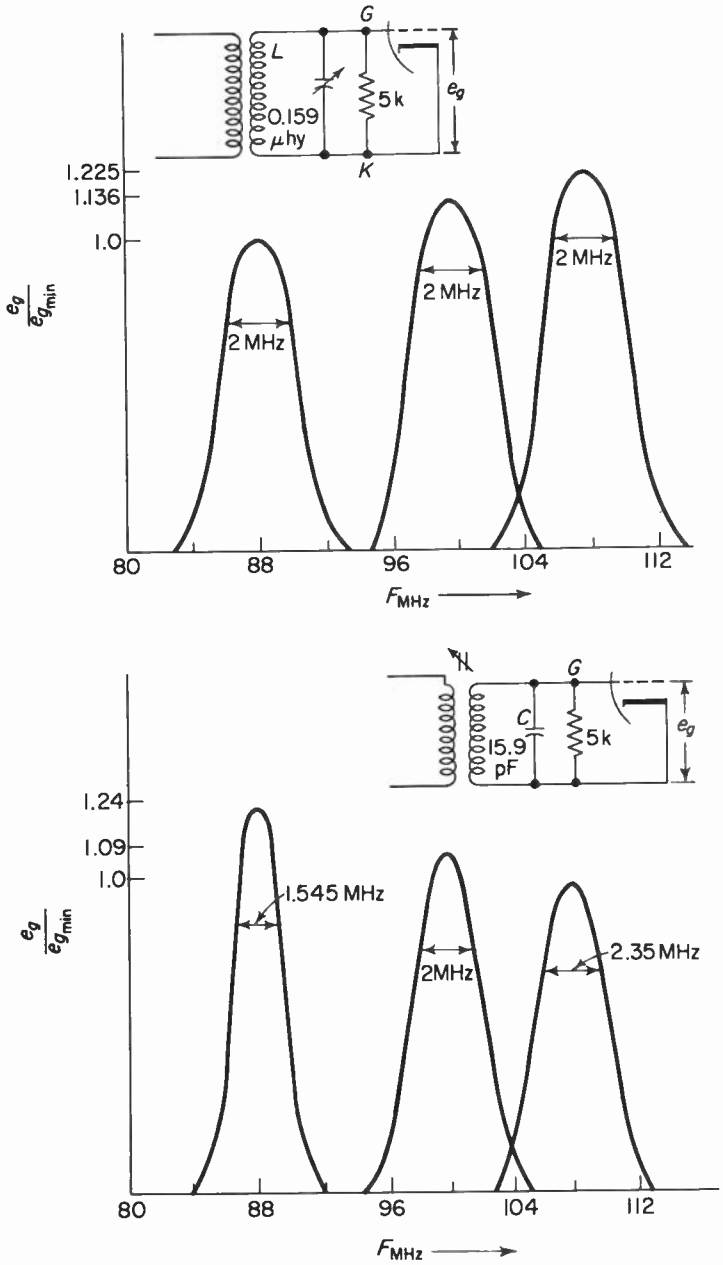
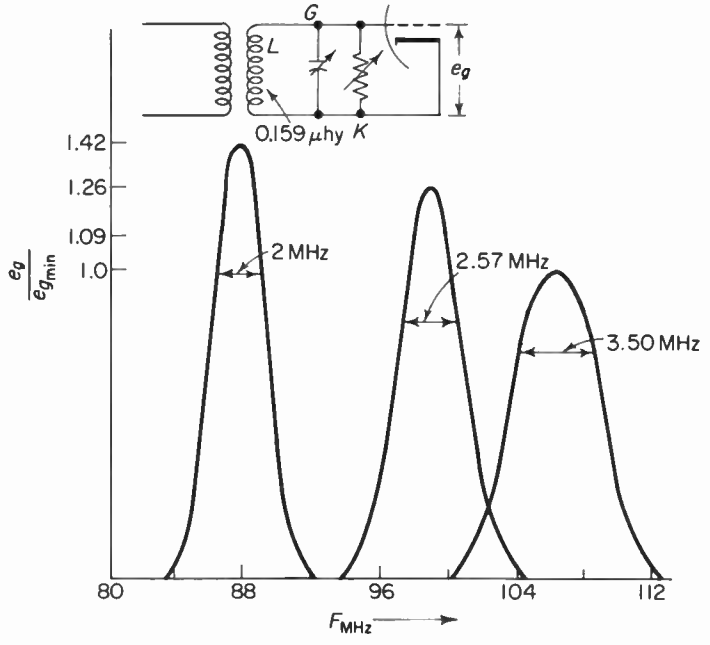
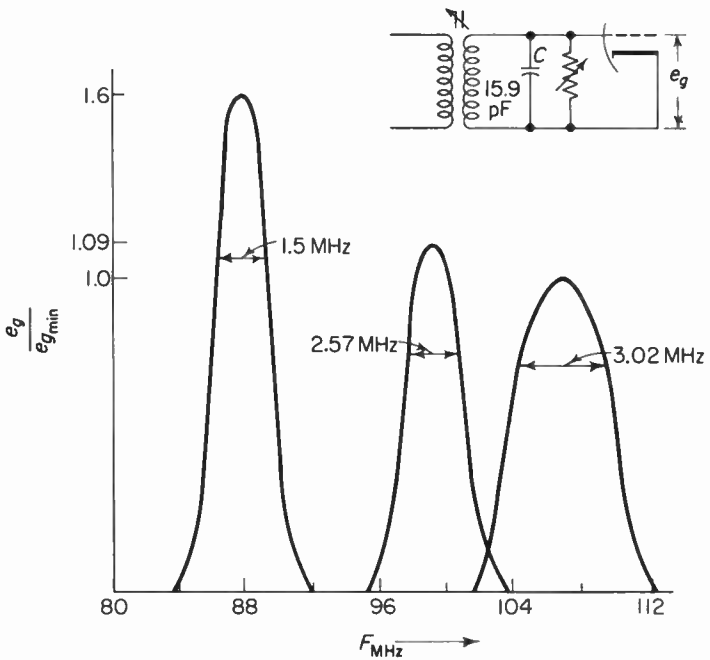



**Fig. 5-39.** These curves indicate the changes in bandpass and input transformer gain of a variable tuned circuit with the input resistance of the tube constant.




Variable capacitor tuning




Variable inductance tuning

**Fig. 5-40.** These curves indicate the changes in bandpass and input transformer gain of a variable tuned circuit with the input resistance of the tube inversely proportional to the frequency squared.

a constant bandwidth (assuming constant loading) throughout the tuning range, and at the same time will display an increasing input transformer gain with an increase in frequency. Variable-inductance tuning will tend to broaden the tuned-circuit bandpass as the frequency of operation increases, and at the same time the input transformer gain will decrease. Figure 5-39 illustrates these changes. It is assumed that the circuit is power-matched and that the tube's input resistance is constant with frequency. In fact the input resistance decreases with an increase in frequency. The total input resistance for a pentode or a neutralized triode can be calculated by

$$R_t = \frac{1}{(4\pi^2 f^2 g_m)(T_{kg}^2 L_k C_{gk})}, \tag{5-38}$$

where $T_{kg}$ is the transit time, $L_k$ is the cathode lead inductance, $C_{gk}$ is the grid-to-cathode interelectrode capacitance, $g_m$ is the tube transconductance, and $f$ is the frequency. It can be seen that a change in frequency of 1.23 times (108 MHz/88 MHz) will cause a decrease in $R_t$ by 0.666 times. Thus, if the input resistance is 5000 ohms at 88 MHz, it will be 3300 ohms at 108 MHz if all other quantities remain constant. The results of input-resistance variation with frequency on the bandpass of the tuned circuit can be seen in Fig. 5-40. Both variable-capacitor tuning and variable-inductance tuning exhibit changes in bandwidth with frequency. If the ratio between the bandwidth at high frequencies and the bandwidth at low frequencies is determined from Fig. 5-40, it will be found that this ratio for variable capacitor tuning is smaller (1.75) than that obtained for variable inductance tuning (1.96). Likewise it will be noticed that variable inductance tuning results in a greater variation in transformer gain then does variable capacitor tuning.

## 5-8. The FET RF amplifier

The field-effect transistor may be used in any of three basic amplifier-circuit configurations. Figure 5-41 illustrates these circuit arrangements. For application as a VHF RF amplifier the FET is most commonly used in the common-source configuration, although the common-gate and common-drain arrangements may be used at somewhat less gain. FET amplifiers may also be used in cascode arrangements similar to those discussed under vacuum-tube amplifiers.

From a DC-circuit point of view, an N-channel depletion-mode MOSFET has characteristics almost identical to the DC-circuit considerations of a vacuum tube, except that vacuum tubes are normally operated at very much greater voltages. However, such things as the methods for obtaining bias for both vacuum tubes and FET's are similar.

Figure 5-42 shows three basic methods for biasing an MOSFET. These are (a) fixed bias, (b) self bias, and (c) combination bias. In fixed bias the

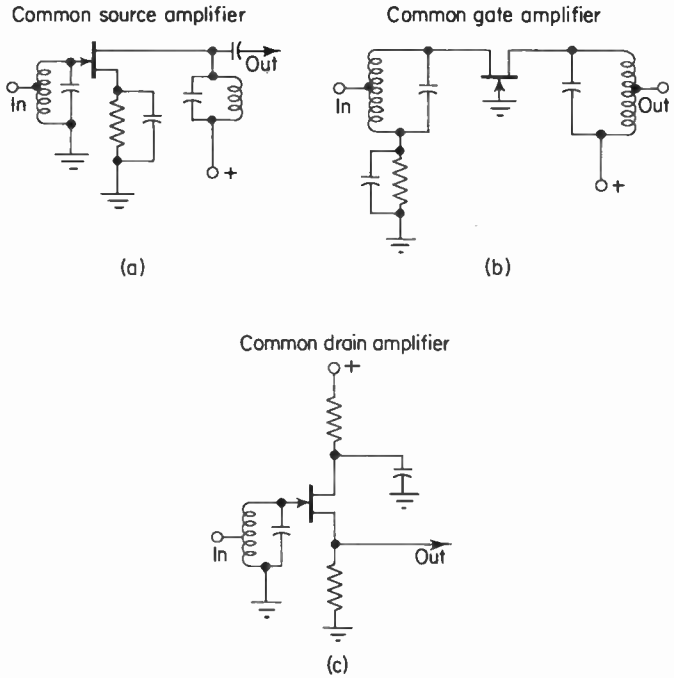



**Fig. 5-41.** Three basic tuned amplifier circuit configurations using junction field effect transistors.

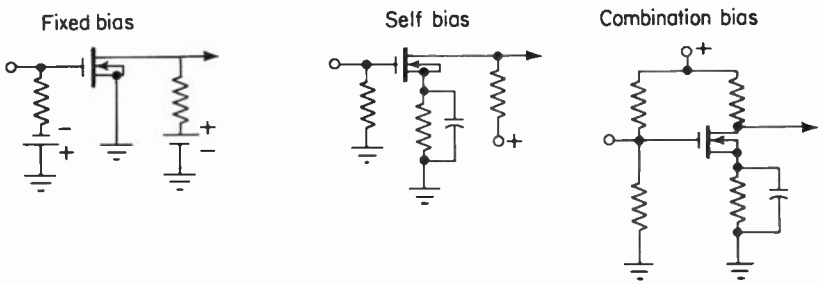



**Fig. 5-42.** Three methods for providing bias for a MOS FET. This type of FET may also be biased at zero.

DC voltage between gate and source is obtained from a separate source. Self bias is produced by the DC component of drain current passing through a resistor ($R_s$) in series with the source. The voltage drop produced across $R_s$ is in fact also between the gate and the source; thus the voltage drop across $R_s$ is the bias for the FET. If feedback is to be avoided, the bias resistor of the FET must be bypassed, just as with vacuum-tube self bias. In combination bias the gate-to-source bias voltage may be compared to the voltage developed across the meter of an unbalanced bridge. One arm of the bridge

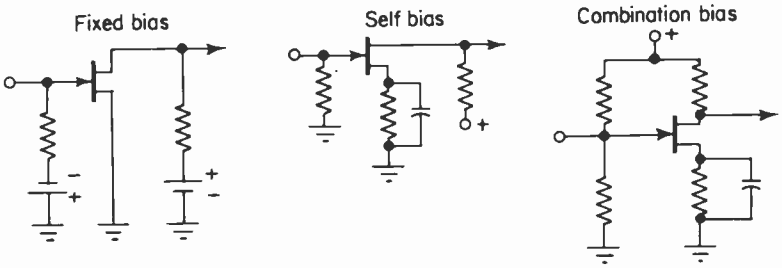



**Fig. 5-43.** Three methods for providing bias for a JFET.

consists of the voltage divider $R_1$ and $R_2$; the other arm consists of $R_s$ and the drain-to-source resistance of the FET. In this case, the unbalanced bridge current will be very small because of the isolation of the gate from the channel. Since the gate and source do not form a junction diode in an MOSFET, it is possible to operate the FET at zero or positive bias. This is limited by the drain current and maximum drain dissipation.

Methods of bias comparable to grid-leak bias of a vacuum tube cannot be used in an MOSFET, since the gate and source do not form a diode and cannot rectify the input signal. The junction FET, however, does have a PN junction between gate and source; thus an input signal of sufficient amplitude can be rectified and the DC component so derived can be used to bias the FET. Fixed, self, and combination bias may also be used in a JFET. Examples of these types of bias are shown in Figure 5-43. An important difference between the JFET and the MOSFET is that the JFET cannot be operated at zero bias, since this would forward-bias the gate-to-source junction and would allow very high gate and drain currents, which might damage the FET or at best cause excessive loading of the input source.

### *Distortion, noise and AGC*

The schematic diagram of a high-frequency common-source MOSFET amplifier is shown in Fig. 5-44. This type of tuned amplifier is characterized by low third-harmonic and cross-modulation distortion—primarily because of the almost perfect square-law characteristic of the FET, which for small-

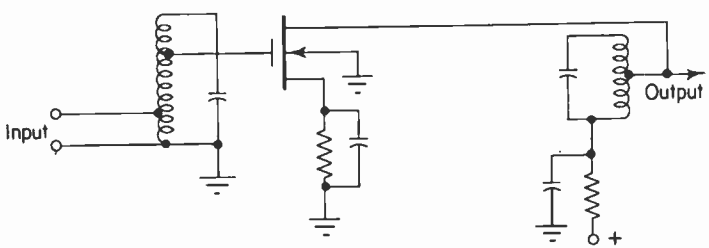



**Fig. 5-44.** A tuned RF amplifier using an MOS FET.

signal operation eliminates third- and higher-order terms. This reduction in distortion plus better selectivity results in improved spurious-response rejection, which in practice may be equal to or better than that achieved with receivers using vacuum tubes for the front end.

Another important characteristic of the FET amplifier is its excellent noise figure. When an FET is operated at the proper drain current and from a signal source whose internal resistance is optimum, its noise performance may be superior to that of most vacuum tubes. The noise factor of a high-frequency FET is given by

$$F = 1 + \frac{\omega}{g_m}(G_{gs} + C_{gd}) + \frac{0.5}{g_m}\left[\frac{G_s - W(C_{gs} + C_{gd})}{G_s}\right]^2, \qquad (5\text{–}39)$$

where

$G_s$ = source conductance,
$g_m$ = low-frequency transconductance,
$\omega$ = frequency of operation,
$C_{gs}$ = gate-to-source capacitance,
$C_{gd}$ = gate-to-drain capacitance.

The optimum source conductance may be determined by the following approximate formula:

$$G_{opt} = \omega(C_{gs} + C_{gd}). \qquad (5\text{–}40)$$

In general the noise factor of the FET may be determined by the same methods used to determine the noise factor of a triode. The only difference is that the value of the equivalent noise resistance of the FET is equal to $2/3g_m$, whereas the equivalent noise resistance of the triode is equal to $2.5/g_m$. Thus, the noise resistance of the FET is one-fourth that of a triode vacuum tube of equal transconductance. It is, therefore, to be expected that the FET will provide better weak-signal reception than can either bipolar transistors or vacuum tubes.

The slope of the static transfer curve ($I_d/E_g$) is defined as the $g_m$ of the FET. Since this curve is square-law in nature, the $g_m$ of the device will decrease as the bias approaches drain-current cutoff. Thus, any shift in bias will affect the gain of the amplifier. This method of varying the gain of an amplifier by varying the bias of an amplifier is called reverse automatic gain control; for all practical purposes it is the same as the method used for controlling the gain of vacuum-tube amplifiers.

### *Power gain and stability*

An high frequencies the input impedance of an FET amplifier decreases. Thus, the input circuit of the FET will draw power from the input source; and therefore the amplifier behaves more like a power amplifier than a voltage amplifier. Since this requires that both input and output parameters be known, design of an FET amplifier at high frequencies will be similar to the design of high-frequency transistor amplifiers, in that $Y$ parameters may be used.

Typical $Y$ parameters for an FET operating at 100 MHz in a common-source circuit configuration might be as follows:

$Y_{is}$ (input admittance with output short-circuited) $= 0.15 + j4$ mmhos,
$Y_{os}$ (output admittance with input short-circuited) $= 0.18 + j9$ mmhos,
$Y_{fs}$ (forward transfer admittance with output short-circuited) $= 7 - j1$ mmhos,
$Y_{rs}$ (reverse transfer admittance with input short-circuited) $= 0 - j0.06$ mmhos.

If a tuned amplifier such as that of Fig. 5-44 is to be designed so that both the input and output tuned circuits have a bandpass of 1 MHz at 100 MHz, then each tuned circuit must have a loaded $Q$ of 100. If the reactance of the inductance is assumed to be 50 ohms, then for a $Q_L$ of 100 the effective shunt resistance across the coil must be 5000 ohms. This shunt resistance will consist of two components. In the case of the input tuned circuit, one shunt component will be due to the input generator (assumed to be 300 ohms) and the other to the input admittance of the FET. In the case of the output tuned circuit, one of the resistive shunt components will be due to the amplifier load and the other to the output admittance of the amplifier. Thus, in designing an amplifier, we must know the input and output admittance. Assuming a load resistance of 2000 ohms (0.5 mmho )and using Eq. (5–24), we find the input admittance of the amplifier to be

$$Y_{in} = (0.15 + j4) - \frac{(7 - j1)(0 - j0.06)}{0.18 + 0.5} = 0.239 + j4.623 \text{ mmhos.}$$

The output admittance may be found by using Eq. (5–25) and assuming a conjugate match between $Y_{in}$ and $Y_s$. Thus $Y_{out}$ is

$$Y_{out} = (0.18 + j0.9) - \frac{(7 - j1)(0 - j0.06)}{0.15 + 0.239} = 0.333 + j2 \text{ mmhos.}$$

At this point the stability of the amplifier may be checked by using Stern's stability criterion, Eq. (5–29a):

$$K = \frac{|(7 - j1)(0 - j0.06)| + (-0.06)}{2(0.15 + 0.239)(0.18 + 0.5)} = 0.8.$$

Since $K$ is less than 1, the amplifier will be stable without neutralization, although in order to maintain stability a mismatch was required in the output circuit ($G_o = 0.333$ and $G_L = 0.5$). Under actual conditions neutralization may be employed as a means of adjusting the skew or shape of the tuned-circuit response curve. The same methods of neutralization discussed for vacuum-tube amplifiers are used for FET amplifiers.

The power gain of the amplifier is found by means of Eq. (5–28):

$$PG = \frac{|7.1|^2}{|0.18 + 0.333|^2} \times \frac{0.5}{0.239} = 26 \text{ db.}$$

We can cause the input tuned circuit to act as a matching transformer and at the same time achieve the desired bandpass by properly tapping the coil and making it an autotransformer. Thus, the 300-ohm source may be converted into a 10,000-ohm resistive shunt component across the coil by finding the proper turns ratio:

$$n = \sqrt{\frac{10\text{k}}{300}} = 5.8.$$

The resistive component of the input admittance is 0.239 mmho or 4200 ohms. This resistive shunt component may also be converted into an equivalent 10,000-ohm resistive component across the coil by finding the proper turns ratio:

$$n = \sqrt{\frac{10\text{k}}{4.2\text{k}}} = 1.54.$$

These two turns ratios indicate that if the transformer coil consists of six turns, the input source will be connected approximately ($n_t/n = n_1$, $6/5.8 \approx 1$) one turn from the bottom of the coil and the input of the FET will be connected approximately four turns from the bottom of the coil. Under these conditions the total effective load across the coil ($X_L = 50$ ohms) will be 5000 ohms (two 10,000-ohm components in parallel). Thus, the loaded $Q$ of the input tuned circuit will be 100. The inductance required for a reactance of 50 ohms at 100 MHz will be 80 nhy. This in turn will require a shunt capacity of 32 pF in order to produce resonance.

The output tuned circuit may be handled in the same way as the input tuned circuit. However, in this example an important difference exists between the input and output matching requirements. In the output circuit of the amplifier a mismatch exists between the resistive components of output admittance and the resistive component of the load, whereas in the input circuit of the amplifier the resistive component of the source matches the resistive component of input admittance.

One way of maintaining an impedance mismatch while providing the proper loading on the output tuned circuit so that the required bandpass is obtained, is to simply connect the output of the transistor directly across the input load of the next stage. Then place this parallel combination at the proper tap on the tuned circuit to provide the desired loading on the tuned circuit. In this example the parallel combination of $G_L$ and $G_{\text{out}}$ is

$$\begin{aligned} G_T &= G_L + G_0 \\ &= .5\text{mmho} + .333 \text{ mmhos} \\ &= .833 \text{ mmhos} \end{aligned}$$

which is the equivalent of 1,200 ohms ($R_T$).

Since a bandpass of 1 MHz at a frequency of 100 MHz requires a tuned

circuit $Q$ of 100 and since the reactance of the coil ($X_L$) is 50 ohms, the effective shunt resistance must be

$$\begin{aligned} R_{\text{eff}} &= Q_L X_L \\ &= 100(50) \\ &= 5000 \text{ ohms} \end{aligned}$$

The tuned circuit is converted into a step-up transformer by connecting the resultant load of 1,200 ohms on to a tap on the coil. See Fig. 5-44. The required turns ratio is

$$n = \sqrt{\frac{R_{\text{eff}}}{R_T}} = \sqrt{\frac{5000}{1200}} = 2.04$$

Thus the tap on the tuned circuit must be located

$$n_1 = \frac{n_t}{n} = \frac{6}{2.04} = 2.9$$

turns from the bottom of the coil.

## 5-9. Summary

**1.** The four important functions of RF amplifiers are: (1) to reduce oscillator radiation, (2) to reduce image and spurious response, (3) to increase receiver sensitivity, (4) to improve the signal-to-noise ratio of the receiver.

**2.** Local oscillator radiation is not detrimental to the receiver from which it is emitted, but is a source of interference to other receivers.

**3.** The FCC limits oscillator radiation to 50 microvolts per meter at a distance of 100 feet.

**4.** Oscillator radiation can be reduced by the use of an RF amplifier that isolates the antenna from the oscillator.

**5.** Owing to the heterodyne action of a superheterodyne receiver, this type of receiver is subject to a wide range of spurious responses.

**6.** Three types of spurious responses are: (1) image response, (2) adjacent and cochannel interference, (3) second-order intermodulation products.

**7.** The tuned RF amplifier provides additional selectivity for the receiver and therefore minimizes spurious responses.

**8.** RF amplifiers used in FM receivers may use vacuum tubes, FETs or transistors.

**9.** Weak-signal amplification requires a low-noise RF amplifier.

**10.** Voltage amplification is directly proportional to the $L/C$ ratio of the output tuned circuit.

**11.** The gain-bandwidth product is used as a figure of merit for vacuum tubes and transistors.

**12.** AVC has two functions in an FM receiver: (1) it aids in the amplitude-limiting process, and (2) it minimizes overloading in the RF and IF amplifier stages of the receiver.

**13.** Vacuum tubes rely on the nonlinearity of their gird transfer characteristics to enable a change in grid bias to cause a change in stage gain.

**14.** Transistors obtain changes in stage gain by shifting the operating point of the stage toward cutoff or saturation. This is called reverse or forward AVC, respectively.

**15.** Pentode RF amplifiers are characterized by good stability, high gain, moderately high input impedance, and high noise level.

**16.** Neutralized grounded-cathode triode RF amplifiers are characterized by low noise, high gain, and good stability. Therefore this type of amplifier is suitable for weak-signal application.

**17.** Owing to the greater input resistance of a neutralized triode (as compared to a pentode), the gain of a triode can be greater than that of a pentode operating under similar conditions.

**18.** The grounded-grid amplifier exhibits low noise, good stability, less gain than a pentode, and low input impedance which results in a broadband input.

**19.** In a grounded-grid amplifier the magnitudes of voltage amplification and power amplification are equal.

**20.** The noise figures of a grounded-grid amplifier and a neutralized grounded-cathode amplifier are about the same.

**21.** The cascode RF amplifier has the noise figure of a triode and the gain of a pentode. The cascode amplifier consists of a grounded-cathode amplifier feeding a grounded-grid amplifier.

**22.** In the cascode amplifier the voltage amplification of the first stage is slightly greater than unity. The second-stage voltage amplification is equal to the second-stage power gain.

**23.** If the plate resistance of a pentode is very much greater than the load resistance, the voltage amplification of the cascode is always less than that of the pentode—assuming that both the pentode and the triode have the same $g_m$.

**24.** Transistor RF amplifiers are usually of the common-base or the common-emitter variety.

**25.** Transistors employed at very high frequencies are constructed to insure that current-carrier transit time between emitter and base is reduced to a minimum.

**26.** Transistor circuit analysis at very high frequencies is facilitated by means of the *Y*-parameter equivalent circuit. These parameters are more easily determined at high frequencies than are other parameters, and the simple calculations are easier to perform.

**27.** Common-base RF amplifiers display regenerative internal feedback. If the feedback is not excessive, the common-base amplifier will deliver greater power gain than the common-emitter amplifier, which at high frequencies exhibits degenerative internal feedback.

**28.** Transistor instability can be avoided by neutralization, mismatch, or unilateralization.

**29.** The input circuit of a weak-signal RF amplifier must fulfill five functions: power match, noise match, minimize insertion loss, provide adequate selectivity, and provide balanced to unbalanced circuit transformation.

**30.** Power matching may be accomplished by inductive or capacitive means.

**31.** The means of tuning the input circuit will to some degree determine the changes in input circuit gain and bandwidth that occur as the receiver is tuned across the FM band.

**32.** The field effect transistor is similar to a triode vacuum tube in at least three ways. These are: 1. they are both three-element devices, 2. they both have high low frequency input impedance and lower high frequency input inpedance, and 3. the output impedance of both is in the same order of magnitude.

**33.** The junction field effect transistors (JFET) and the isolated gate field effect transistors (IGFET) are the two basic types of field effect transistors.

**34.** The JFET is similar to the vacuum tube in that they both require a reverse bias between their input terminals, whereas the IGFET may not require any bias for its proper operation.

**35.** FET's may function in either an enhancement or a depletion mode of circuit operation.

**36.** The output characteristic of an FET is similar in shape to that of a pentode.

**37.** For VHF applications the FET is most commonly used in the common source configuration.

**38.** FET's do not have heaters and require much less operating votage than do vacuum tubes.

**39.** Bias methods for the vacuum tube and the FET are similar.

**40.** The FET amplifier is characterized by low distortion and low noise as compared to a vacuum tube amplifier.

## REFERENCES

1. Lovering, W. F.: *Radio Communication.* David McKay Company, Inc., New York, 1958.
2. Rheinfelder, W. A.: *Design of Low Noise Transistor Input Circuits.* Hayden Book Companies, New York, 1964.

3. Pullen, K. A.: *Handbook of Transistor Circuit Design.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.
4. Henney, K.: *The Radio Engineering Handbook.* McGraw-Hill Book Company, New York.
5. Mural, Frank: "Spurious Multiple Responses in FM Receivers," LB-762, RCA Laboratories Division Industry Service Laboratory, 1948.
6. Gray, C. R., and Lawson, T. C.: "Transistorized FM Front Ends," *Audio*, August 1962.
7. von Recklinghausen, D. R.: "FM Tuner Characteristics and Their Relative Importance," *Audio*, August and September 1963.
8. Watts, H. M.: "Television Front-End Design," *Electronics*, April and May 1949.
9. Lawson, T. C.: "An FM Receiver Using MADT Transistors," *The Solid State Journal*, April 1961.
10. Amos, S. W., and Johnstone, G. C.: "Design for an FM Tuner," *Wireless World*, April and May 1955.
11. Paananen, R.: "Head End Design for FM Tuner," *TV and Radio Engineering*, February and March 1953.
12. Cooke, H.: "Transistorized Entertainment Type FM Receivers," *Semiconductor Products*, March 1959.
13. Englund, J. W., and Thanos, H.: "Application of RCA Drift Transistors to FM Receivers," *IRE Trans. on Broadcast and Television Receivers*, January 1959.
14. Eberhardt, A. E.: "Spurious Response Chart," *Tele-Tech*, December 1952.
15. Janes, R. B.: "The Continuing Revolution in Semiconductors," *RCA Engineer*, November 1962.
16. Reddi, V. G. K.: "Applying Transistor '*Y*' Parameters," *Electronic Industries*, January 1960.
17. Fitchen, F. C.: *Transistor Circuit Analysis and Design.* D. Van Nostrand Co., Inc., Princeton, N.J., 1960.
18. Cheng, C. C.: "Neutralization and Unilateralization," *IRE Trans. PGCT*, June 1955.
19. Chu, Y. Y.: "Unilateralization of Junction Transistor Amplifiers at High Frequencies," *Proc. IRE*, July 1955.
20. Hunter, L. P.: *Handbook of Semiconductor Electronics.* McGraw-Hill Book Company, New York, 1956.
21. Siegal, M. E.: "Investigation of Coupled Circuits for 100–1000 mc. Application," *RCA Engineer*, October–November 1956.
22. Stern, A. P.: "Stability and Power Gain of Tuned Transistor Amplifiers," *Proc. IRE*, March 1957.
23. Linville, J. G., and J. F. Gibbons: *Transistors and Active Circuits.* McGraw-Hill Book Company, New York, 1961.
24. Wallmark, J., and H. Johnson: *Field-Effect Transistors.* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.

## APPENDIX 5-1

This appendix presents the derivation of the grounded-grid amplifier input impedance of Eqs. (5-9) and (5-10). Refer to Fig. 5-22 for identification of symbols used in the following equations.

$$Z_{in} = \frac{e_s}{i_p}.$$

From Fig. 5-22 it can be seen that

$$\mu e_g + e_s = i_p r_p + i_p r_s + i_p R_L.$$

But

$$e_g = e_s - i_p r_s,$$

$$\therefore \quad \mu e_s - \mu i_p r_s + e_s = i_p r_p + i_p r_s + i_p R_L,$$

$$e_s(\mu + 1) = i_p(r_p + R_L + r_s + \mu r_s),$$

$$\therefore \quad i_p = \frac{e_s(\mu + 1)}{r_p + r_s(1 + \mu) + R_L},$$

$$Z_{in} = \frac{e_s}{i_p} = \frac{e_s}{\dfrac{e_s(\mu + 1)}{r_p + r_s(1 + \mu) + R_L}}$$

$$= \frac{r_p + r_s(1 + \mu) + R_L}{\mu + 1},$$

$$Z_{in} = \frac{r_p}{\mu + 1} + \frac{R_L}{\mu + 1} + r_s.$$

If $\mu \gg 1$ and if $R_L \approx r_p$, as may be the case for a triode.

$$Z_{in} = \frac{1}{g_m} + \frac{1}{g_m} + r_s,$$

$$Z_{in} = \frac{2}{g_m} + r_s.$$

## APPENDIX 5-2

This appendix presents the derivation of Eq. (5–16), the voltage amplification of the second stage of a cascode amplifier. Refer to Fig. 5-25 for the equivalent circuit of a cascode amplifier.

$$A_v = \frac{e_o}{e_g},$$

$$e_o = (e_g + \mu e_g)\left(\frac{R_L}{r_p + R_L}\right),$$

$$e_o = e_g(1 + \mu)\left(\frac{R_L}{r_p + R_L}\right),$$

$$A_v = \frac{e_o}{e_g} = \frac{(1 + \mu)(R_L)}{r_p + R_L}.$$

## APPENDIX 5-3

This appendix presents the derivation of Eq. (5–19), the voltage amplification of the cascode amplifier. Refer to Fig. 5-25 for the equivalent circuit of the cascode amplifier. In this derivation it is assumed that the parameters of both tubes are identical.

$$A_{v_1} \text{ of first stage} = \frac{e_g}{e_s}.$$

By proportion (see Fig. 5-25) it can be seen that

$$e_g = \mu e_s\left(\frac{r_p + R_L - \dfrac{\mu e_g}{i_p}}{2r_p + R_L - \dfrac{\mu e_g}{i_p}}\right). \tag{1}$$

But

$$i_p = \frac{\mu e_g + \mu e_s}{2r_p + R_L} \quad \text{and} \quad e_g = \mu e_s - i_p r_p$$

$\therefore \quad \mu e_s = e_g + i_p r_p,$

$$i_p = \frac{\mu e_g + e_g + i_p r_p}{2r_p + R_L} = \frac{e_g(\mu + 1)}{r_p + R_L}. \tag{2}$$

Substituting Eq. (2) in Eq. (1),

$$e_g = \mu e_s\left(\frac{r_p + R_L - \dfrac{\mu e_g}{\dfrac{e_g(\mu + 1)}{r_p + R_L}}}{2r_p + R_L - \dfrac{\mu e_g}{\dfrac{e_g(\mu + 1)}{r_p + R_L}}}\right)$$

$$= \mu e_s\left(\frac{r_p + R_L - \dfrac{\mu(r_p + R_L)}{\mu + 1}}{2r_p + R_L - \dfrac{\mu(r_p + R_L)}{\mu + 1}}\right),$$

$$A_{v_1} = \frac{e_g}{e_s} = \mu\left(\frac{r_p + R_L - \dfrac{\mu(r_p + R_L)}{\mu + 1}}{2r_p + R_L - \dfrac{\mu(r_p + R_L)}{\mu + 1}}\right).$$

This may be further simplified as follows:

$$A_{v_1} = \mu\left(\frac{(r_p + R_L)\left(1 - \frac{\mu}{\mu + 1}\right)}{r_p + (r_p + R_L)\left(1 - \frac{\mu}{\mu + 1}\right)}\right)$$

$$= \frac{\mu}{\frac{r_p}{(r_p + R_L)\left(1 - \frac{\mu}{\mu + 1}\right)} + 1},$$

$$A_{v_1} = \frac{\mu}{\left(\frac{1}{1 + \frac{R_L}{r_p}}\right)\left(\frac{1}{1 - \frac{\mu}{\mu + 1}}\right) + 1}.$$

## APPENDIX 5-4

This appendix presents the derivation of the capacitance-matching equation (5–35). The discussion refers to the circuit shown in Fig. 5-38.

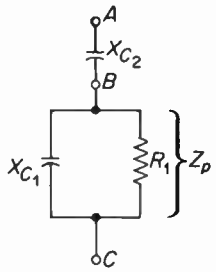



Fig. 1

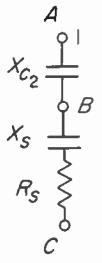



Fig. 2

Let it be assumed that $R_1$ is very much greater than $X_{C_1}$, and that $X_{C_1}$ is very much smaller than $X_{C_2}$. $Z_p$ is approximately equal to $X_{C_1}$. Refer to Fig. 1 and Fig. 2 and convert the parallel circuit ($Z_p$) into an equivalent series circuit as follows:

$$R_s = \frac{Z_p^2}{R_1}, \quad \therefore \quad R_s = \frac{X_{C_1}^2}{R_1}$$

$$X_s = \frac{Z_p^2}{X_{C_1}}, \quad \therefore \quad X_s = X_{C_1}$$

} From the assumptions above

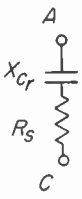



Fig. 3

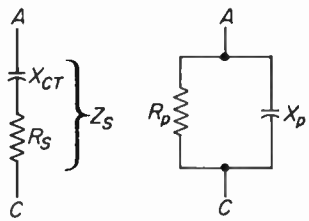



Fig. 4

Combine $X_{C_2}$ and $X_{C_1}$, see Fig. 3:

$$X_{C_T} = X_{C_2} + X_{C_1}.$$

Convert the series circuit of $X_{C_T}$ and $R_s$ into an equivalent parallel circuit as shown in Fig. 4.

$$R_p = Z_s^2/R_s,$$

$Z_s^2$ may be shown to be equal to $X_{C_T}^2$ as follows, $R_s$ equals $(X_{C_1}/R_1)X_{C_1}$,

$$\therefore \quad R_s < X_{C_1} \quad \text{and} \quad X_{C_1} < X_{C_2}; \quad \text{Hence} \quad X_{C_2} \approx X_{C_T}.$$

It follows that $X_{C_T} \gg R_s$,

$$\therefore R_p = X_{C_T}^2/R_s; \text{ but } R_s = X_{C_1}^2/R_1,$$

$$R_p = \frac{X_{C_T}^2 R_1}{X_{C_1}^2} = \frac{(X_{C_1} + X_{C_2})^2}{X_{C_1}^2} R_1$$

$$= \left(\frac{C_1 + C_2}{C_2}\right)^2 \times R_1.$$

Thus $R_p = n^2 R_1$, where $n$ is $\dfrac{C_1 + C_2}{C_2}$.

## APPENDIX 5-5

The derivation of the transfer equations which permit conversion of a series circuit into an equivalent parallel circuit or vice-versa is detailed below.

If both circuits shown in Fig. 1 are to be equivalent to one another their impedance as presented to the generater must be equal. Also the power dissipated or stored

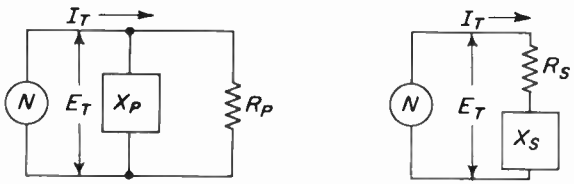



**Fig. 1**

must be the same for each circuit. Thus

$$Z_p = Z_s$$

and

$$P_p = P_s$$

$$\frac{E_T^2}{R_p} = I_T^2 R_s$$

$$\frac{E_T^2}{I_T^2 R_p} = R_s$$

$$\frac{Z^2}{R_p} = R_s$$

Similarly, the reactive power in the parallel circuit equals the reactive power in the series circuit

$$P_p = P_s$$

$$\frac{E_T^2}{X_p} = I_T^2 X_s$$

$$\frac{E_T^2}{I_T^2 X_p} = X_s$$

$$\frac{Z^2}{X_p} = X_s$$

# 6

# FREQUENCY CHANGERS

## 6-1. Introduction

The mixer oscillator stage of a superheterodyne receiver is located between the RF amplifier and IF amplifier stages as shown in Fig. 6-1. Notice that the input to the mixer consists of two signals—one a locally generated oscillator signal and the other an RF signal which may or may not be modulated. The function of the mixer oscillator stage is to change the frequency of the incoming RF signal to a lower frequency, called the intermediate frequency. Thus this stage is often called a frequency changer. In the case of a commercial FM receiver this intermediate frequency (IF) is almost always 10.7 MHz. Under ideal conditions, the IF is held constant for all properly tuned incoming signals by the simple expedient of simultaneously tuning the mechanically ganged RF and oscillator sections of the circuit. Thus, the frequency difference between them is held constant. The process whereby the

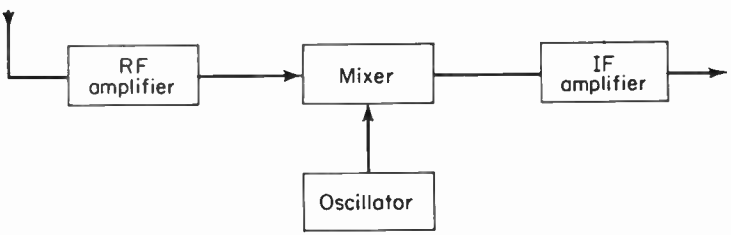



**Fig. 6-1.** The block diagram of the front end of a superheterodyne receiver.

frequency of operation of the RF and oscillator stages remains separated by the IF throughout their tuning range is called tracking.

The ability to reduce its frequency of operation gives the superheterodyne receiver its superiority over other receiver systems. The most important advantages of the superheterodyne are greater stability, higher gain, and greater selectivity. All of these characteristics are due to its lower frequency of operation. The chief disadvantages of lowering the frequency of operation are the increased internally generated noise and the production of spurious responses. These topics were discussed in detail in Chapters 4 and 5.

The function of frequency changing may be performed by three different types of circuits, the block diagrams of which are shown in Fig. 6-2. The mixer of Fig. 6-2(a) is usually defined as a frequency changer consisting of a separate mixer and a separate local oscillator stage. A converter, such as that indicated in the block diagram of Fig. 6-2(b), is also a frequency changer, but it combines the functions of oscillation and mixing in a single device. In the case of vacuum-tube converters the functions of frequency changing and oscillation, although performed in a single envelope, are electronically separate. The autodyne converter of Fig. 6-2(c) also combines the functions of mixing and oscillation, but both are performed by the same active element. Thus, the autodyne converter may be considered a self-oscillating mixer. Unfortunately, a great deal of confusion has surrounded these terms and they are often used interchangeably.

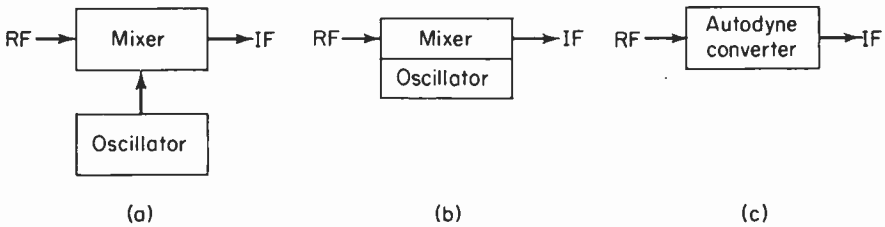



**Fig. 6-2.** The block diagram of three types of frequency changers.

## 6-2. The theory of frequency changing

Frequency changing, which is often referred to as heterodyning, is a form of low-level amplitude modulation. The oscillator acts as the carrier component of the modulation scheme and the RF signal is the modulating signal. The modulation process produces sideband components, which are used as the intermediate frequency. It is the usual practice to keep the oscillator signal level many times larger than the incoming signal, in order to obtain maximum conversion gain. We shall discuss this in detail later in this chapter. The incoming signal, whether it is AM or FM, will modulate

the resultant intermediate-frequency output of the mixer so that the input signal information is preserved but shifted in frequency.

Frequency changing is accomplished by two basic methods: additive mixing and product mixing. In both cases nonlinear circuit elements are necessary for the required frequency translation. Product mixing has also been called multiplicative mixing and is often referred to as a form of linear mixing.

In the case of additive mixing the oscillator and the RF signal are first linearly combined either in series or parallel, as indicated in Fig. 6-3, and then applied to the mixer where frequency conversion takes place. In Fig. 6-3(a) and (b) are shown typical transistor input circuits where the RF and oscillator signals are added either in series or in parallel. Also shown are the equivalent

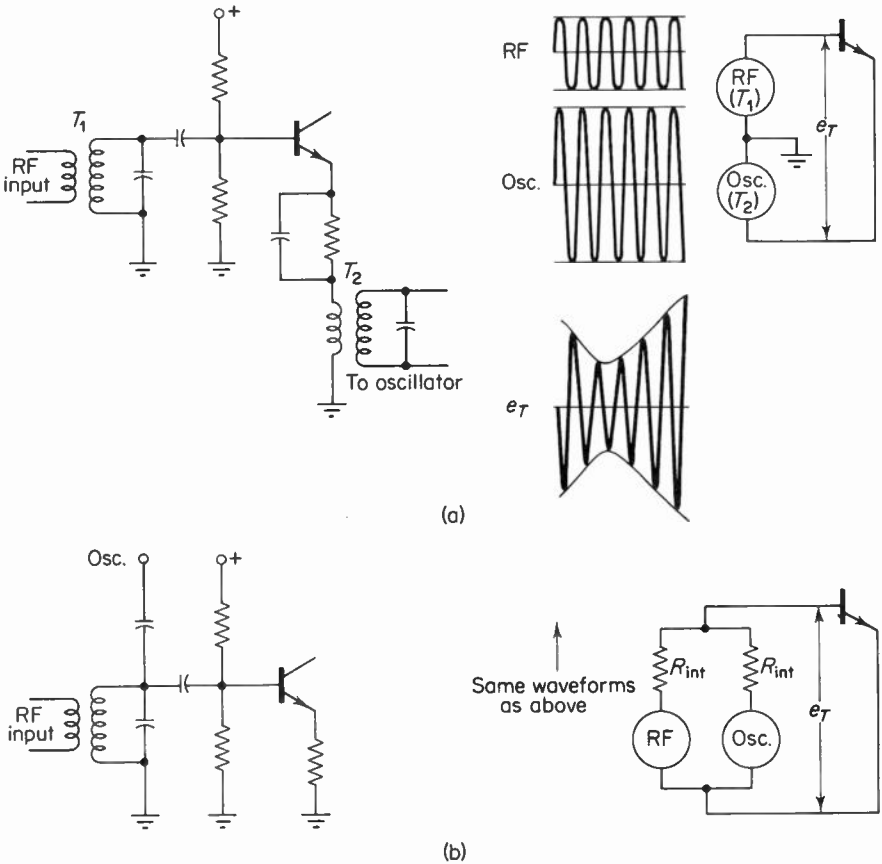



**Fig. 6-3.** Series addition of the oscillator and RF voltages is shown in fig. (a) and parallel addition of these voltages is shown in fig. (b). The waveforms are drawn under the assumption that the generators are feeding a linear load resistance.

circuits and their associated signal waveforms. The usual method of adding the RF and oscillator signals in series is by transformer coupling—that is, the oscillator voltage is induced into the RF signal tuned transformer. Another method of adding input signals in series is to have the RF input transformer connected in series with the secondary of an oscillator stage transformer.

Parallel addition of both signals is usually accomplished by capacitively coupling one or both to the mixer input. In effect this places the oscillator tuned circuit in parallel with the RF tuned circuit. The coupling capacitor connecting the oscillator to the input of the mixer plays an important part in determining the magnitude of oscillator voltage at the mixer input. This is because the RF tuned circuit is capacitive at the oscillator frequency (assuming that the oscillator is higher in frequency than the RF tuned circuit), and, therefore, the coupling capacitor and the tuned circuit constitute a voltage divider for the oscillator signal. Adjusting the size of the coupling capacitor determines the ratio of voltage division and the magnitude of the oscillator input voltage to the mixer.

In both series and parallel addition of input signals the input RF signal is also coupled into the oscillator circuit. This gives rise to a tendency for the local oscillator to synchronize with the incoming signal, often referred to as oscillator pulling. (The theory of pulling is discussed in detail elsewhere in the text.) Oscillator pulling has application in certain FM detector systems. In the context of frequency conversion, frequency pulling becomes a source of difficulty at high frequencies because the RF and oscillator signals are separated from one another by a small percentage. Thus, strong RF signals will force the oscillator frequency to move toward the input radio frequency and thereby reduce the intermediate frequency. Any reduction of intermediate frequency will result in loss of gain, increased harmonic distortion, and increased induced amplitude modulation. Under extreme conditions of pulling the converter may become inoperative. Oscillator pulling also subjects the local oscillator to undesirable frequency modulation by strong interfering signals. Oscillator pulling can be minimized by various bridge-circuit arrangements of the mixer. This will be discussed in detail under the topic of the autodyne converter.

In Fig. 6-3(a) the input RF and oscillator signals are shown graphically being applied to an assumed linear-input volt-ampere characteristic of a transistor. The resultant transistor input voltage ($e_T$) indicates that the input current to the transistor is not distorted, and thus no new signal components are produced. Notice that when two signals are close together in frequency ($< 2:1$), their linear addition results in a signal that has the appearance of an amplitude-modulated signal. To understand this one may resort to a phasor representation, such as that of Fig. 6-4. The phasors in this diagram represent the oscillator and RF signals and are used to show how the re-

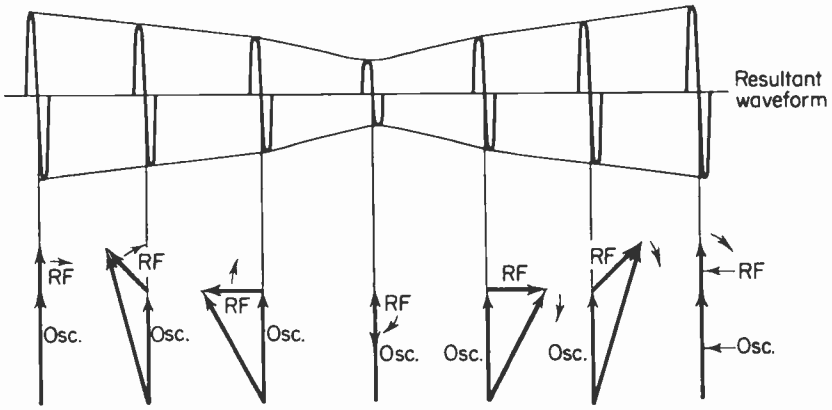



**Fig. 6-4.** The phasor and resultant signal resulting from linear addition of the oscillator and RF input signals.

sultant input signal is formed. The amplitude of a phasor represents the magnitude of a sinusoid; its speed of counterclockwise rotation indicates the frequency of the sinusoid. It is possible to determine the result of phasor addition of two sinusoids of different frequencies, by assuming that one of the sinusoids (oscillator) is held constant and that the other (RF) is moving with respect to it. If the oscillator signal is assumed to be reference and is higher in frequency than the RF (as is usually the case), then the RF phasor will appear to rotate in a clockwise direction with respect to the oscillator. This is because in one rotation of the oscillator phasor, the RF signal will not have completed its rotation and will appear to fall behind the oscillator phasor. The frequency of rotation of the RF signal phasor, with respect to the oscillator, will be the difference in frequency between the RF and the oscillator signals. Thus phasor addition of these signals over a period of time will produce an output whose amplitude and phase vary at the difference frequency, as shown in Fig. 6-4. The resultant signal consists of only the original two components. The average value of this signal is zero, and therefore it does not contain any signal energy at the difference frequency.

To produce frequency translation the signal must be applied to a nonlinear device. Vacuum tubes, transistors, FET's, and diodes are examples of nonlinear devices. When used in this way they are usually biased at or near current cutoff so that the device will act as a rectifier, or possess a square-law characteristic. The nonlinear device rectifies the input signal and produces an output whose average value varies at the difference frequency as shown in Fig. 6-5. In this diagram, the input signal, which consists of the RF and oscillator voltages, is applied to the nonlinear output characteristic of a transistor. This results in rectification, since the negative half of the input signal is made to operate in the region of output-current cutoff. The resultant output signal of Fig. 6-5 is seen to vary at the difference frequency. The

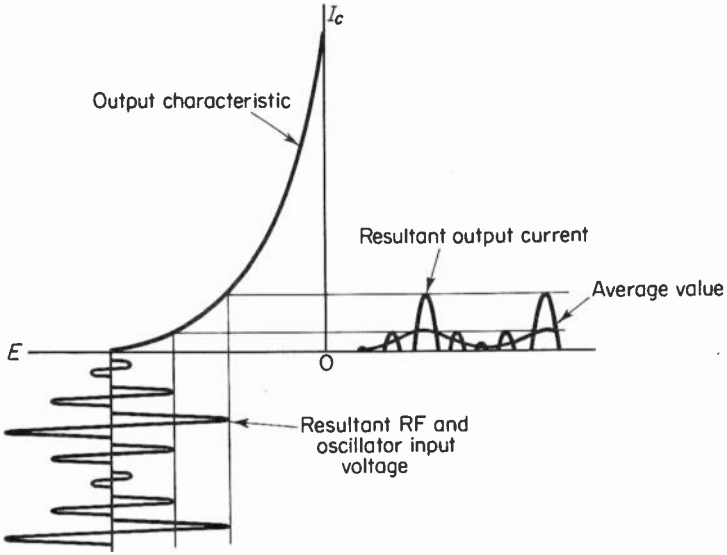



**Fig. 6-5.** The result of applying the combined oscillator and RF input signal to a non-linear device is illustrated in the above diagram.

output signal also contains other frequency components such as the original signals, harmonics of the original signals, and sum and difference frequencies. The desired difference signal is selected from the spectrum of output signals by means of a suitable filter, such as a double-tuned transformer or crystal filter.

The conclusions above may be arrived at mathematically. Thus, if the RF signal is represented by $E_s \cos \omega_s t$ and if the oscillator signal is represented by $E_0 \cos \omega_0 t$, and if the sum ($E_T$) of these input signals is operated over a linear portion of a transistor input-circuit volt-ampere curve, whose equation has the form

$$i_b = b + b_1 E_t, \tag{6-1}$$

then the input current of the transistor may be expressed as

$$i_b = b + b_1(E_s \cos \omega_s t + E_0 \cos \omega_0 t - E_1), \tag{6-2}$$

where $E_s \cos \omega_s t + E_0 \cos \omega_0 t$ is the sum of the input signals and $E_1$ is the DC voltage that determines the operating point of the device. From Eq. (6-2) it can be seen that the input current consists only of the original input signals.

If this current is applied to the input of an ideal NPN transistor whose input resistance between base and emitter is assumed to be linear,† but whose output characteristic is nonlinear and may be represented by

$$i_c = c + c_1 E_t + c_2 E_t^2, \tag{6-3}$$

†In actual transistor mixers the nonlinear input base-to-emitter junction behaves like a diode mixer and the output circuit is used as an amplifier.

where $E_t$ is the sum of the input signals, we obtain

$$\begin{aligned} i_C &= c + c_1(E_s \cos \omega_s t + E_0 \cos \omega_0 t - E_1) + c_2(E_s \cos \omega_s t \\ &\quad + E_0 \cos \omega_0 t - E_1)^2 \\ &= c + c_1(E_s \cos \omega_s t + E_0 \cos \omega_0 t - E_1) + c_2(E_s^2 \cos^2 \omega_s t + E_0^2 \cos^2 \omega_0 t \\ &\quad + 2E_sE_0 \cos \omega_s t \cos \omega_0 t - 2E_sE_1 \cos \omega_s t - 2E_0E_1 \cos \omega_0 t + E_1^2). \end{aligned} \tag{6-4}$$

By use of the trigonometric identities

$$\cos^2 \theta = \frac{1 + \cos 2\theta}{2} \tag{6-5}$$

and

$$\cos \phi \cos \theta = \frac{\cos (\theta + \phi) + \cos (\theta - \phi)}{2} \tag{6-6}$$

the following expression is produced:

$$\begin{aligned} i_C &= c + c_1E_s \cos \omega_s t + c_1E_0 \cos \omega_0 t - c_1E_1 + \frac{c_2E_s^2}{2} + \frac{c_2E_s^2 \cos 2\omega_s t}{2} \\ &\quad + \frac{c_2E_0^2}{2} + \frac{c_2E_0^2 \cos 2\omega_0 t}{2} + c_2E_1^2 + c_2E_sE_0[\cos (\omega_0 - \omega_s)t \\ &\quad + \cos (\omega_s - \omega_0)] - 2c_2E_sE_1 \cos \omega_s t - 2c_2E_0E_1 \cos \omega_0 t. \end{aligned} \tag{6-7}$$

This expression may be separated into its various terms, which represent the DC and AC components of output current. These are:

1. The DC component

$$c - c_1E_1 + c_2\left(\frac{E_s^2 + E_0^2}{2} + E_1^2\right) \tag{6-8}$$

2. The fundamental components

$$f_s \quad (c_1E_s - 2c_2E_sE_1) \cos \omega_s t. \tag{6-9}$$

$$f_0 \quad (c_1E_0 - 2c_2E_0E_1) \cos \omega_0 t. \tag{6-10}$$

3. Harmonic components

$$2f_s \quad \frac{c_2E_s^2}{2} \cos 2\omega_s t \tag{6-11}$$

$$2f_0 \quad \frac{c_2E_0^2}{2} \cos 2\omega_0 t \tag{6-12}$$

4. Sum frequency component, $f_0 + f_s$

$$c_2E_sE_0 \cos (\omega_0 + \omega_s)t. \tag{6-13}$$

5. Difference component, $f_0 - f_s$

$$c_2E_sE_0 \cos (\omega_0 - \omega_s)t. \tag{6-14}$$

The discussion above applies directly to the field-effect transistor, since the volt-ampere characteristics are almost perfectly square-law. If the output characteristic contains higher-order terms (cubic or higher), such as would be the case with vacuum-tube and transistor mixers, additional

frequency components will be produced. These components may give rise to undesired spurious responses. This problem is discussed in Chapter 5.

The discussion above assumed that the device (vacuum tube, transistor, or FET) was biased in a region where a nonlinearity existed. This point does not have to be at or near cutoff for frequency conversion to take place. If the device were perfectly linear except for being unilateral and biased near cutoff, it would also behave as a mixer. Under these conditions the device's characteristic may be represented by an infinite power series, which contains such terms as $c_2E_i^2$. Thus by similar reasoning frequency conversion may be shown to take place.

In the case of product mixing, each signal is applied to a different electrode of the converter. The output-current characteristic of each electrode may be perfectly linear and be represented by $i_c = c + c_1E_0$ and $i_c = b + b_1E_s$, respectively, but the product of these two signals will contain the term $c_1b_1\ E_0E_s$, which is equivalent to the $c_2E_i^2$ term of Eq. (6–3). Thus, frequency conversion will take place. As was pointed out earlier, this type of mixing is often referred to as linear mixing.

## 6-3. Conversion gain

Besides being a frequency changer, the mixer is also an amplifier. The exception is the diode mixer, which produces a conversion loss. The diode mixer's main advantage is its low noise. Conversion gain is defined for transistors in terms of power gain and for vacuum tubes and FET's in terms of voltage amplification. Thus for transistors the conversion gain is

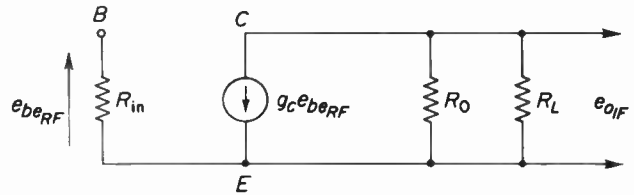
$$\text{C.G.} = \frac{\text{power output at the intermediate frequency}}{\text{power input at the radio frequency}}. \qquad (6\text{–}15)$$

In terms of vacuum tubes or FET's the conversion gain is

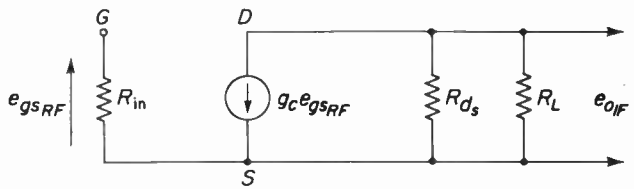
$$\text{C.G.} = \frac{\text{output voltage at intermediate frequency}}{\text{input voltage at radio frequency}}. \qquad (6\text{–}16)$$

The equivalent circuit of the mixer may be used to determine its conversion gain. The equivalent circuit of the mixer is of precisely the same form as that of an ordinary amplifier. Figure 6-6 shows the simplified equivalent circuits for a vacuum-tube, transistor, and FET mixer, which may be used for evaluating the conversion gain of a mixer. In each equivalent circuit is a constant-current generator which produces a current equal to $g_ce_{rf}$, where $g_c$ is the conversion transconductance of the device and $e_{rf}$ is the RF voltage across the input terminals of the mixer. The conversion transconductance can be used as a figure of merit for judging the frequency-changing ability of a given mixer. This is convenient, since the conversion gain of the mixer is the product of the conversion transconductance and the effective load impedance at the difference frequency. Thus, for a transistor mixer the conversion power


Vacuum tube equivalent circuit




Transistor equivalent circuit




Field effect transistor equivalent circuit

**Fig. 6-6.** The equivalent circuits of vacuum tube, transistor, and field effect transistor mixers.

gain may be determined from the simplified hybrid $\pi$ equivalent circuit to be

$$\text{C.G.} = g_c^2\left(\frac{R_0}{R_0 + R_1}\right)^2 R_L R_{in}. \tag{6-17}$$

The conversion gain of vacuum-tube and FET mixers may also be determined by use of their constant-current equivalent circuits. The expression for the conversion gain of a vacuum-tube mixer becomes

$$\text{C.G.} = g_c\left(\frac{r_p R_L}{r_p + R_L}\right). \tag{6-18}$$

The expression for conversion gain for an FET mixer is almost identical to that of the vacuum-tube mixer. It differs only in the designation of the output resistance ($r_{ds}$). Thus

$$\text{C.G.} = g_c\left(\frac{r_{ds} R_L}{r_{ds} + R_L}\right). \tag{6-19}$$

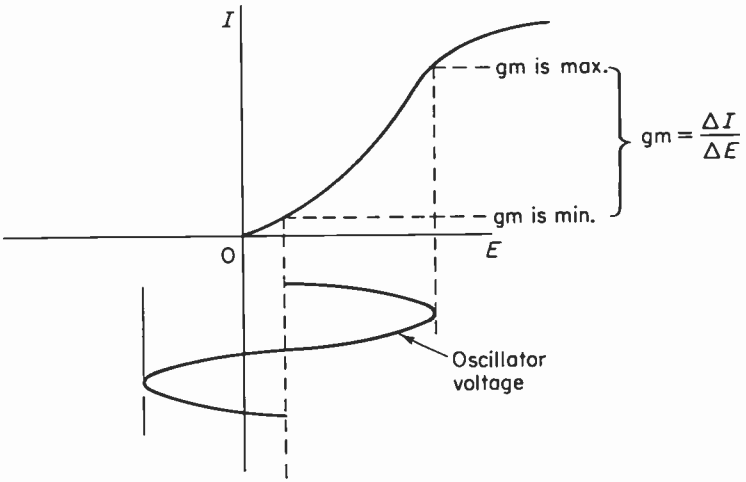



**Fig. 6-7.** The variation in transconductance resulting from large oscillator voltage excursions.

The magnitude of $g_c$ is determined under the assumption that the RF signal voltage is very much smaller than the oscillator voltage. Under these large-oscillator-voltage conditions, the transconductance of the device varies so that as the oscillator voltage sweeps the device into the region of cutoff the transconductance is minimum. See Fig. 6-7. In the region of maximum current, just before approaching saturation, the transconductance will be maximum. It is possible to determine the percentage of the maximum value of transconductance ($g_m$) that corresponds to the conversion transconductance ($g_c$). This may be accomplished by Fourier-analyzing the transconductance as a result of signal application to the device. The general equation derived by these methods for the conversion transconductance at the difference frequency is

$$g_c = \frac{1}{2\pi} \int_0^{2\pi} g_m \cos \omega_0 t \, d(\omega_0 t), \tag{6-20}$$

where $\omega_0$ = radian frequency of the oscillator. In this equation $g_m$ represents the instantaneous value of $g_m$. Thus, $g_c$ is maximum when the mixer bias is set near cutoff and the oscillator AC voltage is of sufficient amplitude to make the instantaneous value of $g_m$ as high as possible. The amplitude of the oscillator voltage is therefore chiefly responsible for the conversion gain of the mixer. Figure 6-8 shows how the conversion gain varies with a variation in oscillator voltage.

The conversion transconductance for a vacuum-tube mixer (triode, pentode or pentagrid converter) may be determined from Eq. (6-20) to be

$$g_c \approx 0.25 g_m, \tag{6-21}$$

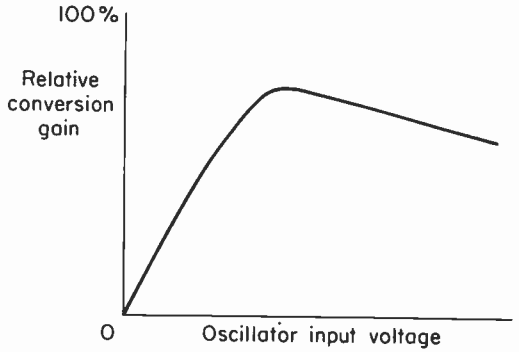



**Fig. 6-8.** The effect of oscillator voltage upon conversion gain.

where $g_m$ is the transconductance of the tube at zero bias ($g_m$ is maximum). This equation indicates that a given tube may provide almost four times more gain as an amplifier than as a mixer.

The conversion transconductance for a transistor mixer, assuming square-wave modulation of the collection current by the input signal, has been found to be

$$g_c \approx .4\, g_m \left( \frac{1}{1 + jr_{b'} \frac{\omega}{\omega_T} \frac{I_e}{27}} \right) \tag{6-22}$$

where

$r_{b'}$ = base spreading resistance,

$\omega$ = angular frequency of operation,

$\omega_T$ = angular frequency at which $h_{fe} = 1$,

$I_e$ = emitter current in ma,

$\frac{I_e}{27} = g_m$, also $\frac{I_e}{27\omega_T} = C_{b'c}$.

The parenthesized term of Eq. (6–22) indicates that the transconductance has associated with it a negative phase angle, and a magnitude-modifying factor. This means that in a common-base circuit where no phase shift might be expected between input and output, a phase shift will exist. If $r_{b'}$, $I_e$, or the ratio of $\omega/\omega_T$ is small, then the conversion transconductance will reduce to

$$g_c \approx 0.4 g_m. \tag{6-23}$$

Unfortunately, at 100 MHz a small $\omega/\omega_T$ ratio would require a transistor whose gain-bandwidth product ($f_t$) is in the range of 1000 to 10,000 MHz. Clearly, then, all other things being equal, Eq. (6–23) will be most useful at relatively low frequencies.

The conversion transconductance of an FET mixer may be found from the expression

$$g_c \approx 0.25 g_m, \tag{6-24}$$

where $g_m$ is the transconductance at zero bias. This equation is based on the assumptions that the input signal voltage is very small and that peak oscillator voltage is biased at and is equal to one-half the drain-current cutoff voltage (pinch-off).

If the oscillator voltage is increased in amplitude to the point where the drain current is a square wave modulated at the oscillator frequency, the conversion transconductance of the FET becomes

$$g_c \approx 0.4 g_m. \tag{6-25}$$

One important disadvantage of increasing the oscillator voltage by a large amount would be the increased spurious responses that would result. This can be attributed to the departure of mixer operation from the almost ideal square-law characteristics of the FET.

## 6-4. Mixer noise

An important source of noise in a receiver is the frequency changer. As a noise source it is capable of from 2 to 100 times the noise output of the same tube, transistor, or FET used as a Class A amplifier. This increased noise level can be attributed to both the mixer input circuit and the mixing device itself.

The noise level developed at the ouput of the mixer depends upon whether or not the input to the mixer is tuned or is preceded by a tuned RF amplifier. If the mixer input is not tuned, then those noise components due to the input circuit and associated with the image frequency and other spurious responses will add to the signal-frequency channel noise, thereby increasing the overall noise level. If the mixer is preceded by a tuned amplifier whose selectivity is sufficient to minimize image and spurious response, a decided reduction in noise level will result. For example, if it is assumed that the signal channel and the image channel contribute equal noise levels, then just the elimination of the image response from a receiver that is free of other spurious responses will reduce the noise level by 3 db.

In addition to the noise introduced into the mixer by its input circuit, the local oscillator also contributes noise to the mixer. Since the bandwidth of the oscillator is very narrow, the noise contribution is quite small and is usually disregarded.

Mixers generate noise by the same means as do amplifiers; however, the mixer output noise will be the result of the signal-channel noise as well as all of the spurious responses of which the mixer is capable. Thus the mixer is to be expected to be more noisy than the same device used as an amplifier.

As pointed out earlier, it is of decided advantage to compare noise performance of amplifiers and mixers in terms of an equivalent noise resistance ($R_n$). The equivalent noise resistances for FET, triode, and pentode mixers and pentagrid converters that generate noise at the intermediate frequency are as follows:

FET mixer:

$$R_n = \frac{2}{g_c}. \tag{6-26}$$

Triode mixer:

$$R_n = \frac{4}{g_c}. \tag{6-27}$$

Pentode mixer:

$$R_n = \frac{I_b}{I_b + I_c}\left(\frac{4}{g_c} + \frac{20I_{c_2}}{g_c^2}\right). \tag{6-28}$$

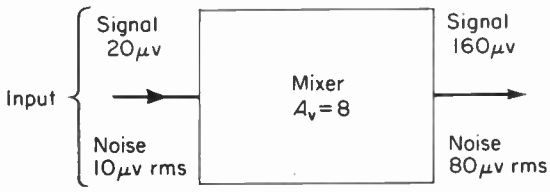
Pentagrid converter:

$$R_n = \frac{20I_b}{g_c^2 I_k}(I_k - I_b). \tag{6-29}$$

In these formulas the values of conversion transconductance ($g_c$), plate current ($I_b$), screen current ($I_{c_2}$), and cathode current ($I_k$) are averaged over a complete cycle of oscillator voltage.

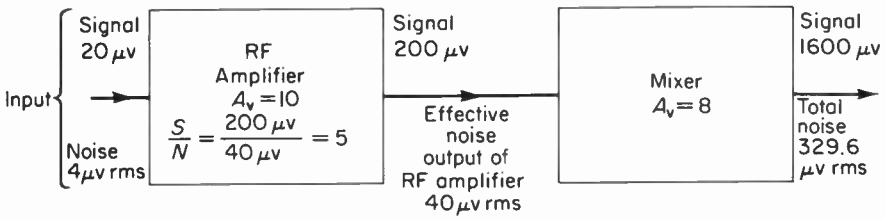
From the equations above we can see that the FET mixer will contribute less second-stage noise than any other mixer listed. Comparison of the equivalent noise resistances also indicates that the fewer the current-carrying elements of a vacuum-tube mixer, the less noise the mixer will generate. Thus the most noise-free vacuum-tube mixer is the triode, and the noisiest is the pentagrid converter. It should be borne in mind that if the mixer is preceded by a tuned RF amplifier that generates less noise than the mixer, the noise contribution of the mixer to the overall system will be negligible and may be disregarded. This improvement in signal-to-noise ratio can be shown by a numerical example that compares the signal-to-noise ratio of a front end with and without an RF amplifier. Consider a receiver that does not incorporate an RF amplifier and whose mixer's conversion gain is 8. See Fig. 6-9. Assume that the input signal is 20 $\mu$v and that the internally generated noise at the output of the mixer is 80 $\mu$v rms. Thus the output of the mixer will consist of a signal of 160 $\mu$v and a noise voltage of 80 $\mu$v rms. The signal-to-noise ratio at the output of the mixer is therefore 160 $\mu$v/80 $\mu$v = 2.

The effect of preceding the mixer with an RF amplifier upon the overall noise figure may be seen in Fig. 6-9(b). If the RF amplifier is assumed to generate only half of the noise generated by the mixer, the noise output of the RF amplifier will be 40 $\mu$v rms. Furthermore, if the amplification of the RF amplifier is 10, the same input signal as applied to the input of the mixer




$$\frac{S}{N} = \frac{160}{80} = 2$$

(a)




$$\text{Total noise output of mixer} = \sqrt{(\text{Noise due to } V_1)^2 + (\text{Noise due to } V_2)^2}$$

$$= \sqrt{(320)^2 + (80)^2}$$

$$= 329.6\,\mu v \text{ rms}$$

$$\text{Overall } \frac{S}{N} = \frac{1600\,\mu v}{329.6\,\mu v} = 4.86$$

(b)

**Fig. 6-9.** This figure illustrates the manner by which the RF amplifier improves the $S/N$ ratio of a receiver.

in the previous case will produce a signal output of 200 $\mu$v. The signal-to-noise ratio at the output of the RF amplifier, without considering the mixer noise contribution, is 200 $\mu$v/40 $\mu$v rms = 5. At the output of the mixer will appear an amplified signal of 1600 $\mu$v and the resultant noise of both the RF amplifier and the mixer. These noise voltages must not be directly added but rather combined by the square root of the sum of the squares. Since the amplified noise due to the RF amplifier at the output of the mixer is 320 $\mu$v rms (8 $\times$ 40 $\mu$v rms = 320 $\mu$v rms) and the mixer noise is 80 $\mu$v rms, the total noise voltage will be 329.6 $\mu$v rms = $\sqrt{(320 \times 10^{-6})^2 + (80 \times 10^{-6})^2}$. This results in an overall signal-to-noise ratio of 1600 $\mu$v/329.6 $\mu$v rms = 4.86. Since the RF amplifier has a signal-to-noise ratio of 5, the effect of the mixer upon the overall signal-to-noise ratio is quite small. Thus, the introduction of the mixer as a noise source results in a degradation of signal-to-noise ratio of only 2.8 per cent. Furthermore, the signal-to-noise ratio with the RF amplifier is more than 100 per cent greater than the signal-to-noise ratio of the mixer alone.

Transistor converters may be either of the common-base or common-emitter circuit configuration. A noise performance comparison by means of equivalent noise resistance is not usually done with transistors, because the equivalent noise resistance is only valid if there is no interaction between the various noise sources and if no feedback exists. Therefore transistor amplifiers and converters are compared by means of noise figure, which is derived from an equivalent circuit containing at least two noise sources. The principal sources of noise in a common-base converter are thermal noise from the base resistance $r_b$, and shot noise from the emitter-base diode. Shot noise from the collector-base diode is considered negligible in its effect upon common-base converter noise figure. The common-emitter converter also has noise introduced by the base resistance and the emitter-base diode, but in addition the collector shot noise makes a contribution. Despite this, Webster† has concluded that while the noise figures of common-base and common-emitter amplifiers are the same, the noise figure of the common-emitter converter is less than the noise figure of the common-base converter. This is the result of higher common-emitter current gain, which by means of the signal source impedance, tends to further the reduction of higher conversion frequency noise.

The superiority of the common-emitter converter is only valid for large values of source resistance. As the source resistance is decreased, the noise figures of both types of converters converge at a relatively high noise figure as compared to the same transistor used as an amplifier.

## 6-5. Vacuum-tube converters

The vacuum-tube converter is a multigrid tube that combines the functions of oscillation and frequency changing in one envelope. The most commonly used converter is the pentagrid converter. A cross-sectional view of a pentagrid converter is shown in Fig. 6-10. Included in this figure is a schematic diagram of a typical circuit that employs the pentagrid converter. From these diagrams it is to be noticed the first two grids form a triode and that the grid (G1) closest to the cathode is used as the grid of an oscillator. The next grid (G2) is returned to the B supply and is used as the plate of the local oscillator. Under normal operating conditions the oscillator section of the tube will modulate the electron stream on its way to the plate. Since grid number three, (G3) the RF signal grid, is usually negative with respect to the cathode, the electron stream will slow down and form a vertical cathode, which pulsates at the oscillator frequency immediately in front of grid 3.

†R. R. Webster, "The Noise Figure of Transistor Converters," *Trans. IEEE PGBTR*, November 1961.

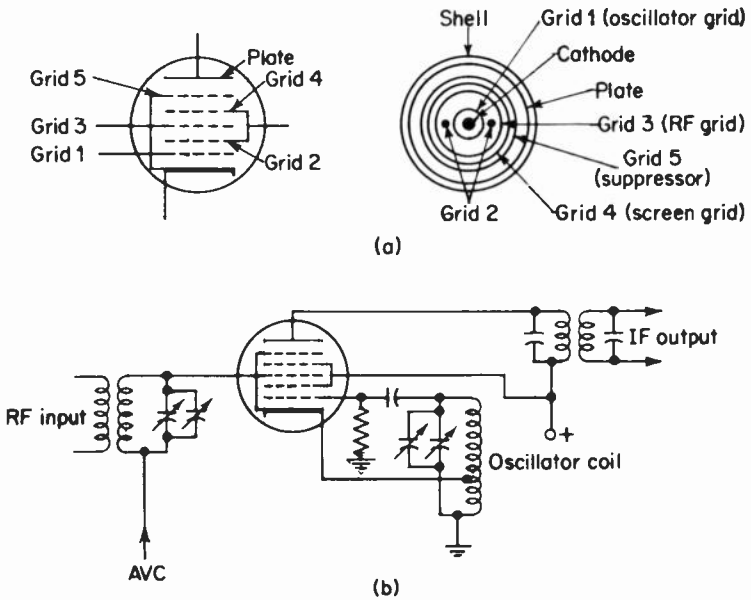



**Fig. 6-10.** The schematic symbol and cross-sectional view of a pentagrid converter is shown in (a). The schematic diagram of a typical pentagrid converter circuit is drawn in (b).

Grid 3 and the other remaining tube elements (screen, suppressor, and plate) constitute a pentode, which is used as a frequency changer. The method of frequency changing used by the pentagrid converter is multiplicative, since the tube's electron current flow is controlled by two sources—the oscillator grid variations and the RF signal input changes.

Unfortunately, the pentagrid converter has three important limitations at very high frequencies: (1) decreased conversion gain as the frequency increases, (2) severe oscillator pulling, and (3) high noise level.

The reduced conversion gain and oscillator pulling are due to undesired coupling between the oscillator grid and the RF signal grid.

The reduction in conversion gain is due to a form of electrostatic coupling known as space-charge coupling. In this form of coupling the pulsating virtual cathode formed in front of G2 induces a voltage, at the oscillator frequency, into the RF signal input grid. In terms of the oscillator grid voltage, space-charge coupling can be thought of as a negative capacitance between oscillator grid and RF signal grid.

The impedance of the tuned circuit connected to the RF grid is capacitive, since the oscillator frequency is assumed to be higher than the RF signal frequency. Thus the simplified equivalent circuit due to space-charge coupling, shown in Fig. 6-11, consists of a generator operating at the oscillator frequency, and a circuit consisting of the negative capacitance in series with the

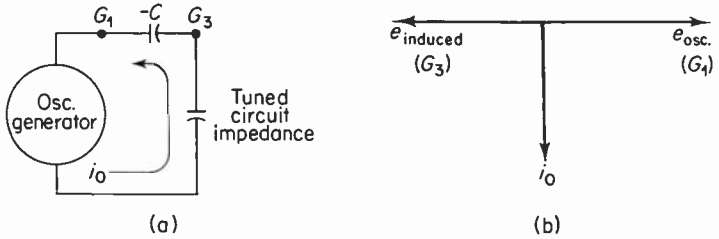



**Fig. 6-11.** The equivalent circuit of a pentagrid converter between grids $G_1$ and $G_3$, which may be used to indicate the effects of space charge coupling, is shown in (a). The phasor diagram in (b) shows the phase relationships which exist in the circuit of (a).

capacitive impedance of the RF signal grid tuned circuit. If it is assumed that the reactance of the negative capacitance (same volt-ampere phase relationships as an inductance) is very much larger than the capacitive resistance of the RF signal grid tuned circuit, the phase relationship between the oscillator grid voltage and the induced oscillator voltage, appearing on the RF signal grid, will be that shown in the phasor diagram of Fig. 6-11(b). From the phasor diagram it can be seen that the induced oscillator voltage on grid 3 is 180 degrees out of phase with the grid 1 oscillator voltage. This tends to cancel the effectiveness of the local oscillator in determining the conversion gain of the converter. The effect of space-change coupling becomes worse as the frequency of operation increases. This can be overcome to some extent by placing a capacitor of the proper value between grids 1 and 3. This positive capacitance will tend to cancel the negative capacitance due to space-charge coupling and thus will increase the conversion gain at high frequencies. Design changes in the structure of the pentagrid converter have also been effective in minimizing the effects of space-charge coupling.

The pentagrid converter is subject to oscillator pulling because of feedback from the RF grid to the oscillator grid via the interelectrode capacitance of the tube. This becomes more severe as the frequency of operation is increased, because the percentage difference between the oscillator frequency and the RF signal frequency becomes smaller; at the same time, increasing the frequency of operation causes the reactance of the interelectrode capacitance to decrease, causing tighter coupling between the grids. The degree of oscillator pulling can be substantially reduced by introducing special structures such as shields and by arranging the grid structure of the pentagrid converter to minimize coupling between the grids.

As has been pointed out earlier, the pentagrid converter is one of the noisiest types of converters. Thus it is not suitable for use in weak-signal receivers unless an RF stage is used. Pentagrid converters alone have been used in lower cost low-sensitivity receivers. However to avoid the limitations of the pentagrid converter, separate mixer and local oscillators are used.

## 6-6. Triode and pentode mixers

The triode mixer is characterized by relatively low noise and fairly high gain. However, depending upon the method of oscillator injection the triode mixer may have a tendency for oscillator pulling as well as oscillator radiation. Figure 6-12 shows two types of triode mixer circuits. In Fig. 6-12(a) the oscillator signal is injected along with the RF signal at the grid; in Fig. 6-12(b) cathode injection is used. In either case oscillator injection may be accomplished with either inductive or capacitive coupling. However, in Fig. 6-12(a) capacitive coupling is used and in 6-12(b) inductive coupling via a transformer is employed. As far as performance of the mixer is concerned, both methods of oscillator injection provide the desired frequency conversion. Unfortunately, grid injection results in oscillator pulling and increased oscillator radiation, since a more direct path from the oscillator to the antenna is provided, and a variation in converter gain occurs during tuning. The last-mentioned effect is due to a variation in converter gain occurs during tuning. The last-mentioned effect is due to a variation in conversion transconductance. This is the result of changes in oscillator injection voltage, corresponding to the changes in voltage division between the impedance of the mixer input and the coupling capacitor of the oscillator. Cathode injection of oscillator voltage limits oscillator radiation and minimizes oscillator pulling, but since the cathode input impedance "seen" by the oscillator is relatively low

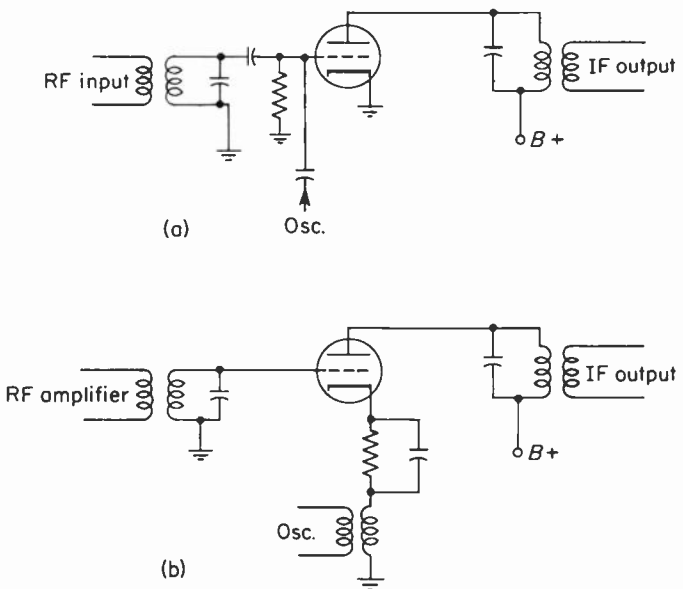



**Fig. 6-12.** Two triode mixers with different methods of oscillator injection. In (a) the oscillator is injected via the grid whereas in (b) oscillator injection is via the cathode.

($\approx 1/g_m$), the mixer will tend to load the oscillator unless transformer matching is used. Furthermore, since the oscillator is tuned to a higher frequency than the RF circuit, the oscillator circuit in comparison to the RF circuit will appear to be inductive.

It was pointed out earlier that an inductive cathode tends to introduce into the input circuit a resistive component of loading that will load the input tuned circuit. Thus cathode oscillator injection is seldom used in practice, since loading of the input tuned circuit of the mixer will reduce the image-rejection abilities of the tuned circuit as well as reducing the overall gain.

The input circuit loading due to an inductive cathode is in addition to other loading phenomena, such as transit-time loading and Miller effect. Transit-time loading is not as severe in a mixer as in the same tube used as an amplifier, because of the reduced transconductance of the mixer. If the capacitive impedance of the IF tuned circuit in the plate of the mixer is appreciable at RF signal frequencies, the Miller effect will cause severe detuning and resistive loading of the mixer grid tuned circuit.

The triode mixer is normally stable, since the mixer output frequency is much lower than that of the input RF signal. Thus, assuming that the input impedance is very small at the intermediate frequency, any feedback at the IF will not be regenerative. Nevertheless, triode mixers are often neutralized or modified so as to increase the input resistance of the mixers. This, of course, will improve selectivity and increase the overall mixer gain. One method of increasing input-circuit shunt resistance is to make the plate circuit of the mixer inductive at the RF signal frequency by placing a small coil in series with the plate of the mixer. This coil will tend to make the plate circuit inductive, thereby introducing a negative resistance into the grid circuit. This, in turn, will in effect increase the shunt resistance in the grid circuit.

An example of a pentode mixer is shown in Fig. 6-13. The pentode mixer is characterized by high gain and a higher noise level than the triode mixer. The higher noise level is of little importance except in weak-signal receivers. The higher overall gain of the pentode is due to its high transconductance and the reduction in input-circuit loading. The reduction in input-circuit loading results from the small interelectrode capacitance between plate and

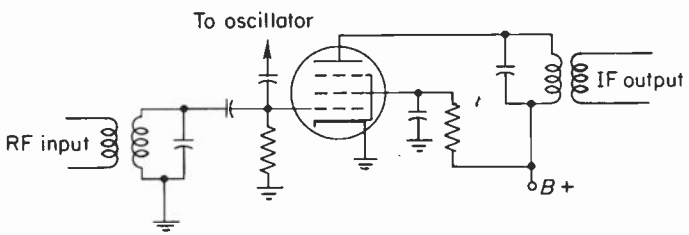



**Fig. 6-13.** A pentode mixer which employs control grid oscillator injection.

grid, which minimizes the Miller effect. Oscillator injection into the cathode circuit is usually avoided, since this would result in resistive loading of the input circuit. Thus, pentode mixers are poorer than triodes with regard to noise but are superior to triode mixers in terms of gain and input loading, and have characteristics like the triode in terms of oscillator radiation and pulling when oscillator injection is applied to the grid.

Both triode and pentode mixers usually use grid-leak bias to set the DC operating point of the tube, because grid-leak bias is self-compensating. For example, if the oscillator voltage should increase and thereby tend to increase the conversion gain, the grid-leak bias will increase, tending to reduce the transconductance of the mixer. Thus the conversion gain remains constant under conditions of oscillator variation.

## 6-7. The vacuum-tube autodyne converter

An autodyne converter is simply a self-oscillating mixer. Thus, all circuit elements used in generating the local oscillator signal also take part in the mixing action. A basic form of the autodyne converter is shown in Fig. 6-14. In this circuit the oscillator portion of the mixer is formed by $L_1$ and $L_2$, where $L_1$ and $C_1$ provide positive feedback to the grid circuit and $L_2$ and its shunt capacitors $C_4$ and $C_6$ form the oscillator grid tank circuit. The RF input from the RF amplifier to the converter is fed to the grid of the mixer via a balanced bridge. The bridge, shown in Fig. 6-14(b), is made up of four circuit and distributed capacitances. One of the arms of the bridge consists of $C_4$ and the interelectrode capacitance of the tube $C_{gk}$. The other arm consists of $C_4$ and a composite capacitance designated $C_x$. $C_x$ is formed by the series combination of $C_2$ and $C_5$. If this bridge is balanced, then the bridge load $L_2$ will be isolated from the RF input voltage. Thus oscillator pulling will be minimized. Furthermore, since the oscillator voltage across $L_2$ "sees" the RF source as the load of the same balanced bridge, little or no oscillator voltage will be coupled into the RF circuit. Therefore oscillator radiation will be reduced.† Some of the components used in this bridge may be used to form a second bridge, which is used to isolate the input circuit from the output circuit and thus insure mixer stability. This bridge is shown in Fig. 6-14(c). The bridge is balanced by constructing one arm of the interelectrode capacitance $C_{gp}$ and a series arrangement of $C_4$, $C_5$, and $C_6$; the other arm is composed of $C_2$ in series with the parallel combination of $C_3$ and the tube's interelectrode capacitance from plate to cathode, $C_{pk}$. The bridge load in this case is the grid-to-cathode circuit of the mixer. When this bridge is

†This refers to oscillator radiation from the antenna and does not take into account oscillator radiation from the chassis itself. This is a function of shielding and adequate bypassing.

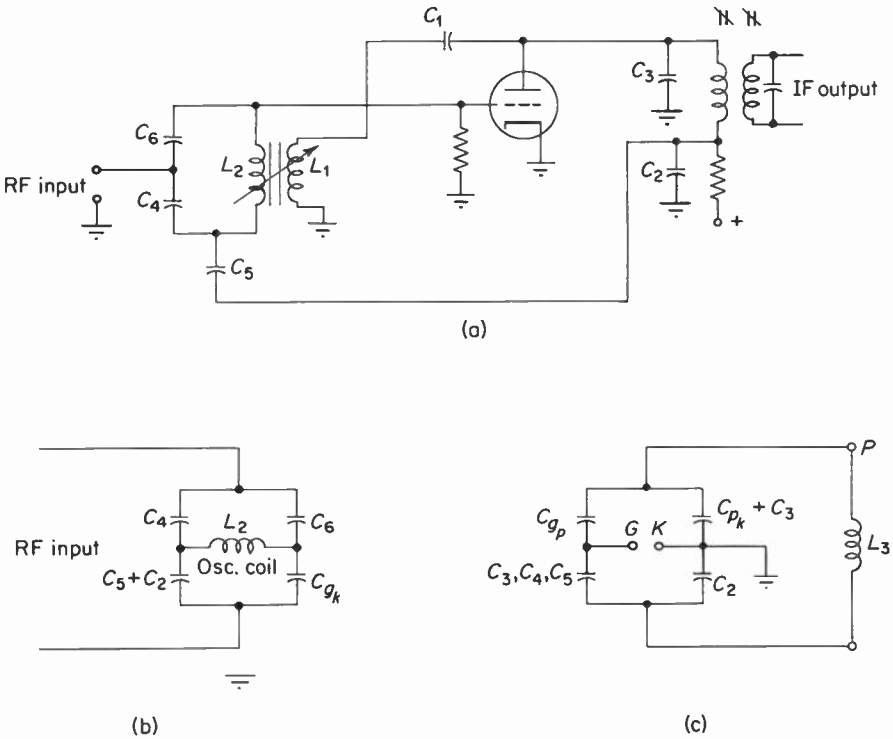


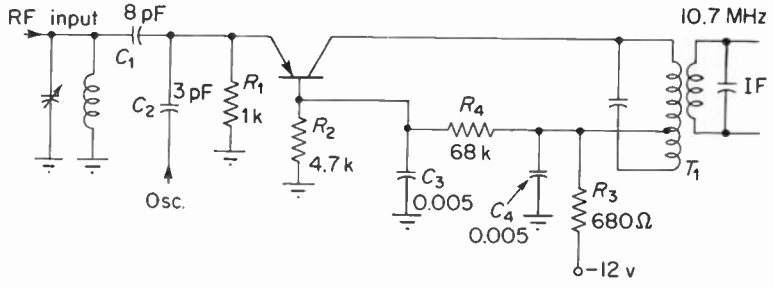

**Fig. 6-14.** A typical schematic diagram of an autodyne converter is shown in (a). (b) and (c) illustrate the bridge circuits found in the converter which minimize undesired coupling.

balanced, the circuit is stable at the IF frequency. Aside from the obvious economic advantages of a self-oscillating mixer, the autodyne converter is capable of the gain and noise figure of a triode mixer.

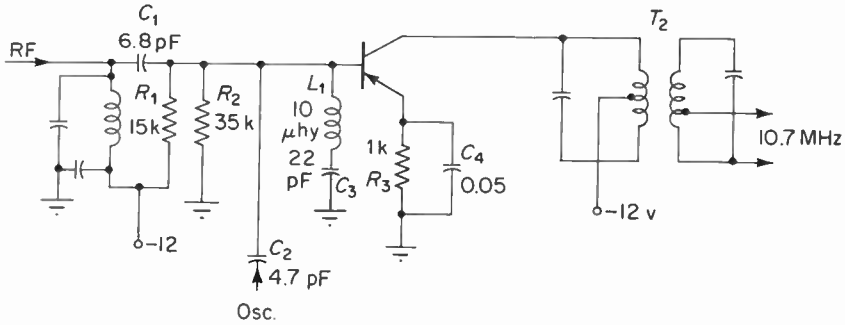
## 6-8. Transistor mixers

Three basic transistor mixer circuit arrangements commonly found in FM receivers are the common-base mixer, the common-emitter mixer, and the self-oscillating transistor mixer, which is also referred to as a converter. Schematic diagrams of these three circuits are shown in Fig. 6-15. In Fig. 6-15(a) both oscillator and RF signals are injected at the emitter of the common-base mixer. The IF is coupled to the next stage by means of a double-tuned transformer, $T_1$. $R_1$, $R_2$, and $R_3$ form a DC voltage divider, which sets the operating conditions of the mixer. $C_3$ and $C_4$ are RF bypass capacitors, which insure that the base of the transistor and the power supply are both at AC ground.

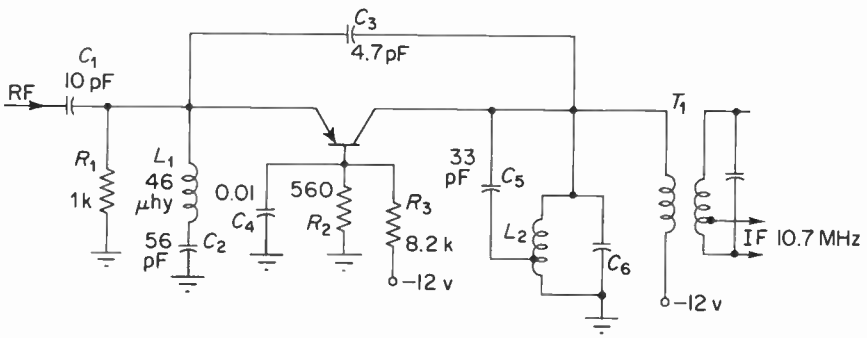
The common-emitter mixer of Fig. 6-15(b) has the RF and oscillator signals injected at the base of the mixer. The RF coupling capacitor $C_1$ is also used as a matching capacitor, which provides a match between the RF amplifier and the input resistance of the mixer. The operating point of the




Common base mixer
(a)




Common emitter mixer
(b)




Transistor converter
(c)

**Fig. 6-15.** Three types of transistor frequency changers.

mixer is determined by resistors $R_1$, $R_2$, and $R_3$. $C_4$ is a bypass capacitor across the emitter resistance, which prevents degeneration and thus a reduction in conversion gain. $C_5$ is a bypass capacitor across the power supply, which owing to its low reactance prevents any high-frequency AC from forming across the capacitor, thereby preventing feedback through the power supply. The intermediate frequency developed by the converter is coupled to the first IF amplifier by means of the double-tuned transformer $T_2$. The primary and secondary of $T_2$ are tapped as a means of matching the output resistance of the transistor and the input resistance of the IF amplifier. $L_1$ and $C_3$ form a series resonant trap that makes the input impedance of the mixer almost zero at the intermediate frequency.

The self-oscillating converter of Fig. 6-15(c) is similar to the common-base mixer of Fig. 6-15(a) insofar as DC operating conditions are concerned. Also similar are the input and output coupling circuits. The major difference is in the simultaneous use of $C_3$, $C_5$, $L_2$, and $C_6$ to form the coupling and tuned-circuit elements of a Hartley oscillator.

In each of the above the mixing action takes place in the emitter-base diode of the transistor. The remaining portion of the transistor can be considered to be an IF amplifier. In the case of the converter the remaining portion of the transistor behaves like both an IF amplifier and the local oscillator.

The common-base mixer is characterized by low conversion gain, a poor noise figure, and low input impedance, which loads both the oscillator and the RF amplifier. Conversely, since the common-emitter mixer is noted for its high conversion gain, low noise figure, and higher input impedance, it is the most frequently used transistor mixer in FM receivers.

Oscillator injection to the common-emitter mixer may be accomplished by inductive or capacitive means either to the emitter or, as shown in Fig. 6-15(b), to the base of the transistor. Base injection is preferred at very high frequencies because fewer stability problems are encountered, and because design of the emitter injection circuits may be difficult. This is because the emitter resistance that helps set the DC operating point of the transistor is usually bypassed for a wide range of signals, which also include the local oscillator signal. If the oscillator is to be injected at the emitter by capacitive coupling, a network must be designed that will be a short circuit for the input frequency and thus avoid degeneration, and at the same time at the local oscillator frequency present a conjugate impedance of the oscillator source impedance. This can become a difficult design problem, since the oscillator and RF signals differ by approximately 10 per cent for an IF of 10.7 MHz.

Assuming that an optimum oscillator injection level has been established, maximum conversion gain can be obtained by conjugate matching of the input and output circuits of the mixer. Proper matching may be had by the use of tapped transformers or by matching capacitors such as $C_1$ of Fig. 6-15(a) and (b). The value of the matching capacitor must be such that the real part

of the load admittance is made equivalently equal to the real part of the transistor output admittance. Figure 6-16 shows the simplified circuit corresponding to the output load of the transistor and its matching capacitor $C_m$.

If it is assumed that all of the reactances in the coupling circuit are lossless, the value of the matching capacitor ($C_m$) may be determined in the following manner. The input power to the interstage circuit ($P_{in}$) must equal the output power ($P_L$) delivered to the load. Thus

$$P_{in} = \frac{e_{in}^2}{R_0} = P_L = \frac{e_0^2}{R_L} = i^2 R_L$$

Assuming that $C_m$ is small and thus $X_{c_m} \gg R_L$, it can be seen that

$$e_{in} \approx i X_{c_m}$$

Hence

$$\frac{i^2 X_{c_m}^2}{R_0} = i^2 R_L$$

$$\frac{X_{c_m}^2}{R_0} = R_L$$

Then

$$\frac{1}{\omega^2 C_m^2 R_L} = R_0$$

Solving for the value of the matching capacitor

$$C_m = \frac{1}{\omega \sqrt{R_L R_0}} \tag{6-30}$$

Transistor amplifiers and mixers are both subject to feedback between output and input circuits. Feedback from the output of the mixer to the input at the IF frequency can result in instability and oscillation. Instability of this sort can result in increased spurious response. To avoid such feedback, neutralization schemes such as those discussed in conjunction with vacuum-tube autodyne converters may be resorted to. Often the components that make up these bridges serve dual purposes or may use some of the distributed

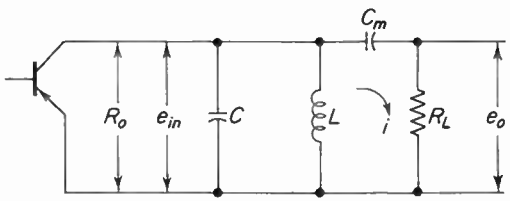



**Fig. 6-16.** A simple high frequency interstage circuit in which the coupling capacitor $C_m$, provides the means for an impedance match.

capacitances of the circuit, and therefore the neutralizing bridge may not be obvious from examination of the schematic diagram. Another method used as a means of eliminating the effects of feedback at the intermediate frequency is shown in Fig. 6-15(b), and consists of a series resonant circuit ($L_1C_1$) tuned to the IF frequency placed between the emitter and ground. This method is often employed when the input impedance at the IF is fairly large owing to the use of relatively high-reactance matching capacitors. Hence adding the series resonant trap tuned to the intermediate frequency will reduce the impedance at the intermediate frequency to almost zero. As a result, feedback at the intermediate frequency from the output to the input circuit is shorted out and has no effect upon circuit operation. In fact, owing to the elimination of the effects of feedback upon the mixer, greater conversion gain is possible.

Transistor converters for use at FM frequencies are usually of common-base configuration for both the oscillator and mixer portions of the circuit. Figure 6-17 shows how the converter of Fig. 6-15(c) may be redrawn to illustrate the oscillator and mixer sections separately. Notice that the mixer portion of the converter is almost identical to the common-base mixer of Fig. 6-15(a). It differs only in the use of a series resonant circuit in the emitter,

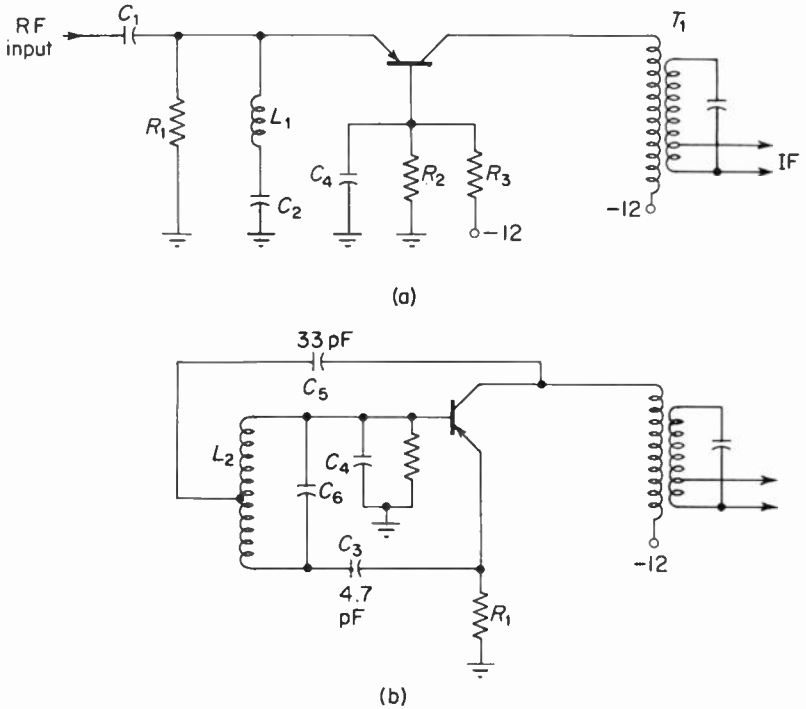



**Fig. 6-17.** The above diagrams show Fig. 6-15(c) redrawn to indicate how the transistor converter consists of a mixer (a) and a Hartley oscillator (b).

the purpose of which has been explained earlier. The oscillator section of the converter is redrawn in Fig. 6-17. In this case the oscillator is a modified Hartley. The capacitors $C_2$ and $C_3$ serve as coupling capacitors, and provide the necessary feedback paths to sustain oscillation. This Hartley is somewhat unusual in that the tap is returned to the collector rather than the emitter.

## 6-9. Field-effect transistor mixers

A typical example of a field-effect transistor used as an additive mixer is shown in Fig. 6-18. In this circuit the RF signal is coupled via a tuned matching transformer $T_1$ to the gate of an N-type junction FET. Reverse bias for the junction FET is obtained by means of the self-bias resistor $R_1$. $C_1$ is a bypass capacitor that eliminates the degenerative feedback voltage across $R_1$, thereby increasing the conversion gain of the mixer. As a means of reducing coupling between RF and oscillator circuits to a minimum, the oscillator is coupled to the source of the FET by means of a step-down transformer. The transformer is used to match the relatively low source impedance to the impedance offered by the oscillator. Owing to the inductance added to the

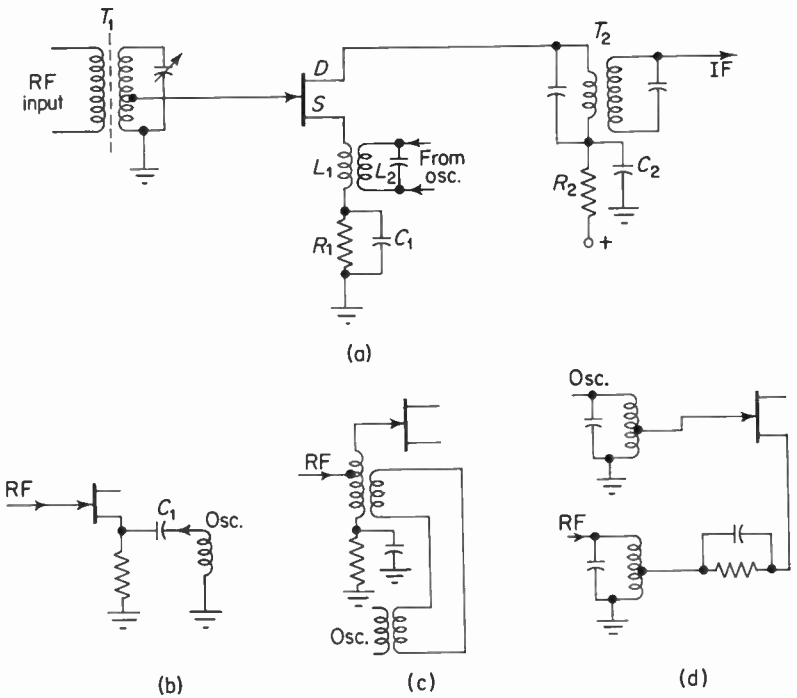



**Fig. 6-18.** The schematic diagram of an FET mixer is shown in (a). The diagrams labeled (b), (c), and (d) show different methods of RF and oscillator injection which may be used with FET mixers.

source, and the feedback due to the interelectrode capacitance between gate and source, this method of injection will tend to introduce a resistive component of loading across the input tuned circuit. This resistive component is part of the input load presented to the matching transformer.

Other methods of injecting the oscillator voltage are shown in Fig. 6-18(b) and (c). In 6-18(b) the oscillator is coupled into the source of the FET by means of a coupling capacitor. The source resistor is left unbypassed; thus the AC component of drain current at the RF frequency will develop a voltage across it. This voltage constitutes undesired feedback, which will tend to reduce conversion gain. Degenerative feedback across $R_1$ due to the components of drain current at the IF frequency will not affect conversion gain, since the coupling capacitor $C_1$ returns the source to ground, via the very low impedance of the oscillator tank coil ($L_1$) at the intermediate frequency.

In Fig. 6-18(c) the oscillator voltage is coupled into the gate and the RF signal is coupled to the source. In series with the source is a bypassed self-bias resistor, which helps set the operating point of the mixer. Since the impedance of the bias network is very small for all frequency components of the drain current, it can be ignored in its effect upon AC circuit operation. The output circuit of the FET mixer consists of a double-tuned matching transformer tuned to the intermediate frequency.

The FET mixer is usually characterized by moderately high conversion gain (somewhat less than a bipolar transistor) and, most important, it exhibits better spurious response-rejection than other types of mixers. The ability of the FET to reject spurious responses is in part due to its relatively high input impedance, which allows the use of highly selective tuned circuits. Most important with regard to spurious response is the FET's square-law transfer characteristic. As long as it is not overloaded, and thereby driven into regions of third and fourth-order curvature, the FET will only generate sum and difference frequencies for those signals which appear at its input. From the above it can be seen that if low spurious response is to be obtained, then the level of the signal reaching the mixer and the oscillator peak-to-peak amplitude must be limited. Also the operating bias must be carefully selected. It has been determined that if the peak-to-peak oscillator voltage is limited to one-half pinch-off voltage ($I_d = 0$) and if the bias is maintained at one-half pinch-off voltage, optimum conditions will exist.

Another factor for obtaining improved spurious-response rejection is effective automatic gain control of stages preceding the mixer. The mixer itself is exempted from the application of AGC, as a means of avoiding oscillator detuning. Effective AGC is necessary as a means of preventing overloading of both the RF amplifier and the mixer stages due to strong signals. Thus spurious responses resulting from cross-modulation or other higher-order products are minimized.

## 6-10. Integrated-circuit mixers

Integrated circuits may be used in place of discrete components in any of the small-signal stages of an FM receiver. Thus, integrated circuits can be used in the RF amplifier, mixer, oscillator, IF amplifier, limiter, detector, and other stages of the receiver. One integrated-circuit chip may include one or more stages, hence introducing economy while at the same time providing excellent performance and improved reliability. Since the internal construction of the usual integrated circuit is largely limited to the use of passive resistive elements and such active devices as transistors; capacitors and the inductances necessary for tuned circuits are usually external to the integrated chip.

For ease of integrated circuit design *Y*-parameters may be used. This information is usually supplied by the manufacturer and is useful for determining power gain, input and output admittances, stability, and voltage gain. The manufacturer's data concerning *Y*-parameters of a given operational mode are usually in the form of graphs, which show how the conductance and susceptance of a given *Y*-parameter vary as the frequency of operation is varied, or as the operating current is varied. An example of this type of graph is shown in Fig. 5-34.

An example of an integrated circuit that is useful as an amplifier, oscillator, mixer, or limiter is shown in Fig. 6-19. Transistors $Q_1$ and $Q_2$ are arranged so that with the proper terminations they can be used as a single-stage differential amplifier. $Q_3$ is a transistor that is used to provide a constant-current source for the differential amplifiers. By changing the integrated circuit terminations, the constant-current source in conjunction with one-

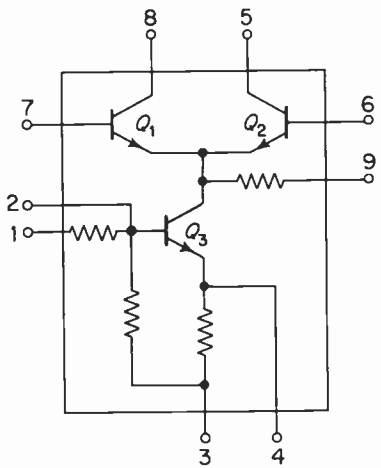



**Fig. 6-19.** The internal schematic diagram and external terminals of an integrated circuit.

half of the differential amplifier may be used as the common-emitter input stage of a cascode amplifier. The half of the differential amplifier used in this combination is arranged as a common-base amplifier. Since integrated circuits may be used from DC to many hundreds of megahertz, the differential or cascode modes of operation may be used as an RF amplifier in the front end of an FM receiver.

It is also possible to use the integrated circuit of Fig. 6-19 as a mixer with separate oscillator input, or as a self-oscillating converter. Examples of these circuit arrangements are shown in Fig. 6-20. In Fig. 6-20(a) is shown the schematic diagram of an integrated-circuit mixer. In this circuit $Q_3$ is used as a buffer amplifier which couples the oscillator signal into a balanced differential mixer. This buffer amplifier minimizes oscillator pulling under strong-signal conditions. The RF signal is fed into the mixer by means of a

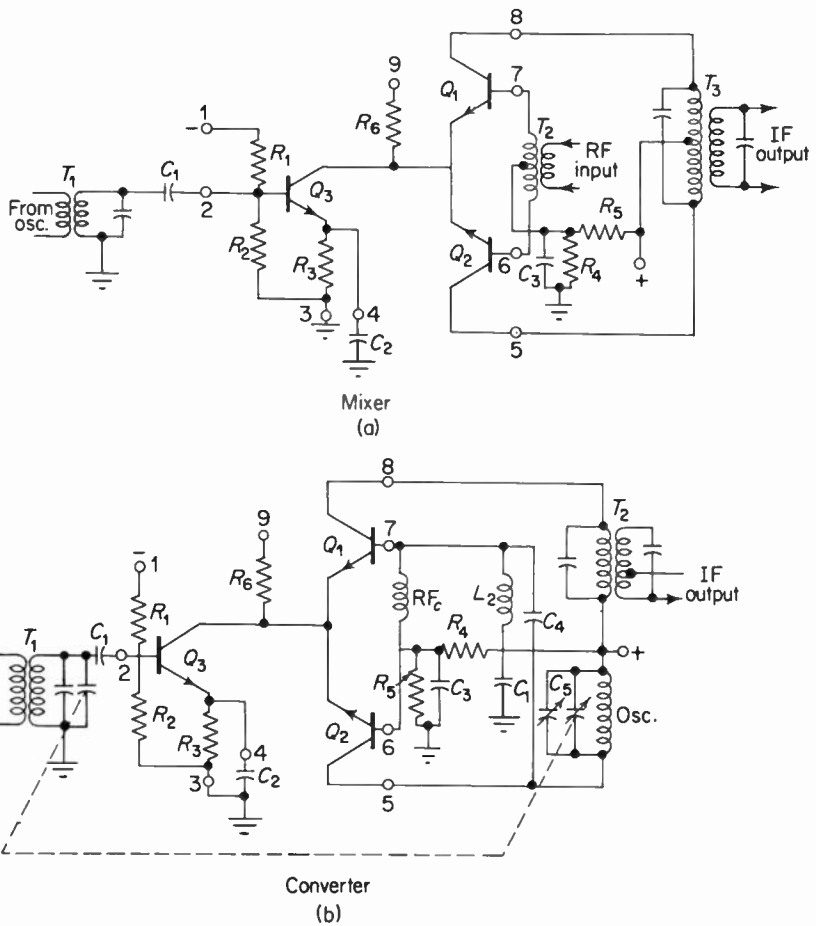



**Fig. 6-20.** The schematic diagram in (a) shows an integrated circuit used as a mixer. In (b) the integrated circuit is used as a self-oscillating converter.

center-tapped coupling transformer $T_2$. The bias for each of the transistors of the differential mixer is obtained through a voltage divider, $R_4$ and $R_5$, which is connected to the center tap of the RF input transformer. The IF output is obtained from a center-tapped transformer $T_3$. The center tap returns to the collector power supply. Since the input and output sections of the differential amplifier are arranged in a push-pull configuration, the output will be free of any in-phase signal appearing at the input. However, the out-of-phase desired RF signal, when applied to the input, will be converted into the intermediate frequency.

An important feature of integrated circuits is the reduction of internal feedback over that in a comparable discrete transistor. For example, at a frequency of 10.7 MHz the reverse transfer admittance ($Y_{12}$) of an integrated circuit in the differential mode may be approximately 140 times lower than that of a common-emitter transistor. At 100 MHz the reverse transfer admittance may be approximately 10 times less than that of a comparable common-emitter transistor. Thus neutralization of RF or mixer circuits is seldom necessary, provided that circuit layout is carefully designed. If improper shielding and layout of the external circuit elements produce excessive feedback paths, instability and oscillation may result.

The self-oscillating converter of Fig. 6-20(b) utilizes the integrated circuit in a somewhat different manner from the circuit of Fig. 6-20(a). Thus $Q_3$ is fed by the RF input transformer and is operated as a class A buffer amplifier, which amplifies and couples the RF signal into the emitters of $Q_1$ and $Q_2$. The combination of $Q_1$ and $Q_2$ is used as a self-oscillating converter. Oscillation is maintained in the manner of a multivibrator by capacitor $C_4$, which provides positive feedback by tying together the base of $Q_1$, which is in phase with the AC voltage at the output of $Q_3$, to the collector of $Q_3$. The tuned circuit $L_1$ and $C_5$ insures sinusoidal oscillations at the desired frequency. The resistors $R_4$ and $R_5$ provide bias for both $Q_1$ and $Q_2$. The RFC (Radio Frequency Coil) connected between the base of $Q_1$ and $Q_2$ provides a DC path for biasing $Q_1$ while at the same time isolating the base of $Q_1$ from the base of $Q_3$, which owing to $C_3$ is at AC ground. $T_2$ is a double-tuned transformer that couples the 10.7-MHz IF to the input of the next stage. This transformer is also used to provide matching between the mixer stage and the input of the next stage. In the base circuit of $Q_1$ is a series resonant circuit, which is tuned to the IF and insures that feedback at the intermediate frequency will not introduce instability at that frequency.

## 6-11. The local oscillator

The designer of an FM receiver has a choice as to where to place the local oscillator frequency. It may be placed either above or below the incoming-signal frequency by an amount equal to the intermediate frequency, thereby

generating the desired IF. However, for an IF of 10.7 MHz, placing the oscillator below the RF signal frequency will cause the receiver to be more responsive to TV image interference. This can be shown in the following manner. Since the RF signal frequencies extend from 88 to 108 MHz, the local oscillator, when placed below the RF signal, will range from 77.3 to 97.3 MHz. The image frequency in this case is 10.7 MHz lower than the oscillator frequency; thus the image response of the receiver will extend from 67.6 to 86.6 MHz. These frequencies correspond to the assigned frequencies for TV channels three through six.

Placing the oscillator above the RF signal has the disadvantage of increased susceptibility to oscillator drift, but the receiver is relatively free of TV image interference. Under these conditions the image response of the receiver falls between 109.4 and 129.4 MHz, which is the range of frequencies assigned to aeronautical air-to-ground services. Thus in most localities image response is minimized by placing the oscillator above the incoming RF signal.

### *Basic oscillator principles*

Oscillators used in FM receivers for frequency conversion are almost universally of the sinusoidal feedback type. Therefore, this review of basic principles will limit itself to this type.

A sinewave oscillator is a high-frequency AC generator. Essentially the oscillator converts the direct-current power-supply energy into high-frequency alternating-current energy. It does this by the use of *LC* tuned circuits, amplification and positive feedback. The *LC* circuit stores and exchanges energy in the form of either magnetic fields or electrostatic fields. The rate at which a stored magnetic field collapses and thereby charges the shunt capacitor and forms an electrostatic field, or vice versa, will determine the frequency of oscillation. Owing to the exchange of energy between the coil and capacitor of the *LC* circuit the voltage developed across the tuned circuit will be sinusoidal. Unfortunately, these sinusoidal oscillations will quickly die out because of power losses, which may be attributed to such things as the resistance associated with the coil, dielectric losses in the capacitor, and radiation losses. All of these power losses may be lumped into an effective resistance of the circuit. To sustain the oscillations of the *LC* circuit these power losses must be supplied by some outside source. The release of energy from the DC power supply for the purpose of replacing the energy lost is the function of the active circuit element (vacuum tube or transistor) of the oscillator. The release of energy from the power supply cannot be haphazard but must occur at the proper time—usually once during each cycle of oscillation. The proper timing of the active element as well as the actual coupling of energy from the output to the input *LC* circuit is accom-

plished by positive feedback. Thus, the basic requirements of an *LC* sinewave oscillator are that it possess (1) an *LC* circuit, (2) an amplifier, and (3) positive feedback of adequate amplitude.

The tuned-plate tuned-grid (*TPTG*), Hartley, and Colpitts oscillators all have the same basic circuit configuration; they differ only in their circuit details. Figure 6-21 shows the three transistor circuits in their usual configuration and also shows them redrawn into T equivalent circuits, indicating their similarities. The circuits of Fig. 6-21 are transistorized, but similar conclusions may be drawn about vacuum-tube or FET oscillators of the same type. Comparing Fig 6-21(a) and (b), we can see that the major difference between the tuned-plate, tuned-grid oscillator and the Hartley oscillator is that in the TPTG oscillator no mutual inductance exists between coils $L_1$ and $L_2$, whereas in the Hartley mutual inductance exists between $L_1$ and $L_2$. Also, the feedback capacitor in the TPTG is the interelectrode capacitance between collector and base; in the Hartley this function is provided by the tuning capacitor $C$. A comparison of the schematics of the Hartley and the Colpitts oscillators shows that coils and capacitors of these oscillators have been interchanged.

The frequency at which these oscillators will oscillate and the necessary conditions of gain and feedback to sustain oscillation may be determined analytically. The usual procedure for vacuum-tube oscillators,† assuming linear operation, is to use the equivalent circuit of the oscillator to determine the conditions that satisfy the Barkhausen criterion for oscillation. The Barkhausen criterion may be stated in a number of different ways; it is usually stated as

$$\beta A = 1 \qquad \text{or} \qquad \beta A - 1 = 0, \tag{6-31}$$

where $\beta$ is the feedback factor and $A$ is the voltage amplification without feedback. Both $\beta$ and $A$ are complex numbers having magnitude and phase angle. Note that any complex number that is equal to zero must consist of a real and an imaginary part, which are both equal to zero. Thus, by evaluating $\beta A$ and setting the real terms of the resulting complex number equal to zero, we may determine the conditions necessary to sustain oscillations. Likewise, setting the imaginary terms equal to zero will allow determination of the frequency of oscillation.

The transistor *LC* oscillator‡ is similar in appearance to the vacuum-tube oscillators. However, the transistor's low-impedance input loads the tuned circuit and makes analysis more difficult. The usual method is to set up the necessary loop equations (typically three loop equations are sufficient) for the complex linear equivalent circuit of the oscillator. The transistor equivalent may be either T, hybrid-pi, or any other type that suits the needs of the designer.

†See Appendix 6-1
‡See Appendix 6-2

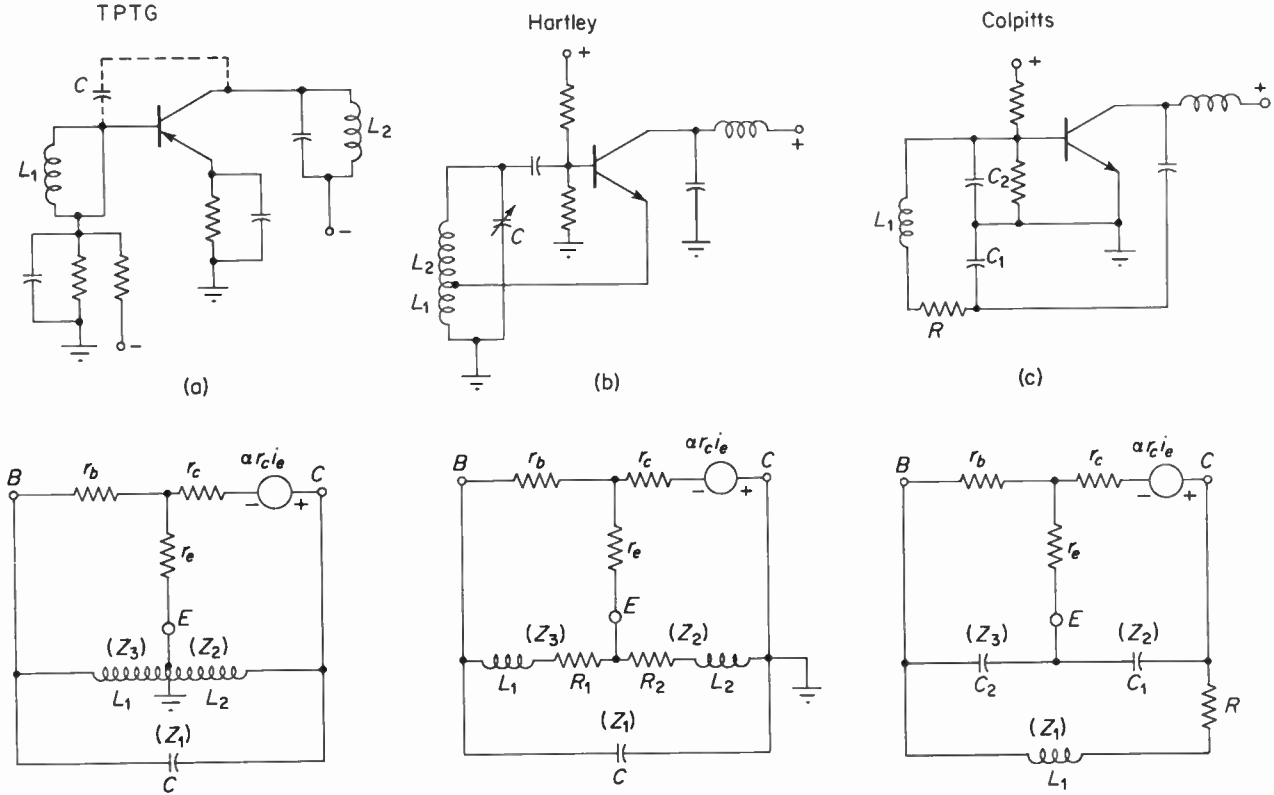



**Fig. 6-21.** Three transistor oscillator circuits and their equivalent circuits.

The solutions that result from solving the simultaneous loop equations are complex numbers. When the real parts of the complex number are made equal to zero, the conditions necessary for sustained oscillation may be determined. When the imaginary terms are made to equal zero, the equation that governs the frequency of oscillation may be found.

Table 6-1 summarizes the conditions necessary to sustain oscillation as well as the frequency at which the oscillation will take place.

### *Types of oscillators*

The local oscillator of receivers using vacuum tubes may be of the Hartley, Colpitts (ultra-audion), or modified Colpitts variety. These oscillators are shown in Fig. 6-22. Triodes are usually used as local oscillators because they require fewer components then do pentodes or tetrodes. Each of these triode oscillators, if properly designed, will oscillate easily and will exhibit good stability. The Colpitts oscillator possesses several advantages that make it a popular choice. First, it uses one untapped coil whereas the Hartley oscillator uses a tapped coil, and the modified Colpitts oscillator uses two untapped coils. Also in the Colpitts oscillator the distributed lead inductance

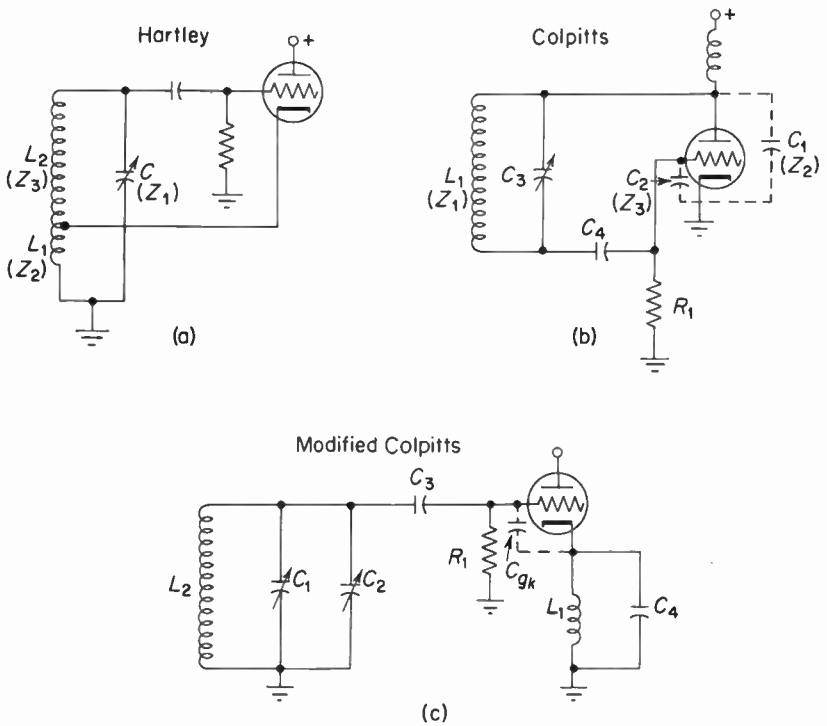



**Fig. 6-22.** Three types of oscillators commonly used in vacuum tube receivers.

**TABLE 6-1**

| Type of oscillator | Output load impedance | $Z_1/Z_3$ (Fig. 6-21) | Frequency of oscillation | Remarks |
|---|---|---|---|---|
| *Vacuum-tube oscillators:* | | | | |
| Hartley | Inductive | Capacitive | $f = \frac{1}{2\pi}\sqrt{\frac{1}{L_T C}\left(1 + \frac{R}{r_p}\right)}$ <br> $L_T = L_1 + L_2 + 2M$ | $g_m = \frac{RL_T C}{(L_1 + M)(L_2 + M)}$ |
| Colpitts | Capacitive | Inductive | $f = \frac{1}{2\pi}\sqrt{\frac{1}{LC_T}\left[1 + \frac{R}{r_p}\left(\frac{C_2}{C_1 + C_2}\right)\right]}$ <br> $C_T = \frac{C_1 C_2}{C_1 + C_2}$; $R$ = coil resistance | $g_m = \frac{R(C_1 + C_2)}{L}$ |
| *Transistor oscillators:* | | | | |
| TPTG | Inductive | Capacitive | $f = \frac{1}{2\pi}\sqrt{\frac{1}{C(L_1 + L_2) + \frac{L_1 L_2}{R}}}$ <br> $R = r_b r_c + r_b r_e + r_e r_c - r_b \alpha r_c$ | $\frac{L_2}{L_1} \approx \frac{\alpha r_c}{r_b + r_e}$ |
| Hartley | Inductive | Capacitive | $f = \frac{1}{2\pi}\sqrt{\frac{1}{CL_T + \frac{L_1 L_2 - M^2}{R}}}$ <br> $L_T = L_1 + L_2 + 2M$ | $\frac{L_2 + M}{L_1 + M} \approx \frac{\alpha r_c}{r_b + r_e}$ |
| Colpitts | Capacitive | Inductive | $f = \frac{1}{2\pi}\sqrt{\frac{1}{LC_T} + \frac{1}{RC_1 C_2}}$ <br> $C_T = \frac{C_1 C_2}{C_1 + C_2}$ | $\frac{C_1}{C_2} \approx \frac{\alpha r_c}{r_b + r_e}$ |

and stray capacitance of the tube become part of the $LC$ tuned circuit, and thus the Colpitts oscillator does not suffer from oscillations at undesired frequencies. Furthermore, the low-impedance paths offered by $C_3$ and $C_4$ to harmonics of the oscillator insure a good sinusoidal waveshape, thus reducing spurious responses. Unfortunately, the Colpitts oscillator tends to have large amplitude variations as the oscillator is tuned through the FM band, thus adversely affecting the conversion gain of the receiver.

The ultra-audion oscillator of Fig. 6-22(b) is a grounded-cathode Colpitts oscillator in which the feedback and voltage-divider capacitors $C_3$ and $C_4$ are the interelectrode capacities of the tubes, $C_{gk}$ and $C_{pk}$, respectively. Under normal operating conditions the grid-to-cathode voltage should be about one-third of the tank voltage. The relative size of $C_{gk}$ and $C_{pk}$ determines this ratio and may be adjusted by adding external capacitance. The frequency of operation is determined essentially by $C_2$ and $L_1$. Grid-leak bias provided by $C_1$ and $R_1$ is used to establish the operating point of the tube. The time constant of $R_1$ and $C_1$ is made larger than the ratio of the tank circuit $Q$ divided by the frequency of operation times pi. This is done to prevent periodic high-frequency interruptions of the oscillator, called "squegging."

The modified Colpitts oscillator of Fig. 6-22(c) can be thought of as a grounded-plate Colpitts oscillator, in contrast to the ultra-audion which uses a grounded-cathode configuration. For this circuit to oscillate, $C_4$ and $L_1$ must form a capacitive circuit at the frequency of oscillation. In many practical circuits $C_4$ is the distributed capacity of the circuit and $L_1$ provides the DC path for current flow. Feedback of the voltage developed across $L_1$ and $C_4$ by the AC component of plate current takes place through the interelectrode capacitance $C_{gk}$. In Chapter 4 it was shown that this type of circuit arrangement is able to generate a negative resistance, whose magnitude is equal to

$$R = -\frac{C_k}{g_m C_{gk}}. \tag{4-9}$$

When this negative resistance is equal to the losses in the circuit, the circuit will oscillate. The fequency at which the circuit oscillates is determined by the tuned circuit $L_2$ and $C_1$.

The local oscillator used in transistor FM receivers is almost universally the ultra-audion arrangement of the Colpitts oscillator. This circuit is shown in Fig. 6-23. The transistor is arranged in a common-base configuration. [See the redrawn circuit in Fig. 6-23(b).] This is accomplished by $C_1$ which provides an AC short between the base and circuit ground. (The position of ground in an oscillator is a matter of convenience, and thus any element of the transistor could be used as the common element.) $C_2$ is the feedback capacitor, which returns part of the output collector voltage to the emitter-base input circuit. This voltage is developed across capacitor $C_3$, which may

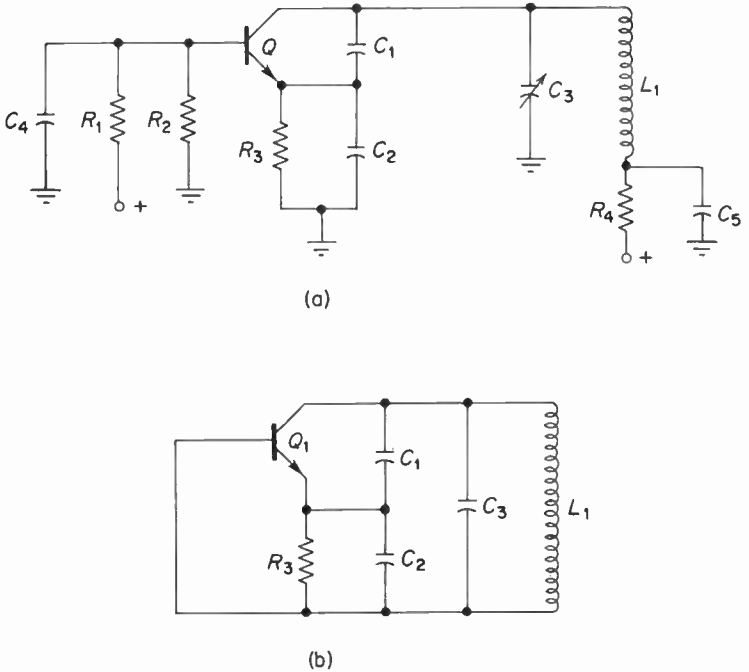



**Fig. 6-23.** A transistor Colpitts oscillator is shown in (a). It is redrawn in (b) to show the AC connections more clearly.

be either lumped or distributed. Resistors $R_1$, $R_2$, and $R_3$ set the operating point of the oscillator. The tuned circuit consisting of $C_2$, $C_3$, $C_4$, and $L_1$ determines the frequency of operation of the oscillator. $C_4$ is part of the main tuning capacitor and is ganged to the main tuning capacitors associated with the RF amplifier and mixer. $C_4$ is usually shunted with a small trimmer capacitor, which is used to adjust tracking at the high-frequency end of the FM band. This trimmer capacitor is adjusted in conjunction with the trimmer capacitors across the RF amplifier and mixer main tuning capacitors. The capacitor $C_5$ is a fairly large bypass capacitor which holds the lower end of $L_1$ at RF ground, thereby tying the tuning capacitors ($C_2$, $C_3$, and $C_4$) across $L_1$.

In general, transistor operation for use as an oscillator is nonlinear and the signal is relatively large; thus small-signal parameters are not very useful. Two important criteria for selecting transistors for use in oscillator circuits are the maximum frequency of oscillation ($F_{max}$) and the amount of available power output at a given frequency. For VHF oscillators the $F_{max}$ transistor rating should be at least 120 MHz and the available power capability should be of the order of 1 to 5 milliwatts. In practice only 0.5 mw is required by the mixer; the reserve is necessary if adequate decoupling from the oscillator is to be achieved.

## 6-12. Frequency stability and automatic frequency control

One of the major problems affecting the FM receiver is distortion caused by local-oscillator frequency drift. The usual symptoms of this defect manifest themselves shortly after the receiver is turned on and a station is properly tuned in. The sound becomes badly distorted, whereupon the receiver may be readjusted for distortion-free reception. Shortly thereafter the receiver may again require readjustment. This continuous readjustment of tuning may continue for a period of a half hour or longer, depending upon the time necessary for the receiver to reach thermal and electrical equilibrium.

This condition is brought about because the local oscillator, whether vacuum tube or transistor, is very sensitive to the initial temperature changes that take place when the receiver is first turned on. In the case of vacuum tubes, the chief cause of frequency drift during the warm-up period is the tendency of tube interelectrode capacitances to become larger. Since the vacuum tube generates considerable heat, the adjacent frequency-controlling tuned-circuit elements tend to expand with increasing temperature. The inductance of a coil is governed by the expression

$$L_{\text{hy}} = \frac{1.26\, N^2\, \mu A}{l} \times 10^{-8} \tag{6–32}$$

where $N$ is the number of turns, $\mu$ is the permeability, $l$ is the length of the coil, and $A$ is the cross-sectional area of the coil. Since the coil's cross section is circular, its area is equal to $\pi$ times the radius squared; thus an increase in temperature will increase the length of the coil to the first power, but the area will increase by the square power. Hence the inductance will increase with an increase in temperature. Likewise, the plate area of the tuning capacitor will expand with an increase in temperature, which will also cause the tuning capacitance to increase. Hence an increase in temperature will cause a decrease in oscillator frequency. Oscillator drift over a period of a half hour due to all causes in a carefully designed but uncompensated oscillator will be approximately 30 kHz at a frequency of 100 MHz. After this initial warm-up time the oscillator frequency will remain constant.

Thermal changes also affect transistor oscillator-frequency stability. Temperature changes will cause the various junction capacitances to change, which in turn will cause the oscillator frequency to change. However, drift due to temperature change is small except in such applications as automobile FM receivers or wherever receivers may experience temperature changes.

Transistor junction capacities are also sensitive to power-supply variations. Thus, where power-supply variations may be expected and hence oscillator drift encountered, some means of correction may be desired.

Figure 6-24 shows how oscillator drift causes distortion. The top line of the figure represents the RF signal frequency. We assume here that the frequency to which the receiver is tuned is 100 MHz. The second line shows the oscillator frequency. Since the IF is 10.7 MHz, the oscillator will be at 110.7 MHz, position *A*. As the temperature increases, the frequency of the oscillator will decrease to the positions marked *B*, *C*, and *D*. The third line illustrates the IF bandpass characteristic, which is centered at the difference frequency between the incoming RF signal and the oscillator frequency. Since the oscillator frequency will drift to a lower frequency, the IF will also become lower in frequency; therefore the signal will shift its position on the IF response to those lower-frequency positions marked *B*, *C*, and *D*. The bottommost line of Fig. 6-24 represents the detector output "S"-curve

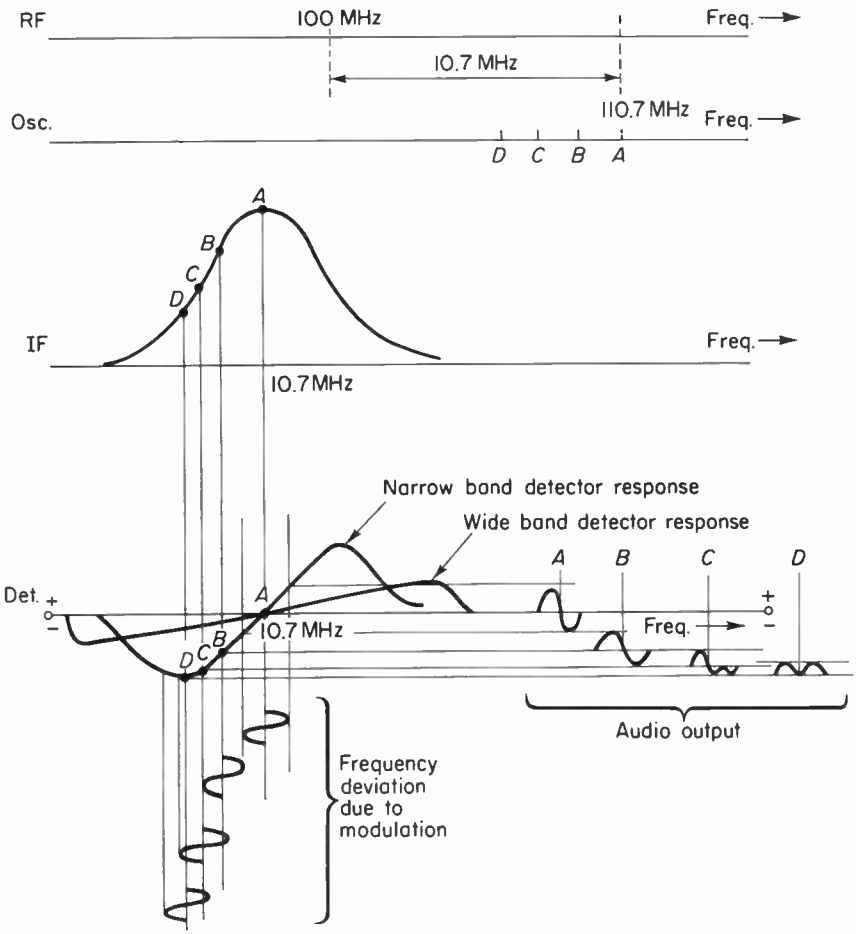



**Fig. 6-24.** The above diagram illustrates the effect upon the audio output of an FM receiver when the local oscillator is subject to drift.

response. Also indicated on this line are the positions of the IF due to oscillator drift. Position *A* corresponds to the desired position of perfect tuning. At this position the IF is centered on the flat portion of the IF response; thus minimum phase distortion is introduced. As a result, distortion due to the production of indirect FM is minimum. Also, since the IF falls in its proper place on the linear portion of the "S" curve, little harmonic distortion is introduced by the detector. Of importance to this discussion is the fact that the average value (DC level) of the audio output is zero when the receiver is properly tuned.

When the oscillator drifts to position *B*, the IF moves to a lower frequency, which now falls on the slope of the IF response curve. In this position, owing to phase distortion, indirect FM will be generated, which will introduce harmonic distortion. Also produced by operation on the slope of the IF response will be amplitude modulation. In position *B* on the "S" curve the frequency deviation due to the modulated signal will carry the signal into the nonlinear regions of the detector response curve. As can be seen from the diagram of the audio output waveform, the negative peak of the output is smaller than the positive peak, indicating harmonic distortion. Note also that the average value of the output of the detector is greater than zero and is negative in polarity. In certain FM detectors the magnitude of this voltage will depend upon input-signal level. Thus strong signals will develop greater average values than will weak signals. If the "S"-curve response were to be reversed (a relatively simple procedure), the average value would be positive as the IF became lower in frequency. The polarity of the average value at the output of the detector may be important when corrective measures such as automatic frequency control are employed.

In positions *C* and *D* the oscillator drift has caused the IF to shift into areas of the IF and detector response curves that produce severe distortion, as illustrated by the audio output waveforms of Fig. 6-24. As is evident from the "S"-curve response, it is possible to minimize distortion due to oscillator drift by using a detector that has a wide-band response. If the detector response bandwidth is made large enough, oscillator drift will usually not be sufficient to shift the signal into nonlinear regions of the detector response. However, distortion and amplitude modulation may still be produced by operation of the signal on the slope of the IF amplifier response.

### *Automatic frequency control*

There are at least three techniques for minimizing oscillator drift. First is careful design, which includes such things as adequate ventilation and careful placement and selection of components. These factors prevent excessive temperature changes during initial warm-up time. Also included in careful design must be good voltage regulation to minimize oscillator drift due to voltage variations.

A second technique, useful primarily in vacuum-tube receivers where large temperature changes are expected, is the use of negative-temperature-coefficient ceramic capacitors. These capacitors may be designed to have positive, zero, or negative temperature coefficients. Thus if a negative- or zero-temperature-coefficient capacitor of the proper size and temperature coefficient is selected and correctly placed in the oscillator circuit, it can counteract the effects of the positive-temperature-coefficient inductance and capacitance of the circuit. An example of the use of such capacitors is shown in Fig. 6-25. In this circuit a 22pF N750 capacitor is used as the grid-leak capacitor. This

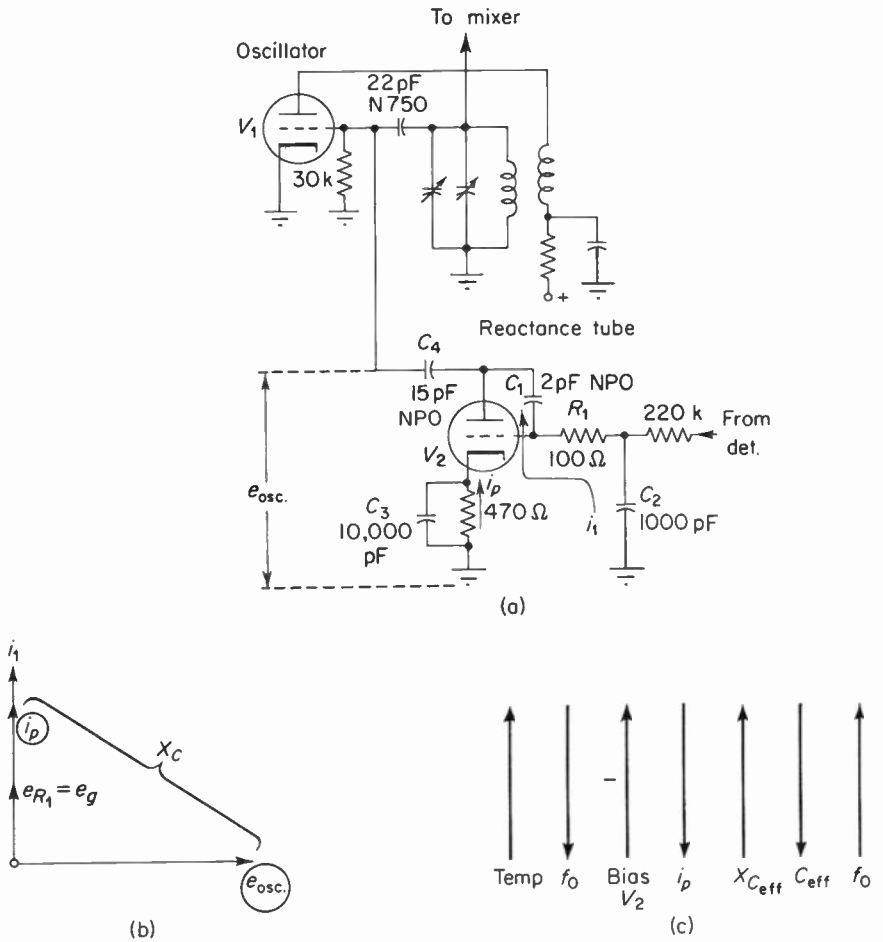



**Fig. 6-25.** The schematic diagram of a vacuum tube oscillator and its associated reactance tube is shown in (a). The phase relationships of the reactance tube is shown in the phasor diagram of (b). The manner by which the reactance tube controls the frequency of the oscillator is shown in (c).

capacitor is in series with the tube interelectrode capacitance $C_{gk}$, and this series combination is in parallel with the frequency-determining $LC$ circuit. Thus the negative temperature coefficient of the capacitor counterbalances the positive temperature coefficient of the interelectrode capacitance.

The third technique for minimizing oscillator drift is the use of automatic frequency control (AFC). Automatic frequency control is basically a DC feedback system consisting essentially of the local oscillator, a frequency-to-DC converter (the FM detector), and a frequency-controlling device. The frequency-to-DC converter generates a DC voltage that is proportional to the frequency drift of the local oscillator. This voltage is then fed back to a device that is able to force the oscillator frequency in such a direction as to reduce the feedback voltage and thereby compensate for oscillator drift. A block diagram showing the flow path of this control system is given in Fig. 6-26. Between the output of the FM detector and the voltage-controlled variable reactance is a low-pass filter whose function is to prevent frequency modulation of the local oscillator by the audio output of the detector. If audio were to frequency-modulate the local oscillator, the effective deviation of the signal would be reduced. Under some conditions this can be a desirable condition. This topic is discussed in greater detail in Sec. 11-6 of the text, hence it will not be discussed here.

Two types of voltage-controlled reactances are commonly used in automatic frequency control systems: (1) the reactance tube and (2) the silicon diode voltage-controlled variable capacitor. A reactance tube† (or transistor) is simply a tube whose circuit is arranged to make the AC plate voltage differ in phase from the AC plate current by 90 degrees. If the plate current leads the plate voltage, the tube appears to the oscillator as a capacitor. If the plate current lags the plate voltage, the oscillator "sees" an inductance. A change of reactance-tube bias will cause a change in $g_m$, which in turn will cause a change in AC plate-current amplitude. Thus by Ohm's law the magnitude of the tube's reactance will change. If the reactance tube is in shunt with the

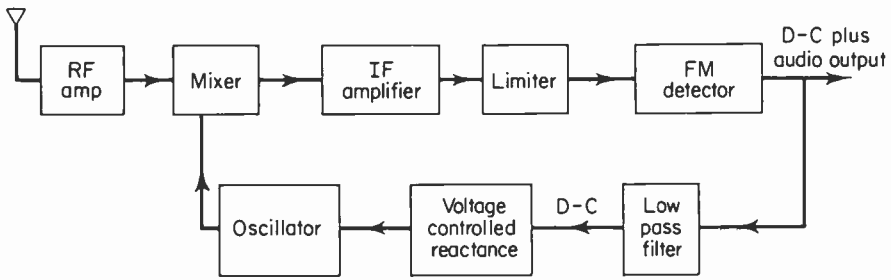



**Fig. 6-26.** The block diagram of a receiver which uses automatic frequency control.

†The reactance tube is analytically treated in Sec. 1-9 of this text.

oscillator tuned circuit, the change in reactance will cause the frequency of the oscillator to change.

An example of a reactance-tube AFC is shown in Fig. 6-25. In this case the tube is made to appear as a capacitive reactance. This is accomplished by the phase-shift network $C_1$ (this may simply be the interelectrode capacitance of the tube) and $R_1$. Notice that the large capacitors $C_2$ and $C_3$ effectively place $R_1$ between grid and cathode of the tube. Since the reactance of $C_1$ is much larger than $R_1$ at the frequency of operation, the circuit is capacitive. $C_4$ couples the oscillator tank voltage to the plate of the reactance tube; therefore the voltage across the tube is the oscillator tank voltage. Figure 6-25(b) is a phasor diagram showing the phase relationships between the voltages and currents of the circuit. Thus the current ($i_1$) passing through the phase-shift network will lead the oscillator voltage by approximately 90 degrees. This current will then produce a voltage drop across $R_1$ that will be in phase with $i_1$. Since this voltage is in fact the grid voltage of the tube, the plate current ($i_p$) will be in phase with it. Therefore the plate voltage will lag the plate current by approximately 90 degrees and will appear to the oscillator as a capacitive reactance.

The manner by which the reactance tube controls the frequency of the oscillator is indicated by Fig. 6-25(c). The arrows indicate the direction of change. Thus, reading from left to right, an increase in temperature ($T$) causes the frequency ($f$) to decrease, which causes the output of the FM detector to produce an increasing negative voltage, which is used as the bias of the reactance tube. The increase in bias causes the plate current ($i_p$) to decrease, which is the same as an increase in effective capacitance reactance ($X_C$). If the reactance of a capacitor is increased, the effective capacitance ($C_{eff}$) must decrease. Since the frequency of the oscillator is approximately equal to

$$f = \frac{1}{2\pi\sqrt{LC}}, \tag{6-33}$$

a decrease in total capacity of the tuned circuit will cause the frequency to increase. This in effect counteracts the original decrease in frequency due to thermal oscillator drift.

### *Silicon diode AFC*

The voltage-controlled silicon-diode variable capacitor has been referred to as a "varicap," "semicap," and "varactor," by various manufacturers. We shall call it a varicap. The varicap is based upon the characteristics of the depletion region of a PN junction. As is well known, when a PN junction is formed, the holes and electrons in the region adjacent to the junction combine and form a zone relatively free of current carriers. This region is, in effect, a good insulator sandwiched between the electron-rich N section and the

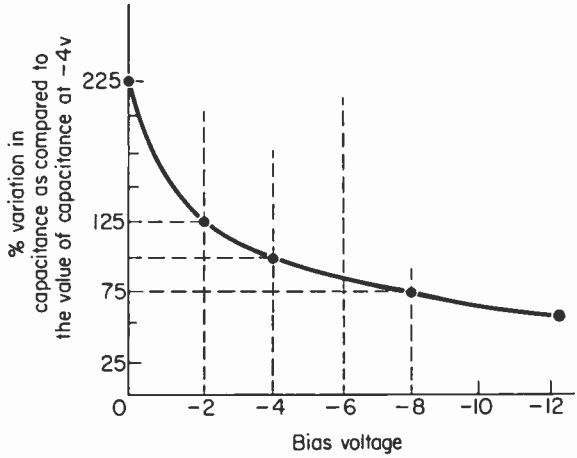



**Fig. 6-27.** The variation of capacitance with bias voltage.

hole-rich P section. The arrangement of an insulator between two conducting areas will exhibit capacitive properties.

If a reverse bias is placed across the diode, the mobile electrons in the N-type material and the mobile holes in the P material will be drawn away from the junction, effectively increasing the thickness of the dielectric. This, in effect, will cause a reduction in the junction capacitance of the silicon diode. Hence it is the property of a PN junction to vary the thickness of the depletion zone when the bias voltage across the diode is varied. Consequently the silicon diode may be used as a voltage-controlled variable capacitor. The capacitance of the diode is inversely proportional to the square root of the voltage across the diode. Figure 6-27 shows how the capacitance varies for changes in bias voltage. This type of characteristic is for a diode with an abrupt PN junction. Such a junction may be obtained by alloying techniques, or a close approximation can be achieved by the use of point-contact diodes. Furthermore, the junction capacitance of the diode, assuming a constant bias, is relatively insensitive to temperature changes, because the effects of changes in resistivity and mobility with temperature tend to cancel one another.

The silicon voltage-controlled variable capacitor can be obtained with values of capacitance as low as 0.5 pF at a bias of −4 volts or as high as 1800 pF at the same bias. The usual diode used in AFC applications will have a −4-volt bias capacitance of approximately 40 pF.

The equivalent circuit of the varicap for frequencies in the range of the FM band is shown in Fig. 6-28. The $Q$ of the varicap may be found by the expression

$$Q = \frac{X_C}{R} = \frac{1}{2\pi f C_d R}, \tag{6-34}$$

where $C_d$ is the diode capacitance and $R$ is the series resistance of the varicap.

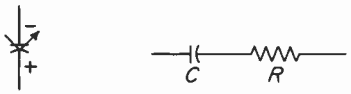



**Fig. 6-28.** The schematic diagram of a varicap and its equivalent.

If this diode is used as part of a tuned circuit, the tuned-circuit $Q$ may be found by the same equation; however, $C$ becomes the total resonant capacitance and $R$ is the total series resistance including the resistance of the coil.

The $Q$ of a high-frequency diode may not be very large, ranging from 20 to 80. One way of increasing the effective circuit $Q$ is by placing a capacitor in series with the varicap. This series capacitor ($C_s$) will increase the $Q$ by a factor of $(C_d + C_s)/C_s$. The addition of the series coupling capacitor increases the reactance of the circuit without appreciably affecting the series resistance of the circuit. Thus the $Q$ is increased. The addition of $C_s$ also tends to reduce the effective capacitance of the diode shunting the tuned circuit, as well as reducing the amount of RF signal energy that appears across the diode. If a small value of $C_s$ is used, this will tend to reduce the effectiveness of the variable diode capacitance upon the frequency response of a tuned circuit; however, a small value of coupling capacitor will tend to reduce the AC signal coupled to the diode and will therefore minimize any self bias that might be generated. Therefore, the range of diode control bias necessary to obtain a given change in capacitance is reduced. Since the frequency sensitivity of the varicap may be shown to be

$$\frac{\Delta F}{\Delta V} = -\frac{F}{4V}\left(\frac{1}{1 + \frac{C_d}{C_s}}\right) \tag{6-35}$$

where $\Delta F$ is the change in frequency, $\Delta V$ is the change in bias voltage, $F$ is the frequency of operation, and $V$ is the bias voltage, it can be seen that $C_s$ should be large to maximize frequency sensitivity.

In general, the sensitivity of a varicap is considerably greater than that of vacuum-tube reactance tubes. Varicaps may be used in AFC systems for either vacuum-tube or transistor receivers. An example of a transistor oscillator that employs a varicap in its AFC system is shown in Fig. 6-29. The oscillator is an "N"-type common-base Colpitts oscillator of the type discussed earlier. The AFC diode obtains its bias from the voltage divider $R_1$ and $R_2$, which also supplies the bias for the oscillator. As can be seen in Fig. 6-29(b), the low reactance bypass capacitors $C_1$ and $C_2$ tie one end of the varicap and one end of the frequency-determining coil together. The other side of the variable capacitor diode is connected to a small coupling capacitor $C_s$ which is tied to the other end of the coil. Thus the varicap is placed across the oscillator $LC$ circuit. The diode capacitance is varied by the error voltage generated by the FM detector. This DC component is coupled to the varicap via a low-pass filter ($R_3$ and $C_3$) which removes the audio component

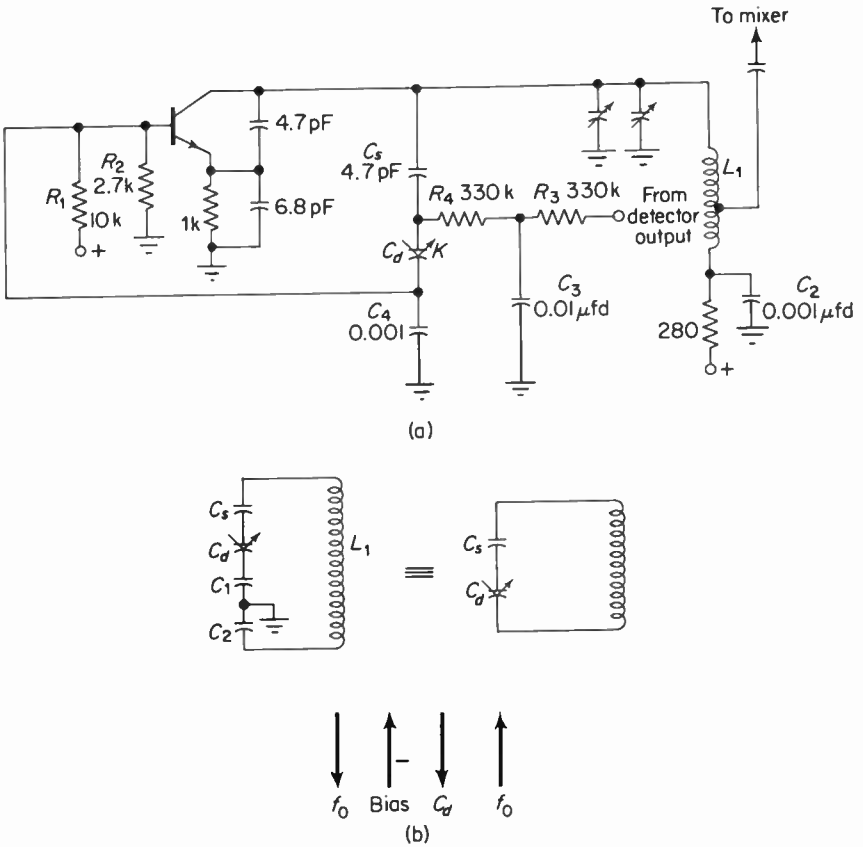



**Fig. 6-29.** The schematic diagram of a transistor oscillator and its varicap controlled AFC is shown in (a). The effective shunt capacitance across the tuning inductance ($L_1$) and the manner by which the frequency on the oscillator is controlled by the varicap is shown in (b).

that also appears at the output of the FM detector. The resistor $R_4$ is used to isolate the varicap from the high-frequency short circuit due to $C_3$.

Figure 6-29(c) shows by means of arrows how the AFC system operates. If the local oscillator were to drift to a lower frequency, the error voltage from the detector would increase the negative bias across the varicap, thereby reducing the total shunt capacity across the tuning inductance. A reduction in tuning capacitance would introduce an increase in the frequency of oscillation, thus tending to correct the original oscillator drift.

## 6-13. Summary

**1.** The mixer is located between the RF amplifier and the IF amplifier of a superheterodyne receiver. Its function is to reduce the incoming signal to a lower frequency called the intermediate frequency.

**2.** Frequency conversion may be performed by multiplicative or by additive mixing.

**3.** The oscillator and RF signals may be coupled together in series or parallel and brought to the mixer by capacitive or transformer coupling.

**4.** Excessive coupling between the RF signal and the oscillator may result in frequency synchronization called oscillator pulling.

**5.** Nonlinear operation is required for frequency changing.

**6.** Frequency conversion is usually accompanied by amplification. A convenient way to compare frequency changers is to determine their conversion gain.

**7.** A mixer is many times noisier than the same tube or transistor used as an amplifier.

**8.** The greatest source of noise in a receiver is the mixer.

**9.** A useful means for comparing the noise generated by amplifiers or mixers is the noise figure.

**10.** The fewer the current-carrying electrodes in a vacuum-tube mixer, the lower the noise level of the mixer.

**11.** The transistor mixer is usually considered to be equivalent to a diode mixer and an amplifier, where the input junction constitutes the diode mixer.

**12.** FET mixers have the advantages of low noise, high gain, and reduced spurious responses owing to their relatively high input impedance.

**13.** Integrated circuits may be used as frequency changers. They may be used as autodyne converters or as separate oscillator and mixer circuits.

**14.** The local oscillator is placed above the frequency of the incoming RF signal and thus minimizes spurious responses.

**15.** An oscillator is a high-frequency generator that converts direct current into alternating current.

**16.** The Colpitts and Hartley oscillators are the most commonly used local oscillators in FM receivers.

**17.** Automatic frequency control is a system of negative feedback that reduces the effects of oscillator drift.

## REFERENCES

1. Herold, E. W.: "The Operation of Frequency Converters and Mixers for Superheterodyne Reception," *Proc. IRE*, February 1942.
2. Langford-Smith, F.: *Radiotron Designers Handbook*. Radio Corporation of America, Harrison, N.J., 1953.
3. Gray, C. R., and T. C. Lawson: "Transistorized FM Front Ends," *Audio*, August 1962.

4. van der Ziel, A.: "Noise in Junction Transistors," *Proc. IRE*, May 1958.
5. Van Duyne, John: "Suppressing Local Oscillator Radiation in TV Receivers," *Tele-Tech*, February 1952.
6. Read, L. W.: "An Analysis of High Frequency Transistor Mixers," *IEEE Trans. on Broadcast and Television Receivers*, May 1963.
7. Hubbard, D. J.: "A Low Cost All Silicon FM Tuner," Application Note 90.27, June 1965, General Electric Company.
8. Pan, W. Y.: "Frequency Characteristic of Local Oscillators," *RCA Review*, September 1955.
9. Pan, W. Y., and Carlson, D. J.: "Practical Aspects of Local Oscillator Stabilization," *RCA Review*, December 1956.
10. Edson, W. A.: *Vacuum Tube Oscillator*. John Wiley & Sons, Inc., New York, 1953.
11. Zeines, Ben: *Principles of Applied Electronics*. John Wiley & Sons, Inc., New York, 1963.
12. Leonard, David: "Improve FM Performance with FET's," *Electronic Design*, March 1, 1967.
13. Carlson, F. M.: "Application Consideration for the RCA 3N128 VHF MOS Field-Effect Transistor," RCA Application Note AN-3193, Radio Corporation of America, Harrison, N.J.
14. Mergner, F.: "P-I-N Diode and FET's Improve FM Reception," *Electronics*, August 22, 1966.
15. Rohde, U. L.: "The Field-Effect Transistor at V.H.F.," *Wireless World*, January 1966.
16. Hugenholtz, E. H.: "One Tube Oscillator Mixer," *Electronics*, January 15, 1960.
17. Fowler, W.: "Transistor Autodyne Converters," *Electro-Technology*, February 1965.
18. Klein, E.: "*Y*-Parameters Simplify Mixer Design," *Electronic Design*, April 1, 1967.
19. The Engineering Staff of Texas Instruments, Inc.: *Transistor Circuit Design*. 1963.
20. Von Recklinghausen, D. R.: "Theory and Design of FET Converters," *Trans. IEEE PGBTR*, May 1965.
21. Sanquini, R. L.: "Integrated Circuits Make a Low-Cost FM Receiver," *Electronics Journal*, August 8, 1966.
22. Kiehn, H. C.: "Application of the RCA-CA3028 Integrated-Circuit RF Amplifier in HF and VHF Ranges," RCA Integrated Circuits Application Note ICAN-5337.
23. Johnson, R. J.: "Drift and A.F.C. in FM," *Radio and TV News*, April 1959.
24. Straube, G. F.: "A Voltage Variable Capacitor," *Electronic Industries*, May and July 1958.
25. Turner, R. P.: "Using the Varicap," *Radio-Electronics*, May 1958.
26. Gibson, S. K.: "Add A.F.C. with a Silicon Diode," *Electronics World*, July 1959.

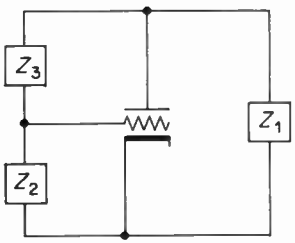
## APPENDIX 6-1

The frequency of operation and the transconductance necessary to sustain oscillation may be derived for a Colpitts vacuum-tube oscillator by applying the Barkhausen criterion to the circuit shown in the accompanying diagram.

$$Z_L = \frac{Z_1(Z_2 + Z_3)}{Z_1 + Z_2 + Z_3}, \tag{1}$$

where $Z_1 = 1/j\omega C_1$, $Z_2 = 1/j\omega C_2$, and $Z_3 = R + j\omega L$.

$$\beta = \frac{Z_2}{Z_2 + Z_3}. \tag{2}$$




Substitute the above into the Barkhausen equation.

$$r_p + Z_L = -\mu\beta Z_L,$$

$$r_p + \frac{Z_1(Z_2 + Z_3)}{Z_1 + Z_2 + Z_3} = -\mu\left(\frac{Z_2}{Z_2 + Z_3}\right)\left[\frac{Z_1(Z_2 + Z_3)}{Z_1 + Z_2 + Z_3}\right],$$

$$\frac{[r_p(Z_1+Z_2 + Z_3) + Z_1(Z_2 + Z_3)](Z_1 + Z_2 + Z_3)}{Z_1 + Z_2 + Z_3} = -\mu\left(\frac{Z_2}{Z_2 + Z_3}\right)\left(\frac{Z_1(Z_2 + Z_3)}{Z_1 + Z_2 + Z_3}\right)\left(\frac{Z_1 + Z_2 + Z_3}{1}\right),$$

$$r_p(Z_1 + Z_2 + Z_3) + Z_1(Z_2 + Z_3) = -\mu Z_2 Z_1,$$

$$r_p(Z_1 + Z_2 + Z_3) + Z_1(Z_2 + Z_3) + \mu Z_1 Z_2 = 0,$$

$$r_p(Z_1 + Z_2 + Z_3) + Z_1(Z_2 + Z_3 + \mu Z_2) = 0,$$

$$Z_1(Z_2 + Z_3 + \mu Z_2) = -r_p(Z_1 + Z_2 + Z_3),$$

$$Z_1[Z_2(1 + \mu) + Z_3] = -r_p(Z_1 + Z_2 + Z_3).$$

Substitute for $Z_1$, $Z_2$, and $Z_3$.

$$\frac{1}{j\omega C_1}\left[\frac{1}{j\omega C_2}(1 + \mu) + (R + j\omega L)\right] = -r_p\left[\frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} + R + j\omega L\right],$$

$$-\frac{1}{\omega^2 C_1 C_2}(1 + \mu) + \frac{R}{j\omega C_1} + \frac{L}{C_1} = -\frac{r_p}{j\omega C_1} - \frac{r_p}{j\omega C_2} - Rr_p - j\omega L r_p.$$

Set reals equal to zero and solve for $g_m$.

$$-\frac{1}{\omega^2 C_1 C_2}(1+\mu)+\frac{L}{C_1}+Rr_p=0,$$

$$-\frac{1+\mu}{\omega^2 C_1 C_2}+\frac{L}{C_1}+Rr_p=0,$$

$$-1-\mu\left(\frac{1}{\omega^2 C_1 C_2}\right)+\frac{L}{C_1}+Rr_p=0,$$

$$1+\mu\left(\frac{1}{\omega^2 C_1 C_2}\right)-\frac{L}{C_1}-Rr_p=0,$$

$$\mu=\left(Rr_p+\frac{L}{C_1}\right)(\omega^2 C_1 C_2)-1.$$

But $g_m=\mu/r_p$; therefore

$$g_m=\frac{\left(Rr_p+\frac{L}{C_1}\right)(\omega^2 C_1 C_2)-1}{r_p}.$$

Set imaginaries equal to zero and solve for $\omega_{osc}$.

$$-j\frac{R}{\omega C_1}-j\frac{r_p}{\omega C_2}-j\frac{r_p}{\omega C_1}+j\omega Lr_p=0,$$

$$-\frac{R}{\omega C_1}-\frac{r_p}{\omega C_2}-\frac{r_p}{\omega C_1}+\omega Lr_p=0.$$

$$\omega_{osc}=\sqrt{\frac{1}{LC}\left(1+\frac{RC}{r_p C_1}\right)}, \qquad \text{where} \quad C=\frac{C_1C_2}{C_1+C_2}.$$

Thus

$$f_{osc}=\frac{1}{2\pi\sqrt{LC}}\sqrt{1+\frac{RC}{r_pC_1}}.$$

## APPENDIX 6-2

In this appendix we will derive the equations for the frequency of operation and the required conditions to sustain oscillation for the transistor Colpitts oscillator of Fig. 6-23. The hybrid-pi equivalent circuit of this oscillator is shown in Fig. 1. In this oscillator the function of $C_2$ of Fig. 6-23 is accomplished by the transistor interelectrode capacitance $C_{b'e}$.
The tuned circuit is assumed to have negligible losses and its impedance is therefore

$$Z_T=\frac{X_C(jX_L)}{jX_L+X_C} \qquad \text{Note} \quad X_c=\frac{1}{j\omega C} \tag{1}$$

$$=\frac{j\omega L}{1+\frac{j\omega L}{X_C}}$$

$$=\frac{j\omega L}{1-\omega^2 LC} \tag{2}$$

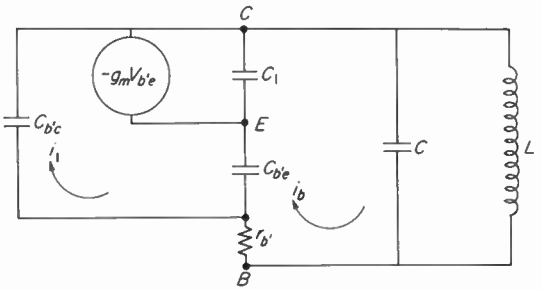



**Fig. 1**

The loop equations for the circuit of Fig. 1 are

$$i_b(r_{b'} + X_{C_{b'e}} + X_{C_1} + Z_T) - i_1 X_{C_{b'e}} g_m V_{b'e} X_{C_1} = 0 \tag{3}$$

$$i_b(r_{b'} + Z_T) + i_1 X_{C_{b'c}} = 0 \tag{4}$$

Solving the last equation for $i_b$ and substituting it into the first equation

$$\frac{-i_1 X_{C_{b'c}}}{r_{b'} + Z_T}(r_{b'} + X_{C_{b'e}} + X_{C_1} + Z_T) - i_1 X_{C_{b'c}} + g_m V_{b'e} X_{C_1} = 0 \tag{5}$$

Multiply through by $(r_{b'} + Z_T)$

$$r_{b'} + X_{C_{b'e}} + X_{C_1} + Z_T + \frac{X_{C_{b'e}}(r_{b'} + Z_T)}{X_{C_{b'c}}} + \frac{g_m V_{b'e} X_{C_1}(r_{b'} + Z_T)}{-i_1 X_{C_{b'c}}} = 0 \tag{6}$$

Notice that $V_{b'e} = X_{C_{b'e}}(i_b - i_1)$ and substitute it into the last equation

$$r_{b'} + X_{C_{b'e}} + X_{C_1} + Z_T + \frac{X_{C_{b'e}}(r_{b'} + Z_T)}{X_{C_{b'c}}} + g_m X_{C_1} X_{C_{b'e}}\left(\frac{i_b(r_{b'} + Z_T)}{-i_1 X_{C_{b'c}}} + \frac{-i_1(r_b + Z_T)}{-i_1 X_{C_{b'c}}}\right) = 0 \tag{7}$$

Notice that the term $i_b(r_{b'} + Z_T)/-i_1 X_{C_{b'c}}$ is the same as equation 4 and therefore equals unity.

$$r_{b'} + X_{C_{b'e}} + X_{C_1} + Z_T + \frac{X_{C_{b'e}}(r_{b'} + Z_T)}{X_{C_{b'c}}} + g_m X_{C_1} X_{C_{b'e}}\left(1 + \frac{r_{b'} + Z_T}{X_{C_{b'c}}}\right) = 0 \tag{8}$$

The frequency of oscillation may be determined by setting the imaginary terms of equation 8 to zero and solving for $f_0$

$$X_{C_{b'e}} + X_{C_1} + Z_T + \frac{Z_T X_{C_{b'e}}}{X_{C_{b'c}}} + \frac{g_m X_{C_1} X_{C_{b'e}} r_{b'}}{X_{C_{b'c}}} = 0 \tag{9}$$

Solve for $f_0$

$$\frac{1}{j\omega C_{b'e}} + \frac{1}{j\omega C_1} + \frac{j\omega L}{1 - \omega^2 LC} + \frac{j\omega L(j\omega C_{b'c})}{(1 - \omega^2 LC)(j\omega C_{b'e})} + \frac{g_m j\omega C_{b'c} r_{b'}}{j\omega c_1 j\omega C_{b'e}} = 0 \tag{10}$$

Rearrange and multiply by $1/j$

$$\frac{-1}{\omega C_{b'e}} - \frac{1}{\omega C_1} + \left(\frac{\omega L}{1 - \omega^2 LC}\right)\left(1 + \frac{C_{b'c}}{C_{b'e}}\right) - \frac{g_m r_{b'} C_{b'c}}{\omega C_1 C_{b'e}} = 0 \tag{11}$$

Multiply by $C_{b'e}$

$$-1-\frac{C_{b'e}}{C_1}-\frac{g_m r_{b'} C_{b'c}}{C_1}+\left(\frac{\omega^2 LC_{b'e}}{1-\omega^2 LC}\right)\left(1+\frac{C_{b'c}}{C_{b'e}}\right)=0 \tag{12}$$

$$\frac{\omega^2 L(C_{b'e}+C_{b'c})}{1-\omega^2 LC}=\frac{g_m r_{b'} C_{b'c}+C_{b'e}}{C_1}+1 \tag{13}$$

Divide numerator and denominator of the left side of the equation by $\omega^2 LC$ and rearrange terms.

$$\omega^2 LC=\frac{\left(\frac{g_m r_{b'} C_{b'c}+C_{b'e}+C_1}{C_1}\right)}{\left(\frac{g_m r_{b'} C_{b'c}+C_{b'e}+C_1}{C_1}\right)+\frac{C_{b'e}+C_{b'c}}{C}} \tag{14}$$

Let $\left(\frac{g_m r_{b'} C_{b'c}+C_{b'e}+C_1}{C_1}\right)$ be represented by $S$, and then solve for $f_0$

$$2\pi f_0\sqrt{LC}=\sqrt{\frac{S}{S+\frac{C_{b'e}+C_{b'c}}{C}}}$$

$$f_0=\frac{1}{2\pi\sqrt{LC}}\sqrt{\frac{S}{S+\frac{C_{b'e}+C_{b'c}}{C}}} \tag{15}$$

If $C \gg C_{b'e}+C_{b'c}$ as is usually the case in practical oscillators

$$f_0=\frac{1}{2\pi\sqrt{LC}} \tag{16}$$

The conditions necessary to sustain oscillation are found by setting the real terms of equation 8 to zero

$$r_{b'}+\frac{X_{C_{b'e}} r_{b'}}{X_{C_{b'e}}}+\frac{g_m X_{C_1} X_{C_{b'e}} Z_T}{X_{C_{b'e}}}+g_m X_{C_1} X_{C_{b'e}}=0 \tag{17}$$

$$r_{b'}\left(1+\frac{C_{b'c}}{C_{b'e}}\right)-\left(\frac{g_m}{\omega^2 C_1 C_{b'e}}\right)(1+\omega C_{b'c} Z_T)=0 \tag{18}$$

$$g_m(1+\omega C_{b'c} Z_T)=r_{b'}(C_{b'e}+C_{b'c})\omega^2 C_1$$

$$g_m=\frac{r_{b'}(C_{b'e}+C_{b'c})\omega^2 C_1}{1+\frac{\omega^2 C_{b'c} L}{1-\omega^2 LC}} \tag{19}$$

The gain-bandwidth product ($f_t$) is equal to

$$f_t=\frac{g_m}{2\pi C_{b'e}} \tag{20}$$

In terms of the hybrid-pi equivalent circuit of Fig. 1 the value of $f_t$ required to sustain oscillation

$$f_t \geqslant \frac{2\pi f_0^2 C_1 r_{b'}\left(1+\frac{C_{b'c}}{C_{b'e}}\right)}{1+\frac{C_{b'c}\omega^2 L}{1-\omega^2 LC}} \tag{21}$$

If $C_{b'c} \ll C$

$$f_t \geqslant 2\pi f_0^2 C_1 r_{b'} \left(1 + \frac{C_{b'c}}{C_{b'e}}\right) \tag{22}$$

However

$$f_{\max} = \sqrt{\frac{f_t}{8\pi r_{b'} C_{b'c}}} \tag{23}$$

Substituting for $f_t$

$$f_{\max} \geqslant \frac{f_0}{2} \sqrt{\frac{C_1(C_{b'e} + C_{b'c})}{C_{b'e} C_{b'c}}} \tag{24}$$

# 7

# INTERMEDIATE-FREQUENCY AMPLIFIERS

## 7-1. Introduction

The advantage of the superheterodyne receiver over other receiver systems lies in its ability to reduce its operating frequency. The reduction in frequency allows the receiver amplifiers to have better selectivity and higher gain than they could if they operated at the incoming signal frequency. Let us look at the reasoning behind these conclusions.

From basic considerations we know that the bandpass ($B$) of a tuned circuit is directly proportional to the frequency of resonance and inversely proportional to its circuit $Q$:

$$B = \frac{f_r}{Q}, \tag{7-1}$$

It then becomes clear that a decrease in tuned-circuit resonant frequency will result in reduced bandpass and thereby improved selectivity. For example, assuming the same $Q$, a tuned circuit operating at a frequency of 10 MHz will have a bandpass ten times narrower than a tuned circuit at 100 MHz.

Amplifier gain may be made higher at lower frequencies because feedback through the interelectrode capacitance of the amplifier is less than it would be at higher frequencies. Therefore, the receiver amplifiers are more stable at lower frequencies. This conclusion is confirmed by the expression for maximum stable gain of a single-stage critically coupled double-tuned

vacuum-tube amplifier. Equation (7–2) expresses the maximum gain that can be obtained before the circuit breaks into oscillation. Note that as the frequency is decreased, the maximum stable gain increases.

$$\text{Maximum stable gain} = \left(\frac{0.79 g_m}{2\pi f C_{gp}}\right)^{1/2}, \tag{7–2}$$

where $g_m$ is the transconductance of the tube, $f$ is the frequency of operation, and $C_{gp}$ is the interelectrode capacitance between grid and cathode.

The gain of a tuned amplifier depends upon the magnitude of its tuned plate-circuit impedance. The impedance of a parallel resonant circuit at resonance is $L/CR$. Since a lower operating frequency requires a larger $L/C$ ratio to provide resonance, it becomes apparent that at lower frequencies the plate-load impedance will be larger than at higher frequencies, assuming the same values of $R$ and $C$ in both cases. Therefore, since the voltage amplification of a pentode amplifier is basically equal to $g_m Z_L$, a given value of $g_m$ a low-frequency amplifier will be able to provide greater voltage amplification than would a high-frequency amplifier.

The output of the mixer stage of a receiver, is a signal that is lower in frequency than the input RF signal and is constant in frequency no matter what the input-signal frequency is. The fixed frequency output of the mixer is called the intermediate frequency (IF). Another advantage of the superheterodyne receiver over other receiver systems is due to the fixed-frequency characteristic of the IF amplifiers. This provides constant gain and selectivity throughout the entire range of input-signal frequencies. The constancy of gain and selectivity is due to the fact that both of these circuit characteristics depend upon the tuned-circuit $Q$. The $Q$ in turn depends upon the frequency-sensitive AC resistance of the tuned circuit. It was pointed out in Chapter 4 that skin effect will increase the effective resistance of a conductor as the frequency is increased. Since the frequency of operation of the IF amplifiers is constant, the $Q$ of the amplifiers' tuned circuits will remain constant, and so the gain and selectivity of these amplifiers will also tend to remain constant as the receiver is tuned across the FM band. This would not be so in a receiver system such as a TRF (tuned radio frequency), where all of the RF stages must tune through the entire range of input-signal frequencies.

Another advantage due to fixed-frequency operation of the IF amplifiers is the economy provided by having all the tuned circuits after the mixer fixed-tuned.

## 7-2. The intermediate frequency

The frequency at which the IF amplifier operates is determined by (1) bandwidth considerations, (2) spurious and direct IF response, (3) stage gain and stability, (4) receiver selectivity.

The minimum frequency that can be used as an IF is primarily determined by the bandwidth occupied by a 100 per cent modulated signal. In theory the spectrum of an FM signal is infinitely wide, but in practice all sideband components of the modulated wave that are less than 1 per cent of the unmodulated carrier amplitude may be disregarded. For monophonic FM transmission and a deviation ratio of 5 this would result in a bandwidth of 240 kHz. This may be reduced in practice, since any harmonic distortion generated by sideband clipping of the sidebands farthest removed from the carrier will not be audible anyway, since these components will be higher than the limit of hearing. Most authorities agree that 210 kHz would be an adequate 3-db bandwidth to provide relatively distortion-free reception and yet possess sufficient adjacent-channel suppression.

A stereophonic FM transmission would require a somewhat greater bandwidth. The maximum allowed deviation of the composite stereophonic signal (disregarding the SCA) is 67.5 kHz (90 per cent of 100 per cent modulation), of which 50 per cent is available for the $L - R$ signal. Thus, 33.75 kHz deviation is used for the $L - R$ and 33.75 kHz is used for the $L + R$ information. For an audio signal of 15 kHz the $L - R$ signal consists of a 23-kHz and a 53-kHz sideband spectrum. Since the total deviation is 33.75 kHz, each sideband must be responsible for approximately 17 kHz deviation. The modulation index for the 53-kHz sideband will be 17 kHz/53 kHz, or 0.32. Consequently, the FM spectrum produced for the highest modulating frequency will consist of only two pairs of sidebands. Thus, the bandwidth occupied by a stereophonic FM transmission will be $4 \times 53$ kHz or 212 kHz.

From the above it can be seen that the overall 3-db bandpass of an FM receiver should be about 210 kHz. If a degree of tolerance is allowed, an overall 3-db bandpass of 240 kHz is recommended.

Since the sideband spectral bandwidth may exceed 200 kHz, the lowest intermediate frequency that can be used must be high enough to allow a circuit bandpass this wide. It can be seen that an IF lower than 100 kHz would be impossible, since at 100 per cent modulation the sidebands would extend past zero frequency, and negative frequencies do not exist.

In selecting an intermediate frequency we must also consider the effect of oscillator pulling. That is, if a low-frequency IF is chosen, the local-oscillator frequency will be almost the same as the incoming RF signal. If there is sufficient coupling between the RF and oscillator sections of the receiver, the oscillator will tend to synchronize with the RF and, therefore, the proper IF will not be generated at the mixer output. To minimize this effect a high-frequency IF should be employed.

### *Spurious responses*

Spurious responses are undesired receiver responses that appear at points other than their assigned frequency on the receiver dial. The more common forms are (1) image response, (2) second-order images, (3) second-order

intermodulation products, and (4) direct IF pickup. The first three spurious responses have been discussed in Chapter 5 in terms of the effect that a tuned RF amplifier has upon spurious response. Four main points were brought out in that discussion. (1) To reduce image and other spurious responses, a tuned RF amplifier must precede the mixer. (2) A high-frequency IF tends to discourage image and spurious responses by causing these signals to fall on the extreme edge of the RF response curve. (3) If the IF selected is greater than half the band of FM signals (half of the 88- to 108-MHz band), all of the image frequencies resulting from this band would fall outside the FM band, thereby eliminating a major source of potential image-producing signals. Finally, point 4 indicates that image response caused by local TV stations can be avoided by placing the oscillator frequency higher than the signal frequency.

*Direct IF pickup*

The IF stages are responsible for most of the gain of a superheterodyne receiver. Since the IF amplifiers are tuned to a discrete range of frequencies, signals radiated at these frequencies can enter the IF amplifiers and be highly amplified. If the interfering signal is strong enough, it can cause objectionable interference at the output of the receiver. Undesired signals at the IF frequency can enter the IF amplifiers by direct pickup, owing to insufficient shielding of the IF amplifiers, or by stray coupling between the antenna input and the IF amplifiers. To minimize direct pickup the intermediate frequency selected should be located in a relatively unused area of the electromagnetic spectrum. For example, the frequencies assigned to aircraft or maritime services might be good choices, since these services do not generally employ high-power transmitters and may be only transient in most localities. Most receivers used for the VHF broadcasting band of 88 to 108 MHz have employed intermediate frequencies of 10.7 MHz. This frequency is almost a universal standard, although frequencies as low as 8 MHz and as high as 20 MHz have been used.

All of the factors for determining the IF listed at the beginning of this section are interrelated and may conflict with one another. Thus, if a low frequency is selected, high gain with stability and good selectivity is possible, but oscillator pulling, sideband clipping, and increased spurious responses may result. If a high frequency is selected, interference due to spurious responses is reduced, but at a possible sacrifice of selectivity, stability and gain.

## 7-3. Gain requirements of an FM IF amplifier

The gain requirements of a receiver are largely determined by the minimum receiver signal input expected. Therefore, FCC specifications concerning transmitter signal levels at specified distances can be used as a guide

to establishing the necessary receiver gain. The FCC has specified that within an FM transmitter's urban service area the minimum signal level generated in a receiving antenna whose effective height is 30 feet should be at least 1000 microvolts. In rural service areas for the same antenna 50 microvolts is to be generated. In these areas, signal output from antennas of lower effective height or from indoor antennas may range from 1 to 10 microvolts.

A receiver that employs a Foster-Seeley discriminator must precede the detector with a limiter. To be effective, the limiter requires an input signal of 1 to 2 volts. A receiver using a ratio detector does not require a limiter and will limit properly with a signal level of 1 volt delivered to the diode of the ratio detector. The ratio detector may also allow reduction of the number of IF stages, because the stage preceding the detector can operate at full gain. A limiter preceding a detector, on the other hand, operates at the IF but contributes very little gain to the receiver.

If it is assumed that the signal input to the limiter is to be 1 volt and if the input to the receiver from the antenna is 5 microvolts, the overall required voltage amplification would be 2,000,000. The antenna input and RF stages can account for a voltage amplification of about 50; therefore, the total IF voltage amplification must be 4000. This could be achieved with three identical stages having a voltage amplification of 15.8 or two stages with a voltage amplification of 63.5 each.

We may determine the power-gain requirements for a transistorized FM receiver in somewhat the same way as the voltage amplification of vacuum-tube receivers. Thus, the power input to a receiver resulting from a 5-micro-

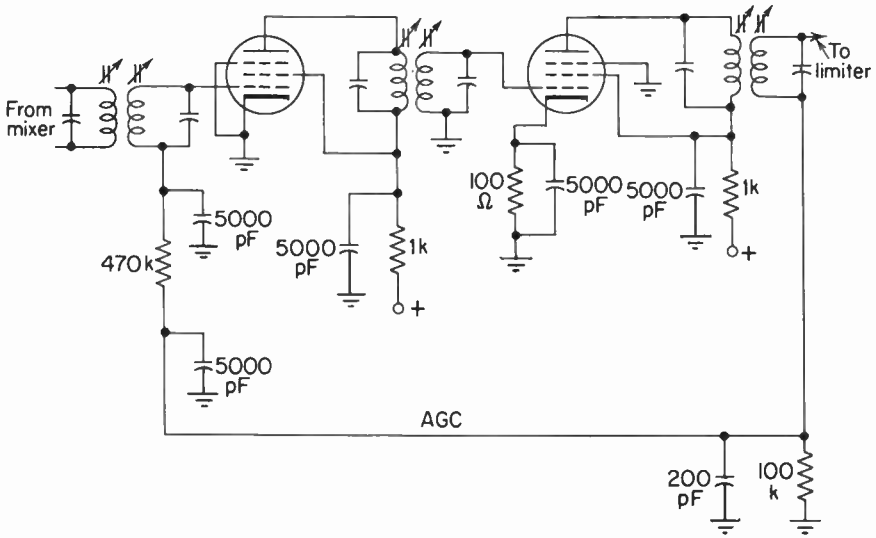



**Fig. 7-1.** A two-stage vacuum tube IF amplifier.

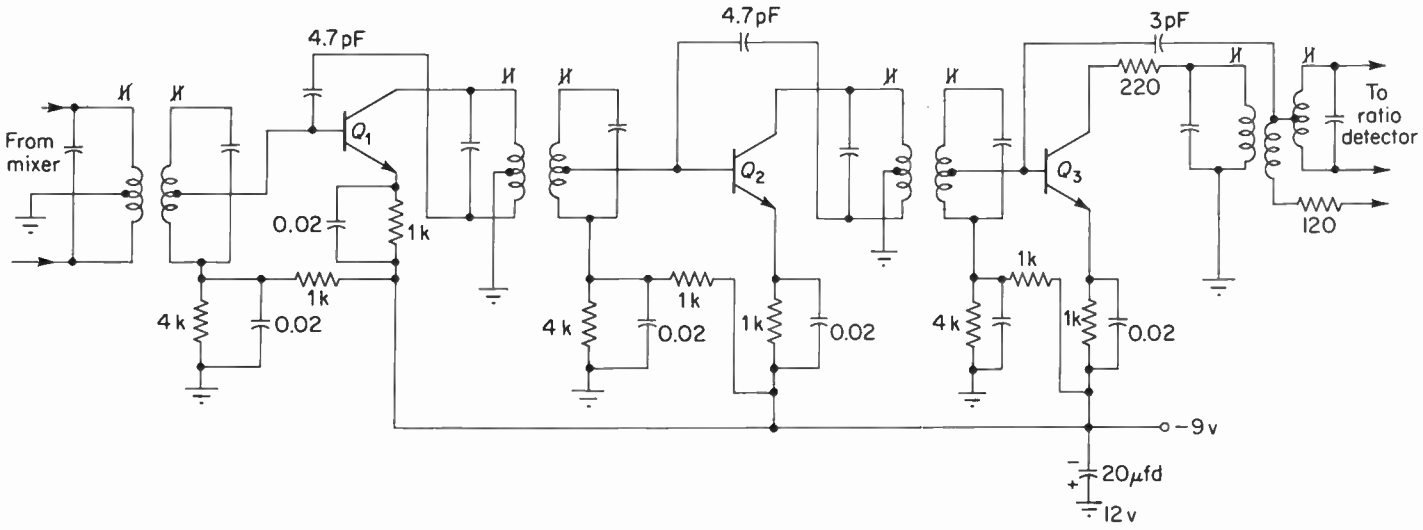



**Fig. 7-2.** A three-stage transistor IF amplifier.

volt antenna signal delivered to a 300-ohm receiver input would be $0.083 \times 10^{-12}$ watt. Assuming that a ratio detector is being used, this would require 1 volt acting on an effective load of approximately 6000 ohms. The power output would then be $0.166 \times 10^{-3}$ watt. The overall power gain required by the receiver may be found from the formula

$$PG = 10 \log \frac{P_o}{P_{in}}. \tag{7-3}$$

Solving Eq. (7–3) indicates that 93 db would be the necessary overall power gain. The RF stages contribute about 20 db. The IF stages must therefore provide the remaining 73 db gain. If three identical IF stages are used, each stage would be designed to supply approximately 24 db gain. A four-stage IF amplifier could develop 73 db gain if each stage contributed about 18 db gain.

The number of IF stages used in an FM receiver depends upon such factors as cost, stability, and adjacent-channel selectivity. If cost is no object, an IF amplifier consisting of many low-gain stages is preferable to few high-gain stages. An amplifier operated at low gain exhibits greater stability; furthermore, detuning of stages that use AGC is less troublesome. A lower stage gain, because it requires a greater number of stages, also necessitates more tuned circuits, so that adjacent-channel selectivity can be improved. For this reason transistor FM receivers may have as many as four stages of IF amplification. A tuned transistor amplifier may require more stages than a vacuum-tube amplifier, because the low input resistance of a transistor amplifier is responsible for tuned-circuit loading that tends to make good selectivity difficult to achieve. Figure 7-1 is the schematic diagram of a two-stage vacuum-tube IF amplifier, and Fig. 7-2 is the schematic diagram of a typical three-stage transistor IF amplifier.

## 7-4. The FM receiver IF amplifier tuned circuits

At least three different single-frequency types of IF amplifiers are found in commercial FM receivers. These are the single-tuned (Fig. 7-3), the double-tuned (Fig. 7-4), and the double-tuned overcoupled amplifier. The double-tuned overcoupled amplifier differs from the doubled-tuned amplifier only in the degree of coupling between windings of the interstage transformer.

These three amplifiers are similar in that they are operated as linear class A, except in those situations where the IF amplifier also acts as a limiter. The differences between these IF amplifiers lie in the characteristics of the interstage coupling networks. These tuned networks are responsible for the $L/C$ ratio, which determines the stage gain. Furthermore, their bandpass characteristics determine the adjacent-channel selectivity as well as the harmonic distortion that appears at the output of the detector.

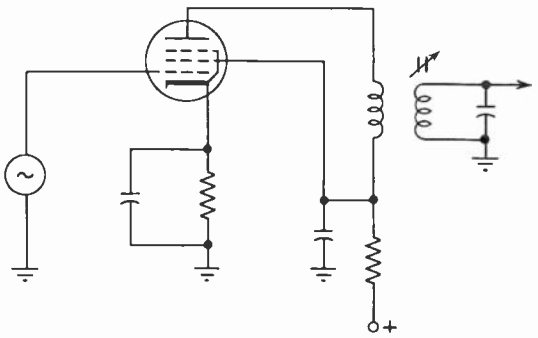


Fig. 7-3. A single-stage single-tuned amplifier.

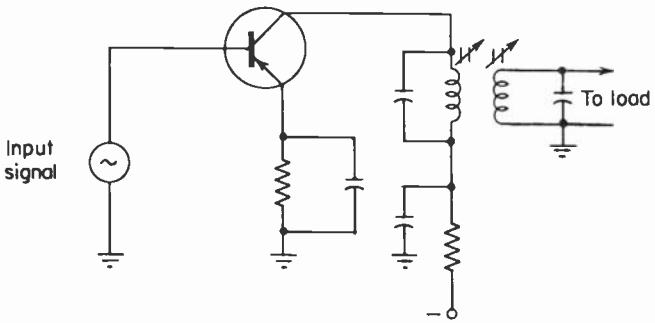



Fig. 7-4. A transistor amplifier which employs double tuned transformer coupling.

The ideal selectivity curve for an IF amplifier is shown in Fig. 7-5. Notice that the passband is perfectly flat, thereby providing uniform amplification for all frequencies within the passband. Notice also that the response is zero for all frequencies outside the passband. Unfortunately, such an ideal response is very difficult to achieve in practice, although close approximations may be realized by proper design of the amplifier interstage network.

As a means of comparison between the ideal rectangular response and a given practical response curve, Deutsch† has defined the "steepness ratio" $S$ as

$$S = \frac{\text{the bandwidth at a relative gain of 0.1}}{\text{the bandwidth at a relative gain of 0.9}}. \tag{7-4}$$

A rectangular response (ideal) would have an $S$ of unity. Thus, the closer a given response is to unity, the more closely the curve resembles the ideal response. It has been shown by Deutsch that the $S$ ratio for a single-stage single-tuned amplifier is 20.5. The $S$ ratio for a double-tuned critically coupled amplifier of the same 3-db bandwidth is 4.53; and for an overcoupled amplifier of 3-db bandwidth, where the coefficient of coupling is twice that of crit-

†Sid Deutsch, *Theory and Design of Television Receivers*. McGraw-Hill Book Company, New York, 1951.

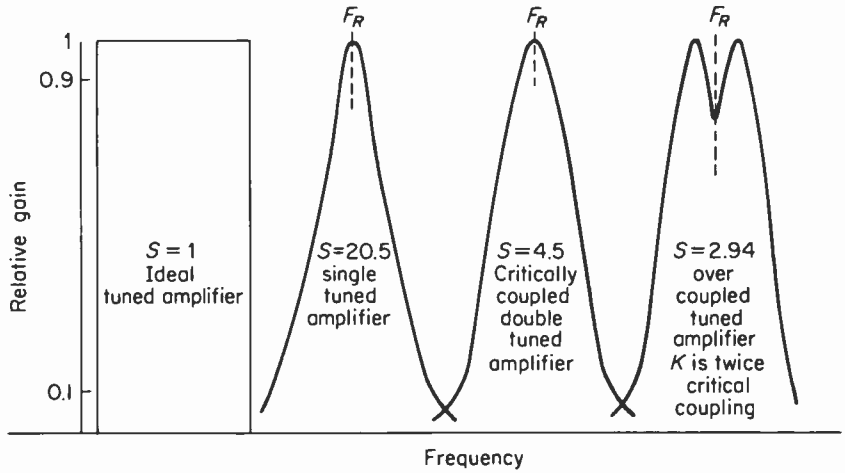



**Fig. 7-5.** Comparison between an ideal tuned amplifier response and that of practical tuned amplifiers.

ical coupling, $S$ is 2.94. It is evident that a critically coupled double-tuned amplifier would be much superior to the single-tuned amplifier in terms of adjacent-channel rejection. In the case of the overcoupled tuned amplifier the $S$ ratio is closest to the ideal of unity, but the double-humped response (see Fig. 7-5) makes alignment difficult as well as introducing considerable distortion. For these reasons the overcoupled double-tuned amplifier is seldom used in practical FM receivers, and we shall not discuss it in detail here.

### *FM receiver distortion*

In conventional AM receivers the operating point of the RF and IF amplifiers determines the degree of harmonic and intermodulation distortion produced at the output of the detector. However the selective tuned circuits reject almost all of the harmonic components generated by nonlinear amplifier operation since these components fall outside the bandpass of the tuned circuits. Distortion that does occur because of nonlinear tube operation is due to third-order effects. These effects introduce cross-modulation and cause the amplification of the stage to vary when the input signal is a modulated carrier. Because the amplification varies between peak and trough of the modulated signal, the output modulation envelope becomes distorted. If it is assumed that the bandpass is broad enough to pass all sideband components, the tuned circuit will not introduce any distortion.

Distortion is introduced into an FM receiver RF and IF stages in a somewhat different manner. Nonlinear operation of the amplifier does not cause any distortion, because the effect of second-harmonic and higher-

order distortion components is to cause amplitude variations in the frequency-modulated signal. These amplitude variations are removed by the limiter, and assuming that the limiter and detector stages introduce no distortion of their own, the output of the detector will be distortion-free. Distortion generated in the IF amplifiers of an FM receiver is due to phase nonlinearity of the tuned coupling circuit. That is, changes in frequency result in nonlinear changes in the phase relationships that exist in the tuned coupling circuits, thereby introducing indirect frequency modulation.

It was mentioned in Chapter 1 that an FM signal theoretically consists of an infinite number of sidebands and a carrier component. Each pair of sidebands is separated from an adjacent pair of sidebands by an amount equal to the audio modulating frequency. Furthermore, all resultant odd-numbered sideband pairs (1, 3, 5, 7, . . . .) will be displaced from the carrier and the resultant even-numbered sideband pairs (2, 4, 6, . . .) by 90 degrees when the modulating signal is at 90 degrees. The amplitude of the carrier and the sideband components depends upon the deviation and the audio modulating frequency (modulation index). The phasor addition of the carrier and all of the sideband components for any given modulation index will always provide a constant resultant carrier. Figure 7-6 shows the development and the phasor addition of the carrier and the significant sidebands for a modulation index of 5.

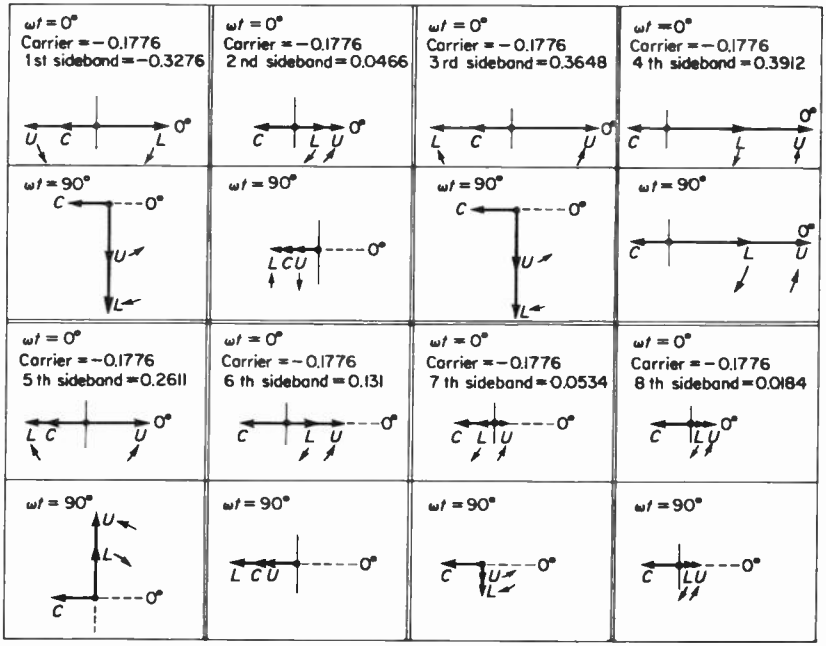
If an FM signal is passed through a linear resistive network of infinite bandpass, none of the sideband components will be shifted in phase or altered in amplitude. Therefore, the output of such a network would be identical to the input, and no distortion will be introduced by the network. This, of course, represents the ideal situation, which cannot be achieved in practice.

A practical IF amplifier must use reactive circuit elements. Therefore, a phase shift must be expected between the input to the circuit and the output. However, if the FM signal is passed through a network that passes all sidebands without attenuation and that introduces a phase shift proportional to frequency,† no distortion will occur. A graph showing such a linear phase characteristic is given in Fig. 7-7. This type of phase characteristic does not introduce any harmonic distortion because the time delay ($d\phi/d\omega$) is constant for all components of the FM signal. A constant time delay for the carrier and all sidebands means that their relative phase relationships will not be disturbed. Thus, neither phase nor amplitude modulation will be introduced.

For example, if a phase linear network introduced a constant time delay of 0.02 microsecond‡ at all frequencies, it would produce the graph shown

†The condition where equal changes in frequency correspond to equal changes in phase.

‡In practice time delays of 1.5 to 2 microseconds are typical.



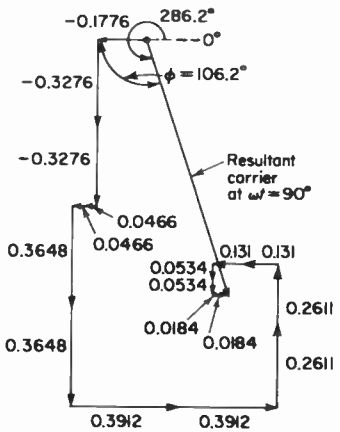




**Fig. 7-6.** Each of the above smaller diagrams indicates the relative magnitude and phase relationships between the carrier and the 8 significant sideband pairs of a modulated FM signal whose modulation index is 5. Each sideband pair is shown with respect to the carrier at two time intervals; (a) when the modulating signal ($wt$) is at 0° and (b) when it is at 90°. In each case the direction of sideband rotation relative to the carrier is indicated. The number of rotations of each sideband pair is relative to the carrier rotation. Thus, the 3rd sideband pair rotates three times faster than the carrier, and the 5th sideband pair rotates five times faster than the carrier, etc. For this modulation index the carrier and first sideband pair are negative, thus they are rotated 180° as indicated. The larger resultant vector diagram shows the magnetude and relative phase of the modulated carrier and the sideband pairs for each of the smaller diagrams above. All phasor diagrams are based upon Eq. 1-14.

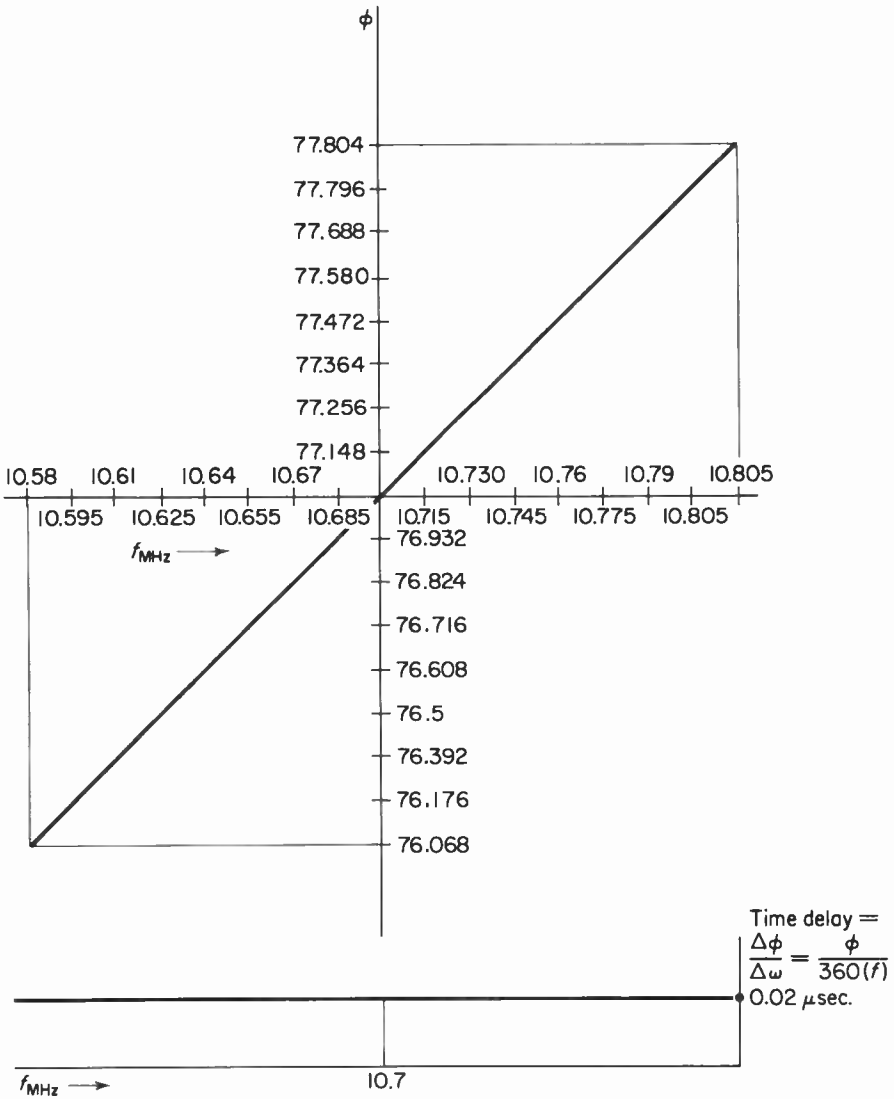



**Fig. 7-7.** An ideal phase linear characteristic and its associated constant time delay.

in Fig. 7-7. Thus at 10.7 MHz the phase shift would be 77.04 degrees, and would be directly proportional with frequency. An FM signal having a modulation index of 5 would have the phase relationships shown in Fig. 7-6. Thus when this signal passes through the phase linear network (described in Fig. 7-7), the phase relationships will shift to those of Fig. 7-8. Note that

the resultant carrier at the output of the phase linear network is undisturbed, in that it is still 106.5 degrees from the carrier component position. Note also that because of the coupling network's phase linearity, the relative phase and amplitude relationships of the resultant sideband components have not been altered.

If the selective tuned circuits used in the IF amplifier produce a phase shift that is not proportional to frequency (see Fig. 7-9), the 90-degree phase difference between the resultant even and odd sideband pairs will be modified. The shift will be larger or smaller depending upon the degree of nonlinearity. As a result, the phasor addition of these components will not produce a resultant carrier of the same phase or amplitude as that of an equivalent FM signal passing through a phase-linear network. This implies that during modulation the resultant carrier amplitude will vary, and thus amplitude modulation at primarily odd harmonics of the modulating frequency will occur.

A properly designed limiter can remove the amplitude modulation but will not be able to remove the indirect FM produced by phase modulation. In this case phase modulation is the consequence of the additional resultant carrier phase shift, which varies throughout the modulating cycle. Since the deviation due to phase modulation is directly proportional to the change in phase angle and the audio frequency ($FM = \Delta\phi f_a/57.3$), the distortion due to phase modulation will increase with modulating frequency and deviation. This distortion occurs as odd harmonics of the input signal. The third harmonic is generally the strongest harmonic component produced. The percentage of odd-harmonic distortion for a critically coupled double-tuned

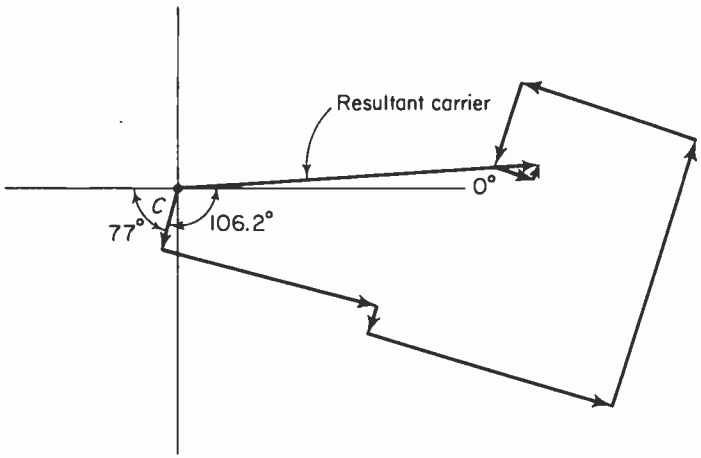



**Fig. 7-8.** The phase relationships between the carrier and sidebands ($MI = 5$) after passing through a network whose phase shift is proportional to frequency.

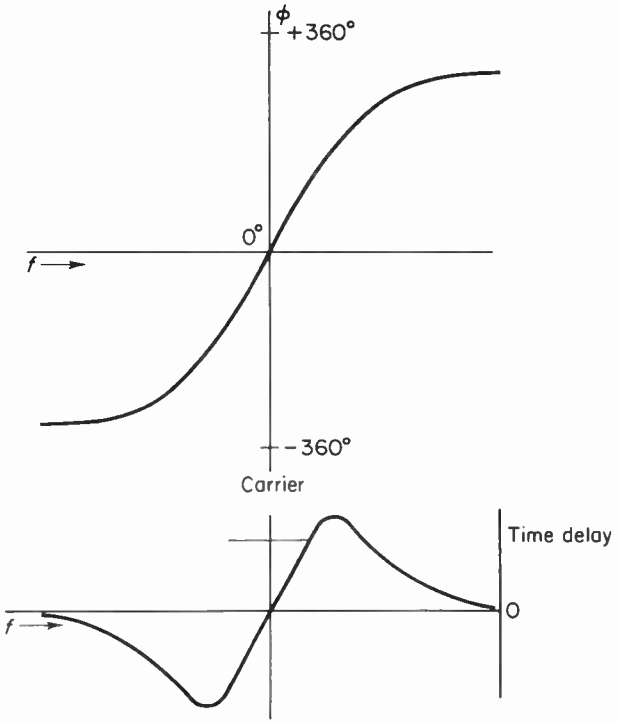



**Fig. 7-9.** A nonlinear phase response and its associated time delay.

amplifier is given by Jaffe† as

$$D_n = \frac{2(\sqrt{1+\rho^2}-1)^n}{\rho^n(\Delta\omega/\lambda)} \times 100, \tag{7-5}$$

where $D_n$ is the percentage $n$th odd-harmonic distortion, $B$ is the bandwidth in kilohertz at the 3-db points, $\rho = \sqrt{2}\,(2\,\Delta\omega/B)$, $\Delta\omega/2\pi$ is the peak frequency swing, and $\lambda/2\pi$ is the modulating frequency.

It has been shown by Ross‡ that for a critically coupled IF transformer, if the product of the loaded $Q_L$ and the ratio

$$\frac{2\text{ (sideband pairs) (maximum audio frequency)}}{\text{frequency of operation}}$$

is numerically equal to 2 or less, then the IF phase response will be substantially linear. Thus, if these conditions are met, minimum harmonic distortion will be introduced by the IF amplifiers.

†D. L. Jaffe, "A Theoretical and Experimental Investigation of Tuned-Circuit Distortion in Frequency-Modulation Systems," *Proc. IRE*, May 1945.

‡Ross, H.A., "The Theory and Design of Intermediate-Frequency Transformers for Frequency Modulated Signals," AWA Technical Review, 6:8 (March 1946).

*Amplitude modulation*

If the bandpass of the IF amplifier is too narrow, phase and amplitude modulation will be generated during the deviation of the FM signal. The amplitude modulation that appears at the output of the IF amplifiers is the result of the basic amplitude-versus-frequency characteristic of a narrow-band filter. Figure 7-10 illustrates how a change in frequency causes a change in output amplitude. It can be seen that the response curve of the IF amplifier is being used as a transfer curve. Notice that as the input frequency becomes lower than the resonant frequency, the output signal across the tuned circuit decreases in amplitude. As the deviating input signal increases above resonance, the output IF amplitude also decreases. Assuming that the FM signal is modulated with a 1-kHz tone, the output of the tuned circuit will be an FM signal whose amplitude varies at a 2-kHz rate. The shape of the modulation envelope will be the same as the shape of the response curve.

The percentage of amplitude modulation introduced by a critically coupled double-tuned amplifier may be determined from the curves developed by Ross and shown in Fig. 7-11. These curves indicate that the percentage

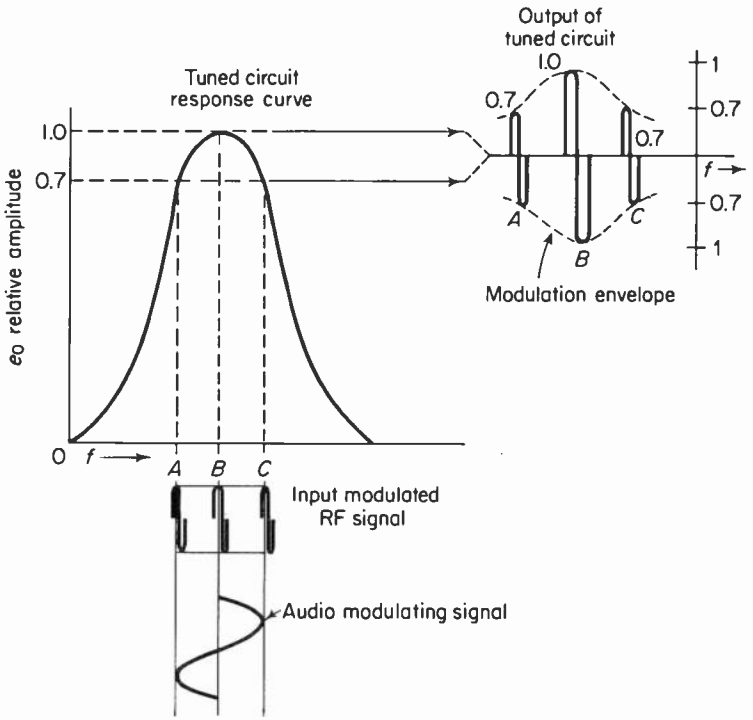



**Fig. 7-10.** The above diagram illustrates how a narrow bandpass tuned circuit generates amplitude modulation from frequency modulation.

of amplitude modulation introduced depends upon the single-stage $Q$ (primary and secondary of the tuned transformer are assumed equal), the number of equal-$Q$ transformers used in the amplifier, and the total deviation ($2\,\Delta f$).

From Fig. 7-11 we can see that when the product of $2\,\Delta f/f_{op}$ and the $Q$ of the tuned circuit is less than 0.4, the percentage of amplitude modulation is almost zero. As the product increases to 2, the percentage of AM increases to 38 per cent for a single transformer to 92 per cent for four tuned transformers.

Let us use the curves of Fig. 7-11 to calculate the percentage of AM introduced in an amplifier employing four tuned transformers operating at a frequency of 10.7 MHz and whose overall bandwidth is 240 kHz.

In order to use these curves the product of the tuned circuit $Q$ and the quotient of twice the deviation ($2\Delta f$) divided by the operating frequency ($f_{op}$) will have to be found. From considerations of cascaded critically coupled double-tuned amplifiers, Terman† has determined that an amplifier with four tuned transformers will have an overall bandwidth of 0.66 of each individual stage. Thus, for this example each stage will have a bandwidth of 240 kHz/0.66

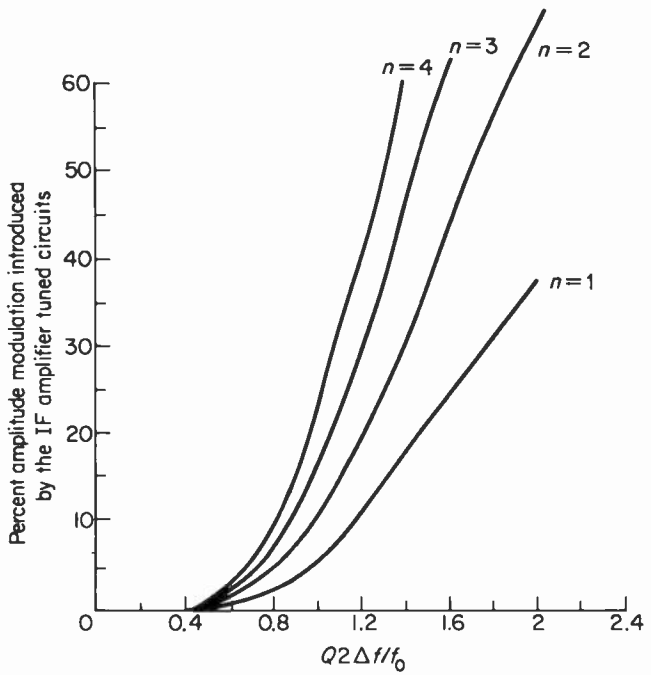



**Fig. 7-11.** The percentage of A-M introduced by $n$ tuned critically coupled transformers. The $Q$ of the primary equals the $Q$ of the secondary.

†Terman, F.E., Electronic and Radio Engineering, 4th ed. McGraw-Hill Book Company, New York, 1955.

or 365 kHz. Elementary consideration of a double-tuned circuit shows that its bandwidth is 1.414 of that of a single-tuned circuit of equal $Q$. The $Q$ of the tuned transformers will be found to be $f_r\ 1.414/B = 41.4$. Assuming 100 per cent frequency modulation, $2\ \Delta f/f_{op}$ equals 0.014; thus the product of $Q(2\ \Delta f/f_{op})$ is 0.577. From Fig. 7-11 the percentage of amplitude modulation introduced in this system is 2.5 per cent.

The amount of introduced AM that can be tolerated will depend upon the ability of the limiter to remove amplitude variations, the deviation ratio, and the application for which the receiver was intended. The generally accepted maximum values of introduced amplitude modulation fall between 20 and 40 per cent. It is most desirable to keep the AM as close to zero per cent as possible, if distortion due to nonlinear phase characteristics is to be avoided.

## 7-5. Vacuum-tube IF amplifiers

Figure 7-12 illustrates a typical schematic diagram of a vacuum-tube double-tuned IF amplifier. Usually this type of amplifier is operated class A, although in some cases it may serve the function of limiter and be operated as an overdriven amplifier. The tube is a pentode, and since the stage is in the grounded-cathode configuration, this insures amplifier stability without neutralization. The frequency of operation is 10.7 MHz, which is the standard frequency used by all commercial receiver manufacturers. Automatic volume control may supply the bias for the stage.

Schematically this circuit is the same as that of IF amplifiers found in AM broadcast receivers; however, it differs from them in bandwidth and frequency of operation. This difference leads to practical difficulties in maintaining amplifier stability. Such things as lead dress and common ground

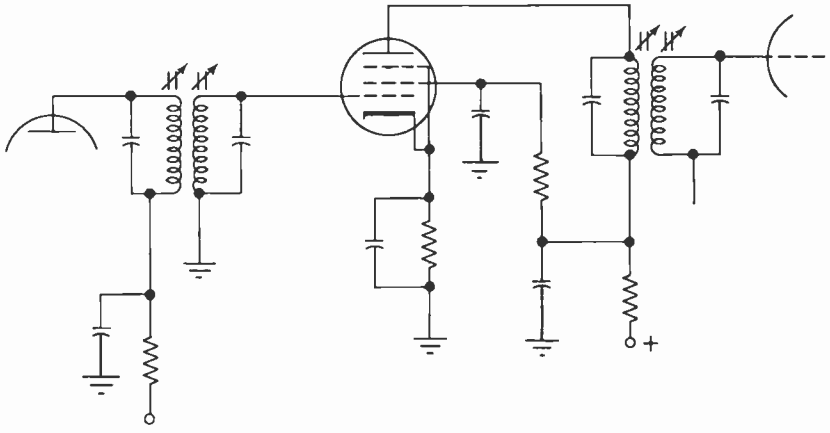


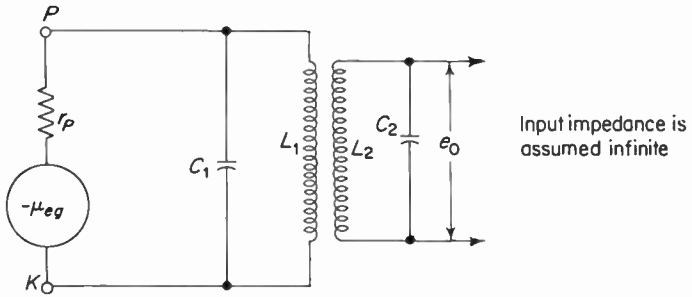

**Fig. 7-12.** Vacuum tube FM IF single stage amplifier.

connections require careful positioning as the frequency of operation increases.

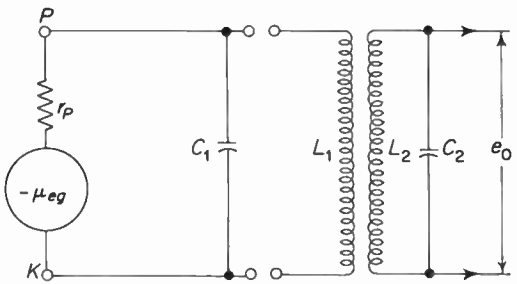
*Voltage amplification*

The voltage amplification of a single-stage critically coupled double-tuned amplifier may be determined from its constant-voltage equivalent circuit.

First the output circuit of Fig. 7-12 is converted into its Thevenin equivalent circuit. This is done by using the constant-voltage equivalent circuit of the vacuum tube shown in Fig. 7-13, and then disconnecting the primary inductance from its resonating capacitance. The resulting series circuit is




Equivalent circuit of the output circuit of a vacuum tube double tuned transformer coupled amplifier




Step one in converting into a Thevenin equivalent circuit

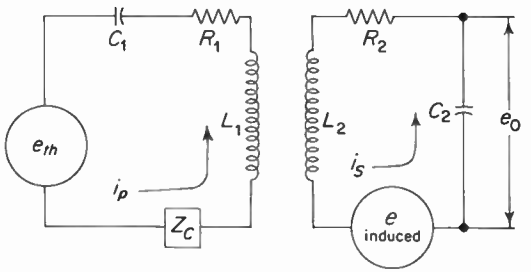



**Fig. 7-13.** Thevenin equivalent circuit of a vacuum tube double tuned transformer-coupled amplifier used in deriving its voltage amplification.

then reduced to a simple Thevenin equivalent. The Thevenin voltage ($e_{th}$) is the voltage across $C_1$ and is found to be

$$e_{th} = \mu e_g\left(\frac{-jX_{C_1}}{r_p - jX_{C_1}}\right)$$

but $r_p \gg X_{C_1}$

$$e_{th} = \mu e_g\left(\frac{-jX_{C_1}}{r_p}\right) = \frac{\frac{1}{j\omega C_1}}{r_p} = \frac{\mu e_g}{j\omega C_1 r_p}$$

but $\mu/r_p = g_m$

$$e_{th} = -jg_m e_g X_{C_1} \tag{7-6}$$

The Thevenin impedance is found by replacing the source ($-\mu e_g$) with its internal impedance ($r_p$) and determining the total impedance looking back toward the source from the primary inductance. This becomes

$$Z_{th} = -\frac{jX_{C_1} r_p}{r_p - jX_{C_1}} \tag{7-7}$$

If it is assumed that $r_p \gg X_{C_1}$ the expression may be simplified to

$$Z_{th} \cong -jX_{C_1} \tag{7-8}$$

The above simplifications result in the equivalent circuit shown in Fig. 7-13.

The AC plate current ($i_p$) is found by Ohm's law and is

$$i_p = \frac{e_{th}}{Z_p + Z_c}$$

where $Z_p$ is the primary impedance and is equal to $R_1 + j\omega L_1 - j/\omega C_1$.

$Z_c$ is the reflected impedance equal to $(\omega M)^2/Z_s$. Where $M$ is the mutual inductance and $Z_s$ is the secondary impedance. Therefore

$$i_p = \frac{e_{th}}{R_1 + j\omega L_1 - j/\omega C_1 + \frac{(\omega M)^2}{Z_s}} \tag{7-9}$$

The secondary current is also determined by Ohm's law and is

$$\begin{aligned} i_p &= \frac{e_{\text{induced}}}{Z_s} \\ &= \frac{j\omega M i_p}{Z_s} = \frac{j\omega M e_{th}}{Z_s\left(R_1 + j\omega L_1 - j/\omega C_1 + \frac{(\omega M)^2}{Z_s}\right)} \\ &= \frac{j\omega M e_{th}}{Z_s(R_1 + j\omega L_1 - j/\omega C_1) + (\omega M)^2} \end{aligned} \tag{7-10}$$

Since the voltage amplification is being determined at resonance, $X_{L_1} = X_{C_1}$. Therefore the secondary current may be simplified to

$$i_s = \frac{j\omega M e_{th}}{Z_s R_1 + (\omega M)^2}$$

but the secondary impedance is resistive and is equal to $R_2$. Hence

$$i_s = \frac{j\omega M e_{th}}{R_1 R_2 + (\omega M)^2} \tag{7-11}$$

The $Q$ of a coil is defined as $X_L/R$, where $X_L$ is the reactance of the coil and $R$ is the series resistance of the coil and associated circuitry. For a series resonant circuit $R$ is the impedance of the circuit. Thus, rearranging the expression for $Q$, the secondary impedance $R_2$ is equal to $\omega L_2/Q_2$, and the primary resistance $R_1$ is equal to $\omega L_1/Q_1$. If the above as well as the expression for mutual inductance ($M = k\sqrt{L_1 L_2}$) is substituted in the expression for secondary current, the result is

$$i_s = \frac{j\omega k\sqrt{L_1 L_2}\, e_{th}}{\dfrac{\omega L_2}{Q_2}\dfrac{\omega L_1}{Q_1} + \omega^2 k^2 L_1 L_2}. \tag{7-12}$$

Replacing $e_{th}$ with $-jg_m e_g/\omega C_1$,

$$i_s = \frac{+j\omega k\sqrt{L_1 L_2} - jg_m e_g}{\left(\dfrac{\omega L_2}{Q_2}\dfrac{\omega L_1}{Q_1} + \omega^2 k^2 L_1 L_2\right)\omega C_1}. \tag{7-13}$$

But $i_s = je_o\omega C_2$, therefore

$$je_o\omega C_2 = \frac{g_m e_g \omega k\sqrt{L_1 L_2}}{\left(\dfrac{\omega L_2}{Q_2}\dfrac{\omega L_1}{Q_1} + \omega^2 k^2 L_1 L_2\right)\omega C_1}, \tag{7-14}$$

$$e_o = \frac{-jg_m \omega k\sqrt{L_1 L_2}}{\left(\dfrac{\omega L_2}{Q_2}\dfrac{\omega L_1}{Q_1} + \omega^2 k^2 L_1 L_2\right)\omega^2 C_1 C_2}, \tag{7-15}$$

$$\frac{e_o}{e_g} = \frac{-jg_m e_g \omega k\sqrt{L_1 L_2}}{\left(\dfrac{\omega L_2}{Q_2}\dfrac{\omega L_1}{Q_1} + \omega^2 k^2 L_1 L_2\right)\omega^2 C_1 C_2}, \tag{7-16}$$

$$\frac{e_o}{e_g} = \frac{-jg_m \omega k\sqrt{L_1 L_2}}{(\omega^2 C_1 C_2)(\omega^2 L_2 L_1)\left(\dfrac{1}{Q_1 Q_2} + k^2\right)}. \tag{7-17}$$

But at resonance $\omega L_1 = 1/\omega C_1$ and $\omega L_2 = 1/\omega C_2$; also $Q_1 = Q_2$ and $L_1 = L_2$; therefore

$$A_v = \frac{e_o}{e_g} = \frac{-jg_m \omega k L}{\dfrac{1}{Q_2} + k^2}. \tag{7-18}$$

Let $k = 1/Q$ (critical coupling); then

$$A_v = \frac{-jg_m \omega L Q}{2}. \tag{7-19}$$

The expression for the voltage amplification of a critically coupled double-tuned amplifier, whose primary and secondary inductance and $Q$ are equal,

is interesting in that it is similar to the expression for a single-tuned amplifier that is divided by the factor 2. Thus, an equivalent single-tuned amplifier tuned to resonance can be expected to have twice the gain of a double-tuned amplifier.

Amplifier stability will be insured and neutralization will not be necessary provided the gain of the double-tuned amplifier does not exceed the maximum stable gain determined by Eq. (7–2).

*Bandwidth*

Most commercial FM receivers incorporate at least two stages of intermediate-frequency amplification and a limiter stage. This would ordinarily mean a total of three double-tuned transformers, not including the detector transformer. It is well known that the overall bandpass of a tuned filter decreases as the number of tuned transformers is increased. Thus, if the number of stages of IF amplification is increased as a means of obtaining the necessary overall gain, the overall bandpass of the amplifier will decrease. This can be offset by increasing the individual stage bandpass by an amount that would provide the desired overall bandpass.

The overall 3-db bandpass ($B$) of a critically coupled double-tuned IF amplifier of identical stages can be determined from the following expression:

$$B = B_2\sqrt[4]{2^{1/n} - 1}, \tag{7–20}$$

where $n$ is the number of identical stages and $B_2$ is the bandwidth of one stage. If the relative bandpass of a single stage is considered to be unity, Eq. (7–20) predicts that two tuned circuits will provide a relative bandpass of 0.8, three tuned circuits a bandpass of 0.71, and four tuned circuits will decrease the overall bandpass to 0.66 of the bandpass ($B_2$) of one tuned circuit.

For example, if a three-stage amplifier required an overall bandpass of 240 kHz, each stage would be designed to have a 3-db bandpass of 240 kHz/0.71 = 338 kHz.

## 7-6. The transistor amplifier

A typical transistor IF amplifier is shown in Fig. 7-14. In many respects this amplifier is similar to the front-end tuned RF amplifiers discussed in Chapter 5. Thus, both RF and IF amplifiers may be operated linearly. However, on strong signals the IF amplifier may be designed to behave as a limiter. This is accomplished by reducing the DC emitter-to-collector voltage of the stages before the detector, so as to shift the AC load line toward zero. Thus, on strong signals the transistor operates between saturation and cutoff, and behaves as an overdriven amplifier.

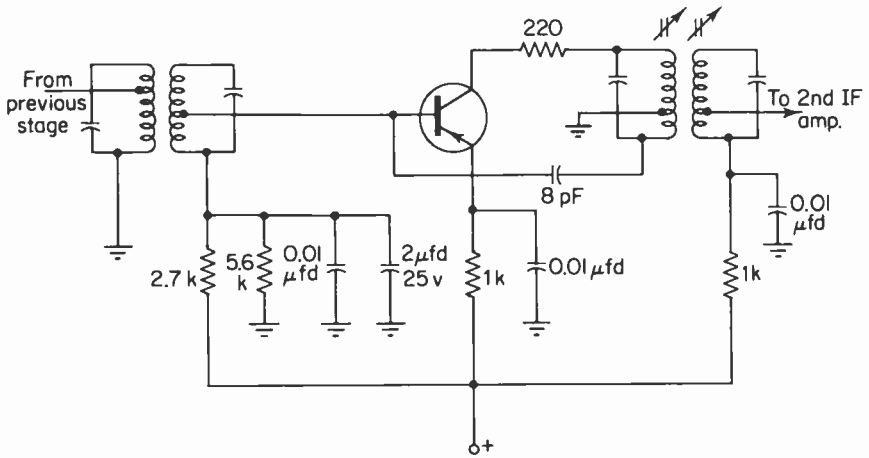



**Fig. 7-14.** A typical neutralized transistor IF amplifier.

Radio-frequency and intermediate-frequency amplifiers are also similar in that both may use either forward or reverse AGC as a means of avoiding overloaded conditions on strong signals. In this way AGC also introduces a certain amount of limiting.

Both types of amplifiers must avoid instability by such techniques as neutralization or mismatching, which in effect reduce the gain of the amplifiers below the critical levels that may cause oscillation. Mismatching or matching is accomplished in both amplifiers by tapped transformers.

Thus, from a schematic point of view both types of amplifiers may appear similar. However, intermediate-frequency amplifiers differ from front-end RF amplifiers in at least five respects:

1. Intermediate-frequency amplifiers operate at considerably higher gain than do front-end radio-frequency amplifiers. A single-stage RF amplifier may be able to provide a power gain of about 15 db; a single-stage IF amplifier may be able to provide a power gain of approximately 25 db.

2. Low-noise IF amplifiers are not as critical for weak-signal response as are low-noise front-end RF amplifiers.

3. The bandwidth and skirt-response characteristics of the IF amplifier are more demanding than those of the RF amplifier. Higher frequency of operation and input-circuit loading result in an RF-amplifier bandpass of several megacycles, whereas in IF amplifiers, the lower frequency of operation and the type and number of tuned circuits used result in a relatively narrow bandwidth and characteristically good skirt response.

4. Intermediate-frequency amplifiers usually employ critically coupled double-tuned transformers as a means of coupling between stages. The number of tuned transformers in the IF section of the receiver is always one more than the number of stages. Thus, a three-stage IF amplifier will have four

tuned transformers between the mixer output and the limiter input. RF amplifiers usually use either single-tuned or impedance-coupled circuits between stages, and only two tuned transformers between the antenna and the mixer input.

5. The signal level at the IF amplifier may be a million times larger than the signal level handled by the RF amplifiers. Thus, under strong-signal conditions, the AC component of collector voltage may be large enough to cause the collector voltage to drop below the base voltage, causing the base-to-collector junction to become forward-biased. As a result, feedback from collector to base during this time will be great enough to cause instability. To prevent such instability in IF amplifiers, a small resistor of about 200 ohms is placed in series with the collector as shown in Fig. 7-14.

### *Transistor intermediate-frequency amplifier gain*

The gain of a transistor amplifier may be determined by converting the amplifier and its coupling circuits into an appropriate equivalent circuit and then using network analysis and algebraic manipulation. The equivalent circuit may also be used to determine the bandpass characteristics, the power matching requirements between stages, the stability of the system, and other important design information.

Transistor amplifiers operating at high frequencies are most conveniently analyzed by converting the transistor amplifier into either a hybrid-$\pi$ or a $Y$-parameter equivalent circuit. An example of the use of $Y$-parameters in RF amplifier design is to be found in Chapter 5. In this chapter the hybrid-$\pi$ equivalent circuit will be used to determine the characteristics of a double-tuned IF amplifier.

Figure 7-15 shows the hybrid-$\pi$ equivalent circuit of a single-stage common-emitter double-tuned amplifier, such as that shown in Fig. 7-14. Also shown in the figure are typical values for the parameters of a high-frequency transistor. Using this equivalent circuit, let us assume that a three-stage intermediate-frequency amplifier is to be designed to meet the following requirements:

1. All three stages are identical and operate at an IF of 10.7 MHz.
2. Overall 3-db bandpass is to be 240 kHz.
3. All coupling transformers are critically coupled.
4. Amplitude modulation produced by the tuned circuits must be less than 3 per cent.
5. Harmonic distortion must meet the Ross criterion.
6. The power gain of each stage must be at least 25 db to provide an overall IF gain of 75 db.
7. Determine from the maximum-stable-gain criterion whether or not neutralization is required.

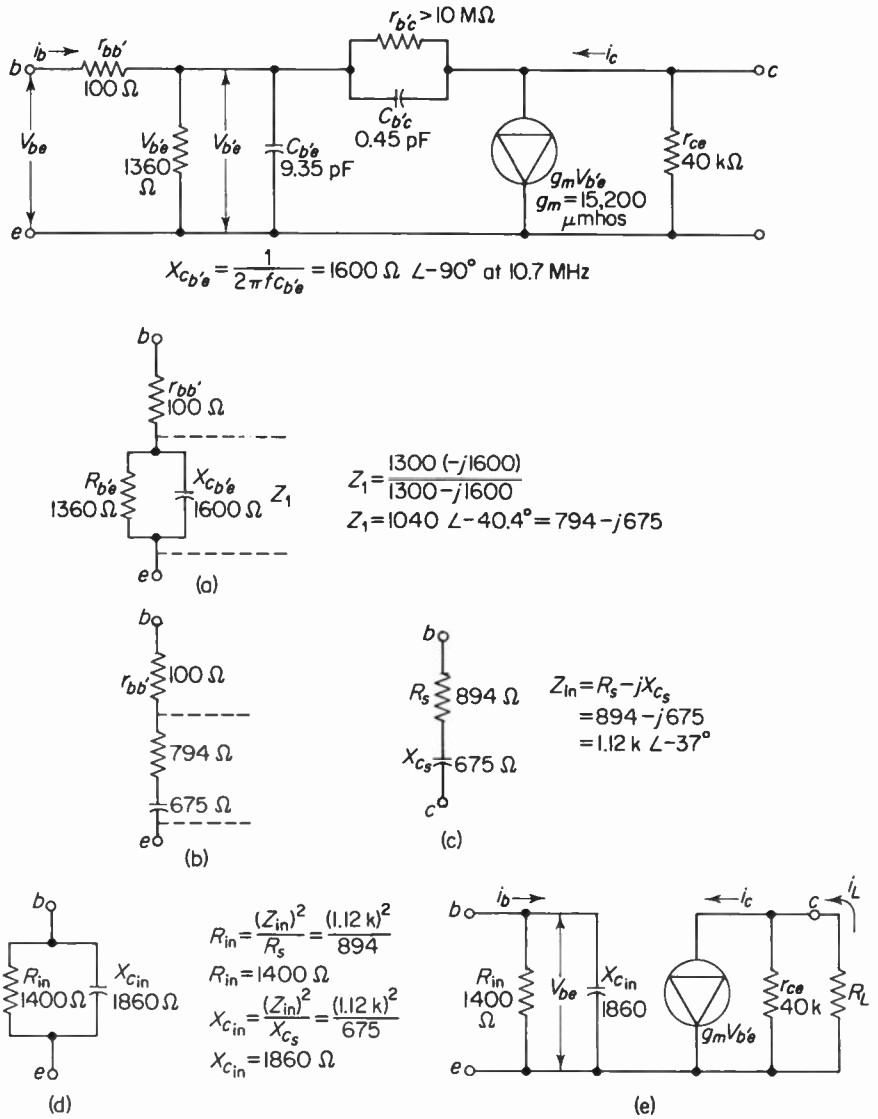



**Fig. 7-15.** The hybrid $\pi$ equivalent circuit of a high frequency transistor. Also illustrated is the step-by-step simplification of the series-parallel input circuit to that of the simple parallel circuit in fig. (e).

The transistor used in this amplifier will be assumed to have the following hybrid-$\pi$ parameters at the frequency of operation and at the DC voltages in the circuit. The lead and base resistance $r_{bb'}$ is 100 ohms, the base-to-emitter resistance $r_{b'e}$ is 1360 ohms, the feedback resistance $r_{b'c}$ is 10 megohms, the collector-to-emitter resistance $r_{ce}$ is 40,000 ohms, the depletion plus

the diffusion capacitance $C_{b'c}$ is 9.3 pF, the $g_m$ is 15,200 micromhos, and the feedback capacitance $C_{b'c}$ is 0.45 pF. This type of information may be obtained from the manufacturer's data sheets for a given type of RF transistor, but often it must be obtained by direct measurement.

Since the overall bandpass is to be 240 kHz, the individual tuned-transformer bandpass may be determined from Eq. (7–20). Application of this formula for three identical stages, and thus four tuned transformers, results in an individual tuned-circuit bandpass of 365 kHz. From this information and the center frequency of 10.7 MHz the necessary loaded tuned-circuit $Q$ can be found. Recall that the bandpass of a double-tuned critically coupled transformer is 1.414 times greater than the bandpass of a single-tuned circuit of high $Q_L(> 10)$. Therefore, the $Q_L$ of the required tuned transformer will be greater than that of a single-tuned circuit by a factor of 1.414. The loaded $Q$ of the tuned transformer will be

$$Q_L = \frac{1.414 f_r}{B} = \frac{1.414 \times 10.7 \times 10^6}{3.65 \times 10^5} = 41.4. \tag{7–21}$$

Thus, the primary and the secondary of the tuned transformer will have a loaded $Q$ of 41.4. Knowledge of the transformer $Q$ allows us to calculate the coefficient of coupling. The coefficient of coupling for a double-tuned critically coupled amplifier in terms of $Q$ may be determined from the fact that critical coupling results in maximum power transfer. For maximum power transfer to occur, the reflected resistance ($R_r$) from the secondary into the primary must equal the resistance associated with the primary. Thus $R_r = R_{s_1}$. However $R_{s_1}$ is also equal to the secondary resistance ($R_{s_2}$), since it is assumed that primary and secondary have equal $Q$'s.

(1) $R_r = R_{s_1}$,

(2) $\dfrac{(\omega M)^2}{R_{s_2}} = R_{s_1}$, but $R_{s_2}$ is equal to $R_{s_1}$

(3) $(\omega M)^2 = R^2$,

(4) $\omega M = R$, but $\omega M = 2\pi f k \sqrt{L_1 L_2}$ and $L_1 = L_2$,

(5) $\omega k L = R$,

(6) $k = \dfrac{R}{\omega L}$,

(7) $k = \dfrac{1}{Q_L}$. (7–22)

For the conditions determined in Eq. (7–21), $K$ is equal to 1/41.4 or 0.024.

As was shown in Fig. 7-14, the transformer of each tuned circuit is tapped as a means of providing the proper loaded $Q$'s and thereby providing a power match. To determine the tapping point on the secondary we must first determine the input impedance of the next stage transistor.

As can be seen from the hybrid-$\pi$ equivalent circuit of Fig. 7-15, the input of this transistor consists of $r_{bb'}$ in series with the parallel combination of $C_{b'e}$ and $R_{b'e}$. To determine the shunt impedance seen by an input generator we must first convert the parallel combination of $C_{b'e}$ and $R_{b'e}$ into an equivalent series circuit. This as well as the subsequent development is shown in Fig. 7-15. Next, the series equivalent impedance is added to $r_{bb'}$ to obtain the total series impedance, which is 1120 ohms $\angle -37°$. Finally the circuit is converted into equivalent shunt components, consisting of a 1400-ohm resistive component and an 1860-ohm capacitive component in shunt with one another. If it is assumed that the tuned circuits are at resonance, the capacitive reactive component will be canceled and only the resistive component ($R_{in}$) need be dealt with.

The next step in determining the proper tapping point of the secondary winding is to determine the load to which $R_{in}$ must match. See Fig. 7-16 for the actual and equivalent circuits of the transformer. The unloaded $Q$ of both the primary and secondary tuned circuits will be assumed to be 62.5. If the inductances of both primary and secondary are assumed to be 2.43 $\mu$hy, the reactance of each coil at 10.7 MHz will be 163.5 ohms. Therefore, the equivalent shunt resistance, which determines the unloaded $Q$ of the coils, is

$$R_s = Q_{uL}X_L = 62.5 \times 163.5 = 10{,}200 \text{ ohms.}$$

The loaded $Q$ of the secondary is 41.4, which can be represented by a 6780-ohm shunt resistance ($Q_L = R_l/X_L$) across the 163.5-ohm inductive reactance of the secondary. The loaded coil is in effect shunted by the shunt equivalent resistance of the coil itself (10,200 ohms) and the resistance due to the shunt load component of the input impedance of the transistor. By transformer action this component appears across the entire secondary winding. If the loaded $Q$ of 41.4 is to be obtained, the parallel combination of all resistive components shunting the secondary must result in a total resistance of 6780 ohms ($R_l = Q_L X_L$).

On the basis of the above, it can be seen that the 10,200-ohm shunt resistance, due to the coil, must be in shunt with a 20,000-ohm resistance to provide a total resistance of 6780 ohms ($1/R_l = 1/10.2\text{ k} + 1/20\text{ k}$). This 20,000-ohm shunt resistance across the secondary consists of two components—the first due to the transformed input resistance of the second stage transistor and second due to the reflected shunt resistance of the primary load. Since the circuit is arranged to have critical coupling and since a power match is desired ($R_o = R_r'$), the reflected shunt impedance from the primary into the secondary will equal the output resistance of the first-stage transistor. The output resistance of the transistor is 40,000 ohms. In order for two shunt components to be equal and have a total resistance of 20,000 ohms, each component must equal 40,000 ohms. One of these components is due to the reflected imped-

ance of the first-stage transistor. The other component of 40,000 ohms is due to the transformed input resistance (1400 ohms) of the second-stage transistor.

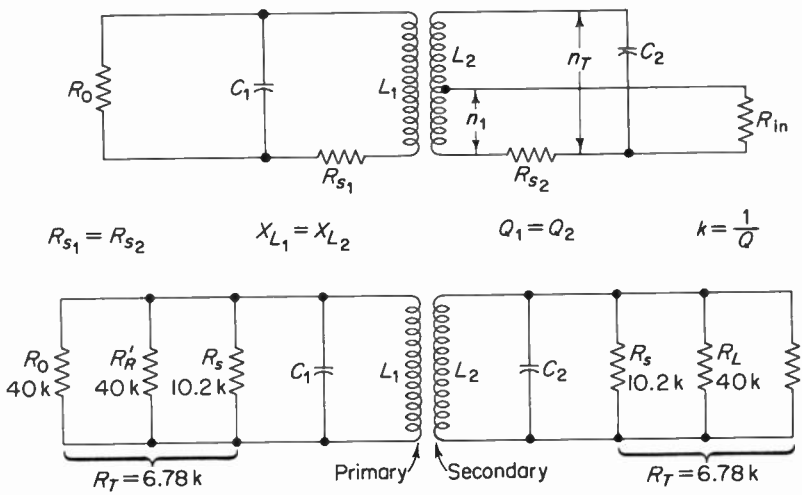
The secondary can act as a step-up transformer for the input resistance of the first-stage transistor by tapping the secondary coil as shown in Fig. 7-16. The turns ratio may be calculated to a good approximation by the turns-ratio formula:

$$N = \sqrt{\frac{R_L}{R_{in}}} = \sqrt{\frac{40\text{ k}}{1.4\text{ k}}} = 5.34 : 1. \tag{7-23}$$

If it is assumed that the 2.43-$\mu$hy coils that make up the transformer consist of 20 turns, the tapping point, $n_1$, may be found as follows:

$$n_1 = \frac{n_t}{n} = \frac{20}{5.34} = 3.75 \text{ turns}. \tag{7-24}$$

In this example the primary does not require a tap, since the reflected resistance from the secondary to the primary (40,000 ohms) is equal to the output resistance $r_{ce}$ (40,000 ohms). Thus, a power match is assured. Note




$R_O$ = the output resistance of the transistor

$R'_R$ = shunt equivalent of reflected resistance from the secondary

$R_s$ = shunt equivalent of the resistance associated with $L_1$ or $L_2$

$R_L$ = the transformed value of $R_{in}$ which appears across the entire secondary

$R_R$ = the reflected resistance in the sec. due to $R_O$ in the primary

**Fig. 7-16.** The diagram labeled (a) illustrates the actual tuned circuit and its associated loads. Diagram (b) illustrates the various shunt components of resistance which determine the loaded $Q$ of the transformer and thus also determine the bandpass of the transformer.

also that the total shunt load in the primary due to the parallel combination of $R_o$, $R_s$, and $R_L$ is equal to 6780 ohms; therefore the loaded $Q$ is 41.4 ($Q_L = R_l/X_L$). This is the same as the loaded $Q$ of the secondary.

If it were necessary to transform the output resistance of the first-stage transistor into a large value, as a means of matching a large reflected impedance from the secondary, the same tapping procedure that was used in the secondary could be followed.

Keep in mind that tapping the transformer also steps up the reactance of the output and input capacitance of the transistor. This in effect reduces the magnitude of these capacitances. Since the transformed capacitances are in effect in parallel with the resonating capacitors of the transformers, they are effectively swamped out. This simply means that the variable inductance of the resonant transformer is adjusted to a slightly smaller value than it would have been under no-load conditions.

### *Power gain*

The transducer gain of the amplifier is defined as $P_L/P_{in}$, where $P_{in}$ is the matched input power of the amplifier, and $P_L$ is defined as the power dissipated in the output load. The true input power is equal to

$$P_{in} = i_b V_{be} \cos\theta = \frac{i_b V_{be} R_{in}}{Z_{in}}. \tag{7-25}$$

The input resistance of the transistor is shown in Fig. 7-15 and consists of $r_{bb'}$ in series with the parallel combination of $R_{b'e}$ and $C_{b'e}$. This parallel combination, called $Z_1$, can be converted into a series equivalent by use of the transfer equations

$$\frac{Z_1^2}{R_{shunt}} = R_{series} \quad \text{and} \quad \frac{Z_1^2}{X_{shunt}} = X_{series}. \tag{7-26}$$

Thus $R_{in}$ is equal to

$$R_{in} = r_{bb'} + \frac{Z_1^2}{R_{b'e}}. \tag{7-27}$$

Since the capacitive reactance is tuned out by the inductance of the tuned circuit, we shall not consider it any further at this point. If matched conditions and a perfect transformer ($R_s = \infty$) are assumed, then $r_{ce}$ must equal $R_L$, and the AC current in the load ($i_L$) must be one-half of the current due to the current generator $g_m V_{b'e}$. This is shown in Fig. 7-15. Thus the power delivered to the output load is

$$P_L = \left(\frac{i_c}{2}\right)^2 R_L = \frac{(g_m V_{b'e})^2}{4} R_L.$$

Power gain $= P_L/P_{in}$, so

$$\frac{(g_m V_{b'e})^2}{4} R_L \times \frac{Z_{in}}{i_b V_{be} R_{in}}, \qquad \text{where} \quad Z_{in} = r_{bb'} + Z_1,$$

but from Fig. 5-15(a) and (e) it can be seen that

$$V_{b'e} = V_{be}\left(\frac{Z_1}{Z_{in}}\right).$$

Squaring both sides,

$$V_{b'e}^2 = V_{be}^2\left(\frac{Z_1}{Z_{in}}\right)^2,$$

$$PG = \frac{g_m^2 R_L Z_{in} V_{be}^2 \left(\frac{Z_1}{Z_{in}}\right)^2}{4 i_b R_{in} V_{be}}$$

$$= \frac{g_m^2 R_L V_{be} \frac{Z_1^2}{Z_{in}}}{4 i_b R_{in}}.$$

But $V_{be}/i_b = Z_{in}$, and so

$$PG = \frac{g_m^2 R_L Z_1^2}{4 R_{in}},$$

but since $R_{in} = r_{bb'} + Z_1^2/R_{b'e}$,

$$PG = \frac{g_m^2 R_L Z_1^2}{4 r_{bb'} + \frac{4 Z_1^2}{R_{b'e}}} = \frac{g_m^2 R_L R_{b'e}}{4\left(\frac{r_{bb'} R_{b'e}}{Z_1^2} + 1\right)}. \tag{7-28}$$

Thus if the $g_m$ of the transistor is 15,200 $\mu$mhos, and if $R_{b'e}$ and $C_{b'e}$ in parallel have an impedance magnitude of 1040 ohms, the power gain of this amplifier is

$$PG = \frac{(1.52 \times 10^4 \times 10^{-6})^2 \times 4 \times 10^4 \times 1.36 \times 10^3}{4\left[\frac{1 \times 10^2 \times 1.36 \times 10^3}{(1.04 \times 10^3)^2} + 1\right]} = 2780.$$

In terms of decibels this is equal to

$$PG = 10 \log 2780 = 34.45 \text{ db}.$$

### *Insertion loss*

Since the tuned transformer is not perfect, it introduces an insertion loss. This loss must be taken into consideration, since the expression for power gain was developed with the assumption of a lossless transformer. To find the actual power gain of the circuit we must determine the loss introduced by the transformer and subtract this from the power gain obtained by Eq. (7–28).

Insertion loss is defined as the actual power ($P_L$) delivered to the load divided by the power ($P_{L_i}$) delivered to the load by an ideal transformer. Using a circuit such as that of Fig. 7-16, where all circuit elements are made into shunt equivalents, we can obtain a useful expression for insertion loss. Let $R_0$ represent the interval resistance of the device (transistor); $R_L$ is the

load resistance, $R_s$ is the equivalent shunt resistance associated with the tuned-circuit inductance $L$, and $C$ is the resonating capacitance. Thus,

$$P_{L_i} = \frac{E^2}{R_L} = \frac{I_o^2 R^2}{R_L} \qquad \text{where } R \text{ is } \frac{R_0 R_L}{R_0 + R_L} \text{ and } R_s = \infty;$$

$$P_L = \frac{E^2}{R_L} = \frac{I_o^2 R_T^2}{R_L} \qquad \text{where } R_T = \frac{1}{\frac{1}{R_0} + \frac{1}{R_s} + \frac{1}{R_L}}.$$

Thus,

$$\frac{P_L}{P_{L_i}} = \frac{I_o^2 R^2}{R_L} \times \frac{R_L}{I_o^2 R_T^2} = \frac{R^2}{R_T^2} = \left(\frac{R}{R_T}\right)^2.$$

Now

$$Q_L = \frac{R_T}{X_L} = \frac{1}{\left(\frac{1}{R_0} + \frac{1}{R_s} + \frac{1}{R_L}\right) X_L} \quad \text{and} \quad X_L = \frac{R_s}{Q_{uL}},$$

$$\therefore \quad Q_L = \frac{1}{\left(\frac{1}{R_0} + \frac{1}{R_s} + \frac{1}{R_L}\right) \frac{R_s}{Q_{uL}}}$$

$$= \frac{Q_{uL}}{\left(\frac{1}{R_0} + \frac{1}{R_s} + \frac{1}{R_L}\right) R_s},$$

$$\frac{Q_{uL}}{R_s} = \frac{Q_L}{R_0} + \frac{Q_L}{R_s} + \frac{Q_L}{R_L},$$

$$\frac{Q_{uL}}{R_s} - \frac{Q_L}{R_s} = Q_L\left(\frac{1}{R_0} + \frac{1}{R_L}\right)$$

but $1/R_0 + 1/R_L = 1/R$, and so

$$\frac{Q_{uL} - Q_L}{R_s} = \frac{Q_L}{R},$$

$$R = \frac{Q_L R_s}{Q_{uL} - Q_L}.$$

But

$$\frac{P_L}{P_{L_i}} = \left(\frac{R}{R_T}\right)^2 = \left[\left(\frac{Q_L R_s}{Q_{uL} - Q_L}\right)\left(\frac{1}{R_0} + \frac{1}{R_L} + \frac{1}{R_s}\right)\right]^2$$

$$= \left[\left(\frac{Q_L R_s}{Q_{uL} - Q_L}\right)\left(\frac{Q_{uL} - Q_L}{Q_L R_s} + \frac{1}{R_s}\right)\right]^2,$$

$$\frac{P_L}{P_{L_i}} = \left(\frac{Q_L R_s}{Q_{uL} - Q_L} \times \frac{Q_{uL} - Q_L + Q_L}{Q_L R_s}\right)^2 = \left(\frac{Q_{uL}}{Q_{uL} - Q_L}\right)^2. \qquad (7\text{–}29)$$

In terms of decibels the insertion loss due to an imperfect coupling transformer is

$$\text{loss in db} = 20 \log \left[\frac{Q_{uL}}{Q_{uL} - Q_L}\right]. \qquad (7\text{–}30)$$

For this example, the insertion loss becomes

$$20 \log \frac{62.5}{62.5 - 41.4} = 9.45 \text{ db.}$$

The total power gain of a single stage is the total power gain less the insertion loss.

$$\text{Power gain} = 34.45 - 9.45 = 25 \text{ db.}$$

The overall power gain of three identical stages will then be 75 db.

### *Maximum stable gain*

To determine whether the amplifier requires neutralization we must determine whether the overall power gain exceeds the maximum stable gain of the amplifier. As pointed out in Chapter 5, the maximum stable gain can be determined from

$$MSG = \frac{g_m}{1.25\omega C_{ob}}. \tag{5-23}$$

In terms of this example,

$$MSG = \frac{1.52 \times 10^{-3}}{1.25 \times 6.28 \times 1.07 \times 10^{-7} \times 0.45 \times 10^{-12}} = 400,$$

$$MSG = 10 \log 400$$

$$= 26.03 \text{ db.}$$

Since the stage gain is 25 db, this amplifier should be stable for the conditions indicated and no neutralization will be necessary. Of course, if the amplifier is unstable and it is desired to avoid neutralization, a mismatch between $R_L$ and $R_0$ can be produced that will reduce the gain below the maximum-stable-gain level.

### *Harmonic distortion*

By using the Ross criterion we can determine whether or not the amplifier will depart sufficiently from phase linearity to introduce significant harmonic distortion. The Ross criterion for a critically coupled double-tuned transformer is that the ratio

$$\frac{2(\text{sideband pairs})(\text{maximum audio frequency})Q_L}{\text{frequency of operation}}$$

must numerically be less than 2 for minimum harmonic distortion. In this example, assuming an IF of 10.7 MHz, and a modulation index of 5, the number of sideband pairs will be 8 assuming a modulation frequency of 15 kHz. Thus,

$$\frac{2(8)(1.5 \times 10^4)(4.14 \times 10^1)}{1.07 \times 10^7} = 0.0928.$$

From this last example it is clear that operating the tuned circuits at a high IF will insure low harmonic distortion.

*Amplitude modulation*

Using the curves of Fig. 7-11, we may determine the percentage of amplitude modulation introduced by the four tuned circuits of the amplifier. The product of $Q_L\, 2\Delta f/f_{op}$ is found to be

$$\frac{41.4 \times 2 \times 75 \times 10^3}{10.7 \times 10^6} = 0.058.$$

From the curves it is found that for four tuned-circuit transformers the percentage of amplitude modulation is less than 1 per cent. This again points up the need for a high IF if amplitude modulation is to be kept low.

## 7-7. AM-FM IF amplifiers

Many different types of commercial FM receivers are available. The block diagrams of the most common types are summarized in Fig. 7-17. Figure 7-17(a) illustrates the block diagram of that type of receiver classified as a tuner. Tuners are receivers that require an outboard audio-amplifier section separate from the RF-to-detector portions of the receiver. The most common

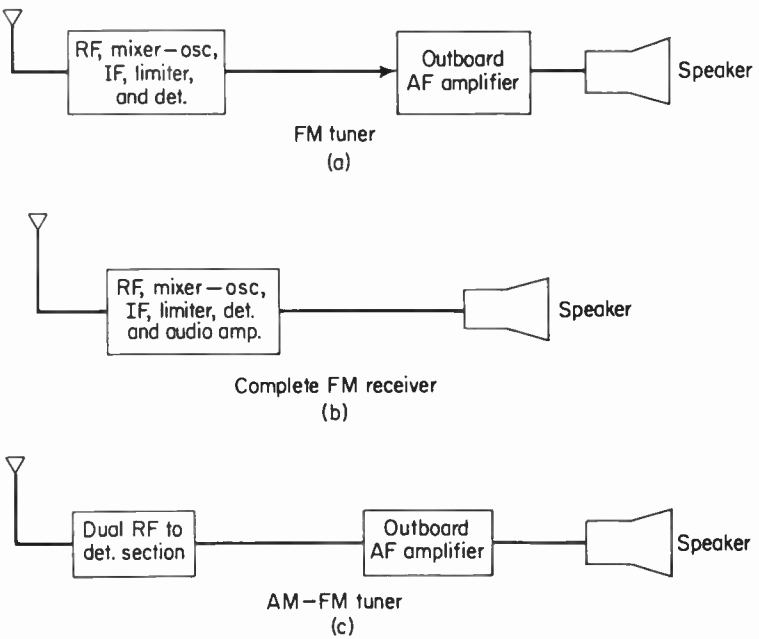



**Fig. 7-17.** A comparison of the three types of FM receivers.

FM receiver is the complete receiver, which contains both the RF and the audio sections. The block diagram of this type of receiver is shown in Fig. 7-17(b). As shown in Fig. 7-17(c) some tuners and complete receivers are designed to receive both FM and AM broadcast bands; these receivers may be classified into two types. As illustrated in Fig. 7-18, one type requires completely separate RF, mixer-oscillator, IF, and detector sections for both FM and AM. This type employs a common audio amplifier. The other type, which is much more common, uses separate RF, mixer-oscillator, and detector sections but employs a composite IF amplifier and common audio amplifiers.

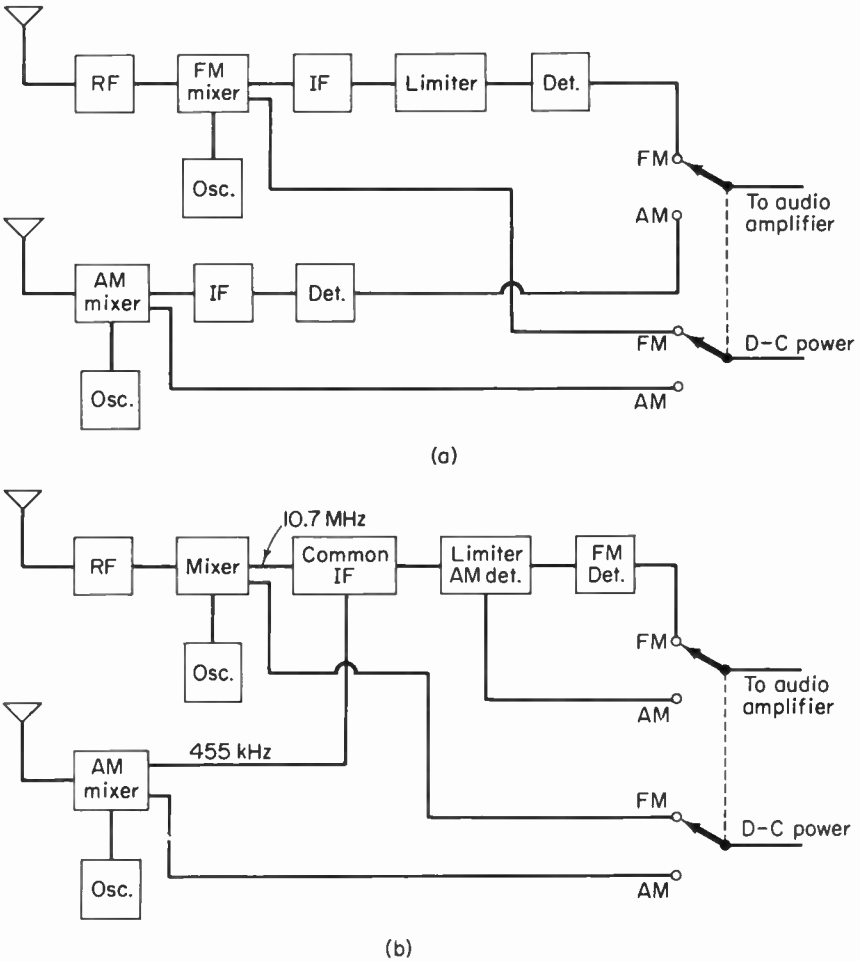



**Fig. 7-18.** Two types of AM-FM receivers. In fig. (a) both AM and FM portions of the receiver are complete. In fig. (b) the IF is common to both AM and FM, also the FM limiter is the AM detector when the receiver is in the AM mode of operation.

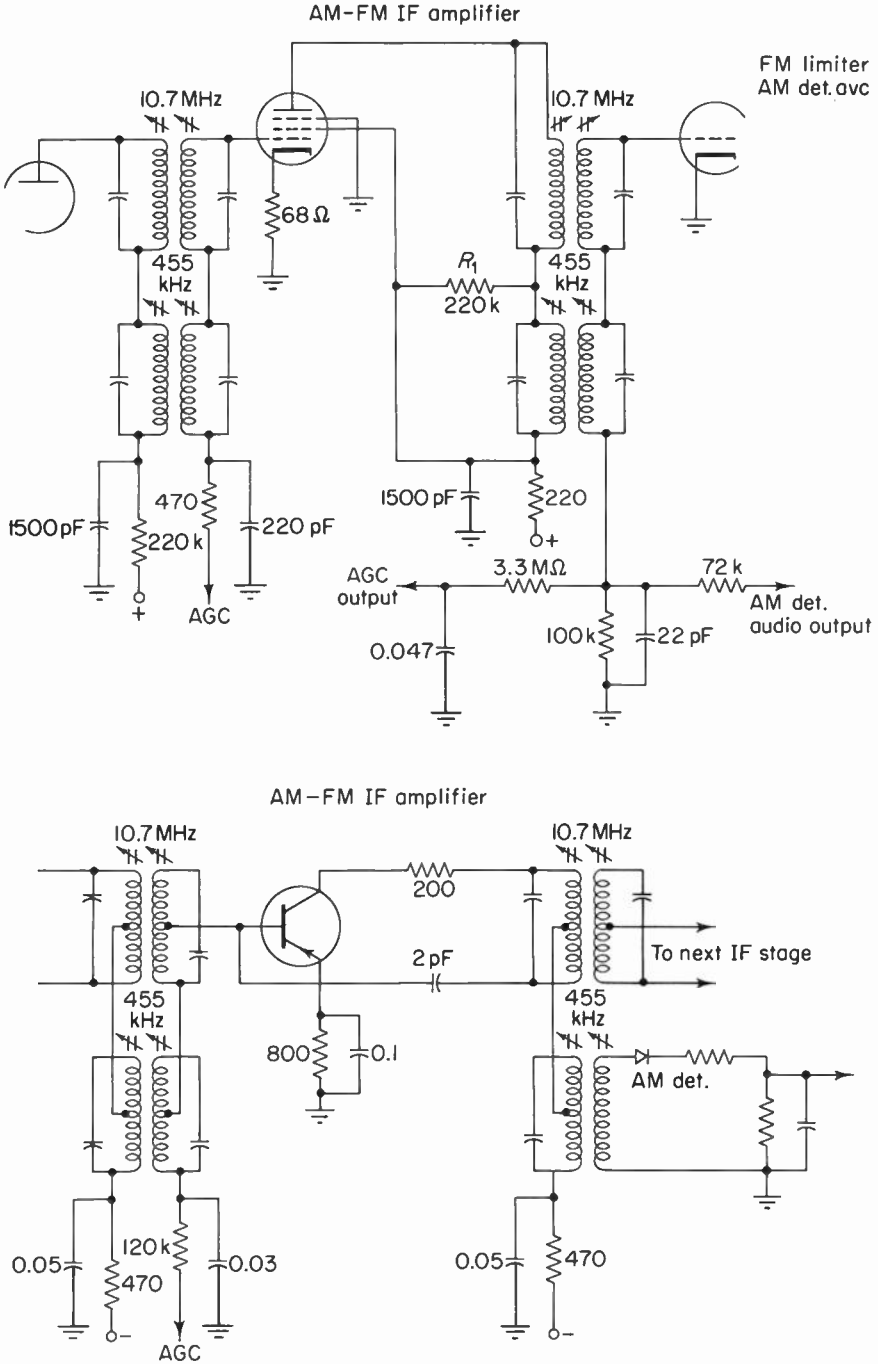



**Fig. 7-19.** Typical vacuum tube and transistor dual purpose IF amplifiers.

Selection of the desired mode of reception is accomplished by switching circuits located in the DC portions of both the FM and the AM mixers. This switch activates the desired mixer and allows an IF to be generated. For AM reception the IF is usually 455 kHz; for FM reception the FM mixer-oscillator is activated and generates an IF of 10.7 MHz when a signal is being received. The input to the audio section is also ganged to the AM-FM selector switch to provide the activated channel audio amplification.

The schematic diagram of a transistor and a vacuum-tube dual IF amplifier is shown in Fig. 7-19. As we can see by comparing Figs. 7-1, 7-14, and 7-19, the single-purpose amplifiers and the dual-purpose amplifiers are fundamentally the same. The two types of circuits have similar DC characteristics but differ in the number of coupling circuits in the input and output sections of the amplifiers. In the dual IF amplifier the input and output coupling transformers are connected in series with one another. One of them is tuned to 10.7 MHz, the FM IF, and the other is tuned to 455 kHz, the AM IF. Since the resonant frequencies of the transformers are very far apart (20:1), the tuned circuits have very little effect upon one another. Thus, when the IF is 10.7 MHz, the tuning capacitor of the 455-kHz tuned transformer acts as a low-impedance return to AC ground, since its reactance is very small at this frequency. Likewise, when the frequency of the IF is 455 kHz, the reactance of the inductance associated with the 10.7-MHz IF transformer is very small and the transformer is very far removed from its resonant frequency. Therefore, the output of the transformer is nil.

The 455-kHz amplifier tends to have greater gain than the 10.7-MHz amplifier because of its greater $L/C$ ratio. Also to be considered is the added inductance due to the 10.7-MHz transformer, which tends to make the circuit potentially unstable. Thus, the gain of the stage, for 455 kHz, is kept somewhat lower than might be otherwise expected. The 10.7-MHz IF usually requires two to four stages of amplification for proper FM reception. The 455-kHz IF seldom requires more than two stages of amplification before it is detected.

A frequent practice in the case of vacuum-tube arrangements is to use the grid-leak-bias section of the limiter as the AM detector when the receiver is operating in the AM mode. In transistor receivers the AM detector function is derived from a separate diode detector circuit. Typical circuits are shown in Fig. 7-19.

## 7-8. The integrated-circuit IF amplifier

The monolithic (one stone) integrated circuit has brought a revolution into the field of electronics. One of the first consumer products to show the effects of this revolution has been the FM receiver. In particular, the IF amplifier, limiter, and detector circuits have been integrated into a single

monolithic silicon chip. Since these devices can operate at rather high frequencies (> 300 MHz), entire FM tuners can be integrated on one or more silicon integrated-circuit chips.

A monolithic integrated-circuit is a small piece of doped crystalline silicon, upon which have been selectively deposited, by oxidation, diffusion, and evaporation, layers of silicon dioxide, impurity elements, and metal films, so arranged and modified by acid etching as to form interconnecting active and passive circuit elements, which in turn form complete circuits. A cross-sectional view of an integrated circuit is shown in Fig. 7-20. Passive-circuit elements are those elements which either store or dissipate signal energy. Active devices can be used to control signal energy and can perform the functions of amplification, oscillation, and the like. Integrated circuits can have incorporated into their structure both passive and active circuit elements such as resistors, capacitors, transistors and FET's.

An important difference between integrated-circuit design and lumped-circuit design is that in integrated circuits the cheapest, most easily made components are active circuit elements. Thus they are used to replace passive elements wherever possible. In contrast, lumped-circuit design is based upon the premise that the active circuit elements are the most expensive circuit elements and their use is to be avoided wherever possible. As a result, the integrated circuit does not duplicate the lumped circuit for any given design, and in fact the two will probably be completely different for the same circuit function.

The remarks above do not mean that integrated-circuit resistors and

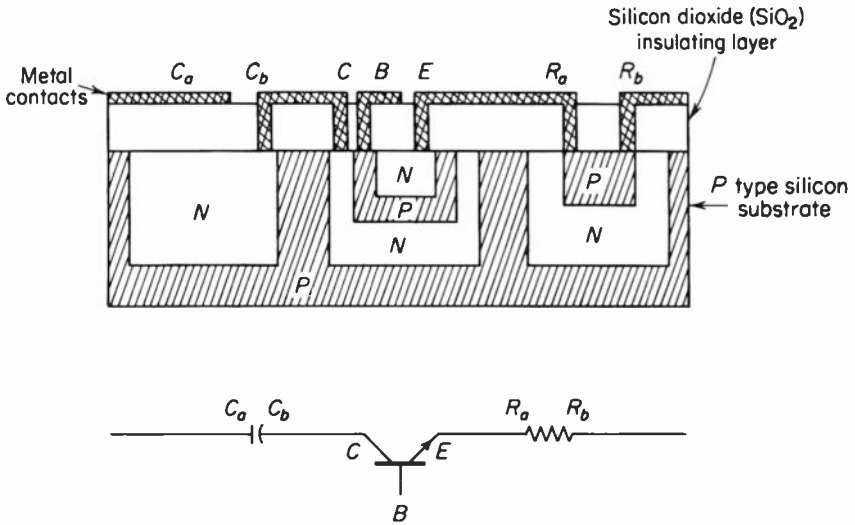



**Fig. 7-20.** A cross-sectional view of an integrated circuit consisting of a capacitor, an *NPN* transistor, and a resistor connected in series.

capacitors are seldom fabricated by integrated-circuit techniques. They are, however one 10-pF capacitor occupies the same space as three or four transistors, and one resistor is equivalent in area to two transistors. Thus transistors are the most economical elements in integrated technology.

In integrated circuits that employ capacitors, the capacitors are usually less than 50 pF. Typical resistors can be made up to approximately 30,000 ohms, although absolute values are difficult to reproduce and have a tendency to vary. Thus, designs are sought that are governed by resistor ratios rather than particular resistor values.

Except for the single-frequency application, where a given inductive reactance may be formed by active circuits, inductors are the one circuit element that cannot be formed as an integral part of the integrated circuit, and must be used as lumped components outboard to the integrated chip. Thus, almost all integrated-circuit designs used in tuned-amplifier applications employ external *LC* tuned circuits to obtain the receiver's selectivity and bandpass. However, it is not entirely necessary for tuned circuits to be used at all. A number of designs have appeared that attempt to eliminate the inductive element required for tuning. In one design the RF signal falling in the range of 88 to 108 MHz is converted by ordinary mixers to a lower IF centered at about 120 kHz. All of the significant sidebands associated with the incoming signal will appear in a frequency range falling between 0 and 240 kHz, assuming a desired bandpass of 240 kHz. Since integrated circuits can be designed that can easily pass this range of signal frequencies, it is possible to incorporate into the design *RC* filters that provide rapid attenuation at the high-frequency end of the bandpass. Frequencies below 0 Hz do not exist, so a high degree of selectivity and adjacent-channel attenuation is possible. The use of inductive elements for tuning may also be avoided by the use of crystal filters, which do not require alignment and which can provide skirt response almost equal to the ideal case of Fig. 7-5.

### *Differential amplifiers*

The schematic diagrams of commercial linear integrated circuits available for IF amplification indicate a number of different approaches to linear amplification. Some designs use common-emitter amplifiers which are direct-coupled and employ negative feedback. Other common designs have used emitter-coupled amplifiers as differential amplifiers. The schematic diagram of each basic circuit is shown in Fig. 7-21.

It has already been pointed out that integrated capacitors require considerably more surface area than do resistors or transistors on a given silicon chip. Therefore, as a means of minimizing the use of integrated-circuit capacitors, direct-coupled amplifier circuits called differential amplifiers have been widely used. In the past, cascaded direct-coupled amplifiers have been avoided because any small change in operating conditions in the low-level

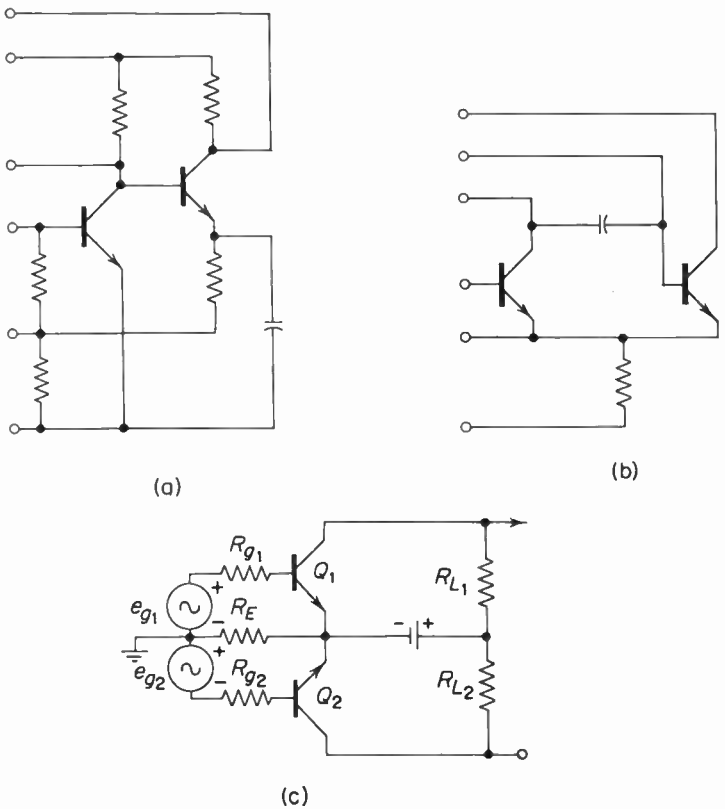



**Fig. 7-21.** Fig. (a) illustrates the circuit found on an integrated circuit chip, and consists of a direct coupled two-stage amplifier. Circuit (b) is a differential amplifier suited for IF amplification. Circuit (c) is drawn to show that the differential amplifier is similar in circuit arrangement to a push-pull amplifier.

stages would be amplified and cause major changes in operating conditions in the high-level stages. DC operating conditions such as bias are affected by temperature and power-supply variations. The differential amplifier is a DC amplifier designed to minimize these problems and at the same time provide substantial gain for low-level signals.

The small-signal voltage amplification of a basic differential amplifier, such as the one in Fig. 7-21c, operated in the differential mode (the condition where signals in phase opposition are applied to both inputs and signals are obtained from both outputs) is approximately

$$A_v = \frac{2R_L}{R_E} \qquad \left(\text{assuming } \frac{R_g}{h_{FE}} \ll R_E\right). \tag{7-31}$$

If either one of the inputs or outputs is returned to AC ground making the device a single-ended amplifier, the voltage amplification becomes approx-

imately equal to $R_L/R_E$. Thus, the voltage amplification depends upon a resistive ratio and not upon the absolute values of the resistors. Owing to the close proximity of all elements in an integrated circuit, temperature changes are the same throughout the circuit. Therefore, integrated differential amplifiers are relatively insensitive to temperature variations since the $R_L/R_E$ ratio will remain constant.

The input resistance for the differential input is

$$R_{\text{in}} \approx (h_{FE} + 1)R_E. \tag{7-32}$$

This formula is similar to the equation for input resistance of a common-collector amplifier. The input resistance of a single-ended input differential amplifier is

$$R_{\text{in}} = (h_{FE_1} + 1)2R_E + \frac{R_g}{h_{FE_2} + 1}. \tag{7-33}$$

*Darlington amplifier*

The last two equations indicate that the input resistance of the amplifier may be increased by increasing $h_{FE}$ of the transistor. One way of increasing $h_{FE}$ is to use a circuit arranged in the Darlington configuration. See Fig. 7-22. Notice that in this circuit two DC coupled transistors are arranged so that the output of the first appears between the base and the collector of the second stage. The current gain of an equivalent transistor that could replace this arrangement is found as follows. Referring to Fig. 7-22,

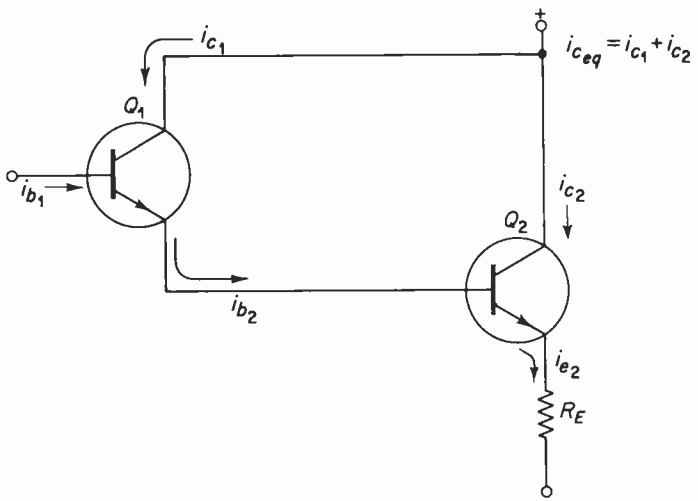



**Fig. 7-22.** The Darlington circuit which is useful in producing high input resistance amplifiers. This arrangement may also be used with differential amplifiers. Also shown is the direction of current (hole) flow.

$$
\begin{aligned}
h_{FE_{eq}} &= \frac{i_{c_{eq}}}{i_{b_1}} = \frac{i_{c_1} + i_{c_2}}{i_{b_1}}, \\
i_{c_2} &= h_{FE_2} i_{b_2}, \\
i_{b_2} &= i_{e_1}, \\
i_{b_2} &= i_{c_1} + i_{b_1}, \\
i_{c_1} &= h_{FE_1} i_{b_1}, \\
i_{b_2} &= h_{FE_1} i_{b_1} + i_{b_1}, \\
i_{b_2} &= (h_{FE_1} + 1) i_{b_1}, \\
i_{c_2} &= h_{FE_2}[i_{b_1}(h_{FE_1} + 1)], \\
i_{c_{eq}} &= i_{c_1} + i_{c_2}, \\
i_{c_{eq}} &= h_{FE_1} i_{b_1} + i_{b_1} h_{FE_2}(h_{FE_1} + 1), \\
h_{FE_{eq}} &= \frac{i_{c_{eq}}}{i_{b_1}} = h_{FE_1} + h_{FE_2}(h_{FE_1} + 1).
\end{aligned}
\tag{7-34}
$$

If

$$
\begin{aligned}
h_{FE_1} &= h_{FE_2}, \\
h_{FE_{eq}} &= (h_{FE})^2 + 2h_{FE}, \\
h_{FE_{eq}} &\approx (h_{FE} + 1)^2.
\end{aligned}
\tag{7-35}
$$

Assuming that this equivalent transistor is used in place of the Darlington configuration, the input resistance of the differential amplifier will be

$$R_{in} = (h_{FE_{eq}} + 1)R_E. \tag{7-36}$$

For example, if two identical transistors have a beta of 60 and are used in the Darlington configuration, where the emitter resistor $R_E$ is 1000 ohms, the input resistance of the amplifier will be

$$[(61)^2 + 1] \times 1 \times 10^3 = 3.72 \text{ megohms}.$$

### *Commercial circuits*

A schematic diagram of an integrated circuit used as an IF amplifier and limiter is shown in Fig. 7-23(a) and (b). Also shown in this figure is the schematic diagram of the internal circuit of the integrated circuit. The IF amplifier limiter circuit of Fig. 7-23(a) is simplicity itself. Outboard to the integrated circuit are a number of decoupling filters leading to power sources used to establish the operating conditions of the device, and the tuned circuits necessary to provide the required selectivity. The schematic diagram of Fig. 7-23(b) shows that the integrated circuit consists of eleven resistors, a two-stage differential amplifier ($Q_1$, $Q_2$, $Q_3$, $Q_4$), and one limiter stage ($Q_5$ and $Q_6$). Located between each of the differential amplifiers is an emit-

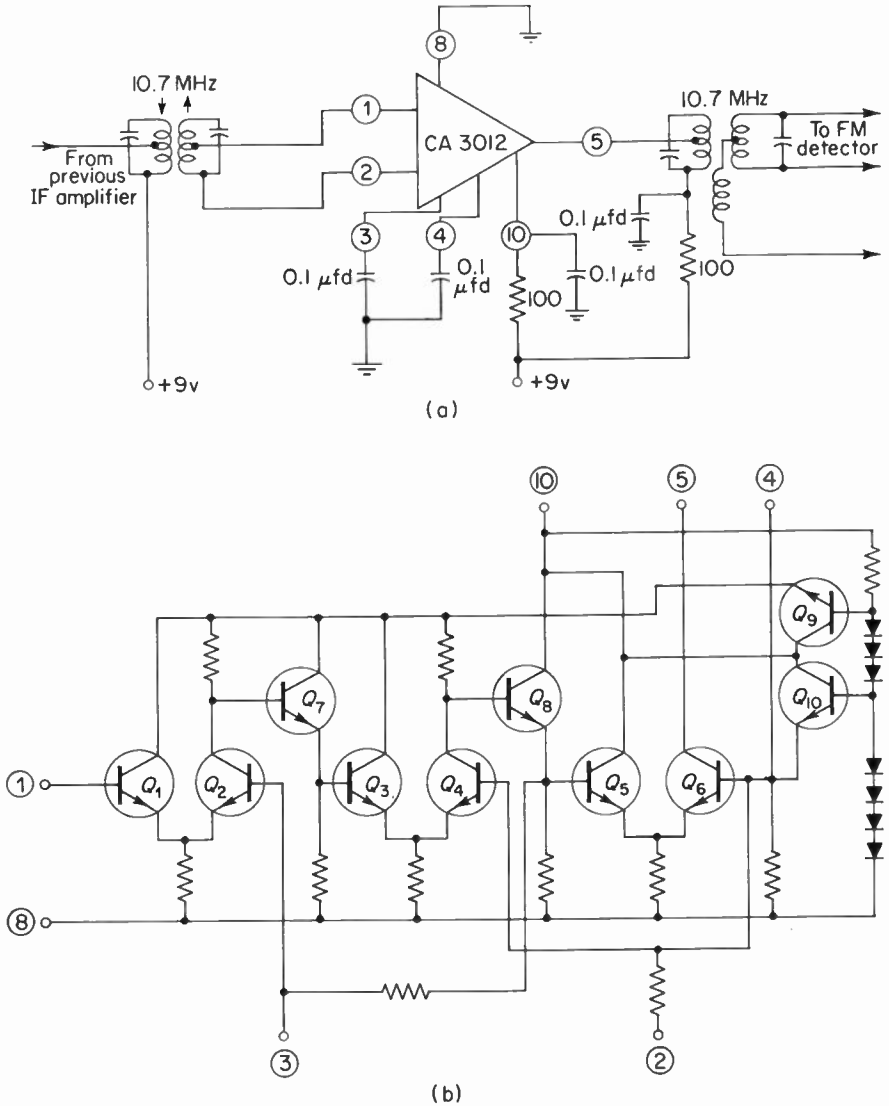



**Fig. 7-23.** An integrated circuit IF amplifier and limiter. Diagram (a) shows the integrated circuit and its external connection. Diagram (b) is the internal circuit of the integrated circuit chip. (Courtesy RCA)

ter follower ($Q_7$ and $Q_8$), which provides matching and allows direct coupling to be used between stages. Also part of the integrated circuit is a regulated power supply consisting of seven diodes and two transistors ($Q_9$ and $Q_{10}$). This regulated power supply is used to keep the gain of the differential amplifiers relatively constant under conditions of external power-supply voltage

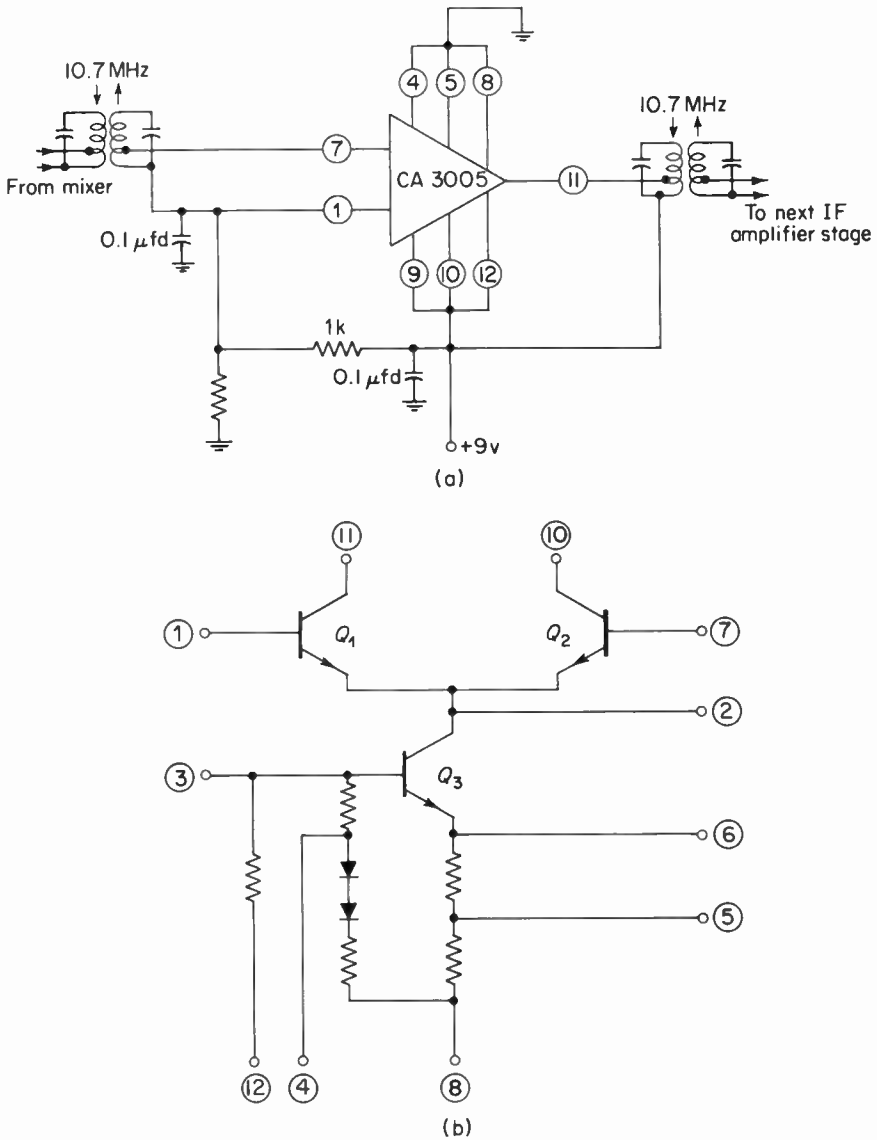



**Fig. 7-24.** (a) The integrated amplifier is used as a differential amplifier. This diagram shows the external circuit connections. (b) Shows the internal circuit of the integrated circuit. In the differential mode $Q_3$ supplies a constant current to transistors $Q_1$ and $Q_2$. $Q_1$ and $Q_3$ may be connected in cascode by joining terminals 7 and 10.

changes. The diodes are used as a means of providing temperature stabilization.

Another somewhat simpler circuit used for IF amplification is shown in Fig. 7-24(a) and (b). Part (a) shows the schematic diagram of the integrated circuit and its associated external circuit. Part (b) shows the schematic representation of the integrated circuit itself. In this circuit there are three transistors, two diodes, and five resistors. The three transistors are arranged so that by proper selection of the integrated-circuit output terminals the circuit may be used either as a differential amplifier and a current sink, or as a cascode amplifier. The cascode arrangement is usually used in RF and mixer applications.

In the IF amplifier of Fig. 7-24(a) the integrated circuit is arranged so that $Q_1$ and $Q_2$ are used in the differential mode. Under strong-signal conditions the differential amplifier will become overdriven and act as an excellent symmetrical limiter. Thus, the IF amplifier may also be used as a self-limiting amplifier. Notice also that the input and output circuits of a given stage are effectively separated from each other by the use of two transistors. Thus feedback and the possibility of oscillation are reduced to the point where neutralization is not necessary even at fairly high gains (28 db).

The constant-current-sink transistor ($Q_3$) is used to supply the emitter current for the differential amplifier transistors ($Q_1$ and $Q_2$). When used in this capacity, the current sink acts as a temperature-stabilizing circuit.

## 7-9. Summary

**1.** The function of the IF amplifier is to amplify signals obtained from the mixer that are lower in frequency than the original input signals.

**2.** Amplification at low frequency results in two advantages: (1) better selectivity, (2) improved amplifier stability.

**3.** Intermediate-frequency amplifiers are fixed-frequency amplifiers, and thus gain is constant with input frequency; also, they do not require elaborate variable-tuning arrangements.

**4.** The choice of intermediate frequency depends upon bandwidth, stage gain, required selectivity, and the desired ability to reject spurious responses.

**5.** The gain requirements of a receiver depend upon expected input-signal levels. Typical values of voltage amplification for vacuum-tube IF amplifiers are 200,000. For transistors a power gain of 73 db is necessary.

**6.** Three methods of interstage coupling are used in commercial receivers: (1) single-tuned, (2) critically coupled double-tuned transformers, (3) overcoupled tuned transformers.

**7.** From the steepness ratio it can be shown that the double-tuned

overcoupled amplifier comes closer to the ideal response than any of the methods of coupling listed above. The single-tuned amplifier would have the poorest steepness ratio.

**8.** In AM RF amplifiers distortion is due to nonlinear operation of the active elements of the circuit. In FM RF amplifiers distortion is primarily due to nonlinear phase relationships in the tuned circuits.

**9.** The frequency-discrimination characteristics of the tuned circuits can produce amplitude changes that introduce amplitude modulation into the frequency-modulation signal.

**10.** Vacuum-tube IF amplifiers generally use the grounded-cathode circuit configuration and are similar to AM-receiver IF amplifiers.

**11.** The voltage amplification of a vacuum-tube amplifier used as a double-tuned critically coupled amplifier is one-half that of the same tube used as a comparable single-tuned amplifier.

**12.** The bandwidth of a commercial receiver IF amplifier should have an overall bandwidth of 240 kHz for optimum selectivity, distortion, and AM-rejection characteristics.

**13.** Transistor IF amplifiers may be designed by means of the hybrid-$\pi$ equivalent circuit. From this equivalent circuit power gain, bandpass, matching requirements, and stability may be determined.

**14.** AM-FM receivers may be found in two basic forms: (1) completely separate AM and FM sections, (2) receivers that have common IF amplifiers.

**15.** Integrated circuits are used in FM receivers to perform the functions of RF amplifier, mixer-oscillator, IF amplifier, and detector.

**16.** Integrated circuits employ direct-coupled amplifiers called differential amplifiers.

## REFERENCES

1. Deutsch, Sid: *Theory and Design of Television Receivers.* McGraw-Hill Book Company, New York, 1951.
2. Jaffe, D. L.: "A Theoretical and Experimental Investigation of Tuned-Circuit Distortion in Frequency-Modulation Systems," *Proc. IRE*, May 1945.
3. Ross, H. A.: "The Theory and Design of Intermediate-Frequency Transformers for Frequency Modulated Signals," *AWA Technical Review*, **6:** 8 (March 1946).
4. Terman, F. E.: *Electronic and Radio Engineering*, 4th ed. McGraw-Hill Book Company, New York, 1955.
5. Seely, S.: *Electron Tube Circuits.* McGraw-Hill Book Company, New York, 1958.
6. Mergner, F. L.: "A Survey of Performance Requirements and Design Tech-

niques for Highest Quality FM Multiplex Reception," *J. Audio Engineering Society*, January 1965.

7. von Recklinghausen, D. R.: "FM Tuner Characteristics and Their Relative Importance," *Audio*, August 1963.
8. von Recklinghausen, D. R.: "Silicon Transistor I.F. Amplifier for FM Tuner," *Electronics World*, October 1965.
9. Walker, D.: "Solving Hi-Fi FM Tuner Design Problems," *Electronics World*, May 1965.
10. Redmond, K.: "Rapid Design Procedure for a Transistorized Double Tuned Bandpass IF Amplifier," *IEEE Trans. on Broadcast and Television Receivers*, November 1963.
11. Hittinger, W. C., and Sparks, M.: "Micro Electronics," *Scientific American*, November 1965.
12. Sanquini, R. L.: "Integrated Circuits Make a Low-Cost FM Receiver," Electronics, August 8, 1966.
13. Lancaster, D.: "Linear Integrated Circuits: What's Available?," *Electronics World*, November 1966.

# 8

# LIMITERS

## 8-1. Basic considerations and requirements

The basic requirement for any form of limiting is to remove all forms of amplitude modulation from the received signal and provide an output that is independent of both input amplitude *and* rate of change of input amplitude. This is a rather difficult requirement, and not all limiters can perform equally well under all conditions of operation.

A limiter stage (if one is provided) always follows the last IF amplifier and therefore precedes the detector stage. Sometimes one limiter stage can prove adequate if the received signal is strong, but two limiters in cascade is the preferred arrangement. It will be shown that two limiters provide more than twice the improvement over one to the extent that noise rejection (or, more specifically, the limiting sensitivity) is improved by a factor of about five.

The types of limiters to be discussed in this chapter are the overdriven (grid-leak and transistor) limiter and the dynamic limiter. A third type, the 6BN6 gated-beam tube, is fully discussed in Chapter 11.

## 8-2. The ideal limiter

The ideal limiter should take an input voltage that is varying in both frequency and amplitude and provide an output of constant amplitude that retains the frequency deviations of the transmitted carrier. This is shown in Fig. 8-1. The variations of the input signal amplitude can arise from

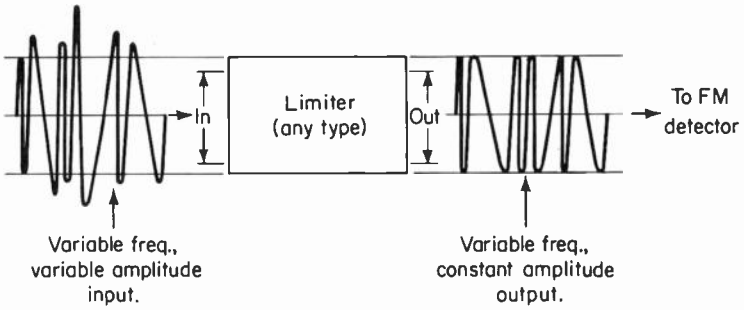



**Fig. 8-1.** What the ideal limiter should do in an FM receiver.

a number of causes, such as noise (random and impulse types), an interfering carrier, fading, IF selectivity, and so on.

## 8-3. The grid-leak limiter

The grid-leak limiter is widely used and can provide adequate AM rejection under most conditions. It does, however, have certain shortcomings. Before we undertake its analysis, we shall review certain aspects of vacuum-tube characteristics, since these will prove to be very pertinent.

Consider first a pentode of the sharp-cutoff variety and its associated control-grid-to-plate transfer characteristic. The characteristic is shown in Fig. 8-2 for a number of different values of screen voltage. The curves show that for small values of screen voltage, the grid voltage required for plate-current cutoff is small. The grid base (the amount of grid voltage required for plate-current cutoff) can thus be varied to suit the needs of a particular

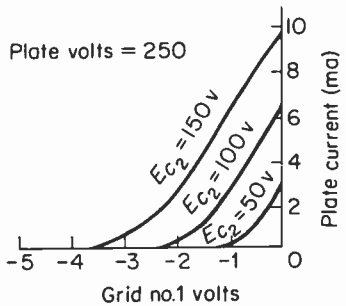



**Fig. 8-2.** Transfer curves for a sharp cut-off pentode for three values of screen voltage. (Static curves)

application. Notice also that for small values of screen voltage, the quiescent (no-signal) plate current is low, which makes for higher efficiency since the plate dissipation will be lower. The screen grid of a pentode thus acts like the plate of a conventional triode with respect to its effect on the shape of the transfer curve.

Consider this same pentode operated with a relatively normal value of screen voltage, but this time with the plate voltage maintained constant at the assumed value. Notice that the grid base is essentially unaffected by changes in plate voltage, but that the extent of plate-current saturation is very definitely changed. In this respect the plate voltage should be more correctly referred to as the "plate supply voltage," since we are not referring to the plate-to-cathode voltage. Thus, for low values of plate supply voltage, the current can reach saturation with a smaller grid voltage than otherwise. A graph of plate current versus grid voltage for different values of plate supply voltage is shown in Fig. 8-3(a).

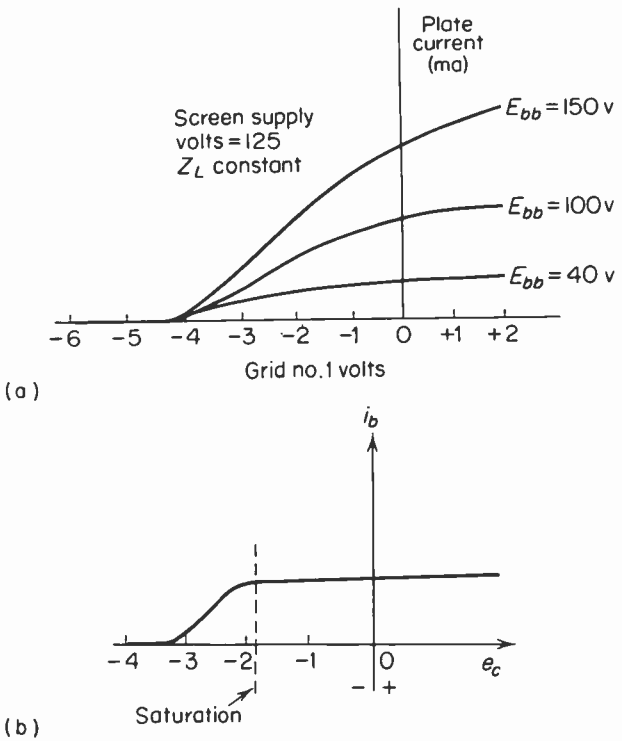



**Fig. 8-3.** (a) Transfer curves for three different values of plate *supply* voltage. For very low values of plate supply voltage, the plate current saturates in the negative grid region. (Dynamic curves). (b) Dynamic transfer characteristic using low plate and screen voltages. The plate current commences to saturate in the negative grid region.

By operating the sharp-cutoff pentode with reduced screen and plate supply voltages, the shape of the transfer curve can be greatly modified as shown in Fig. 8-3(b).

The next item concerns grid-leak bias. This method is very commonly used in oscillators, grid-leak detectors, sync clippers, and a host of other applications. Consider the simple circuit shown in Fig. 8-4. Here, the output of the generator, designated $e_s$, is sinusoidal and of constant amplitude. The generator is very often a series tuned circuit consisting of the secondary of a tuned transformer. The DC resistance of the secondary coil is assumed to be zero, the resistor has a value typically between 50 and 100 kilohms, and the reactance of $C$ is considered to be negligible at the frequency of interest. The cathode is at ground potential and all contact bias effects are ignored.

A better understanding can be had if the discussion centers around the transfer characteristic for this tube, which is shown in Fig. 8-5. The input signal is seen to be operating about the initial zero-bias point.

Consider the effect when the input signal begins to rise from zero toward maximum positive volts. Since the initial bias is zero, the control grid is immediately driven positive with respect to the cathode, resulting in the flow of grid current. The capacitor begins to charge (via $e_s$) toward the peak amplitude of the applied voltage. When the grid of a tube is positive with

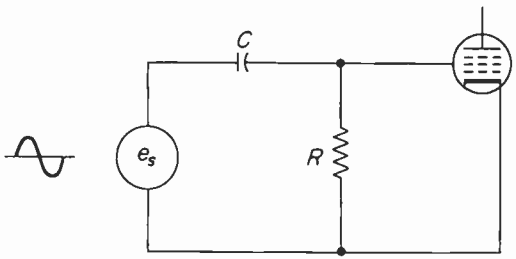



**Fig. 8-4.** The basic circuit for grid-leak bias. The control grid-cathode portion of the circuit performs as a diode detector with an $RC$ load.

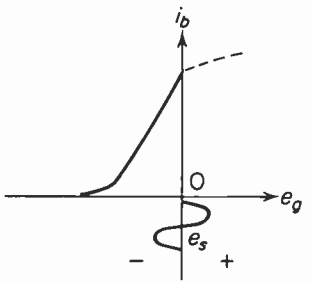



**Fig. 8-5.** The transfer curve showing the first cycle of $e_s$ operating at zero bias. Successive cycles will cause an increase in bias and a shift in the operating point. (Dynamic curve)

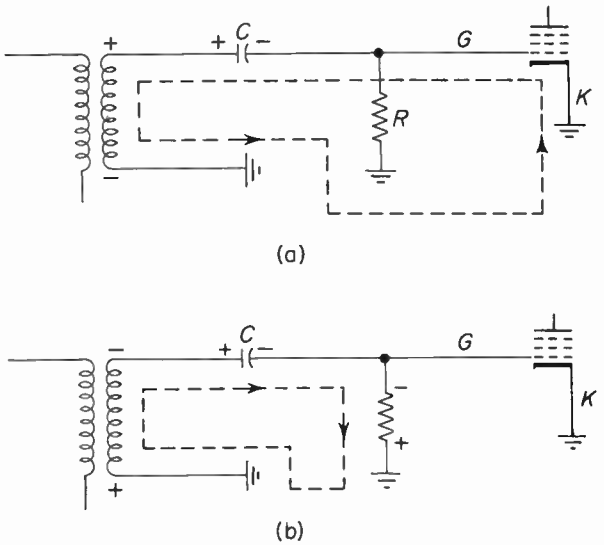
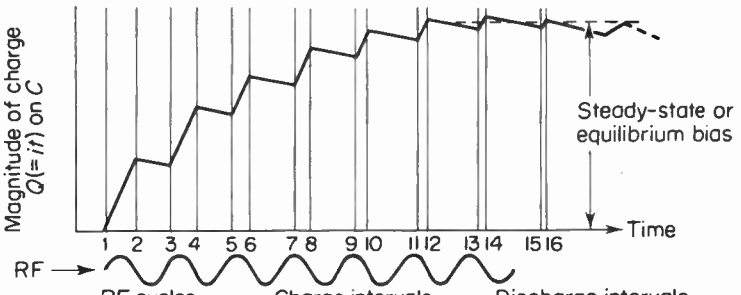



**Fig. 8-6.** The equivalent circuits for grid-leak bias. (a) Shows the charge path for *C* when the grid is positive with respect to the cathode. (b) Shows the discharge path for C when the grid is negative with respect to the cathode. It should be emphasized that when the grid is positive with respect to the cathode, grid current does not flow through *R* since the grid-to-cathode path does not constitute a source of *emf*.

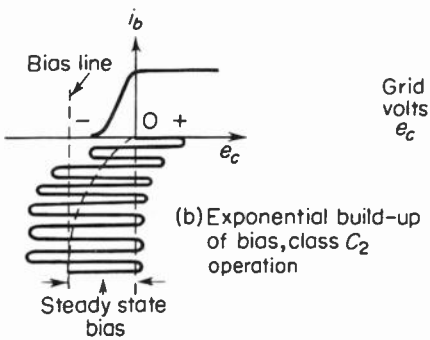
respect to the cathode, the grid-to-cathode path resistance is low (1 to 5 kilohms) and the capacitor can charge rapidly. The charging circuit for approximately the first quarter-cycle of the applied voltage is shown in Fig. 8-6(a). Note that *R* is not *directly* involved at this time. For the balance of the input cycle the grid is negative with respect to the cathode, and thus the grid-to-cathode resistance is considered infinite. A fundamental criterion for maintaining grid-leak bias is that the discharge time constant for the bias network be long compared to the time of one cycle of the input signal. This will insure that the bias will not vary (to any great extent) over the IF cycle. For an FM limiter application, the discharge time constant should be at least ten times the period of one cycle of a 10.7-MHz "carrier" (the standard IF). The discharge circuit for *C* is shown in Fig. 8-6(b).

On successive RF cycles, the capacitor builds up a charge until the charge in coulombs gained (on the positive signal peaks) is equal to the charge in coulombs lost (during the remainder of the IF cycle when the grid is negative with respect to the cathode). This point, and thereafter, is referred to as equilibrium bias, and the bias remains steady at a value approximately equal to the peak carrier amplitude. The bias buildup and steady-state waveforms are shown in Fig. 8-7 with the steady-state bias indicated as approximately twice cutoff.

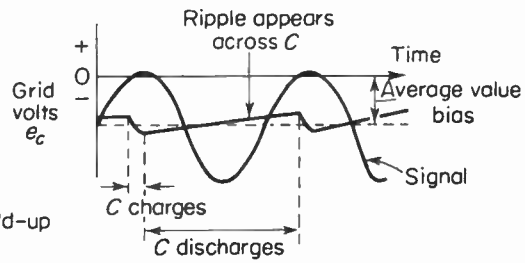



| RF cycles | Charge intervals | Discharge intervals |
|---|---|---|
| 1-3 1st | 1-2 | 2-3 |
| 3-5 2nd | 3-4 | 4-5 |
| 5-7 3rd | 5-6 | 6-7 |
| 7-9 4th | 7-8 | 8-9 |
| etc. | 9-10 | 10-11 |
| | 11-12 | 12-13 |
| | 13-14 | 14-15 |
| | 15-16 | 16-17 |
| | etc. | etc. |

(a) Charge build-up on $C$ of fig. 8–6. The capacitor takes more charge when grid current flows than it loses when grid current does not flow. The charging current becomes smaller on successive RF cycles since a larger negative charge keeps "piling up" on the grid side of $C$. See text.




(b) Exponential build-up of bias, class $C_2$ operation




(c) Steady-state waveforms for grid-leak bias

**Fig. 8-7.** The build-up and steady-state conditions for grid-leak bias. (b) Shows that the positive peaks of the signal are "clamped" to the zero grid voltage axis. This is more fully explained in the text. The magnitude of the saw-tooth ripple across C of part (c) is somewhat exaggerated for clarity.

Equilibrium bias is reached in a fraction of a second. The time it takes to reach this steady-state condition is a function of the charge time constant or the product of $C$ and the forward conduction resistance of the grid-to-cathode portion of the tube. The forward resistance of the grid-to-cathode is of the order of 4000 ohms for a typical case, making the charging time constant approximately 0.2 $\mu$sec. It thus takes a few cycles to build up the

bias, but once the bias reaches steady state, it remains approximately equal to the peak input signal amplitude.

Another important concept worth mentioning is the matter of "grid clamping." This comes about as follows. When the control grid is driven positive with respect to the cathode, grid current flows and the forward conduction resistance of the grid-to-cathode is low. In other words, the grid and cathode are behaving like a forward-biased diode. If the conduction resistance is of the order of 1000 ohms, then the grid-to-cathode voltage is 0.5 volt (approximate). This is equivalent to the control grid's being "held" at approximately cathode potential—or the grid is said to be "clamped" to approximately zero volts on the positive peaks of the applied voltage. To indicate how the positive peaks of signal remain clamped to zero even when the amplitude of the applied voltage changes, refer to Fig. 8-7(b). Assume that the bias has reached the steady-state value and that a slight decrease in amplitude of the input voltage takes place (such as by signal fading). The magnitude of grid current (and therefore bias) must tend to *decrease*, since the positive peaks will not drive the grid positive as they did before the decrease in signal. This will discharge $C$ and shift the bias line to the right. Since the input signal always operates about the bias line, the entire wave must shift to the right. The positive peaks are once again clamped to the zero grid voltage axis. Now, assume a slight increase in signal amplitude (such as caused by noise). This results in increased grid current, since the signal peaks drive the grid further positive. The increased grid current charges the capacitor to a higher voltage, the bias increases, and the bias line must now shift to the left. The entire wave shifts to the left, the positive signal peaks are again clamped to zero grid voltage, and a new steady-state condition results. Thus, if the time constant of the bias network is long compared to the time of one cycle of the input signal *and* short compared to the time it takes for the carrier amplitude to change owing to a disturbance, the positive signal peaks will be clamped to zero grid voltage. This is because the bias will vary directly and practically instantaneously with the input signal amplitude.

The above discussion of clamping is based on the ideal arrangement, and it will be shown that the bias may not be able to instantaneously readjust under all disturbance conditions. If the bias is unable to follow rapid fluctuations in carrier amplitude, limiting will fail. However, it should be recognized that *if* the positive signal peaks can be held to approximately zero grid voltage, no plate-current amplitude changes will take place with changes in the amplitude of the input signal (this assumes that the peak signal exceeds half the cutoff bias).

Consider now in detail the operation of a grid-leak limiter and how it provides an output voltage that is independent of input amplitude and rate of change of input amplitude. A typical circuit is shown in Fig. 8-8.

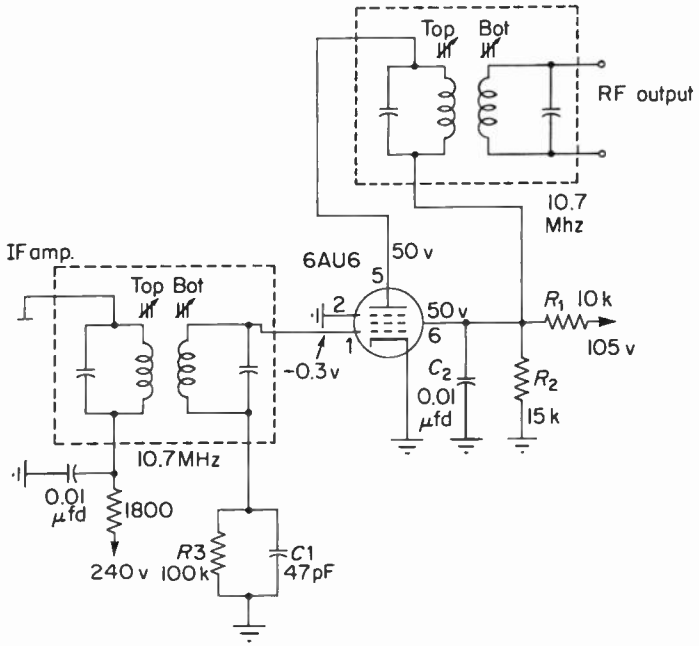



**Fig. 8-8.** A representative grid-leak limiter circuit.

The tube is a sharp-cutoff type operating with reduced screen and plate voltages. This arrangement provides a working grid base of approximately 2 volts, while plate-current saturation commences at approximately −0.5 volt at the grid. The plate supply voltage is derived from a voltage divider and is reduced to 50 volts. This same divider furnishes the screen voltage, which is also 50 volts. The divider stabilizes the screen and plate voltages against changes in bias, tube replacements, aging, loading variations, and any other small changes that take place *within* the stage. The divider network does *not* stabilize against variations in the power supply, since if the supply voltage were to change, the voltage across $R_2$ would also change in proportion. However, if the magnitude of direct current through $R_2$ is made large compared to the screen current, then a change in screen current will hardly be felt as far as total current through $R_1$ is concerned. Thus, the voltage drops across $R_1$ and $R_2$ are held reasonably constant.

The bias network consists of $R_3$–$C_1$ and has a discharge time constant of 4.7 μsec. This time is long compared to the period of a 10.7-MHz signal and short compared to the period of a 15-kHz disturbance.

Consider the transfer characteristic for this tube under the conditions previously described. This is shown in Fig. 8-9, together with an input signal that is varying in both frequency and amplitude. The steady-state bias as a result of grid-leak action is approximately 0.75 volt. The curve shows that

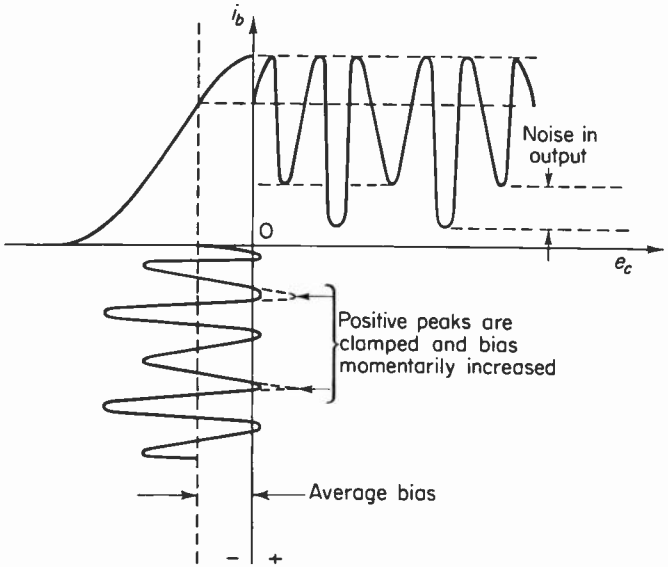



**Fig. 8-9.** If the signal amplitude is too small to drive the grid below cut-off, variations in the amplitude of plate current result in a noisy output. This is true whether the positive peaks are clamped to zero or not. The change in bias due to the sudden increase in amplitude is not shown for reasons of clarity.

the tube is acting essentially as a class $A_1$ amplifier. That is, changes in the amplitude of the input will produce changes in plate-current amplitude even if the positive peaks are clamped to zero volts, since the negative peaks are not capable of driving the tube below cutoff. The output-current waveform is thus both frequency and amplitude-modulated and the noise peaks have not been removed. The FM detector audio output will be noisy, since no limiting has taken place.

Consider now the case where the input-signal amplitude increases the bias to approximately twice cutoff. The positive peaks are again clamped to zero, but this time the negative signal peaks drive the tube well below cutoff. Any change in input-signal amplitude above the value shown will not result in any change in the peak-to-peak amplitude of the plate current. This is because the negative peaks drive below cutoff while the positive peaks are clamped to zero (aided by the fact that any increase above −0.5 volt will result in plate-current saturation). This is shown in Fig. 8-10.

The period of time during which plate current flows for the condition of Fig. 8-9 is essentially the full 360° of the input cycle. The conduction time corresponding to the signal shown in Fig. 8-10 is less than a full cycle. If the input signal increases above that shown, the conduction time for plate current is even smaller. The plate-current waveform can be analyzed into

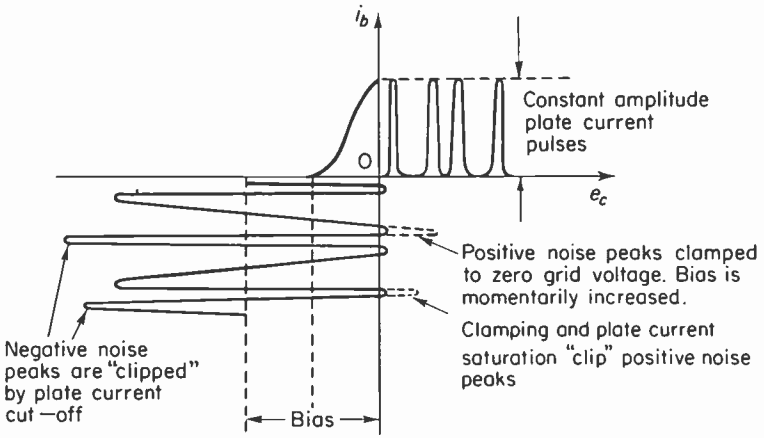



**Fig. 8-10.** Noise rejection in a grid-leak limiter when the peak input signal exceeds the cut-off (gridbase) of the tube. The plate conduction angle here is less than 180° of the input cycle.

a DC component, a fundamental-frequency sinewave component, and harmonics of this fundamental component. As the plate conduction angle becomes smaller with increasing signal amplitude, the amplitude of the fundamental-frequency component becomes smaller. A graph of input-signal amplitude as a function of the fundamental-frequency component of plate current is shown in Fig. 8-11. Note that with small input signals (signal amplitudes less than the grid base of the tube), the output rises linearly with signal increase and then reaches a threshold (point *A*). Beyond the threshold, the output is essentially independent of the input level. The output does tend to drop off slowly beyond this level, but careful design can minimize this reduction.

Note that the plate-current pulses comprise a complex waveform. Figure 8-8 indicates a tuned transformer in the plate circuit of the limiter. This transformer is tuned to the IF, which constitutes the fundamental-frequency

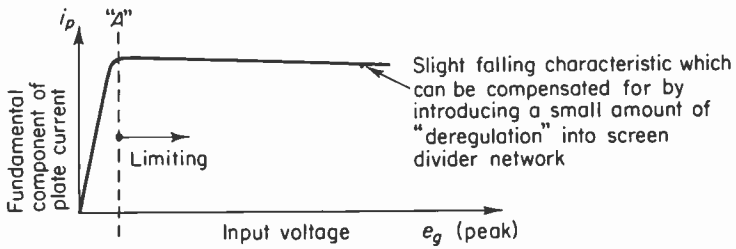



**Fig. 8-11.** The limiting action (above the threshold of limiting "*A*") is as a result of the fact that the fundamental component of plate current varies little with increasing signal amplitude.

component of plate current; therefore, its impedance to the higher-ordered harmonics will be low. This is particularly so if the circuit $Q$ is high, which generally it is. Thus, the DC component is completely removed while the harmonic components are greatly attenuated, since the stage gain is extremely low at these frequencies. The resultant waveshape at the IF output terminals is then a sinewave, since this is the waveshape of the fundamental component. The frequency deviations of the input wave have been retained, since the fundamental-frequency component of plate current changes with the instantaneous deviation of the FM carrier.

## 8-4. Time-constant considerations

It has been mentioned that in order to neutralize any amplitude variations of the carrier, the bias should be able to change instantly (the circuit should be fast-acting) in order to clamp the positive peaks to the zero grid-voltage axis. This implies that the charge and discharge time constants of the bias network should be as small as possible. This will insure that on a sudden increase in amplitude (*upward* AM) the capacitor will charge quickly to the signal peak to increase the bias; and on a sudden decrease in carrier amplitude (*downward* AM) the capacitor will quickly discharge through the grid-to-cathode path of the tube when the grid is more positive than the cathode. Since the resistance of the grid-to-cathode portion of the tube ($R_{gk}$) is essentially fixed for a given tube, only the value of the capacitor can be varied to meet time-constant requirements. With respect to the discharge time constant, either $C$ or $R$ can be varied, and these should be made as small as possible consistent with avoiding certain problems.

Using Fig. 8-12, it will be seen that if $C$ is reduced drastically, its reactance will be high. A voltage divider is then formed with the input capacitance of the tube, and a loss of IF signal to the tube will result. Limiting action will be adversely affected, owing to insufficient drive. If the value of $R$ is made very small to decrease the discharge time constant, the damping imposed on the tuned circuit that drives the limiter will be increased. The damping imposed† on the tuned circuit for the configuration of Fig. 8-8 is approximately $R/2$. This is equivalent to placing a resistance equal to half the grid-leak resistance across the secondary winding of the input transformer. The equivalent circuit for this condition is shown in Fig. 8-12(a). In addition, this resistance reflects back a resistance in parallel with the primary winding of the same transformer, which must then affect the stage sensitivity. In summary, then, a reduction in $R$ causes a resistance to be reflected back across the primary; the tuned-circuit $Q$ is reduced and a reduction in sensitivity results. If the limiter is of the type shown in Fig. 8-6, the damping imposed

†The derivation for this equivalent damping resistance is given in Chapter 10.

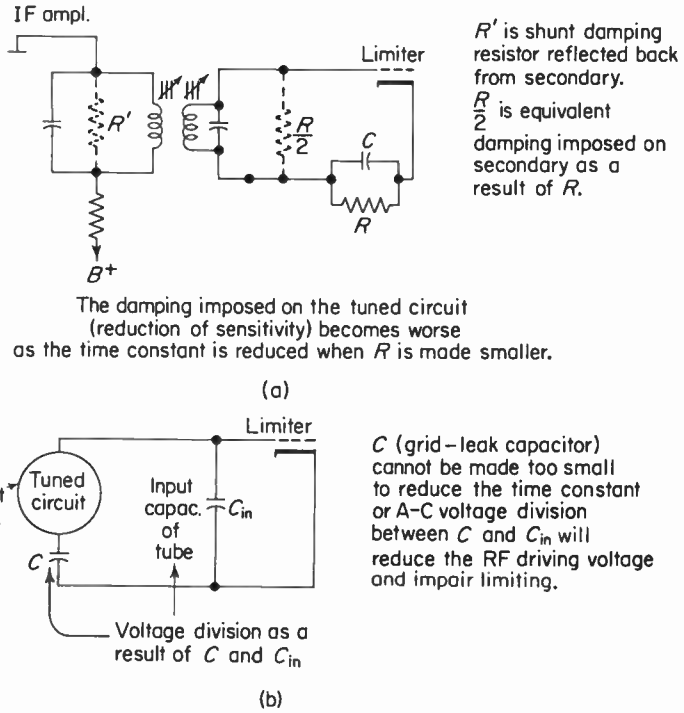



**Fig. 8-12.** The equivalent circuits showing why there are lower limits to the grid-leak bias network time constant. Practical time constants range from about 1.5 to 20 usecs.

on the tuned circuit is equivalent to placing a resistor (equal to the grid-leak resistor) in parallel with $R/2$, or

$$R_{\text{total damping}} = \frac{R\left(\frac{R}{2}\right)}{R + \frac{R}{2}} = \frac{R}{3}. \tag{8-1}$$

This arrangment thus imposes more damping on the driving circuit than the circuit of Fig. 8-8, but the same comments as to $R$ and $C$ apply here also.

A grid-leak limiter cannot completely suppress rapid changes in signal amplitude, particularly on very rapid *downward* AM. This is because the peak signal amplitude suddenly falls below the value necessary to drive the grid above plate-current cutoff. This will result in plate-current cutoff, and the output will be distorted. The capacitor must discharge through the grid resistor until the bias is sufficiently reduced to permit plate current to flow again, but this may take a number of IF cycles. During this time, however, the plate current is zero and remains so until the grid is driven above cutoff. This form of distortion, called "diagonal clipping," is shown in Fig. 8-13; it is analogous to what happens in a diode detector whose load is unable

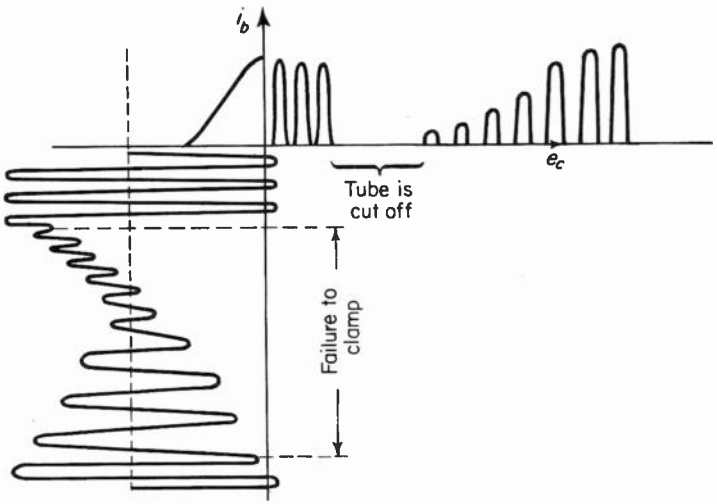



**Fig. 8-13.** The effect of a rapid downward change on the limiter output when the grid circuit time constant is too long to permit instantaneous bias readjustment.

to follow rapid changes in the modulation envelope. It can be minimized if the bias-network discharge time constant is made approximately 2 to 3 $\mu$sec. This low value can effectively reject amplitude variations occurring at a 150-kHz rate caused by an undesired modulated carrier operating on the *same* channel frequency as the desired carrier with a power ratio of desired to undesired signal of about 3 db.

There is inherently no problem on rapid *upward* AM, since the capacitor charges quickly through the grid-to-cathode forward resistance. The time constant is sufficiently small to permit almost instantaneous readjustment of the bias for clamping purposes.

Thus, a compromise must be reached in terms of damping, adequate IF sensitivity, a fast-acting bias network, and so on. Input-circuit time constants of the order of 1.5 to 20 $\mu$sec are used, depending upon such factors as the type of disturbance to be rejected and the number of limiters used.

## 8-5. Cascaded Limiters

Before discussing the advantages of a double limiter (two limiters in cascade), let us review some basic concepts relating to amplification. In a class $A_1$ linear amplifier using a pentode with the cathode at ground potential, the amplification of the stage is given by $A = g_m Z_L$. Since limiters often operate class $C_2$ (or $B_2$), the expression above is *not* valid since the operation is nonlinear. In a linear class $A_1$ amplifier, having an amplification of 15, a 1-volt input produces a 15-volt output. If we increase the input to 3 volts

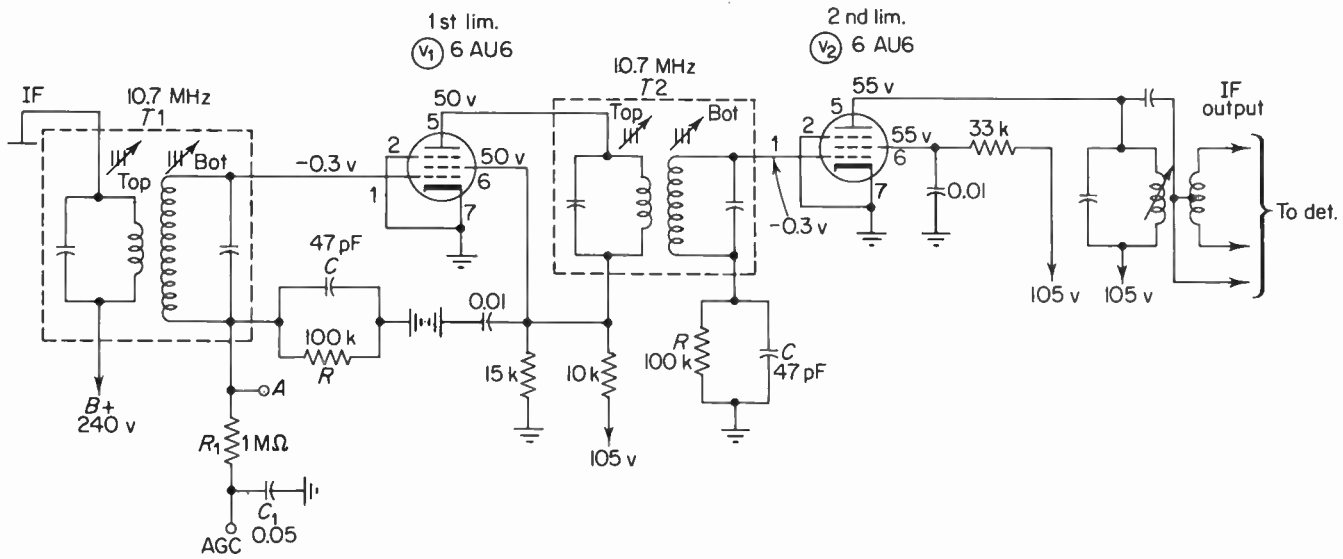



**Fig. 8-14.** A two-stage limiter arrangement.

and still operate linearly, the output will be 45 volts. This relationship remains valid until the stage as a whole becomes saturated or overloaded (class $C_2$, for example). With a saturated limiter, Fig. 8-11 shows that beyond the limiting threshold (point *A*), any further increase in input-signal amplitude produces essentially no further increase in output. Thus, the conventional concept of stage gain is not valid when operating beyond saturation. Instead, we shall refer to the stage gain at point *A* but not to the right of this point, since a 1-volt input at the limiting threshold might produce a 3-volt output while a 30-volt input will yield the same output. The low values of *threshold gain* for a limiter came about as a result of the low values of plate and screen voltage. It is emphasized that the primary purpose of a limiter is to amplitude-limit and not to provide high gain.

A reasonable value for the threshold gain of a single-stage limiter is of the order of 3. The limiting sensitivity of a single-stage limiter is of the order of 1 volt; that is, 1 volt peak signal is required to drive the limiter to point *A* on Fig. 8-11. Actually, to secure adequate limiting under all conditions of operation, signals of the order of 10 volts or more are required, which implies high gain for the IF amplifiers.

Assume now that two limiters are connected in cascade as shown in Fig. 8-14. In this arrangement, the limiter stages are transformer-coupled. The bias time constants in this circuit are the same for both stages, and it is assumed that each stage has a threshold gain of 5 with a limiting sensitivity of 1 volt. Limiting will commence at 1 volt for each stage. If only a single limiter were used with a threshold gain of 5, an input signal of 1 volt would be required to just saturate the stage (corresponding to point *A* of Fig. 8-11). If we now cascade two limiters, a 0.2-volt input to the first stage is amplified and appears as a 1-volt input to the second limiter. This is amplified by a factor of 5 (this becomes 5 volts), which is adequate to just saturate the second limiter. The limiter circuit as a whole (both stages) can now limit with an input signal of only 0.2 volt where a single limiter would require a 1-volt signal. The limiting sensitivity is thus improved by a factor of 1/0.2 or 5. Two limiters in cascade therefore provide more than twice the sensitivity of a single limiter and thus give good noise rejection on weak signals. Also, since the input to the first limiter has already been substantially amplitude-limited, the requirements on the second stage can be somewhat relaxed (that is, its screen voltage can be raised slightly for increased gain so as to increase the input to the detector and provide a higher audio output).

## 8-6. Automatic gain control (AGC)

In order to minimize the possibility of RF and IF amplifier overload, some receivers use the signal-developed bias as a voltage source for controlling the sensitivity of the RF and IF amplifiers. Since the magnitude of grid-

leak bias is determined by (a) the bias time constant, (b) the potentials at the electrodes of the tube, and (c) the input-signal amplitude, then for a given time constant and fixed operating potentials, the grid bias is proportional to the signal amplitude. Since the bias may vary at an audio rate or even faster under conditions of heavy impulse interference, it is necessary to filter these variations and provide a steady direct voltage approximately equal to the peak carrier. In other words, the DC component at point *A* of Fig. 8-14 is proportional to the carrier level, but it also contains an AC component that is varying in accordance with the modulation envelope. This is normal limiter operation. For purposes of AGC, however (as well as tuning-indicator, squelch, and other applications), these variations must be removed. A low-pass filter will remove them, and the remaining DC component will thus be available for AGC purposes. This AGC voltage must not vary even if the carrier amplitude is changing as slowly as 100 Hz and this implies a filter time constant that is long compared to the lowest rate of change of the modulation envelope. The values used for the time constant are typically 0.05 second. If we assume that the modulation envelope is varying at 50 Hz due to noise, the period of the disturbance is $\frac{1}{50}$ or 0.02 second. This is reasonably long, but it is still short compared to the filter time constant. The filtering will thus be effective down to about 50 Hz.

This voltage, being proportional to the IF level, is also suitable for activating tuning indicators ("magic eye" tubes, meters, and so on). This is another reason for providing a well-filtered AGC voltage in order to prevent the tuning eye from "winking" at an audio rate. If, however, the AGC filter time constant is made excessively long to improve the filtering, the receiver will be difficult to tune; in tuning from a strong to a weak station, the weak station may not be heard (it will simply be passed over) because the AGC voltage takes too long to readjust.

If a VTVM is connected across $C_1$ of Fig. 8-14 and an unmodulated 10.7-MHz signal is injected in at the mixer input, the IF tuned circuits can be adjusted for a maximum indication and IF alignment can be performed without a sweep generator.

## 8-7. Input circuit detuning

It has been mentioned that grid current on the positive peaks imposes damping on the preceding tuned circuit, and power is thereby "taken" from the primary circuit. Since the power taken by the equivalent damping resistance is proportional to the square of the input voltage, the damping will become worse with increasing signal. Damping causes an increase in the bandwidth owing to the reduction of $Q$. This reduction in sensitivity on weak to medium signals can be prevented by the introduction of a delay in developing the bias. A small bypassed resistor in the cathode leg can provide

the delay, since the input-signal amplitude must exceed the fixed cathode bias prior to grid-current flow. Thus, for weak signals, the stage performs like a class $A_1$ amplifier for maximum sensitivity and there is no damping imposed on the tuned circuit for small signals. A circuit arrangement for providing "delayed AGC" is shown in Fig. 8-15. The peak carrier signal must exceed 0.5 volt before AGC voltage is developed.

Another area of concern is the effect that variation in bias has on the input capacity of the limiter tube. It was mentioned earlier that changes in bias result in a shift in the position of the space charge associated with the cathode. This space-charge shift gives rise to a change in the capacity between grid and cathode ($C_{gk}$). An increase in the DC bias on the grid due to an increase in signal causes a *decrease* in $C_{gk}$ and a reduction in the total tuning capacitance of the tank feeding the limiter. This changes the resonant frequency by causing the tuned circuit to resonate at a higher frequency than it would otherwise. On a weak signal input, the reverse occurs. In order to minimize these effects, a large value of physical tuning capacitance may be employed or the response may be broadened so that a slight amount of detuning will not affect the overall tuning characteristic.

The discussion above is particularly important with regard to alignment. If the alignment is carried out with a strong input signal (resulting in a large bias), the circuit will not be properly aligned for weak signals, since the bias change will detune the circuit. Strong signals are able to cope with noise and slight misalignment more effectively than weak signals. Therefore, it is important that the alignment be performed using a low-level signal. This will favor weak signals so that they can more efficiently override noise.

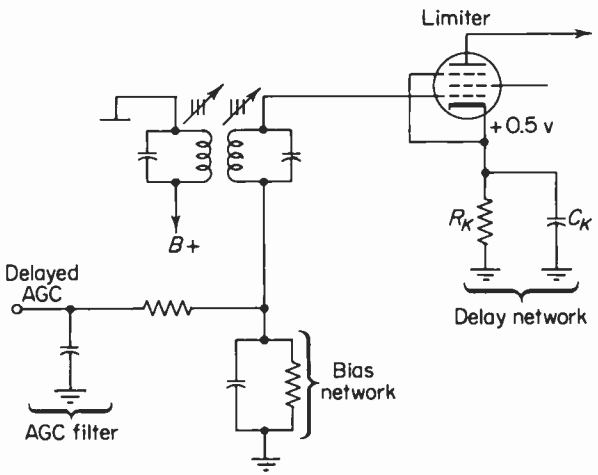



**Fig. 8-15.** Circuit arrangement for delaying the AGC voltage until the peak carrier exceeds 0.5 volt. This circuit will not properly drive a tuning indicator.

## 8-8. Miscellaneous considerations

A bias time constant of the order of 2 $\mu$sec can cope with cochannel interference reasonably effectively, but under conditions of heavy impulse interference the bias may not be able to follow the rapid changes in the modulation envelope. For signal peaks exceeding a certain value, no further change in the output can take place since the plate current remains saturated. Thus, the plate-current saturation characteristic *supplements* the grid clamping on the positive peaks to neutralize the disturbance. In addition, the lower the bias network time constant, the more effectively the stage can reject impulse noise (the impulses are of short duration—about 7 $\mu$sec). A time constant of about 1 to 3 $\mu$sec can reject impulse noise with a fair degree of effectiveness, provided the carrier amplitude is reasonably large compared to the noise amplitude. If the signal is small, then after the first noise burst, the tube will be cut off.

Another point concerns the sequence of events when the input-signal amplitude increases. The direct component of tube current decreases, the screen current tends to decrease, the screen voltage tends to rise, and the grid base is lengthened. This causes an increase in the amplitude of the fundamental component of plate current. Thus, by means of careful screen-grid bias network design, the falling characteristic shown in Fig. 8-11 can be compensated for.

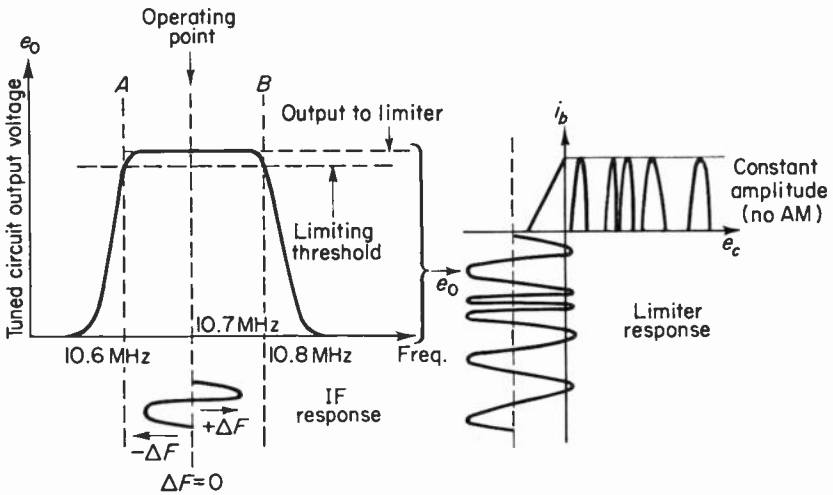



**Fig. 8-16.** A desirable tuned circuit response (IF) for optimum limiting. The passband is flat for 200 kHz (to handle a carrier swing of 150 kHz), the circuit is properly aligned and all carrier deviations result in equal outputs above the threshold of limiting. Positive deviations of the carrier (above center frequency, $+\Delta F$) give the same output voltage as negative deviations ($-\Delta F$). The input to the limiter is shown to the right of the tuned response characteristic and the limiter output contains no AM.

Improper alignment (as well as very narrow, single-peaked IF stages) can also have a marked effect on the AM rejection capabilities of the limiter stages. Assume that the tuned response of the circuit just prior to the first limiter is as shown in Fig. 8-16. The response is flat-topped and its passband is 200 kHz between points *A* and *B*. If no noise appears at the input, the output due to carrier-frequency deviations will be a constant-amplitude variable-frequency IF voltage, since all frequencies within the passband are amplified equally. The detector output will be distortionless and conditions are optimum. Assume, now, that the tuned circuit feeding the limiter is misaligned *or* the front-end tuning dial is incorrectly set. This will shift the operating point on the IF passband to point *C* as shown on Fig. 8-17. For small frequency deviations this will make little difference, since the tuned-circuit output is still above the limiting threshold. If, however, the degree of modulation is increased at the transmitter and the carrier deviation is increased to the point where the frequency swings beyond point *A*, both AM and spurious FM will be introduced. The AM is due to the fact that when the carrier deviates below center frequency, these deviations result in less IF output than for positive deviations of the carrier. Since larger negative deviations of the carrier result in still smaller IF outputs, a point will be reached where the

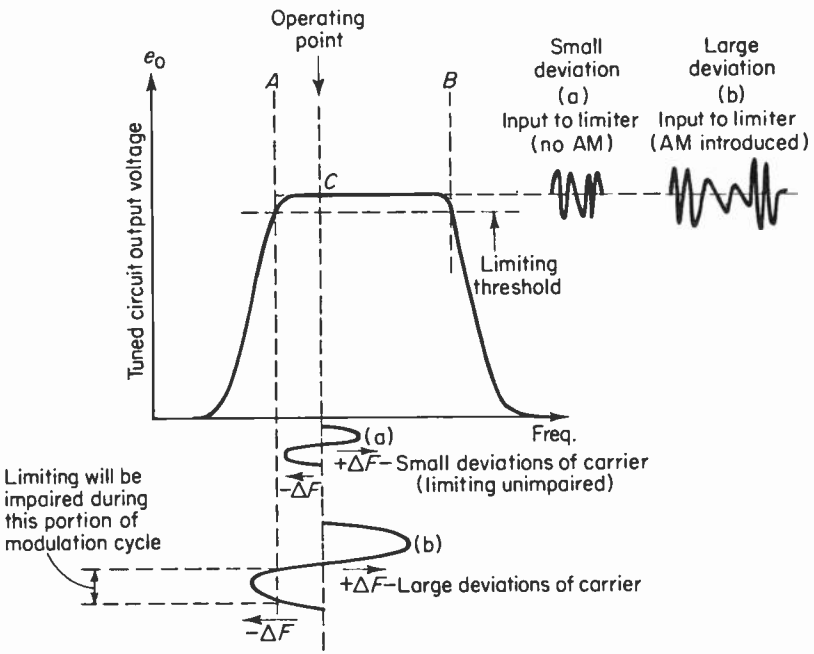



**Fig. 8-17.** The AM introduced as a result of misalignment or receiver improperly tuned. For small deviations, the output remains above the threshold of limiting but for large deviations, the output of the IF amplifier falls below the threshold level during a portion of the modulation cycle. See Fig. 8-18 for the limiter output.

instantaneous IF output will go below the limiting threshold and limiting will be impaired. The limiter input for large carrier swings under conditions of an off-resonant IF is shown in Fig. 8-18. Notice that the limiter output contains amplitude modulation due to operation below the limiting threshold during a portion of the modulation cycle when the deviation is below center frequency.

The importance of good phase linearity cannot be overstressed. If the tuned-circuit passband is very narrow or if the receiver is improperly tuned or aligned, large carrier deviations will result in a shift in the phase of the applied signal. As discussed in an earlier chapter, phase shift is equivalent to a small frequency deviation which limiters cannot remove from the input. This spurious FM is demodulated by the detector stage and appears as a distorted output. This form of distortion is most severe at high modulation levels, since at these times the instantaneous frequency at the input to the IF stage is furthest removed from the 10.7-MHz resonant frequency. With an adequate IF passband and a properly aligned and tuned receiver, the difference between the center of the response and the points corresponding to maximum excursions of the carrier is small and little phase shift is introduced. The detector output will thus be undistorted, since the spurious FM introduced (spurious AM will be rejected) will be negligible compared to the signal-carrier deviations.

Transformer-coupled limiters provide high gain and good selectivity, but they are more expensive and their alignment is not as simple as for the

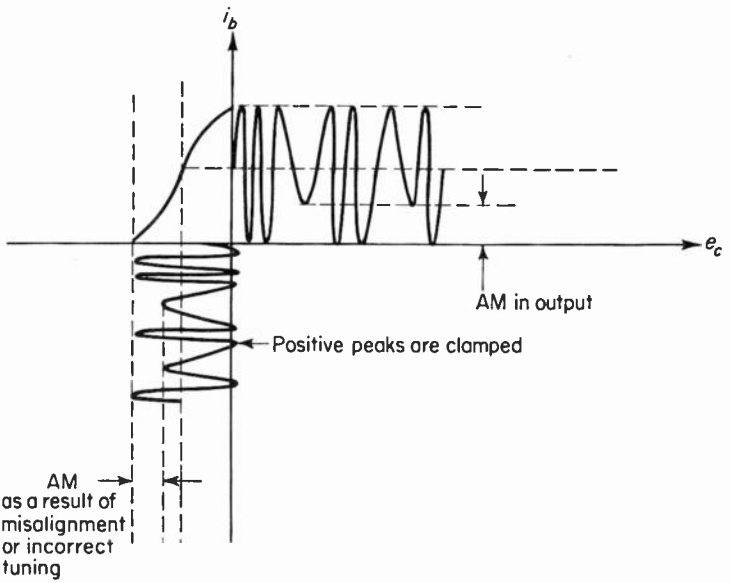



**Fig. 8-18.** Limiter plate current output contains AM due misalignment.

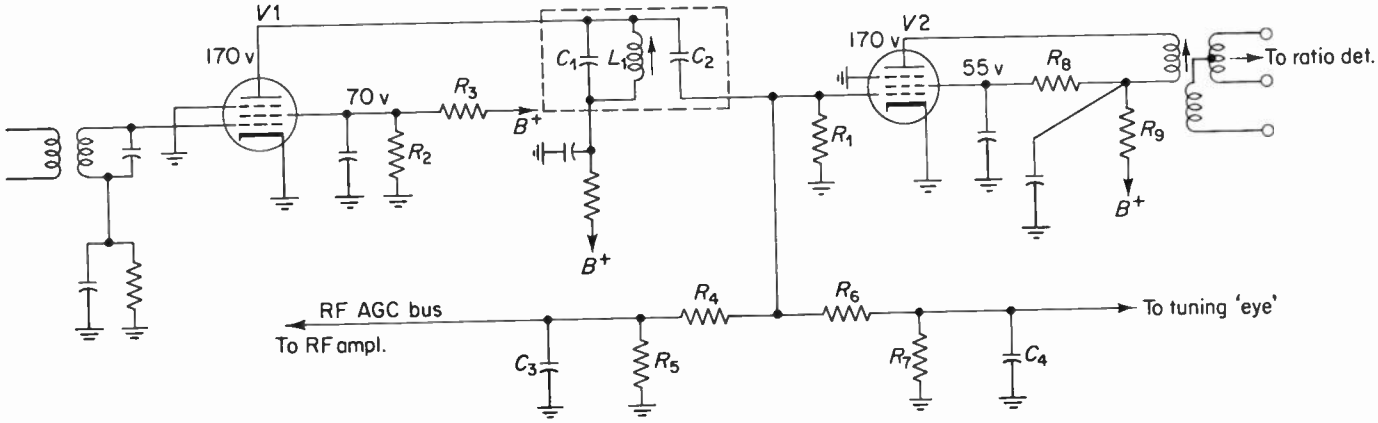



**Fig. 8-19.** An impedance-coupled cascaded limiter circuit.

impedance-coupled arrangement. It provides a very symmetrical response with minimum ripple within the passband. This is necessary for minimum phase shift and AM introduced by the tuned-circuit selectivity. A double-tuned transformer-coupled limiter arrangement generally provides a response that is flatter near resonance as well as steeper sided, but for the same resonant circuit and tube employed, the impedance-coupled (single-tuned) circuit provides twice the amplification at resonance of the double-tuned transformer circuit. This assumes that the double-tuned arrangement is critically coupled (or $k = 1/Q$). Since adequate selectivity can be provided ahead of the first limiter, impedance coupling between limiters is a very desirable method owing to its economy, high gain, and relative ease of alignment. Also, since the last limiter stage can provide damping of the tuned circuit due to grid current (especially on strong signals), the response will be somewhat flattened to provide a more linear phase characteristic. If grid-current loading does not provide the optimum damping for good phase linearity, resistive damping across the coupling coil may be employed.

Figure 8-14 shows a representative transformer-coupled cascaded limiter circuit. The damping imposed on the secondary tuned circuit is approximately 50 kilohms, giving a flat-topped response with good phase linearity and high gain. A circuit arrangement employing impedance coupling is shown in Fig. 8-19 with the coupling impedance as well as the coupling capacitor enclosed in a shield. This is good practice and will minimize the possibility of harmonic radiation back to the front end of the receiver. $L_1$ is adjusted by means of a single ferrite-core (powdered-iron) slug providing a wide tuning range, higher $Q$ and good temperature stability.

The second limiter ($V_2$) provides AGC as well as tuning-eye voltage. Both voltages must be filtered to remove changes due to AM (noise, fading, and so on); this is provided by $R_4$ and $C_3$ for AGC filtering and $R_6$ and $C_4$ for tuning-eye voltage filtering. Since the working grid base of the RF amplifier is approximately 5 volts, the AGC voltage on the RF AGC "bus" cannot exceed this value. In order for 5 volts to appear across $R_5$, approximately 20 volts must appear across the series combination of $R_4$ and $R_5$. This 20 volts is accordingly the developed grid-leak bias of $V_2$. If the grid base of $V_2$ is approximately 4 volts, 20 volts peak should insure good limiting at high signal levels.

## 8-9. The dynamic limiter

The limiter arrangement previously described has the major disadvantage that below a certain threshold level, limiting action stops. In other words, below about 2 volts peak input to the limiter, the receiver does not "quiet" at all. Also, under conditions of heavy impulse-type interference as well as multipath reception, the peak input signal must exceed 2 volts (by a factor

of about 4 or 5) in order to respond to rapid downward modulation. The maximum modulation that a grid-leak limiter can handle is the ratio of the difference between the peak signal (without noise) and the minimum signal required to saturate the stage to the peak signal (without noise). This is a variable quantity; thus the limiter cannot be relied upon to perform well under all conditions. In strong-signal metropolitan areas this is not too serious a problem, but in rural areas it must be given special consideration.

The limiter type to be discussed now is based upon the concept of variable damping of a tuned circuit. That is, if a tuned circuit comprises an amplifier load that is damped by means of a resistor, the magnitude of the composite impedance decreases and the stage gain decreases accordingly. If this shunt damping resistance can be made to vary inversely with signal amplitude, then the gain will decrease with increasing signal and increase with decreasing signal. Thus, the output can be made independent of changes in the input over a relatively wide range.

The basic circuit of a dynamic limiter is shown in Fig. 8-20. First, consider the *RC* network in the cathode of the diode. If its discharge time constant is made large compared to the period of the lowest audio frequency, then the voltage across the combination cannot readily change with changes of the modulation envelope. Let us take the case of a constant-amplitude carrier and consider what happens both in the diode and driver load.

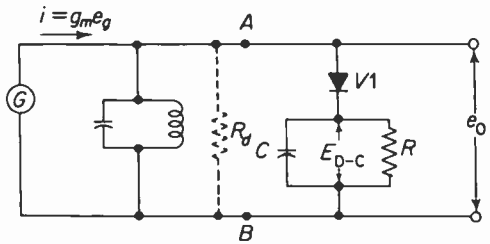



**Fig. 8-20.** Basic circuit of dynamic limiter. "*G*" is a constant current generator and the forward resistance of the diode is considered to be zero.

The input carrier waveform is as shown in Fig. 8-21(a). The potential across the *RC* combination is simply the rectified voltage and is approximately equal to the peak amplitude of the voltage across the tuned circuit. That is, *C* charges to the peak positive signal and, since the *RC* time constant is very long compared to the period of one IF cycle, the voltage remains constant. That is, practically no charge leaks off during the time that the diode is nonconducting (the diode is biased off for most of the IF cycle) and not passing current. The waveform across the *RC* combination is shown in Fig. 8-21(b). Since the power dissipated by *R* is integrated over the entire IF cycle, the damping imposed on the tuned circuit is likewise integrated. The magnitude of the equivalent damping resistor is given by *R*/2, designated

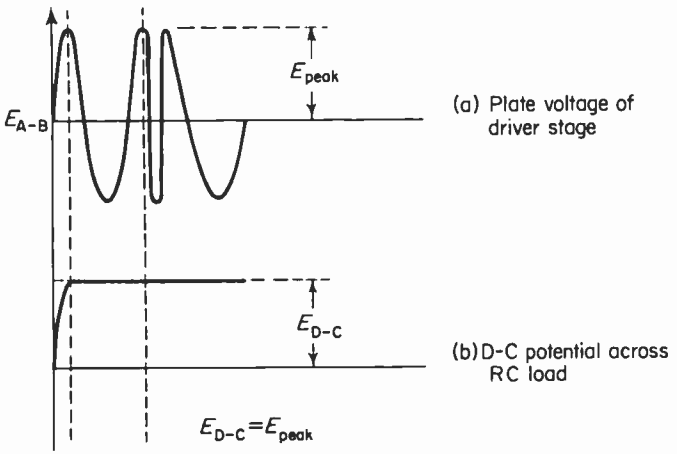



**Fig. 8-21.** The capacitor C charges to the peak amplitude of the input. The damping under these conditions (equilibrium) is equivalent to a resistor $R_d = R/2$ placed across the tuned circuit.

$R_d$. If the tuned circuit alone has an impedance that is high compared with $R_d$, then the gain of the driver stage will be almost wholly determined by the magnitude of $R_d$ (not the tuned-circuit impedance).

Assume that the carrier-input amplitude increases. Since the *RC* combination behaves very much like a battery, its voltage cannot readily change and the capacitor charging current simply increases. The diode and the associated *RC* network then behave like a "black box" whose voltage is maintained constant with increasing current. This is equivalent to a reduction in $R_d$, since the power taken from the generator must increase. If $R_d$ decreases, the total load impedance of the driver stage must decrease. Since the stage gain is given by $g_m Z_L$ and $Z_L$ is now reduced, the gain must decrease and the input change is effectively canceled.

If the input-signal amplitude decreases, the diode is cut off owing to the reverse bias provided by the *RC* network. This removes $R_d$, and the gain rises to cancel the reduction in input-signal amplitude. The degree of *downward* AM rejection is, in effect, a function of how high the gain of the driver stage can rise above the equilibrium gain (the stage gain under constant-input-signal conditions) and is determined by the ratio of $R_d$ to the tuned-circuit dynamic impedance ($QX$).

An advantage of the dynamic limiter over the grid-leak bias type is that the dynamic limiter operates on the basis of a variable threshold level while the threshold level for the former is fixed. As an example, suppose the peak carrier level is 1 volt. The diode load capacitor will charge to approximately 1 volt, and the static damping imposed on the tuned circuit is $R/2$. This resistance will vary the damping inversely with signal level to maintain a constant output. If the carrier level should now "permanently" increase (if we tune to a stronger station, for example), this long-term change will

simply cause $C$ to charge to a higher value, and once again there will be variable damping imposed on the tuned circuit to stabilize the output. If the carrier level should drop to a low value (if we tune to a weak station, for example), the long-term change in amplitude will result in the capacitor's slowly discharging through $R$. The load voltage will again readjust to the "new" peak carrier. Limiting will again take place as previously described. Each of the conditions above represents a threshold level for limiting (each level is different). The minimum threshold for dynamic limiting is based on considerations of diode linearity and efficiency.

At very low signal levels, the diode ceases to act like a linear device, and its forward conduction resistance $r_c$ is high. Under these condition its forward voltage drop will be high, and this will detract from the bias across the $RC$ load, since the capacitor will not charge to the peak signal. The damping imposed on the tuned circuit will decrease and, since the *downward* AM rejection capabilities of the dynamic limiter are a function of how low $R_d$ is compared to $QX$, the limiting ability of the circuit is impaired. This is shown in Fig. 8-22, where $r_c$ is shown as an equivalent resistance in series with

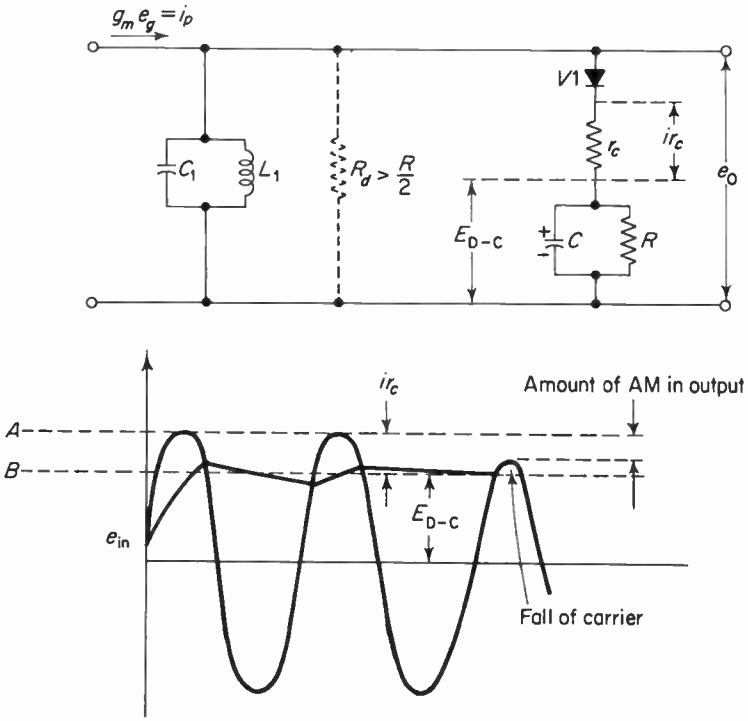



**Fig. 8-22.** The condition corresponding to a relatively large forward drop across the diode. The carrier amplitude can drop from $A$ to $B$ and no limiting will take place since variable damping does not come into play until point $B$ is reached.

the diode and load circuit. The rectified current will develop a voltage drop across the diode, and the capacitor will not, in this case, charge to the peak carrier. The difference between the peak carrier (point $A$) and the average voltage across the bias network (point $B$) is the amount of carrier reduction for which practically no limiting takes place. This is because variable damping of the tuned circuit essentially does not take place until the peak carrier falls below point $B$. A high-conductance diode should therefore be employed. Germanium diodes display a lower forward resistance than vacuum types, they are smaller and can be mounted in the same shield as the tuned circuit to minimize IF radiation, they do not require heater power and are thus free from heater-cathode leakage. In addition, their reverse resistance (which is of the order of 1 megohm) actually supplements the limiting characteristic. Since the reverse resistance of a crystal diode is finite, the diode is never truly "cut off." This means that a degree of limiting can take place on a rapid fall of the carrier where a vacuum type might be cut off.

We noted earlier that the downward AM rejection capabilities of the dynamic limiter were a function of the damping imposed on the tuned circuit, and that the greater the degree of damping imposed under static conditions, the greater the amount of *downward* AM rejection. This is expressed† as a percentage,

$$\text{per cent } \textit{downward} \text{ AM rejected} = M = 100\% \left(1 - \frac{Q'}{Q}\right), \qquad (8\text{-}2)$$

where $Q$ = the tuned-circuit $Q$ without the damping imposed by the diode and its load, and $Q'$ = the circuit $Q$ with the tuned circuit shunted by the diode and its load. If the unloaded $Q$ is made approximately 100 and the loaded $Q$ is made approximately 25, the per cent of *downward* AM rejected is 75.

A graph showing the limiting characteristic is given in Fig. 8-23. The dotted curve labeled $iZ_t$ is a plot of the input current versus output voltage and shows how the gain rises linearly with increasing input current. This shows that no limiting takes place, since the diode and associated load current are not connected across the tuned circuit. The curve of input current versus output voltage under equilibrium conditions (constant carrier level) is the static characteristic and is labeled $iZ_tR'/(Z_t + R')$. Curves $A$, $B$, and $C$ are the dynamic characteristics of the limiter for three different values of average carrier current. Curve $A$ represents a high average carrier current and $C$ is low. The intersection of the static load line and the dynamic characteristics represents the three average operating points for three different average inputs. For curve $A$, the threshold point is point $X$. This means that the carrier current can fall from point $X$ to point $X'$ and the IF output voltage remains essentially constant. Curve $B$ indicates that limiting com-

†See Appendix 8-1 for the derivation of this expression in addition to a sample design.

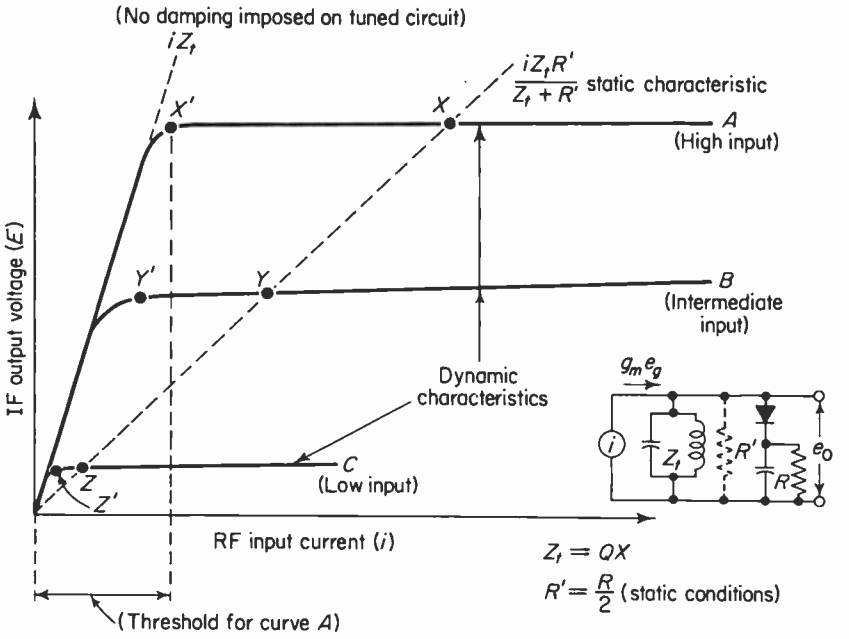



**Fig. 8-23.** The limiting characteristic for a dynamic limiter showing the effect on the output for different input carrier currents.

mences at a lower IF level but the range of dynamic limiting is less ($Y$ down to $Y'$). Thus, the percentage of *downward* AM rejection must be less. Curve $C$ indicates that the threshold level is still lower but the range of dynamic limiting is quite small ($Z$ down to $Z'$).

The limiting characteristic can be improved (at the expense of gain) by reducing the value of the diode load resistor. This will reduce the slope of the static characteristic and shift the operating points to the right on Fig. 8-23, increasing the range of operation. This cannot be done without limit, for a number of reasons. First, the forward conduction resistance $r_c$ will become an appreciable percentage of the diode load resistor, and the limiting characteristic will not tend to be as flat. Second, a reduction of $R$ will reduce the discharge time constant, and the diode bias will not remain steady with rapid changes in the modulation envelope. This will result in output voltage variations. The reason is that if the current into the diode circuit increases and causes an increase in the voltage across the load circuit, the damping does not vary. It is this *variation in damping* that provides the dynamic limiting.

The advantages of this type of limiter are as follows. Limiting is effective down to very low values of input voltage and is limited by the diode conduction resistance $r_c$. This resistance tends to increase at very low values of applied voltage because the diode is more efficient at high input levels.

Since the limiting characteristic does not depend upon the rapid charge and discharge of a capacitor, it does not "block" on rapid fall of the carrier. That is, the IF output is not cut off as would be the case for a grid-leak type (see Fig. 8-13).

The disadvantages of the dynamic limiter are as follows. Since the bandwidth of the tuned circuit is

$$B = \frac{f_r}{Q}, \tag{8-3}$$

variations of $Q$ with AM cause variations in the bandwidth of the driver stage. Its selectivity characteristic is thus poor, particularly under conditions of *upward* AM. Also, since the load capacitor will readjust to long-term changes in the carrier amplitude, the audio output does not remain fixed with carrier-level changes. An effective AGC arrangement can, however, minimize this last effect. Also, since the dynamic limiter reduces the equilibrium gain of the driver stage by a factor of $R_d/(R_d + Z_t)$, this loss must be compensated for by means of an additional IF stage or possibly by driving the existing IF amplifiers harder. This may cause problems of instability. A dynamic limiter then may be more expensive if all factors are considered.

In *downward* AM rejection capabilities the dynamic limiter is comparable to a ratio detector,† somewhat inferior to a grid-leak biased limiter at very high input levels, and somewhat superior to it at low to intermediate levels. Its tuning characteristic, with respect to side response, is very similar to that of a ratio detector (see Fig. 10-51) and gives less interstation noise than a Foster-Seeley arrangement.

Improved limiting can be had by providing a dual diode arrangement as shown in Fig. 8-24. Symmetrical clipping of positive and negative noise

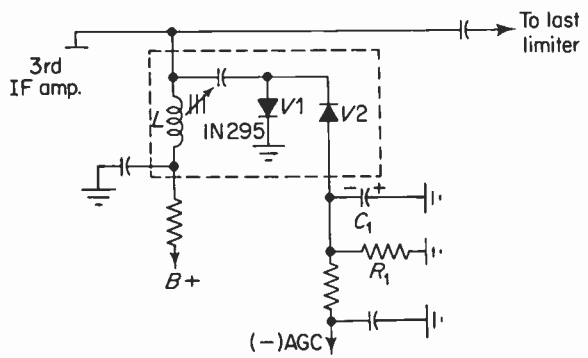



**Fig. 8-24.** A dual-diode limiter (clipper). In addition to more effective limiting this circuit provides *twice* the normal AGC voltage due to voltage doubler action.

†A ratio detector also limits amplitude by virtue of variable damping.

pulses is realized by the use of two diodes, $V_1$ clipping the positive pulse and $V_2$ clipping the negative pulse. The discharge time constant of $R_1C_1$ is sufficiently long to maintain the proper bias for both diodes and also provides an AGC voltage proportional to the average carrier level. The circuit shown is in the form of a half-wave voltage doubler. Therefore, the AGC bias is approximately equal to twice the peak carrier. This voltage is used to control the sensitivity of the RF and IF amplifiers and maintains a constant audio output over a wide range of input signals. The approximate equivalent circuit is shown in Fig. 8-25.

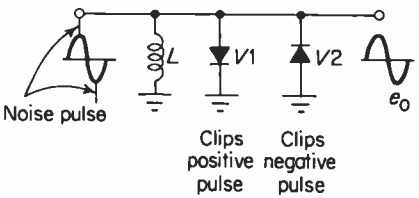



**Fig. 8-25.** Symmetrical clipping of noise pulses by both diodes give better rejection of impulse noise.

## 8-10. Transistorized limiters

Transistors have the advantages of small size, no heater supply, very long life, portability, and sensitivities and quieting levels comparable to vacuum tubes. One of the main drawbacks of transistors (particularly the RF and IF sections) is the generation of intermodulation products, which is closely related to overloading. Transistors will overload at much lower signal levels than vacuum tubes. However, the use of AGC and careful design will tend to minimize this problem.

The basic principle involved in transistorized limiters is that of an overdriven amplifier. That is, the collector current is alternately driven between cutoff and saturation. This concept is valid for grid-leak biased vacuum-tube limiters as well, but the signal conditions and biasing are, of course, different. Consider the basic circuit shown in Fig. 8-26. The transistor chosen is a PNP unit, but the explanation is valid for NPN units provided the battery polarities are reversed. The emitter-base junction is forward-biased by

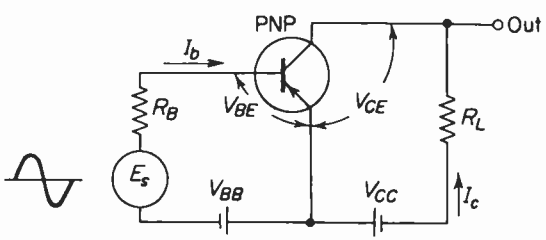



**Fig. 8-26.** Simplified circuit used to explain limiting action of transistor.

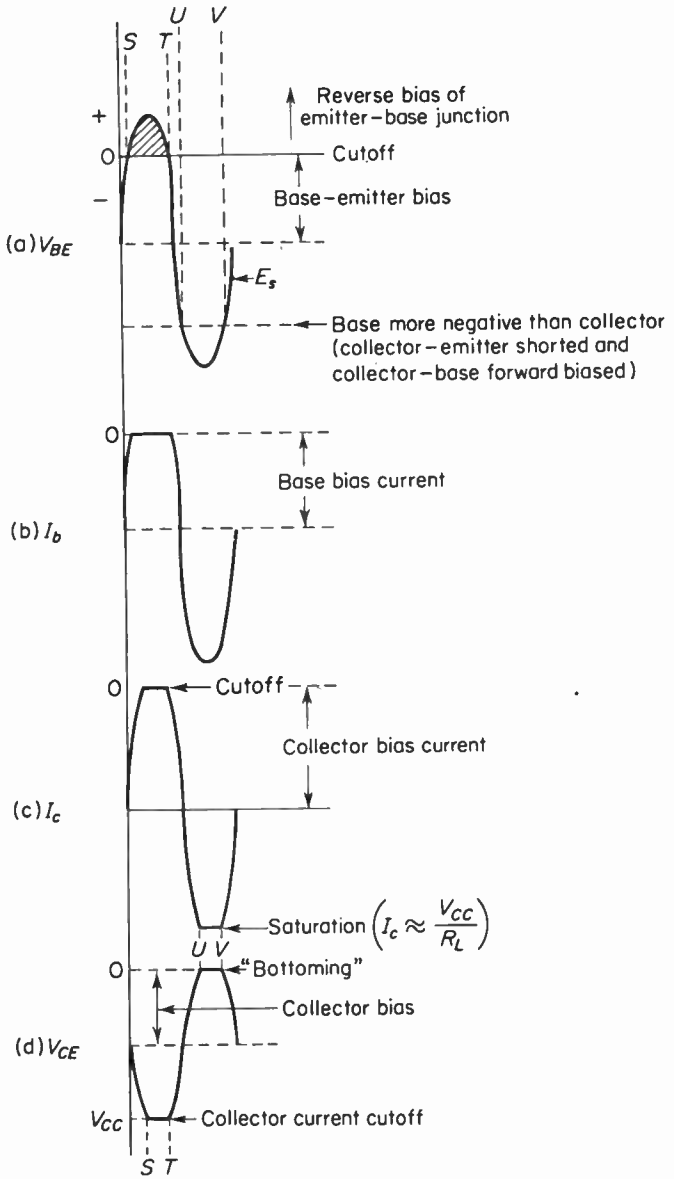



**Fig. 8-27.** Waveforms for circuit of Fig. 8-26 showing how amplitude-limiting is affected by both saturation and cutoff of collector current. The peak positive signal must always exceed the emitter-base bias while the peak negative signal should always exceed the collector-base bias. If the second requirement is not met, a second limiter can clip the negative noise peaks due to polarity inversion of the signal.

means of battery $V_{BB}$ while the collector-base junction is reverse-biased via $V_{CC}$. All leakage currents are neglected (in higher-quality units, $I_{co}$ would be extremely small) and a load resistor $R_L$ of some appropriate value is assumed. A sinewave generator is inserted in the emitter-base circuit and the base and collector currents are shown with the indicated direction.

The associated waveforms are shown in Fig. 8-27. As the input signal, $E_s$, swings more positive, the base current decreases toward zero. In a transistor, when the emitter-base voltage reaches zero, the transistor is cut off. This is shown at points $S$ and $T$. If the input signal should swing any more positive, the emitter-base junction will become reverse-biased and partial rectification will take place. Thus, between points $S$ and $T$, the base current is "clamped" at zero. Since the collector current is given by $I_c = \beta I_b + (1 + \beta)I_{co}$, where $\beta = dI_c/dI_b$, it follows that the collector current will also be zero between points $S$ and $T$ ($I_{co}$ is assumed to be zero). Therefore, the collector current is also "clamped" at zero between these points. Since the collector current is zero, the voltage drop across $R_L$ must be zero. The collector voltage then rises to $V_{CC}$. Any further positive increase in $E_s$ cannot increase the collector voltage, since $V_{CC}$ is the maximum available voltage in the collector circuit. Thus, positive-going *upward* AM is clipped by virtue of collector-current cutoff.

As $E_s$ swings in the negative direction, the emitter-base forward bias is increased. This increases the base and collector currents, and in turn the $R_L$ voltage drop increases. As the load drop increases, the collector voltage ($V_{CE}$) decreases. Point $U$ is reached when the base becomes more negative than the collector, so that the collector-base junction becomes forward-biased. This results in a virtual short circuit across the collector-emitter junction. This is the point of collector-current saturation. Thus, between points $U$ and $V$, the collector current is maximum while the collector voltage has reached zero or the collector has "bottomed." Therefore, negative-going *upward* AM is clipped by virtue of saturation or collector-voltage "bottoming." Should the positive peaks of $E_s$ fail to drive the emitter-base junction to cutoff or should *downward* AM reduce the junction voltage below cutoff, limiting will be impaired. Also, the peak negative carrier amplitude should be capable of driving the base more negative than the emitter, so that on *downward* AM the collector current will remain saturated. This will insure collector-voltage bottoming and effective clipping of noise peaks.

A schematic of a two-stage transistor limiter is shown in Fig. 8-28. The transistors employed are PNP alloy diffusion types (indium and antimony solder are alloyed with germanium under carefully controlled heat). Since transistors are low-impedance devices, power gain is more important than voltage gain. Proper impedance matching is thus important, so transformer coupling is used here to couple $X_1$ to $X_2$. Base bias for the two transistors is provided by means of two voltage dividers. $R_1$-$R_2$ is used for $X_1$ while $R_5$-$R_6$ is

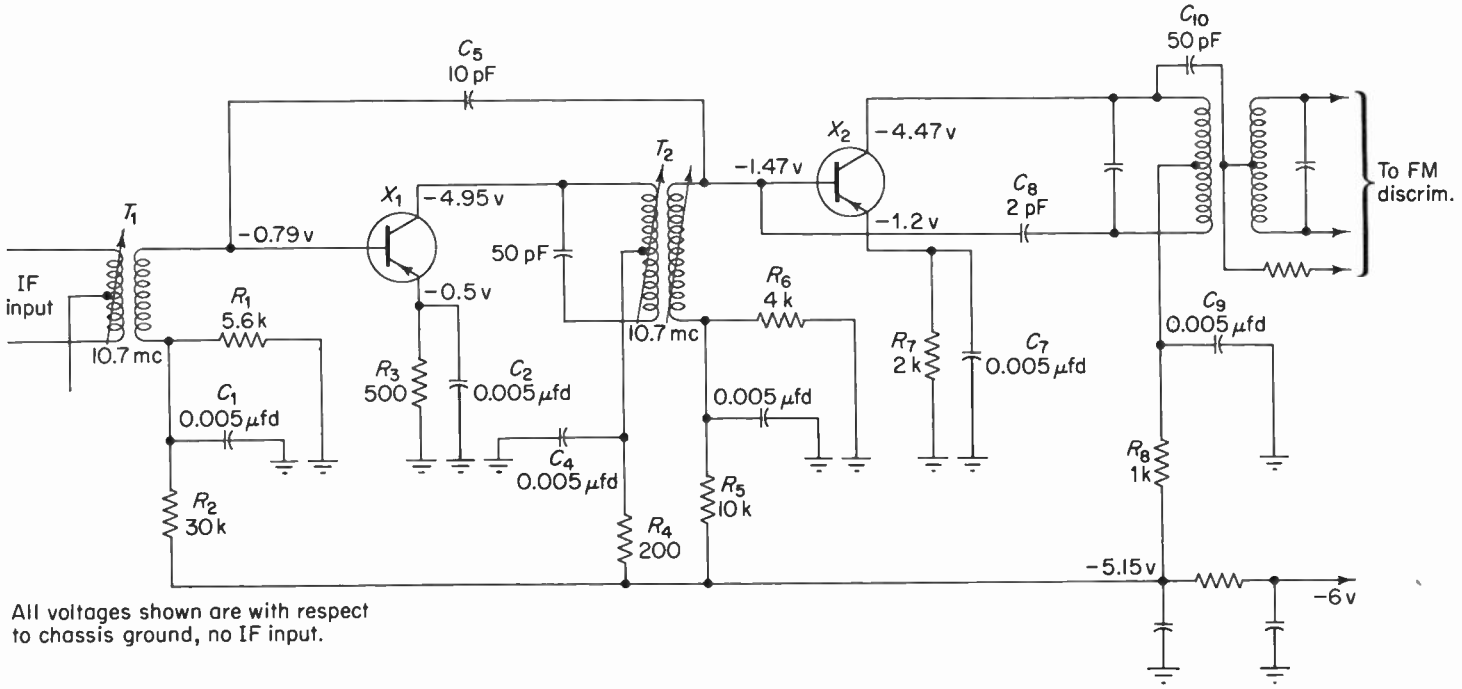



**Fig. 8-28.** A dual transistorized limiter circuit (Courtesy of Realtone Electronics, Inc.)

used to bias $X_2$. The emitters of both transistors contain bypassed resistors for purposes of DC stabilization against changes in temperature, which if not corrected would result in damage to the transistor.

The impedance between collector and base (internal) can result in feedback from output (collector) to input (base). Since the load impedances of both transistors are tuned, they can shift the phase of the feedback voltage and cause regeneration (or degeneration). The internal feedback impedance can be considered as a capacitance of $C_r$ connected between collector and base. If an external capacitor ($C_5$ and $C_8$) is connected as shown, it will feed back to the base a voltage that is equal and opposite in phase to cancel the internal feedback. This is particularly important at the high frequencies used here.

An examination of the schematic will show that the biases used for $X_1$ and $X_2$ are different. The collector-base bias for $X_1$ is 4.16 volts while that of $X_2$ is 3 volts. Thus, $X_2$ will limit more readily than $X_1$, while both will effectively limit on strong signals. $X_1$ may thus act as a conventional IF amplifier on very weak signals and provide sufficient amplification in order to overdrive $X_2$. If the input signal to $X_1$ falls below the level necessary to overdrive $X_2$, limiting will be adversely affected.

## 8-11. Summary

**1.** A limiter removes amplitude modulation from a received carrier and provides an output that is independent of the input amplitude and rate of change of input amplitude.

**2.** A grid-leak limiter limits by means of positive-peak clamping, plate-current cutoff, and plate-current saturation.

**3.** Reducing electrode voltages provides more effective limiting on weak signals.

**4.** In order to reject short-duration impulse noise, the bias time constant should be short in order to keep the peak positive signal clamped to approximately zero grid bias.

**5.** A grid-leak limiter damps the preceding stage because grid current causes power to be taken from the tuned circuit of the driver stage.

**6.** The threshold gain of a single limiter is much lower than the gain of an IF amplifier.

**7.** The limiting sensitivity is greatly improved by the use of two limiters in cascade.

**8.** Two limiters in cascade provide improved performance, since the output of the first stage is already substantially amplitude-limited.

**9.** Bias time constants of the order of 1.5 to 20 $\mu$sec are normally employed.

**10.** A larger time constant raises the gain of the driver stage owing to reduced damping.

**11.** A grid-leak limiter may fail under conditions of rapid *downward* AM because the bias capacitor cannot discharge rapidly enough.

**12.** Impedance-coupled limiters provide high gain and ease of alignment.

**13.** A grid-leak limiter is a fixed-threshold device; below this threshold it behaves like an ordinary IF amplifier.

**14.** The plate-voltage waveform of a limiter with a tuned plate load is sinusoidal because the plate-load impedance is very low at the harmonic frequencies.

**15.** A grid-leak limiter develops bias by means of grid-current rectification on the positive signal peaks. This bias is proportional to the average carrier level.

**16.** The signal-derived bias of a grid-leak limiter can be used for alignment, tuning indicators, squelch circuits, and other applications requiring a negative voltage proportional to the signal level.

**17.** A grid-leak limiter operated under large-signal conditions can be considered as an overdriven amplifier.

**18.** A dynamic limiter limits on the basis of variable damping imposed on a high-$Q$ tuned circuit.

**19.** A dynamic limiter provides no voltage amplification.

**20.** A dynamic limiter can handle almost any amount of *upward* AM.

**21.** The amount of *downward* AM rejection for a dynamic limiter depends upon the extent of tuned-circuit damping under equilibrium conditions.

**22.** The selectivity characteristic of a dynamic limiter is poor, since the passband varies with the degree of damping.

**23.** A dual-diode dynamic limiter provides better AM rejection than a single-diode arrangement. Each diode can clip, respectively, positive and negative noise pulses.

**24.** Dynamic limiters provide an AGC voltage proportional to the average signal level.

**25.** Dynamic limiters are variable-threshold devices and can limit down to small signals where the diode no longer behaves like a linear rectifier.

**26.** Crystal diodes are often used as dynamic limiters and can be mounted in the same can as the tuned circuit to minimize IF radiation.

**27.** Dynamic limiters perform very well at high input levels (see Fig. 8-23).

**28.** Transistorized limiters reject AM by virtue of collector-current cutoff and saturation (overdriven amplifier).

**29.** In order to realize proper impedance match for optimum power gain, transistor limiters are often transformer-coupled.

**30.** Transistors provide good sensitivity, low current drain, and compactness.

## REFERENCES


1. Tibbs, E., and Johnstone, G. G.: *Frequency Modulation Engineering*. John Wiley & Sons, Inc., New York, 1956.
2. Loughlin, B. D.: "Performance Characteristics of FM Detector Systems," *Tele-Tech*, January 1948.
3. Freeland, E. C.: "F-M Receiver Design Problems," *Electronics*, January 1949.
4. Baghdady, E. J.: "FM Demodulator Time-Constant Requirements for Interference Rejection," *Proc. IRE*, February 1958.
5. Johnstone, G. G.: "Limiters and Discriminators for FM Receivers," *Wireless World*, August 1957.


## APPENDIX 8-1

Derivation to determine the per cent of AM rejection of a dynamic limiter. The following is based on rejection of downward AM only since upward AM is not generally a problem with dynamic limiters or ratio detectors.

$e_{nl}$ = output, no limiting = $iQX$

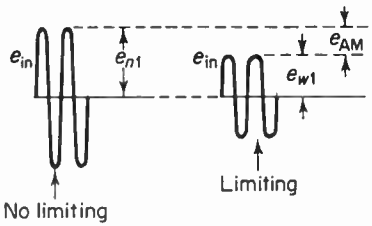
$e_{wl}$ = output, with limiting

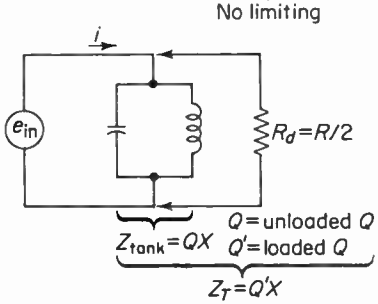
$= iZ_t = iR_dQX/(QX + R_d)$

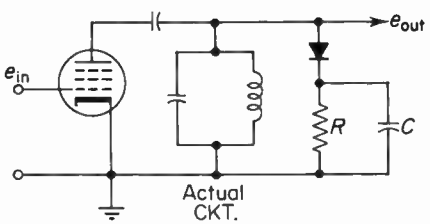
$e_{AM}$ = amount of $AM = e_{nl} - e_{wl}$

% of $e_{AM}$ rejected = $M = (e_{nl} - e_{wl}/e_{nl})\,100\%$

$$1.\quad M = \frac{iQX - \dfrac{iQXR_d}{QX + R_d}}{iQX} = \frac{QX - \dfrac{QXR_d}{QX + R_d}}{QX}$$



2. $M = 1 - \dfrac{R_d}{QX + R_d}$ divide each term on right side by $iQX$.

3. $M = \dfrac{QX + R_d - R_d}{QX + R_d} = \dfrac{QX}{QX + R_d}$

4. $Z_{\text{tank}}$ without damping $= QX$.

5. Circuit $Q$ with damping $= Q' = \dfrac{Z_T}{X} = \dfrac{1}{X}\left[\dfrac{QXR_d}{QX + R_d}\right]$

6. $Q' = \dfrac{QXR_d}{X(QX + R_d)}$

7. $Q'X = \dfrac{QXR_d}{QX + R_d}$ multiply both sides by $X$.

8. $M = \dfrac{QX - Q'X}{QX}$ divide each term on right-hand side by $X$.

9. $M = \dfrac{Q - Q'}{Q} = 1 - \dfrac{Q'}{Q}$.

Now, if $R_d = R/2$ and is made equal to one-eighth of the dynamic tuned-circuit impedance, then $R_d = QX/8$ and $QX = 8R_d$. The per cent of downward AM rejected is

$$M = \frac{QX}{QX + R_d} = \frac{8R_d}{8R_d + R_d} = \frac{8}{9} \times 100 \text{ per cent} \cong 89 \text{ per cent.}$$

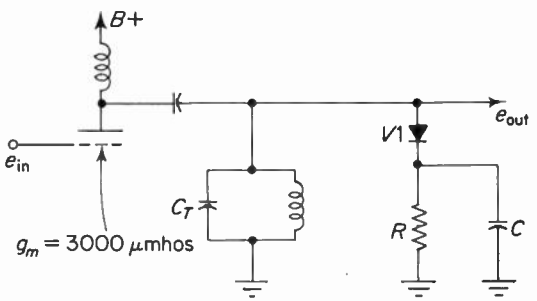
The accompanying drawing shows a sample design, where $C_T = 30$ pfd, $f = 10.7$ mhz, $Q_{\text{coil}} = 100$, $g_m = 3000$ $\mu$mhos, $V1$ is an ideal diode, and $r_c = 0$ ohms.

1. $QX$ = dynamic impedance $= (10^2 \times 10^{12} \times 10^{-6})/(6.28 \times 10.7 \times 30) \cong 50{,}000$ ohms.

2. Undamped gain $= A = g_m QX = 3000 \times 10^{-6} \times 50 \times 10^3 \cong 150$.

3. To reject 90 per cent downward AM, set $M = 0.9$ or $0.9 = QX/(QX + R_d)$. Then, $R_d = 50/9 \cong 5500$ ohms.

4. Damped gain $= g_m QXR_d/(QX + R_d) = (3000 \times 10^{-6} \times 50 \times 10^3 \times 5.5 \times 10^3)/55{,}500 \cong 15$.



5. Since $R_d = R/2$ or $R = 2R_d$, $R = 11{,}000$ ohms.

6. If $RC$ is to be $10T$, where $T = 1/f_{\text{lowest}}$, then $RC \cong 0.15$ sec. Thus, $C = 0.15/R = 0.15/11{,}000 \cong 13.5$ $\mu$fd.

7. The circuit can now reject 90 per cent downward AM, or the change in output due to a 90 per cent change in the input is only 10 per cent.

8. The "AM reduction factor" is equal to

$$\frac{\text{per cent change in the input}}{\text{per cent change in the output}} = \frac{90 \text{ per cent}}{10 \text{ per cent}} = 9, \qquad \text{or } 9{:}1.$$

9. This corresponds to $20 \log 9 \cong 19$ db.

# 9

# FM DETECTORS (SLOPE TYPES)

## 9-1. Basic requirements

Detection or demodulation is defined as the process of recovering the intelligence from a high-frequency wave. In the case of amplitude-modulation receivers, the detectors most often used are diodes. Any nonlinear circuit element can be used to restore the intelligence (audio signal, for example) while the load circuit, acting as a low-pass filter, removes the high-frequency component from the demodulated output. A circuit that responds to envelope variations of the carrier would fulfill the basic requirements for an AM detector. A typical arrangement is shown in Fig. 9-1(a). The waveform shown in Fig. 9-1(c) is the input voltage to the detector and consists of the modulated carrier. This type of circuit is referred to as a peak or envelope detector, since the output voltage essentially follows the peaks (or envelope) of the applied carrier. The reader may wonder at this point why AM detectors are mentioned at all in a book concerning frequency modulation. We shall show, however, in discussing FM detectors, that peak detection is an integral part of the FM detection process.

A discussion of the peak rectifier would not be complete without a reference to the load circuit consisting of $R$ and $C$ in Fig. 9-1(a). The time constant for this circuit is important, since it determines how well the voltage across the capacitor follows the envelope variations. If the discharge time constant

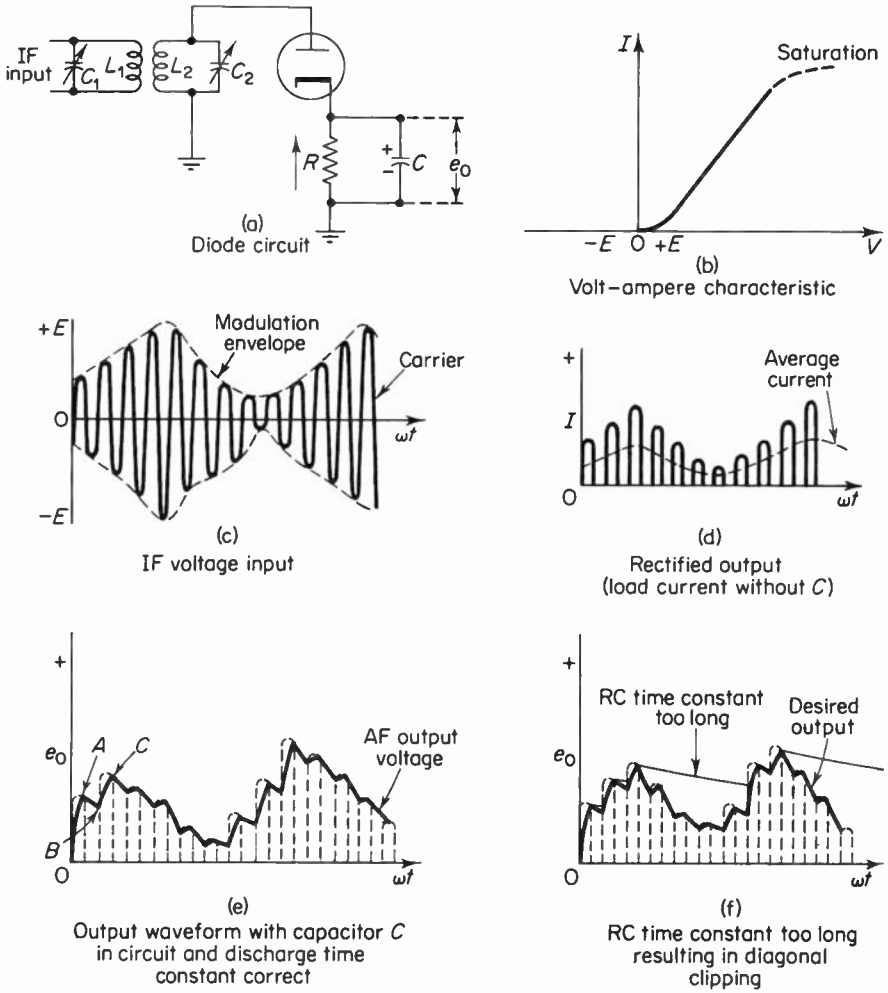



**Fig. 9-1.** Basic detection process.

is very long compared to the period $T$ of the IF carrier, the load voltage will not respond to carrier-frequency variations. The circuit will then maintain the load voltage when the diode is not conducting. This is desirable, since the average output will remain high. However, the discharge time constant should be chosen such that the capacitor voltage will change at the envelope rate, which in turn corresponds to audio. This requires that the load be "quick-acting" with respect to changes of the modulation envelope. A time constant that is short compared to the period of the highest expected modulation frequency would be necessary. A 100-microsecond discharge time constant is a reasonable compromise for AM-type detectors.

If the discharge time constant is too long, the capacitor voltage will not be able to follow rapid changes in the envelope. A form of distortion results, which is known as *diagonal clipping*. Figure 9-1(f) shows how this takes place. During the interval $A$–$B$, the carrier is dropping rapidly while the capacitor discharges too slowly through $R$. Consequently, the load voltage does not follow the peaks between $A$ and $B$.

Since the diode conducts only when the applied voltage exceeds the standing bias on the capacitor, the conduction angle is short. If $C$ is omitted, the conduction angle is essentially a half-cycle, the average load voltage is reduced, and the average power taken from the tuned circuit is less. This will result in an increase of sensitivity, selectivity, and circuit $Q$, since the diode, together with its load, imposes less damping on the tuned circuit.

A diode detector also distorts very badly on weak signals as a result of a bend at the lower end of its volt-ampere characteristic [see Fig. 9-1(b)]. This means that considerable amplification is required prior to detection.

## 9-2. Slope detection†

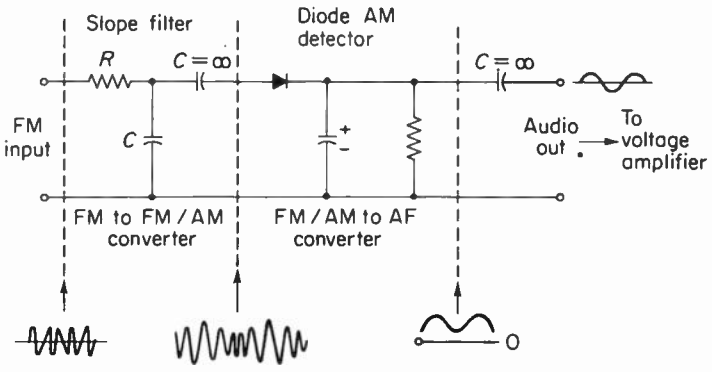
It has already been pointed out that the FM carrier deviates as a function of the amplitude of the modulating signal. In order to recover the information from the carrier, we must translate frequency variations into amplitude variations. Thus, the basic requirement of any FM detector is to provide an audio output voltage that is *linearly* proportional to the instantaneous frequency deviation of the carrier. It is also necessary that the rate of change of these variations be at the frequency of the audio signal.

The simplest approach to FM demodulation is by means of a network as shown in Fig. 9-2. It is simply a low-pass filter with a frequency versus amplitude characteristic as shown in Fig. 9-2(b). If the input FM carrier is centered at point $A$, the output voltage will vary, since the reactance changes inversely with frequency. The output waveform is both amplitude- and frequency-modulated. Since the peak-to-peak amplitude of the modulation envelope is proportional to the instantaneous frequency deviation, the envelope *amplitude* becomes a measure of the deviation to one side of the center frequency. The output voltage is then applied to an AM detector, which is insensitive to FM but produces a rectified output that varies linearly with frequency deviation.

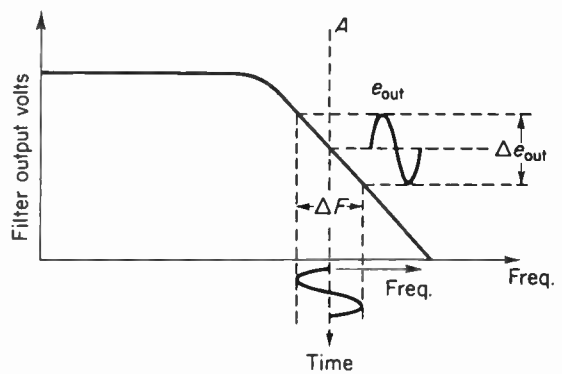
Since the input carrier deviates along the slope of the characteristic, it is termed a "slope detector." Its main drawbacks are low sensitivity (volts/cycle) and no protection at all against residual AM.

A somewhat improved form is shown in Fig. 9-3, together with its response curve. This arrangement provides somewhat more sensitivity than the

†A slope detector is sometimes referred to as a linear time-rate demodulator.




(a)
Circuit and pertinent waveforms




(b)
Filter response. Conversion sensitivity = $\frac{\Delta e_{out}}{\Delta F}$

**Fig. 9-2.** Elementary slope detector and associated response (a basic approach to the detection of an FM carrier). The conversion sensitivity of this simple filter is inadequate for most applications.

previous circuit because its slope is steeper. The FM carrier is centered at point *B*. The region between *A* and *C* is required to be reasonably linear over the deviation range of the carrier. This range for television is 25kHz on either side of the rest position. Additional range is provided by the fact that the tuned-circuit response does not change abruptly below *A* or above *C*. As the carrier deviates from its mean position, the output voltage varies in a reasonably linear manner, provided the carrier does not swing outside the linear portion of the "slope." It should be noted that the output of this circuit is varying in both amplitude *and* frequency, so that the additional process of AM detection must be performed. This, of course, is where the AM detector

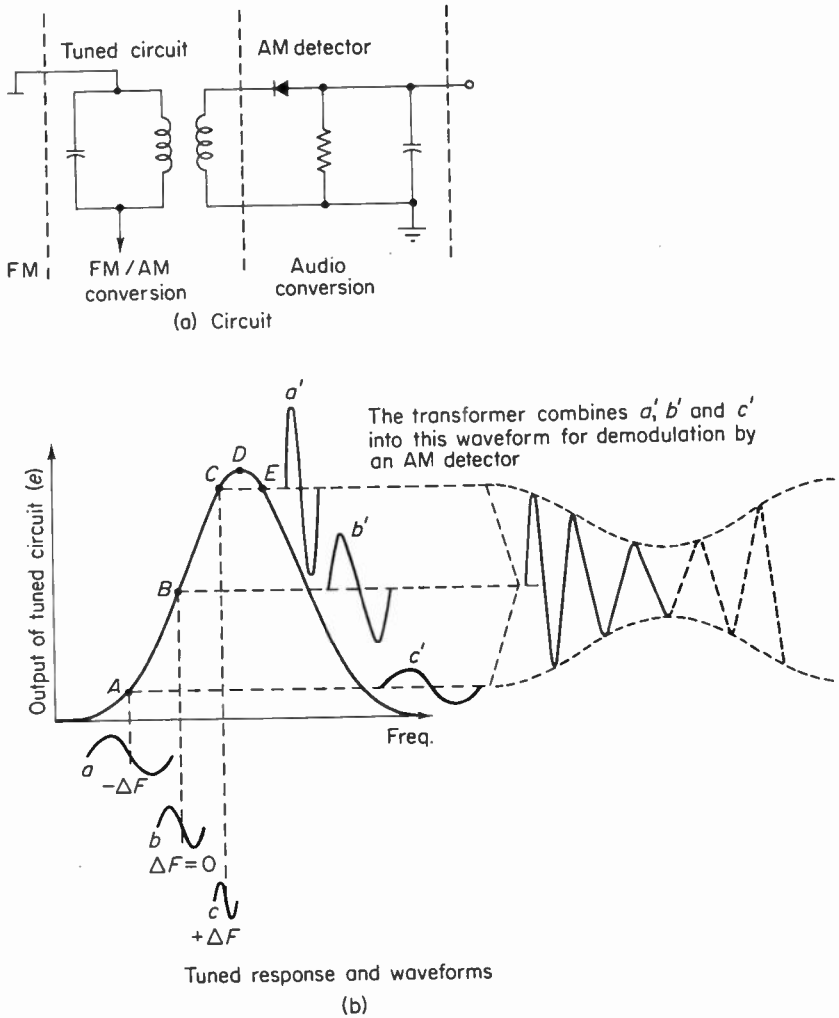



**Fig. 9-3.** Slope detection using a tuned circuit. Waveforms *a*, *b* and *c* are constant amplitude, instantaneous frequency carriers. Waveforms *a'*, *b'* and *c'* correspond to instantaneous outputs of tuned circuit for frequencies *a*, *b* and *c* respectively.

comes into use. An AM detector will not respond to frequency variations but will respond to carrier amplitude changes. A diode detector may be used to demodulate the output by responding to the signal peaks and ignoring the frequency deviations. Since the peaks are varying at the original audio-frequency rate, the modulation recovery is complete. The FM to AM to audio translation process depicted is characteristic of FM detectors.

There are a number of reasons why this type of detector cannot be used in wide-band applications. One limitation concerns the length of the linear

portion of the tuned response, *A–C*. It will not handle large carrier excursions (75 kHz on either side of center) without extending into the nonlinear region, which in turn would cause severe harmonic distortion. Also, the selectivity of the average FM receiver is not great enough to prevent a carrier from falling on position *F*. Should this happen, interference with the desired carrier at position *B* would result. The circuit shown is then limited to narrow-band systems such as television, mobile radio, military, and so on.

The possibility of a spurious carrier appearing at point *F* in an intercarrier television receiver is very remote, since the sound intercarrier IF (4.5 MHz) is obtained by beating the sound and picture frequencies in the second (video) detector. Also, the sound IF response in a television receiver is narrow compared to the IF response of a broadcast tuner because of the lower IF used and the smaller carrier swing employed (25 kHz on either side of rest position).

Another consideration concerns the centering of the IF on the linear portion of the detector response curve for distortion-free audio. Should an FM receiver be slightly mistuned or should the local oscillator drift, severe distortion would result, since the deviation would swing the carrier into the nonlinear region. In a television receiver the sound intercarrier IF is fixed, since the 4.5-MHz carrier spacing is established at the transmitter and retained at the receiver. The IF is therefore accurately centered on the detector bandpass regardless of oscillator drift or slight misadjustment of the fine tuning control on the television receiver.

Up to this point, the discussion has centered on recovery of the modulation without mention of the possibility of amplitude variations caused by noise and other disturbances not related to the original signal. An increase in carrier amplitude will result in an increase of output from the tuned circuit. That is, if the driver stage is a class $A_1$ IF amplifier, any increase in input (noise, for example) will simply produce a proportional output. This will be demodulated along with the frequency deviations and appear as audible noise. It is therefore necessary to perform noise limiting when employing any type of slope detector.

## 9-3. Travis detector

It was mentioned that slope detectors are limited to narrow-deviation systems. This is because the linear range of the slope of the tuned circuit is small, and large carrier swings would result in severe harmonic distortion. The *Travis* detector, to be described, operates on a "push-pull" basis. From elementary power-amplifier theory the reader will recall that this type of circuit eliminates even-ordered harmonic distortion. In addition to reduced distortion, it will be seen that this detector has a much greater linear working range. The Travis detector is employed in mobile radio (narrow-band) com-

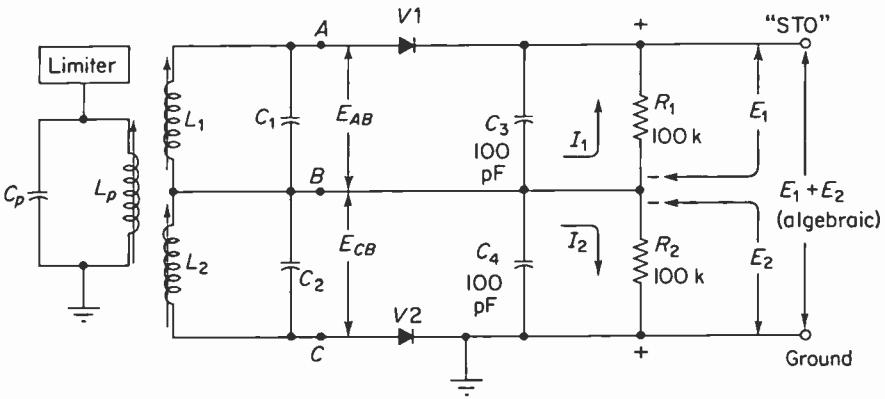



**Fig. 9-4.** Generalized form of Travis Detector (primary may or may not be tuned).

munication receivers, and an understanding of it will give the reader valuable background for the analysis of other types of detectors to follow.

A representative circuit is shown in Fig. 9-4, indicating a total of three tuned circuits following the limiter. Some versions show the primary as being untuned, but the basic operation is the same for both cases. If $V_1$-$R_1$-$C_3$ and $V_2$-$R_2$-$C_4$ are considered as separate circuits, then each represents a diode peak detector, each being driven by its respective tuned circuit ($L_1$-$C_1$ and $L_2$-$C_2$). The primary circuit ($L_p$-$C_p$) is broadly tuned to accept the full carrier swing. It is assumed that the carrier will swing a total of 150 kHz, which means the primary tuned bandwidth is of the order of 200 kHz. $L_1$ and $L_2$ are inductively coupled to $L_p$, but there is essentially no magnetic coupling between $L_1$ and $L_2$. $L_1$ and $L_2$ are considered as a center-tapped secondary, each coil and its associated capacitor being tuned to a different resonant frequency. The detection process that takes place here results in the same type of translation that occurs with the slope detector. The transformer tuned circuit converts a variable-frequency, constant-amplitude carrier into a variable-amplitude, variable-frequency output, which is then used to drive the diode envelope detectors. The differential voltage output at "STO" represents the modulation or audio.

Figure 9-5 shows the response of each tuned circuit with each inverted relative to the other. This inversion is necessary in order to obtain a differential output. $L_1$-$C_1$ resonates at $F_1$ and $L_2$-$C_2$ resonates at $F_2$, the center frequency appearing midway between both. The composite characteristic curve is obtained by combining both curves between the limits $F_L$ and $F_U$. The resultant curve (joining the points $U$-$V$-$W$-$O$-$X$-$Y$-$Z$) is referred to as an "S" curve and is typical of *all* FM detectors. This characteristic is shown separately in Fig. 9-6. If the carrier deviates instantaneously beyond points $W$ or $X$, severe distortion would result, since the S curve is no longer

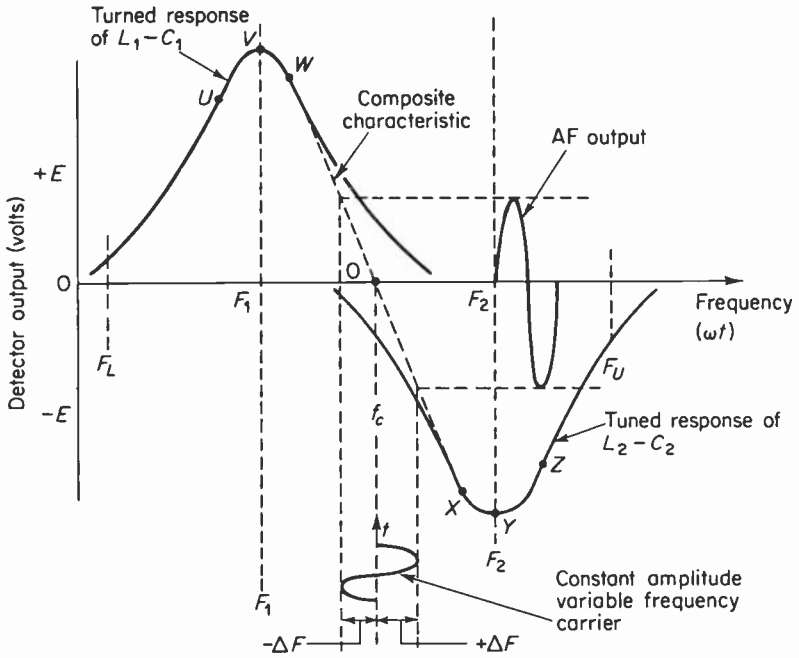



**Fig. 9-5.** Tuned circuit response of $I_1 - C_1$ and $L_2 - G_2$. The composite curve (shown dotted) is obtained by algebraically adding the instantaneous ordinates at points along the frequency abscissa. The length of curve *W-O-X* is considered to be linear.

linear. For a zero carrier deviation, there should be no audio output. Under this condition, the instantaneous potential from *A* to *B* must be equal to the potential from *B* to *C*. This is based on the tuned response shown in Fig. 9-5. At this instant the carrier frequency is $f_c$, and since the driving voltages for the diodes are equal and opposite in phase, the instantaneous potential across $R_1$ will be equal to the potential across $R_2$. However, since the audio output is taken from across both resistors, the differential output must be zero.

The discharge time constant of both $R_1$-$C_3$ and $R_2$-$C_4$ is 10 microseconds. This is long compared to the period of a 10.7-MHz IF. Therefore, the potentials across $R_1$ and $R_2$ are approximately equal to the peak driving voltages at *A–B* and *B–C*. The forward voltage drops across the diodes are small enough to be ignored. It should be noted that a DC voltage measured across either load resistor indicates the presence of a carrier and therefore could be utilized in troubleshooting.

Should the carrier deviate upward, the output of $L_1$-$C_1$ will be less than that of $L_2$-$C_2$, which means the driving voltage for $V_2$ will exceed that for $V_1$. The composite curve of Fig. 9-5 indicates that there will be an instantaneous negative output from STO to ground. This comes about because the recti-

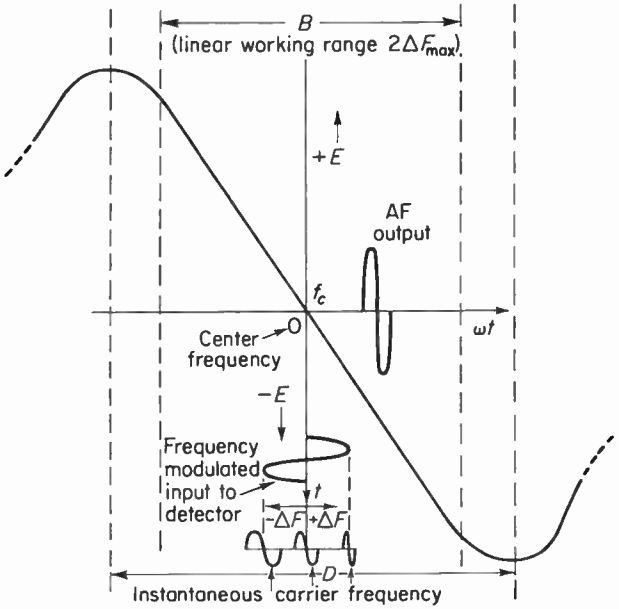



**Fig. 9-6.** Final response ("*S*" curve) of output of both peak detectors. The linear bandwidth *B* should exceed twice the system deviation (at least 200 kHz for commercial FM broadcast applications). The frequency displacement *D* is referred to as the "peak separation."

fied voltage across $R_1$ in this case is smaller than that across $R_2$, since $E_{AB}$ is smaller than $E_{CB}$. If the carrier frequency should shift downward from center position, then the output of $L_2$-$C_2$ will be less than that provided by $L_1$-$C_1$. The driving voltage for $V_1$ will exceed the driving voltage for $V_2$ and $C_3$ will charge to a higher peak voltage than $C_4$. The rectified voltage across $R_1$ is higher than that across $R_2$ and the differential voltage at STO is positive with respect to ground. Thus, a change in the instantaneous carrier frequency results in a linear change in output voltage. The *magnitude* of this output voltage is proportional to the frequency deviation, and the *rate* at which this output voltage varies is equal to the modulation frequency.

Up to this point, we have assumed that the carrier amplitude has been constant and only its frequency has been varying with modulation. However, suppose that the voltage across $L_p$ is not varying in frequency but that its amplitude is varying at a 1-kHz rate due owing to the presence of *noise* at the input to the receiver. A further assumption is that the receiver is accurately tuned to the center of the detector bandpass. Under these conditions, both peak rectifiers will produce equal output voltages across their respective loads. Since both diodes and their loads comprise AM detectors, they will both respond to these 1-kHz "audio" variations. However, since $E_1$ is equal to $E_2$, the differential output is zero. If the receiver is manually detuned, the

operating point is shifted off-center and there will be an audible noise output. The reason is that $E_{AB}$ will not be equal to $E_{CB}$, and the result will be a non-zero differential output voltage. This is also an indication that the inter-station noise using this type of detector will be quite high. This is true whether there is a carrier present or not, since receiver noise is perfectly capable of providing an audible hiss in the output.

Assume now that the detector is again accurately center-tuned but that this time the carrier is both frequency and amplitude-modulated. An examination of Fig. 9-7 shows that since the percentage modulation of $E_{AB}$ is greater

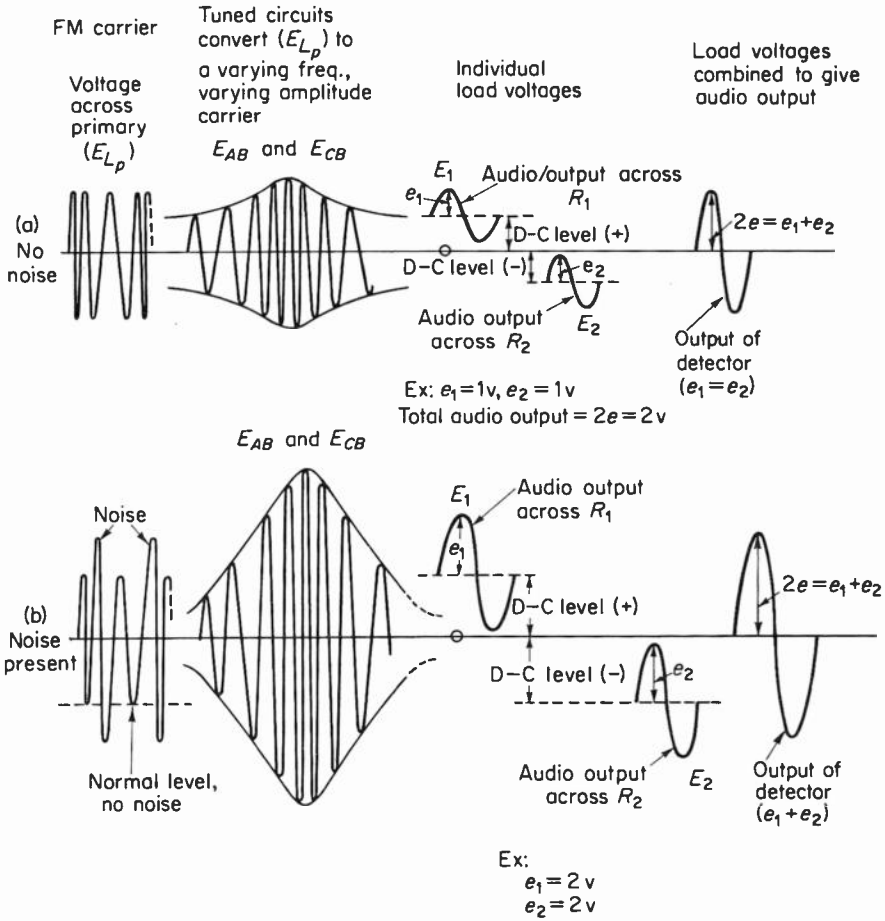



**Fig. 9-7.** Voltage waveforms indicating FM-AM-audio-conversion and effect of noise on output of detector. This indicates the need for pre-limiting when using a Travis detector. The detector output ($e_1 + e_2$) for case (b) includes the noise component as well as the signal component.

for case (b) where there is noise present than for case (a) where there is no noise present, the variable component of $E_1$ will be increased from 1 volt to 2 volts. Since the variable components are in phase, the audible output for case (b) is 4 volts total with half of this output due solely to noise. This means that a Travis detector, like the slope detector, requires prelimiting since it responds readily to noise. There is available at the detector output a DC voltage that is proportional to the amount of detuning. This voltage could be used for (1) driving a tuning indicator, (2) AFC, and (3) alignment and troubleshooting.

## 9-4. Summary

**1.** A diode is a nonlinear circuit element.

**2.** Demodulation involves recovering the intelligence from a high-frequency carrier.

**3.** Diode detectors are universally used in AM broadcast receivers.

**4.** All FM demodulators must *linearly* convert carrier-frequency variations into audio frequencies, where the instantaneous frequency deviation determines the *amplitude* of the resultant audio while the rate at which the carrier deviations take place determines the frequency of the resultant audio.

**5.** The slope detector demodulates the FM carrier by means of a tuned circuit whereby the carrier-frequency deviations result in different IF output voltages. These amplitude variations are then converted into audio-frequency variations by a diode peak rectifier.

**6.** The slope detector is used in some commercial television sound strips.

**7.** The slope detector requires prelimiting in order to function satisfactorily.

**8.** A Travis detector demodulates an FM carrier by virtue of tuned circuits that convert the frequency variations into amplitude variations, followed by two peak rectifiers that convert the amplitude variations into audio-frequency variations.

**9.** The Travis detector requires prelimiting to perform satisfactorily.

**10.** The Travis detector can reject noise only when the circuit is accurately center-tuned or when the disturbance occurs when the instantaneous carrier is at center-frequency.

**11.** The linear bandwidth of the slope detector is less than that of a Travis detector. A Travis detector is more difficult to align.

## REFERENCES

1. Hund, A.: *Frequency Modulation.* McGraw-Hill Book Company, New York, 1942.
2. Department of the Army Technical Manual, TM11-668, Superintendent of Documents, Washington, D.C., 1952.

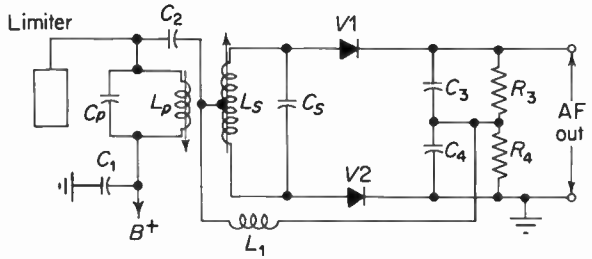
3. Johnstone, G. G.: "Limiters and Discriminators for FM Receivers," *Wireless World*, January 1957.
4. Black, H. S.: *Modulation Theory*. D. Van Nostrand Co., Inc., Princeton, N.J., 1953.
5. Scott, R. F.: "New Sound Detector System for TV," *Radio-Electronics*, February 1959.
6. Burns, C.: "New G-E FM Detector," *Radio & TV News*, June 1958.
7. Dome, R. B.: "Inexpensive Sound for Television Receivers," *Electronics*, February 1959.

# 10

# PHASE-SHIFT TYPE DETECTORS

## 10-1. Foster-Seeley discriminator

A widely used FM demodulator is shown in Fig. 10-1. This circuit is similar to the Travis detector in many respects, but the method used to obtain the diode driving voltages is entirely different. An important advantage of the Foster-Seeley over the Travis detector is its ease of alignment.

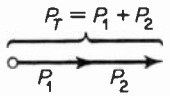



$C_1 = 0.01\,\mu fd$
$C_2 = 33$ pF
$C_p = 33$ pF (does not include "strays" which might amount to 15 $\mu\mu fd$)
$C_s = 33$ pF ( " " " " " " " " " " " " )
$C_3 = 100$ pF
$C_4 = 100$ pF
$R_3 = 100$ k
$R_4 = 100$ k
$L_p = 2.5\,\mu h$, $L_s = 4.5\,\mu h$ } Slug-tuned, ferrite cores
$L_1 = 1$ mh

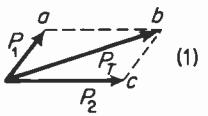
**Fig. 10-1.** Foster-Seeley Detector. Values shown are typical for an intermediate frequency of 10.7 MHz.

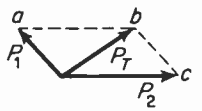
In order to better understand the operation of the phase-shift detector, we shall briefly review phasors. The discussion that follows should call up enough background information so that the reader may accurately follow subsequent discussions.

If two quantities are *in phase* with one another, their vector sum can be obtained by simple addition of their respective magnitudes as shown in Fig. 10-2(a). If, however, they are *out of phase*, their resultant sum can be gotten by means of a *parallelogram of forces* method as shown in Fig. 10-2(b); an example is shown in Fig. 10-2(c). It is assumed that a voltage has been induced *in* coil $L$ with a $Q$ of 50. The induced voltage ($e_i$) drives a current $i$ around the loop that includes the coil (with its small resistance, $R$) and capacitor $C$. Since the secondary circuit is *series* resonant, the loop current develops three voltage drops, labeled $e_r$, $e_i$, and $e_c$. Since the reactances are equal, their respective voltage drops are equal. The sum of all three must equal $e_i$. At resonance, the reactive drops cancel and the circuit is purely

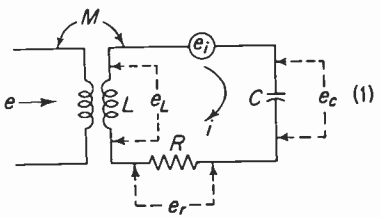



(a) Obtaining the resultant $P_T$ when $P_1$ and $P_2$ are in phase.



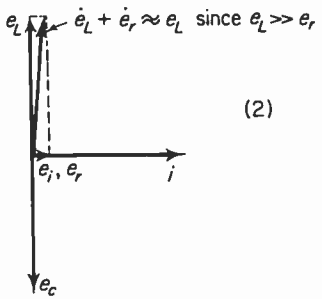




(b) Two examples of obtaining the resultant $P_T$ when $P_1$ and $P_2$ are out of phase. (Line a–b should be drawn parallel to $P_2$. Line b–c should be drawn parallel to $P_1$).




Conditions at resonance:

1. $X_L = X_c$
2. $e_r = e_i = iR$
3. $e_i = e_r + e_L - e_c = e_r$
4. $e_c = e_L = iX_L = \frac{e_i}{R} X_L = Qe_i$
5. $Q > 10$




Resonant phasor

(c)

Secondary circuit and associated diagram for condition of resonance.

**Fig. 10-2.** Basic phasors.

resistive. The induced voltage generator thus "sees" only a resistor $R$, and the voltage across the resistor is $e_i$. Since this is a high-$Q$ circuit, the voltage across $C$ is approximately equal to the voltage across the coil, since the coil resistance is so much smaller than the coil reactance. The in-phase component of coil voltage ($e_r$) is shown to be very much smaller than the reactive component $e_l$. There is no convenient way of measuring $e_i$ directly, since a voltage measurement made across $C$ will indicate $e_c$ and not $e_i$.

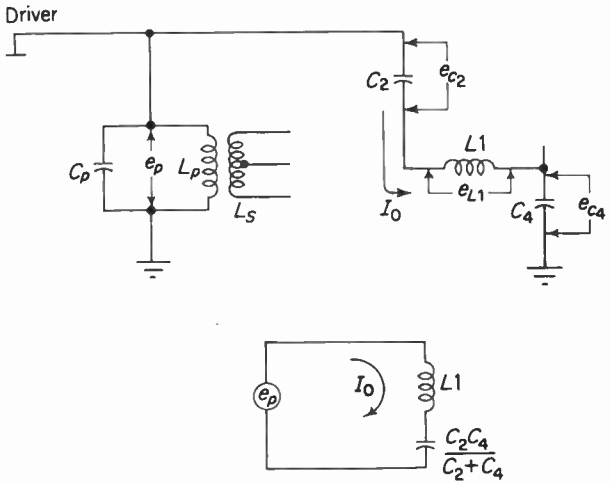
It has been mentioned that the purpose of the detector is to convert carrier-frequency deviations in a *linear manner* into audio. The FM-AM-audio conversion in the Foster-Seeley circuit is also accomplished in two distinct steps. It will be shown that the transformer converts, by means of phase shifts, a varying-frequency carrier into an AM signal, which in turn is AM-demodulated by means of a pair of diodes. The Foster-Seeley discriminator will now be considered in detail.

## 10-2. Reference voltage

An inspection of Fig. 10-1† shows that $C_1$ effectively places the bottom end of the primary tuned circuit at IF ground potential. The top of the primary tank is connected (via $C_2$) to the center-tap of the secondary coil $L_s$. The reactance of $C_4$ at the same frequency is approximately 1500 ohms. The reactance of $L_1$ is approximately 67,000 ohms. These three components are effectively connected in series across the primary winding as shown in the equivalent circuit of Fig. 10-3(a). Since the inductive reactance of $L_1$ is so much larger than the reactance of either capacitor, the voltage across $L_1$ (by voltage division) is almost equal to the primary voltage $e_p$. Since $L_1$ is untuned, the voltage across it is in phase with $e_p$. This is shown in the phasor diagram, Fig. 10-3(b). It is therefore assumed that a voltage appears across $L_1$ which is essentially equal to and in phase with $e_p$. This voltage ($e_{L_1}$) is absolutely essential to the operation of the detector and is referred to as the reference voltage.

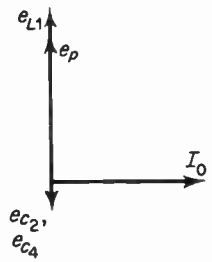
Figure 10-4 shows the equivalent circuit for the primary and secondary tuned circuits. Both circuits are resonant at 10.7 MHz and are driven by IF voltage $e_p$. It is initially assumed that this voltage is a constant-amplitude, variable-frequency "carrier." Three distinct conditions will be investigated. If the carrier is at center frequency, the amplitude of the modulating voltage is zero. This corresponds to the condition of $\Delta F_c = 0$. A second condition will be where the carrier frequency deviates above 10.7 MHz owing to a positive change in the modulating voltage. This condition corresponds to $+\Delta F_c$. A third condition will be where the carrier deviates below 10.7 MHz and will be referred to as $-\Delta F_c$. If the detector frequency versus modulation output response is linear over the range of at least 2 $\Delta F$, then the output will be un-

†This is only one form of the Foster-Seeley detector. Others will be discussed.




The equivalent circuits showing how coil $L1$ is effectively placed across the primary of the detector transformer

(a)




The phasor diagram showing how $e_{L1}$ is in phase with and of approximately the same magnitude as $e_p$

(b)

**Fig. 10-3.** The derivation of the reference voltage $e_{L1}$.

distorted. It will be shown that a linear detector bandwidth in excess of 200 kHz is necessary because of such factors as mistuning, oscillator drift, high modulation levels, and so on.

## 10-3. Diode driving voltages

It is assumed that $C_2$ and $C_4$ in Fig. 10-4 are both zero reactances at 10.7 mhz. This places $L_1$ across the primary winding as shown. The secondary loop current $i_s$ develops a voltage drop $e_c$ across $C_s$, but this voltage is *not* used to drive the diodes, as will be shown.

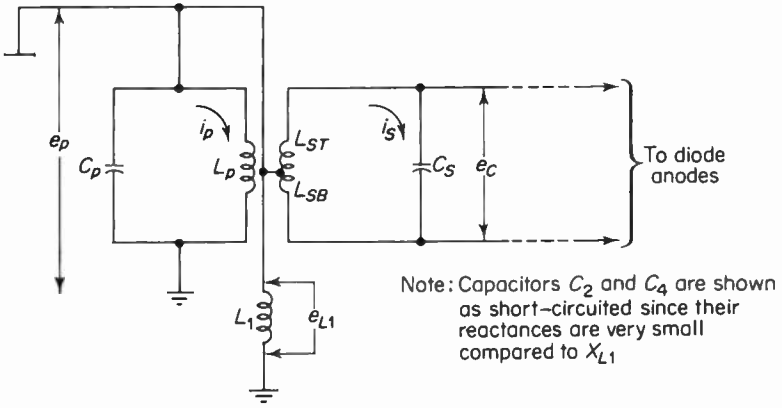



**Fig. 10-4.** IF equivalent of tuned circuits. It is assumed that $Q_p = Q_s = 50$.

Figure 10-5 shows the equivalent circuit for deriving the upper and lower diode driving voltages, $e_{ud}$ and $e_{ld}$. $L_s$ is center-tapped and the upper and lower halves are labeled accordingly. An instantaneous direction of current flow has been assumed such that the bottom of the secondary winding is positive with respect to the top end. If the instantaneous potential from bottom to center-tap is designated $e_t/2$ (where $e_t$ is a reactive *drop* and not the induced voltage), then the voltage from the top of the secondary to the center-tap must be 180° out of phase with $e_t/2$. The upper half-secondary voltage drop

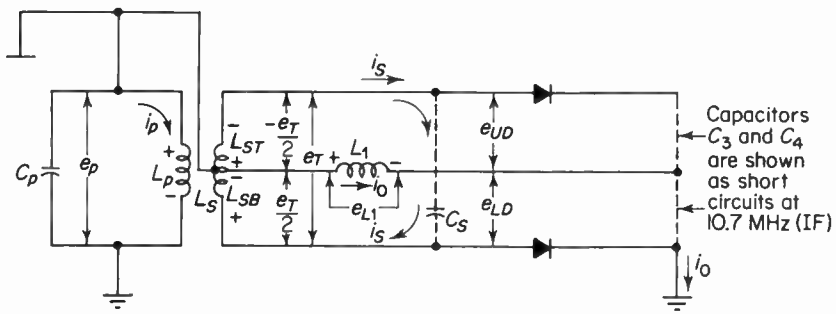



**Fig. 10-5.** Equivalent circuit for deriving diode driving voltages.

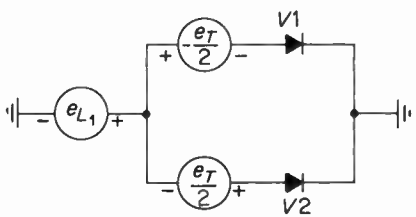



**Fig. 10-6.** Simplified IF equivalent of diode driving circuit. Voltages $e_t/2$, $e_{L1}$ and $-e_t/2$ have been replaced by generators having instantaneous polarities as shown.

is labeled $-e_t/2$. Since $L_1$ is essentially across the primary winding, a constant-amplitude IF voltage appears across it and is labeled $e_{L_1}$. The simplified driving circuit for both diodes is shown in Fig. 10-6 with all necessary voltages replaced by equivalent generators.

## 10-4. Basic resonant diagram

The carrier is at rest and the secondary tuned circuit is resonant. Figure 10-7 shows $i_p$ lagging the primary voltage $e_p$ and shows that $L_1$ is essentially in parallel with the primary circuit. Therefore, $e_{L_1}$ is equal to and

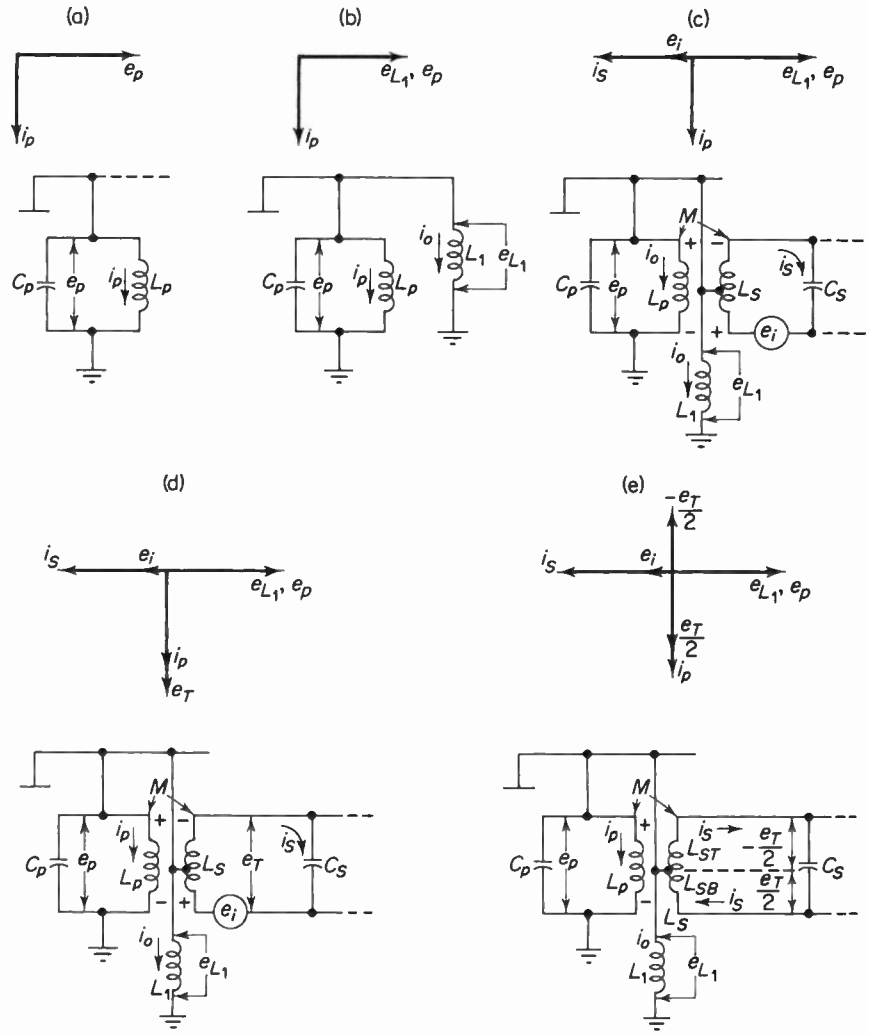



**Fig. 10-7.** Development of final phasor diagram and simplified circuits for condition of $\Delta F_c = 0$.

in phase with $e_p$. Fig. 10-7(c) shows $e_i$ 180° out of phase with $e_p$ and in phase with $i_s$. Since the secondary tuned circuit is resonant, the induced voltage generator "sees" a pure resistance, because the reactances cancel. The secondary circuit is resistive and, therefore, current and voltage are in phase.

Figure 10-7(d) shows that $i_s$ develops a reactive voltage drop $e_t$ across $L_s$ that leads the current through it by 90°. There is also a reactive drop across $C_s$, but this plays no active part in terms of the diode driving voltages and so is not included. Since the secondary winding is center-tapped [Fig. 10-7(e)], the voltage drops across $L_{st}$ and $L_{sb}$ (the upper and lower halves of the secondary winding, respectively) are equal in magnitude. The *instantaneous* polarity of the voltage from bottom to center-tap is assumed positive. The potential at the top end with respect to the center-tap is negative, and is $-e_t/2$. It is therefore permissible to show $-e_t/2$ 180° out of phase with $e_t/2$. Figure 10-7(e) shows the final diagram. The phasor to be used in subsequent discussions is shown in Fig. 10-8. The voltage applied to each diode is then the vector sum of the voltages across $L_1$ and across the appropriate half of the secondary winding. The diode driving voltages are

$$e_{ud} = e_{L_1} + \left(-\frac{e_t}{2}\right), \tag{10-1}$$

$$e_{ld} = e_{L_1} + \frac{e_t}{2}, \tag{10-2}$$

and their additions must be performed vectorially.

At the resonant frequency of $L_p$-$C_p$ and $L_s$-$C_s$, $e_{ud}$ and $e_{ld}$ are equal and therefore the rectified currents (containing the original modulation) flowing in $R_3$ and $R_4$ of Fig. 10-1 are equal. Since the current flow directions through each resistor are opposite to each other, the *average* rectified voltages across $R_3$ and $R_4$ are equal and opposite and therefore cancel. The net voltage from STO to ground is zero. This is as it should be, since the instantaneous carrier frequency is 10.7 MHz and corresponds to *zero modulation* with no audio output.

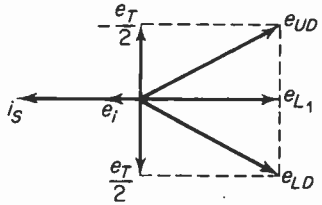



**Fig. 10-8.** Phasor relationships for the Foster-Seeley circuit when the applied carrier is at the resonant frequency of $L_s - C_s$, ($\Delta F_c = 0$). The magnitude of $e_t$ is greatly exaggerated for clarity.

## 10-5. Above resonance

The carrier frequency *instantaneously* deviates above rest and the secondary tuned circuit is off-resonant. Since the secondary tuned circuit is *series* tuned, the induced voltage of Fig. 10-7 will now "see" a resistance in series with an *inductive* reactance. This is because the inductive reactance is now larger than the capacitive reactance and hence the reactances do not cancel. In an inductive circuit, current lags voltage by an angle that is proportional to the ratio of $X_L/R$. Figure 10-9 shows this with $i_s$ lagging $e_i$ by an angle $\theta$. Since the secondary current still flows through an inductance, $i_s$ must still remain in quadrature with $e_t$. Therefore, the half-secondary voltages $e_t/2$ and $-e_t/2$ must shift off-normal by an equal amount in order to retain the quadrature relationship with $i_s$. If voltage additions are again made vectorially for $e_t/2$, $-e_t/2$, and $e_{L_1}$, it will be found that $e_{ud}$ is now larger than $e_{ld}$. The upper diode is driven harder than the lower diode and therefore the rectified output of the upper diode is greater. Figure 10-11 shows the instantaneous direction of current flow through the load resistors and the instantaneous polarity of the sound take-off terminal STO will be positive to ground. It should be noted that the diode load-circuit discharge time constant is too long (10 $\mu$sec) to permit the potentials across either load resistor to respond to carrier-frequency changes. Both output capacitors will charge to the peaks of $e_{ud}$ and $e_{ld}$ and remain there so long as the driving voltages remain constant in magnitude. If the driving voltages vary in amplitude at an audio rate due to frequency modulation, the diodes and their respective load circuits will peak-rectify, since the load time constant is short enough to permit the output voltage to change in-step with the modulation. Figure 10-11(b) shows the output waveform for an input carrier that is 50 kHz above center frequency.

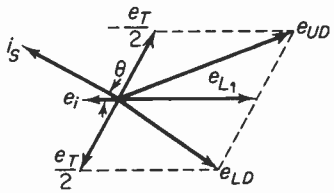



**Fig. 10-9.** Signal frequency is higher than center frequency ($+\Delta F_c$). $\theta$ is proportional to $\Delta F_c$.

## 10-6. Below resonance

The carrier frequency *instantaneously* deviates below rest and the secondary tuned circuit is off-resonant. For this condition, the induced-voltage generator $e_i$ "sees" a resistance in series with a *capacitance*, because the capaci-

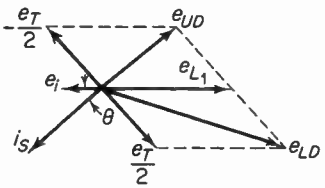



**Fig. 10-10.** Signal frequency is lower than center frequency, $(-\Delta F_c)$. $\theta$ is proportional to $\Delta F_c$.

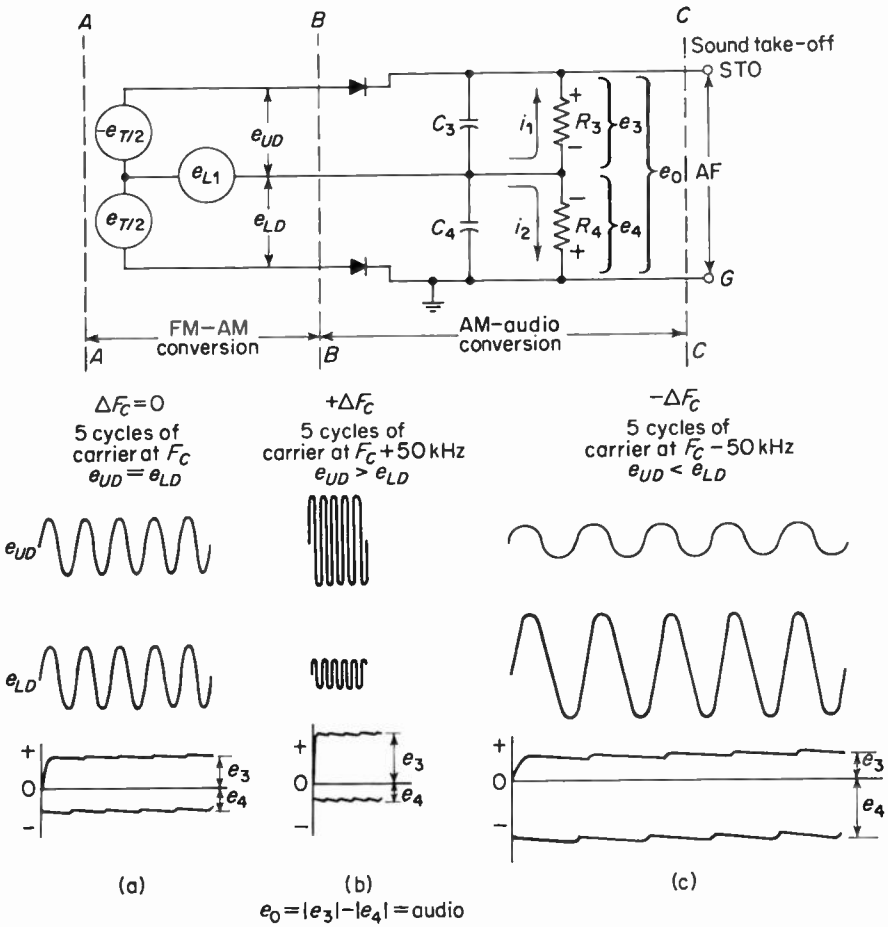



**Fig. 10-11.** Waveforms for all three conditions of modulation. The instantaneous output $e_0$ is equal to the *arithmetic* difference between the load voltages $e_3$ and $e_4$.

tive reactance is larger than the inductive reactance. Thus, $i_s$ leads the induced voltage by an angle $\theta$, the angle being proportional to the amount of frequency deviation. The lower diode driving voltage $e_{ld}$ (vector sum of $e_{L_1}$ and $e_l/2$) is now greater than $e_{ud}$ (vector sum of $e_{L_1}$ and $-e_l/2$). See Fig. 10-10. Thus, the rectified output voltage across $R_4$ is greater than that across $R_3$. Figure 10-11 shows that the polarity at the sound take-off (STO) is negative to ground. A plot of frequency deviation versus discriminator output will yield the response ("S" curve) shown in Fig. 10-12. Should the carrier deviate beyond points $P$ and $M$, the limits of the tuned transformer are such that the output begins to drop off toward zero.

The phasors of Figs. 10-8, 10-9, and 10-10 refer to a *single*, specific carrier frequency resulting in a *single* DC output voltage from STO to ground. When program material is frequency-modulating the carrier, the DC output voltage of the detector is varying at an audio rate and therefore Fig. 10-13 is more indicative of what actually is happening.

Assume that the carrier is shifting between the limits of $+\Delta F_c$ and $-\Delta F_c$ as shown in Fig. 10-13. If the carrier is at rest position or if the deviation is instantaneously zero, the arithmetic difference of $e_3$ and $e_4$ is zero and there is no variable component at the output. If modulation is present, the voltages across the load resistors will vary at an audio rate; these voltages are labeled

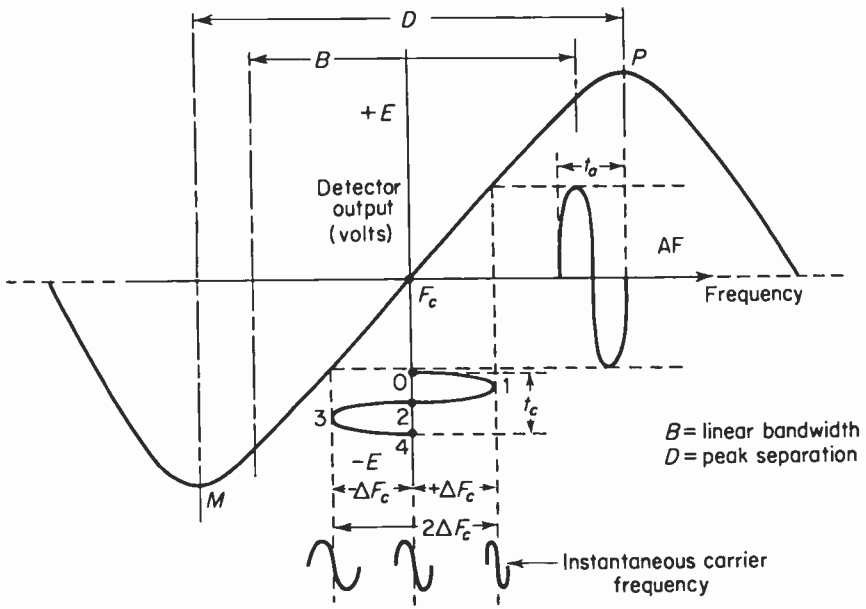



**Fig. 10-12.** Discriminator response ("*S*" curve) for Foster-Seeley Detector. The time $t_c$ (from point 0 to point 4) is the same as the time for one *AF* cycle ($t_a$).

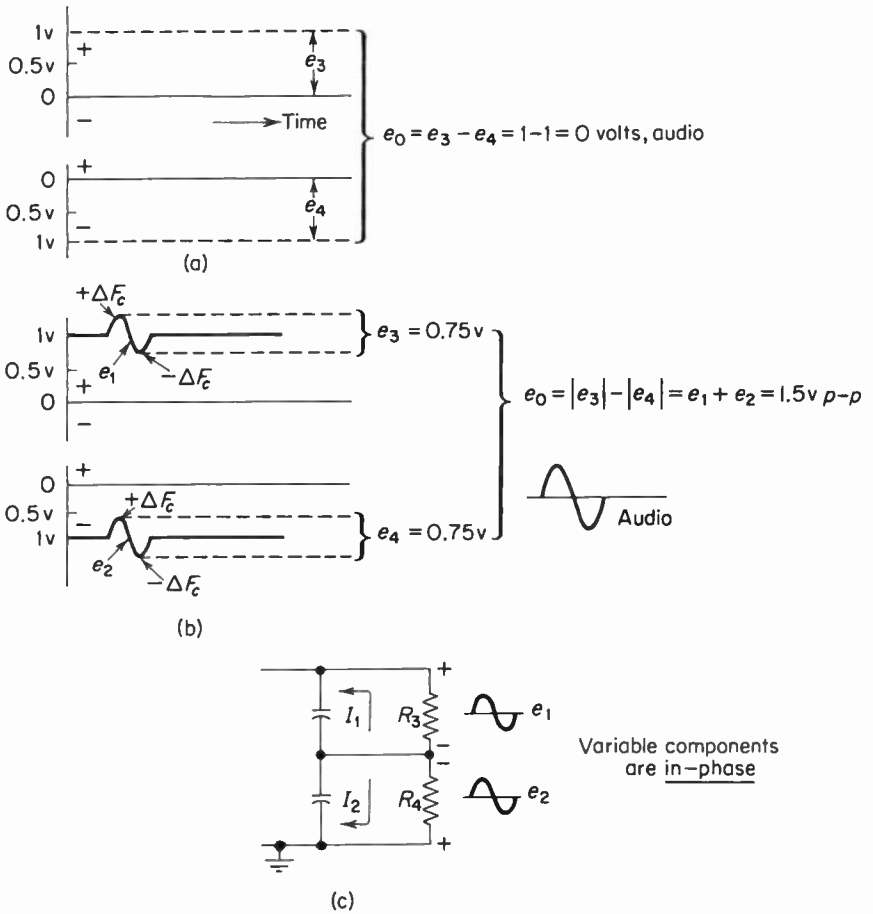



**Fig. 10-13.** Waveforms showing how the DC components across the diode loads cancel and how the variable components add. (c) Shows the fixed polarities at the load circuit with mid-point of the diode load resistors *always* negative to ground with a carrier present.

$e_1$ and $e_2$. These components are superimposed on a DC level (whose magnitude for this example is 1 volt). When the carrier deviates above center, the potential across $R_3$ (with respect to the midpoint of both resistors) becomes more positive and reaches a maximum of 1.375 volts. At the same time, the potential across $R_4$ (with respect to ground) becomes increasingly more positive and reaches a maximum of 0.625 volt. Therefore, at the instant of peak frequency swing, the total instantaneous audio output voltage is 1.375 − 0.625 = 0.75 volt (peak). When the carrier shifts downward in frequency, the potential across $R_3$ goes in a more negative direction and reaches a minimum value of 0.625 volt. Simultaneously, the potential across $R_4$ becomes

increasingly more negative and reaches a value of 1.375 volts. The total instantaneous audio output is 1.375 — 0.625 = 0.75 volt (peak).

We may summarize as follows: the audio output of a Foster-Seeley discriminator is equal to the arithmetic *difference* of the rectified output voltages; the audio output is also equal to the arithmetic *sum* of the variable components across the respective diode load resistors; the direct components of voltage across the respective diode load resistors are *subtractive;* the variable components ($e_1$ and $e_2$) are *in* phase and are thus *additive*. It can also be seen in Fig. 10-13 that there exists an audio output voltage across either diode load resistor, but if the audio is taken from across one resistor only, any noise-rejection property of the circuit is immediately lost. This will be discussed in a subsequent section.

Since the rate of change of frequency deviation is the same as the frequency of the original audio voltage, the rate of change of the instantaneous DC voltages at the discriminator output corresponds to the original audio frequency. Thus, the frequency and amplitude characteristics have been retained. This is shown in Fig. 10-12. The carrier goes through a complete frequency excursion in time $t_c$ from point 0 to point 4 and, at the same time, the audio voltage goes through a complete cycle in time $t_a$.

## 10-7. Noise-suppression factors

It can be shown mathematically that the AM suppression ratio† for good noise limiting should be of the order of 30 to 40 db. This means that the power output of the detector due to FM should be approximately 40 db above the power output due to AM with both inputs simultaneously modulated at the same percentage but by different audio frequencies. A Foster-Seeley detector provides about 9 db suppression, which is completely inadequate for good sound reproduction.

Figure 10-14 shows graphically why this detector requires prelimiting (at least one stage and preferably two). The upper of the two waveforms shows the potential across the upper diode load resistor and is labeled $e_3$. The lower waveform is the potential across the lower resistor and is labeled $e_4$. It is assumed first that the carrier is *not* deviating from center frequency and therefore the operating point for both sinewaves is at point *X*. At this instant, a noise spike appears and causes conduction through *both* diodes to increase. The potential $e_3$ increases in a positive direction to a value of, say, 1.5 volts. At the same instant, the heavy current through the lower diode causes the potential to increase, but in a more negative direction, and $e_4$ reaches a value

†This ratio is obtained by simultaneously frequency-modulating the receiver input at 22.5 kHz (or 30 per cent of system deviation) at 1 kHz audio and amplitude-modulating the input at 30 per cent modulation at 400 Hz. The AM suppression ratio is then the ratio of the desired output (FM) to undesired output (AM) expressed in decibels.

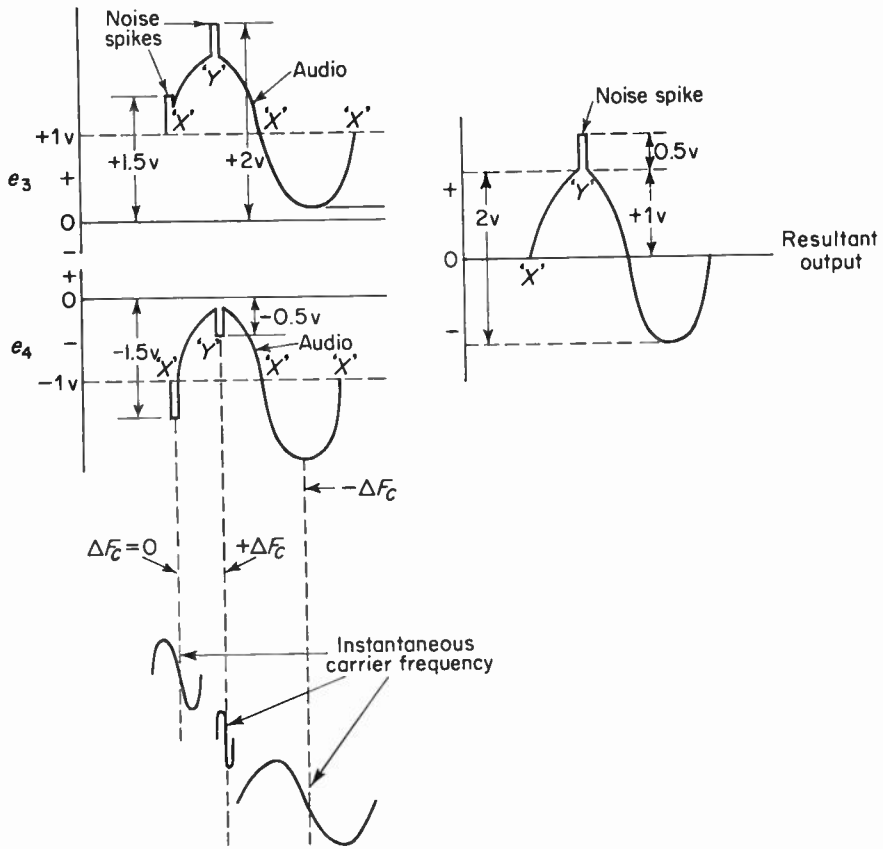
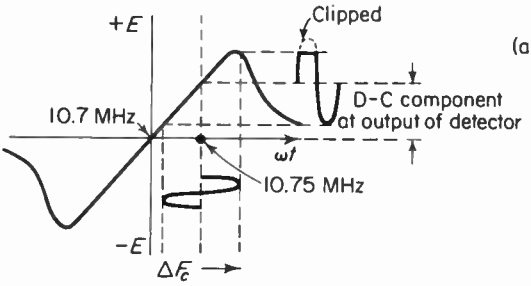


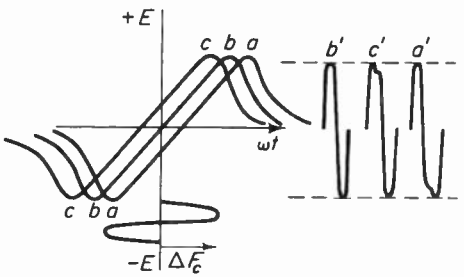

**Fig. 10-14.** Graphical representation showing why a Foster-Seeley Detector requires pre-limiting. Resultant output at "*Y*" = 1.5 v and at "*X*" = 0 v. Therefore, Foster-Seeley circuit can reject noise only when carrier is at rest position ("*X*") when instantaneous modulation is zero.

of −1.5 v. Since the instantaneous demodulated output is the arithmetic difference between the voltages, the total output is 1.5 − 1.5 = 0 v. The disturbance is thus effectively suppressed provided two conditions are met: noise spikes must occur at points in the modulation cycle when the instantaneous carrier frequency is at rest position *and* the detector must be center-tuned so that the net DC output voltage is zero when the carrier is at rest position.

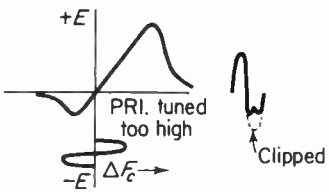
Let us examine the case when the noise spikes occur at times in the modulation cycle when the carrier is "off center." Such a case might be when the carrier is above rest frequency. This corresponds to point *Y* in Fig. 10-14. In this instance, the disturbance raises $e_3$ to a peak positive potential of 2 volts, while at the same time the spike increases $e_4$ but in a more *negative*




(a) Receiver mistuned causing the local oscillator to operate at an incorrect frequency. This results in an incorrect value of IF. Oscillator set 50 kHz too high for this example.

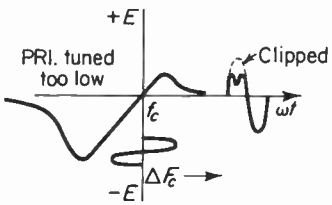


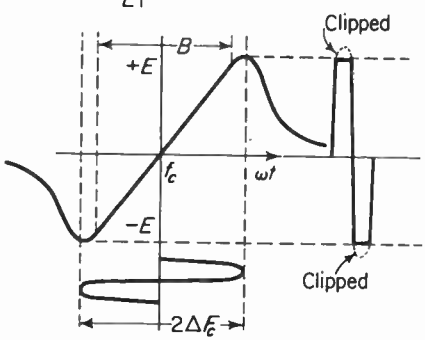

(b) Detector secondary circuit mistuned (improper alignment).
*a* – det. sec. tuned too high
*b* – " " " correctly
*c* – " " " too low
Curves *b'*, *c'* and *a'* represent AF outputs for conditions *b*, *c* and *a* respectively.




$f_c$ = center frequency

(c) Primary tuned circuit of detector transformer mistuned (improper alignment, open primary shunt capacitor). Sound very weak and distorted, particularly on large carrier swings.




(d) Distortion caused by overmodulation or insufficient linear bandwidth. This will occur on modulation peaks or when linear bandwidth is less than $2\Delta F_c$ (max.)

**Fig. 10-15.** Types of distortion and possible causes.

direction. The net potential across the entire load circuit is the arithmetic difference between $e_3$ and $e_4$. Thus, the demodulated output is $2 - 0.5 = 1.5$ v, of which 0.5 v represents noise remaining in the output. Thus, noise due to auto ignition, rapid carrier fading, beats between the desired and an undesired carrier, and any other source of interference that *amplitude*-modulates the FM carrier will be rejected only if the disturbance occurs when the carrier is *not* being instantaneously frequency-modulated *and* the steady-state DC output voltages across the respective diode load resistors are equal.

The two conditions just mentioned establish the need for proper alignment and some form of tuning indicator to insure that the detector is center tuned at all times. Noise rejection will then be optimum, but neither a tuning indicator nor proper alignment is sufficient. At least one stage (preferably two) of prelimiting is essential when using a Foster-Seeley discriminator. Figure 10-15 indicates forms of distortion and their possible causes. A compromise detector design is given in Fig. 10-16.

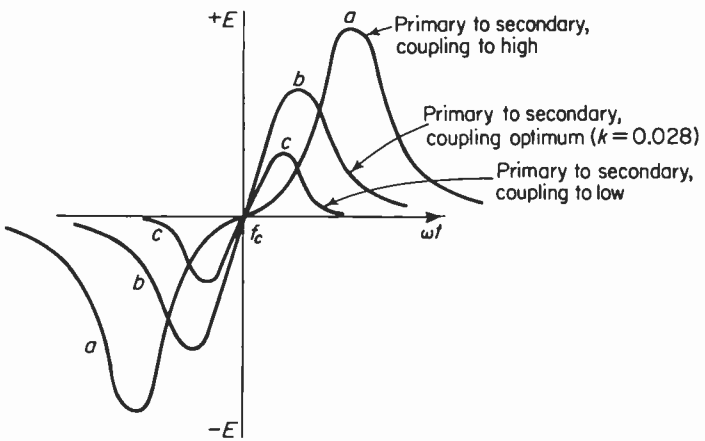



**Fig. 10-16.** If primary to secondary magnetic coupling is too "tight," detector response becomes very non-linear. If coupling is too "loose," detector output drops off considerably. Optimum coupling gives maximum output with good linearity extending over a range of at least $2\Delta F_c$(max.) where $\Delta F_c$(max.) $= \pm 75$ kHz.

## 10-8. Tuned-circuit damping and effect on circuit $Q$

When a resistance is placed across a tuned circuit, the circuit is said to be *damped.* When a tuned circuit is thus "loaded down," its effective or working $Q$ (which is not necessarily the coil $Q$) is reduced below its original value. It will be recalled that when a generator drives a resonant tuned circuit, the generator "sees" an impedance that is resistive, and the magnitude of the impedance is

$$Z = QX, \tag{10–3}$$

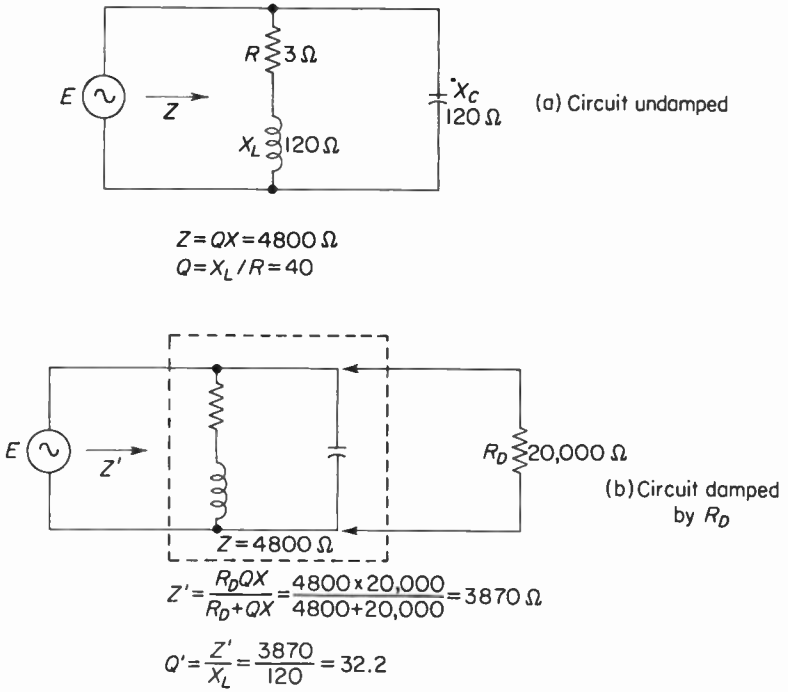



**Fig. 10-17.** Example of damping and reduction of circuit $Q$.

where $Q = X_L/R = X_C/R$. Figure 10-17(a) shows a typical arrangement (undamped) where the coil $Q$ is 120/3 or 40. The impedance presented to the generator is 120 × 40 or 4800 ohms, pure resistance. Assume now that a resistor $R_d$ is placed across the tuned circuit as shown in Fig. 10-17(b). The value of impedance that the generator now "sees" is the parallel combination of 4800 ohms and 20,000 ohms, which is 3870 ohms. The effective or working $Q$ is then 3870/120 = 32.2 (an 8 per cent reduction). Thus, the effective $Q$ is directly proportional to the magnitude of the damping resistor $R_d$, and the $Q$ approaches the unloaded $Q$ as $R_d$ approaches infinity. Also, since bandwidth $B = f_c/Q$, the tuned-circuit bandwidth is inversely proportional to the magnitude of $R_d$, where $f_c$ is the resonant frequency of the tuned circuit. To sum up: the effective or working $Q$ of a tuned circuit at resonance is

$$Q_{\text{eff}} = \frac{R_{\text{eq}}}{X}, \tag{10–4}$$

where $R_{\text{eq}}$ = the parallel combination of the resonant tank impedance (resistive) and $X$ = the inductive reactance of the coil.

Let us examine the damping of a tuned circuit used to drive an envelope detector. An actual circuit of the simple diode detector is shown in Fig. 10-18(a). It will be recalled that the diode conducts only on the positive carrier peaks and that the potential across the load is approximately equal

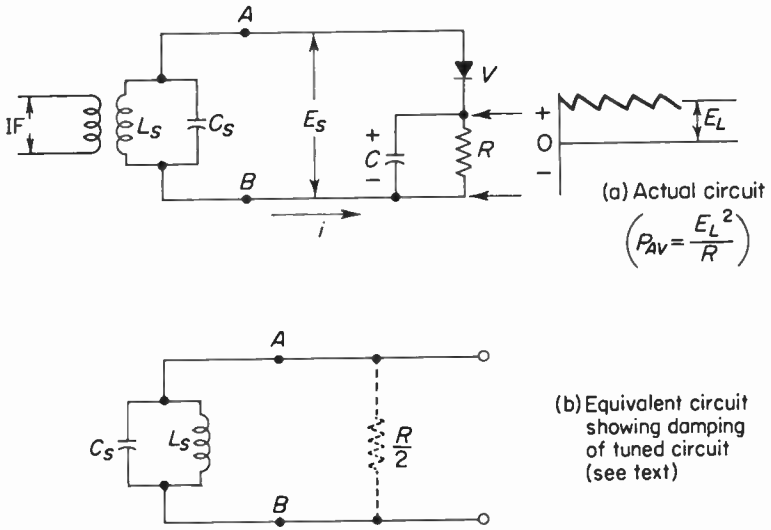



**Fig. 10-18.** Damping of tuned circuit used to drive simple AM diode detector.

to the *peak* carrier voltage $E_s$. If the small forward drop across the diode is neglected, little error is introduced, since the forward resistance is small compared to $R$. The power dissipated by $R$ is then $E_L^2/R$. This power is taken from the driving source (the tuned circuit), so we place an equivalent resistance across the tuned circuit and then consider the power taken by this hypothetical resistor on the basis of the average power dissipated per IF cycle. The average power is the ratio $E_{\text{rms}}^2/R$. However, $E_s = \sqrt{2}\,E_{\text{rms}}$, so the average power taken from the source is

$$P_{\text{av}} = \frac{(\sqrt{2}\,E_{\text{rms}})^2}{R} = \frac{2E_{\text{rms}}^2}{R}.$$

Dividing both numerator and denominator of the right side by 2 gives

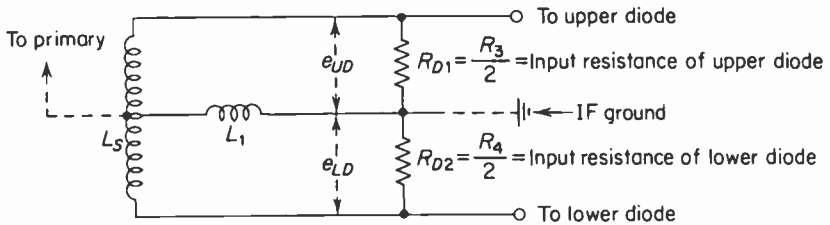
$$P_{\text{av}} = \frac{E_{\text{rms}}^2}{R/2}. \tag{10–5}$$

This is the general form of voltage²/resistance, but the resistance as indicated is not $R$ but $R/2$. The hypothetical resistance $R/2$ shown in Fig. 10-18(b) is then the input resistance presented by the diode circuit to the tuned circuit. On the basis of $Q_{\text{eff}} = R_{\text{eq}}/X$, one can readily see how the diode circuit damps or loads down the source, resulting in a reduction of $Q$ and a widening of the bandpass. The analysis above is then an argument for making the diode load resistance reasonably large to maintain sensitivity.† Also, owing to volt-

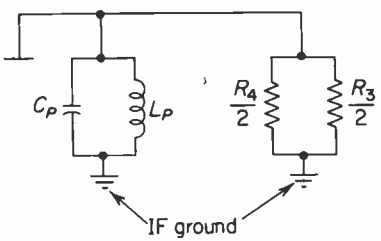
†There are many applications where the tuned circuit is intentionally damped to obtain a wider bandwidth. Chapter 7 goes into this thoroughly.

age division between the diode and $R$, the load resistance should be large compared to the forward resistance of the diode.

The previous idea is now extended to the Foster-Seeley detector. Design considerations generally indicate that for optimum operation, the primary and secondary $Q$'s should be equal and the primary inductance should equal the secondary inductance (or perhaps be slightly smaller). Equal primary and secondary $Q$'s assumes that the tuned-circuit loadings are identical. Figure 10-11 shows that each diode plus its load is driven by a generator designated $e_{ud}$ (for the upper diode) and $e_{ld}$ (for the lower diode). $R_3$ draws power from $e_{ud}$, and an equivalent resistance $R_{d_1}$ (equal to $R_3/2$) can be placed across $e_{ud}$. The same situation is valid for the lower diode. These conditions are shown in Fig. 10-19. With regard to the entire secondary winding, both input resistances are effectively in *series*, and the loading on the secondary is equal to half of the total diode load resistance ($R_3 + R_4$). The damping imposed on the *full* secondary tuned circuit is 100,000 ohms. Figure 10-19(b) shows the simplified damping circuit for the primary winding. Here, the secondary winding resistance and the resistance of coil $L_1$ are neglected. The parallel combination of $R_3/2$ and $R_4/2$ is effectively across $C_p$-$L_p$, and the loading on the primary is thus equal to 25,000 ohms. The damping imposed on the primary side is thus *four* times the damping on the secondary side. If the primary and secondary $Q$'s are to be the same, then the secondary $Q$ must be reduced. Some commercial designs incorporate a resistance in shunt with the




(a) Damping imposed on entire secondary = $R_4$




(b) Damping imposed on primary side = $R_4/4$

**Fig. 10-19.** Damping imposed on detector transformer.

secondary winding. It should be emphasized that any circuit component bridged across the sound take-off will affect the circuit $Q$ and hence the sensitivity and selectivity. This also includes meters, tuning indicators, filter circuits, deemphasis networks, and so on.

## 10-9. Foster-Seeley modifications

The circuit shown in Fig. 10-1 is not the only type used. In fact, the arrangement shown in Fig. 10-20 is much more common and somewhat less expensive. The operation of the circuit is exactly as explained for the circuit of Fig. 10-1, except that the unbalancing voltage $e_{L_1}$ is derived in a different manner. This reference voltage is obtained by means of $R_3$ and $R_4$ in parallel for IF purposes with the equivalent driving circuit shown in Fig. 10-21. $R_3$ and $R_4$ thus furnish the unbalancing voltage instead of a coil. The recovered audio is again the arithmetic difference between the voltages across $R_3$ and $R_4$. This last point is applicable to *all* Foster-Seeley circuits.

A few important differences do exist between the two circuits of Figs. 10-1 and 10-20. Figure 10-22 shows the equivalent damping circuits for both

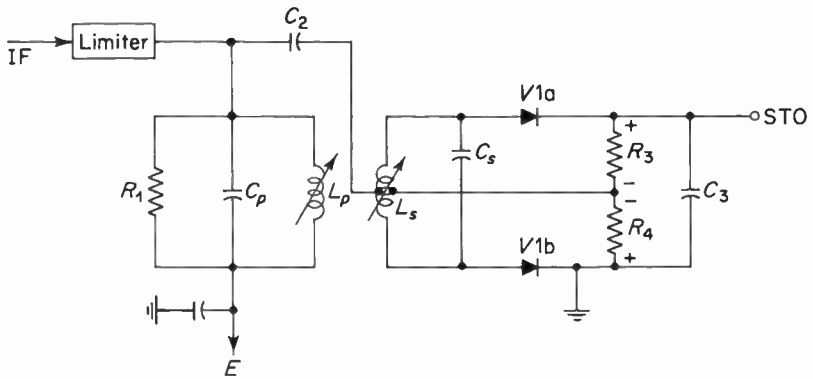



**Fig. 10-20.** Practical form of Ratio Detector.

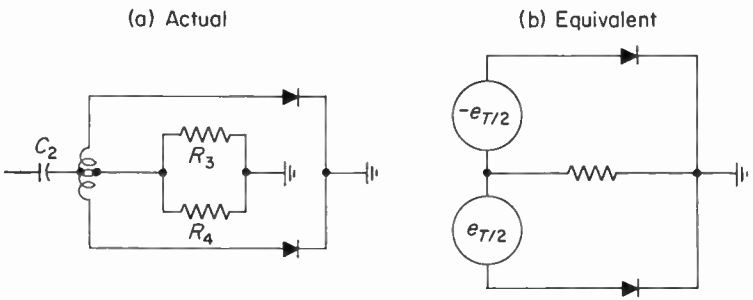



**Fig. 10-21.** Driving circuit for diodes of Fig. 10-20.

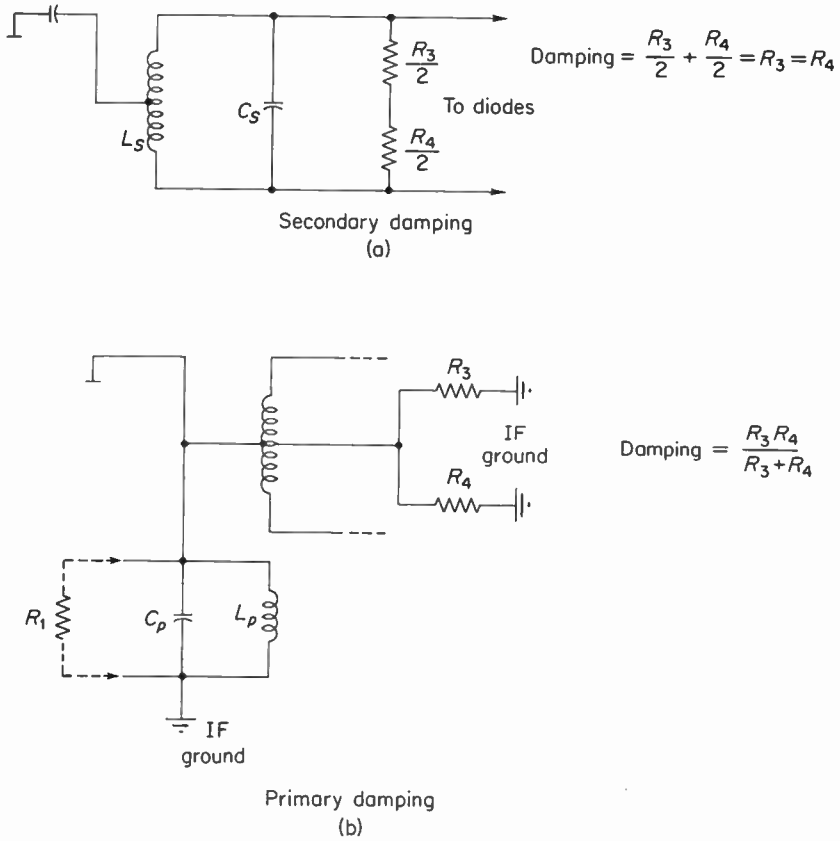
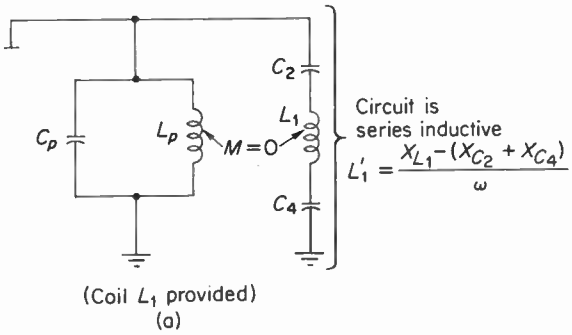


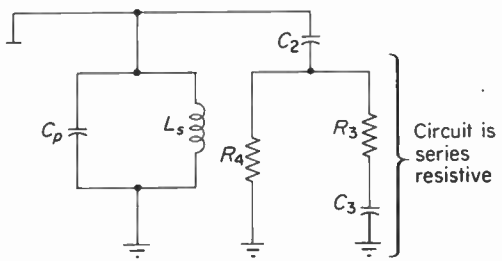

**Fig. 10-22.** Equivalent circuits for primary and secondary damping for circuit of Fig. 10-20. To equalize $Q$'s, $R_1$ is added across $L_p - C_p$.

primary and secondary circuits. The secondary side is damped by a diode input resistance of 100,000 ohms. This is the same value of damping imposed on the secondary of Fig. 10-1. Figure 10-22(b) shows the primary damping to be 50,000 ohms instead of 25,000 ohms as was the case for the primary of Fig. 10-1 All other conditions being equal, the primary $Q$ for the previously described circuit is lower. This is the reason for the additional damping (in the form of resistor $R_1$) across the primary winding. The presence of $R_1$ also reduces the plate-load impedance of the preceding limiter and consequently reduces its gain. This is not serious, since there is sufficient gain to meet receiver requirements.

Another circuit difference concerns the effect that $L_1$ has on the choice of the primary tuning capacitor. If the coil is retained, then the primary IF equivalent circuit is as shown in Fig. 10-23(a). Notice that the series branch $C_2$-$L_1$-$C_4$ is inductive, since the combined series reactance of $C_2$ and $C_4$ is very




(Coil $L_1$ provided)
(a)




Primary circuit inductance essentially unchanged

(Coil omitted)
(b)

**Fig. 10-23.** Omission of $L_1$ increases primary circuit inductance.

small compared to $X_{L_1}$. This is equivalent to an inductance $L_1$ in parallel with $L_p$. When the two coils are placed in shunt, the net inductance is less than the inductance of either coil (assume that there is no mutual coupling between the coils). Therefore, in order to maintain the same resonant frequency, the primary tuning capacitor must be greater than if $L_1$ were not present. Note the difference in primary capacitance between the circuits of Figs. 10-1 and 10-20.

A further modification worth noting is shown in Fig. 10-24. Note again the omission of $L_1$. A review of Fig. 10-25, showing the IF equivalent circuit, will help one understand the operation of this modified circuit.

The primary current $i_p$ induces a voltage *into* the center-tapped secondary winding $L_s$. The upper portion is $L_{st}$ and the lower portion is $L_{sb}$. Each coil can then be considered as having an induced-voltage generator $e_t$ in series with it. These generators force a secondary curent $i_s$ to flow around the loop consisting of $L_{st}$, $L_{sb}$, $C_{st}$, and $C_{sb}$. The reactive voltage drops are labeled, as before, $e_t/2$ and $-e_t/2$. $L_{st}$ series-resonates with $C_{st}$, and $L_{sb}$ series-resonates with $C_{sb}$. When a tuned circuit is series resonant, it represents a very

low resistance to the driving voltage (in this case $e_i$). Since both series resonant circuits are in parallel (with respect to points $A$ and $B$), the resistance from $A$ to $B$ is very small and is given by

$$R_{AB} = \frac{\left(\frac{e_i}{i_s}\right)\left(\frac{e_i}{i_s}\right)}{\frac{e_i}{i_s} + \frac{e_i}{i_s}} = \frac{e_i}{2i_s} \tag{10–6}$$

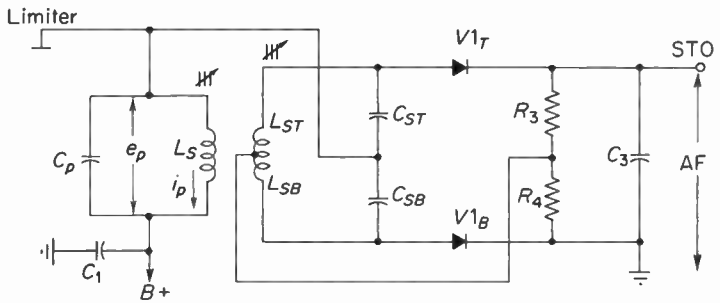



**Fig. 10-24.** Another alternative to Foster-Seeley circuit. Coil $L_1$ is again omitted in addition to one IF by-pass capacitor.

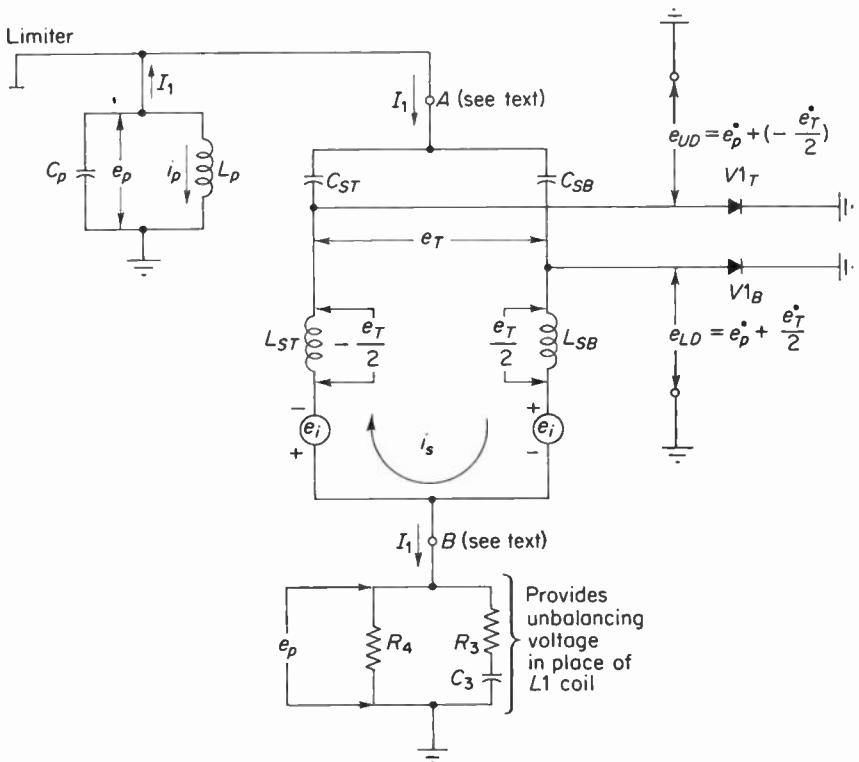
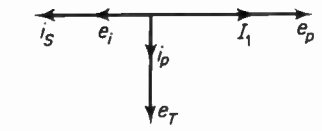
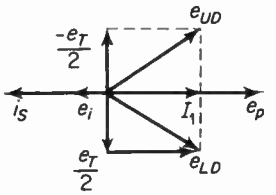



**Fig. 10-25.** Equivalent IF circuit for obtaining diode driving voltages.

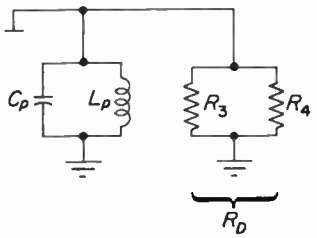



(a) Basic vector diagram



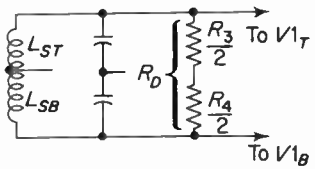

(b) The inductive drops ($e_{T/2}$ and $-e_{T/2}$) are 180° out of phase with respect to the center tap on the secondary coil $L_S$

**Fig. 10-26.** Resonant phasors for circuit of Fig. 10-24.




$$R_D = \frac{R_3 R_4}{R_3 + R_4} = \frac{R_3}{2} = \frac{R_4}{2}$$

Primary
(a)




$$R_D = \frac{R_3}{2} + \frac{R_4}{2} = R_3 = R_4$$

Secondary
(b)

**Fig. 10-27.** Equivalent circuits for damping of primary and secondary of circuit in Fig. 10-24.

This places point $B$ at the same IF potential as point $A$, and therefore the potential across the parallel combination of $R_3$ and $R_4$ is the primary voltage $e_p$. From the standpoint of the diodes, however, the total driving voltage for each tube is the vector sum of $e_p$ and the reactive drops across each half-secondary. The diagram shown in Fig. 10-26 is essentially the same as before, and the equivalent damping circuits are shown in Fig. 10-27. The primary and secondary damping are the same as the arrangement shown in Fig. 10-20.

## 10-10. Design aspects

Certain design aspects related to $C_2$ (Fig. 10-20) are worth noting. Its value is fairly critical, and a substitute should be of the same value. The reason is as follows: if the substitute is very much smaller than the value indicated, the magnitude of the unbalancing voltage will be reduced owing to a larger

drop across $C_2$ with less voltage across the parallel combination of $R_3$-$R_4$. This will result in a smaller audio output voltage for a given frequency deviation. Figure 10-28 shows the effect of making $C_2$ smaller than it should be, resulting in three different audio outputs. The designations $AF$, $AF'$, and $AF''$ correspond to three different outputs for large, small, and very small values of $C_2$, respectively.

The question then arises as to why $C_2$ is not made very large in order to do away with this problem altogether. The answer lies with the load circuit and how it could be severely unbalanced if $C_2$ were increased without limit. Assume that the audio frequency to be recovered is 10 kHz. If the audio component across either load resistor is reduced, owing to the substitution of a large capacitor for a small one, the total output will be reduced (in addition to being distorted). Figure 10-29 shows a "sneak" low-impedance path from the midpoint of $R_3$-$R_4$ through a "large" $C_2$, through $L_p$ (practically zero reactance), and through $C_1$ to ground. In other words, the resistors "see" different impedances for audio. The load circuit is severely unbalanced, resulting in distortion of the audio. This will distort the "S" curve, since the

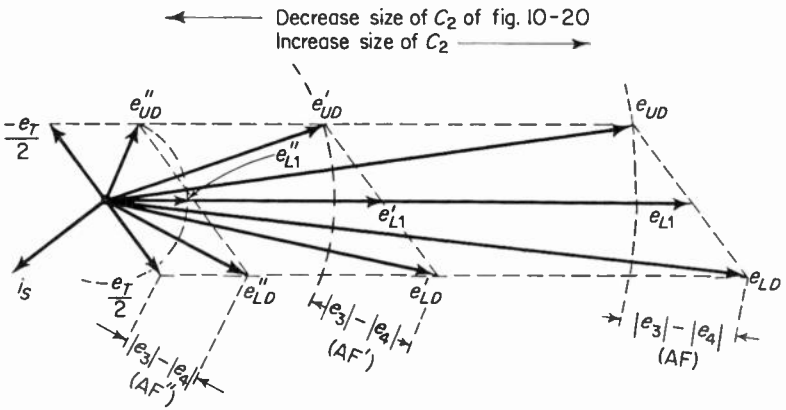



**Fig. 10-28.** Phasor diagram showing effect on audio output for different values of $C_2$. It is assumed that primary voltage is held constant. ($AF > AF' > AF''$)

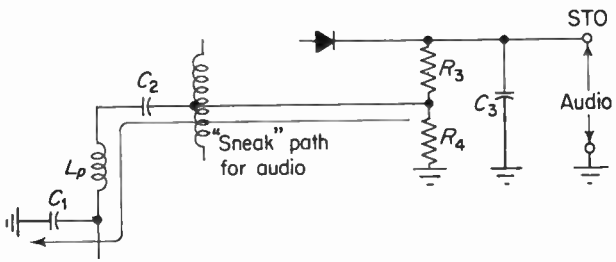



**Fig. 10-29.** Simplified circuit showing low impedance across $R_4$ due to incorrect value (too large) of $C_2$. This will unbalance the detector load.

instantaneous DC voltages across the respective loads will be different for equal positive and negative frequency deviations. $C_2$ should therefore be no larger than necessary to insure that the audio output circuit is effectively isolated from the primary side of the detector transformer.

$R_3$ and $R_4$ should be balanced to within at least 5 per cent to maintain proper detector linearity and maximum noise rejection. $C_3$ should not be too large, or diagonal clipping will occur owing to too long a time constant. This will distort the high modulation frequencies, since the capacitor will be unable to follow *rapid* envelope variations.

## 10-11. DC control voltage for AFC, tuning indicators, and alignment

In an earlier discussion of oscillators we referred to a DC voltage needed to control the reactance modulator, which in turn "pulled" the local oscillator. This voltage is readily available at the detector output, since the differential DC output voltage is proportional to the amount of frequency departure from the IF. This voltage must be filtered so that the audio component does not frequency-modulate the local oscillator and reduce the detector output. If this DC component is separated from the audio component by means of a low-pass filter, such as shown in Fig. 10-30, it can be used as the control voltage for a reactance modulator (tube or variable-capacitance diode) for automatically controlling the oscillator frequency (AFC). This DC potential is also present if the receiver is improperly tuned, since this will also result in an incorrect value of IF.

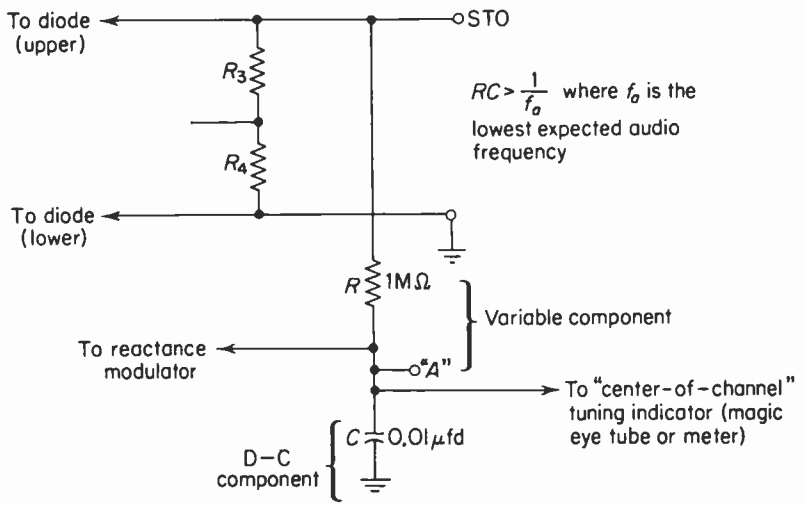



**Fig. 10-30.** Access point for AFC and tuning eye voltage. Voltage at point "*A*" should be well filtered.

The choice of components here will have some effect on the low-frequency audio response, since, if the potential across $C$ (Fig. 10-30) is permitted to vary at a low frequency rate (this will happen if the time constant is too short), the local oscillator will be likewise frequency-modulated. The frequency deviation (at the input to the IF amplifiers) will be reduced for the lower audio frequencies, and therefore the detector will produce less output voltage for these frequencies.

If the potential at point $A$, Fig. 10-30, is to be used for controlling a "center-of-channel" tuning indicator, the modulation components must be filtered out to prevent the magic eye from "winking." The $RC$ network shown could thus perform double duty in providing a pure DC voltage whose magnitude and polarity (with respect to chassis ground) are a function of the magnitude and direction of detuning.

## 10-12. Deemphasis and frequency response

Chapter 2 mentioned the necessity for *preemphasizing* the higher modulation frequencies, since these components produced smaller carrier deviations. The standard preemphasis time constant was given as 75 $\mu$sec, and a typical network was shown as a simple $LR$ combination. In order to bring the audio components back into balance, the complementary effect must now take place in the receiver. If the energy content of the higher modulation frequencies is increased at the transmitter, these components must now be attenuated in the receiver. The *deemphasis* network is generally placed across the output of the detector, as shown in Fig. 10-31. The time constant for the deemphasis network is 75 $\mu$sec, and the network consists of a single-section low-pass filter. The deemphasis circuit can affect the frequency response of the receiver, while the high-frequency (above about 1500 Hz) response is a function of the time constant.

A typical arrangement is shown in Fig. 10-31 and is labeled $R_d$-$C_d$-$C_w$. $R_d$ and $C_d$ are physical components, while $C_w$ represents stray wiring and shield capacitance to ground. The cable capacitance cannot always be ne-

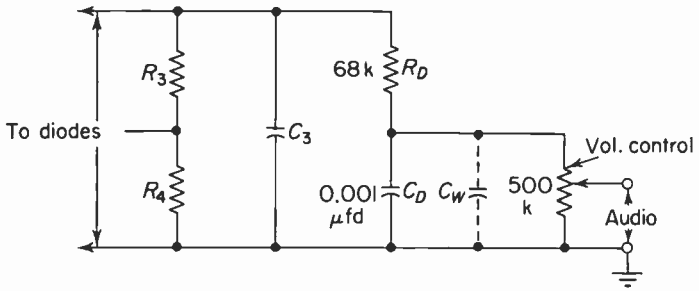



**Fig. 10-31.** $R_D$, $C_D$ and $C_W$ comprise deemphasis network.

glected, since the signal is generally carried to an output jack or to a cathode follower stage, neither of which is always placed physically close to the detector output. $C_w$ is then simply added to $C_d$, since they are both in parallel. The design of the deemphasis network can have a marked effect on the audio response; therefore the choice of components should not be taken lightly if the response is to be flat or at least within a few decibels of the standard deemphasis curve shown in Fig. 10-33.

Figure 10-32 shows the deemphasis network consisting of $R$ and $C$, where $C = C_d + C_w$. The input voltage to the network is labeled $e_{in}$ and is obtained directly from the detector output. The deemphasized audio output voltage $e_o$ is then applied to a cathode or emitter follower stage or to an audio output jack. The input voltage is

$$e_{in} = i\sqrt{R^2 + \frac{1}{\omega^2 C^2}}, \tag{10-7}$$

where $\omega = 2\pi f_a$. The output voltage is

$$e_o = \frac{i}{\omega C}. \tag{10-8}$$

Dividing Eq. (10–8) by Eq. (10–7) gives the network gain, which is

$$A = \frac{e_o}{e_{in}} = \frac{1}{\omega C\sqrt{R^2 + \frac{1}{\omega^2 C^2}}}. \tag{10-9}$$

Simplifying, we have

$$A = \frac{1}{\sqrt{\omega^2 C^2 R^2 + 1}} \tag{10-10}$$

or

$$A = (\omega^2 R^2 C^2 + 1)^{-1/2}. \tag{10-11}$$

The voltage gain in decibels (network response) is

$$\begin{aligned} \text{db} = 20 \log A &= 20 \log (\omega^2 R^2 C^2 + 1)^{-1/2} \\ &= 10 \log (\omega^2 R^2 C^2 + 1). \end{aligned} \tag{10-12}$$

If we let $RC = T$, then

$$\text{db} = -10 \log (\omega^2 T^2 + 1). \tag{10-13}$$

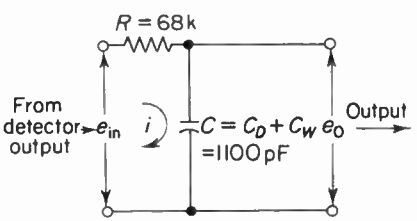



**Fig. 10-32.** Deemphasis network and voltages used to derive response of filter in terms of $RC$.

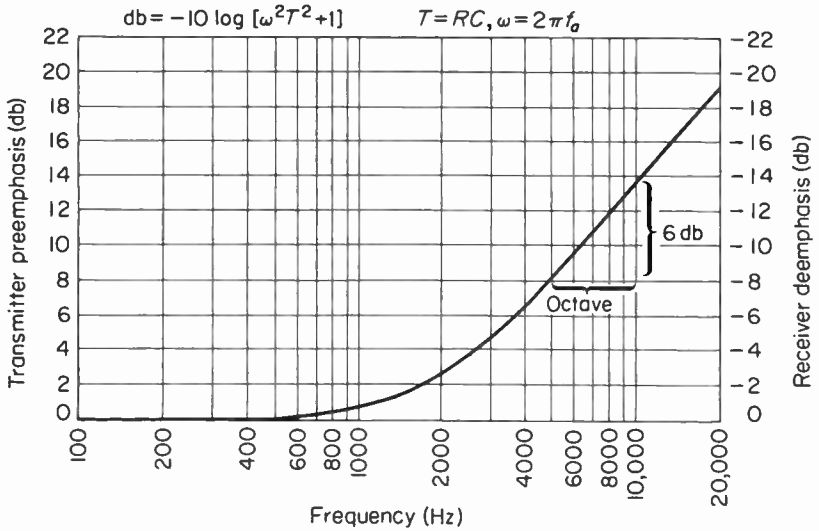



**Fig. 10-33.** Preemphasis (rising) characteristic and deemphasis (falling) characteristic for a time constant of 75 μsec.

Thus, the system response in decibels is a function of the *time constant* of the deemphasis network. A typical value for $R_d$ is 68,000 ohms, while for $C_d$ it is 0.001 μfd. Cable capacitance, stray wiring, and the input capacity of the following stage could contribute about 100 pfd to bring the total shunt capacitance to about 1100 pfd. This will provide a deemphasis time constant of 75 μsec.

Let us look at an example of how the time constant affects the tuner response. Assume that we make a plot of Eq. (10–13) for different audio frequencies and a time constant of 75 μsec. The curve will be as shown in Fig. 10-33. Note that for an audio frequency of 8 kHz, the preemphasis network at the transmitter boosted the amplitude of this component about 12 db above the level of a 500-Hz component. At the receiver, this same 8-khz tone is deemphasized 12 db, which means that it was brought down to the original 500-Hz level (0 db). If this same frequency (8 kHz) is transmitted and the time constant is increased to 150 μsec due to excessive cable capacitance, wiring dress, and so on, the amount of attenuation in this case is, from Eq. (10–13), −17.6 db. This represents a difference of about 5.6 db or an excessive reduction of high-frequency response. It should be noted from Fig. 10-33 that the rate of attenuation is very small for frequencies below about 1500 Hz, and that from about 1500 Hz on up, the rate of attenuation is fairly constant. The rate of attenuation is given by

$$\text{rate} = \text{no. of db/octave} = \text{slope of linear portion of curve}, \quad (10\text{–}14)$$

where an octave refers to a frequency *change* of 2 to 1 (or 1 to 2). For the curve shown, the rate is approximately 6 db/octave, which means that for

every increase in frequency of 2 to 1 (from 4 to 8 kHz, for example), the higher component is attenuated by a factor of 6 db. At the transmitter, the exact opposite occurs—namely, every time the audio frequency doubles, the 8-kHz tone is *boosted* by 6 db.

The reduction in noise output of an FM receiver over an AM receiver for a standard 75-$\mu$sec deemphasis network is approximately 7 db. This improvement does not include gains due to higher transmitter efficiency and basic impulse and fluctuational noise reduction. The total improvement in signal-to-noise ratio for FM over AM is of the order of 28 db, as discussed in Chapter 2.

The previous mathematical analysis can be summed up as follows. As the frequency of the audio voltage ($e_{in}$) increases, the reactance of $C_d$ decreases. Since $R_d$ and $C_d$ comprise a voltage divider, the portion of $e_{in}$ appearing across $C_d$, which is given by

$$e_o = \frac{e_o X_c}{\sqrt{R^2 + X_c^2}}, \tag{10–15}$$

decreases. As the audio frequencies decrease, the reactance of $C_d$ increases and therefore $e_o$ increases. Thus, below about 1500 Hz very little attenuation takes place, but above this frequency the *rate* of voltage reduction is approximately 6 db/octave. If the deemphasis network is chosen carefully (all shunt capacitance should be considered), the audio response will be flat. This network is always placed at the output of the detector in order to immediately attenuate the higher audio frequencies, which might otherwise cause regeneration should they reach a later stage.

## 10-13. Ratio detector

The slope, Travis, and Foster-Seeley detectors all require some form of prelimiting, since they all respond to AM as a result of noise. The circuit to be described next does not require a limiter preceding it, since it is, to a large degree, self-limiting. However, its overall performance can be improved considerably if one or more stages of limiting are provided. Most high-quality tuner designs do just that in order to optimize the AM rejection and wideband capabilities.

We shall discuss modulation recovery first and then the AM rejection properties. A complete waveform analysis will follow, coupled with sufficient information to permit the reader to fully understand the operation and alignment of the ratio detector.

Figure 10-34 shows a preliminary form of the ratio detector, which will be used for the initial discussions. Modifications will be made later and explained, but the arrangement shown will allow the reader to follow the explanations more easily.

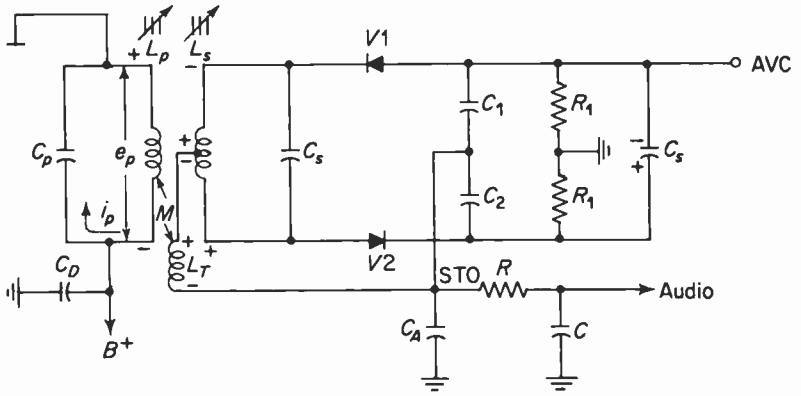



**Fig. 10-34.** Preliminary form of balanced ratio detector.

The major differences between the Foster-Seeley detector and Fig. 10-34 lie in (1) the reversal of one diode, (2) the presence of a third winding in the detector transformer, $L_t$ (referred to as a tertiary coil), and (3) the large electrolytic capacitor $C_s$ connected across the entire diode load. The diodes could be either a dual high-vacuum diode or a pair of matched germanium or silicon diodes. The advantage of using germanium (or silicon) is the absence of a heater, which eliminates the problem of heater-cathode leakage and its attendant hum. The operation of the detector is as follows. Winding $L_t$ consists of a few turns of wire tightly wound around the "cold" or $B^+$ end of the primary winding. This effects an impedance match between the high impedance of the preceding stage and the relatively low impedance of the diode load, which reflects a damping resistor of approximately 76,000 ohms in parallel with the primary winding. If a direct connection were to be made from primary, via a capacitor, to secondary (as in the Foster-Seeley), the primary damping would be excessive owing to the very low value of diode load resistance. The tertiary winding then plays the same role as the IF choke in the Foster-Seeley circuit. By means of induction, the primary induces a voltage into $L_t$ which is then used to unbalance the circuit when the carrier deviates from rest position. Figure 10-34 shows that the secondary winding is center-tapped. Therefore, the same phase relationships exist here as in the Foster-Seeley circuit, with one or two minor exceptions.

The primary voltage and current are in quadrature, with current lagging voltage. The primary current generates a varying primary flux $\phi$, which induces the voltage into $L_s$. The secondary induced voltage $e_{in}$ drives a current around the tuned secondary loop, which in turn develops three reactive drops. The capacitive drop is not directly used to drive the diodes so it is ignored for the present. The other two drops are $e_t/2$ and $-e_t/2$, which are equal in magnitude but are in phase opposition with respect to the center-tap. So far, the actions are exactly as explained for the Foster-Seeley

circuit. The tertiary coil is tightly coupled to the primary winding, so a voltage is induced in $L_t$. The voltage induced in it is given by (for $k = 1$)

$$e \text{ across } L_t = \frac{e_p N_t}{N_p} = e_p \sqrt{\frac{L_t}{L_p}}, \tag{10–16}$$

where the tertiary voltage is equal to the primary voltage multiplied by the turns ratio. The ratio of $e_t$ to $e_p$ will be referred to as $a$, so that the tertiary-coil equivalent generator is then $ae_p$. Approximately one-sixth of the primary voltage appears across the tertiary coil, all equivalent driving voltages being shown in Fig. 10-35. This $ae_p$ voltage is used, in conjunction with the half-

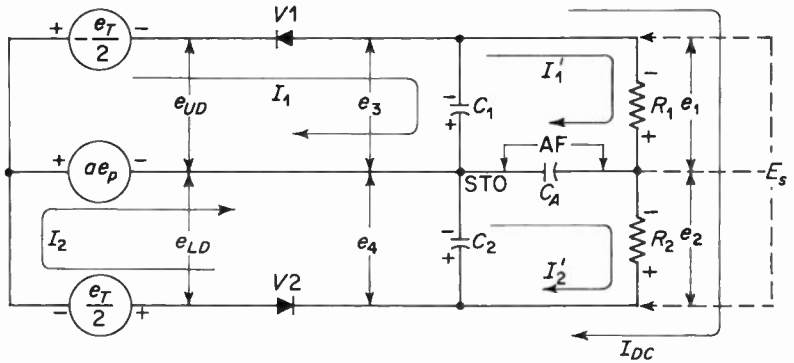



**Fig. 10-35.** Driving circuit (equiv.) for Fig. 10-34. Assume that $R_1R_2$, $C_1C_2$, and $|e_T/2| = |-e_T/2|$. $a$ = fraction of primary voltage appearing across tertiary coil $L_T$.

secondary voltages $e_t/2$ and $-e_t/2$, to drive the diodes in a manner very similar to the Foster-Seeley detector. The reader should recognize that the diagrams shown in Fig. 10-36 are essentially the same as those in Figs. 10-8, 10-9, and 10-10.

Figure 10-36(a) shows the resonant condition (carrier frequency at center position) where $i_s$ and $e_t$ are in phase owing to a resistive secondary tuned circuit. Thus, the vector sum of $e_t/2$ and $ae_p$ drives the lower diode, while the vector sum of $-e_t/2$ and $ae_p$ drives the upper diode.

Therefore,

$$e_{ud} = ae_p - \frac{e_t}{2} \tag{10–17}$$

and

$$e_{ld} = ae_p + \frac{e_t}{2}. \tag{10–18}$$

Since both voltages are equal in magnitude, the rectified currents $I_1$ and $I_2$ must be equal, since the diode loads are symmetrical and equal. Under the condition that the stabilizing voltage $E_s$ is held constant, both load capacitors

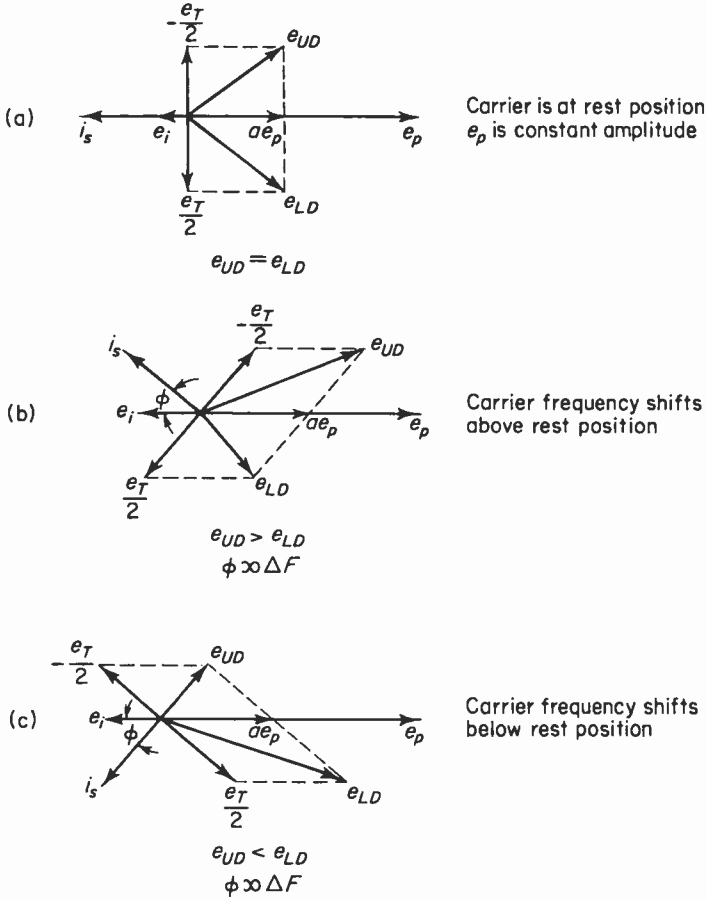


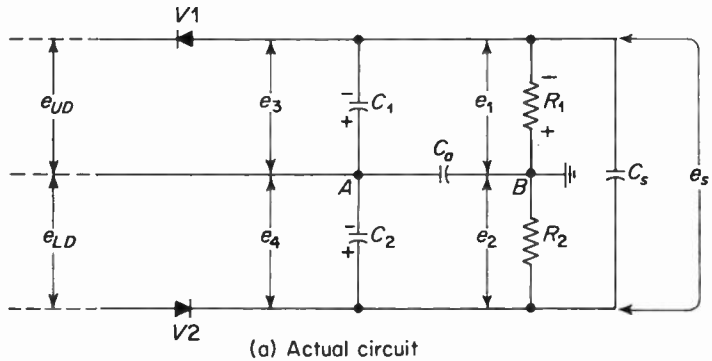

**Fig. 10-36.** Phasors for three conditions of modulation and $A$ constant amplitude FM carrier.

($C_1$ and $C_2$) charge to the peak values of $e_{ud}$ and $e_{ld}$, respectively. The current flowing through the audio load impedance $C_a$ is made up of two components ($I_1'$ and $I_2'$) and these two currents flow in *opposite* directions. Since $e_{ud} = e_{ld}$ when the deviation is zero ($\Delta F = 0$), these two currents are equal in magnitude and therefore the differential current ($I_1' = I_2'$) must be zero. Since the audio output voltage is ($I_1' = I_2'$)$Z_a$, the output voltage is zero for a condition of zero frequency deviation.

If the carrier frequency instantaneously swings *upward* ($+\Delta F$), then it is seen that $e_{ud}$ exceeds $e_{ld}$, the upper diode is driven harder, and $C_1$ charges to a higher potential than $C_2$. Thus, $I_1'$ is greater than $I_2'$ and the differential current through $C_a$ develops an output voltage $E_{ab}$. This voltage is positive to ground with a magnitude proportional to the amount of carrier deviation.

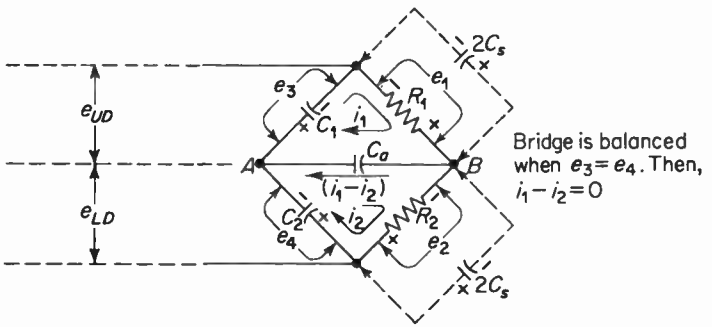
If the carrier frequency momentarily *decreases* below resonance, the secondary tuned circuit appears capacitive. Therefore, the secondary loop current leads the induced voltage and, at the same time, the half-secondary voltage drops ($e_t/2$ and $-e_t/2$) shift off-normal as shown in Fig. 10-36(c). This results in a larger driving voltage for the lower diode, which in turn causes $C_2$ to charge to a higher potential than $C_1$. The differential current in $C_a$ now reverses direction and charges $C_a$ in the opposite direction. The voltage across $C_a$ is now negative to ground and again proportional to the amount of carrier deviation.

The audio recovery for the ratio detector is thus similar to that of a Foster-Seeley, except that the magnitude of the audio voltage for the ratio detector is half† as much for similar conditions (see Fig. 10-37). The reason




(a) Actual circuit

$R_1=R_2$, $C_1=C_2$, $e_3+e_4=e_1+e_2=e_s=$ constant




(b) Bridge equivalent

$C_s$ can be represented by $2C_s$ across each load resistor.

$$\text{Audio output} = e_{AB} = \frac{e_4 - e_3}{2}$$

**Fig. 10-37.** Bridge circuit for derivation of audio output.

†See Appendix 10-1.

for calling this circuit a *ratio* detector is that variations in the detector output, which occur as the carrier deviates, take place as a result of the fact that the *ratio* of $e_3$ to $e_4$ varies while the sum voltage ($e_3 + e_4$) remains constant over a considerable range.

The waveshapes taken across several selected portions of the ratio detector are shown in Fig. 10-38. These curves are based on a 60-Hz sweep rate with the center frequency set at 10.7 MHz. These waveshapes indicate normal performance and can thus be used as a troubleshooting aid. It is assumed that the carrier input is devoid of spurious AM. It should be noted that if the stabilizer

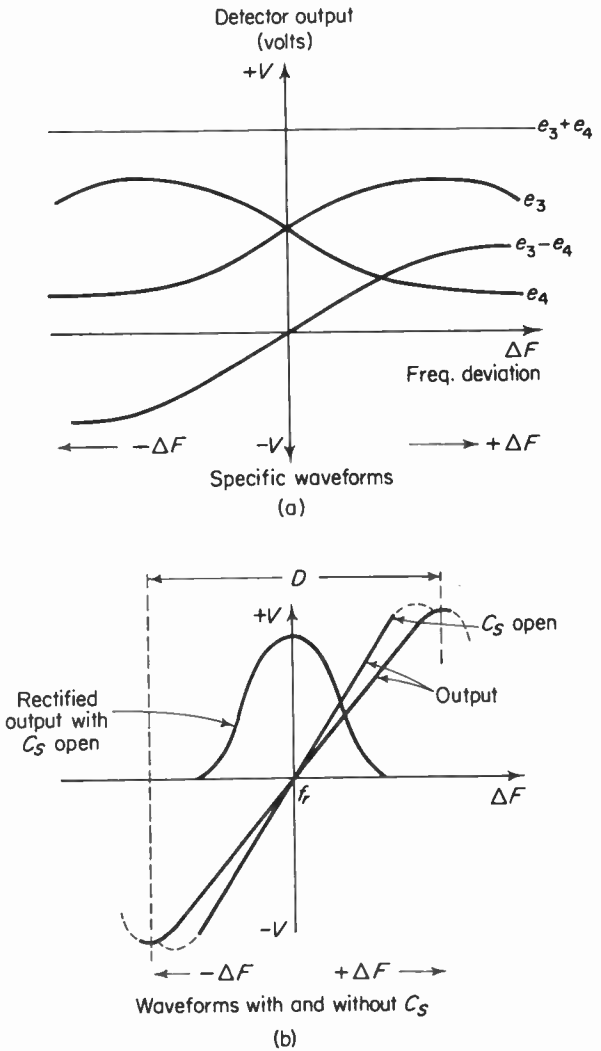



**Fig. 10-38.** Ratio detector wave-forms.

capacitor $C_s$ is open [Fig. 10-38(b)], a slight reduction in the peak separation $D$ takes place but the quality of the "S" curves is retained. This reduction in peak separation is due to the fact that the loading on the primary side of the detector transformer is somewhat reduced, causing an increase in primary $Q$ and a slight decrease in bandwidth. This, in turn, results in a slight increase in the slope of the "S" curve. This raises the detector sensitivity and gives a slightly greater output for a given carrier deviation.

These changes are generally not of sufficient magnitude to change the behavior of the ratio detector in the presence of a station carrier and, unless the input contains noise, the detector circuit as a whole will be unaffected. Since the ratio detector is essentially two diodes in series feeding a common load, it can be readily converted into an AM detector by simply opening $C_s$. If this is done, then the load discharge time constant becomes approximately 5 to 10 $\mu$sec. Thus, if an FM signal is coupled to one of the preceding IF or limiter stages and an oscilloscope is placed across both load resistors ($R_1$ and $R_2$), the result will be the "rectified output" shown in Fig. 10-38(b). The pattern is representative of an IF response, since the diodes are connected (with respect to the load circuit) series-*aiding* and therefore the voltage across both resistors is $e_1 + e_2$. The curve is not, strictly speaking, an IF response, since it must include the effects of the detector transformer. However, its behavior is close enough to an IF response for use in alignment of the IF (and limiter) tuned circuits.

The previous discussion was based on the fact that the carrier contained no amplitude modulation. This is not true in practice, since AM can appear at the input to the detector from a number of sources. Man-made noise (such as auto ignition), atmospheric noise (static), and the selectivity of the tuned circuits ahead of the detector all contribute to AM. None of the above-noted signals bears any relationship to the *original* modulation, and up to this point these disturbances can be eliminated by means of limiters. A ratio detector is self-limiting to a large degree; therefore, under several circumstances, its AM rejection properties are considered adequate.

## 10-14. Ratio-detector limiting

The principle of self-limiting by a ratio detector is based upon the concept of *variable* damping or loading imposed on a tuned circuit. The reader will recall that this is precisely what occurred for the dynamic limiter already discussed.

An important difference between the dynamic limiter and the ratio detector is shown in Fig. 10-39. Here, it is seen that if the dynamic limiter diode is cut off on rapid *downward* AM, limiting fails but there still will be an output. On the other hand, if the ratio detector diodes are cut off, there will be no output—or at best a very distorted output. There will be a charge redistribu-

tion not controlled by the original modulation between the three capacitors, and thus the output is distorted.

The ratio-detector diodes are cut off on *downward* AM because the applied carrier falls below the reverse bias voltage $E_s$. This condition is referred to as 100 per cent stabilization, since the *entire* rectified output is used to reverse-bias the diodes. If resistors $R_3$ and $R_4$ [Figs. 10-39(b) and 10-40] are now inserted, we see that only 84.5 per cent of the detected voltage is stabilized (13,600/16,100 = 0.845). Variable damping can now reduce input AM (either upward or downward) without diode cutoff. The degree of *down-*

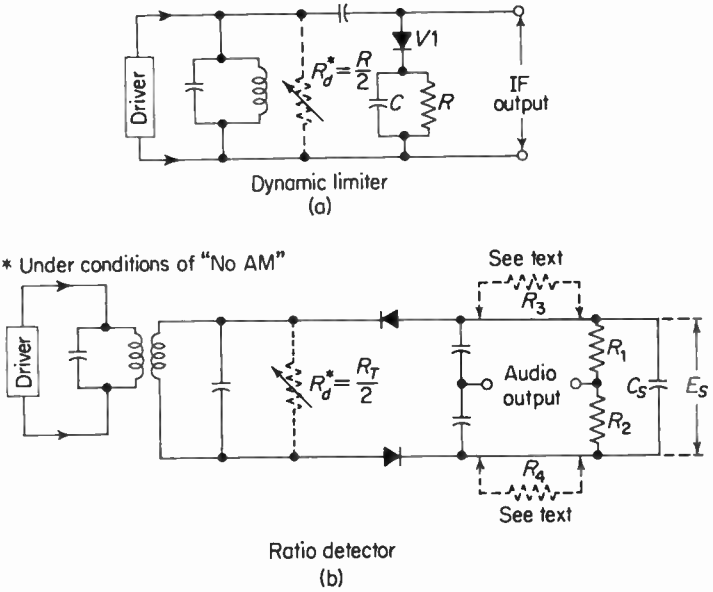



**Fig. 10-39.** The AM rejection mechanisms for the dynamic limiter and ratio detector are essentially identical.

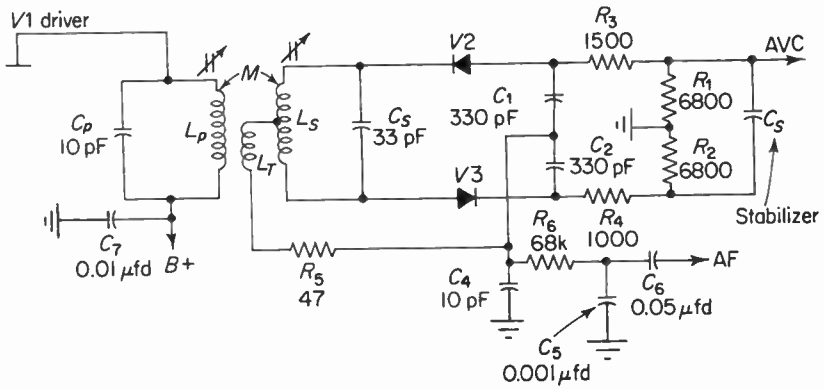



**Fig. 10-40.** Practical form of balance ration detector.

*ward* AM rejection is, as in the dynamic limiter, a function of the unloaded $Q$'s of the primary and secondary windings, since the maximum driver gain cannot exceed the gain without diode loading. The unloaded transformer $Q$ should be as high as possible (consistent with bandwith and *stable* gain requirements of the driver stage )in order for the gain to rise on *downward* AM.

Thus, we see that variations in $R_d$ (with signal amplitude) produce variations in the tuned-circuit $Q$ in a direction to reduce the AM.

The load discharge time constant $(R_1 + R_2)C_s$ should be long compared to the period of the *lowest* expected envelope change. This will produce a steady diode bias and permit variations in $R_d$.

The property of dynamic limiting is based on the fact that the stabilized voltage remains *constant*. This voltage will not remain constant for disturbances that occur at very low frequencies (say 15 Hz or less). Aircraft flutter occurring at a few hertz will therefore not be suppressed, since the time constant is too short to maintain a steady bias voltage. The audio output will then fluctuate and limiting will cease. It appears that the solution would be to increase the time constant to, say, 0.4 sec in order to make the load circuit even more "sluggish." This provides a very poor tuning characteristic, since in tuning from a strong to a weak signal the weak signal may be "passed over." This is because the stabilized voltage may not fall rapidly enough to prevent diode cutoff and subsequent distortion. A typical value for the load time constant is 0.2 sec. A partial solution to the problem above is the use of an efficient AGC system for varying the IF (and RF) stage gains. This AGC voltage can be readily obtained from the detector load circuit (see Fig. 10-40) and, since it indicates relative signal strength, it can also be used to activate tuning devices.

Resistors $R_3$ and $R_4$ also aid AM rejection on *upward* AM. They are sometimes referred to as compensation or "inefficiency" resistors.† In a ratio detector, as the input signal dynamically increases, the diode current increases and the output voltage decreases. Thus, the circuit is overcompensated. This overcompensation becomes "worse" as the diode efficiency increases. If we now insert $R_3$ and $R_4$ (in series with the diodes), they effectively act to *reduce* the diode efficiency. Now, with an increase in input signal, the voltage drops across these resistors increase the effective diode bias and compensate for the reduced output due to increased input.

Thus there is an optimum operating level, above which is produced a *reduction* in output with increase in input (overcompensation), and below which is produced an *increase* in output with increase in input (undercompensation). The choice of $R_3$ and $R_4$ is then a compromise, since the operating level can vary.

†B. D. Loughlin, "The Theory of Amplitude-Modulation Rejection in the Ratio Detector," *Proc. IRE*, **40**, 293 (March 1952).

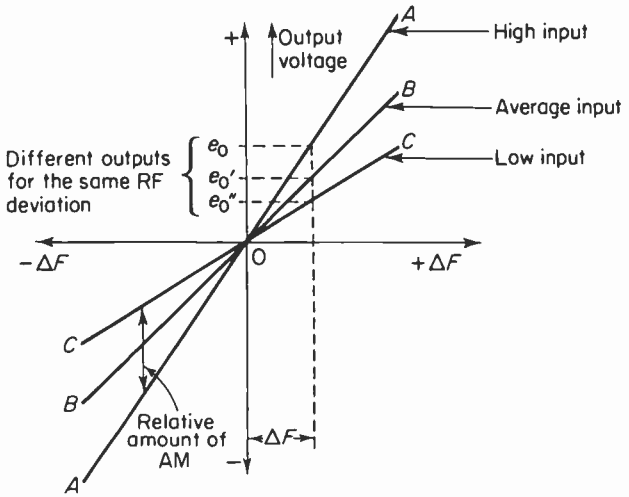



**Fig. 10-41.** Effect of AM on the output of a ratio detector due to variations in the slope.

Since the ratio detector rejects noise by varying the tuned-circuit $Q$, it follows that the slope of the "S" curve must change with variations in the input voltage. This is shown in Fig. 10-41. Curve $A$ is for a large input and curve $C$ is for a small input. If the carrier is at center, the spurious AM output will be zero, since it is assumed that the detector is balanced and the outputs cancel. If the carrier is not at center, then there will be different amounts of audio output voltage for the same carrier deviation. The magnitude of the amplitude variations is thus proportional to the amount that the carrier shifts off center. Since the amounts of spurious AM introduced in this manner are the same for equal increments in the carrier deviation, it is referred to as the "balanced AM component."

By careful adjustment of $R_3$ and $R_4$, the effects of the balanced component can be minimized. If $R_3 + R_4$ is increased, variations in tuned-circuit loading will be reduced due to a given amount of AM. This will permit $ae_p$, $e_t/2$, and $-e_t/2$ (Fig. 10-35) to have their magnitudes changed more and their phase position changed less. Decreasing $(R_3 + R_4)$ will have the opposite effect, permitting their magnitudes to change less and their phase positions to change more. This concept is shown in Fig. 10-42 and explains the effects of changes in the phase sensitivity of the ratio detector. The solid phasors refer to the applied voltages *without* AM present. The dotted phasors refer to the applied voltage *with* AM present. The carrier deviation in the absence of noise causes a shift in $e_t/2$ and $-e_t/2$ by an amount $\phi$. For the same carrier deviation, this time with AM present, $e_t'/2$ and $-e_t'/2$ are larger than for the previous case, but the phase angle $\phi'$, for the same carrier shift, is smaller. The change in phase angle is just enough to compensate for the increased

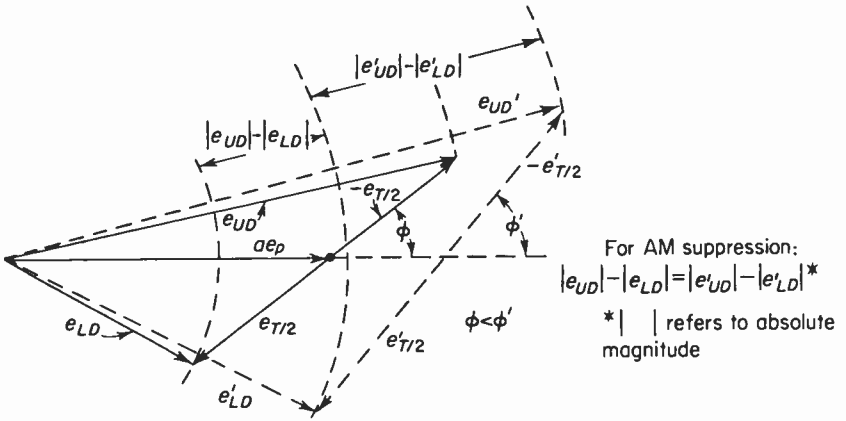



**Fig. 10-42.** This shows how, in a ratio detector, increases in the applied voltage $ae_p$, $e_t/2$ and $-e_t/2$ are prevented from changing the detector voltage, $e_{ud}-e_{ld}$.

length of $e'_t/2$ and $-e'_t/2$ and therefore results in maintaining the audio output voltage constant. Thus, there is an optimum value of $(R_3 + R_4)$ for which AM of the input signal will not affect the output.

There is another spurious output of a ratio detector that is of constant amplitude and independent of carrier deviation. It is known as the "unbalanced component of AM" and is due to variations in the diodes' input impedance over the operating cycle (this variation of input impedance varies with the signal amplitude), insufficient stored energy in the tuned circuits feeding the diodes, and insufficient IF bypassing of the detector load. Any of these conditions plus any unbalance in the transformer, diodes, or circuit constants may result in a detuning of the secondary winding. If, for example, the diode input impedances are not equal, the transformer reactive loadings are not the same across both halves of the secondary and this will result in an effective shift of the secondary center-tap away from its true electrical center. Figure 10-43 shows the effect on the detector characteristic of secondary detuning due to one of the above causes. Figure 10-44 shows the combined effect of both the balanced and unbalanced components of AM. In order to reduce the effects of the unbalanced component of AM, resistors $R_3$ and $R_4$ are made unequal. This adds a component to the output voltage that can be made to cancel the unbalanced component due to the reactive unbalance. This is accomplished by adjusting $R_3$ and $R_4$ so that the unbalanced component is reduced to zero with simultaneous FM and AM applied. The *sum* of the resistors must be held constant during this adjustment in order to minimize the balanced AM component. This, then, is the reason for stabilizing only part of the diode load. If the diode load were entirely stabilized by $C_s$, the differential drops across $R_3$ and $R_4$ would *not* vary with AM and no such cancellation could take place. In addition, a small resistor ($R_5$ of Fig. 10-40) is connected in series with the tertiary winding for the purpose of

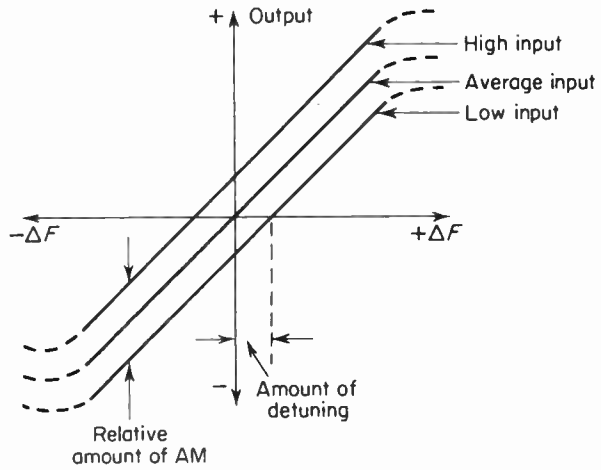



**Fig. 10-43.** Effect on the output of a ratio detector when AM causes secondary detuning. This is the unbalanced component of AM.

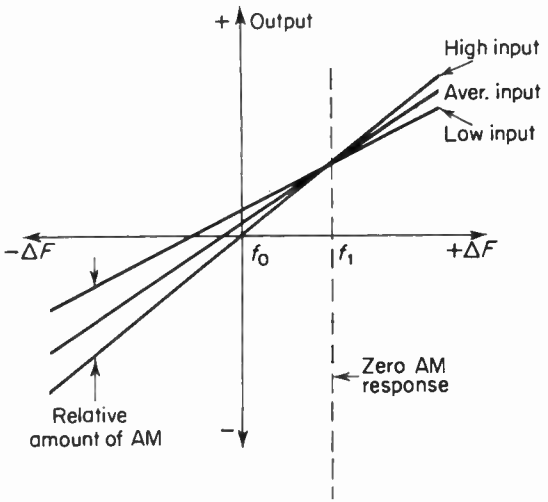



**Fig. 10-44.** The combined effect of balanced and unbalanced components of AM. The point of zero AM response has been shifted to a point other than the center frequency.

modifying the peak diode currents to aid in reducing the unbalanced component of AM at high signal levels.

## 10-15. Specific design criteria

It has been mentioned that the primary reference voltage $ae_p$ is obtained by close-coupling the tertiary coil to the primary winding and that approximately one-sixth of the primary voltage is induced in $L_t$. Thus, the turns

ratio, secondary to primary, is 1 to 6. In this case the "secondary" is the tertiary winding. If the coefficient of coupling $k$ is unity (which is essentially the case here), then the equivalent loading on the primary can be deduced. The loading on the entire secondary, assuming ideal diodes, is equal to $R_d = (R_1 + R_2 + R_3 + R_4)/2 = 9000$ ohms (approx.). This is equivalent to 4500 ohms connected across each half-secondary. The tertiary coil is represented by an $ae_p$ generator, which then "sees" two 4500-ohm resistors in *parallel*. The loading on the tertiary coil is therefore equal to 2250 ohms. Since unity coupling exists between the primary winding and the tertiary coil, the following expression is valid for the resistance reflected back in parallel with the primary winding:

$$Z_p = Z_t\left(\frac{N_p}{N_t}\right)^2, \qquad (10\text{–}19)$$

where

$Z_p$ = the resistance reflected in parallel with the primary winding,
$Z_t$ = the damping resistance imposed on the tertiary winding,
$N_p$ = the number of primary turns,
$N_t$ = the number of tertiary turns.

Thus

$$Z_p = (2250)\left(\frac{6}{1}\right)^2 = 79{,}000 \text{ ohms (approx.)}.$$

This relatively high value of shunt damping resistance is necessary in order to maintain high primary sensitivity. This is why a tertiary coil with relatively few turns is used. The conditions above are shown in Fig. 10-45. The very fact that $R_d$ *varies* with AM accounts for the noise-rejection properties of the ratio detector.

It was also mentioned that the tertiary coil is wound at the $B^+$ or "cold" end of the primary coil. This minimizes the possibility of capacitively coupling IF voltage from the primary to the tertiary coil. The IF potential at this end of the primary coil is almost zero; therefore there will be little transfer of energy via the capacitance between windings. If additional energy should be transferred via this path, it would be equivalent to additional damping and a *reduction* in the primary damping resistance. This would reduce the primary sensitivity and make a proper impedance match difficult.

The total primary tuning capacitance, including stray wiring, is about 15 pfd. The primary *unloaded* $Q$ is approximately 70 and falls to about 40 under average loading conditions. The primary $Q$ should be as high as possible consistent with the proper bandpass and the stable gain of the preceding stage. If the $Q$ is made too high, the maximum stable gain of the preceding stage may be exceeded and oscillations might result. The secondary winding is wound in a "bifilar" arrangement. Thus, the "slug" penetrates each

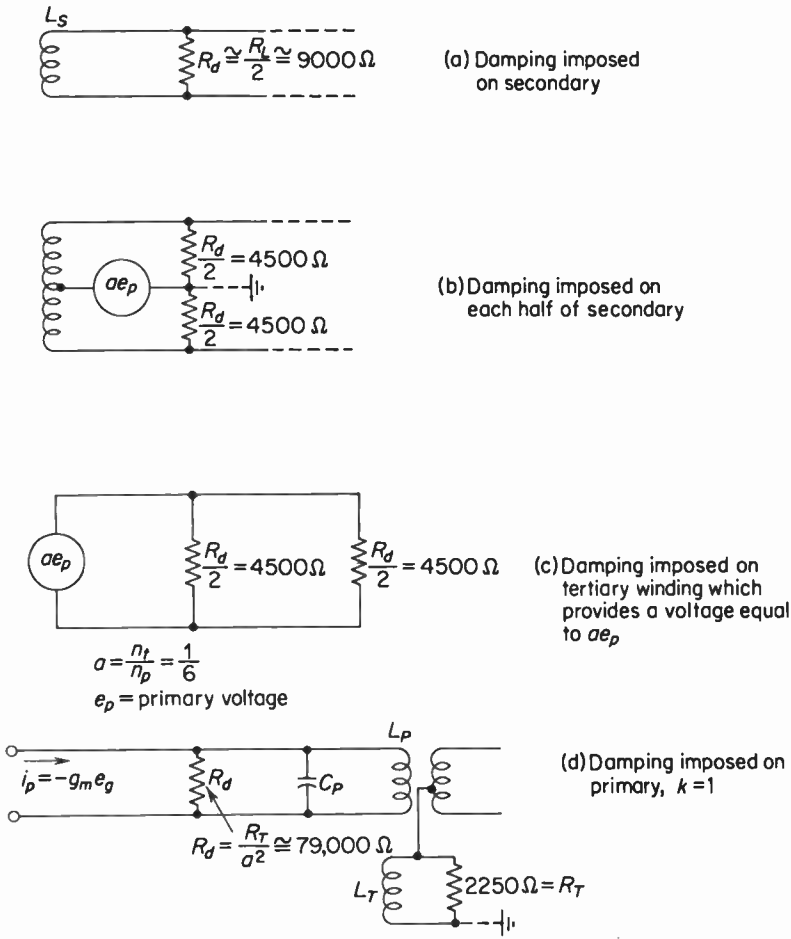



**Fig. 10-45.** Equivalent circuits used to derive damping imposed on both secondary and primary windings.

half of the winding by the same amount and, as a result, variations in the position of the slug do not unbalance the two halves of the secondary winding. This type of construction assures uniform coupling of each half to the primary and good coupling between the half-secondaries. The secondary tuning capacitance is approximately 50 pfd (including wiring), and the secondary *unloaded* $Q$ is approximately 100. The secondary dynamic impedance is given by

$$Z_s = QX = (100)(330) = 33{,}000 \text{ ohms (approx.)}.$$

The *loaded* or working $Q$ is determined as follows. Since the total equivalent secondary damping resistance is 9000 ohms, this appears in parallel with

$Z_s$, which is approximately 33,000 ohms. Thus, the secondary impedance under load is

$$Z_{sl} = \frac{(33{,}000)(9000)}{33{,}000 + 9000} = 6700 \text{ ohms (approx.)}.$$

The working or *loaded* secondary $Q$ is then

$$Q' = \frac{Z_{sl}}{X} = \frac{6700}{300} = 20 \text{ (approx.)}.$$

The diode load resistance is generally selected to reduce the loaded secondary $Q$ to a value of one-fourth or less of its unloaded $Q$. If this resistance is further reduced, the *downward* AM capabilities of the detector will be improved with a reduction in sensitivity.

It was previously mentioned that a ratio detector is not immune to long-term changes in the carrier level due to aircraft reflections and the like. However, since the stabilizer capacitor $C_s$ provides a voltage proportional to signal strength, this voltage (after adequate filtering) could be used to AGC-control one or more stages and hence minimize variations in the audio output. This voltage could also be used in alignment and troubleshooting, since it is proportional to incoming signal strength. Figure 10-46 shows a typical circuit arrangement for obtaining both AGC tuning-indicator voltage. Note the decoupling networks required to minimize possible feedback from one stage to another via a common power-supply impedance.

It should be apparent that the bandwidth of a ratio detector is, in general, much greater than that of a Foster-Seeley discriminator. The wider passband is due primarily to the large amount of damping imposed on the detector transformer, which in turn is required for efficient noise rejection. A typical Foster-Seeley discriminator is of the order of 400 kHz (peak separation), while that of a ratio detector is approximately 600 kHz (peak separation). Very wide-band commercial designs employ ratio detectors having peak separations of 1 MHz or more.

An important aspect of the ratio detector concerns its output voltage

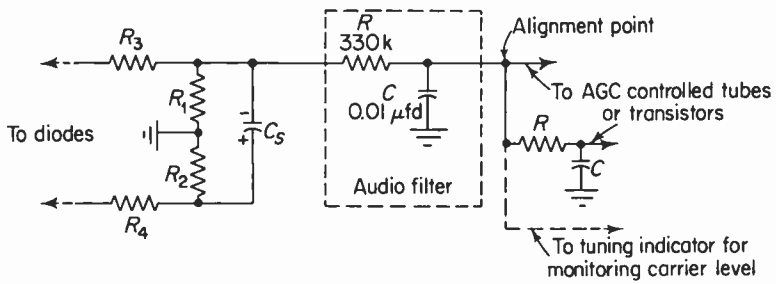



**Fig. 10-46.** Schematic showing access point for obtaining AGC or tuning eye voltage.

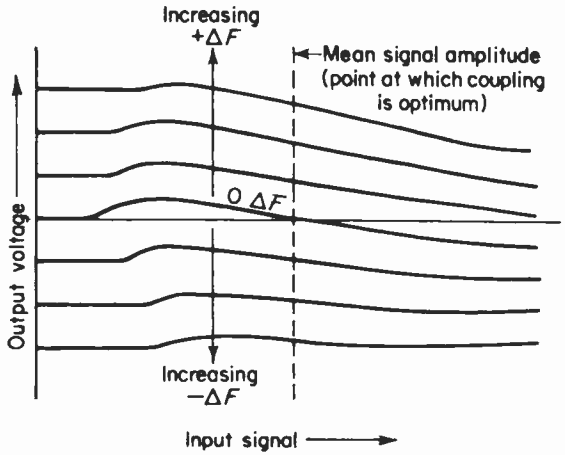



**Fig. 10-47.** Typical input-output voltage curves for a ratio detector. The variations of the zero deviation curve with changes in RF input are due mainly to the development of harmonics at the IF in the stages preceding the detector and variations in the input reactances of the diodes.

versus input voltage for a given deviation. Figure 10-47 shows typical curves that indicate output variations beyond some arbitrary signal level. Experiments with the ratio detector† have shown that amplitude variations in the primary of the detector transformer were such that an increase in applied signal level produced an increase in damping of the secondary and an *increase* in primary impedance. This, in turn, altered the primary-secondary phase relationship and thus resulted in a reduction of output. This situation is analogous to the case of overcompensation previously mentioned. If the input-signal amplitude could be made more constant, the phase relationships would remain fixed (except for frequency deviations). Therefore, this would provide a more uniform output. The stage preceding the detector is then often operated as a saturated limiter on strong signals. On weak signals it will operate as a linear IF amplifier.

A number of very wide-band broadcast tuners use limiters to drive the detector. Since, in a wide-band circuit, heavy tuned-circuit damping is employed, the primary $Q$ (and sensitivity) is very low. Under these conditions, the primary working $Q$ will not be determined by the loading‡ and it will be difficult for the gain of the preceding stage to rise when the carrier falls. Thus, a wide-band ratio detector cannot effectively reject large amounts of downward AM. Virtually all wide-band ratio detectors are therefore preceded by at least one stage of efficient limiting.

†RCA License Bulletin, LB-645, "Ratio Detectors for FM Receivers," S. W. Seeley.

‡This is because the equivalent diode loading resistance reflected across the tuned circuit is comparable to its dynamic impedance.

## 10-16. Unbalanced ratio detector

A modification of the ratio detector is shown in Fig. 10-48 together with representative component values for an IF of 10.7 MHz. The equivalent circuits for modulation recovery are shown in Fig. 10-49. The differential current $i_d$ flowing through $C_a$ is zero when the carrier deviation is zero, since the diode driving voltages are equal. When the carrier deviates off normal, one diode is driven harder than the other, resulting in a differential current in $C_a$. This develops the audio voltage across $C_a$, which is then applied to a deemphasis network consisting of $R_d$ and $C_d$. The diode load time constant is approximately 0.18 second, which is adequate for most applications. The

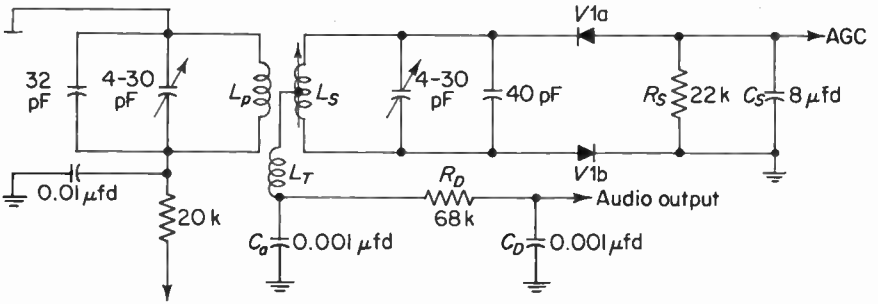
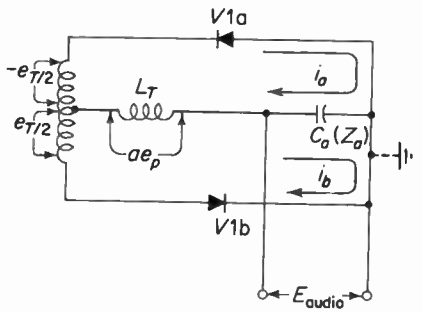


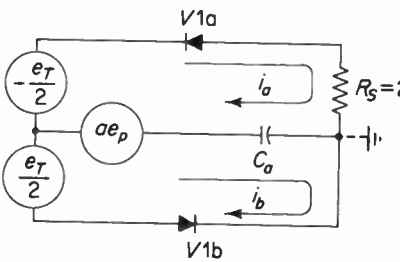

**Fig. 10-48.** Unbalanced ratio detector.




Anode of V1a is at IF ground potential

(a) Equivalent circuit for audio recovery

| IF $\Delta F=0$, | IF $\Delta F\neq 0$, |
|---|---|
| $i_a=i_b$ | $i_a>i_b$ |
| $i_D=i_a-i_b=0$, | $i_D=i_a-i_b\neq 0$, |
| $E_{AF}=i_DZ_a=0$ | $E_{AF}\neq 0$ |




(b) Circuit showing the effect of an open $C_S$. Intermediate frequency return path for V1a is thru a 22 k resistor.

**Fig. 10-49.** Equivalent circuits for unbalanced ratio detector.

magnitude of voltage obtained from this circuit is twice that of the balanced variety, since it is derived from the *entire* diode load instead of from only one of the two split resistors. For the same value of diode load time constant, the balanced detector will reject low-frequency AM more effectively than the unbalanced. The audio output is the same for both circuits, with the unbalanced circuit being somewhat more economical.

An examinaton of Fig. 10-49(b) shows that if the stabilizer $C_s$ should fail to open, the return paths for both diodes would be through different impedances. This is because the large stabilizer capacitor, under normal conditions, is considered to be a zero impedance for the fundamental frequency of diode current. The return path for the upper diode is now through a 22-k resistor, and this will distort the detector response and affect noise rejection. This is not the case for the balanced detector, since the diode load impedance is still balanced to ground. Practically no change in audio quality will result but noise rejection will, of course, be poor.

A point worth noting about the unbalanced ratio detector concerns the DC potential at the sound take-off point with respect to ground with zero carrier deviation. For the balanced ratio detector with zero deviations, the potential at the STO is zero. The equivalent circuit (Fig. 10-50) shows that the voltmeter is effectively connected across the lower diode. Assuming that the DC resistance of the coils is zero and that the diodes are replaced by two resistors, the voltmeter will indicate exactly one-half the stabilizer voltage, provided the circuit is properly aligned. Thus, this form of detector cannot directly activate a conventional "center-of-channel" tuning indicator, since there is no change of polarity with carrier deviations.

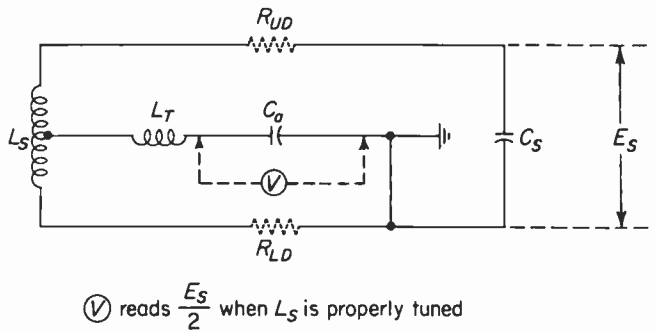
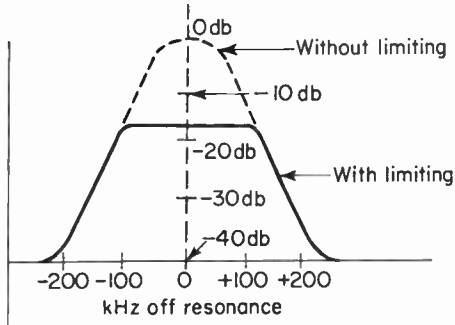


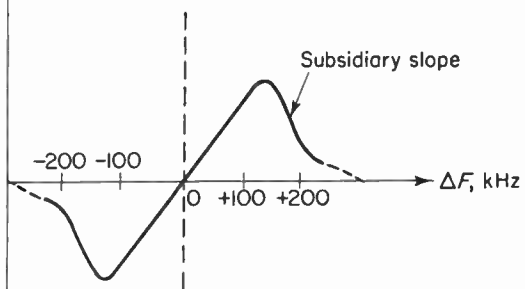

**Fig. 10-50.** DC equivalent circuit under "zero" balance conditions.

## 10-17. Detector side response

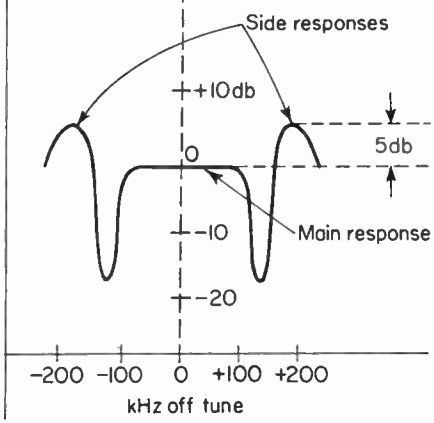
When an FM receiver is detuned, large amounts of AM are introduced owing to the slope detection along the steep IF selectivity curve. In a Foster-Seeley arrangement, the input to the limiter falls below the limiting level and, therefore, the AM cannot be effectively suppressed. A response will thus



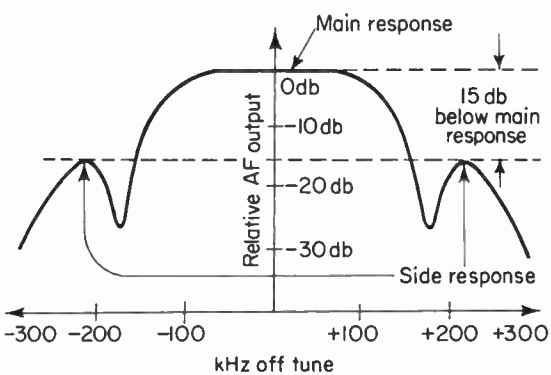

(a) IF response




(b) Detector characteristic with a 300 kHz peak separation




(c) Typical tuning characteristic for a Foster–Seeley detector. Side responses are approximately 5 db higher than main response




(d) Ratio detector tuning characteristic showing side responses about 15 db below main response

**Fig. 10-51.** Side responses for two different types of FM detectors. (a), (b) and (c) are for a Foster-Seeley plus a grid leak bias limiter.

be obtained on either side of resonance. The magnitudes of these responses can be greater than the main response and will depend upon the steepness of the subsidiary slopes of the "S" curve, since output for a given deviation is proportional to detector sensitivity [see Fig. 10-51(c)].

One way of reducing the amplitude of these side responses is to reduce the IF selectivity. This will have the effect of introducing less AM for a given carrier deviation. Another approach is to use a wide-band ratio detector. Also, if the peak separation is wider than the widest part of the IF response, the side responses will fall completely outside of the IF passband. Figure 10-51(d) shows a typical tuning characteristic for a ratio detector and indicates side responses that are approximately 15 db below the main response. This detector thus provides an easier tuning characteristic and relatively little interstation noise, since most receiver noise, between stations, is efficiently rejected by the ratio detector. Also, the effect of AFC is to broaden the receiver tuning characteristic (regardless of the type of detector used), and thus the side responses are effectively reduced.

## 10-18. Summary

**1.** The Foster-Seeley discriminator is a balanced type very suitable for wide-band FM receivers.

**2.** The Foster-Seeley circuit converts carrier-frequency variations into audio frequencies by first converting FM to AM and then converting AM to audio. This is done by means of a phase-shift circuit followed by a pair of balanced peak rectifiers.

**3.** The Foster-Seeley circuit can provide AFC as well as tuning-eye voltage.

**4.** The Foster-Seeley linearity is excellent over the required working range.

**5.** The Foster-Seeley circuit requires prelimiting, since it cannot reject AM at all times.

**6.** The Foster-Seeley circuit is simple to align with only two adjustments to be made.

**7.** Preemphasis and deemphasis are used in FM broadcast and television sound to increase the signal-to-noise ratio for the higher modulating frequencies so that the carrier-frequency deviations are equalized for the entire audio band.

**8.** The standard preemphasis and deemphasis time constant is 75 $\mu$sec.

**9.** The ratio detector is self-limiting.

**10.** Both the ratio detector and the Foster-Seeley discriminator are classed as phase-shift discriminators. Their operation is based upon the behavior of tuned circuits whose impedance characteristic varies with changes of carrier frequency.

**11.** The ratio detector can reject larger amounts of "upward modulation" than "downward modulation."

**12.** The ratio detector can provide AGC, AFC, and tuning-eye voltages.

**13.** The sensitivity of the ratio detector is higher than for a Foster-Seeley circuit, since the minimum threshold level of the ratio detector is limited primarily by diode nonlinearity at very low signal levels.

## REFERENCES


1. Langford-Smith, F.: *Radiotron Designer's Handbook*. Wireless Press, Sydney, Australia 1953.
2. Terman, F. E.: *Electronic and Radio Engineering*. McGraw-Hill Book Company, New York, 1955.
3. Seeley, S., and Avins, J.: "The Ratio Detector," *RCA Review*, June 1947.
4. Johnstone, G. G.: "Limiters and Discriminators for FM Receivers," *Wireless World*, January 1957, February 1957, May 1957.
5. Loughlin, B. D.: "The Theory of Amplitude-Modulation Rejection in the Ratio Detector," *Proc. IRE*, March 1952.
6. Loughlin, B. D.: "Performance Characteristics of FM Detector Systems," *Tele-Tech*, January 1948.
7. Mayo, C. G., and Head, J. W.: "Foster-Seeley Discriminators," *Electronic and Radio Engineer*, February 1958.
8. Mayo, C. G., and Head, J. W.: "Combined Limiter and Discriminator," *Electronic and Radio Engineer*, March 1958.
9. Baghdady, E. J.: "Theory of Stronger-Signal Capture in FM Reception," *Proc. IRE*, April 1958.
10. Sturley, K. R.: "The Ratio Detector—How It Works," *Wireless World*, November 1955.
11. Travis, C.: "Automatic Frequency Control," *Proc. IRE*, October 1935.
12. Johnson, L. W.: "F-M Receiver Design," *Wireless World*, October 1956.
13. Seeley, S. W.: "Ratio Detectors for F-M Receivers," RCA Laboratories, Industry Service Bulletin, LB-645, September 1945.
14. Sturley, K. R.: *Radio Receiver Design*, Parts I and II. John Wiley & Sons, Inc., New York, 1945 (Part I), 1953 (Part II).
15. Baghdady, E. J.: "FM Demodulation Time-Constant Requirements for Interference Rejection," *Proc. IRE*, February 1958.


## APPENDIX 10-1 Ratio detector output voltage

Based upon the bridge circuit of Fig. 10-37b, the output voltage $e_{AB}$ is derived as follows:

Taking the mesh currents,

$$(i_1 \text{ mesh}), \quad i_1\left(R_1 + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_a}\right) - \frac{i_2}{j\omega C_a} = 0 \tag{1}$$

$$(i_2 \text{ mesh}), \quad i_2\left(R_2 + \frac{1}{j\omega C_2} + \frac{1}{j\omega C_a}\right) - \frac{i_1}{j\omega C_a} = 0 \tag{2}$$

Factoring $(i_1 - i_2)$ from each of the above equations gives

$$\frac{i_1 - i_2}{j\omega C_a} + i_1 R_1 + \frac{i_1}{j\omega C_1} = 0 \tag{3}$$

and

$$\frac{i_1 - i_2}{j\omega C_a} - i_2 R_2 - \frac{i_2}{j\omega C_2} = 0 \tag{4}$$

Let $(i_1 - i_2)/j\omega C_a = e_{AB}$, $i_1/j\omega C_1 = e_3$, $i_2/j\omega C_2 = e_4$ and $i_1 R_1 = i_2 R_2$. Substituting these identities back in Eqs. (3) and (4) and solving each equation for $e_{AB}$, we have

$$e_{AB} = -i_1 R_1 - e_3 \tag{5}$$

and

$$e_{AB} = i_2 R_2 + e_4 \tag{6}$$

Adding Eq. (5) to Eq. (6), we have

$$2e_{AB} = e_4 - e_3 \tag{7}$$

or

$$e_{AB} = \frac{e_4 - e_3}{2} \tag{8}$$

This shows that the audio output voltage from a ratio detector is half that of a Foster-Seeley discriminator.

# 11

# QUADRATURE-TYPE DETECTORS FOR TELEVISION SOUND

## 11-1. The 6BN6 gated-beam detector

The 6BN6 tube is extensively used as a sound detector in TV receivers. It can provide efficient limiting, FM demodulation, and audio voltage amplification all within a single tube with no need for a relatively complicated tuned transformer. Its input circuit is tuned by a single coil (either by a physical capacitor or the input capacity of the tube); therefore it is a very economical arrangement. By itself, it can be used as a broad-band limiter followed by a wide-band detector and will provide excellent overall performance.

A description of the physical construction of the tube will help make its operation clearer. Figure 11-1 shows two views of the tube with the important parts labeled. The tube consists of (a) two control grids—one called the limiter grid and the other the quadrature grid, (b) an indirectly heated cathode, (c) an accelerator (similar to a screen grid in a conventional pentode), (d) an anode or plate, and (e) an electron lens system. The limiter grid (the one closest to the cathode) and the quadrature grid are well shielded from each other. This results in a very low value of interelectrode capacitance and greatly minimizes certain problems of instability. The electron beam is formed as follows. Aperture *A* and the first focusing electrode shape the

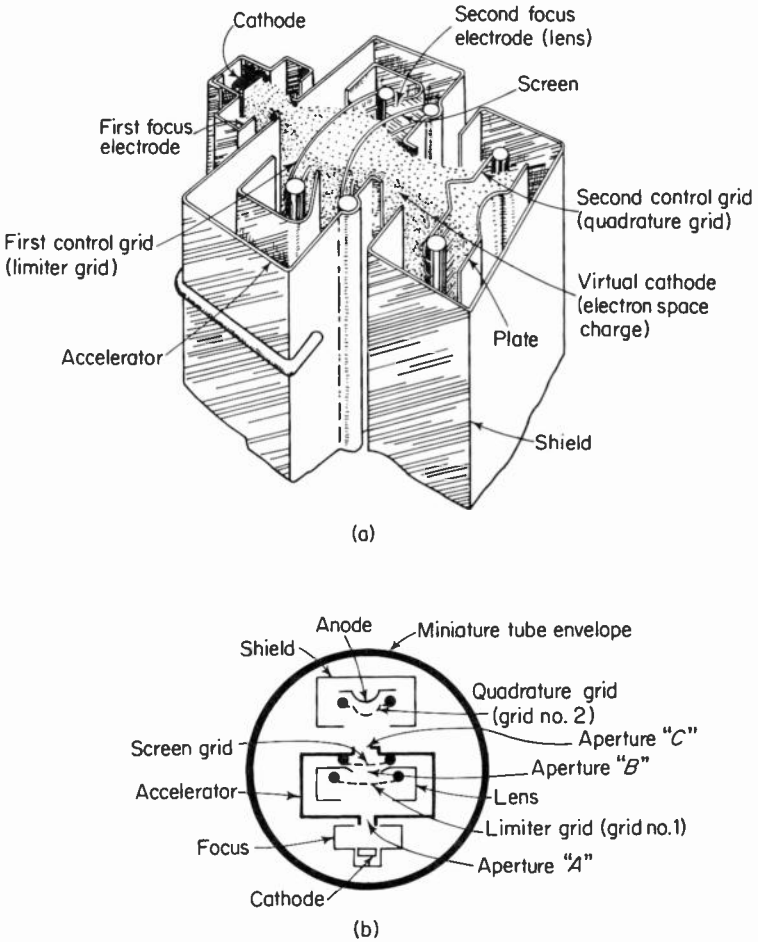



**Fig. 11-1.** Schematic representation of gated beam tube construction, showing position and effects of tube electrodes.

beam and project it to the limiter grid as a thin sheet of electrons with a very high electron density in the vicinity of the aperture. The electrons, in the vicinity of the limiter grid, approach the grid almost perpendicular to it. This results in a very high transconductance and a very steep transfer characteristic, as shown in Fig. 11-2. After passing through the limiter grid, the electron beam is again shaped (converged) by means of the slight curvature of the limiter grid, lens aperture *B*, and the curved screen grid. The beam is now projected through the slot in the shield to the quadrature grid. The quadrature grid also displays a steep transfer characteristic, as shown in Fig. 11-3. The focus, lens, and shield electrodes are connected internally to the cathode and are therefore at the same potential. The accelerator and plate

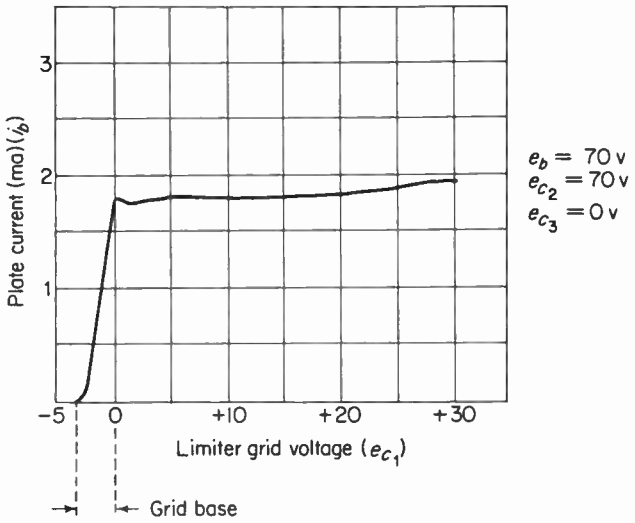



**Fig. 11-2.** Limiter grid-plate transfer characteristic. If tube is biased midway between saturation and cut-off, about 1.25 volts of grid signal would be able to switch the plate current on or off, that is, "gating" the full plate current.

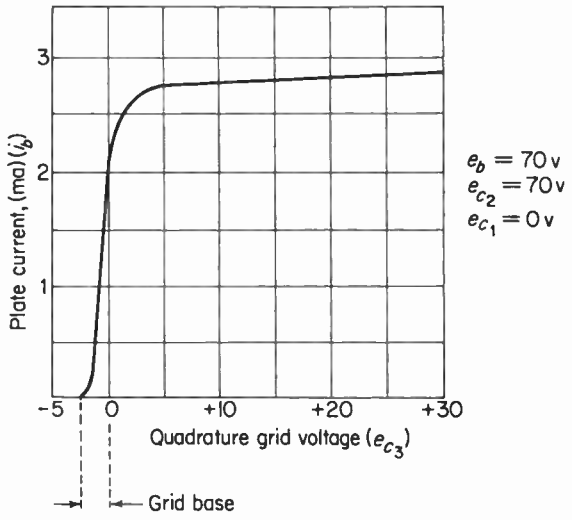



**Fig. 11-3.** Quadrature grid-plate transfer characteristic. The quadrature grid is also capable of gating the full plate current on and off with a 1.25 volt signal applied to this grid.

are connected to a source of positive potential as in a conventional pentode.

An examination of Figs. 11-2 and 11-3 shows that both grids are of the very sharp-cutoff variety with a small dynamic range. The grid base for both

is of the order of 2.5 volts. Thus, very little grid driving voltage is required for plate-current cutoff. Notice, too, that when the grid voltage swings in the positive direction, the plate current rises abruptly from zero to maximum and remains saturated regardless of how much further positive the grid voltage may go. This "step" characteristic should immediately suggest its use as a very efficient limiter.

## 11-2. Gating or switching action of the control grids

Consider two simple switches connected in series with a battery and a lamp, as shown in Fig. 11-4. Basic theory tells us that the lamp $L$ lights only when *both* switches are closed and remains lighted as long as *both* switches remain closed. If one switch is opened and the other switch remains closed, the lamp will go out. This basic action takes place in the plate circuit of the 6BN6 tube in the same manner. If the quadrature grid is biased at, say, +2 volts, Fig. 11-3 indicates that it can "pass" the full plate current provided the limiter grid first "passes" current to the quadrature grid. Let us assume that the quadrature grid is acting like a closed switch and that it is to remain "closed." Also, a sinewave of approximately 2 volts peak is applied to the limiter grid. An examination of Fig. 11-5 shows that the plate-current pulses are of rectangular form, owing to the sharp-cutoff characteristic of the limiter grid. As soon as the limiter-grid voltage swings in the negative direction, plate current is almost instantly cut off. It is assumed that the limiter grid is cathode-biased midway between saturation and cutoff.

A few items might be mentioned at this point. First, when the limiter-grid voltage swings a few volts negative, the electron beam cannot pass and plate current stops (the limiter-grid "switch" is effectively open). The repelled

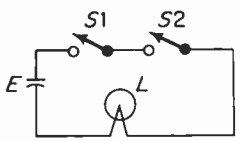



**Fig. 11-4.** Lamp $L$ will light and will remain lighted only if *both* switches are closed. If either or both switches open, lamp will go out.

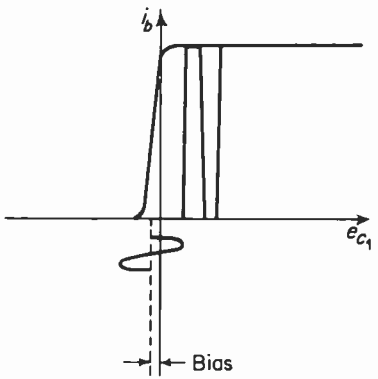



**Fig. 11-5.** Plate current waveform with a sinewave input to limiter grid.

electrons "fall back" toward the cathode. However, because of the small size of aperture *A* [Fig. 11-1(b)] the probability that an electron will find its way back through the narrow opening is small. Also, since the electron density is maximum in the vicinity of the aperture, the dense negative "pile-up" causes the electrons to fall back like a water spray and again miss the narrow opening. These electrons are collected by the accelerator and return to the cathode via the power supply, as shown in Fig. 11-6.

The reader should appreciate that an action is occurring here that does not occur in a conventional vacuum tube: this tube is capable of cutting off its own plate current by means of *cathode bias*. This action is possible because cathode current is equal to plate current plus screen current. In this tube, however, whatever doesn't reach the plate is simply collected by the screen grid. In a conventional pentode, if the control-grid bias is made negative enough to cut off plate current, the cathode current drops to zero. Thus, in a conventional tube, plate-current cutoff requires some method of biasing other than a cathode resistor. This point is important in terms of servicing this type of circuit, since an apparently correct DC voltage read across the cathode resistor does not necessarily mean that plate current is flowing.

Let us assume now that the limiter grid swings a few volts in the positive direction. This permits the beam current to pass through the limiter-grid wires to the quadrature grid. Since the quadrature-grid "switch" was assumed to be closed, the electrons will reach the plate and the external circuit. If we now assume that the quadrature grid is biased below cutoff, then the approaching electrons are repelled from the quadrature grid and again fall back toward a relatively small opening [aperture *C* of Fig. 11-1(b)]. Most

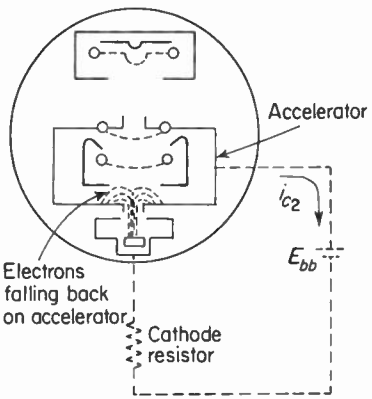



**Fig. 11-6.** Conditions in the vicinity of the limiter grid if its voltage is more than a few volts negative. With plate current cut off, accelerator current still flows through the cathode resistor.

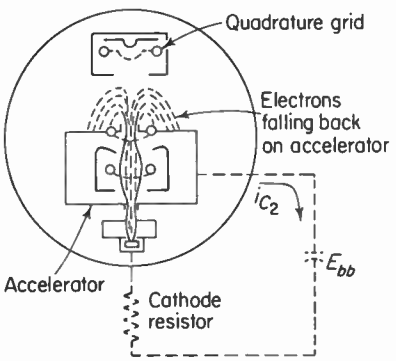



**Fig. 11-7.** Conditions in the vicinity of the quadrature grid if the limiter voltage is a few volts positive and the quadrature grid "switch" is open. With plate current cut off, accelerator current still flows through the cathode resistor.

of these "backward-falling" electrons are collected by the accelerator anode and return to the cathode via the power supply, as shown in Fig. 11-7. If the quadrature grid is again biased above cutoff, these electrons will reach the plate circuit as previously described. Therefore, either grid is capable of switching the plate current on or off and *both* grids must be biased above cutoff before plate current will flow. The speed of switching is limited only by the inertia of the electron beam, which we can consider to be zero. The switching speed will then be taken as "instantaneous." This unique tube is not limited to FM sound applications but can be used in such operations as clipping, limiting, coincidence detection in television, radar, computers, and so on. Our discussion will be limited to its use as a limiter and FM sound detector; other uses have been covered in the literature.

## 11-3. The gated-beam tube as a limiter

In the chapter on limiters we saw that one of the drawbacks of the grid-leak limiter is its relative inability to follow rapid downward modulation, owing to grid-circuit time-constant considerations. For satisfactory operation, the developed grid-leak bias across the grid-leak capacitor must change instantaneously with any sudden change in the amplitude of the

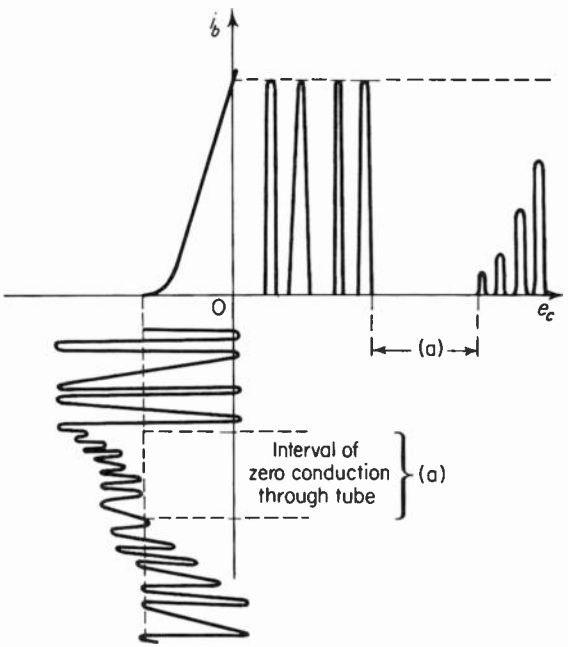



**Fig. 11-8.** A grid leak limiter will "block" on rapid downward modulation unless the grid circuit time constant is made small. The action is analagous to diagonal clipping in a diode detector.

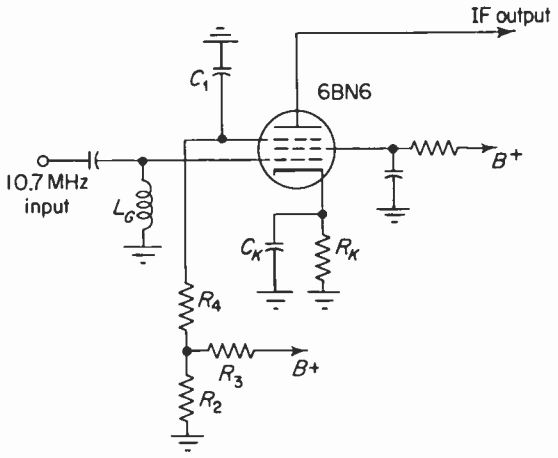



**Fig. 11-9.** Circuit using the gated beam tube as a straight limiter. The coil $L_g$ and input capacity of the tube shunt resonate to form a high input impedance.

applied FM carrier. This is necessary to keep the peak positive signal clamped to zero bias. When the amplitude of the carrier suddenly drops, the capacitor must rapidly discharge through the grid-leak resistor—but, since the amplitude is dropping, it may go below that which is required to maintain conduction through the tube. This is shown in Fig. 11-8. The 6BN6 is free of such problems, since grid-leak bias is not used and therefore the energy-storage problem of a grid-leak capacitor is not involved.

Figure 11-9 is the circuit of the gated-beam tube wired as a straight limiter. To obtain the largest output, the quadrature grid should be connected to the plate (or a source of positive potential). To limit at the smallest possible carrier input, the quadrature grid should be connected to ground. The circuit shown indicates the condition for high output with a slight sacrifice in limiting sensitivity. To understand how this circuit performs as an amplitude limiter, we shall assume that the quadrature grid is acting like a closed "switch," since its potential is positive with respect to the cathode. $R_2$ and $R_3$ comprise a voltage divider with about 4 volts appearing across $R_2$. With about 2.5 volts at the cathode of the tube, the quadrature-grid "bias" is approximately 1.5 volts, and this effectively maintains the quadrature "switch" closed. Now, let us examine the grid-to-plate transfer characteristic with an FM carrier applied to the limiter grid. This is shown in Fig. 11-10.

With sufficient drive at the limiter grid to alternately switch the tube between cutoff and plate saturation, the plate-current waveform is essentially rectangular owing to the "shearing" effect of the limiter grid. Should a sharp burst of noise appear, it will be clipped either by plate-current cutoff (on the negative swing) or by plate saturation (on the positive swing). Thus, the disturbance is rejected by means other than a change in grid bias, and the rejection is independent of capacitor storage effect. Since the noise-rejection

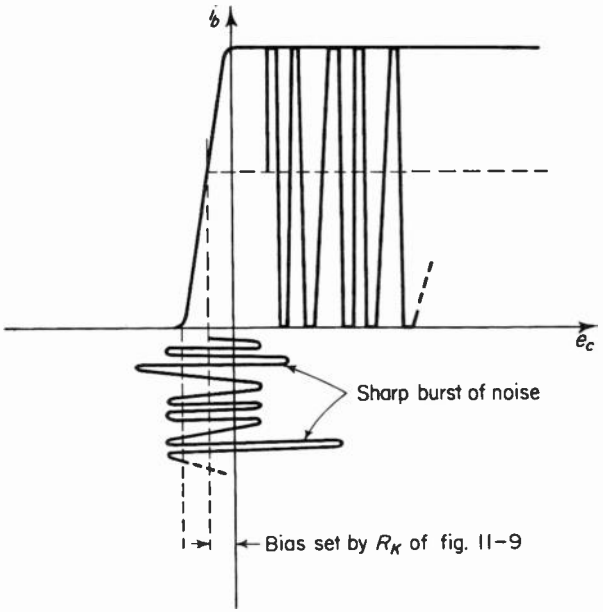



**Fig. 11-10.** Transfer curve of limiter grid for 6BN6 tube with an applied FM carrier. The sharp cutoff characteristic, together with the tube biased midway between saturation and cutoff, provides for excellent limiting of AM.

qualities of this tube do not depend upon the time constant of an *RC* network, the action can be considered as instantaneous.

The grid coil ($L_G$ of Fig. 11-9) is a very low resistance and, therefore, when the grid is driven positive, the charge on the 220-pf coupling capacitor cannot build up any bias since the capacitor discharges through the coil almost instantaneously. Therefore, no bias is developed in the grid circuit of this limiter. Also, the design is such that when the grid is driven positive with respect to the cathode, grid current does flow, but the grid cannot take more than its share of the total space current. The flow of current levels off at about 500 $\mu$a, with about 2 volts of signal applied to the limiter grid. This means that the loading imposed on the preceding circuit is negligible. As a matter of fact, the grid-to-cathode resistance increases with increasing grid signal (above 2 volts), because the grid current saturates while the grid voltage increases. The minimum grid-to-cathode resistance is about 20,000 ohms with a grid voltage of about 1 volt. The approximate voltage amplification of this arrangement is of the order of

$$A_v = g_m Z_L = 3000 \times 10^{-6} \times 10^4 = 30. \qquad (11\text{–}1)$$

Equation (11–1) is valid only under class $A_1$, grounded-cathode conditions. Under actual conditions (above the limiting threshold of about 1 volt), this tube is operating essentially class $AB_2$. Therefore, owing to plate-satura-

tion effects, the voltage amplification will be somewhat less. The calculation above assumes a transconductance of 3000 micromhos and a plate-load impedance of 10,000 ohms.

## 11-4. The gated-beam tube as a combined limiter-detector

In order to understand how this tube demodulates an FM carrier, we must grasp the principles underlying space-charge coupling in a multigrid tube (pentagrid converters, gated-beam tubes, and so on). The basic concept of this phenomenon is that when an electron cloud approaches a conductor comprising "free" electrons, the charge associated with the cloud repels the electrons away from the end of the conductor nearest the space charge. This leaves a charge deficiency at that point resulting in a net *positive* charge. Should the space charge now recede from the conductor, the free electrons will migrate back toward the end of the conductor nearest the space charge. This action is shown in Fig. 11-10, with the electrons in the vicinity of the conductor designated as $q$. As this charged cloud approaches the conductor, it will repel an electron (actually many electrons) $A$ out of the conductor. This small induced current thus develops an emf across $R$. Should the charge $q$ recede, $A$ will migrate back to neutralize the positive charge $B$. Thus, a small oscillatory current could be induced in the conductor, provided the density of the space charge could be varied. This "resistor" ($q$) could be replaced by an impedance such as a coil, capacitor, or even a tuned circuit.

Assume now that the density of the space charge is to be varied in step with an applied voltage on one grid while the space charge is in the vicinity of *another* grid. The applied voltage is sinusoidal and, as indicated in Fig. 11-11, will be applied between grid 1 and the cathode. This voltage will be designated $e_L$. The induced current flowing in the external circuit of the other grid will be designated $i_q$. This second grid is called the quadrature grid.

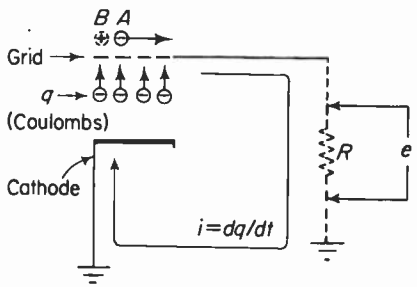



**Fig. 11-11.** Basic mechanism of space charge coupling. A small current $i$ is induced in a conductor by means of an approaching and receding electron charge $q$.

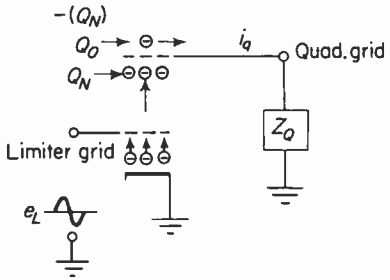



**Fig. 11-11**(a). Diagram showing the electron charge in the vicinity of the quadrature grid. $Z_Q$ takes the form of a parallel tuned circuit.

Grid 1 is called the limiter grid and its applied voltage is the FM carrier ($e_L$). The charge density in the vicinity of the quadrature grid is designated $Q_N$ while the charge *on* the quadrature grid will be destinated $Q_O$.

The coupling mechanism is as follows. Since the cathode can provide an unlimited source of electrons, the tube is said to be operating "space-charge limited." The cathode current is assumed to remain essentially constant even if plate current should stop flowing because of cutoff by either or both grids. Now, as $e_L$ starts to swing in the positive direction, the space-charge density in the vicinity of the quadrature grid increases. The "cloud" becomes more *negatively* charged and, since it repels free electrons, the charge *on* the quadrature grid becomes less negative. At this point, it can be stated that $e_L$ is 180° out of phase with $Q_N$ and in phase with $Q_O$. Also, it can be stated that $Q_O = -Q_N$. As $e_L$ reaches its maximum positive value, $Q_N$ will reach maximum density owing to the accelerating effect of $e_L$. Also, $Q_O$ will reach minimum density because a maximum number of electrons will be repelled out of the quadrature grid into the external circuit. As $e_L$ starts to swing in the more negative direction, fewer electrons will be accelerated from the cathode and the charge density in the vicinity of the quadrature grid will be reduced. Therefore, $Q_N$ becomes less negative and repels fewer electrons out of the quadrature grid. $Q_O$ increases in density or becomes more negative. As $e_L$ reaches its most negative value, $Q_O$ will reach maximum density while $Q_N$'s density will be minimum. Thus, a change in quadrature grid-charge density comes about as a result of a potential applied to the limiter grid. Figure 11-12 shows the phasor diagram and the waveforms established up to this point.

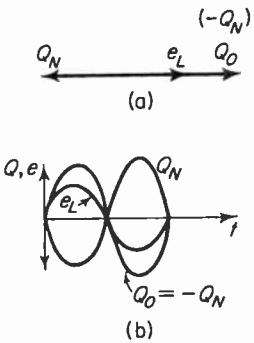
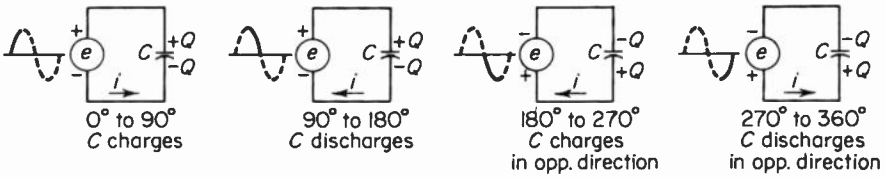
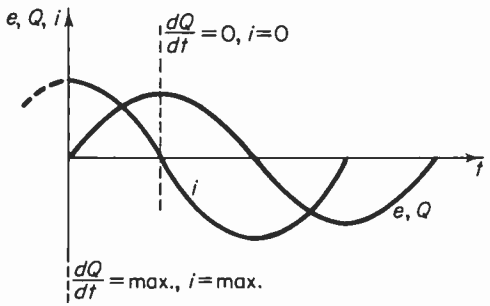



**Fig. 11-12.** Phasors and waveforms relating the applied voltage, charge near the quadrature grid ($Q_N$) and the charge on the quadrature grid ($Q_O = -Q_N$).

The next important quantity is $i_q$. This current flows as a result of the above "coupling of




(a)




(b)

**Fig. 11-13.** (a) Shows the response to a sinewave for a conventional capacitor and how the charges on the plates "follow" the applied voltage. (b) Shows the corresponding waveforms. The charge and applied voltage lag the capacitor current by 90°.

energy" from limiter to quadrature grid. A basic formula in electrostatics is

$$i = \frac{dQ}{dt}. \tag{11-2}$$

If a capacitor is connected across a source of AC potential, the current that flows "through" the capacitor $C$ will be maximum when the charge is changing most rapidly. Since the charge is "following" the applied voltage in phase, the current will be maximum when the applied voltage is changing most rapidly. Therefore, the current will lead both the voltage and charge by exactly 90°. This is shown in Fig. 11-13. Notice that the charge on either plate follows the polarity of the applied voltage with respect to the generator terminals. Figure 11-13(b) shows the associated waveforms and indicates the points at which the current is minimum and maximum. This action occurs with "normal" conventional capacitors. In the gated-beam tube, however, something very different happens. Since $Q_N$ and $Q_O$ change with $e_L$, it is perfectly permissible to assume that $Q_N$ and $Q_O$ act exactly like the plates of a "capacitor." This capacitor is not a physical component, but it does behave like one.

Referring to Fig. 11-14(a) through (d), as $e_L$ swings positive, $Q_N$ becomes

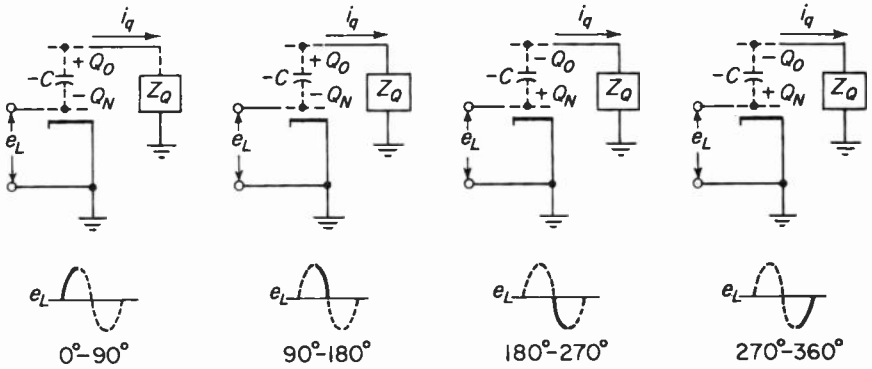



**Fig. 11-14.** (a) Through (d) show the presence of a negative capacitance ($-C$) which acts as the "coupling" element between the limiter and quadrature grids in. A negative capacitance has positive reactance which decreases with an increase in frequency. A negative capacitance therefore takes the form of $-1/j\omega c$.

more negative while $Q_o$ becomes less negative. As $e_L$ swings negative, the reverse occurs. Thus, the "plates" charge and discharge differently compared to capacitor $C$ of Fig. 11-13. This is equivalent to a small capacitance connected between limiter and quadrature grid. However, since it is not performing like a conventional "positive" capacitance, it is termed a negative capacitance ($-C$) (see Appendix 11-1). This parameter is no less significant than a negative resistance, a concept which has been accepted for many years. Therefore, ($-C$) exists in all space-charge coupled tubes operating as *space-charge limited* devices. Pentagrid converters are examples of such tubes. In this case, however, the ($-C$) generally is troublesome and can be eliminated by connecting a small conventional capacitor between the affected grids. The magnitude of this negative capacitance is of the order of $-2$ to $-0.5$ pf.

The quadrature-grid current $i_q$ is, of course, an induced current, since no "physical" contact takes place between the limiter and quadrature grids. The basis for $i_q$ is the fact that if there is a *change* of charge, there will be an electron flow. The magnitude of this induced current (occasionally referred to as displacement current) will depend upon the rate of change of charge with respect to time. Since $Q_o$ is equal to $-Q_N$, the quadrature current $i_q = dQ_o/dt$. First, the phase angle between $i_q$ and $Q_o$ must be 90°, since the current is a maximum when the rate of change of charge is maximum. Since there is a negative sign preceding "$dQ_o/dt$," $i_q$ must *lag* $Q_o$ by 90°. This is shown in the waveforms of Fig. 11-15. At instant $A$, the slope of the $Q_o$ curve is positive and maximum. This means that an increase in charge in the positive direction takes place with an increase in time. At point $A$, $i_q$ is maximum negative. At point $C$, the slope of the $Q_o$ curve is negative and

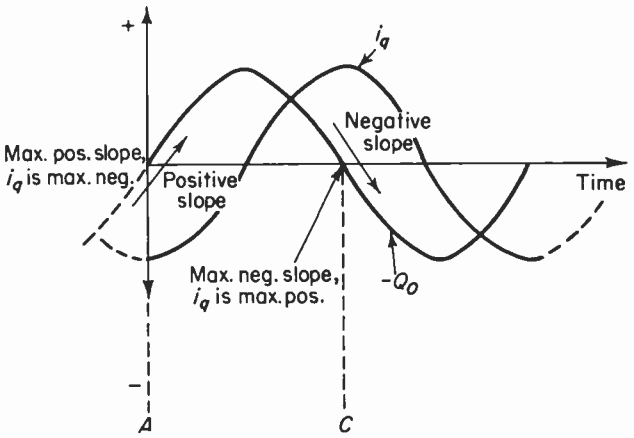



**Fig. 11-15.** The quadrature grid current $i_q$ is shown to lag the charge $Q_O$ on the quadrature grid by 90°.

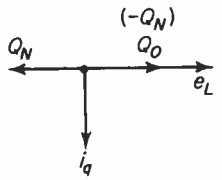



**Fig. 11-16.** The phasor diagram showing that the quadrature grid current $i_q$ lags the limiter grid voltage $e_L$ by 90°.

maximum. At point *C*, then, $i_q$ is maximum positive. The phasor diagram is shown in Fig. 11-16.

If an impedance is placed from quadrature grid to ground, a voltage will appear across that impedance. This voltage is termed the quadrature voltage, $e_q$. If this impedance is purely resistive, then the current and voltage will be in phase. If this impedance is inductive, the current will lag, while if the impedance is capacitive, the current will lead. This impedance takes the form of a parallel tuned circuit, as shown in the complete schematic of Fig. 11-17 where the 6BN6 tube is being used as a combined limiter and FM detector.

Consider an unmodulated carrier appearing at the limiter grid. This is designated $e_L$. Under this condition, the quadrature tuned circuit (which is tuned to the IF carrier) appears resistive and $i_q$ and $e_q$ are in phase. This is shown in the phasor of Fig. 11-18. The *Q* phasors have been omitted, since they will not be used to describe the circuit operation. If both grids are driven above cutoff, the plate current is switched on, but both grids must be above cutoff at the *same time*. Thus, the plate current must "wait" for the quadrature voltage to swing above cutoff before plate conduction begins. It must "wait" for $e_q$, since $e_q$ lags $e_L$ by 90° when the carrier deviation is zero. It will be assumed that the magnitude of the carrier has a peak amplitude much greater than the grid base of either grid and that both $e_L$ and $e_q$ are capable of driving the tube alternately into plate-current saturation and cutoff. The grid-voltage waveforms are as shown in Fig. 11-19, along with the plate-current pulses. Note that under conditions of zero modulation, the plate current starts to flow 90° after the start of the cycle and flows for approximately one-quarter of the cycle. Thus, the fundamental component

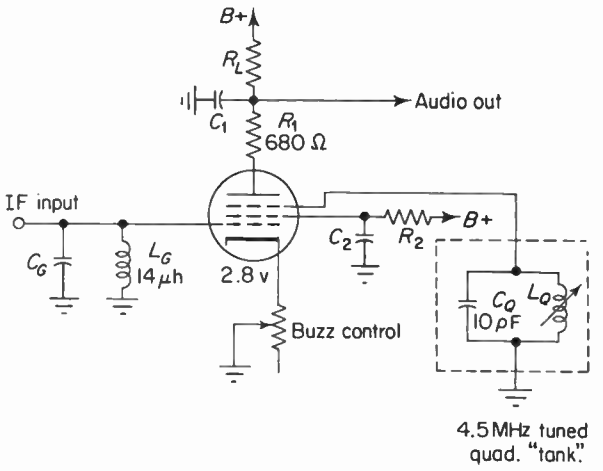



**Fig. 11-17.** The 6BN6 gated beam tube for use as a combined limiter-detector for TV sound. The IF input is the 4.5 MHz sound intercarrier IF. $L_G$ and $C_G$ shunt resonate (together with the input $C$ of the tube) to form a high impedance input.

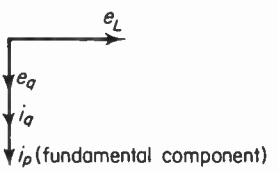



**Fig. 11-18.** The finalized phasor for the 6BN6 circuit under resonant conditions. The quadrature voltage normally lags the limiter voltage; thus the name quadrature grid.

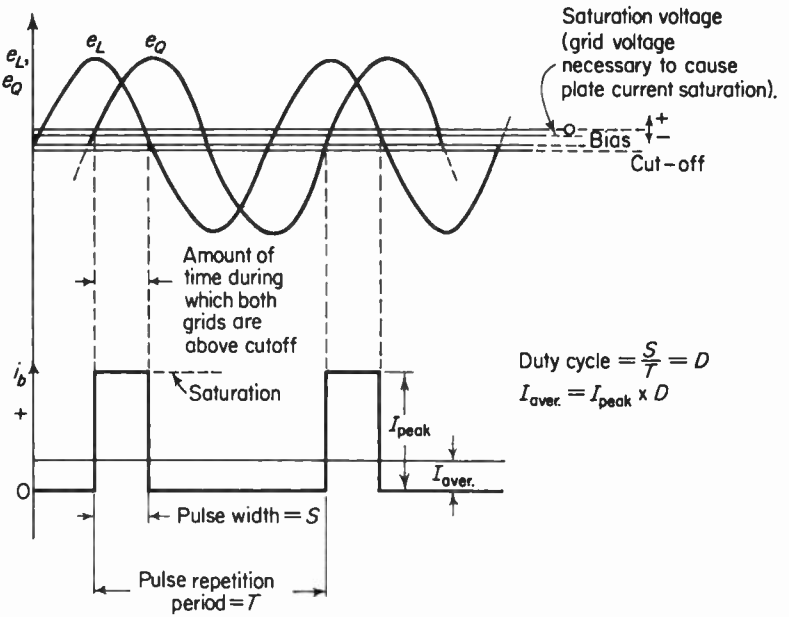



**Fig. 11-19.** The grid voltages and plate current waveforms under conditions of zero carrier deviation.

of plate current lags $e_L$ by approximately 90° and is shown in the phasor diagram of Fig. 11-18. The duty cycle (the ratio of the pulse width to the pulse repetition period) is then 25 per cent. So long as the carrier deviation remains zero, the duty cycle will remain at 25 per cent. Therefore, the width of the plate-current pulses will remain constant as shown in Fig. 11-19. An examination of the plate-load circuit of Fig. 11-17 shows the presence of an *RC* network. Since the plate-current wave is a complex waveform, it can be shown (by means of a Fourier analysis) to consist of a DC component (an average value), a high-frequency fundamental sinewave component, and harmonics (in this case, odd-ordered). The average plate current is given by

$$I_{av} = I_{peak} \times D, \tag{11-3}$$

where $D$ = duty cycle.

Since the duty cycle remains constant for a zero carrier deviation, the *average* load current must remain constant. No audio output voltage is produced across $RL$, and the fundamental carrier component and harmonics are bypassed by $C_1$.

Assume, now, that the carrier frequency is instantaneously above the center frequency. The "forced frequency" of oscillation of the quadrature tank does not change, since its $Q$ is of the order of 100. The frequency of $e_L$, however, is now greater than 4.5 MHz. The quadrature tank impedance no longer "appears" as a resistance to the limiter-grid driving circuit. When a parallel tuned circuit is driven by a voltage ($e_L$) whose frequency is above the resonant frequency, the impedance of the tuned circuit becomes *capacitive* with respect to the input circuit. The quadrature current and voltage are no longer in phase. In this case, $e_q$ must lag $i_q$ by an angle $\theta$. Thus, $e_q$ lags $e_L$ by an angle greater than 90°. Since the plate current must "wait" for $e_q$ to swing the quadrature grid above cutoff, the plate current will be delayed for more than 90° of the IF cycle. The amount of time for plate-current flow is now reduced, since the amount of time that both grids are above cutoff is reduced. Since plate current now flows for a short time, the duty cycle is reduced.

Figure 11-20 shows the grid and plate-current waveforms for the case of a positive deviation of the FM carrier. Part (c) shows the corresponding phasor. The average component of plate current is given by Eq. (11–3) and the duty cycle is now reduced. There will be a reduction in the average plate current and therefore an increase in the average plate voltage. Since this represents a *change* in load voltage, there will be an output *different* from that obtained for zero deviation. Thus, the circuit is capable of discriminating between zero and positive deviations of the carrier. This change of load voltage is a change in the DC level; the carrier component and harmonics are bypassed by $C_1$.

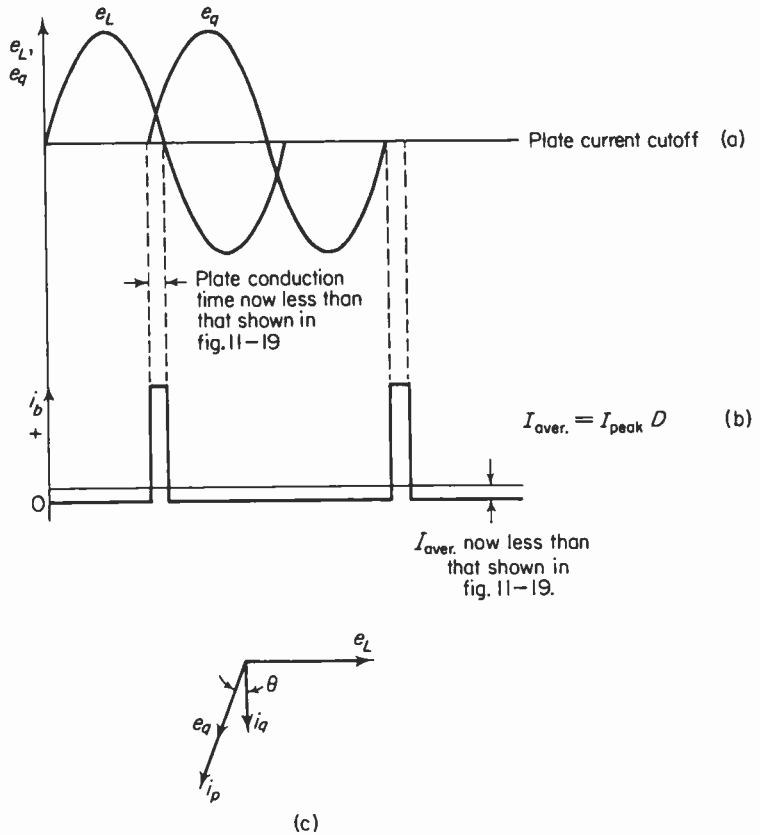



**Fig. 11-20.** The grid voltages and plate current waveforms under conditions of a positive deviation of the carrier (above 4.5 mm).

At this point it would be well to consider that the magnitude of this change in load voltage is determined by the amount of change of plate current, which is directly proportional to duty cycle. The duty cycle, in turn, is related to the phase angle between $e_q$ and $e_L$, the angle being determined by the amount of carrier deviation off center. The carrier deviation is a function of the amplitude of the audio modulating voltage. Thus, the amplitude of the audio modulating voltage determines the magnitude of change in load voltage, while the *frequency* of the modulating voltage determines the *rate* at which these changes in load voltage take place. The case where the FM carrier deviates below the center frequency is shown in Fig. 11-21.

Since the plate current is averaged over a great many cycles, it is difficult to show sufficient IF cycles corresponding to one complete audio cycle. However, if the time base of the plate-current pulses is expanded, the time for a part of one audio cycle can be illustrated as in Fig. 11-22. The points

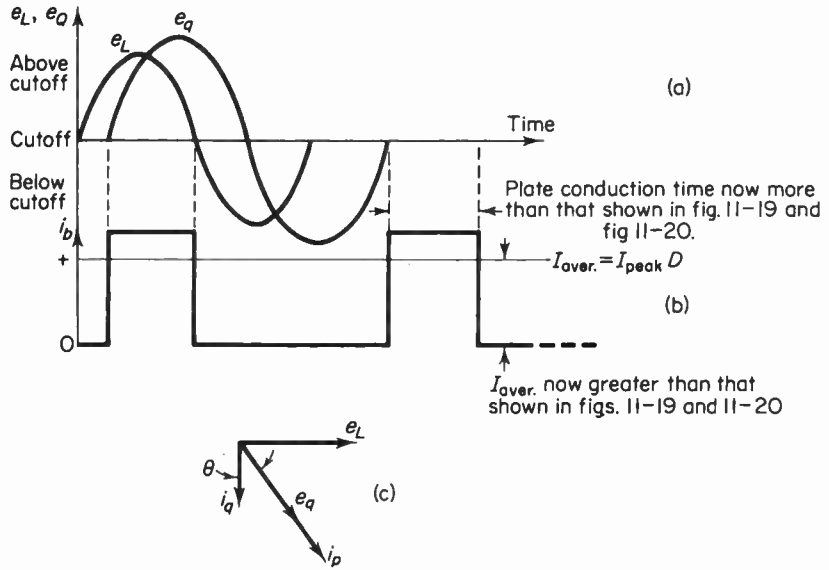



**Fig. 11-21.** The grid voltage and plate current waveforms corresponding to a negative deviation of the carrier.

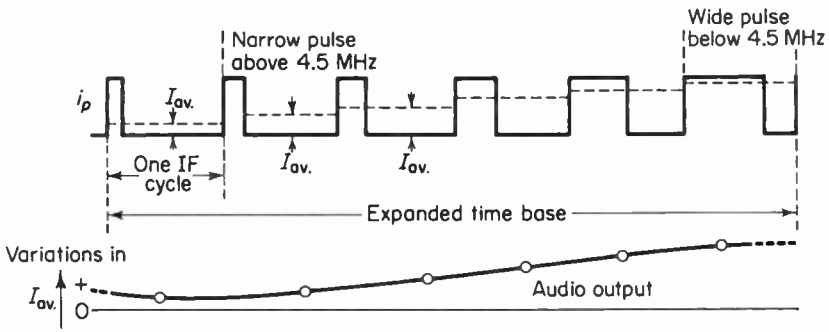



**Fig. 11-22.** The audio output is shown for the small portion of an audio cycle.

representing the average value of plate current are joined together to show a small portion of one audio cycle.

The *RF* filter in the plate circuit of the 6BN6 tube consists of $R_L C_1$. This network filters out the carrier component (plus harmonics) and retains the average component, which varies with carrier deviation. This filter has a time constant that is long compared to the time of one cycle of the carrier (0.25 $\mu$ sec approximately) and comparable to the period of the audio. Its ability to filter IF is thus comparable to that of the load filter of an AM detector.

$C_1$ in the schematic of Fig. 11-17 is also part of the deemphasis network.

The resistive part of the network is the plate resistance of the tube. The value of $C_1$ is adjusted to give the proper high-frequency rolloff corresponding as closely as possible to the standard 75-$\mu$sec deemphasis time constant.

## 11-5. Miscellaneous considerations

Owing to the very effective shielding between the limiter and quadrature grids, the interelectrode capacitance is of the order of 0.004 pF (or less). This is, relatively speaking, a very small value, and it will not sustain feedback from the quadrature to the limiter grid at the IF. Therefore, conditions in the quadrature grid do not affect conditions in the limiter-grid circuit, and the IF stability is excellent.

The magnitude of the developed audio voltage depends upon many factors (carrier deviation, quadrature tuned-circuit damping, supply voltages, and so on), but under average conditions the circuit can be expected to provide 10 to 15 volts rms audio for a 1.25-volt rms IF input signal. This output is under conditions of a 25-kHz carrier deviation; it would be somewhat less for a correspondingly reduced swing. However, it is still adequate to directly drive a power amplifier without the need for a first audio voltage amplifier. This is an economical arrangement, since the 6BN6 performs limiting, FM demodulation, and audio amplification. Thus, the entire sound strip of a TV receiver could consist of a single stage of sound IF amplification, the beam-gated tube followed by a power amplifier, plus the saving of at least two transformers.

The transfer characteristic (Fig. 11-23), is based on an IF of 4.5 MHz and a maximum deviation of 25 kHz. Note the absence of the sharp bends

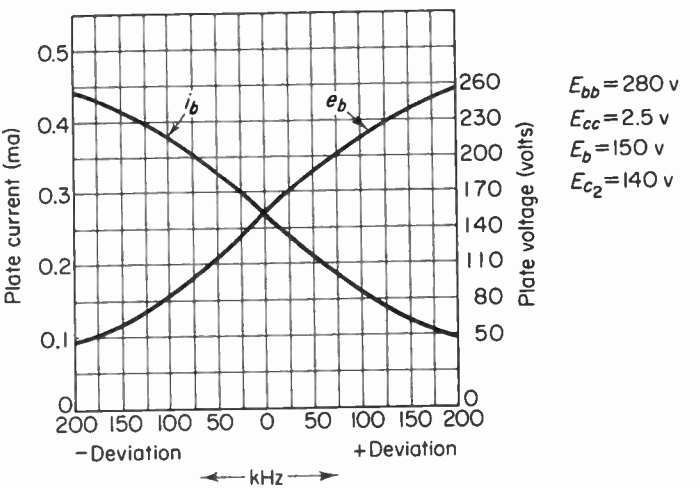


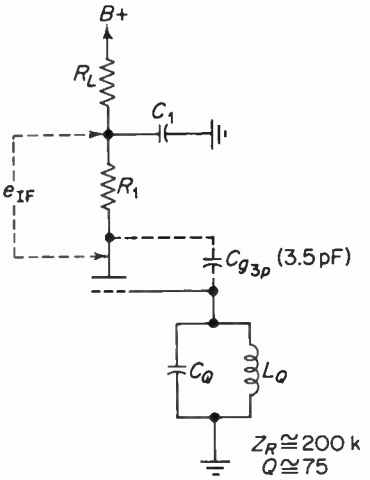

**Fig. 11-23.** 6BN6 detector response for 4.5 MHz IF. This is a point by point plot with $e_b$ taken across $C_1$ of Fig. 11-17.

found in phase-shift discriminators. This characteristic would eliminate side response (mentioned earlier) if this detector were to be used in an FM broadcast receiver. In TV sound, however, side response is not a problem, since the intercarrier IF is set at the transmitter and therefore the tuned circuits are always properly centered (barring any gross misalignment). Any AM introduced as a result of slope detection in the sound IF is of little importance anyway, since the 6BN6 output is very low in the region where slope detection takes place. Note, also, that for zero deviation the DC potential at the sound take-off (high side of $C_1$ in Fig. 11-17) is not zero. Unlike the balanced discriminators previously discussed, the 6BN6 makes no voltage available for AFC purposes or for a center-of-channel tuning indicator.

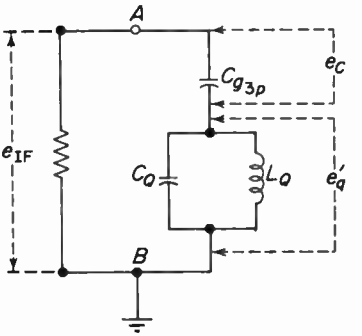
An examination of Fig. 11-17 will show the presence of a small resistor designated $R_1$. The purpose of this resistor is twofold. Since one end is at IF ground potential, while the other end is tied to the plate of the tube, the carrier component of plate current must flow through this resistor. It therefore comprises the IF plate load for the tube and controls the IF gain of the stage. Consider the interelectrode capacitance between plate and quadrature grid designated $C_{g3p}$. This capacitance displays a reactance of approximately 10,000 ohms. The quadrature tank resonant impedance is of the order of 200,000 ohms with a working $Q$ of about 75. As far as the plate circuit is concerned, the equivalent circuit is shown in Fig. 11-24(a) with a further simplification in (b). The IF component of plate voltage thus appears between points $A$ and $B$. The network to the right of points $A$ and $B$ forms a voltage divider, with most of the voltage appearing across the quadrature tank and in phase with the quadrature current. The quadrature current, plate current, space-charge induced voltage, and the component fed back to the quadrature grid ($e_q$) must all be in phase. This feedback component supplements the original $e_q$ and thus aids in driving the tuned circuit, as shown in Fig. 11-24(c).

Since $C_{g3p}$ is part of the quadrature-circuit total tuning capacitance with a resistor ($R_1$) in series with it, the presence of this resistor would damp the circuit as well as supply energy to it. This is, in fact, what it does. However, the energy lost is more than made up for by energy feedback via $C_{g3p}$. The net result is improved AM rejection and a low value of harmonic distortion owing to improved linearity.

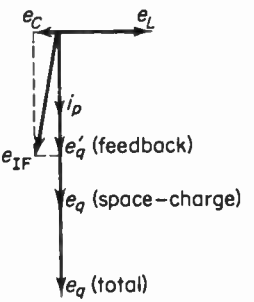
The working range of the discriminator characteristic is largely determined by the bandwidth of the quadrature tuned circuit. Since the quadrature-circuit bandwidth is a function of damping imposed on it by the linearity resistor, its value determines to some extent the linearity and bandwidth of the transfer characteristic. Omitting the resistor causes an increase in the slope of the response, with somewhat more output for a given carrier deviation. Inclusion of the resistor improves linearity at a slight loss in output.




(a)
Equivalent circuit for RF




(b)
Further simplification of part (a)




(c)
The phasor diagram showing the feedback component in phase with the space-charge induced component

**Fig. 11-24.** The circuits and phasor diagram showing how additional energy is coupled to the quadrature grid circuit for the purpose of improving both linearity and AM suppression.

The harmonic distortion with the resistor included is of the order of 1 per cent.

In light of the previous discussion, the plate and quadrature circuit capacities play an important part in the operation of this detector. Therefore, lead dress should not be disturbed when working on this circuit (repair, alignment, and the like), since changes in the distributed capacitance may adversely affect stability, linearity, quadrature tuning capacity, and so on.

The resistor in the cathode circuit affects noise rejection. It takes the form of a potentiometer (in some circuits, it is bypassed) for varying the operating bias. The optimum bias (for both grids) is midway between plate-current saturation and cutoff. This will insure symmetrical clipping of positive and negative noise spikes. This is particularly true of weak signals, as shown in Fig. 11-25. The transfer curve shown is equally applicable to the quadrature grid, since the cathode bias affects its operating point as well. In practice, the control is set for minimum noise output with a very weak signal input to the detector. This will permit the locally generated receiver noise to predominate, thereby utilizing the maximum limiting sensitivity of the stage. Its correct adjustment will minimize 60-Hz sync buzz, which is sometimes encountered in an intercarrier television receiver under conditions of video amplifier overload.

Since the adjustment determines the operating bias and since there is slight curvature of both ends of the transfer curve, spurious AM may appear at the plate due to plate detection unless the buzz control is carefully adjusted.

Noise rejection at high signal inputs does not insure optimum rejection with weak signals. Bias adjustment should always be made under the weakest signal conditions with noise always present. After the noise adjustment has been made, adjust the quadrature coil for best sound. Adjust the grid coil ($L_g$ of Fig. 11-17), if it is variable, for maximum audio output.

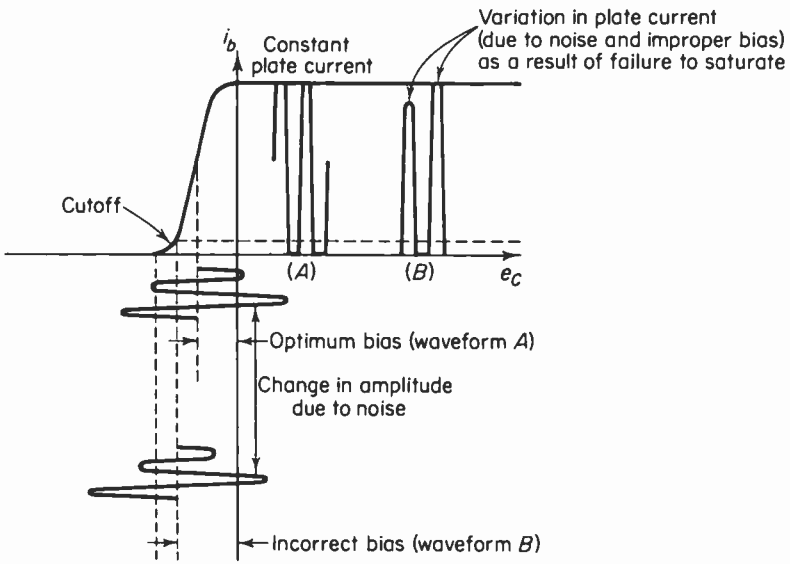



**Fig. 11-25.** This shows the effect of an improperly adjusted buzz control. Failure to saturate or failure to drive to cutoff will impain the AM rejection qualities of 6BN6.

It has been mentioned that the working range of the discriminator characteristic is largely determined by the quadrature tuned circuit. One of this circuit's most important characteristics is its working $Q$ ,which is in the order of 75. If a tuned response is plotted for two circuits that are resonant at the same frequency but that have different values of $Q$, the curve will appear as shown in Fig. 11-26(a). The plot shows how the quadrature tank impedance ($Z_q$) varies with frequency. At the center frequency ($f_r$) the impedance is very high and is given by

$$Z_q = QX, \tag{11-4}$$

where $X$ is the reactance of either the coil or the capacitor. The "circulating"

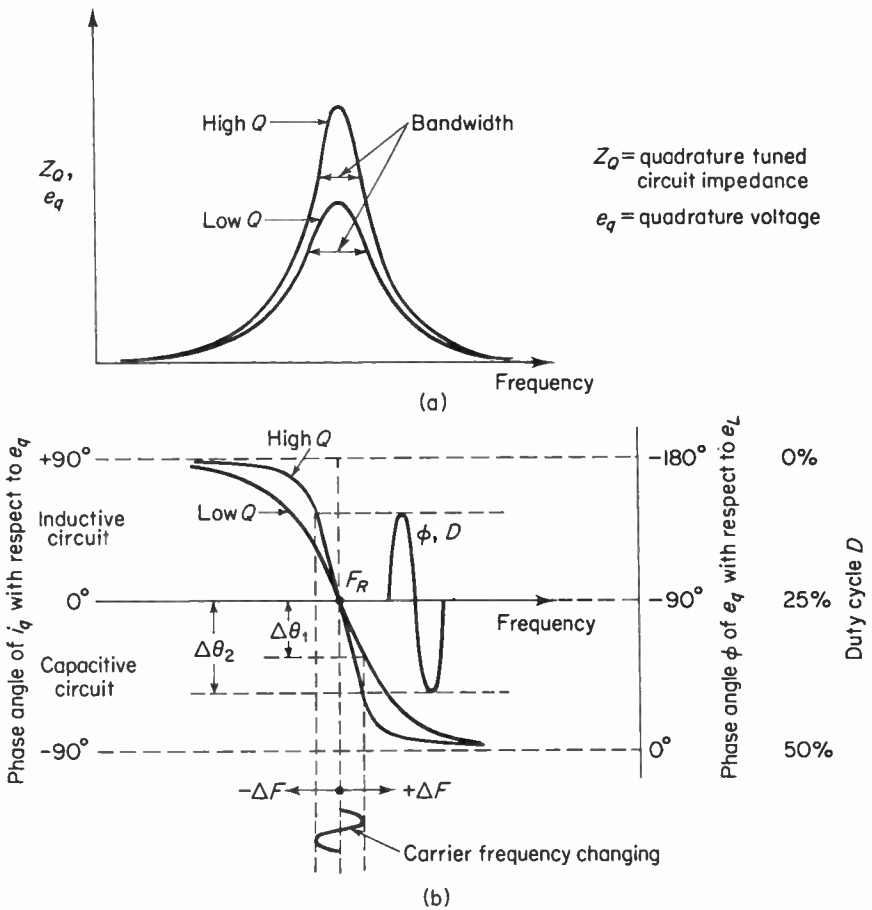



**Fig. 11-26.** Tuned circuit response for a low and a high $Q$ circuit (a) and the corresponding phase shift characteristics (b). The ratio of $\Delta\theta$ to $\Delta F$ should be large to provide a large output. This requires a high $Q$ circuit (i.e., $\Delta\theta_2 > \Delta\theta_1$).

tank current is extremely small and the power factor is unity. Note that the resonant rise in impedance is greater for the higher-$Q$ circuit. In Fig. 11-26(b) a plot is made of the variation in phase angle $\theta$ between $i_q$ and $e_q$ with changes in frequency. The right-hand ordinate represents variations in the phase angle $\theta$ with changes in frequency. Again, two curves are shown for the low- and high-$Q$ case corresponding to the tuned-circuit response of part (a). Observe that the slopes of the phase-shift curves are different, particularly within the working passband of the tuned-circuit response of part (a). Assuming that the carrier-frequency deviations are kept within the linear range of the phase-shift curve, then we can see that changes in frequency will give rise to *linear* changes in phase angle. This will, of course, insure that the linearity of the "S" curve will be maintained and harmonic distortion will be minimized. Also, if the circuit $Q$ is high, the quadrature-circuit phase sensitivity will be high. High phase sensitivity means that a small frequency deviation produces a large change in the phase angle between $e_q$ and $e_L$. Alternately, it can be said that high phase sensitivity produces a larger audio output for a given frequency deviation. Thus, the amount of phase shift for a given frequency deviation is proportional to the tuned-circuit $Q$. Therefore, the value of the linearity resistor $R_1$ of Fig. 11-17 does have a bearing on the circuit performance.

A plot of the limiting characteristic for a 6BN6 detector and a Foster-Seeley detector preceded by a single limiter is shown in Fig. 11-27. The curve shows that the 6BN6 is comparable if not slightly superior. The limiting threshold level for the 6BN6 is of the order of 4 volts rms before the audio output begins to level off, while that of the Foster-Seeley plus limiter is approximately 1 volt rms. However, beyond the limiting levels, the 6BN6 has a slight edge in performance. The judicious use of some form of delayed AGC would tend to level off the curve of the Foster-Seeley arrangement, and some broadcast tuner designs do make use of this arrangement.

A comparison of both circuits for AM rejection is shown in Fig. 11-28.

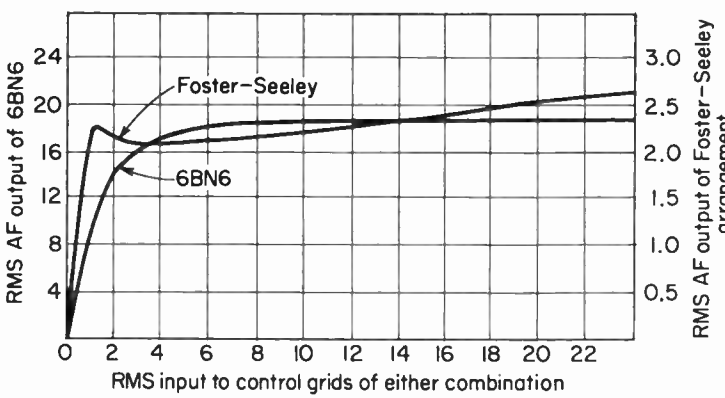



**Fig. 11-27.** Limiting characteristics of 6BN6 vs. Foster-Seeley plus limiter.

Here, it is seen that the Foster-Seeley arrangement is somewhat better, particularly in the working range between 5 to 20 volts rms input. This test is made by simultaneously injecting AM and FM signals, both of the same amplitude and percentage modulation, and noting the amount of attenuation (in decibels) of the AM with respect to the FM. The Foster-Seeley arrangement (from about 2.5 volts rms) can reduce the AM about 40 db below the FM throughout the working range. This corresponds to a 100-to-1 reduction of the AM over the FM. The 6BN6, however (above about 5 volts rms input), provides approximately 30 db rejection of AM, which is comparable to a ratio detector without a limiter. The 6BN6 gated-beam detector can thus be summarized as follows: good sensitivity, a fixed limiting threshold (about 3.5 volts rms), constant audio output (about 15 volts rms) for signals above the limiting level, and acceptable linearity. AM rejection is somewhat below that of the Foster-Seeley but comparable to the ratio detector where the ratio detector is not preceded by a limiter. Rejection of impulse-type noise by the 6BN6 is comparable to that of the phase-shift detectors.

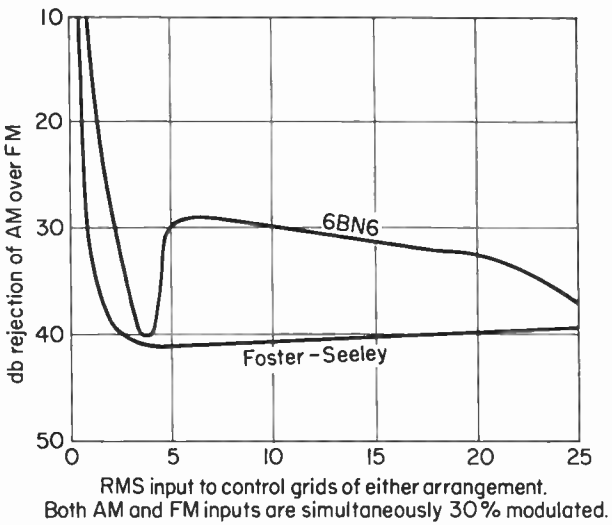



**Fig. 11-28.** AM rejection characteristic of 6BN6 vs. Foster-Seeley plus limiter. The simultaneously applied AM and FM signals are of equal value throughout test.

## 11-6. The 6DT6 locked-oscillator quadrature detector

This tube was developed by the Radio Corporation of America for use as a sound detector in television receivers, and in some respects it resembles the 6BN6 detector. It is self-limiting and can furnish sufficient audio voltage

to drive a power amplifier directly. It is a dual-controlled pentode similar to the 6BN6 in that either of two grids is capable of plate-current switching as well as AM limiting. The AM rejection is, in part, due to the sharp-cutoff characteristics of two grids—a control grid to receive the IF input and a suppressor grid analogous to the quadrature grid of the 6BN6. Both grids have working grid bases of the order of 4.5 volts. A typical circuit is shown in Fig. 11-29, where we see some basic differences between this and the gated-beam detector circuit. The important differences are (a) the presence of an *RC* network in the suppressor-grid circuit, (b) a fixed resistor in the cathode circuit, and (c) the lack of a linearity resistor in the plate circuit.

We shall first give a general description of the circuit and then consider the details. The design makes use of two modes of operation. The first mode occurs at low input-signal levels and is referred to as "locked-oscillator operation." With a weak IF input signal to the control grid, the circuit is designed to oscillate in a manner similar to a tuned-grid, tuned-plate oscillator (TGTP). The oscillation frequency is determined by the components in the input and suppressor circuits as well as the instantaneous frequency of the IF carrier. The suppressor acts like the "plate," while the control grid acts like the control grid of a conventional TGTP oscillator.

Energy is coupled from control grid to suppressor grid by means of space-charge coupling exactly as in the 6BN6 tube, and this develops a voltage at the suppressor grid that lags the voltage at the control grid by 90°. These

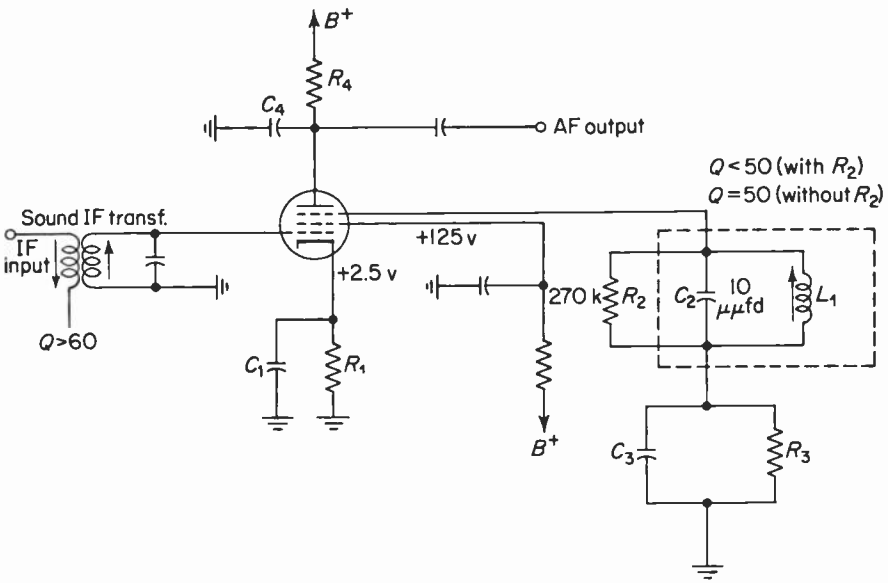



**Fig. 11-29.** A typical circuit arrangement using the 6DT6 tube as a television sound detector. The IF is 4.5 MHz (sound intercarrier).

voltages are designated as $e_g$ and $e_c$. There is a voltage gain of approximately 3 between control and suppressor grids, and therefore $e_g$ is approximately $3e_c$ under weak-signal conditions. A certain amount of energy developed in the suppressor circuit is then coupled to the control-grid circuit via the interelectrode capacitance between both grids. The phase of this feedback voltage is such that the entire circuit now oscillates, while energy is fed through the tube in two different directions (and through two different paths). Space-charge coupling from control to suppressor grid energizes the suppressor tuned circuit, while feedback via the interelectrode capacitance sustains oscillations by increasing the impedance of the input tuned circuit. The oscillation frequency is the same as the IF carrier frequency. The circuit now oscillates in step with the incoming signal and is said to be operating in the locked-oscillator mode.

As the carrier deviates above and below the center frequency, the phase of the suppressor-grid voltage $e_g$ shifts, relative to the control-grid voltage $e_c$. This results in changes in the width of the plate-current pulses, exactly as in the case of the 6BN6 gated-beam tube previously described. Variations in plate current give rise to an audio output voltage, the magnitude of which is linearly related to carrier-frequency deviation over a bandwidth sufficient to accept the full 50-kHz frequency swing encountered in television FM sound.

If the carrier deviation extends beyond the lock-in range of the oscillating detector, the circuit will not be able to remain in step with the deviations of the IF carrier. This results in a condition known as "breakout," which can be observed at the extremes of the detector characteristic in Fig. 11-30. The breakout phenomena represent the difference beats between the deviating carrier and the oscillating detector. Thus, if the carrier deviates beyond

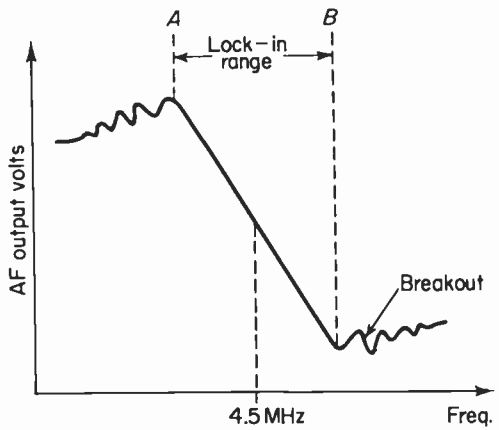



**Fig. 11-30.** The 6DT6 detector characteristic showing the beats occuring outside the lock-in range *A-B*. The breakout represents the difference beat between the oscillating detector and the RF carrier.

*A–B*, the oscillator will not remain synchronized with the IF carrier and will therefore oscillate at a fixed frequency (4.5 MHz). So long as the carrier-frequency deviations remain within the lock-in range *A–B*, the circuit will be forced or pulled into step with the input signal.

For moderate to strong input signals, the magnitude of control-grid current is such that the input impedance is reduced (due to grid-to-cathode damping) below the level required to sustain oscillations. The circuit now demodulates the FM carrier on the basis of quadrature grid detection exactly as for the 6BN6 circuit. This is the directly driven quadrature mode for moderate to strong inputs.

The sensitivity of the circuit is high, since a relatively weak input signal can control relatively strong oscillations, while its AM-rejection capabilities are comparable to those of the 6BN6 tube. Its circuitry is simple and its alignment is straightforward.

## 11-7. 6DT6 tube characteristics and associated data

The 6DT6 tube is a dual-controlled pentode having two sharp-cutoff grids, either of which is capable of plate-current cutoff. The control grid-to-plate transconductance is approximately 600 $\mu$mhos and the suppressor grid-to-plate transconductance is of the order of 500 $\mu$mhos. For a plate supply of 250 volts, the magnitude of bias for plate-current cutoff is of the order of 4.5 volts for either grid. These values are under class $A_1$ conditions and would vary somewhat with different signal levels. With a 0.15-volt rms input to the driver stage, the audio output is approximately 18 volts rms for a deviation of 25 kHz from the center frequency of 4.5 kHz. The detector bandwidth under the above conditions is approximately 100 kHz for a total harmonic distortion of 10 per cent.

It has been stated that under weak-input-signal conditions this circuit oscillates and that the configuration resembles a tuned-grid, tuned-plate oscillator. Actually, oscillations occur even when the IF carrier amplitude is zero. To understand this, let us examine the equivalent circuit of Fig. 11-31. The bias and other DC conditions have been omitted for the sake of simplicity.

Assume that there is some sort of signal "disturbance" in the control-grid circuit—perhaps noise or a *weak* IF carrier from the input circuits ahead of the detector. The disturbance induces a voltage into the secondary coil and causes a current to develop a voltage drop across the input to the tube. This voltage is space-charge coupled, via the $(-C)$ between both grids, to the suppressor-grid circuit, and it develops a voltage from suppressor grid to ground. Because of the nature of the negative capacitance and the design of the tube, a relatively small change in $e_c$ produces a relatively large change

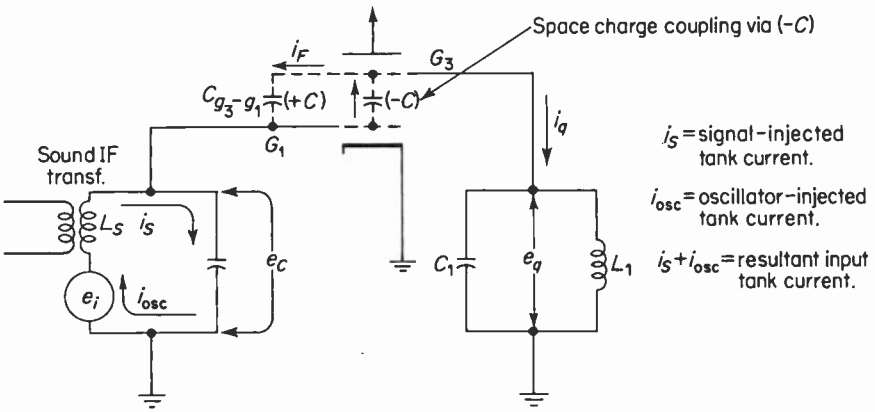
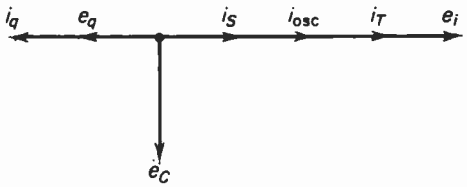



**Fig. 11-31.** Equivalent circuit showing the $(-C)$ space charge coupling path and the oscillatory feedback path via $C_{g_3-g_1}$. The circuit takes the general form of a TGTP oscillator.

in the charge density in the vicinity of the suppressor grid to produce a large change in suppressor-grid voltage. These voltages are 90° out of phase, and for the present we shall assume that the carrier deviation is zero. The voltage gain between control and suppressor grid is approximately 3 (under weak-signal-input conditions), and this amplification is sufficient for purposes of oscillation. A portion of $e_q$ is fed back to the control-grid tank by way of the interelectrode capacitance $C_{g3\text{-}g1}$, and this energy causes an oscillatory current to flow in the input tank. This current is designated $i_{osc}$, since it is the component of tank current due to the oscillations. Thus, two components of current flow in the control-grid tank circuit—one due to the input disturbance designated $i_s$ and the other due to the circuit oscillations designated $i_{osc}$. When the frequency of the IF carrier is the same as the resonant frequency of the oscillating circuit, these two currents are in phase and add to increase the output of the oscillator.

One of the important conditions for oscillation in the TGTP configuration is that the plate-circuit (in this circuit, the suppressor grid represents the "plate") resonant frequency be slightly higher than the resonant frequency of the grid tank. This will cause the plate tank to be inductive, and the plate circuit will reflect back a negative resistance into the grid circuit. A negative resistance is equivalent to a generator supplying power to the input circuit.

Assume that the circuit is oscillating at the frequency of the incoming carrier. The resonant-circuit phasor (Fig. 11-32) indicates that the oscillator-injected current $i_{osc}$ is in phase with the signal-injected current $i_s$. Since both quantities are rotating in a counterclockwise direction at the same angular velocity, the two currents simply add to increase the oscillator output. Suppose, however, that the incoming carrier deviates upward in frequency.




$i_T = i_S + i_{osc}$ = total tank current

**Fig. 11-32.** The phasor corresponding to the case where the signal frequency and the resonant frequency of the oscillator are the same.

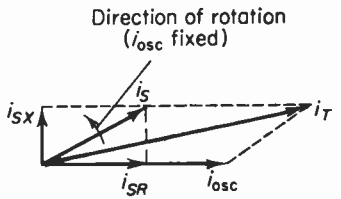



**Fig. 11-33.** The phasor corresponding to a sudden increase in signal frequency ($+\Delta F$). $i_s$ speeds up in the direction shown.

In order to keep the oscillating detector locked with the carrier, we must increase the frequency of the oscillator. This comes about as follows. Since the frequency of $i_s$ is now greater than that of $i_{osc}$, $i_s$ will rotate faster. To simplify the diagram, all other phasors will be omitted and $i_s$ and $i_{osc}$ will be retained. Also, $i_{osc}$ will be maintained in a fixed position and $i_s$ will rotate counterclockwise at an angular velocity ($2\pi f$) equal to the difference frequency between both phasors. Thus, $i_s$ speeds up and now leads $i_{osc}$. This is shown in Fig. 11-33, where $i_s$ now consists of two components. The in-phase component, $i_{sr}$, is resistive and plays no direct part in changing the frequency. The reactive component, $i_{sx}$, is equivalent to a *leading* current in the grid tank. This is equivalent to placing a capacitor across the grid tank, which causes a momentary *decrease* in the oscillator frequency. Since we are maintaining the oscillator phasor in a fixed position relative to the signal, we must therefore speed up the signal.

An instant later, the signal phasor will be in a position as shown in Fig. 11-34. Here, the reactive component of signal current $i_{sx}$ lags the oscillator current and injects a lagging current into the grid tank. This is equivalent to connecting an inductance in parallel with the grid tank. This causes the oscillator frequency to increase, and so the oscillator will lock with the signal. Hereafter, the signal phasor remains as shown in Fig. 11-35 and injects just enough inductance in parallel with the grid tank to lock the oscillator in step with the incoming carrier.

Figures 11-35 and 11-36 show the phasor diagrams corresponding to the condition of a carrier-frequency decrease.

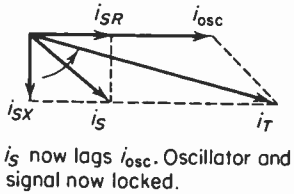



**Fig. 11-34.** The phasor corresponding to an instant after that shown in Fig. 11-33. A *lagging* current is injected into the grid tank.

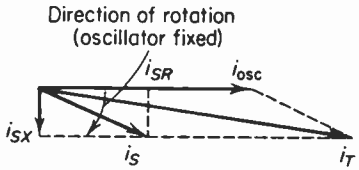



**Fig. 11-35.** The phasor corresponding to a sudden decrease in signal frequency ($-\Delta F$). $i_s$ slows down and thus rotates in the direction shown.

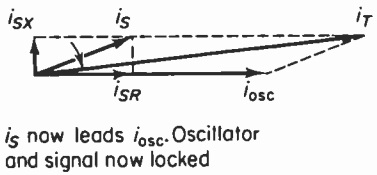



**Fig. 11-36.** The phasor corresponding to an instant after that shown in Fig. 11-35. A *leading* current is injected into the grid tank.

Since the suppressor-grid tank is physically part of the oscillator, it follows that the phase of $e_q$ changes with respect to $e_c$ with deviations of the carrier as it did in the 6BN6 gated-beam detector. The quadrature relationship thus exists with zero deviation and is other than 90° when the deviation is other than zero.

The demodulation of the carrier is a function of the amount of time that plate current flows—that is, the width of the plate-current pulse. The pulse width determines the average plate current, which in turn determines the average plate voltage. The average plate voltage thus varies as a function of carrier deviation, while the rate of change of plate voltage is determined by the frequency of the modulating voltage. Figure 11-37 shows the voltage conditions on both grids, the plate-current pulse, and the corresponding phasor for a positive deviation of the carrier. Notice that the suppressor grid is being biased at approximately cutoff by virtue of grid-leak bias developed by $e_q$. The cathode bias adds to this to set the operating point of the suppressor grid. Grid-leak bias is used in order to minimize grid current, since the positive peaks of $e_q$ are clamped to approximately zero grid voltage. Excessive grid current would damp the circuit heavily and reduce its weak-signal sensitivity (thereby reducing the audio output). The control

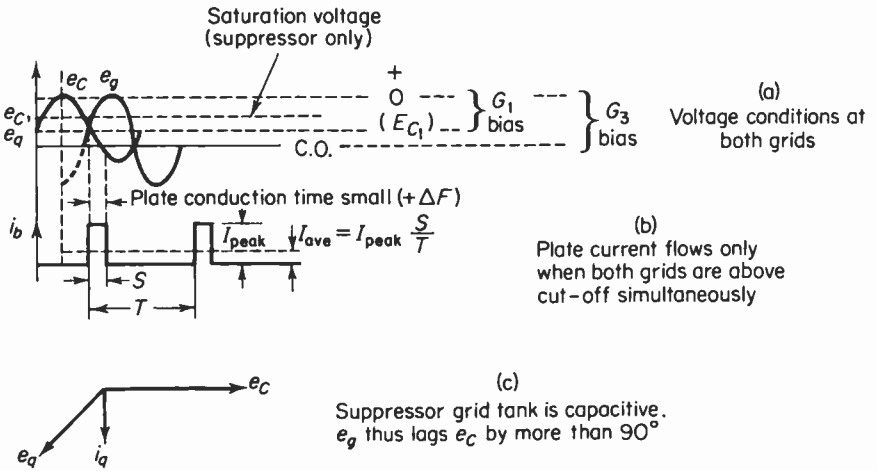



**Fig. 11-37.** The voltage conditions at both grids, the plate current pulse and the phasor diagram corresponding to a positive deviation of the carrier. The duty cycle $D$ is short and the average plate current $I_{ave}$ is small.

grid is cathode-biased approximately midway between plate-current cutoff and saturation. Since the suppressor grid is capable of cutting off the plate current, the plate current must again "wait" for the suppressor grid to be brought out of cutoff by $e_g$. Owing to the sharp-cutoff characteristic of the suppressor grid, the plate current is again shaped as a rectangular pulse, the width depending upon the amount of time that both grids are above cutoff.

Since the plate-current waveform is nonsinusoidal, it consists of a DC component in addition to a fundamental component (plus harmonics). The DC component develops the audio output, while the carrier component and harmonics are bypassed by $C_4$. The plate is thus at IF ground potential, the IF stage gain is zero, and no IF feedback occurs.

Since $e_g$ is adequate from the standpoint of space-charge coupling, no linearity resistor is required for supplementing the internal space-charge coupled voltage. It is for this reason that an external damping resistor ($R_2$ of Fig. 11-29) is occasionally placed across the suppressor-grid tank to provide the appropriate bandwidth necessary for optimum linearity. Some versions of this circuit omit this resistor for greater audio output, since the phase sensitivity will be greater without it (see Fig. 11-26).

Noise rejection in this tube comes about in a number of ways. First, in the event of an increase in carrier amplitude (sufficient to drive the control grid positive with respect to the cathode), the control grid will draw current. This reduces the grid-to-cathode resistance, and therefore the input circuit is damped to minimize the effect of the carrier increase. Some sensitivity

is lost, however, since the input-circuit bandwidth is increased. Second, the suppressor-grid characteristic (for a control-grid bias of 2 volts) shown in Fig. 11-38 indicates that plate current begins to saturate in the negative-grid region. Therefore, a sudden increase in $e_g$ will result in no increase in plate current on the positive peaks, while negative noise spikes are clipped by plate-current cutoff. Thus, the suppressor grid contributes greatly to noise suppression. Third, there is the matter of the cathode circuit. The cathode resistor is capacitor-bypassed. However, there is a component of voltage across this capacitor corresponding to *amplitude* variations of the IF, which serves to introduce degeneration and thereby maintain the plate current constant with changes in the input.

Experience with this circuit has shown that a fixed value for the cathode resistor for optimum linearity and AM rejection can be incorporated. This simplifies the circuit somewhat over that for the 6BN6 gated-beam detector, since with the 6DT6 tube no bias control is needed and no adjustment must be made for optimum conditions.

In order to maintain high sensitivity in the locked-oscillator mode, the input impedance presented to the control grid should be as high as possible so that the magnitude of $C_{g3\text{-}g1}$ does not have to be too large to sustain oscillations. This requirement can be met by feeding the detector tube from a step-up transformer and by resonating the secondary with the wiring and input capacitances of the tube. High grid-circuit impedance also improves limiting at high signal levels owing to clipping and input circuit damping, both as a result of control-grid current.

The purpose of the $R_3C_3$ network of Fig. 11-29 is to develop grid-leak bias for both the locked-oscillator and direct-driven quadrature modes of operation. The time constant of this network is such that the positive peaks

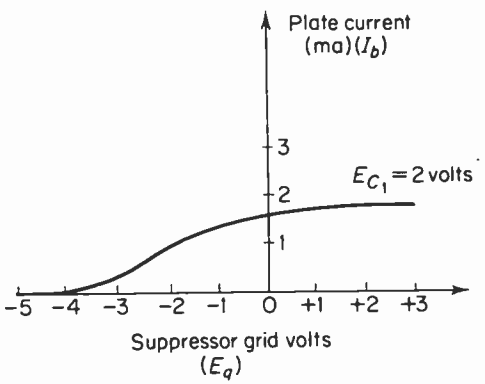



**Fig. 11-38.** The suppressor-to-plate transfer characteristic showing (1) the short working grid base to clip negative noise peaks and (2) the beginning of plate current saturation in the negative grid region which aids in clipping positive noise peaks.

of $e_q$ are clamped to zero grid voltage; consequently, the magnitude of suppressor-grid current is small. Were it not for the self-regulating mechanism of this form of bias, excessive suppressor-grid current on strong signals would cause severe loading of the associated tuned circuit, a reduction in audio output, and increased nonlinearity of the detector response (see Fig. 11-26 and note the effect of a reduced $Q$).

The direct-driven quadrature mode (moderate to strong signals) is essentially the same as for the 6BN6 gated-beam detector. A moderate to strong input signal at the control grid will cause sufficient grid current to damp the input tank. The damping is due to the reduction of control grid-to-cathode resistance on the positive signal peaks. This reduces the input-circuit impedance below the level necessary to sustain oscillations. The forward gain between control and suppressor grid falls below unity, and space-charge coupling continues to take place as previously discussed. Demodulation of the FM carrier is now exactly as for the 6BN6 detector via changes in the phase angle between $e_c$ and $e_q$, variations in the width of the plate-current pulse, and so forth.

The alignment of this detector is simple and straightforward and can be performed with an off-the-air signal. The quadrature coil in the suppressor-grid circuit is adjusted for maximum undistorted sound at a normal input level (any local TV channel will suffice). This will minimize the effects of any input-circuit detuning, since grid current will flatten the input-circuit response if the signal level is high enough. Then, disconnect one of the two TV antenna lead-in wires and adjust *all* input-circuit coils for minimum distortion and noise and maximum audio output. Reconnect the antenna wire and the alignment is complete.

## 11-8. Summary

**1.** The 6BN6 gated-beam tube uses an electron optical system for controlling the plate current.

**2.** Both grids (limiter and quadrature) are easily capable of cutting off plate current with small input signals.

**3.** The voltage for driving the quadrature grid is obtained by space-charge coupling from the limiter grid and normally lags the limiter-grid voltage by 90°.

**4.** Carrier deviations cause the phase angle between the two grid voltages to vary, resulting in variations in the width of plate current, thus giving rise to an audio output voltage.

**5.** The 6BN6 tube is a highly efficient limiter by virtue of the sharp-cutoff grids.

**6.** The 6DT6 locked-oscillator quadrature detector is a very sensitive type, since a low-level input signal can control a relatively strong oscillator.

**7.** The 6DT6 circuit is similar to the 6BN6 circuit in that the suppressor grid derives its working voltage via space-charge coupling from the control or input grid.

**8.** The 6DT6 circuit limits AM by virtue of input damping, cathode degeneration, and a sharp-cutoff suppressor-grid characteristic.

**9.** The 6DT6 circuit oscillates on weak signals, and the suppressor-grid phase varies about the quadrature phase to modify the plate-current width as for the 6BN6 circuit. Variations in the width of the plate-current pulses give rise to an audio output.

**10.** On the basis of a strong input carrier, input-circuit damping due to grid current stops the oscillations, and the carrier is demodulated on a conventional quadrature basis.

**11.** The amount of audio output for both types of quadrature detectors (for a given carrier deviation) depends upon the phase sensitivity of the quadrature-grid tuned circuit.

**12.** The 6BN6 gated-beam tube has been used as a very effective wide-band limiter in some commercial FM broadcast tuners.

## REFERENCES


1. Department of the Army: "F-M Transmitters and Receivers," TM 11-668. Superintendent of Documents, Washington, D.C., 1952.
2. Adler, R.: "A Gated Beam Tube," *Electronics*, February 1950.
3. Avins, J., and Brady, T. J.: "A Low-Cost Sound Detector for Television Receivers," RCA Laboratories, Industry Service Bulletin, LB-1000, October 1955.
4. Johnstone, G. G.: "Limiters and Discriminators for FM Receivers," *Wireless World*, June 1957.
5. Johnson, L. W.: "The Gated-Beam Valve," *Wireless World*, January 1957.
6. Hasse, A. P.: "New One-Tube Limiter-Discriminator for FM," *Tele-Tech*, January 1950.
7. General Electric Engineering Bulletin ET-B28, "The Gated Beam Tube and Its Application in Intercarrier Television."
8. Adler, R.: "A Study of Locking Phenomena in Oscillators," *Proc. IRE*, June 1946.
9. Beers, G. L.: "A Frequency-Dividing Locked-In Oscillator Frequency Modulation Receiver," *Proc. IRE*, December 1944.
10. Corrington, M. S.: "Locked-in Oscillator for TV Sound," *Electronics*, March 1951.
11. *Sylvania News Bulletin*, **23**: 4 (April 1956), "Gated Beam Discriminator—Type 6BN6."

## APPENDIX 11-1

This appendix considers negative capacitance and quadrature grid-current phase in the 6BN6 tube. If $e_L$† increases ($+de_L$), the charge density ($Q_N$) in the vicinity of the quadrature grid increases. This corresponds to a negative charge increment ($-dQ_N$). From fundamental considerations

$$i = \frac{dQ}{dt} = C\frac{de_c}{dt}. \tag{1}$$

Multiplying both sides by $dt$, it follows that

$$dQ = C\,de_c. \tag{2}$$

Therefore,

$$dQ_N = -C\,de_L \tag{3}$$

and

$$C = -\frac{dQ_N}{de_L}.^{\ddagger} \tag{4}$$

An increase of charge ($-dQ_N$) in the vicinity of the quadrature grid causes a proportional charge $dQ'$ to flow out of the quadrature grid. Thus,

$$dQ' = +K\,dQ_N, \tag{5}$$

where $K$ is a constant having the unit of farad. Substituting in (5) the value ($-C\,de_L$) from (3) for $dQ_N$, we obtain

$$dQ' = -KC\,de_L \tag{6}$$

or

$$KC = -\frac{dQ'}{de_L}. \tag{7}$$

The quantity $KC$ is negative, has the unit of farad, and is the equivalent of a negative capacitance of magnitude $KC$ connected between limiter and quadrature grid. Also, the quadrature current is given by

$$i_q = \frac{DQ'}{dt} = -KC\frac{de_L}{dt}. \tag{8}$$

Letting $e_L = E_m \sin \omega t$, then

$$i_q = \frac{-KC\,dE_m \sin \omega t}{dt}. \tag{9}$$

†$e_L$ is assumed to be sinusoidal.

‡This is equivalent to a negative capacitance connected between limiter grid and the space charge ($Q_N$) in the vicinity of the quadrature grid.

Differentiating (9),

$$i_q = -KC\omega E_m \cos \omega t. \tag{10}$$

Rearranging the right-hand side of (10),

$$i_q = \omega KC(-E_m \cos \omega t). \tag{11}$$

Since $i_q$ represents a negative cosine function, $i_q$ must then lag $e_L$ by 90°.

# 12

# FM TUNING INDICATORS

## 12-1. Introduction

The quality of the audio output of an FM receiver depends to a large degree upon proper tuning. Proper tuning means that the IF falls on the center of its response curve and that the signal also falls on the zero or null point on the detector S curve. With the advent of multiplex, proper tuning has become essential for the achievement of good channel separation.

Proper tuning presents difficulties because of the wide and relatively flat bandpass requirements of the IF, the limiter, and the linear response of the detector stages. This type of response results in a tuning situation where the operator must tune for least distortion rather than maximum volume as is the case with AM receivers. As can be seen in Fig. 12-1(a), proper tuning causes the IF to fall on the center of the IF response and at the same time to fall on the most linear portions of the detector S curve. Volume is not necessarily maximum at this point, because the audio output level depends upon the slope of the S curve. For a receiver that employs a ratio detector, the output level of the detector depends upon the input-signal strength, as is the case with AM receivers. However, a receiver employing a Foster-Seeley discriminator as an FM detector will exhibit little noticeable variation in output level for variations in (RF) input-signal levels, provided that the threshold of limiting is exceeded.

The effect of improper tuning can be seen in Fig. 12-1(b). Note that during

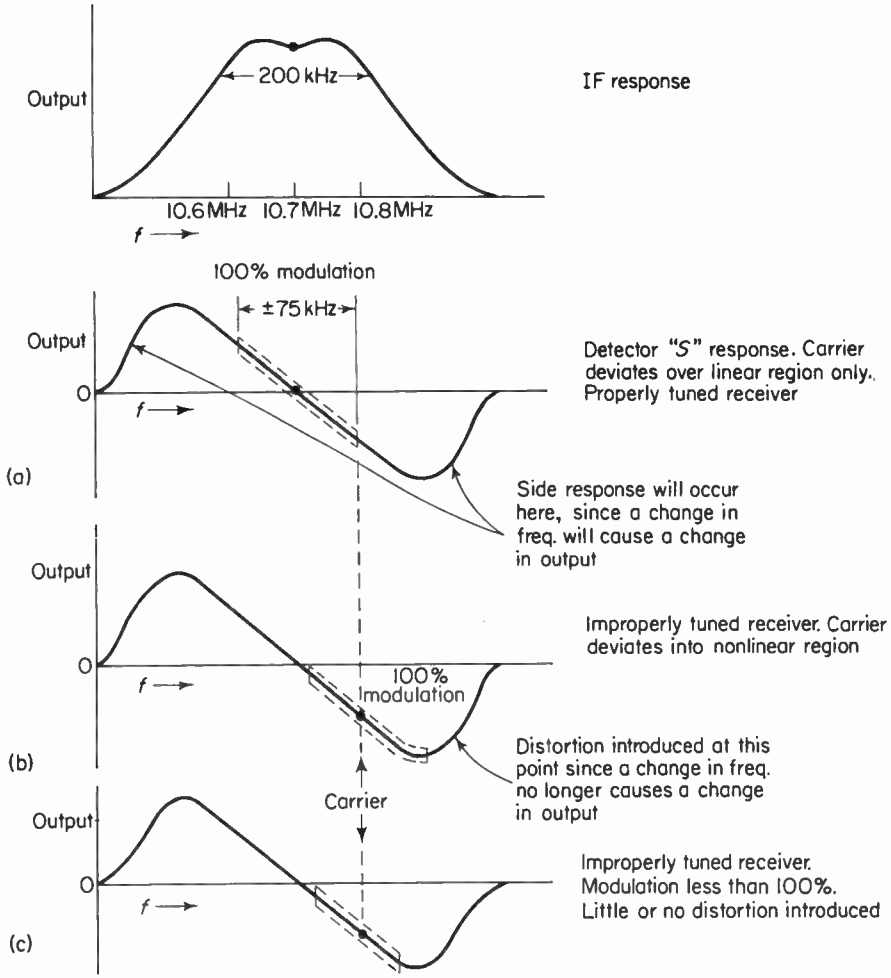



**Fig. 12-1.** (a) Illustrates the IF response and the detector *S* curve response for a properly tuned receiver. (b) Illustrates the *S* curve response for an improperly tuned receiver. The percentage of modulation is 100%; note that on the peaks of carrier swing distortion is introduced. (c) Is the same as (b), except that the incoming signal is modulated less than 100%, thus little distortion is introduced.

maximum deviation (maximum audio level at the transmitter) the carrier deviates into the peak or curved portion of the S curve. This will introduce rather obvious harmonic and intermodulation distortion. Figure 12-1(c) represents the same tuning error except that the deviation is less than maximum (low audio level at the transmitter). Note that the entire signal falls on the linear portion of the detector S curve and, therefore, little or no distortion will be introduced. It can be seen that if the transmitted program

material consists mainly of low-level audio, the operator may detune slightly and be unaware of the error.

An additional problem associated with tuning an FM receiver is that of side response. Side response refers to the condition where more than one response to an FM channel may occur. These responses may occur above and below the main response. See Fig. 12-1(a). It is, therefore, quite possible in certain types of receivers for the inexperienced operator to tune to the side response rather than the main response. Receivers that have relatively narrow peak separation of their detector S curve are prone to side response. Receivers that use wide-band (as much as 3 MHz) limiters and ratio detectors have no side response at all, because the output of the IF amplifiers will be zero long before the peaks of the detector S curve are reached and passed.

#### *Automatic frequency control*

One method for insuring proper tuning as well as drift-free reception is to make the local-oscillator frequency depend upon the output of the FM detector. Such a system is called automatic frequency control (AFC). Basically, the system is a feedback device. The detector's S curve determines the magnitude and direction of the tuning error. For example, if improper tuning should lead to the situation depicted in Fig. 12-1(b), then the output of the detector will consist of distorted audio and a DC voltage proportional to the error in tuning. The DC component is fed back to a reactance tube or a similar device, which forces the local oscillator to shift in frequency in the direction necessary to make the difference between the oscillator and the incoming RF approximately equal to the IF (10.7 MHz). Thus, AFC will force the receiver to be properly tuned. Theoretically AFC should remove the need for a tuning indicator, except that for the average listener it is more satisfying to use a tuning indicator and thereby know that the receiver is tuned properly. Consequently, many receivers on the market employ both AFC and a tuning indicator. Needless to say, there are also many receivers with neither AFC nor a tuning indicator. Such economy receivers are tuned by ear, and so improper tuning may be the rule rather than the exception. AFC may not be able to provide proper tuning in a situation where the desired signal is weak and is close (in frequency) to a strong signal. Under such conditions, the stronger signal may prevent reception of the weak signal. Thus, most receivers employ an AFC defeat switch and tune in the weak signal by means of the tuning indicator.

## 12-2. Methods of indication

FM tuners and receivers use at least three methods for tuning indication: (1) peak, (2) null, and (3) a combination of both. See the block diagram of Fig. 12-2 for the location of the take-off points for these methods.

The peak method of tuning depends upon the principle that when the receiv-

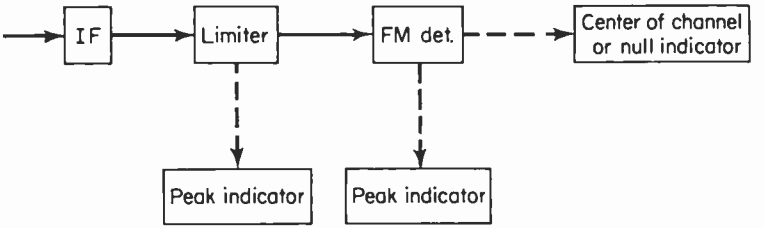



**Fig. 12-2.** A partial block diagram of an FM receiver showing the take off points for the peak or null type tuning indicator.

er is properly tuned to a station, the grid-leak bias of the limiter stage, the voltage across the stabilizer capacitor of the ratio detector, and the voltage across half the load of the Foster-Seeley discriminator are maximum. A variety of devices may be employed to display this fact. The obvious disadvantage of this system is that the bandpasses the of IF, limiter, and detector stages are approximately flat-topped, thus maximum indication can occur over a wide range of frequencies, and exact tuning may be difficult. The advent of multiplex has made the IF bandpass requirements even more critical.

The use of the null or zero voltage developed at resonance, and appearing at the output of the Foster-Seeley discriminator or ratio detector, is the sharpest method for determining the correctness of tuning. The reason is that the zero-voltage output of the S-curve response occurs only at the IF (10.7 MHz) rather than over a range of frequencies,

A number of tuners on the market have two simultaneous indicators—one for the relative strength of the incoming signal (peak) and the other for center-of-channel tuning (zero).

It should be understood that no matter what method of tuning indication is employed, it is only as good as the overall alignment of the receiver.

*Tuning indicators*

Manufacturers of FM tuners and receivers have relied upon various devices to show proper tuning. These devices fall into three categories: (1) tubes that make use of fluorescent screens for the display of tuning, such as the 6E5, 6AF6, 6AL7, DM70, EM81, and EM84 tubes, (b) neon-bulb displays, which indicate correctness of tuning by changes in brightness, (c) meters that read peak voltages or meters that can read zero at the center of their scale and can, therefore, indicate voltage polarity (thus direction of frequency error) as well as degree of detuning.

## 12-3. The 6E5

The 6E5, sometimes called the "magic eye," is one of the oldest tuning indicators, and its basic principle is still being used. The tube is a triode as well as a display device. Figure 12-3(a) is a cutaway diagram of the 6E5;

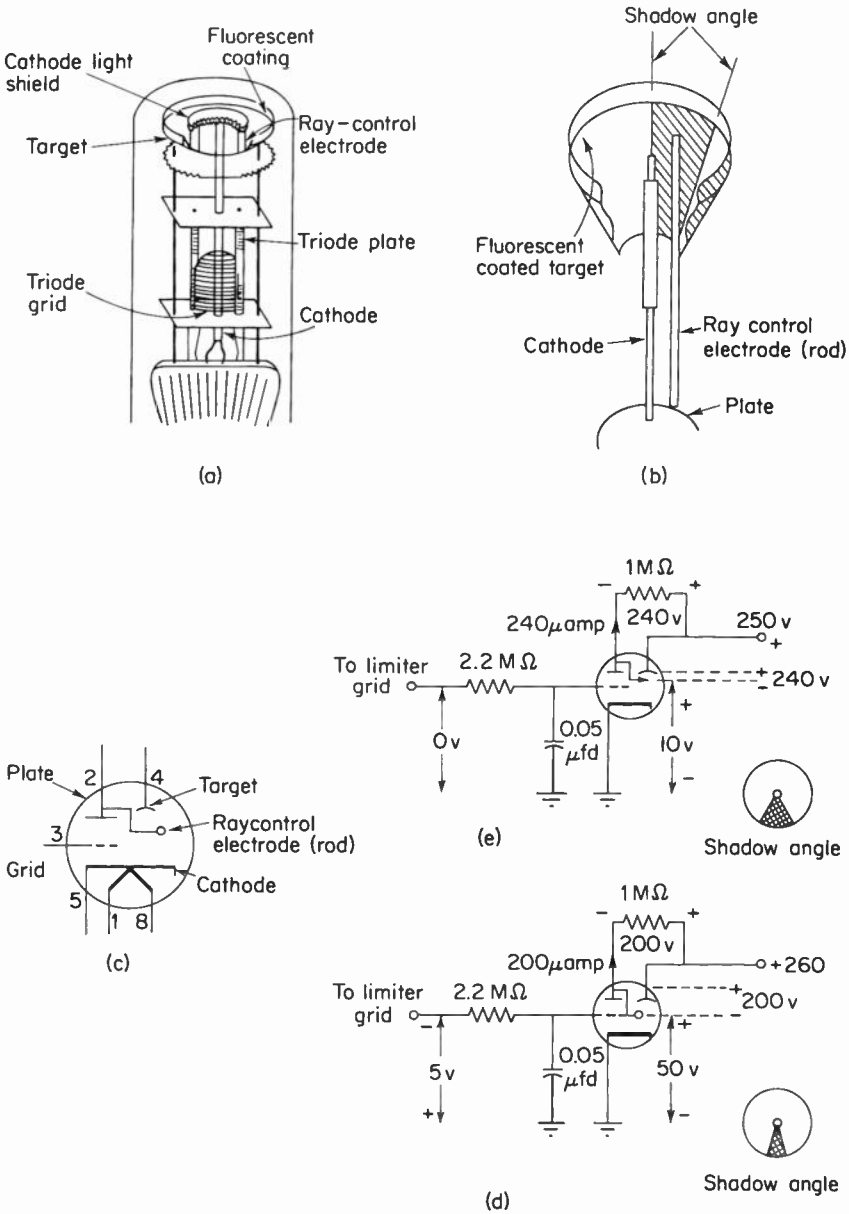



**Fig. 12.3.** (a) Is a cutaway diagram of the 6E5, illustrating the triode and fluorescent display portions of the tube. (b) Is an enlarged diagram of the fluorescent display portion of the tube. (c) Is the schematic representation of the tube. (d) and (e) show typical circuits and also illustrate the changes in voltage, current and shadow angle that takes place in a tube for changes in grid bias.

the schematic representation is shown in Fig. 12-3(c). With the help of Fig. 12-3(b) we can see that the plate is connected (internally) to a rod that extends up through the fluorescent target. The cathode of the triode also extends into this region. Electrons leaving the cathode (in the region of the target )are accelerated by the positive potential of the target, and upon striking the target (which is connected to $B^+$) convert their kinetic energy into light energy (green). The plate of the triode and thus the "ray-control electrode" are at a lower potential than the target because of the voltage drop across the plate load resistor. This difference of potential results in an electrostatic field between rod and target, which guides the electrons in a path that avoids the rod. Thus fewer electrons leaving the cathode, and in line with the rod, reach the area of the target in front of the rod. This wedge-shaped region, in contrast to the surrounding area, is dark and is called the shadow.

The greater the potential difference between the rod and target, the greater will be the shadow angle. If no potential difference exists between rod and target, then the shadow angle will be zero. The potential difference is varied by changing the bias on the grid of the triode section of the tube. For example [see Fig. 12-3(d)], if the grid is made more negative (as would occur when the receiver is properly tuned to a station), the plate current will decrease and the plate voltage will increase. Thus, the ray-control electrode will become more positive (with respect to ground) and the difference of potential between rod and target will become smaller. Consequently, the shadow angle will become smaller (the eye closes). Of course, if the grid is made less negative [Fig. 12-3(e)], as would be the case when tuning between stations, the plate current will increase with a resulting decrease in plate voltage. Thus, the voltage from rod to target will increase, and the shadow angle also will increase (eye opens).

As noted in Fig. 12-3(d) and (e), the voltage feeding the tuning indicator is obtained from the limiter's grid-leak bias. Of course, if the tuner employs a ratio detector, the voltage developed across the stabilizing capacitor may be employed to supply the grid voltage of the tuning indicator. Between the grid of the tuning indicator and the voltage source for the grid is a low-pass filter. The filter consists of a fairly large capacitor (0.05 $\mu$fd) placed from grid to cathode of the tuning indicator, and a large resistor (2 megohms) placed between the voltage source and the grid of the tuning indicator. See Fig. 12-3(d). One purpose of this filter is to prevent the tuning eye from loading down circuits into which it is connected. Loading may be caused by velocity or gas currents, which may change the quiescent conditions in either the limiter or the ratio detector. If the tuning indicator were connected directly to the grid of a limiter stage, the input capacity of the tube (6E5) would short-circuit the IF at that point. The 2-megohm series resistor effectively eliminates these loading effects by increasing the input resistance of the tuning indicator.

Finally, the ratio detector, and the limiter acting as a slope detector, will provide a bias for the tuning indicator in which there will exist audio-frequency components. If these audio components are allowed to appear between the grid and cathode of the 6E5 they would cause the eye to flicker, or in the case of higher audio frequencies they will cause the edge of the shadow to be indistinct. To prevent these effects the time constant of the low-pass filter must be long. Typical values are from 0.01 to 0.1 sec.

This circuit arrangement of the 6E5 provides only for a peak indication of tuning, with the inherent drawbacks that have already been listed. Various manufacturers of FM receivers have employed the 6E5 for zero or null tuning indication. The means by which this is done will be discussed in conjunction with another tuning indicator. All of the fluorescent display indicators operate on similar principles, so that the circuitry employed in any one type may be employed in any other.

## 12-4. The 6AF6

The 6AF6 is a tuning indicator that is capable of twin indications. That is, it has a fluorescent target about the same size and made of the same material as that of the 6E5 (thus the display is in green light), but this tube has two completely independent ray-control electrodes. Thus the 6AF6 is capable of producing two wedge-shaped shadows in the same manner as the 6E5 produces one. See Fig. 12-4. The main difference between the 6E5 and

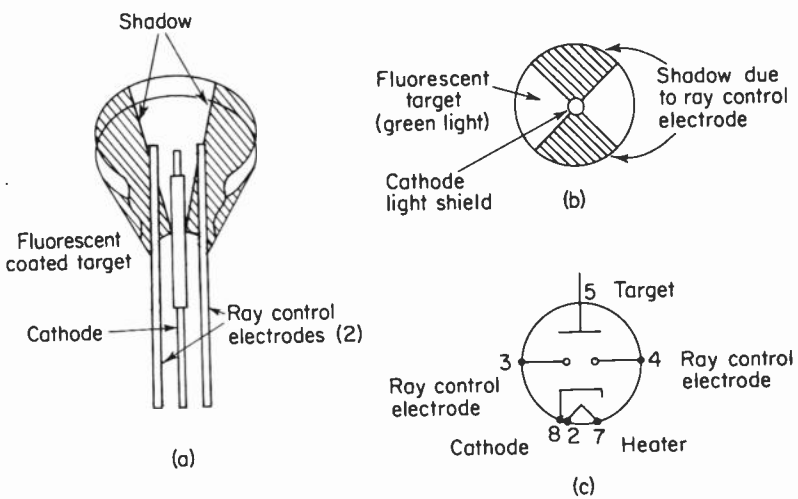



**Fig. 12-4.** The basic internal structure of the 6AF6 is shown in (a). The type of visual display is shown in (b), and the schematic diagram of the tube is illustrated in (c).

the 6AF6 is that the 6E5 has built into it a triode. Therefore, the 6AF6 requires external amplifiers for its operation. The 6AF6 is a peak or center-of-channel tuning indicator and finds its widest application in AM-FM tuners where tuning indications are needed for both AM and FM. An example of how the 6AF6 is employed is shown in Fig. 12-5. Notice that instead of separate amplifiers, existing circuits are employed to drive the ray-control electrodes. It can be seen that, just as in the 6E5, the ray-control electrode must be positive with respect to ground, so in the 6AF6 it must also be about 60 or 70 volts positive. Thus, these voltages are obtained from the plate circuit of an FM limiter and the plate circuit of an AM IF amplifier. Note that when the FM portion of the receiver is energized, the input signal to the limiter will develop grid-leak bias. Therefore, the DC plate current of the receiver will decrease, reducing the voltage drop across $R_1$. This results in an increase in the DC voltage from the ray-control electrode to ground, and the "eye" will close. The same reasoning is employed when the AM section of the receiver is energized, except that the grid bias of the last AM IF amplifier is obtained from the AGC bus.

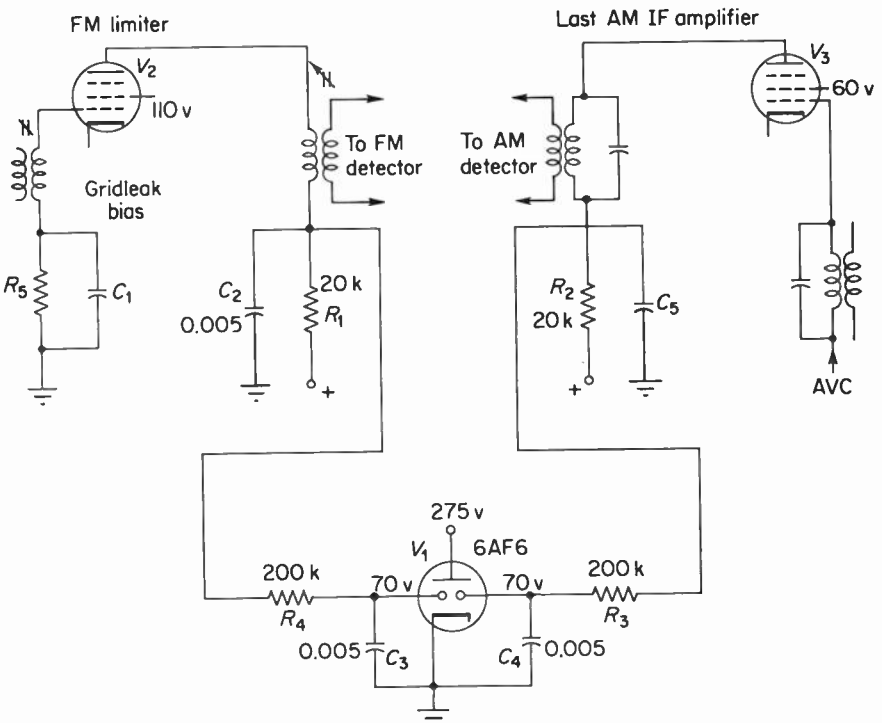



**Fig. 12-5.** A circuit utilizing the 6AF6 as a tuning indicator. Notice that the ray control electrodes are driven by separate dc amplifiers. These amplifiers are also being utilized as an AM IF amplifier and an FM limiter.

## 12-5. The 6AL7

The 6AL7 tuning indicator can be used as a peak or center-of-channel tuning indicator or as a null-type indicator. It appears most often in FM receivers in a capacity that utilizes both peak and null methods of tuning indication, and we shall limit our discussion here to this method.

Figure 12-6 is a sketch of the internal construction of the 6AL7. We can see at least six principal parts: the cathode, the boundary deflector, the comparison deflector, the FM control deflector, the space-charge grid, and the fluorescent target. The cathode's function is simply to emit electrons, which, as can be seen in Fig. 12-6, must pass between all three deflector electrodes. If any of the deflectors is made positive with respect to the cathode, the diverging electron beam will be deflected toward that electrode, increasing the beam's divergence. On the other hand, if any of the electrodes is made negative with respect to the cathode, the electrons, which are diverging because of their like charges, will be forced away from the electrode, and thus divergence of the electron beam will be reduced. The space-charge grid is generally at the same potential as the cathode, or it may even be slightly negative with respect to the cathode. Thus, the electrons in the region between cathode and space-charge grid will be moving rather slowly. As a result, the deflector electrodes can act upon the electron beam for a longer period of time (as compared to the situation with no space-charge grid), with a subsequent increase in deflection sensitivity. The space-charge grid is brought out to pin 1 at the socket and thus its potential may be varied. This results in a possible means of control of the brightness of the fluorescent display, since the grid will control the number of electrons that strike the screen.

The fluorescent screen is made of willemite, which when struck by electrons produces a green display. For maximum brightness, its voltage is adjusted to about 250 volts. The target (willemite screen) is coated with a thin layer of tin chloride, which provides a conducting surface for the return path of the incident electrons.

The tuning-indicator display obtained with this tube is seen in Fig. (12-6(a) and (c) to consist of four parts. The vertical division of Fig. 12-6(a) is obtained by means of a metal divider mounted on the willemite screen. The horizontal display area labeled $P_3$ is controlled by the boundary deflector, which as seen in Fig. 12-6(e) obtains its voltage from the grid-leak bias developed by the first limiter stage. As can be seen from Fig. 12-6(c), top view, the boundary deflector will vary the deflection of the entire bottom region of the display, since it is a solid strip. Thus, when the receiver is properly tuned, a large negative voltage will be applied to this electrode, which will repel the electron beam back to the cathode; thus the area of $P_3$ will shrink. See Fig. 12-6(d)(3). The maximum area of display for $P_3$ will occur when the boundary electrode is at zero potential, as would occur when the receiver is completely off tune.

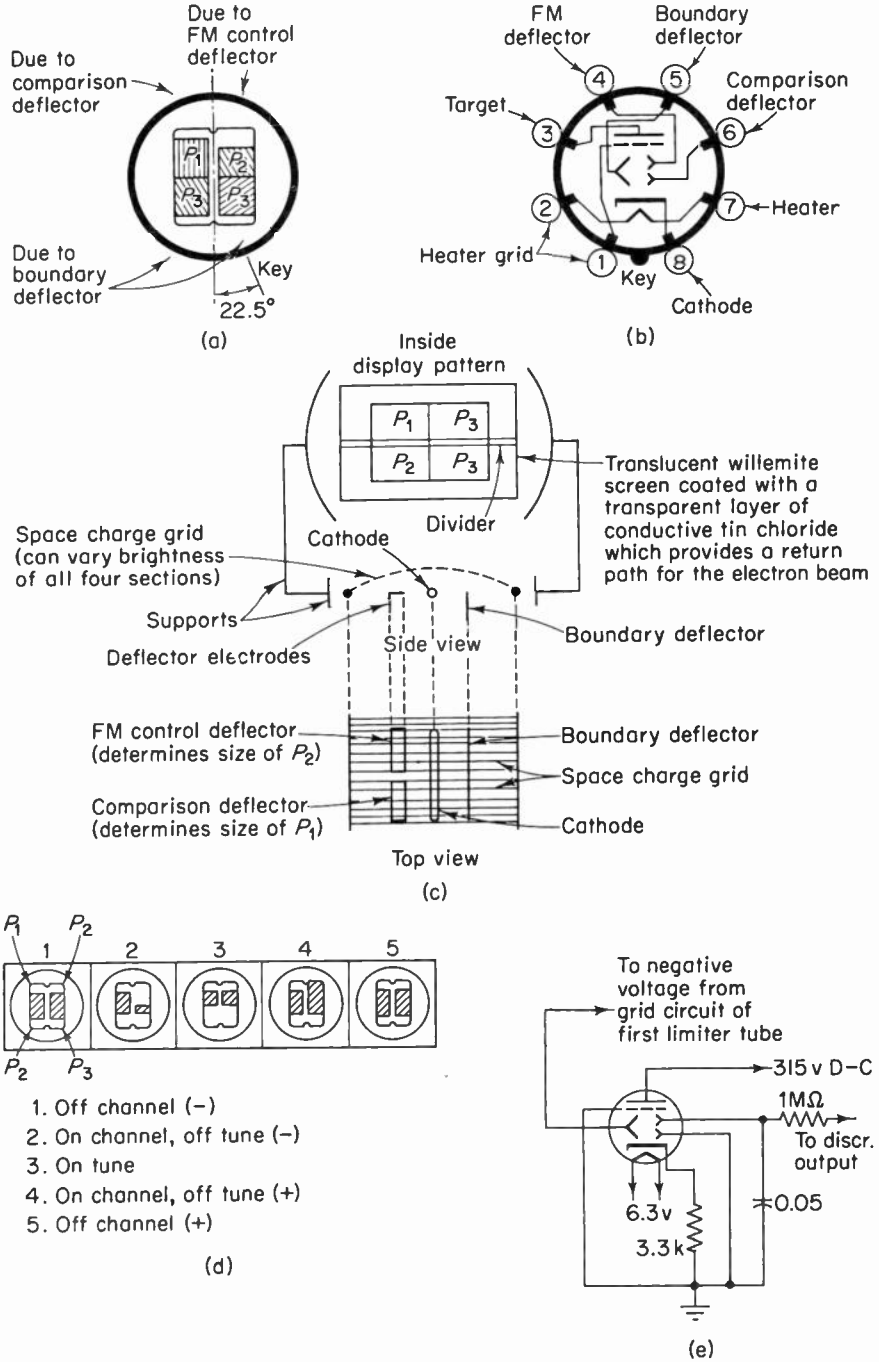



**Fig. 12-6.** (a) Is the fluorescent indication as seen on the face of the 6AL7. (b) Schematic diagram of 6AL7. (c) Sketch of the internal construction of the 6AL7. (d) Changes in display with changes in tuning. (e) Typical circuit arrangement for the 6AL7.

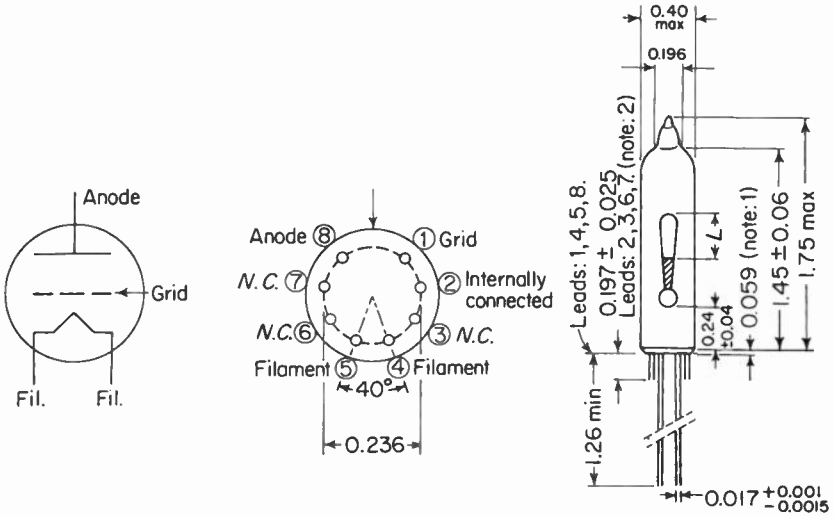
See Fig. 12-6(d)(1) and (5). Of course, for intermediary tuning positions the size of $P_3$ will look like Fig. 12-6(d) (2) and (4).

The comparison deflector ,which controls the area labeled $P_1$, is returned to ground [Fig. 12-6(e)]; thus its potential with respect to the cathode does not vary. The area of display, $P_1$, is therefore constant, and is used for comparison purposes. See Fig.12-6(d)(1) through(5).

The FM deflector is similar to the comparision deflector in both size and position with respect to the cathode. See Fig. 12-6(c), top view. The area of display that it controls is labeled $P_2$ [Fig. 12-6(a)]. The voltage applied to the electrode is obtained from the output of the balanced ratio detector or the Foster-Seeley discriminator. Since these points are also the audio take-off points, an *RC* filter is employed to prevent these audio components from reaching the FM deflector electrode. See Fig. 12-6(e). The effect on $P_2$ of tuning through a station is shown in Fig. 12-6(d). Notice that in diagram 1 (off channel —) $P_2$ is the same size as $P_1$; thus the output of the discriminator must be zero. In diagram 2 (on channel —) the output of the discriminator is negative; thus the electron beam is deflected back toward the cathode and the size of $P_2$ is reduced. When the receiver is in tune, the output of the discriminator is zero; thus $P_2$ will again be the same size as $P_1$. This is shown in diagram 3. Diagram 4 illustrates the condition where the receiver is off tune and the ouput of the discriminator is positive. A positive FM deflector electrode will result in increased deflection as compared to $P_1$. Thus, the area of $P_2$ is larger than that of $P_1$. When the receiver is again completely off tune, as in diagram 5, the output voltage of the discriminator will be zero and the area of $P_2$ will again equal $P_1$.

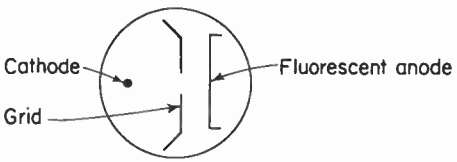
## 12-6. The DM70

Figure 12-7 shows a drawing of the DM70, and in the same figure will be seen a cutaway diagram of the tube. The tube has a subminiature envelope (diameter 10.1 mm; length less leads 38 mm) and thus may serve the dual purpose of tuning indicator and illuminated dial pointer. The tube consists of three structures: a filament; a perforated plate (grid), the perforation shaped like an exclamation mark; and an anode, which is coated with a fluorescent material. The fluorescent material when excited by electron bombardment glows with a greenish light. The tube is so mounted that the vertical filament is closest to the observer, and the observer is looking through the grid at the fluorescent anode. It might seem that the filament would obscure the fluorescent screen, but it does not, because it is made of very thin oxide-coated material and at the temperature at which it is operated it is barely visible. The filament may have either a DC or AC heater supply; it requires a source of 1.4 volts and draws 0.025 amp. If AC is used for the filament power supply 60-Hz hum may become a problem, because the



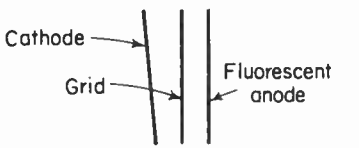

$L$ = length of light bar

(a)




Transverse cross section

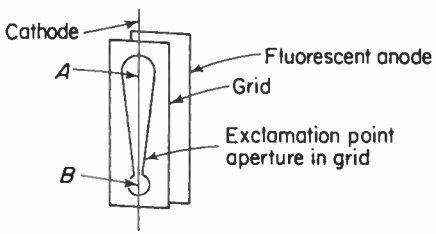
(b)




The converging structure as seen in the longitudinal cross section is required in order to cause area $A$ and area $B$ (see drawing below) to extinguish at the same time, when a large enough bias is applied to the grid

Longitudinal cross section

(c)




(d)

**Fig. 12-7.** Schematic and internal structure of the DM70. (a) Schematic and external dimensions. (b), (c) and (d) Illustrate the internal structure. [(a) *courtesy of Amperex*)]

temperature of the filament will vary as the power-supply current varies, and thus the emission from the filament will not be constant. Consequently, the plate current of the DM70 will contain a large 60-Hz component. The manufacturer therefore recommends that the plate voltage of the DM70 should not be obtained from a source such as the screen grid of one of the receiving tubes, since this type of connection will introduce hum into that tube, Furthermore, hum may be introduced, by induction, into the grid of the DM70. Since the grid may obtain its control voltage from those portions of the receiver which are sensitive to 60-Hz hum, it is recommended that a 6.8-megohm resistor be placed in series with the grid of the DM70. This resistor will act in series with a capacitor, thereby forming an AC voltage divider. Since the reactance of the capacitor at 60 Hz will be much smaller than 6.8 megohms, the voltage developed across the capacitor will be quite small. The capacitor serves a dual function in that it also acts as part of the low-pass filter between the DC tuning error-voltage take-off point and the grid of the DM70. See Fig. 12-8.

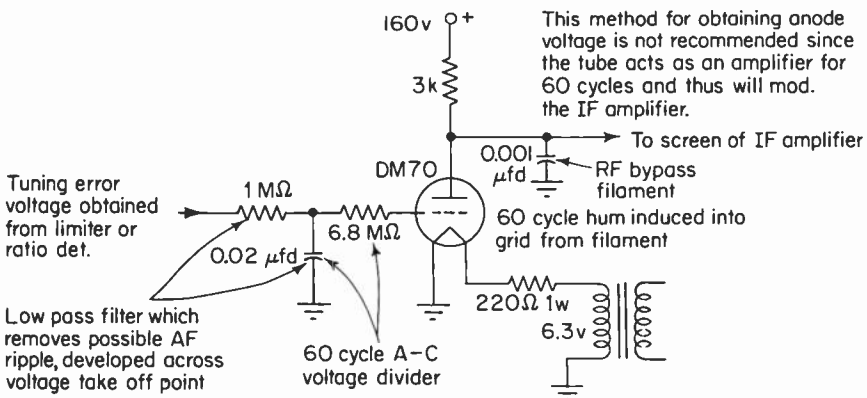



**Fig. 12-8.** Illustrates methods for reduction of hum which may be introduced when the filament is operated from an ac source.

This tuning indicator is a peak-voltage or center-of-channel indicator. Thus, the voltage required at the grid of the DM70 will be obtained from the ratio-detector stabilizer capacitor or from the limiter grid-leak bias.

A circuit that might be used in an FM receiver is shown in Fig. 12-9. In this circuit, the tuning error voltage is obtained from the grid-leak bias developed in the limiter stage of the receiver. The circuit that connects the bias network to the grid of the DM70 is a low-pass filter, including a resistor that acts as a means of hum reduction as pointed out earlier.

Operation of the circuit is reasonably simple. When the receiver is not tuned to a signal, the voltage developed across the grid-leak bias network

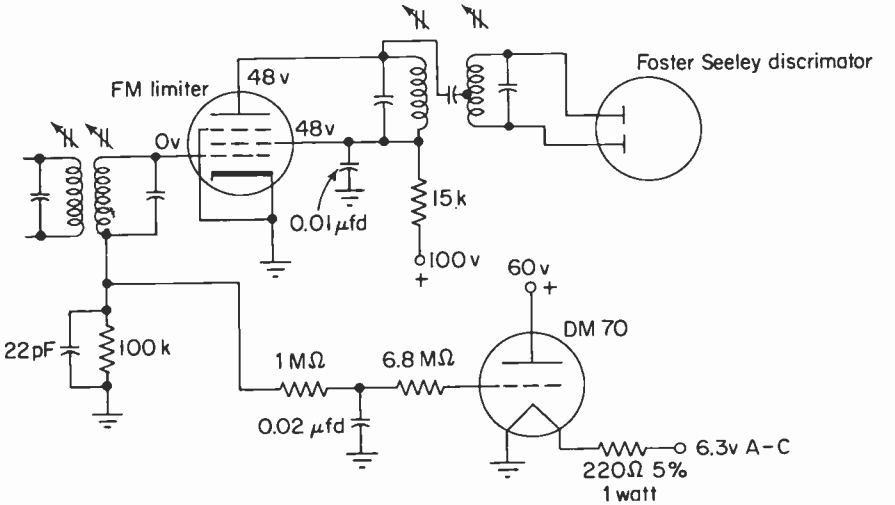



**Fig. 12-9.** A circuit illustrating a possible method of connection for the DM70.

will be minimum.† Thus the grid-to-cathode bias of the DM70 will also be minimum. The grid structure with its exclamation-point aperture will offer little or no opposition to the movement of electrons from the cathode to a plate that is positive with respect to the cathode. Electrons will, therefore, be able to pass through the aperture throughout its entire length (14 mm), and the observer will see the fluorescent glow of the anode in the shape of an exclamation point. On the other hand, when the receiver is tuned to a station, the grid will be biased to a negative value with respect to the cathode, and fewer electrons will be able to reach the plate. The electrons that reach the fluorescent anode will do so by passing through the widest aperture of the grid. That part of the anode behind the narrow part of the aperture will not receive enough electrons to fluoresce. Consequently, the length of the fluorescent bar will be shortened. The only regions that will be glowing will be the wide areas *A* and *C* [Fig. 12-7(d)]. If the signal is very strong (−7 to −10 volts), even areas *A* and *C* will be extinguished. Thus the fluorescent area of the DM70, in contrast to other types of tuning indicators, diminishes in size when the receiver is properly tuned.

Since the deflection process is electrostatic, it may be subject to interference from stray external electrostatic fields. To reduce the possibility of such interference the internal surface of the glass envelope is given a conductive coating, which is returned to ground via an internal connection to the

†Velocity effects will cause a minimum negative voltage to be developed across the bias network.

filament. Thus the tube is surrounded by a transparent electrostatic shield that does not impede the observation of the tuning display.

## 12-7. The 6BR8/EM80

A fluorescent tuning indicator popular because of its esthetic properties is the EM80. This center-of-channel indicator displays a fanlike fluorescence (see Fig. 12-10) that becomes larger as the tuning error voltage increases.

Figure 12-11 gives a schematic diagram of the EM80 as well as a diagram of the internal structure. The tube consists of two basic parts: a triode amplifier and a fluorescent indicating system. Both are activated by a common cathode. The indicator portion of the tube is mounted above the cathode and the triode section is mounted below the cathode. This arrangement avoids possible mutual interference of one system upon the other, while permitting the use of a single cathode.

The indicating portion of the tube is made up of five parts; they are, in order from the cathode: (1) the space-charge grid, which acts as a shield,

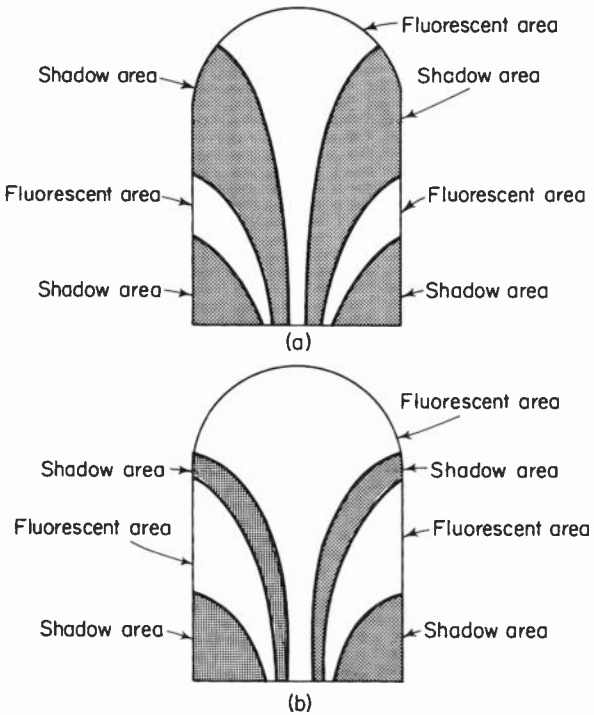
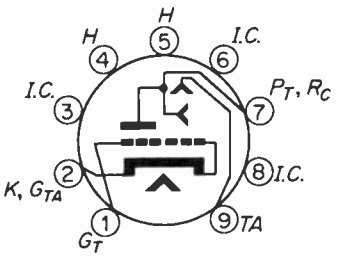


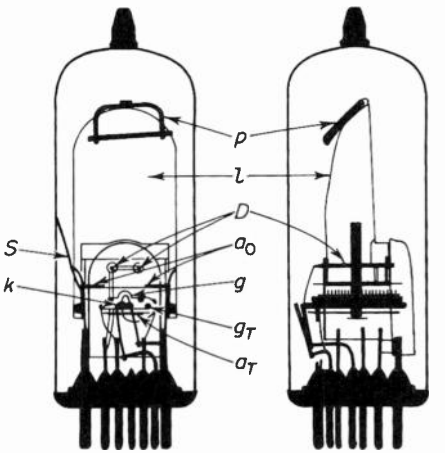

**Fig. 12-10.** (a) Illustrates the display pattern of the EM80 for conditions of weak or no signal reception. (b) Represents the display pattern when a signal is being received.




Pin connections

1 – Grid, triode
2 – Cathode, target grid
3 – Internal connection
4 – Heater
5 – Heater
6 – Internal connection
7 – Plate of triode, ray control electrode
8 – Internal connection
9 – Target

(a)




(b)

Radiographs of the EM 80 ($p$ = getter; $l$ = luminescent screen; $D$ = deflection electrodes; $a_0$ = accelerator electrodes; $g$ = space charge grid; $k$ = cathode; $g_T$ = grid of the triode section; $a_T$ = anode of the triode section; $S$ = spring)

**Fig. 12.11.** (a) Shows the schematic diagram and pin connections of the EM80. (b) Illustrates the internal structure of the EM80 by means of an Xray photograph, *courtesy of Amperex.*

(2) an accelerator electrode, which is at the same potential as the fluorescent screen and thus accelerates the electrons on their way to the screen, (3) and (4) two deflection electrodes, which are internally connected to the plate of the triode and cause deflection in a manner similar to that of the 6E5, and (5) the fluorescent screen itself, which converts the kinetic energy of the electrons into a greenish light.

Under no-signal conditions the triode section of the EM80, shown in a typical circuit arrangement in Fig. 12-12, conducts heavily. As a result the plate voltage is low, since there is a large voltage drop across the plate-load resistor. The deflection electrodes are, therefore, less positive than the screen. Thus, an electrostatic field is set up between the screen and the de-

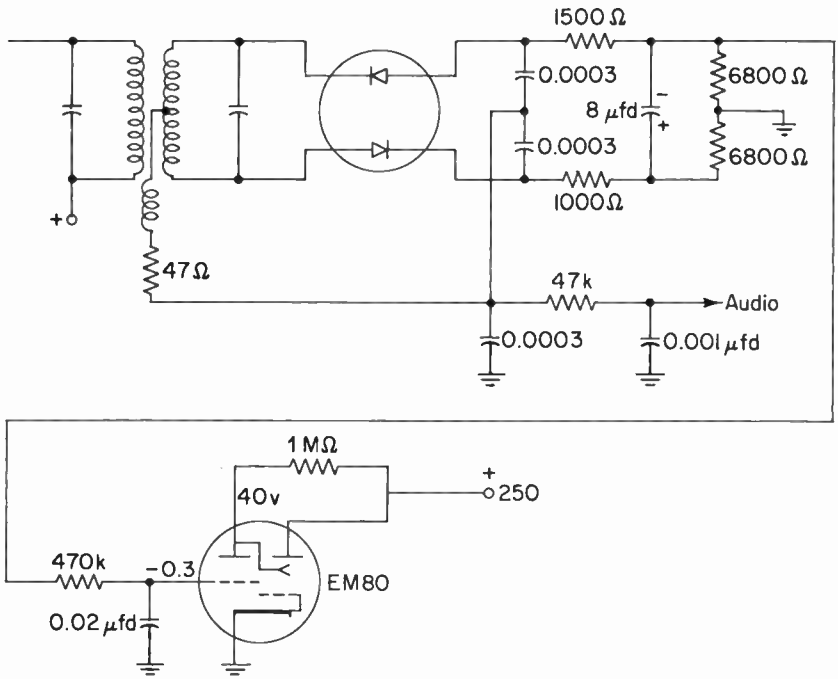



**Fig. 12-12.** The EM80 is shown in a typical circuit arrangement. The tuning error voltage is obtained from a balanced ratio detector.

flection electrode, which results in a display of maximum shadow areas [Fig. 12-10(a)]. When the receiver is tuned to a station, the grid voltage of the triode section is now negative. Less plate current flows, and so the voltage drop across the plate-load resistor is reduced, with a consequent rise in plate-to-cathode voltage. Since the difference of potential between the deflection electrodes (internally connected to the triode plate) and the fluorescent screen is now reduced, their associated electrostatic field must also be less intense. This reduction in electrostatic-field intensity causes the shadow areas to become smaller [Fig. 12-10(b)]. The condition of best tuning occurs when the shadow areas are smallest, or conversely when the fluorescent areas are largest.

The inner surface of the glass envelope is coated with a transparent conductive layer. This layer connected to the fluorescent screen by means of a spring, serves two purposes. First, it shields the tube from external stray electrostatic fields, which might disturb the indication. Second, it prevents the glass envelope from developing a positive charge (due to secondary emission) which might produce undesirable effects.

The input circuit to the grid of the EM80 is the same as that for the 6E5. Therefore, the comments made in that discussion apply here.

## 12-8. The 6FG6/EM84

The EM84 is a center-of-channel fluorescent indicator that provides a display of white fluorescence rather than the almost standard green of the other types. This indicator has been given various trade names; thus it may be called "Micro tune" or "Acro tune" depending upon the manufacturer. The display pattern is in the form of a narrow white horizontal bar, in the center of which is a vertical shadow that changes in width as a signal is tuned through. See Fig. 12-13. The sketches show that when the signal is weak or nonexistent, the shadow is of maximum width. When the receiver is tuned to a signal of moderate strength, the shadow width will decrease. Of course, the stronger the signal the narrower the shadow, until under very strong-signal conditions the shadow area will disappear.

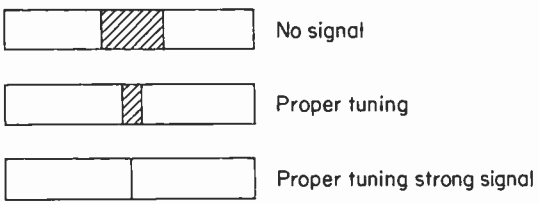



**Fig. 12-13.** Fluorescent pattern of the EM84 for different conditions of tuning.

The schematic diagram of the EM84 as well as a diagram of the internal construction of the tube is given in Fig. 12-14. We can see that basically this tube is the same as the 6E5. That is, they both contain a triode section, a ray-control electrode, and a fluorescent screen for tuning indication. The main difference between the two tubes is in layout. In the EM84 the cathode and the ray-control electrode are positioned somewhat like a motion picture projector, thereby casting a beam of electrons upon a fluorescent screen which is placed at right angles to the beam. As in the 6E5 the degree of deflection of the electron beam depends upon the intensity of the electrostatic field between the ray-control electrode and the fluorescent screen (target). Therefore, if the potential between the screen (which is at $B^+$ potential) and the ray-control electrode (plate of triode) is zero, then the deflection must also be zero. Since the deflection of the electron beam is zero, no shadow area will appear on the screen of the indicator. This condition will occur when the receiver is tuned to a station. As in almost all of the other fluorescent indicators the negative tuning error voltage (obtained from the limiter or the ratio detector) will be applied through a low-pass filter to the grid of the tuning indicator. This will cause the plate voltage of the triode section to rise, reducing the potential difference between the ray-control electrode and the target. This action, as

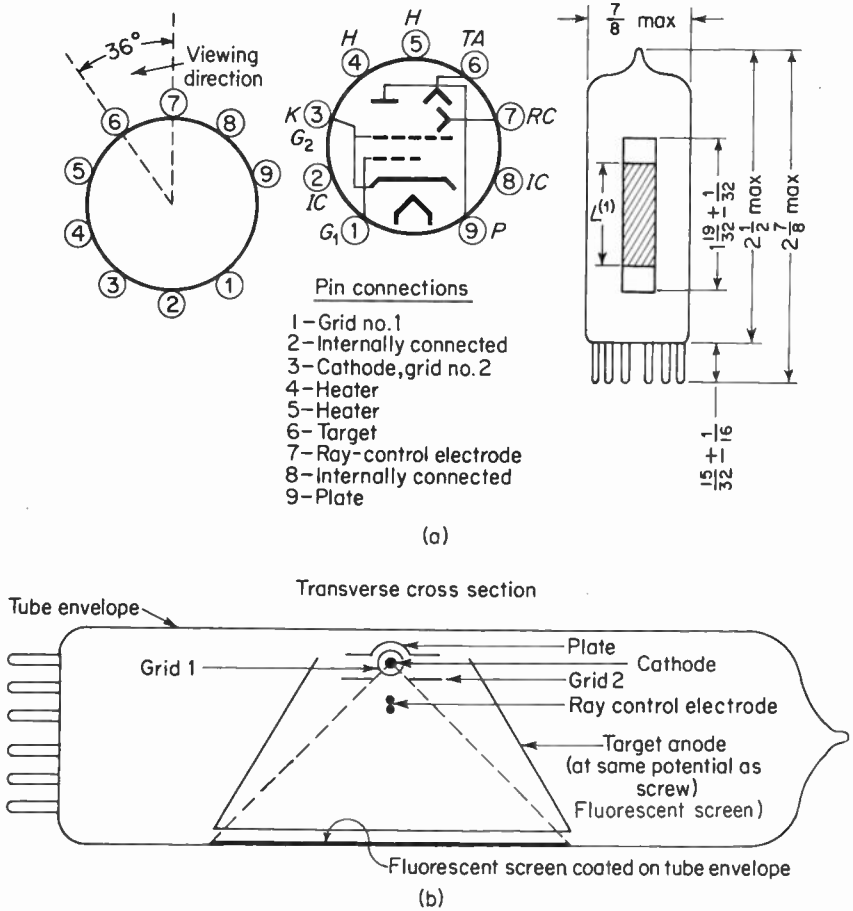



**Fig. 12-14.** (a) Schematic symbol for the EM84, together with a diagram of physical dimensions, *courtesy of Amperex*. (b) A transverse cross section of the EM84 showing the internal construction of the tube.

pointed out above, will cause the shadow area to diminish in size. Naturally, under no-signal conditions the reverse action occurs and the shadow area becomes larger.

A typical circuit using the EM84 is shown in Fig. 12-15.

The inner surface of the EM84 is coated with a transparent conductive layer that acts as an electrostatic shield.

The EM84, like other fluorescent indicators, may be employed as a null-type indicator if so desired. For this purpose the grid of the tuning indicator must be fed by a tuning error voltage obtained from the audio take-off point of the ratio detector or the Foster-Seeley discriminator. As is well known, when a receiver employs these detectors and is properly tuned, the DC output

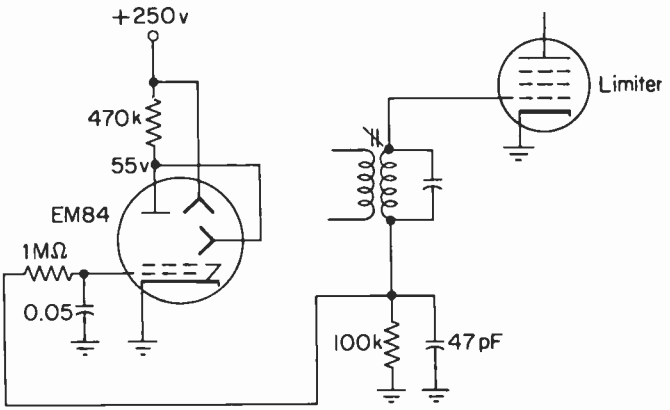



**Fig. 12-15.** Typical circuit employing the EM84.

of the detectors is zero. The fluorescent indicator, to be used as a null detector, must be arranged to give a definite indication when the tuning error voltage applied to the grid is zero. The only definite indication occurs just after the shadow area has become so narrow it disappears and just before the display indicates an overlap in the fluorescent pattern. Thus, any circuit using the EM84 as a center-of-channel indicator will have to be so designed that for zero grid input voltage the shadow area will be a very thin, almost nonexistent line. Of course, on either side of proper tuning the indicator display will either overlap or open up to a wider shadow.

A circuit designed to provide such a display is shown in Fig. 12-16(a). In this circuit a form of combination bias (fixed and cathode) provided by $R_1$ and $R_2$ is used to set the grid bias of the triode section of the tube to approximately plate-current cutoff. As has been pointed out, this will result in zero potential difference between the fluorescent screen and the ray-control electrode, with the result that the shadow area will be at most a very thin line [see Fig. 12-16(c)]. $R_1$ is made variable in order to permit adjustment of bias at any time during the life of the receiver. (Readjustment may be made necessary by aging of components and the like.) When the receiver is not tuned properly, and the output of the detector is a positive voltage, the grid-to-cathode voltage of the EM84 will decrease. Plate current will now flow, and the shadow area will widen. This is illustrated in Fig. 12-16(b). When the receiver is properly tuned, the grid bias of the tuning indicator will return to zero. Since this is the normal operating condition of the tube, the display pattern will once again become a thin line. At some point in the tuning process the output of the detector will be negative. This voltage is coupled to the grid of the tuning indicator; therefore, the triode section of the indicator is driven further into cutoff. It should be noted at this time that cutoff is not a point but is a range of voltages. Therefore, the plate voltage of the triode

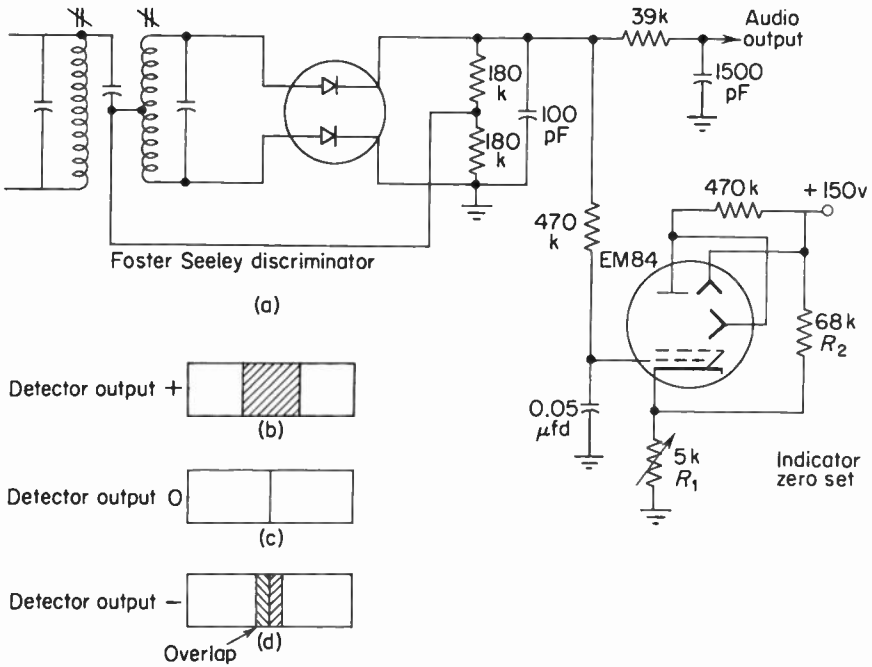



**Fig. 12-16.** (a) A circuit in which the EM84 is used as a center of channel indicator. (b), (c) and (d) Illustrate the changes that take place as the receiver is tuned through a station.

will become somewhat larger as the grid voltage becomes more negative. This will cause the shadow area of the indicator to become thinner and eventually to disappear and be replaced by an overlapping of the fluorescent display. See Fig. 12-16(d).

## 12-9. Neon-bulb tuning indicators

A simple form of tuning indicator that has been used in both the United States and England, is the neon-bulb indicator. These indicators are generally peak-reading, and they display proper tuning by changing in brightness or by becoming extinguished.

A popular type of neon-bulb indicator is shown in Fig. 12-17. Its circuit essentially consists of a variable resistor ($V_1$), which controls the current through the neon bulb and thereby controls its brightness. Under no-signal conditions the voltage from grid to cathode of $V_1$ is minimum, because the tuning error voltage obtained from the limiter's grid-leak bias or from the ratio detector's stabilizer voltage is almost zero. Thus the neon bulb will be at maximum brightness, since the current through both the triode and the

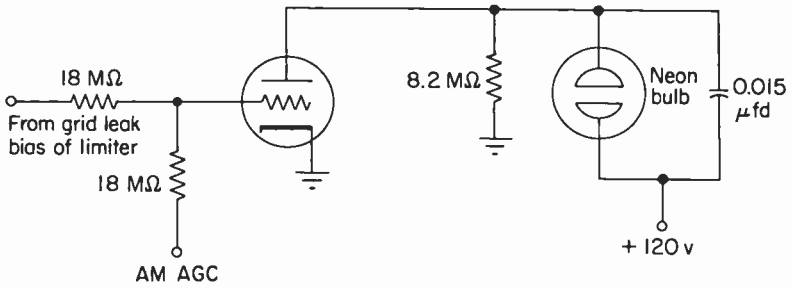



**Fig. 12-17.** A neon-bulb tuning indicator.

neon bulb will be maximum. When a signal is properly tuned, the negative tuning error voltage between grid and cathode of $V_1$ will be maximum (peak). This increased bias level will cause less plate current to flow through the neon bulb, and its brightness will diminish. (Under very strong-signal conditions the bulb may visibly flicker, because the bulb, $R_1$, and $C_1$ act as a relaxation oscillator.) Since the eye is more sensitive to changes in light intensity at low than at high illumination levels, the fact that the tuning center occurs at minimum light levels permits greater accuracy of tuning.

## 12-10. Tuning meters

For ease of reading and for possible greater accuracy of tuning (richness of style should not be ignored) many manufacturers of FM tuners have been using meters as tuning indicators. These meters may be employed as either center-of-channel or peak-signal indicators. Center-of-channel indication is provided by a meter whose movement is designed so that the zero or pointer rest is at the center of the scale. Signal-strength (peak) meters are full-scale deflection meters of conventional design. Most manufacturers confine themselves to one or the other type of indicator, but some have utilized both methods on the same chassis. This arrangement leaves no room for doubt as to the perfection of alignment of the receiver, but is more of an eye-catcher than a functional necessity. The main drawback associated with tuning meters is their expense; thus meters have not been as widely used as fluorescent indicators.

Meters may be found in FM receivers in any of three circuit arrangements: (1) the simple voltmeter, (2) the separate-tube VTVM, (3) the existing-circuit VTVM.

The simplest method, and one often used in transistor receivers, is that of employing a very sensitive meter (50 $\mu$amp or less) in conjunction with a multiplier resistance (100 k or 500 k) as a voltmeter. The problem is that the meter tends to load any circuit into which it is connected. Thus, if this arrangement is used as a center-of-channel tuning indicator, it may

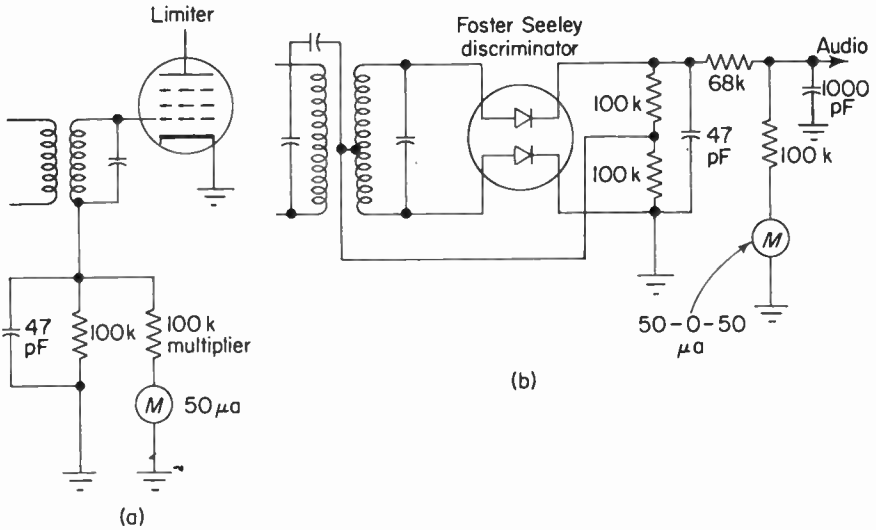



**Fig. 12-18.** (a) A simple voltmeter connected as a peak reading indicator. Notice that as a result of the multiplier resistance the time constant of the grid leak circuit is reduced. (b) A simple voltmeter connected as a zero type indicator. Loading is also a problem here.

tend to change the detector circuit to the point where additional distortion and reduction of output would be the result. Naturally, the smaller the current demand by the meter upon the circuit into which it is connected, the smaller the effect the meter will have on the circuit. Meters of high sensitivity are rather expensive and, therefore, are seldom employed. Possible methods of connection are illustrated in Fig. 12-18.

A solution to the problem of loading is offered by the vacuum-tube voltmeter. As is well known, the impedance presented by the input terminals of the VTVM may be of the order of 10 megohms or more. Therefore, the circuit into which it is connected will change very little, provided it is of low impedance by contrast. Such a circuit is seen in Fig. 12-19. This circuit is basically a Wheatstone bridge, with the meter movement as the cross arm of the bridge. The triode vacuum tube acts as the variable-resistance arm of the bridge. Changing the grid voltage of the tube causes the DC resistance of the tube to vary, causing a change in current through the meter. The resistance $R_3$ is used to set the static balance of the circuit. This is generally done with the grid shorted to ground so as to provide zero input voltage to the grid of the vacuum-tube voltmeter (VTVM).

If the VTVM is to be used as a signal-strength tuning meter, the meter should be of the full-scale deflection type, and the balance control $R_3$ is set for zero reading of the meter. When the FM receiver is tuned to a station, the tuning error voltage (obtained from the limiter or ratio detector and

The tuning error voltage is obtained from a limiter or a ratio detector when a meter is employed as a signal strength meter. The output voltage of a Foster Seeley discriminator or the output of a ratio detector, supplies the tuning error voltage when a meter is to be employed as a center of channel indicator.

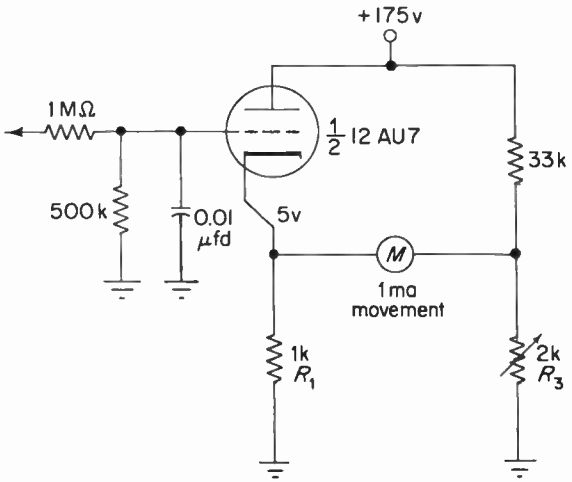



**Fig. 12-19.** A Wheatstone bridge VTVM tuning indicator.

delivered to the VTVM via a low-pass *RC* filter) will be maximum negative. The plate current of the tube will decrease and the bridge will become unbalanced. The degree of unbalance will be indicated on the meter, thereby showing the magnitude of the signal and the correctness of tuning.

If it is desired to use the VTVM as a center-of-channel tuning indicator, the meter movement will be of the zero-center-scale type. As was done with the peak-tuning meter, $R_3$ is adjusted under no-signal conditions for bridge balance—or, what is the same thing, zero indication. The DC tuning error voltage, obtained from the output of the Foster-Seeley or ratio detector, is fed to the grid of the VTVM, where (depending upon the correctness of tuning) the bridge balance will be altered. When the receiver is properly tuned and the tuning error voltage is zero, bridge balance is not upset and thus the meter remains at the zero-center-scale position. Above or below this point of proper tuning, the tuning error voltage will be positive or negative. These voltages will unbalance the bridge in one direction or the other, depending upon whether the current in the tube increases or decreases.

Notice that the tube and $R_1$ constitute a cathode-follower amplifier. The main advantages to be gained by such an arrangment is that of high input impedance and circuit stability. That is, changes in B supply voltages, aging of the tube and of components, and the like will have less effect upon the circuit than would an amplifier (used as a VTVM) without feedback.

Many manufacturers of AM-FM tuners employ VTVM tuning meters using such circuits as the RF amplifier, converter, or other existing amplifier circuits for the VTVM portion of the tuning indicator. The obvious advantage of such a practice is to reduce manufacturing costs. Two circuits that use existing circuits as tuning-meter vacuum-tube voltmeters are illustrated in Fig.

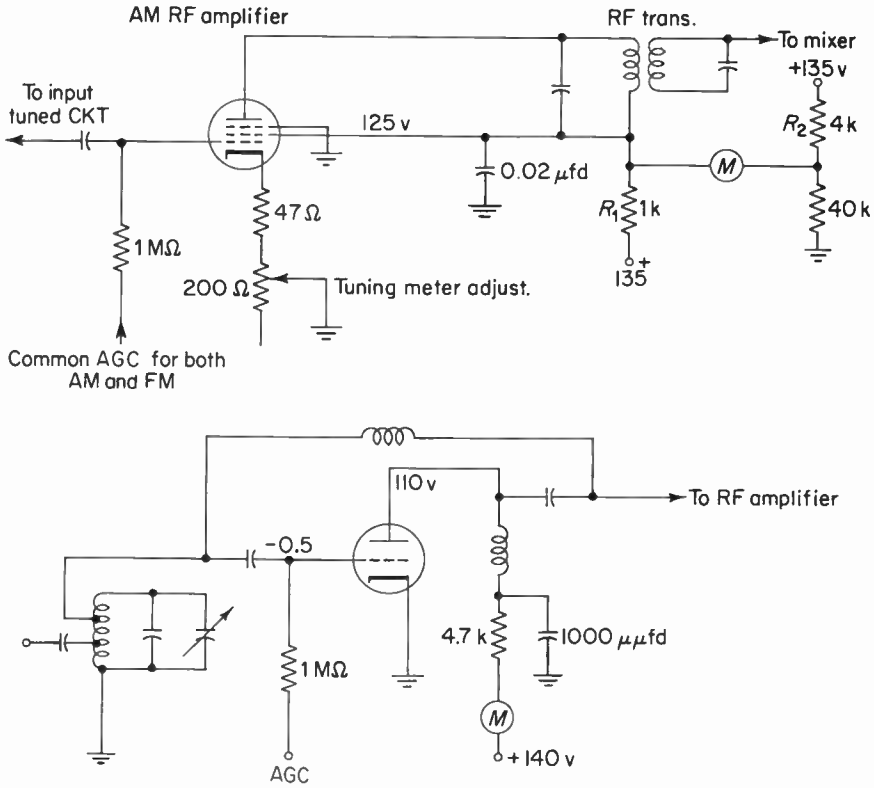



**Fig. 12-20.** (a) A VTVM tuning meter utilizing the AM RF amplifier as the variable arm of a Wheatstone bridge. (b) The average current of a RF cascode amplifier is determined by the level of ac voltage and is measured by the meter.

12-20(a) and (b). In Fig. 12-20(a) the cathode-to-screen portion of the AM RF amplifier is utilized as the vacuum-tube portion of the VTVM bridge. The bridge is placed in balance by adjusting the variable cathode resistor so that the voltage drop across $R_2$ is of the same magnitude as that across $R_1$. Thus, the difference of potential across the meter is zero. When the receiver is tuned to a station, the AGC bias applied to the AM RF amplifier will increase, which in turn causes a decrease in screen current, thereby unbalancing the bridge. If the meter is connected properly, the needle will deflect upward, indicating that the signal has been tuned to the center of the passband.

The AGC bus is fed by both the AM detector and the FM limiter. Therefore, the tuning meter may be used as a signal-strength indicator for both AM and FM.

Not all tuning meters employ the Wheatstone bridge as a means of activating the meter movement. A case in point is shown in Fig. 12-20(b). In this circuit the meter is used as a peak reading meter and is in series with the plate circuit of the cascode RF amplifier. The grid circuit is tied to the AGC bus, and thus the bias of the tube varies with signal strength. When a station is tuned in, the AGC voltage (obtained from the first limiter) will rise, and the plate current of the tube will decrease proportionately. This change in plate current will be seen on the tuning meter as a deflection from the meter's normal position.

## 12-11. Summary

**1.** The quality of the audio output of the FM receiver is determined to a large measure by proper tuning.

**2.** Generally the operator of an FM receiver tunes the receiver for least distortion and noise.

**3.** AM receivers are tuned for maximum audio level as well as minimum distortion.

**4.** Improper tuning occurs when the IF falls on a region of the S curve that is nonlinear.

**5.** Improper tuning of the receiver introduces harmonic and intermodulation distortion.

**6.** Narrow-band FM detectors are subject to possible tuning difficulty as a result of side response.

**7.** Wide-band FM detectors do not experience side response at all.

**8.** AFC ensures proper tuning by means of a feedback system. This system corrects an error in tuning by providing an error voltage that forces the local oscillator to shift in frequency, in the proper direction to reduce the tuning error to a very small amount.

**9.** There are three methods for tuning indication: (1) signal-strength (peak), (2) center-of-channel (zero), and (3) systems that combine both peak and zero methods.

**10.** Tuning indicators are only as good as the overall alignment of the receiver.

**11.** Tuning indicators fall into three categories: (1) fluorescent indicator tubes, (2) neon-bulb displays, (3) meters.

**12.** The fluorescent indicator tubes are almost always used as signal-strength indicators, the fluorescent display area being directly proportional to signal strength. One exception is the DM70, which provides a fluorescent indication that is inversely proportional to the signal level.

**13.** Filters are employed between the tuning error-voltage take-off points and the grid of the tuning indicator in order to prevent the indicator from loading the sources of the tuning error voltage. These filters also prevent

audio components from being applied to the tuning indicator, and thereby prevent possible flicker of the fluorescent indicators.

**14.** Tuning error voltage, which is applied to the signal-strength tuning indicator, is obtained from the limiter grid-leak bias or from the ratio-detector stabilizing capacitor. The Foster-Seeley discriminator cannot provide this voltage, since it is preceded by a limiter and its output is constant.

**15.** Center-of-channel tuning indicators obtain the tuning error voltage from the output of either the ratio detector or the Foster-Seeley discriminator.

**16.** Most of the fluorescent indicators contain built-in triodes that provide additional sensitivity.

**17.** Neon bulbs provide a cheap and simple method of tuning indication.

**18.** Meter movements make excellent tuning indicators because they provide ease of reading and greater possible accuracy.

**19.** Tuning meters may be used for either peak or center-of-channel indications.

**20.** Tuning meters may be found in one of three circuit arrangements: (1) the simple voltmeter, (2) the Wheatstone bridge VTVM, (3) the existing-circuit simple vacuum-tube voltmeter.

## REFERENCES

1. Burstein, H.: "Adding an FM Tuning Indicator," *Radio and Television News*, March 1957.
2. Bailey, F.M.: "An Electron-Ray Tuning Indicator for Frequency Modulation," *Proc. IRE*, October 1947.
3. Liddell, W.E.: "A TV Tuning Indicator," *Radio Electronics*, June 1957.
4. Syrjala, J.: "Tuning Eye or Tuning Meter," *Radio and Television News*, January 1956.
5. Reed, L.: "Outboard Signal-Level Indicator," *Electronics World*, November 1960.
6. Kiver, M. *FM Simplified.* D. Van Nostrand Co., Inc., Princeton, N.J., 1960.
7. "FM Tuning Indicator" (neon bulb), *Radio Electronics*, February 1956.
8. "FM Tuning Meter," *Radio Electronics*, October 1956, p. 137.

# 13

# MISCELLANEOUS TOPICS

## 13-1. Squelch or muting circuits for interstation noise suppression

If an FM broadcast receiver is accurately tuned to a station carrier, whether the carrier is frequency-modulated or not, the balanced form of FM detector is capable of rejecting any amplitude modulation provided the limiting threshold is exceeded. On the other hand, if the receiver is tuned between two carriers, the limiters obviously cannot be saturated since the carrier level at this time is zero. Random noise generated at the front end of the receiver and amplified by the IF amplifiers (and limiters) emanates as a loud hissing, sizzling sound from the loudspeaker. The degree of noise output depends upon a number of factors, not the least being the type of detector employed. A ratio detector can reject a certain amount of this noise by virtue of its self-limiting capabilities, but a Foster-Seeley arrangement will provide little rejection. A number of commercial designs provide interstation noise suppression to give quieter tuning between stations. Most of these designs use the principle of biasing off a stage somewhere in the signal path when the receiver is not tuned to a station, making use of the limiter bias or some other voltage that is present with a signal and not present between stations. These arrangements are referred to as "muting," "squelch," "silencers," and the like.

The basic idea of muting is illustrated in Fig. 13-1. Here a squelch tube, $V_1$, has its control grid tied to the limiter bias point, so the developed bias at the limiter must vary the bias of $V_1$. The cathode of $V_1$ is returned to the

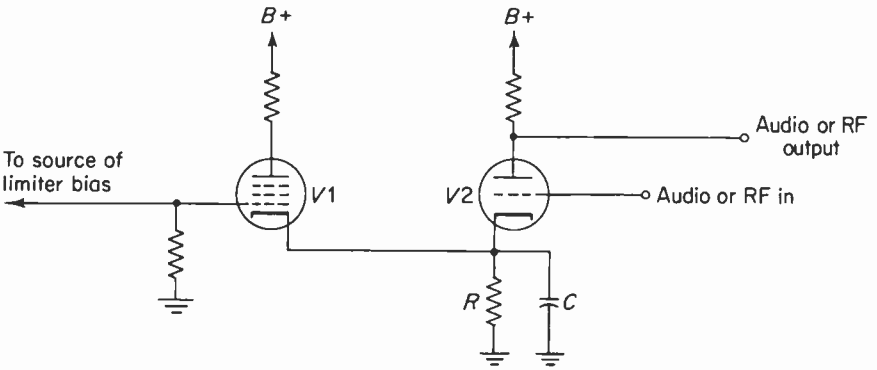



**Fig. 13-1.** A basic form of squelch or muting circuit. $V_1$ cuts $V_2$ off when the limiter bias is zero.

cathode of a tube in the signal path (the last limiter tube, the audio amplifier tube, and so on), with both cathodes returning to ground through a common bias network. Since limiter bias is proportional to RF carrier amplitude, the bias of $V_1$ will be reduced, because its control grid is tied to the source of negative grid-leak bias. This will result in heavy current flow through $R$. This increases the bias sufficiently to cut $V_2$ off, and no output appears at the plate of $V_2$. No hiss can be heard, since the audio circuits have no input. Should the receiver be tuned to a carrier, the limiter bias will increase to cut off $V_1$. The magnitude of current through $R$ will decrease. This will reduce the bias, and $V_2$ will be brought out of cutoff. The signal (audio *or* RF) is now permitted to reach subsequent stages.

This circuit should give the reader some idea of how interstation noise suppression might be accomplished. Its drawbacks will become apparent as the discussion progresses.

## 13-2. A more sophisticated muting circuit

The functional arrangement of Fig. 13-2 uses the principle described above, but it is more effective since it provides sharper cutoff of the signal circuit along with means to vary the point of cutoff relative to the strength of the incoming carrier.

The configuration shown uses a 6BN6 gated-beam tube ($V_1$) as a limiter and also provides a DC voltage proportional to the carrier level for controlling conduction through a muting triode ($V_2$). The plate voltage of $V_2$ thus varies as a function of limiter bias and in turn controls the potential at point $A$. Since the quadrature grid of $V_1$ is tied to point $A$ (via the muting switch), the bias for this grid must vary with the potential at point $A$. Either grid of a 6BN6 tube is capable of sharply cutting off the plate current if either grid is operated below cutoff. This characteristic is used in this circuit.

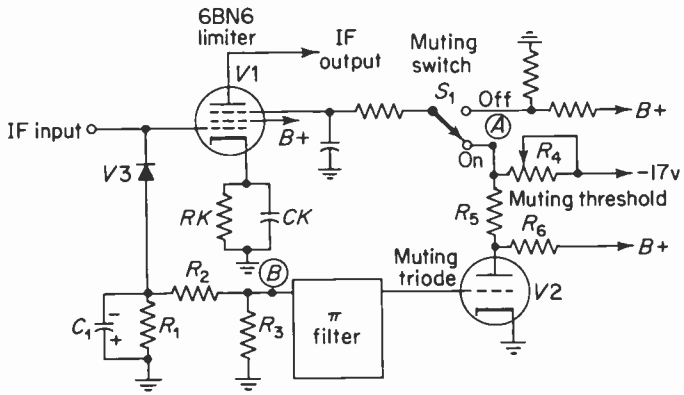



**Fig. 13-2.** A form of squelch or muting which utilizes the sharp cut-off characteristic of the 6BN6 quadrature grid.

With the muting switch in the OFF position, the quadrature grid is switched to a source of fixed positive bias and the muting circuit is effectively disabled. This would be desirable for tuning in a very weak station requiring the highest possible sensitivity. Now, assume that the muting switch is in the ON position. This switches the quadrature grid to point *A*. The control grid of the muting triode is connected to point *B* (via a low-pass filter). Since the gated-beam tube operates with cathode bias only, some means must be available to provide a negative DC voltage proportional to the carrier. The semiconductor diode, $V_3$, conducts on negative carrier peaks to provide a negative voltage across $R_1$-$C_1$. This potential is applied to the $R_2$-$R_3$ voltage divider and is used to vary the bias (and plate conduction) of $V_2$. Thus, the voltage at point *B* is negative to ground and proportional to the average carrier level.

Assume that the tuner is adjusted between stations. Since there is no carrier voltage present at the cathode of $V_3$, its operating bias is essentially zero (neglecting contact bias) and it conducts heavily. Point *A* now becomes sufficiently negative (owing to the −17-volt supply) to cut off $V_1$. The noise inevitably present at its control grid cannot pass through to the AF stages and the receiver is "quiet" when tuned between stations.

Assume, now, that the receiver is tuned to the center of a channel and that the carrier strength is sufficient to cause $V_3$ to conduct heavily on the negative peaks. The rectified DC component of voltage at point *B* is now large enough to cut off $V_2$. The plate voltage of $V_2$ will now rise toward the 180-volt B-supply. This, in turn, causes the potential at point *A* to become sufficiently positive to bias $V_1$ into conduction. The IF signal can now pass through $V_1$ and subsequent stages, and the audio output will be heard.

The purpose of the threshold control is to vary the amount of negative

voltage at point $A$ so that the cutout (or cut-in) point for $V_1$ can be changed to correspond to the amplitude of the input carrier.† For reception of very weak signals, where the carrier signal is not much greater than the noise level, it would not be desirable to cut off $V_1$. As the $R_4$ control reduces the amount of negative voltage at point $A$, the conduction through $V_2$ will not be capable of reducing the plate voltage sufficiently to bias off the limiter tube on weak signals. If the muting control is adjusted to increase the negative potential at point $A$, a strong input carrier will be required to reduce $V_2$ conduction sufficiently to override the negative bias at point $A$. This action will permit $V_1$ to go into conduction. Thus, the threshold at which muting takes place can be varied by means of $R_4$, or the muting can be completely disabled by throwing switch $S_1$ to OFF. This will connect a fixed forward bias to the quadrature grid of the limiter.

A form of squelch circuit used in some forms of military equipment makes use of the property of biasing off the discriminator with no signal present at the detector input. This is shown in Fig. 13-3. $R_1$ and $R_2$ form a voltage divider. A positive voltage is applied to the cathodes of the discriminator diodes and, with no signal present at the discriminator input, the diodes are reverse-biased and no noise is heard, since no signal appears at the sound take-off point. As soon as a signal appears at the input, the diodes are driven into conduction (provided the peak input voltages to the diodes exceed the reverse bias) and the demodulated audio appears at the sound take-off point. Resistor $R_1$ is variable in order to provide a squelch level consistent with signal amplitude. The higher the reverse bias, the larger must be the input signal to cause diode conduction.

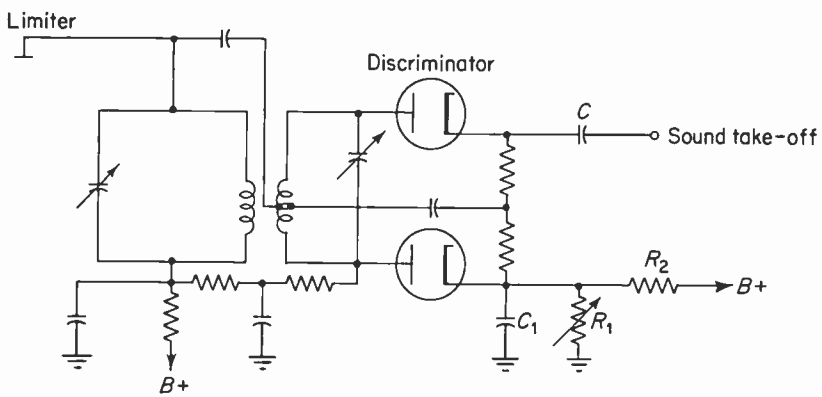



**Fig. 13-3.** An interstation squelch circuit which holds the discriminator diodes biased off in the absence of a signal input. A signal overrides the reverse bias and allows the audio to appear at the sound take-off point.

†This is the major drawback of the circuit in Fig. 13-1.

## 13-3. Negative feedback for distortion reduction

A number of sources in an FM tuner can contribute various forms of distortion. Narrow-band IF amplifiers and detector, nonlinearities in the detector response, and modulation peaks in the transmitted program can all contribute to harmonic and intermodulation distortion. This is particularly important in tuners arranged for multiplex decoding, since much higher modulation frequencies must be handled with minimum distortion. For example, Fig. 8-26 (Chapter 8) shows severe audio clipping due to either overmodulation at the transmitter or insufficient detector linear bandwidth. A method used to reduce distortion due to overmodulation and the like is presently incorporated in a number of commercial FM designs and is described below.

In the section on limiters, it was pointed out that the frequency deviations of the IF should remain within the portion of the IF passband that is subject to limiting. If this were not the case, AM would be introduced as a result of the IF selectivity. If the IF passband is marginal, then large carrier swings will result in a noisy output from the detector, since the limiters will not be saturated on weak RF signals. The limiters will simply flatten the very top of the selectivity curve.

If the linear working range of the detector is small, then large carrier excursions will result in clipping due to operation beyond the linear region of the "S" curve.

Thus, if some means could be devised to reduce the effective carrier deviation at the *receiver*, these problems could be minimized and distortion would be reduced. A reduction in deviation results in a reduction in the audio output voltage, but a slight adjustment of the volume control on the main amplifier will compensate for this loss.

The basic method used is shown in Fig. 13-4 and consists of a reactance modulator ($V_1$), an emitter follower ($V_2$), a ganged switch, and the associated components. The modulator is placed across the local oscillator and translates changes in voltage to changes in the frequency of the local oscillator. To reduce the deviation of the IF signal, the signal frequency modulates the local oscillator *in phase* with the deviations of the transmitted carrier. Since the local oscillator frequency no longer remains constant (it is constant for normal oscillator operation), and since the IF is equal to the instantaneous difference between the RF carrier and the local oscillator frequency, the IF deviations will be reduced in proportion to the amount of frequency modulation of the oscillator. Figure 13-5 will help to clarify this discussion. Part (a) shows what happens in a tuner where the local oscillator frequency is fixed (for a given station) and a station carrier of 100 MHz is selected. Here, the local oscillator is set to 110.7 MHz, and as the RF carrier deviates $\pm 75$ kHz on either side of the 100 MHz position, the peak-to-peak IF devia-

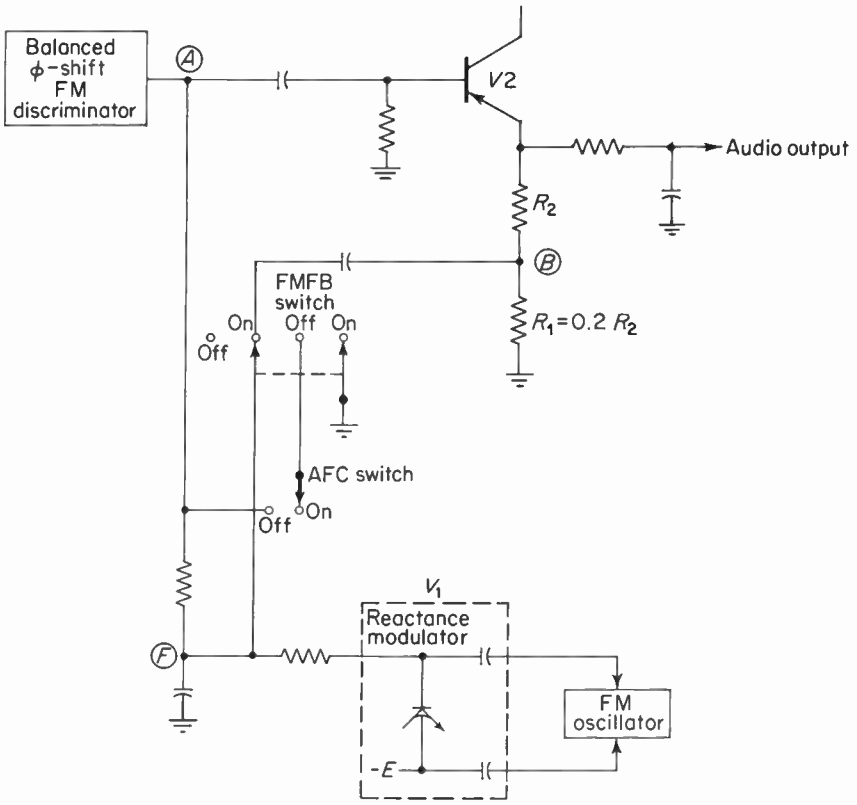



**Fig. 13-4.** A partial schematic showing how the FMFB voltage can be obtained.

tion is 150 kHz. Part (b) shows how the IF peak-to-peak deviation is reduced to 50 kHz by deviating the local oscillator $\pm 50$ kHz on either side of its 110.7 MHz position. The oscillator is made to deviate by a fixed amount when the feedback loop is closed. This is now equivalent to a reduction in carrier deviation at the transmitter.

The FM feedback voltage (FMFB) required to frequency-modulate the oscillator (via the modulator $V_1$) is an AC audio voltage and is derived from a resistive voltage divider ($R_1$-$R_2$). It is approximately 17 per cent of the total available audio output. That portion across $R_1$ is coupled to point $F$ in the schematic and appears at the control point of the reactance modulator to vary its capacitance and thus effect changes in the frequency of the local oscillator. The emitter follower minimizes the loading effect that the reactance modulator would introduce if it were shunted directly across the discriminator output. The AFC error voltage is also applied to the reactance modulator for correcting any drift of the local oscillator. The control voltage

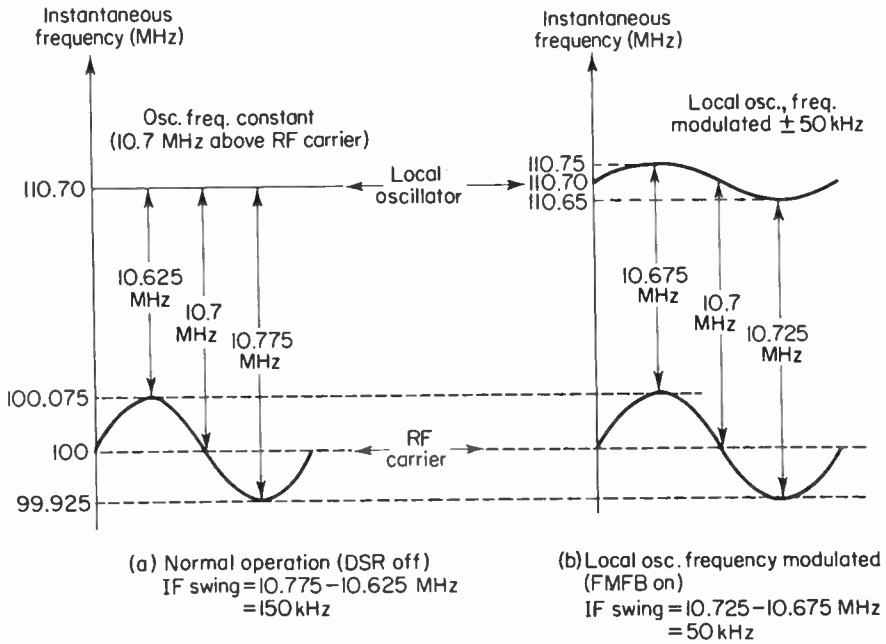



**Fig. 13-5.** The effect on the peak-to-peak IF deviation (frequency swing) as a result of frequency modulating the local oscillator. The conditions illustrated above result in a reduction of the effective modulation index and thus an equivalent reduction in the number of sidebands. The local oscillator is heterodyned on the high side of the RF carrier (standard practice for all FM tuners).

of the reactance modulator thus consists of an AC component superimposed on a DC component (the AFC error voltage), with the AC component frequency-modulating the oscillator at an audio rate while the DC component is correcting for drift or mistuning.

The ganged ON-OFF switch must be provided for two reasons. First, since the AC feedback-voltage variations must be in phase with the carrier deviations, detuning of the oscillator must be avoided to minimize possible regeneration and a spurious output. Thus, AFC must always be ON when FMFB is ON. Second, in order not to upset certain frequency-response correction networks in the tuner, the deemphasis time constant must be modified when the FMFB is ON. This is performed by capacitor $C_1$, which is switched in when FMFB is ON.

Since this arrangement is a negative-feedback system, the magnitude of FMFB used to modulate the oscillator must be chosen carefully. Too little voltage will not accomplish the necessary reduction in IF deviation. Too much voltage may result in instability and severe reduction in the audio output voltage.

## 13-4. Cathode followers

FM tuners are a class of receivers that require separate audio amplifiers. Thus, the tuner must be connected to the audio amplifier or tape recorder by means of coaxial cable, shielded to prevent hum pickup. The shielding introduces into the circuit a great deal of distributed capacity (20 to 100 pF per foot of cable), which will shunt the high-impedance input of the amplifier. As has already been pointed out, both the Foster-Seeley and the ratio detector have fairly high source impedances (tens to hundreds of thousands of ohms). Thus, assuming a direct connection from detector to audio preamp, the cable with its large distributed capacitance and the detector source impedance will act as an AC voltage divider. See Fig. 13-6.

At low frequencies, the distributed capacitive reactance ($X_c$) will be large, and little or no voltage division will take place between the source resistance and the distributed capacitance. The voltage across the distributed capacitance, which is delivered to the audio preamp, will be of approximately the same magnitude as that developed by the detector. At high audio frequencies $X_c$ will decrease. Therefore, the voltage across $Z_{source}$ (detector's equivalent internal impedance) will increase, and less voltage will be available for use at the grid of the preamp.

Because direct coupling via a coaxial cable between the receiver's detector and audio amplifier will result in high-frequency distortion, some means must be found that will permit the use of long connecting cables and at the same time minimize the loss of high audio frequencies. The best approach would be to reduce the source impedance of the detector to zero. Thus, as the frequency is changed, $X_c$, the reactance of the distributed capacity, will vary, but the voltage across it will remain constant since no voltage division can take place. Theoretically, the frequency response under these conditions would be perfect. Of course, it is impossible to reduce the source impedance to zero, but its magnitude can be reduced from thousands to hundreds of ohms by means of a vacuum-tube impedance transformer called a cathode follower. Such a reduction in the effective source impedance (as "seen" by the shunt distributed capacity) will result in improved frequency response.

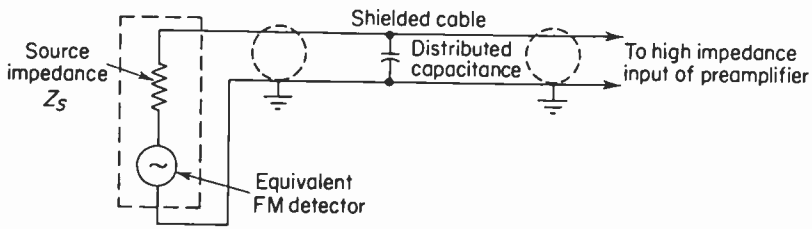



**Fig. 13-6.** The internal impedance of the FM detector and the distributed capacitance of the shielded cable act as an ac voltage divider.

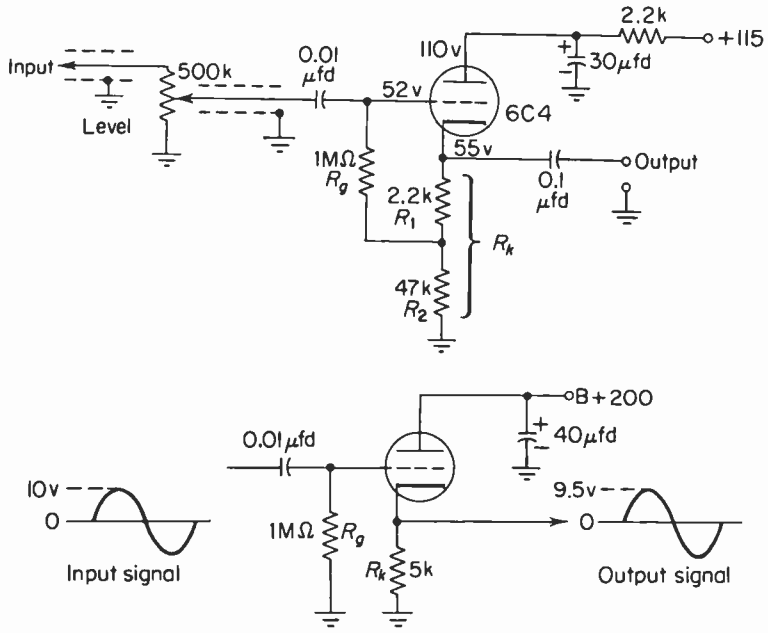



**Fig. 13-7.** (a) Basic cathode follower circuit illustrating that the voltage amplification is less than one and that the output signal voltage is in phase with the input signal voltage. (b) The circuit of a typical commercial cathode follower.

Figure 13-7 illustrates two typical types of cathode-follower arrangements. In both cases the plate returns directly to $B^+$, and since $B^+$ is returned to ground for AC via the power-supply output filter capacitor, the plates are therefore considered to be at AC ground. Thus, the circuit is often called a grounded-plate amplifier. The cathode is returned to ground through the load ($R_k$), across which the bias and the AC output voltage are developed. The voltage (DC) developed from cathode to ground may be quite large (20 to 80 volts), the exact value depending upon the tube, $R_k$, and $E_{bb}$. If the AC output voltage is measured with respect to ground, it will be found to be in phase with the input voltage. Thus in this amplifier, unlike the grounded-cathode amplifier, there is no polarity inversion between input and output voltages. If the cathode is used as reference for both the input signal and the output voltage, it is found that they are series-opposing and that the AC voltage developed between grid and cathode is their difference. See Fig. 13-8. This represents a unique case of 100 per cent negative feedback: all of the *output voltage* is returned to the input. Therefore, this amplifier is considered to be operating under conditions of negative voltage feedback.

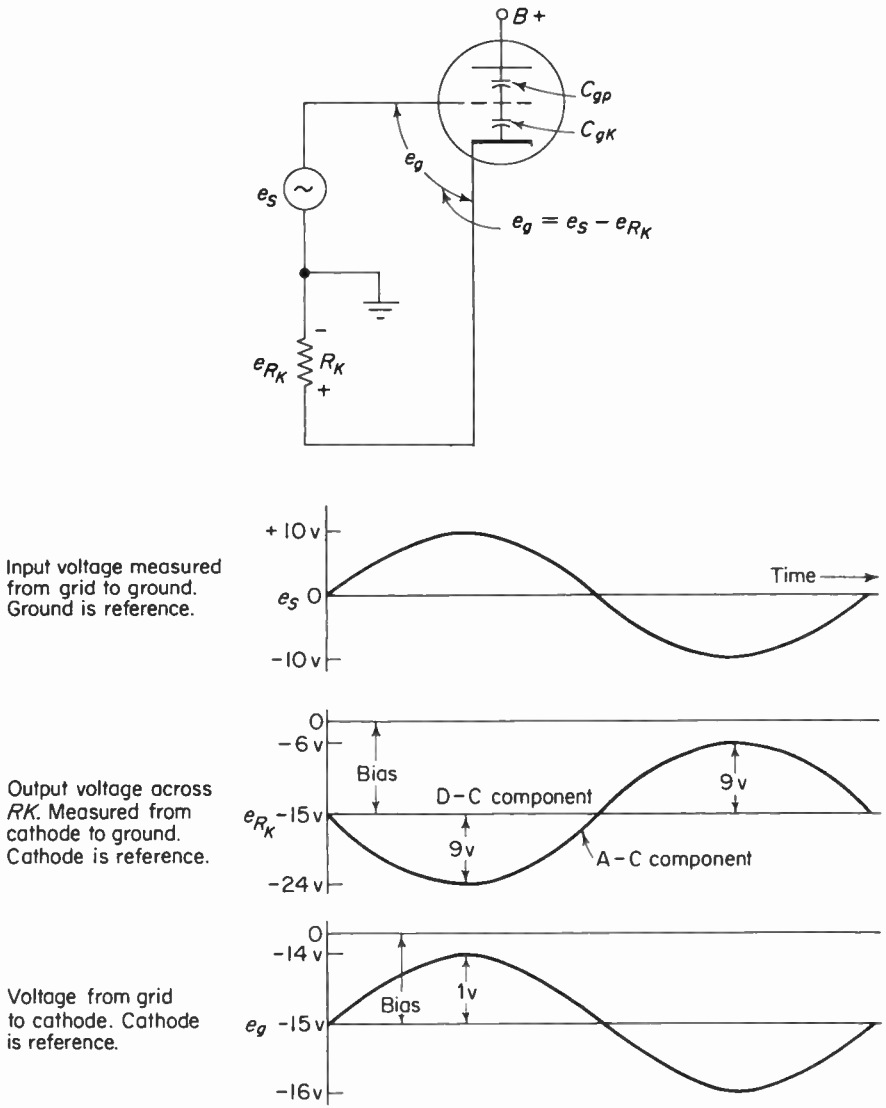



**Fig. 13-8.** Illustrates the phase and amplitude relationships that exist between the various tube elements. Note that overall voltage amplification may be found by dividing $e_{RK}$ by $e_s$ and is .9.

With this enormous degree of feedback, the voltage amplification is always less than one and is equal to

$$A_v = \frac{\mu R_k}{r_p + R_k(\mu + 1)}. \tag{13-1}$$

In order to express the $A_v$ (voltage amplification) of the stage in the same terms used for the grounded-cathode amplifier,

$$A_v = \frac{\mu R_L}{r_p + R_L}, \tag{13-2}$$

the values of $r_p$ and $\mu$ are modified. Thus

$$A_v = \frac{\mu R_k}{r_p + R_k(\mu + 1)} \tag{13-1}$$

changes to

$$A_v = \frac{\left(\frac{\mu}{\mu + 1}\right) R_k}{\left(\frac{r_p}{\mu + 1}\right) + R_k}. \tag{13-3}$$

Note that $R_k$ in both the numerator and denominator of Eq. (13–3) are unencumbered by additional terms. It then appears that as a result of negative voltage feedback, $r_p$ and $\mu$ have been reduced by a factor of $1/(\mu + 1)$.

### *Output impedance*

By definition the output impedance of an amplifier is the internal impedance of a Thevenin equivalent circuit of the amplifier. The equivalent circuit of a cathode follower is shown in Fig. 13-9(a) and is developed from the equation for voltage amplification. That is,

$$\frac{e_o}{e_s} = A_v = \frac{\left(\frac{\mu}{\mu + 1}\right) R_k}{\left(\frac{r_p}{\mu + 1}\right) + R_k}; \tag{13-3}$$

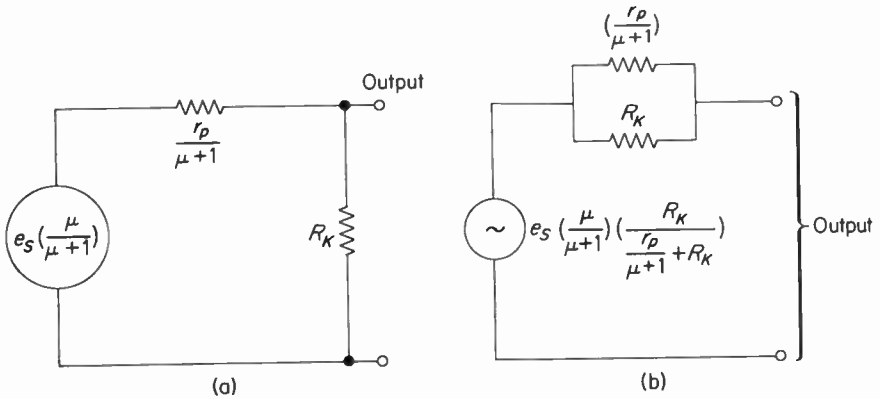



**Fig. 13-9.** (a) Is the equivalent circuit of a cathode follower based on Eq. 13–3. (b) Represents the Thevenin equivalent of (a).

therefore

$$e_o = e_s\left(\frac{\mu}{\mu + 1}\right)\left(\frac{R_k}{\frac{r_p}{\mu + 1} + R_k}\right). \tag{13-4}$$

Note that the final expression for finding the output voltage is that of a generator in series with a two-element voltage divider. The equivalent generator is, as indicated above, $e_s[\mu/(\mu + 1)]$, whereas the series voltage divider is made up of $R_k$ and $r_p/(\mu + 1)$. The Thevenin equivalent of this circuit is shown in Fig. 13-9(b). As can be seen, the output impedance is thus the parallel combination of $R_k$ and $r_p/(\mu + 1)$. If $R_k$ (the parallel combination of all circuit elements, AC or DC, connected from cathode to ground) is much smaller than $r_p/(\mu + 1)$, it then becomes evident that the output impedance is approximately equal to $R_k$. For most conditions of operation the output impedance will be somewhat smaller than $R_k$. The equation for the output impedance ($Z_o$) is therefore (assuming a resistive $R_k$)

$$Z_o = \frac{Z_2 Z_3}{Z_2 + Z_3} \quad \text{or} \quad \frac{\frac{r_p}{\mu + 1} R_k}{\frac{r_p}{\mu + 1} + R_k}$$

which is

$$\cong \frac{\frac{1}{g_m} R_k}{\frac{1}{g_m} + R_k}. \tag{13-5}$$

A given tube may not always be able to provide a desired output impedance. This is because the largest value of $Z_o$ obtainable from a given tube occurs when $R_k$ is infinite, since $R_k$ is in parallel with $r_p/(\mu + 1)$. If the desired value of $Z_o$ exceeds this limiting value, then the required value of $R_k$ will be negative and therefore impossible to achieve. The expression for $R_k$ in terms of $Z_o$ makes this conclusion apparent:

$$R_k \cong \frac{Z_o}{1 - Z_o g_m}. \tag{13-6}$$

It can be seen that if $Z_o g_m$ is greater than 1, the value of $R_k$ will be negative.

### *Input impedance*

The input impedance, as "seen" by the input generator, is considered to be made up of a resistance and a capacitance in parallel. The interesting thing about the input impedance of a cathode follower is that it can be increased to many times the input impedance of the grounded-cathode amplifier, assuming that the same tube is used in both cases. This can be attributed to feedback from cathode to grid through $C_{gk}$, the grid-to-cathode inter-

electrode capacitance. This effect can be thought of as a sort of reverse Miller effect.† The nature of the input impedance depends upon the type of circuit employed. For example, if the circuit of Fig. 13-7(a) is employed, only the shunt capacitive component of the input impedance is reduced ($X_c$ increased). The resistive term can never be larger than $R_g$ since it is in parallel with it. Figure 13-7(b), on the other hand, can reduce the effective $C_{in}$, and at the same time increase the value of $R_{in}$ to many times $R_g$. We can understand this by recalling that the input and output voltages are of opposite polarity when measured between grid and cathode. See Fig. 13-8. Thus the AC current resulting from these two voltages, and flowing through $C_{gk}$, is reduced, as compared to the grounded-cathode amplifier, because the magnitude of the voltage across $C_{gk}$ is less than the applied input voltage. Since the current drawn from the input generator is less, the effective reactance of $C_{gk}$ and the effective resistance of $R_g$ appear to have increased. Thus, for any given frequency, $C_{gk}$ is effectively smaller and $R_g$ is *effectively* larger. The magnitude of the total input capacitance ($C_{in}$) is equal to

$$C_{in} = C_{gp} + C_{gk}(1 - A_v) \tag{13-7}$$

whereas the effective value of the input resistance is

$$R_{in} = \frac{R_g}{1 - A_v}. \tag{13-8}$$

It can be seen that if the voltage amplification ($A_v$) were made equal to 1, $C_{in}$ would equal $C_{gp}$, and $R_{in}$ would become infinite. Thus the closer $A_v$ is to unity, the smaller the degree of loading introduced by the cathode follower upon the input generator.

It is apparent from Fig. 13-7(a) that $R_k$ must satisfy both the requirement for output impedance and the requirement for bias. This is sometimes difficult to arrange. The circuit in Fig. 13-7(b) provides the proper value of $R_k$ for the desired values of $Z_o$ and $Z_{in}$, and at the same time $R_g$ taps $R_k$(which is equal to $R_1 + R_2$) for the proper bias for tube operation.

The equations given in this text are derived for the circuit of Fig. 13-7(a), but if $R_1$ of Fig. 13-7(b) is small compared to $R_2$, these equations also apply.

## 13-5. The emitter follower

The characteristics of the vacuum-tube cathode follower may be approximated by the transistor amplifier when it is arranged in a common-collector—or, as it is often called, the emitter-follower—circuit configuration. A typical circuit is shown in Fig. 13-10. The emitter follower is characterized by a relatively high input impedance, a voltage amplification approaching unity, no phase reversal between input and output signal voltages, and

†See Chapter 4.

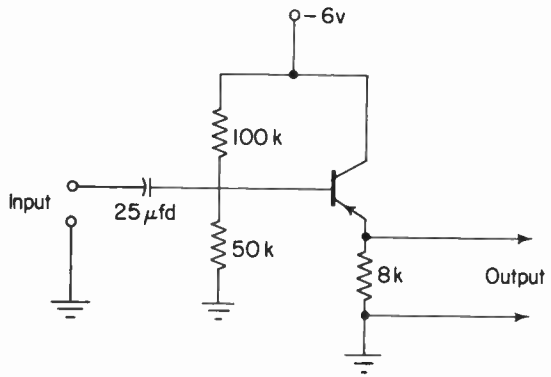



**Fig. 13-10.** A common collector amplifier.

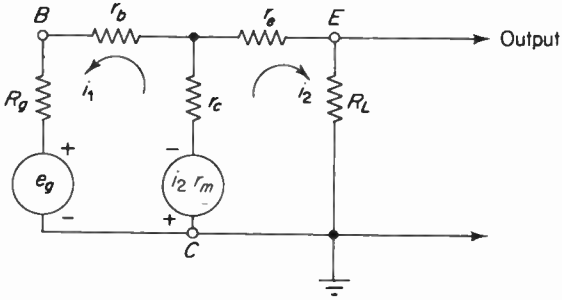



**Fig. 13-11.** The *T* equivalent circuit of an emitter follower.

a low output impedance. In general the emitter follower is used in the same applications as the vacuum-tube cathode follower.

The emitter follower may conveniently be analyzed by means of a T-model equivalent circuit such as that of Fig. 13-11. In the equivalent circuit $r_b$, $r_e$, $r_c$, and $r_m$ are the T parameters of the transistor equivalent circuit. The resistance $R_g$ is the Thevenin equivalent resistance of the input circuit, and $R_L$ is the load resistance. The usual method of analysis is to write two loop equations for the equivalent circuit of Fig. 13-11 and then solve for the desired quantities by means of simultaneous equations or by determinants. The complete analysis will be found in Appendix 13-1. A summary of the results of this analysis is given below. The summary includes the approximate formulas for each exact expression as well as typical values.

Current gain:

$$A_i = -\frac{r_c}{R_L + r_e + r_c - r_m} \approx \frac{-1}{1-\alpha} \approx -50. \tag{13-9}$$

Voltage amplification:

$$A_v = \frac{r_c R_L}{(R_g + r_b + r_c)(R_L + r_e + r_c - r_m) - r_c(r_c - r_m)} \approx 1. \tag{13-10}$$

Input resistance:

$$R_{in} = r_c + r_c - \frac{r_c(r_c - r_m)}{R_L + r_e + r_c - r_m} \approx \frac{R_L}{1 - \alpha} \approx 50 \text{ kilohms.} \tag{13-11}$$

Output resistance:

$$R_o = r_e + r_c - r_m - \frac{r_c(r_c - r_m)}{R_g + r_b + r_c} \approx r_e + (1 + \alpha)(R_g + r_b) \approx 1 \text{ kilohm.} \tag{13-12}$$

Power gain:

$$PG = \frac{4R_g R_L r_c^2}{[(R_g + r_b + r_c)(R_L + r_e + r_c - r_m) - r_c(r_c - r_m)]^2} \approx \frac{1}{1 - \alpha} \approx 15 \text{ db.} \tag{13-13}$$

## 13-6. Summary

**1.** Squelch or muting circuits are sometimes provided to suppress the hiss normally heard when tuning between stations.

**2.** Muting circuits often take the form in which a stage in the signal path is cut off by the absence of limiter bias or some other similar arrangement.

**3.** Muting circuits should act instantaneously, so that in tuning from a strong to a weak station the "squelch-activating" voltage will not suppress the weak station.

**4.** Tuner distortion due to overmodulation or slight mistuning can be reduced by means of frequency modulation of the local oscillator to reduce the IF deviation. This is equivalent to a reduction of the modulation index at the transmitter.

**5.** Long coaxial cables from the FM tuner to the audio amplifier may degrade high-frequency response.

**6.** A reduction of the detector source impedance minimizes the problem of coaxial-cable capacitance.

**7.** The cathode or emitter follower is an impedance transformer that can change a high source impedance to a low source impedance.

**8.** The voltage amplification of a cathode or emitter follower is less than one.

**9.** Negative voltage feedback causes the plate resistance of the tube to be effectively reduced.

**10.** Input and output voltages, measured with respect to ground, are of the same polarity.

**11.** The magnitude of $Z_o$ is obtainable by calculating $R_k$ in parallel with $r_p/(\mu + 1)$.

**12.** The largest value of $Z_o$ obtainable from a tube is $r_p/(\mu + 1)$.

**13.** Attempts to increase $Z_o$ beyond the tube's capability will necessitate a negative value for $R_k$. This is an impossible demand.

**14.** The transistor equivalent of the cathode follower is the emitter follower.

## REFERENCES

1. Chaffee, J. G.: "The Application of Negative Feedback to Frequency-Modulation Systems," *Proc. IRE*, May 1939.
2. Sturley, K. R.: "*Radio Receiver Design*," Vol. 2. John Wiley & Sons, Inc., New York, 1953.
3. Rifkin, M. S.: "A Graphical Analysis of the Cathode-Coupled Amplifier," *Communications*, December 1946.
4. Fitchen, F. C.: Transistor Circuit Analysis and Design, D. Van Nostrand Co., Inc., New York, 1960.

## APPENDIX 13-1

Using the equivalent circuit of Fig. 13-11, we may obtain the following loop equations:

$$i_1(R_g + r_B + r_C) + i_2(r_C - r_m) = e_g, \tag{1}$$

$$i_1(r_C) + i_2(R_L + r_e + r_C - r_m) = 0. \tag{2}$$

We may solve for the currents $i_1$ and $i_2$ by means of determinants as follows:

$$i_1 = \frac{\begin{vmatrix} e_g & r_C - r_m \\ 0 & R_L + r_E + r_C - r_m \end{vmatrix}}{\begin{vmatrix} (R_g + r_B + r_C) & (r_C - r_m) \\ r_C & R_L + r_E + r_C - r_m \end{vmatrix}}$$

$$= \frac{e_g(R_L + r_E + r_C - r_m)}{(R_g + r_B + r_C)(R_L + r_E + r_C - r_m) - r_C(r_C - r_m)},$$

$$i_2 = \frac{\begin{vmatrix} (R_g + r_B + r_C) & e_g \\ r_C & 0 \end{vmatrix}}{D} = \frac{-e_g r_C}{D}.$$

By definition the output current divided by the input current is the current gain of the amplifier.

*Current gain:*

$$A_i = \frac{i_2}{i_1} = \frac{-r_c}{R_L + r_E + r_C - r_m}.$$

The input resistance between the base and the collector terminals of the amplifier must be the total resistance "seen" by the input generator ($e_g/i_1$) less the resistance associated with the generator.

*Input resistance:*

$$R_i = \frac{e_g}{i_1} - R_g = r_B + r_C - \frac{r_C(r_C - r_m)}{R_L + r_E + r_C - r_m}.$$

The voltage amplification by definition is the output voltage ($i_2R_L$) divided by the input voltage ($e_g$).

*Voltage amplification:*

$$A_v = \frac{i_2R_L}{e_g} = \frac{r_CR_L}{(R_g + r_B + r_C)(R_L + r_E + r_C - r_m) - r_C(r_C - r_m)}.$$

If the input resistance ($R_i$) is equal to the generator resistance ($R_g$), maximum power is transferred. The input power under these conditions is

$$P_i = \frac{\left(e_g \dfrac{R_i}{R_i + R_g}\right)^2}{R_i}$$

but $R_i = R_g$, therefore

$$P_i = \frac{\dfrac{e_g^2R_g^2}{4R_g^2}}{R_g} = \frac{e_g^2}{4R_g}$$

The power gain of an amplifier is defined as the power output divided by the input power. Thus

$$PG = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{i_2^2R_L}{\dfrac{e_g^2}{4R_g}} = \frac{4R_gR_Li_2^2}{e_g^2},$$

which may be expanded to

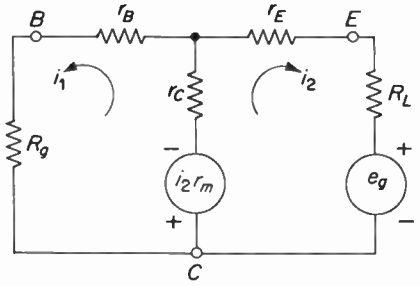
$$PG = \frac{4R_gR_Lr_c^2}{[(R_g + r_b + r_c)(R_L + r_e + r_c - r_m) - r_c(r_c - r_m)]^2}.$$

We may find the output resistance of an amplifier by shorting the input generator and then determining the resistance "seen" by a generator placed in series with the load.

*Output resistance:*

$$i_1(R_g + r_b + r_C) + i_2(r_C - r_m) = 0, \tag{1}$$

$$i_1(r_C) + i_2(R_L + r_E + r_C - r_m) = e_g, \tag{2}$$




$$i_2 = \frac{\begin{vmatrix} R_g + r_B + r_C & 0 \\ r_C & e_g \end{vmatrix}}{\begin{vmatrix} R_g + r_B + r_C & (r_C - r_m) \\ r_C & R_L + r_E + r_C - r_m \end{vmatrix}}$$

$$= \frac{e_g(R_g + r_B + r_C)}{(R_g + r_B + r_C)(R_L + r_E + r_C - r_m) - r_C(r_C - r_m)},$$

$$R_{\text{out}} = \frac{e_g}{i_2} - R_L$$

$$= r_E + r_C - r_m - \frac{r_C(r_C - r_m)}{R_g + r_B + r_C}.$$

# 14

# STEREOPHONIC BROADCASTING

## 14-1. Introduction

The Federal Communications Commission issued the transmission standards for FM Multiplex broadcasting on April 19, 1961, and FM stations began transmitting on June 1 of that year. A brief outline of the system will be given first, and a more detailed discussion will follow.

Multiplexing is a technique whereby two (or more) separate programs are accommodated by one carrier. This single carrier is then received by an FM (or AM) tuner with provision for decoding or separating the programs and feeding them to separate amplifier and speaker systems. These two "programs" could be separate microphones, each referred to as a channel. Let us refer to the left microphone as $L$ and the right microphone as $R$. The standards specify that the sum of the left and right channels $(L + R)$ shall frequency-modulate the station carrier up to a maximum of 90 per cent of full system deviation or 67.5 kHz. This "sum" signal is picked up by the conventional monophonic FM tuner and demodulated in the normal manner with practically no degradation of quality. The difference between the two channels $(L - R)$ is used to amplitude-modulate another ultrasonic carrier (in this case, it is referred to as a subcarrier). The subcarrier is suppressed, and the resultant AM sidebands frequency-modulate the main carrier up to a maximum of 90 per cent of full system deviation.† In order to permit

†The maximum deviation of the main carrier by $(L + R)$ and by $(L - R)$ is 90 per cent. However, the 90 per cent peak deviation is never caused by $(L + R)$ or $(L - R)$ at the same time. The sum of the two signals will never be more than 90 per cent.

demodulation of the AM sidebands at the receiver, a pilot carrier is transmitted, since the AM subcarrier is suppressed.

The FM tuner with the proper decoding equipment takes these two signals ($L + R$ and $L - R$) and appropriately mixes them to obtain two separate stereo signals as follows: If $(L + R)$ is added to $(L - R)$, then $(L + R) + (L - R) = 2L$. If $(L + R)$ is subtracted from $(L - R)$, then $(L + R) - (L - R) = 2R$. Disregarding the coefficient 2 (it indicates relative volume), then the $L$ signal would drive the left loudspeaker while the $R$ signal would drive the right loudspeaker. In any event, $(L + R)$ would be used by the conventional monophonic receiver for full-fidelity reproduction, thus making the system fully compatible with existing equipment.

A block diagram of an FM stereophonic transmitter is given in Fig. 14-1. Both the left ($L$) and right ($R$) channels are fed to 75-$\mu$ sec high-pass filters, where they are preemphasized. The two channels are then fed into a matrix circuit for mixing. The output of the matrix consists of the sum and difference of the two channels. The $(L + R)$ channel is fed to the frequency modulator to frequency-modulate the main carrier. The $(L - R)$ channel is fed to a balanced modulator. This modulator also receives the 38-kHz subcarrier, which is the second harmonic of the 19-kHz pilot carrier. The $(L - R)$ signal now amplitude-modulates the 38-kHz subcarrier, and the balanced modulator suppresses the 38-kHz subcarrier. The output of the

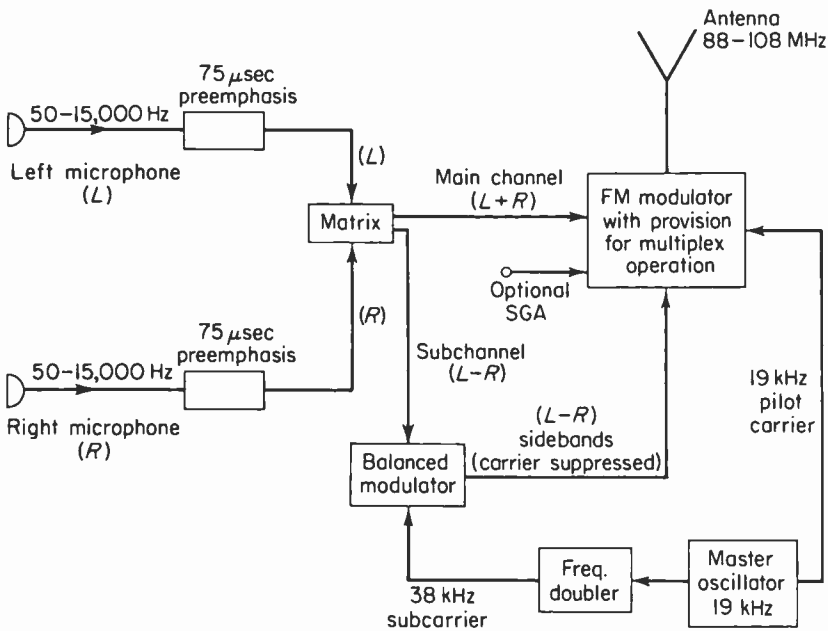



**Fig. 14-1.** Conceptual block diagram of stereophonic transmitter for FM multiplexing.

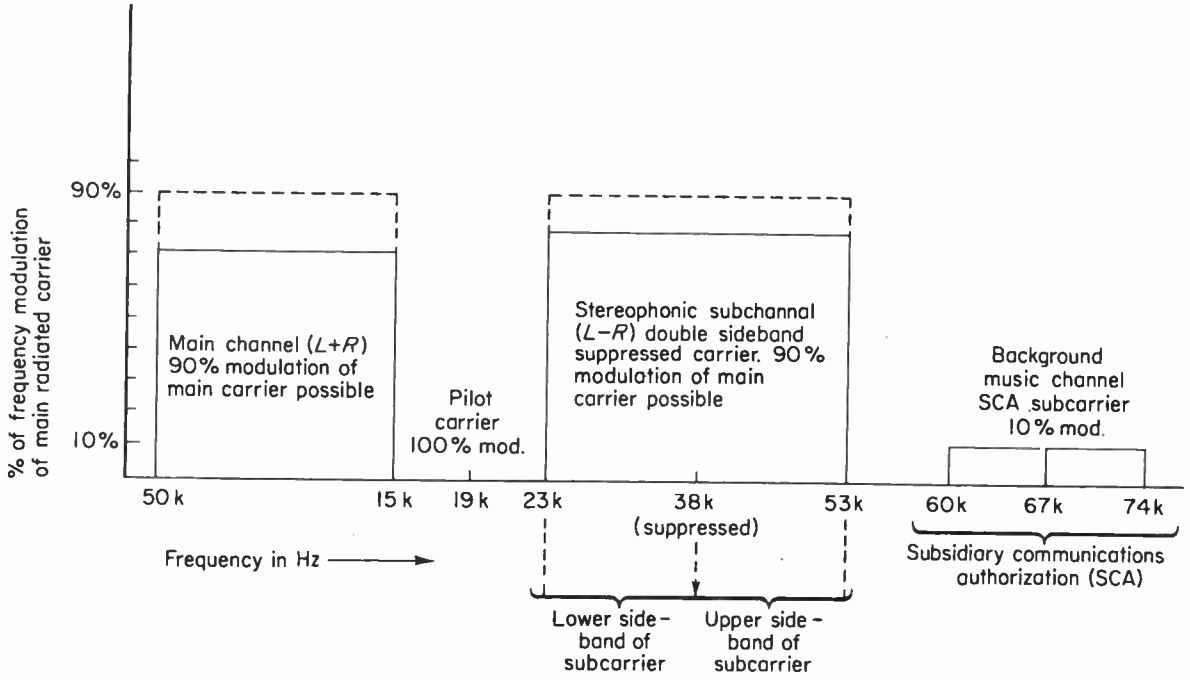



**Fig. 14-2.** Baseband spectrum at output of FM demodulator (before deemphasis). The main and subchannel modulation percentages must each be reduced 10% if an SCA subcarrier is simultaneously employed.

balanced modulator consists only of the $(L - R)$ sidebands, which are fed to the FM modulator. These sidebands in turn frequency-modulate the main carrier along with the sum signal. The 19-kHz master oscillator provides the 19-kHz pilot carrier, which is fed to the main FM modulator for frequency-modulating the main carrier. The standards limit pilot-carrier modulation of the main carrier to approximately 10 per cent of full system deviation or 7500 hz. Thus, the signal at the transmitter antenna is a station carrier (88 to 108 MHz) which contains all the stereo information: the $(L + R)$ signal, the $(L - R)$ signal, and a 19-kHz pilot carrier (necessary for demodulation of the difference signal at the receiver).

An FM station may also transmit background music for restaurants and the like simultaneously with multiplex stereo. These broadcasts will use a higher subcarrier (67 kHz), which wil lfrequency-modulate the main carrier up to a maximum of approximately 10 per cent of full system deviation. Figure 14-2 shows the base-band spectrum at the output of the FM detector (before deemphasis). Notice that the 19-kHz pilot carrier appears in a clear channel at reduced amplitude 4 kHz above the $(L + R)$ audio and 4 kHz below the $(L - R)$ lower sideband. This allows easier filtering of the pilot carrier and its subsequent use in demodulating the difference signal.

## 14-2. Interleaving or nesting

Interleaving of the sum and difference signals in multiplexing is one of the most interesting aspects of this system and can be best explained by the use of some simple waveforms. As will be shown, the use of interleaving will permit 90 per cent of maximum deviation by the main channel as well as 90 per cent of maximum deviation by the subchannel, the other 10 per cent in each case remaining for the 19-kHz pilot carrier. As a result, when one channel is producing peak main-carrier deviation while the other channel is producing zero deviation and vice versa, the monophonic listener experiences a signal-to-noise degradation of less than 1 db.

Figure 14-3 should help in clarifying this principle. Assume that the left microphone input is a sinewave and the right input is a series of pulses. The right input pulse is used for illustrative purposes only and will rarely, if ever, be encountered in actual cases. Figure 14-3(a) shows that the peak amplitude of the sinewave can deviate the carrier a maximum of 45 per cent of full system deviation (75 kHz) with no right input (this is in accordance with the FCC standards.) Also, the right input can deviate the carrier 45 per cent of maximum with no left input (this is also in accordance with the FCC standards). Figure 14-3(c) shows that the sum channel $(L + R)$ can modulate the carrier up to 90 per cent of maximum deviation. Thus one component of the sum signal, Fig. 14-3(c), is for modulating the main carrier and the other is the $(L - R)$ sidebands envelope. The difference between the

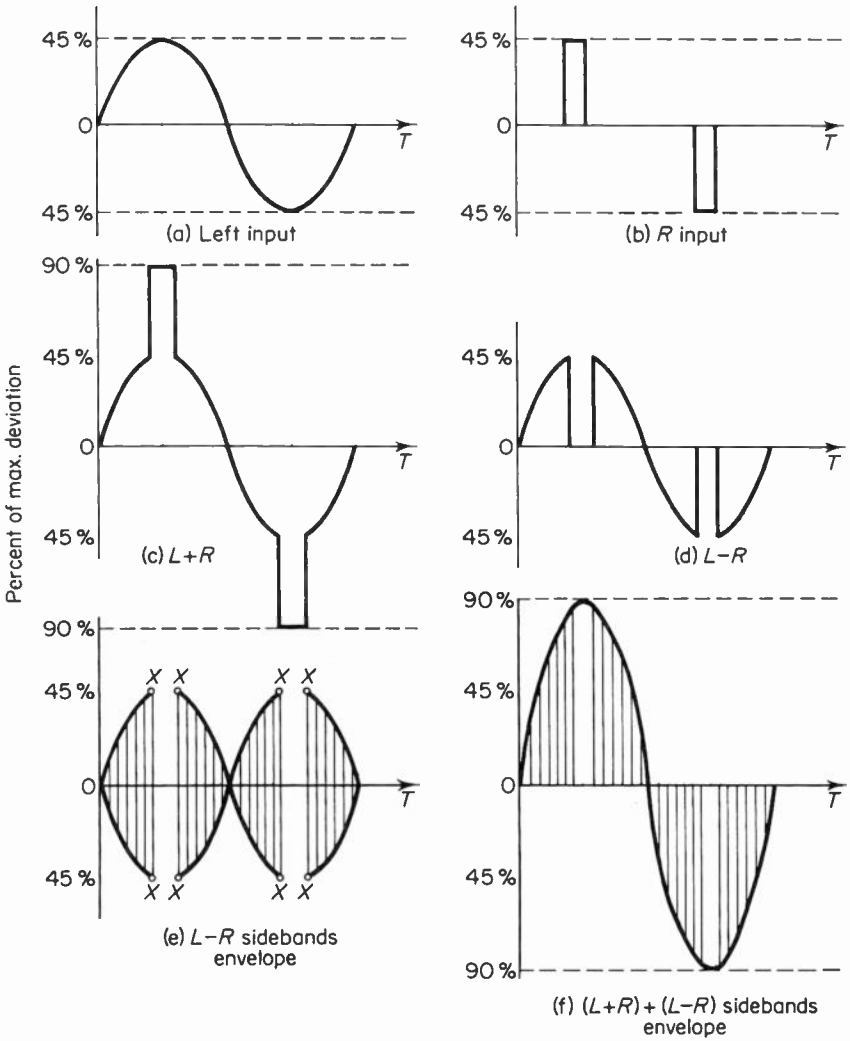



**Fig. 14-3.** Waveforms showing interleaving effect as explained in text. The pilot carrier is omitted for clarity but will deviate the carrier a maximum of 10%.

left and right inputs is shown in Fig. 14-3(d), but this is *not* used to modulate the main carrier. Instead, this difference signal amplitude modulates a 38-kHz subcarrier, and the subcarrier is suppressed in the balanced modulator as previously mentioned. The resultant sidebands envelope is shown in Fig. 14-3(e). The signal is also capable of deviating the carrier 90 per cent of maximum (the "x" points, for example). Now, if the $(L + R)$ [Fig. 14-3(c)] ordinates are added to the $(L - R)$ sidebands ordinates *above* the zero base

line [Fig. 14-3(e)] and then to the $(L - R)$ sidebands ordinates *below* the zero base line, the wave in Fig. 14-3(f) is obtained. This, then, is the composite signal applied to the modulator.

Notice that the peak amplitude of the composite is no greater than the

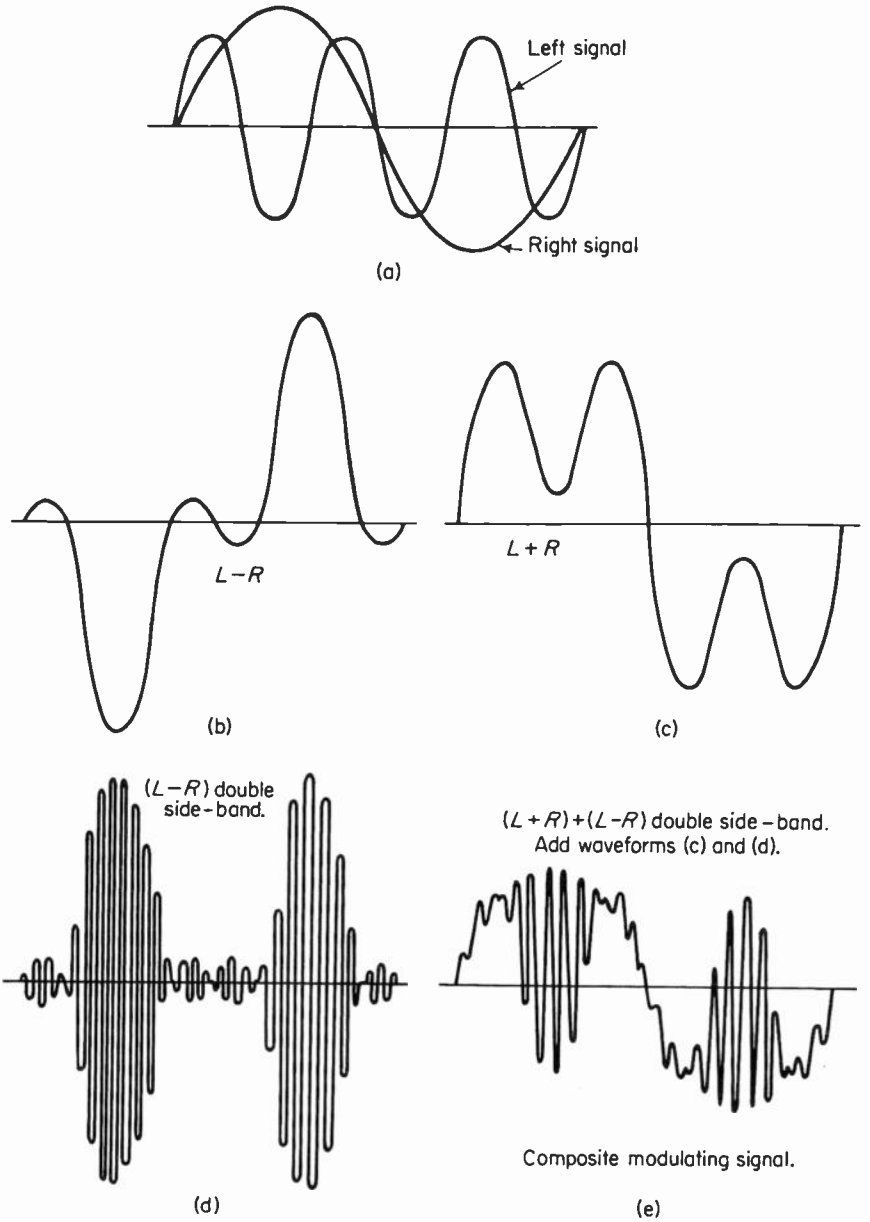



**Fig. 14-4.** Evolution of the composite signal for modulating the main carrier when the left and right signals are sinewaves of different frequencies.

peak amplitude of either the $(L + R)$ or the $(L - R)$ sidebands envelopes. Thus, the carrier deviation cannot exceed 90 per cent of maximum. Also, observe that the $(L - R)$ peak "interleaves" or "nests" in the $(L - R)$ sidebands envelope signal. Therefore, if the subcarrier is suppressed, the main and subchannels can provide maximum main-carrier deviation that is limited only by the requirement for the pilot carrier. Another advantage for suppressing the subcarrier is that the carrier, in "normal" double-sideband broadcast techniques, utilizes approximately two-thirds of the total radiated power (for 100 per cent modulation). When the degree of modulation is less than unity, the sidebands power will be proportional to the square of the modulation factor. As a result, the fraction of the total sidebands power decreases rapidly as the modulation factor is reduced. A large modulation factor would then be desirable. This, however, places severe design restrictions on the AM subchannel detectors in the multiplex receiving equipment. If, however, the carrier is suppressed, we can place more energy in the information-bearing sidebands—but what is just as important, when the subcarrier is regenerated in the receiver, it can be reinserted at a level that permits efficient AM detection. This level is chosen to provide an equivalent modulation factor of approximately 0.3. Distortion can thus be reduced to acceptable levels.

Figure 14-4 shows the various waveforms when the left and right audio signals are sinewaves, the frequency of the left signal being three times that of the right signal. A careful examination of the composite modulating signal [Fig. 14-4(e)] will show that the $(L - R)$ sidebands' positive peaks trace out the left signal waveform while the negative peaks trace out the right signal waveform. This is when both signals are of the same polarity. When the left and right signals are of opposite polarity, there is an abrupt 180° phase reversal of the sidebands and the positive peaks now trace out the right signal waveform while the negative peaks trace out the left signal waveform. This is for the first half-cycle of the right signal. For the second half-cycle, the positive and negative sidebands' peaks trace out the right and left signals, respectively, when the left and right signals are of the same polarity. When the left and right signals are of opposite polarity, the sidebands signal reverses phase and the positive peaks of the sidebands once again trace out the left signal waveform while the negative peaks trace out the right signal waveform. The composite signal of Fig. 14-4(e) is used to frequency-modulate the main carrier along with the 19-kHz pilot carrier plus an optional SCA subcarrier.

## 14-3. Envelope detection

Figure 14-5 shows the more important "blocks" (to the right of the dotted line) of a representative *MX* receiver. The monophonic (single-channel) audio output is available from the regular FM tuner deemphasis network, and this consists of the $(L + R)$ signal. This sum signal is available at the

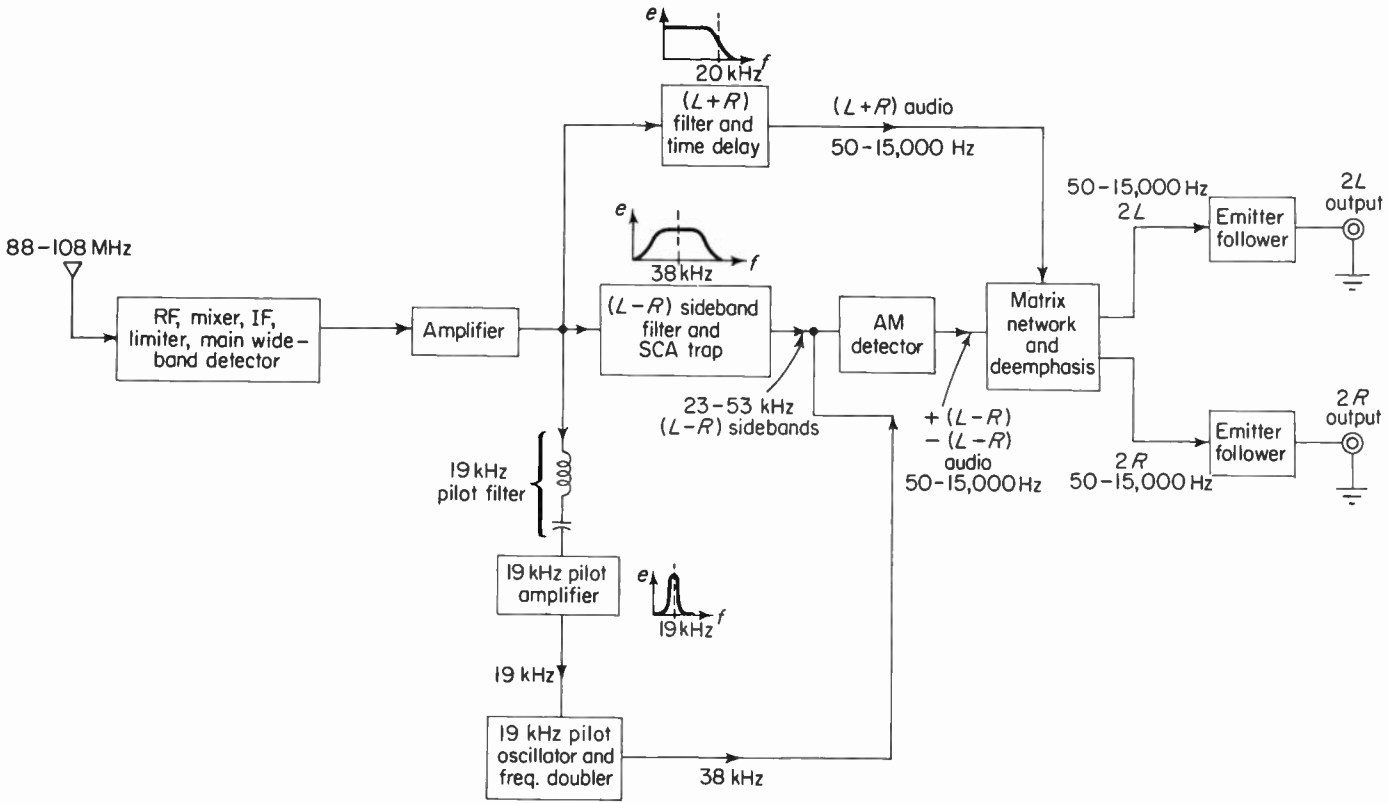



**Fig. 14-5.** Block diagram of a representative FM multiplex (mx) receiver.

output of all monophonic tuners for reasons of compatibility. The composite stereo signal appears at the *MX* jack and consists of the sum signal, the difference-signal sidebands, the 19 kHz pilot carrier, and the 67-kHz (SCA) subcarrier. All this information is further amplified by the first stage in the *MX* section. A 19-kHz pilot-carrier filter selects this pilot frequency, which, in turn, is amplified by the pilot amplifier. The amplified pilot carrier is "injected" into a 19-kHz free-running oscillator loop to lock the oscillator in step with the pilot carrier. The output of this oscillator is tuned to the second harmonic of 19 kHz and this 38-kHz signal corresponds to the original suppressed subcarrier at the transmitter. We shall return to this "recovered" subcarrier shortly.

The composite signal is also fed from the first amplifier into a bandpass filter, which recovers the $(L - R)$ sidebands. These extend from 23 to 53 kHz. This filter also rejects the 67-kHz SCA subcarrier. The $(L - R)$ sidebands are then fed into an AM detector. But in order to demodulate an AM carrier, the carrier component is required as well as the sidebands; this is where the 38-kHz signal comes in. The input to the AM detector thus consists of a 38-kHz "carrier" and the $(L - R)$ difference sidebands. To realize the proper stereo separation, the output of the AM detector must consist of the positive difference audio, $+(L - R)$, and negative difference audio, $-(L - R)$. These two demodulated signals are then fed into a matrix. An $(L + R)$ filter (with cutoff at about 20 kHz) selects the sum signal and, along with the positive and negative signals, we electrically obtain:

$$(L + R) + [+(L - R)] = 2L$$

and

$$(L - R) + [-(L - R)] = 2R.$$

An examination of Fig. 14-6 will help one understand how this matrix network mixes these signals to provide left and right outputs. This matrix is typical; and if the reader understands how this one works, he should have no difficulty with other networks of a similar nature.

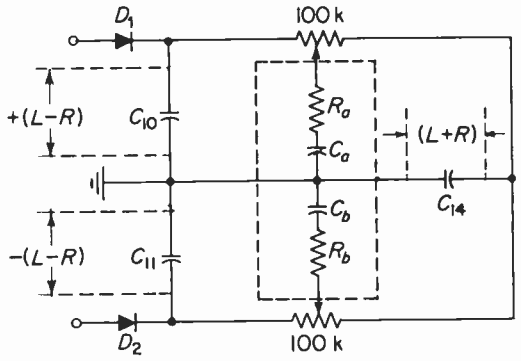



**Fig. 14-6.** Matrix network including deemphasis.

The matrix network is shown as a bridge equivalent (Fig. 14-7) along with the respective signals involved in the mixing process. Assume that the 100-kilohm dimension controls are set at midrange. Thus, $C_a$, $R_a$, and the left half of the upper control (50 kilohms) form a deemphasis network for the $(L - R)$ signal. The $(L + R)$ signal is deemphasized by the upper right half of the dimension control, $C_a$ and $R_a$. The $(L + R)$ signal is also deemphasized by the lower right half of the dimension control, $C_b$ and $R_b$. These deemphasis networks (like all low-pass filters) will attenuate all high-frequency components such as residual 38-kHz subcarrier and its harmonics, 19-kHz pilot components, the SCA subcarrier (67 kHz), and so on. The matrix is put in the form of a bridge network for clarity, and the degree of bridge balance will determine the degree of stereo separation. This is the reason for providing the variable 100-kilohm dimension controls. The currents shown are for a particular instant in time and, for illustrative purposes, we can say that $C_{10}$ "pushes" a current around the upper left-hand loop in the direction shown. This current is the result of the $(L - R)$ signal. Notice that it flows through $C_a$. $C_{14}$ "pushes" a current around the upper right-hand loop. Notice that this current also flows through $C_a$ and in the same direction as the $(L - R)$ current. These currents are thus additive to give a $2L$ output from the "$L$" terminal to ground. Also, $C_{11}$ "pushes" a current around the lower left-hand loop as a result of the $-(L - R)$ signal. This current and the current as a result of the $(L + R)$ signal flow through $C_b$ in opposite directions and are therefore subtractive, giving a $2R$ output

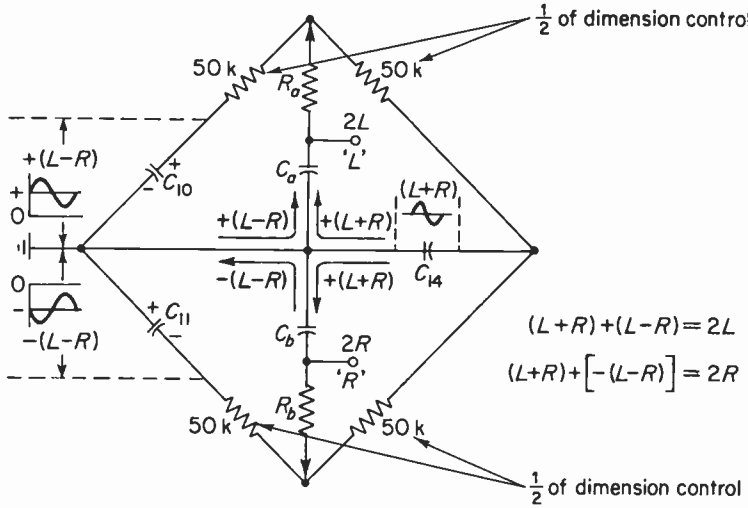



**Fig. 14-7.** Bridge equivalent of Fig. 14-6. The outputs (2L and 2R) across $C_a$ and $C_b$ are determined by the *algebraic* sum of their respective currents. The 100 k "dimension" controls provide for a precise degree of bridge balance.

from terminal "$R$" to ground. If the bridge is perfectly balanced, then the output from terminal "$L$" to terminal "$R$" will be

$$(L - R) + [-(L - R)] = 0 \quad \text{volts.}$$

The dimension controls are provided as a front panel control for varying the degree of separation between channels and are normally set at their center positions. A slight touch-up will result in a more precise and fuller stereo separation at all frequencies. The 50-kilohm separation control at the input to the $(L + R)$ audio filter provides the proper amount of $(L + R)$ signal for matching the $(L - R)$ subchannel and is factory-set.

Networks are also provided, after demodulation, to attenuate any residual 38-kHz signals (and their harmonics) so as not to degrade stereo tape recordings. This is an important function, since the tape head bias frequency can beat against the subcarrier or its harmonic and give rise to annoying beat tones. The above networks also reduce "hiss" in noisy locations by virtue of their low-pass nature.

The two audio signals are then deemphasized and fed to two cathode or emitter followers. The low-output-impedance "followers" permit long cable lengths to the audio amplifiers and thus prevent cable capacitance from degrading high-frequency response.

The block diagram of Fig. 14-5 represents one possible method for decoding stereo signals. It makes use of matrixing (signal separation and algebraic addition) and envelope detection. An alternate method employs a "time division" or sampling technique, which is discussed in the next section.

## 14-4. Time-division multiplex (sampling)

In effect, what the Federal Communications Commission approved was an equation that defines the transmitted signal. The Commission does not establish *how* the composite signal is to be generated and radiated so long as it meets the requirements set forth in the equation. One technique (there are a number of possibilities) had its origin in multiplex carrier telephony and utilizes a high-speed switching and sampling arrangement. Consider Fig. 14-8 as one way to accomplish this job. Two inputs, $L$ and $R$, are shown going to a radio or telephone link. These signals are switched rapidly between the two inputs at a rate *no less* than twice the highest frequency of either signal. The output of the link (the receiver in multiplex reception) is also switched alternately between the two channels at the same rate *and* in synchronism with the input. Both ends of the link must be synchronized in order to prevent $L$ signals from appearing in the $R$ output and $R$ signals from appearing in the $L$ output. The synchronization of the "switches" in a commercial system is accomplished by means of the 19-kHz pilot carrier, which is doubled at the receiver to 38 kHz. Thus, the switching rate is chosen

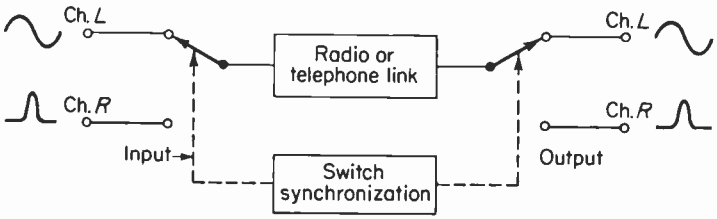



**Fig. 14-8.** Simplified time-division multiplex arrangement.

as 38 kHz, which is at least twice the highest modulation frequency (15 kHz).

Mathematically, the sum signal appears as audio modulation of the main radiated carrier, while the difference signal appears as suppressed-carrier amplitude-modulation of a series of odd harmonics of the switching signal. The switching waveform is assumed to be a square wave. To prevent radiation outside of the 200-kHz channel allocation, all harmonics except the fundamental component of the stereophonic subcarrier are suppressed. Figure 14-9 shows the waveforms that might be used in a time-division multiplex system. The waveforms chosen are only for illustrative purposes.

The $L$ and $R$ inputs are alternately switched into the link at a rate equal

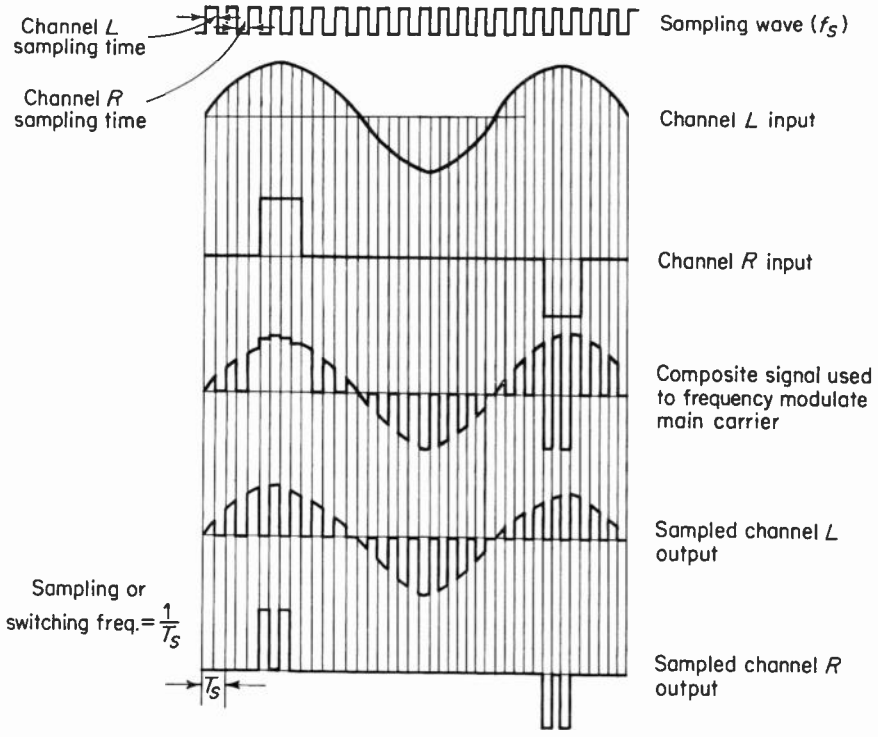
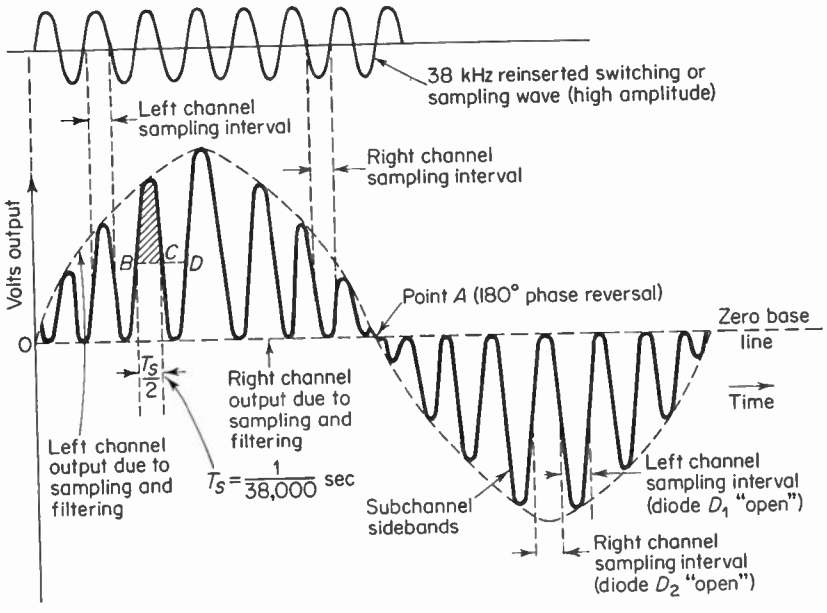



**Fig. 14-9.** Waveforms used to illustrate the principle of time-division multiplex. The sampled outputs are seen to be very good replicas of the inputs.




Cook & Liff 14-10

**Fig. 14-10.** Composite signal input to demodulator. Left channel signal only. See Fig. 14-11. The amplitude of the reinserted 38-kHz voltage must be much larger than the peak resultant of the sidebands.

to $1/T_s$, or at least twice as high as the maximum modulation component. This insures a good sampling of the inputs at the other end of the link. Since the highest modulation frequency is 15 kHz, the switching rate is thus 38 kHz and permits a "half-frequency" pilot (19 kHz) to fall 4 kHz above the highest main-channel frequency and 4 kHz below the lowest subchannel frequency.

One effective way to demodulate the composite signal is to sample it in the receiving equipment by means of a pair of 38-kHz synchronized switches (diodes), which alternately permit "viewing" the composite signal during the correct intervals. Figure 14-10 shows the composite wave to be sampled. Here, we have a left only signal modulating the main carrier. The left signal is the sinewave "envelope," while the zero baseline represents the zero-amplitude right signal. Later, a right signal will be included on the composite wave, but it is simpler to start with one signal only. Notice that alternate peaks of the subchannel sidebands trace out the left and right channels. Thus, the sampling points correspond to the peaks of the subcarrier. At point *A* there is an abrupt 180° phase reversal of the reinserted subcarrier, which corresponds to suppressed carrier AM.

This signal is then applied to the circuit of Fig. 14-11. Diodes D1 and D2 are considered to be switched on and off by the 38-kHz sampling wave. D1 is closed for a $T_s/2$ interval, where $T_s/2$ is the half-period of the sampling

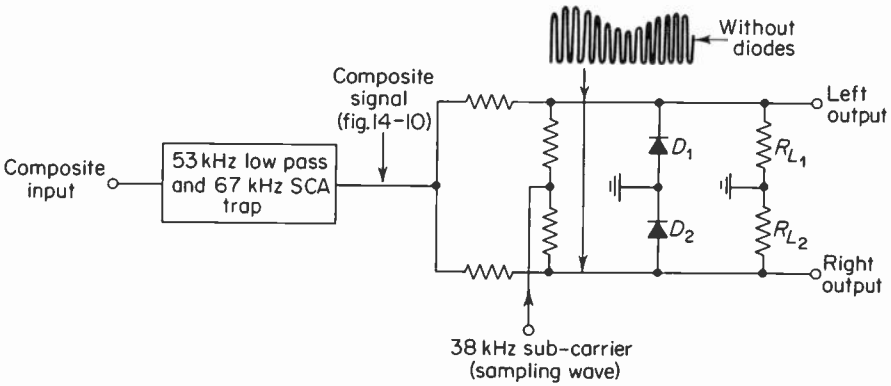



**Fig. 14-11.** Basic circuit showing demodulation by means of time-division (sampling). The diodes are switches which short their respective loads during alternate half-cycles of the 38-kHz switching wave.

wave, or from instant $C$ to instant $D$. Thus, during the time from $C$ to $D$, D1 short-circuits the load $R_{L_1}$ and the load now "sees" zero volts. During this same interval, D2 is reverse-biased by the sampling wave and is therefore open. The $R_{L_2}$ load can thus look at or sample the composite signal, and it also "sees" zero volts, but this is what $R_{L_2}$ *should* see since there is no right signal modulating the carrier. At instant $B$, D1 is reverse-biased, and from $B$ to $C$ it is open. Thus, $R_{L_1}$ can view the composite signal and it "sees" a pulse of voltage corresponding to the crosshatched area. At the same time, D2 is forward-biased and it short-circuits its associated load. Thus, $R_{L_2}$ "sees" zero volts from $B$ to $C$ since this is not the interval for sampling the right channel signal. Summing up, it can be said that a series of pulses appears across $R_{L_1}$ as shown in Fig. 14-15, while zero volts appear across $R_{L_2}$. Subsequent filtering of the waveform will remove the high-frequency components, leaving only the sinewave envelope, which is the channel-$L$ envelope of Fig. 14-10. The reader should recognize the correspondence between the sampled output wave of Fig. 14-12(a) and the waveform of Fig. 14-9 labeled "sampled channel-$L$ output." Since the demodulation is complete, there is no need for *separate* matrixing as already described under "envelope detection." Matrixing and signal separation are already part of the demodulation process in the time-division technique.

Consider the case where two signals will modulate the carrier. The left signal will be a sinewave and the right signal will be a square wave. Since the left and right signals are independent, they may be chosen at random. The sum of the $(L + R)$ signal and $(L - R)$ sidebands is shown as the composite signal at the input to the sampling demodulator. The reinserted 38-kHz subcarrier is also shown; for clarity, the amplitude of the pulse is somewhat smaller than that of the sinewave, and the pilot signal is omitted. The final

waveform to be discussed is shown in Fig. 14-13. Here, alternate peaks of the subcarrier fall on points that again trace out the envelopes of the original left and right signals. The left and right output waveforms are shown in Fig. 14-14 in a "ladder" sequence with the left and right sampling

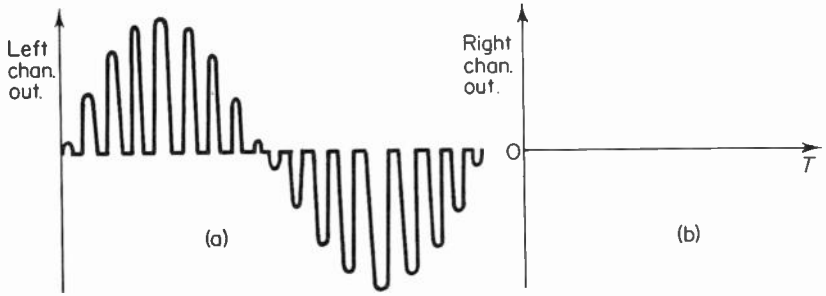



**Fig. 14-12.** Left and right channel outputs of circuit of Fig. 14-10. Right output is zero since only a left signal modulates the main carrier for this example. Compare part (a) with Fig. 14-9.

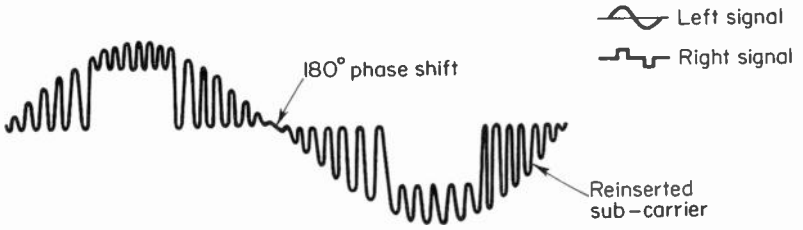



**Fig. 14-13.** Composite signal to be sampled where both left (sinewave) and right (pulse) signals modulate the carrier. Alternate peaks trace out the envelopes of the two signals.

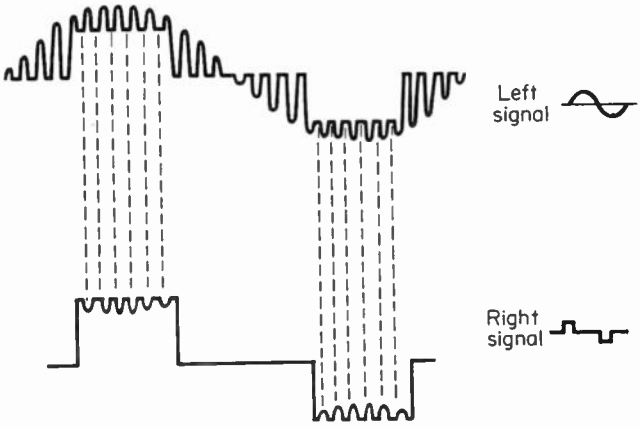



**Fig. 14-14.** Composite signal of Fig. 14-13 to give left and right outputs before filtering.

intervals labeled accordingly. Since the outputs both contain large 38-kHz components, very effective filtering of the signals must be provided.

An approach to the problem of providing effective filtering is to make use of a balanced bridge arrangement. Balance is with respect to the 38-kHz carrier component. Thus, the reinserted carrier component can be removed by the bridge. This will ease the filtering requirements and make demodulation more effective.

The circuit shown in Fig. 14-15 makes use of two balanced bridge circuits that alternately switch the respective load resistors across the composite input. Bridges A and B both contain four matched diodes, which alternately short their respective load resistors when they are forward-biased by the large-amplitude 38-kHz reinserted subcarrier. The load resistors again "view" the composite signal alternately when their respective diodes are reverse-biased or open.

Consider the instant when the polarities are as shown. Notice that the cathodes of D1 and D2 are negative and the anodes of D3 and D4 are positive in the A bridge. All four diodes are thus forward-biased and they all "close." This connects point $A$ to ground and short-circuits $R_{L_1}$. The left channel now "sees" zero volts. At the same time, the anodes of D5 and D6 are negative and the cathodes of D7 and D8 are positive in the B bridge. These four diodes thus are reverse-biased and open. The right-channel load resistor $R_{L_2}$ is *not* short-circuited and is therefore free to "view" the composite signal at point $B$. On the next half-cycle of the 38-kHz subcarrier the polari-

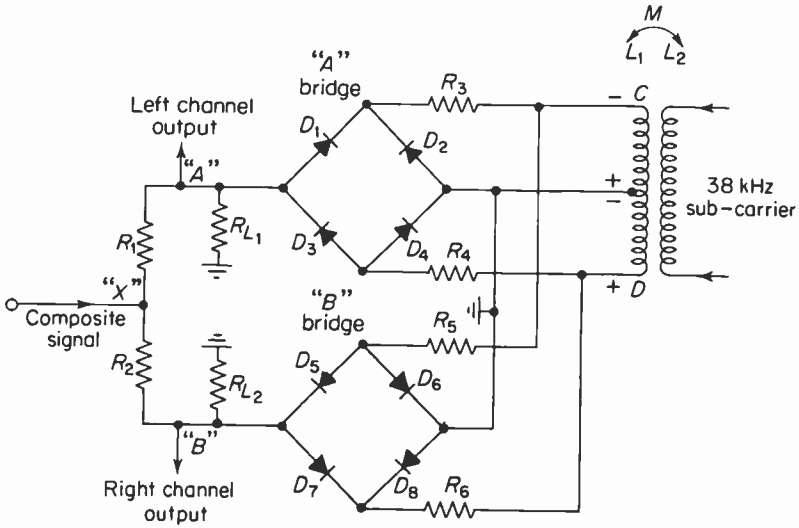



**Fig. 14-15.** Balanced bridge demodulators for sampling the composite signal. The large 38-kHz component does not appear across the output load resistors (see Fig. 14-16).

ties are reversed, so that the B bridge now short-circuits $R_{L_2}$ and the A bridge is open to permit $R_{L_1}$ to sample the composite signal at point $A$.

$R_1$ and $R_2$ are necessary to prevent short-circuiting the composite signal point "X," which is common to both bridges. Resistors $R_3$ through $R_6$ prevent either bridge from short-circuiting the 38-kHz supply points $C$ and $D$.

Figure 14-16 shows how the bridge circuits balance out the large 38-kHz reinserted subcarrier component. If $L_1$ is accurately center-tapped and the diodes are well matched, $I_1 = I_2$. The differential load current due to the 38-kHz subcarrier is zero, since the currents are opposite and thus cancel. When the generator polarities reverse, all four diodes are open and $I_1$ and $I_2$ (due to the subcarrier) are each zero. Once again, the differential current in $R_L$ due to the subcarrier is zero. Conventional low-pass deemphasis networks in subsequent stages can filter out any residual subcarrier components.

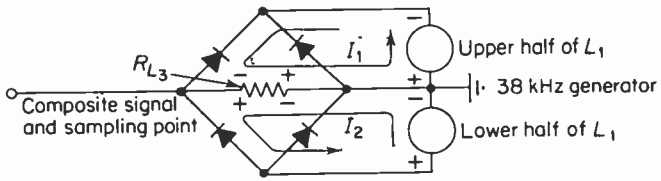



**Fig. 14-16.** Bridge cancellation of the 38-kHz component. $I_1$ and $I_2$ are the load currents due to the reinserted subcarrier. These currents are equal and opposing to give zero differential current. This assumes perfect bridge balance (matched diodes, perfectly center-tapped $L_1$, etc.).

## 14-5. Channel separation aspects†

If a left-only signal is applied to the FM exciter at the transmitter, the output of the multiplexer receiver should be "left" information only. In practical cases, however, the right output will contain some left information and the left output will contain some right information when a right-only signal is applied to the FM exciter. If a left-only signal is applied to the exciter, then the ratio of the left output to the right output is a measure of the degree of separation. In terms of decibels, separation is

$$S = 20 \log \frac{E_L}{E_R}, \tag{14-1}$$

where $E_L$ is the left output due to a left-only signal at the exciter input and $E_R$ is the right output. Ideally, $E_R$ should be zero and the separation would then be infinite db. In the previous matrix equations, it was always assumed that the magnitude of the sum signal was equal to the magnitude of the

†The degree of stereophonic separation is a function of how much "left-channel" information appears at the demultiplexed right output jack when a left-only signal modulates the carrier.

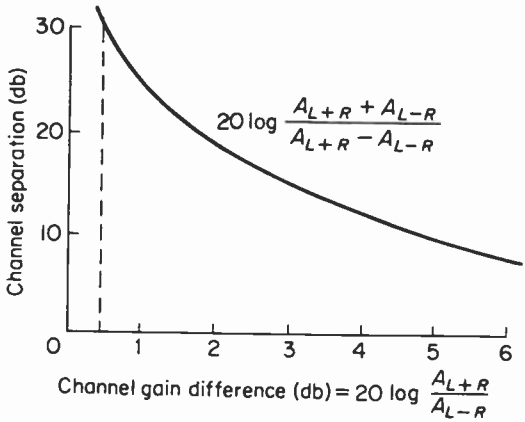



**Fig. 14-17.** Channel separation vs. channel gain difference. A subchannel to main channel relative phase shift of zero is assumed.

difference signal or $(L + R) = (L - R)$. Thus, if the matrix bridge was perfectly balanced and the subchannel and main-channel inputs were equal, the left output appeared at the left output jack and the right output appeared at the right output jack. There would then be no crosstalk between the channels, and perfect separation would be the result. If the gain in the main channel is called $A_{L+R}$ and the gain in the subchannel is called $A_{L-R}$, then the separation is also given by

$$S = 20 \log \frac{A_{L+R} + A_{L-R}}{A_{L+R} - A_{L-R}} \text{ db.} \tag{14-2}$$

Equation (14–2) is graphically shown in Fig. 14-17. Notice that in order to at least meet the requirement at the transmitter, for 30 db separation, the gain difference should not exceed 0.5 db. If the separation is to be at least 20 db, the gain difference between main and subchannels must not exceed 1.7 db. These requirements are difficult to meet, particularly at the higher modulation frequencies. Let us see why.

The FM detector output circuit (prior to stereo demodulation) can be represented by an equivalent generator with internal resistance $R_g$ whose output terminals are shunted by a capacitance $C_s$. $C_s$ does vary with different receivers, length of connecting cable, and such, but typical values range from 50 to 450 pfd. $R_g$ ranges between 5 and 50 kilohms and also varies with the receiver and other factors. Figure 14-18 shows this equivalent circuit to be in the form of a deemphasis network or low-pass filter. The time constant of the network varies between 2 and 10 $\mu$sec, again depending upon the receiver. The detector output level in decibels is given by the following expression, which is derived in Chapter 10:

$$\text{db} = -10 \log_{10} (\omega^2 T^2 + 1). \tag{14-3}$$

A graph (Fig. 14-19) of this equation is plotted for a time constant of 5

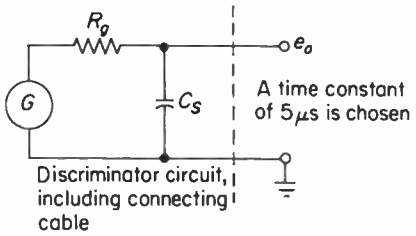



**Fig. 14-18.** The equivalent circuit of the tuner discriminator. This circuit behaves like an equivalent deemphasis network whose time constant varies with different tuners and connecting cables.

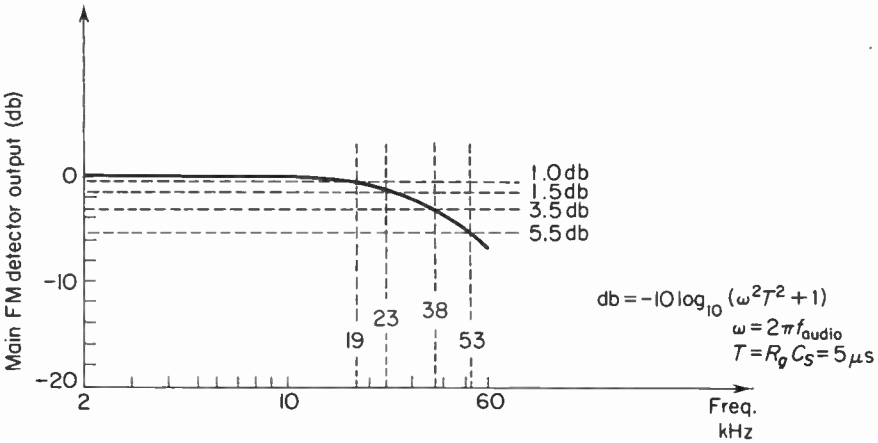



**Fig. 14-19.** Detector output (dB) vs. modulating frequency for an equivalent time constant of 5 μsec.

μsec and a range of frequencies from 3 to 60 kHz. Modulation frequencies up to at least 53 kHz must be included, since the subchannel upper sideband extends out to 53 kHz if the highest modulation frequency of 15 kHz is considered. Notice that at 53 kHz the discriminator output is reduced by 5.6 db, while at 23 kHz (the lower sideband of the subchannel) the reduction is only 1.5 db. This indicates attenuation of the subchannel sidebands relative to the $(L + R)$ main channel. In addition, there is amplitude reduction of 1 db for the 19-kHz pilot carrier plus a small phase shift of the pilot carrier, which manifests itself as a small shift in the phase of the regenerated subcarrier relative to the subchannel sidebands. If the main-to-subchannel gain difference, as indicated above, is not equalized, a maximum channel separation of only 14 db can be realized. This figure will improve as the modulation frequency decreases; nevertheless, some form of compensation must be provided.

Two methods (there are no doubt other ways) can be used to equalize the channel gain difference. One way is to incorporate a *preemphasis* net-

work at the stereo demodulator input such that its high-frequency boost just matches the high-frequency rolloff of the main FM detector output. Another method is to provide a level control that adjusts the magnitude of the main-channel input to the matrix network. This potentiometer, often referred to as a "separation" control, reduces the $(L + R)$ audio signal such that it equals the $(L - R)$ audio and thus equalizes the channel-gain difference. Another method is to vary the balance and separation controls until a left-channel input produces no output in the right channel.

Another point to consider is the phase error between the reinserted 38-kHz subcarrier and the $(L - R)$ sidebands. If this exceeds $\pm 35$ degrees, the separation between channels is reduced to 20 db. If the phase shift is held to no more than 35 degrees, the separation is maintained at 20 db and the distortion is negligible—*provided* the amplitude of the reinserted subcarrier is approximately three times the peak sideband resultant.

The previous discussion on channel separation assumed a perfect phase correspondence between the $(L + R)$ audio main channel and the $(L - R)$ sidebands envelope subchannel. That is, the zero crossings for both signals occurred at exactly the same time. In practice, this generally will not be true. Instead, a relative phase shift, $\theta$, will occur as shown in Fig. 14-20. Now, if a left-only signal is applied to the modulator, we should not have any residual left signal appearing in the right-channel output. It will be recalled that infinite channel separation occurs only if no information appears in the right channel when a left-only signal is fed to the modulator. Let us see what happens if there is a relative phase shift between the main and subchannel signals.

Using Fig. 14-20, we add the left-only audio signal [part (a)] to the $(L - R)$ sidebands envelope [part (b)] to obtain the composite signal [part (c)]. Notice the difference between the composite signal of part (c) and the waveform of Fig. 14-10. A phase shift between channels causes the baseline to fluctuate by an amount equal to "$A$." Reference to Fig. 14-10 shows that the positive peaks of the first half-cycle of the composite signal describe the left audio signal, while the negative peaks of the first half-cycle describe the right audio signal. Also, the negative peaks of the second half-cycle describe the remaining half-cycle of the left audio signal, while the positive peaks describe the remaining half-cycle of the right signal. Since the baseline in Fig. 14-10 was constant, the subchannel detector produced zero output in the right channel. In Fig. 14-20, however, the subchannel detector will produce an output in the right channel equal to "$A$" and an output in the left channel equal to "$B$." The channel separation in decibels is then

$$S = 20 \log \frac{\text{peak-to-peak amplitude of composite signal}}{\text{peak-to-peak variation of baseline}} = 20 \log \frac{B}{A}. \tag{14-4}$$

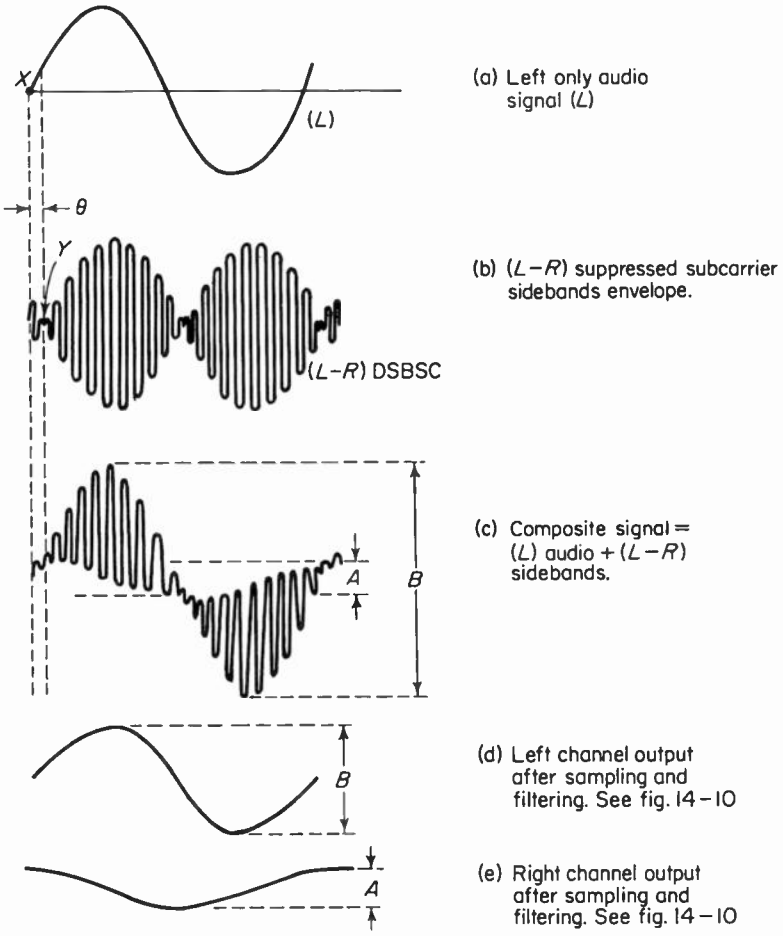



**Fig. 14-20.** The sidebands envelope zero crossing $Y$ lags the left audio zero crossing $X$ by an angle $\theta$. This produces an output $A$ in the right channel with a left only signal. In order to maintain $A$ 30 dB or more below $B$, the main-to-subchannel differential phase shift $\theta$ should not exceed $\pm 3°$ from 50 to 15,000 Hz.

In terms of the angle $\theta$, the channel separation can be stated as

$$S = 20 \log \sqrt{\frac{1 + \cos\theta}{1 - \cos\theta}} \text{ db}, \tag{14-5}$$

where the main-to-subchannel differential phase shift $\theta$ is as defined in Fig. 14-20. The relationships above assume that the main and subchannel signals are accorded the same "treatment" as far as amplitude is concerned, or $(L + R) = (L - R)$. Thus, the channel separation is degraded as a result of

either phase or amplitude nonlinearities and usually by a combination of both. If we now combine the effects of both amplitude and phase differences between channels into a single expression for channel separation, we may state the separation as

$$S_c = 20 \log \sqrt{\frac{A_d^2 + 2A_d \cos\theta + 1}{A_d^2 - 2A_d \cos\theta + 1}} \quad \text{db}, \tag{14-6}$$

where $\theta$ is the phase shift as defined in Fig. 14-20 (degrees), and $A_d$ is the ratio of the main to subchannel gains or $A_{L+R}/A_{L-R}$ (or $A_{L-R}/A_{L+R}$).

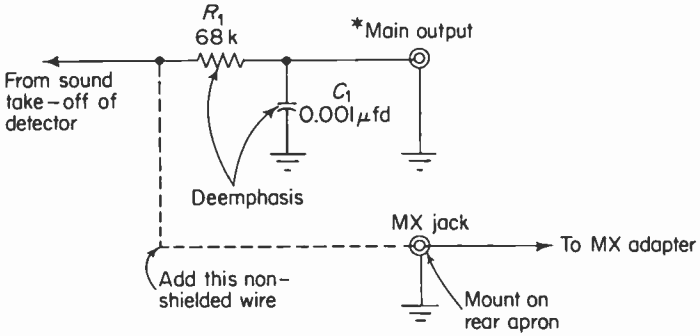
For example, compute the channel separation in decibels for a channel phase difference of $\pm 3°$ (the "nominal" maximum permitted at the transmitter) and a channel gain difference of 1.035. $A_d$ is then 1.035, or the main and subchannel gains are within $\pm 3.5$ per cent of each other (the maximum allowed). Then

$$S_c = 20 \log \sqrt{\frac{1.035^2 + 2.07 \cos 3° + 1}{1.035^2 - 2.07 \cos 3° + 1}} = 27.1 \quad \text{db}.$$

It is stressed that any phase or amplitude differences introduced by the *receiver* equipment will simply magnify the problem by reducing the channel separation in accordance with the above expression.

## 14-6. Miscellaneous aspects

Many FM tuners in use can be easily converted to multiplex operation by means of a few minor modifications. This is not to say that every tuner can provide high-quality multiplex performance, since this depends upon such factors as detector bandwidth and linearity, IF phase-shift characteristics, tuner output voltage, and so on. As far as phase linearity is concerned, the standards of transmission relating to the transmitter suggest that the maximum allowable phase shift between the main and subchannels should not exceed plus or minus 3 degrees. This is not a simple requirement to meet, particularly at frequencies above 10 kHz. In addition, only 3 degrees of phase shift between the transmitted pilot and the locally generated subcarrier can be tolerated. If the standards at the transmitter are maintained (and there is every reason to believe that they are), it is only natural that the tuner should not degrade stereo separation. Thus, minimum "ripple" in the IF passband with a flat-topped response is highly desirable. Also, since the tuner main detector must be capable of detecting high-frequency components (53 kHz, for example) with minimum distortion, it should be free from diagonal clipping effects due to the load time constant. A wideband detector would help greatly in this respect. Phase-shift discriminators will provide excellent results if their load impedances (the connecting cable, the input impedance of the first stage in the multiplex demodulator, and so




*If main output is no longer needed, additional wire (shown dotted) and MX jack are not required. Simply clip out $C_1$ and short $R_1$. Connect main output jack to adapter input jack via short cable.

**Fig. 14-21.** FM tuner modification for multiplex reception.

on) are kept high. Since most *MX* adapters provide a reasonably high input impedance, the shielded connecting cable must be kept as short as possible. The use of very low-capacitance cable is a wise investment for chassis interconnection.

The modification is simply made by locating the deemphasis network at the tuner output (generally, but not always, a 68-kilohm, 0.001-microfarad combination) and connecting a wire (nonshielded) from the "input" side of the deemphasis network to a jack (phone-type), which should be mounted on the back panel. When mounting this jack, place it as close as possible to the signal take-off point. The modification is shown in Fig. 14-21. If the tuner already has a multiplex jack, the only modification to be made (assuming that the tuner performance is acceptable) is to short any coupling capacitor that may be in the *MX* feed line. This is particularly important when the capacitor is small, as is often the case. Failure to do this will result in severe low-frequency attenuation.

If one wishes to improve the high-frequency response of a discriminator at minimum cost, it can be simply done by changing the load components as shown in Fig. 14-22. A comparison between the values shown and those of a conventional discriminator will indicate that the discharge time constant has been reduced. This minimizes diagonal clipping (the capacitor can now more readily follow signal peaks in the modulation envelope) at the expense of additional damping imposed on the tuned circuit feeding the diodes. This slight loss in sensitivity is not serious.

Most *MX* adapters require approximately 0.1 volt (more or less) pilot voltage at the input jack for locking the 38-kHz oscillator. An examination of the specification for the tuner (to be adapted) will frequently give the nominal audio output for a 33 per cent carrier deviation. This is generally in the range of 2 volts. Since the pilot modulates the main carrier 10 per

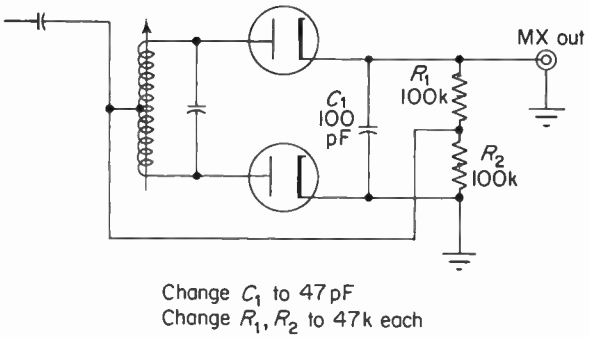



**Fig. 14-22.** Minor detector modification for improving high-frequency response. Be sure that $R_1$ and $R_2$ provide a reasonably close match (5% is reasonable). Values shown in schematic are typical for many discriminators.

cent, simply divide this number by 10 to obtain the amount of pilot voltage available to lock the oscillator in the adapter.

A high-gain, directional antenna is a necessity for receiving multiplex. A piece of wire hanging from the back of the tuner is *not* acceptable and a nominal investment in a good antenna will pay dividends. If the antenna installation is already available but the tuner is located in an outlying area, the only solution to the problem is a higher-sensitivity tuner if the existing equipment is not sensitive enough.

One last precautionary measure should be mentioned. Be certain that the FM tuner alignment is as correct as possible. Misalignment of the IF strip and detector will surely degrade the phase response (and with it, separation).

## 14-7. A transistorized adapter for decoding multiplex signals

Figure 14-23 shows a four-transistor adapter that provides small size, requires a simple power supply (small current drain), and gives high input impedance, high breakdown voltage, and long trouble-free service. The units shown are NPN silicon planar types,† while the diodes used in the subchannel detector circuit may be general-purpose video detector diodes. The circuit shown does not include the power supply, which can be provided by a battery or a separate power supply, or the unit may be powered from the main ampli-

†The planar process is an oxide-masking process whereby a silicon oxide "mask" is grown on the silicon slice prior to the emitter and base diffusion operation. The oxide serves as a mask to protect the PN junctions and is removed after diffusion from the emitter and base portions. Planar types have extremely low leakage currents, improved current gain (beta) at low currents, and good stability.

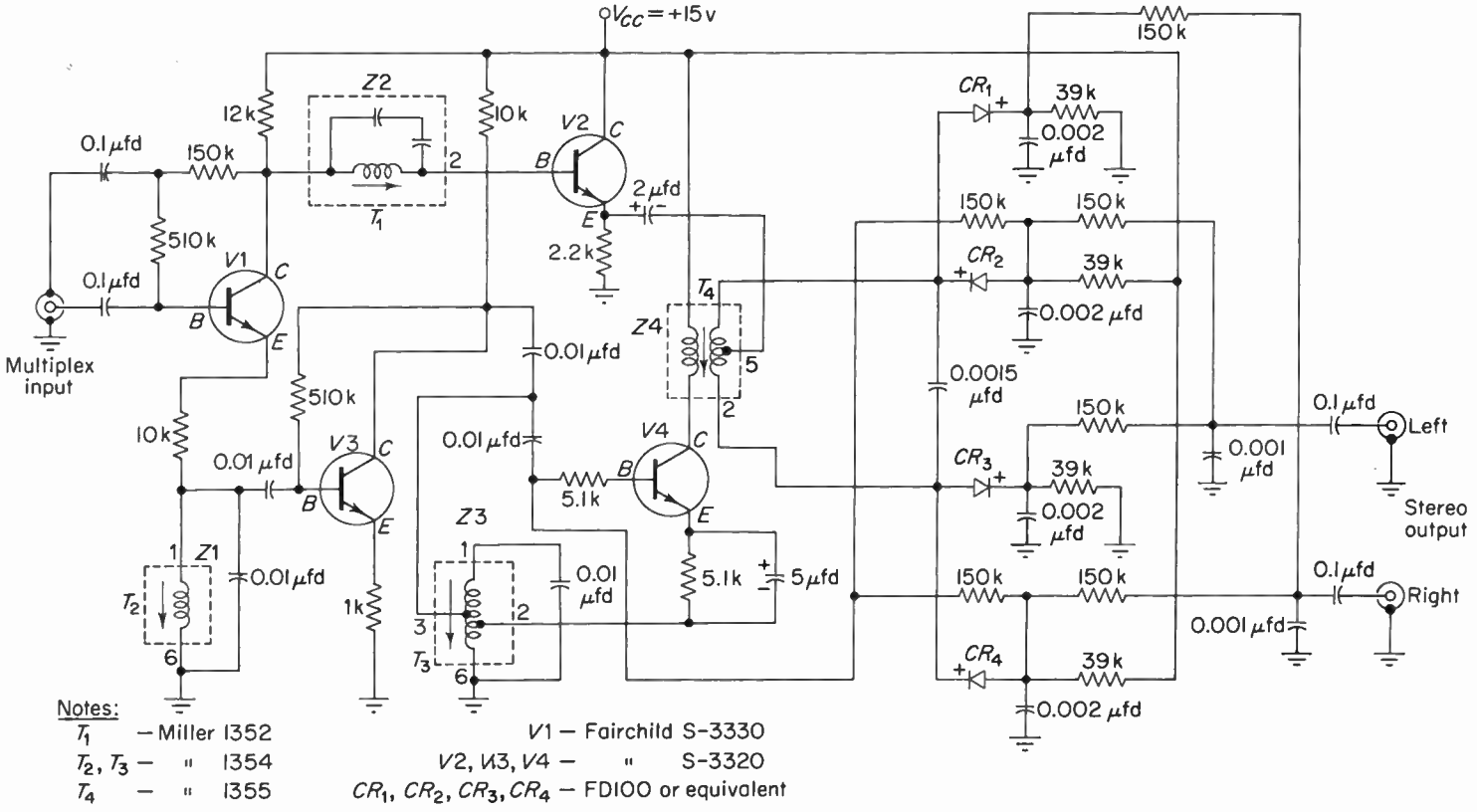



**Fig. 14-23.** Schematic of transistor multiplex adapter. This adapter utilizes the time division or sampling technique for subchannel detection. The unit provides very good performance and requires no external controls. (Courtesy of *Electronics World*, reprinted by permission.)

fier chassis. The power-supply current drain is approximately 7 milliamperes.

Transistor $V_1$ provides a fairly high input impedance (250 kilohms shunted by 20 pfd) and is of the common-emitter type. The parallel resonant tank in the emitter circuit is tuned to 19 kHz, so that a large amount of negative feedback is provided to reject the pilot. Very little 19-kHz pilot voltage thus appears at the collector terminal. However, a large 19-kHz pilot component appears across the emitter tank. This signal is capacitively coupled to the base of $V_3$ for amplification. In addition, the composite signal (less pilot) appears at the collector of $V_1$. A 67-kHz SCA rejection filter ($Z_2$) attenuates this subcarrier, so that the signal appearing at the input to $V_2$ (a high-input-impedance emitter follower) will consist only of the $(L + R)$ 50 to 15,000-kHz main-channel sidebands envelope. The composite signal is then capacitively coupled to the center-tap of the $T_4$ transformer.

The filtered and amplified 19-kHz pilot voltage appearing at the collector of $V_3$ is injected into the tuned circuit ($Z_3$) to lock the 19-kHz oscillator ($V_4$). Nonlinear operation of $V_4$ generates a 38-kHz component (second harmonic) in the oscillator circuit, which is selected by the tuned circuit of $Z_4$. Thus, at opposite ends of the secondary of $T_4$, we have a large-amplitude push-pull 38-kHz subcarrier for switching the diodes on and off in synchronism with the transmitter. Four switching diodes are used here to sample the composite signal in a time-division multiplex arrangement.

The bridge equivalent for the left and right subchannel detectors is shown in Fig. 14-24 together with the push-pull 38-kHz reinserted subcarriers and composite multiplex signals. The composite multiplex signal consists of the $(L + R)$ main-channel and $(L - R)$ DSBSC signals, and this generator is labeled "*MX*." The signal for this example consists of a "left-only" signal. Thus, the right-channel output should be zero. Waveform "A" is applied to diodes $CR_3$ and $CR_4$. Waveform "B" is applied to diodes $CR_2$ and $CR_1$. $CR_3$ rectifies "A" to produce the upper portion of the modulation envelope with $R_1$-$C_1$ comprising the load for $CR_3$. $C_1$ charges quickly to the positive peaks of "A" to produce "C." $C_2$ charges quickly to the negative peaks of "A" to produce "D." $C_3$ charges quickly to the negative peaks of "A" to produce a 38-kHz ripple (zero audio-modulation component), while $C_4$ charges quickly to the positive peaks of "B" to produce a 38-kHz ripple (zero audio-modulation component). "C" and "D" are fed to the $R_5$-$C_5$ and $R_7$-$C_5$ deemphasis networks (low-pass filter). The 38-kHz components are filtered out, leaving "E," which is 1L. "F" and "G" are also fed to deemphasis networks ($R_6$-$C_6$ and $R_8$-$C_6$), where the 38-kHz component is filtered out, leaving a zero "H."

This adapter can provide 25 db separation at 1 kHz, 19-kHz pilot carrier rejection of 20 db below 1 volt, 38-kHz subcarrier rejection of 30 db below 1 volt, and a required signal input of 0.5 to 1 volt rms relative to 100 per cent modulation of the main carrier.

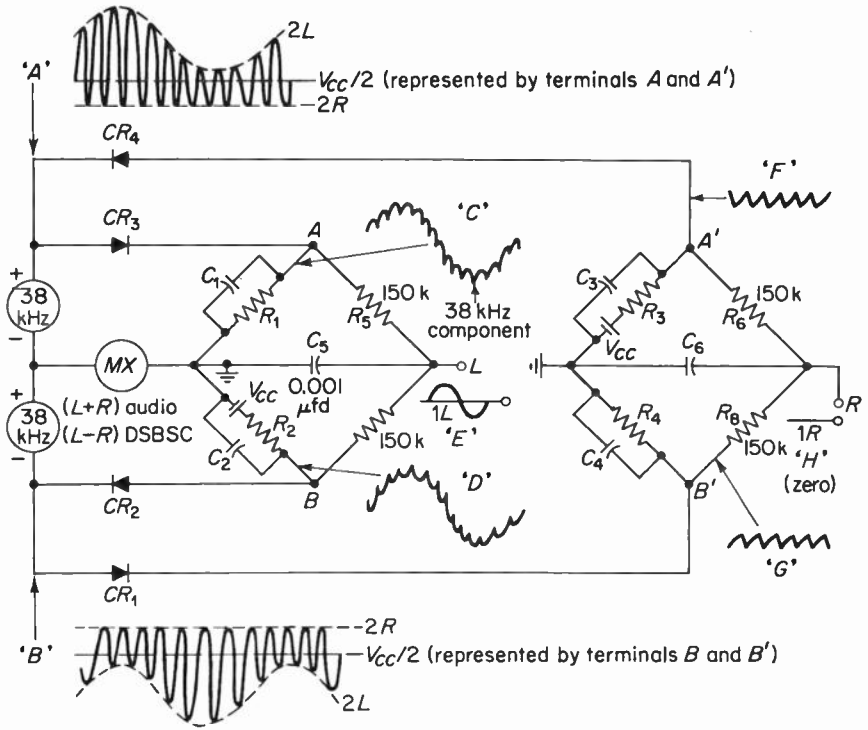



**Fig. 14-24.** The bridge equivalent of the subchannel detector of Fig. 14-23. A "left only" signal is modulating the 38-kHz suppressed subcarrier. Matched diodes and bridge balance ensure optimum channel separation as well as maximum rejection of subcarrier components.

## 14-8. Summary

**1.** FM multiplexing involves the modulation of an FM carrier by two or more independent signals.

**2.** The main and subchannels are decoded to provide two output signals, which are fed to two audio-amplifier chains.

**3.** The purpose of the 19-kHz pilot carrier is to lock an oscillator in the receiver.

**4.** The sum of the left and right signals can frequency-modulate the main carrier up to a maximum of 90 per cent of system deviation.

**5.** The difference channel amplitude-modulates a suppressed 38-kHz subcarrier. The resulting signal frequency modulates the main carrier up to a maximum of 90 per cent of system deviation.

**6.** Interleaving prevents both signals (the sum signal and the sidebands envelope) from overmodulating the main carrier.

**7.** The composite signal can be generated by either sum and difference matrixing or by time-division multiplexing. Both methods are compatible.

**8.** IF and detector phase and amplitude responses should be flat in order not to distort the higher modulation frequencies.

**9.** One decoding method uses a pair of matched diodes to decode the subchannel. This is followed by a balanced bridge matrix for separating the left and right channels.

**10.** Another method uses a pair of balanced bridge demodulators, which alternately sample the left and right audio signals. This is the time-division method.

**11.** The 38-kHz reinserted subcarrier and the 67-kHz SCA subcarrier should be carefully filtered out in order not to degrade tape recordings.

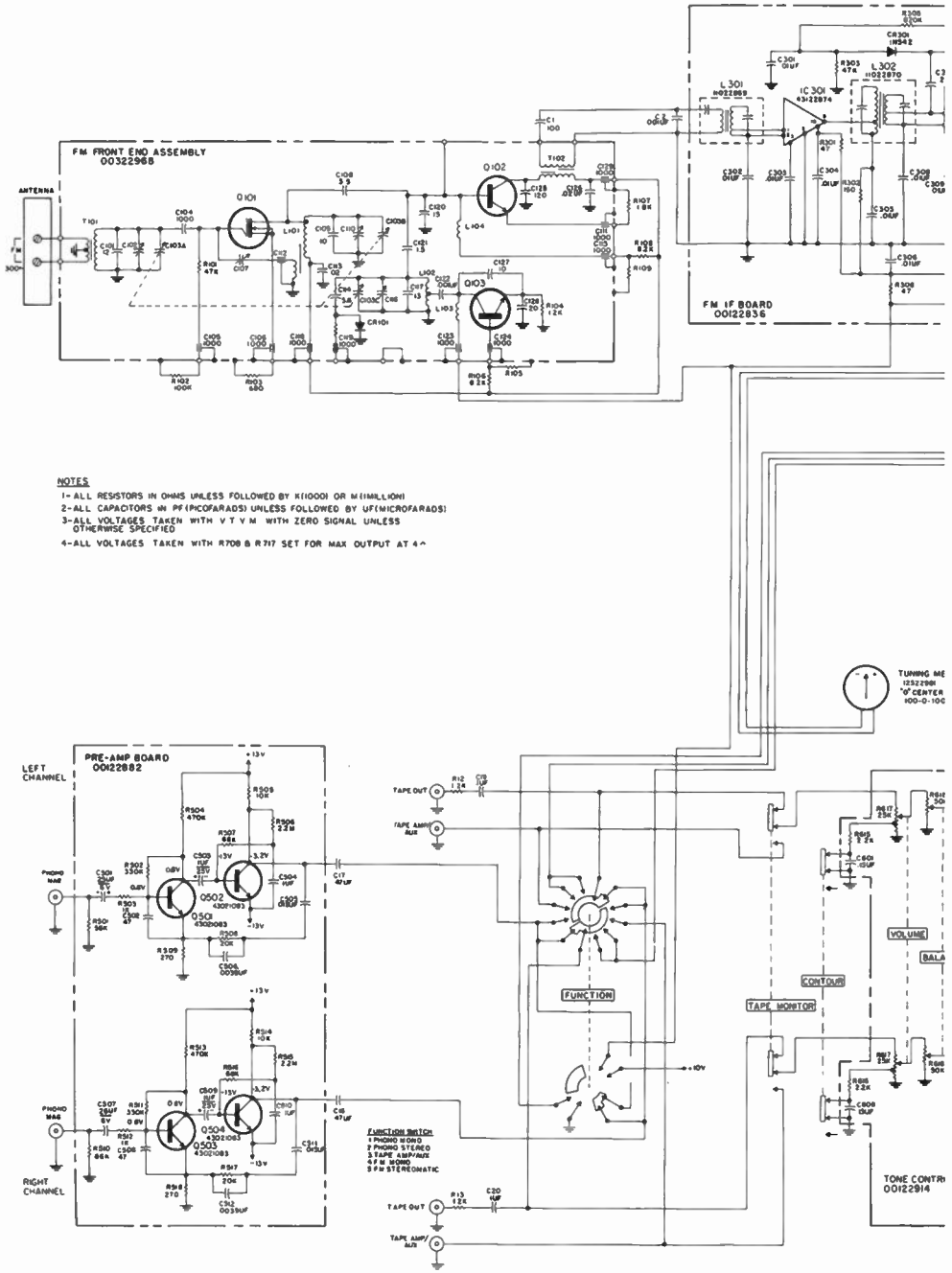
**12.** Fully integrated FM multiplex tuners incorporate the multiplex decoding circuitry on the same chassis as the main FM tuner.

**13.** Existing FM tuners can be easily modified to provide multiplex signals by utilizing existing jacks and disabling the built-in deemphasis network.

## REFERENCES

1. Csicsatka, A., and Linz, R. M.: "Factors Affecting Overall Performance of FM Stereophonic Receivers," *IRE Trans. on Broadcast and Television Receivers*, **BTR-7** (July 1961).
2. DeVries, A. J.: "Design of Stereophonic Receiver for a Stereo System in the FM Band Using an AM Subcarrier," *IRE Trans. on Broadcast and Television Receivers*, **BTR-7** (July 1961).
3. Eilers, C. G.: "Stereophonic FM Broadcasting," *IRE Trans. on Broadcast and Television Receivers*, **BTR-7** (July 1961).
4. Von Recklinghausen, D. R.: "Stereophonic FM Receivers and Adapters," Reprint of paper given at Western Electronic Show and Convention, August 1961.
5. Coffman, B. C.: "Spectrum Analysis of Time Division Stereo," *Electronic Industries*, September 1960.
6. Dome, R. B.: "Stereophonic System for Television Broadcasting," *Electronics*, December 16, 1960.
7. Saslaw, D.: "What Hath FCC Wrought," *Audio*, June 1961.
8. Csicsatka, A., and Linz, R. M.: "FM Stereo—The General Electric System," *Audio*, June 1961.
9. Eilers, C. G.: "FM-Stereo: Time-Division Approach," *Audio*, August 1961.
10. Shottenfeld, R., and Abilock, S.: "Signal Sampling for FM Stereo," *Audio*, December 1961.
11. Von Recklinghausen, D. R.: "A Multiplex Stereo Generator," paper given at the 13th annual meeting of the Audio Engineering Society, October 1961.
12. Schwartz, M.: *Information Transmission, Modulation, and Noise*. McGraw-Hill Book Company, New York, 1959.

The schematic diagram on pages 518 and 519 shows FM receiver. In this receiver the RF amplifier uses an MOSFET and the intermediate frequency amplifier–limiter stages are integrated circuits similar to that shown in Fig. 7-23. *(Courtesy of Harmon-Kardon, Inc.)*



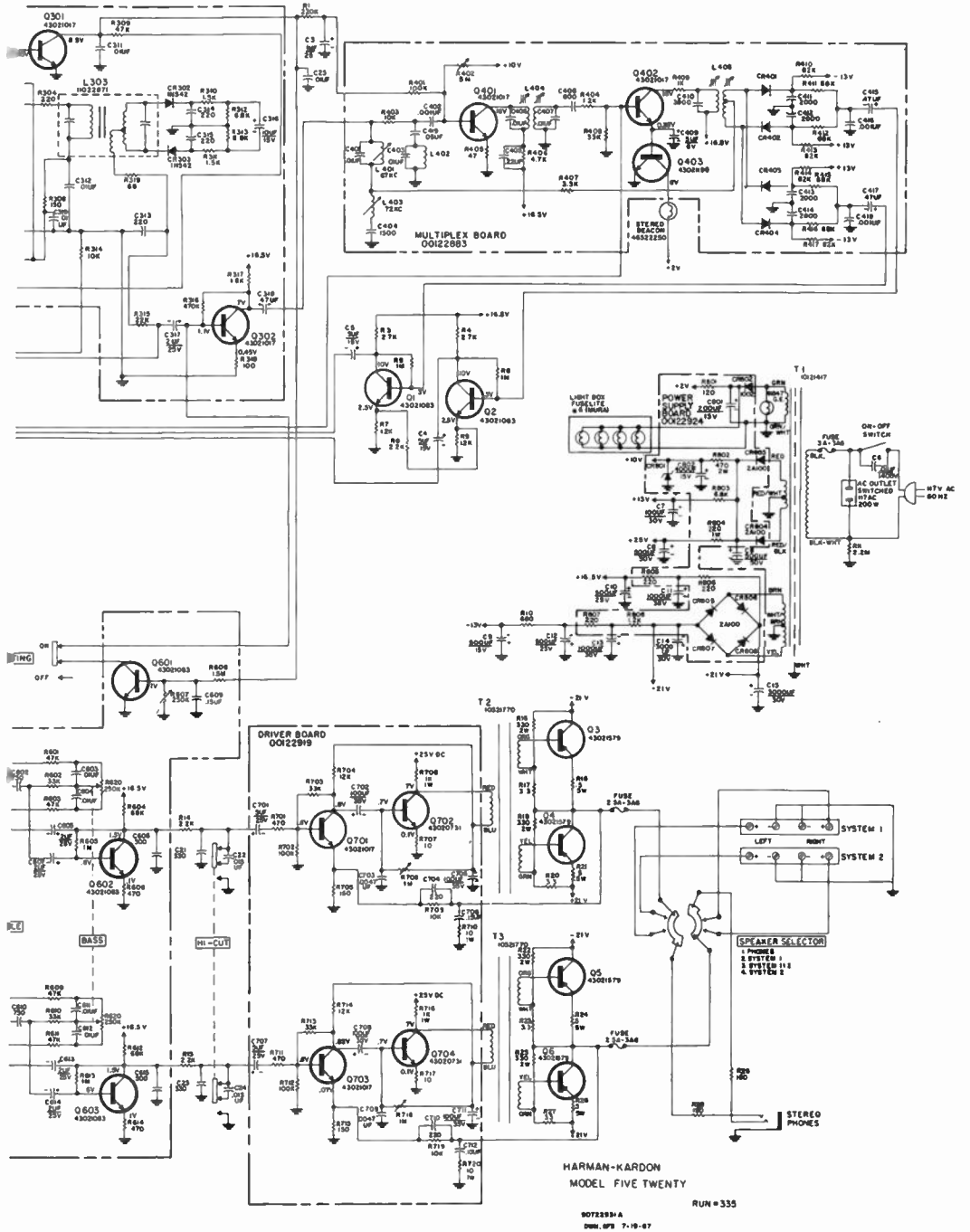

NOTES

1- ALL RESISTORS IN OHMS UNLESS FOLLOWED BY K(1000) OR M(1MILLION)

2-ALL CAPACITORS IN PF(PICOFARADS) UNLESS FOLLOWED BY UF(MICROFARADS)

3-ALL VOLTAGES TAKEN WITH V T V M WITH ZERO SIGNAL UNLESS OTHERWISE SPECIFIED

4-ALL VOLTAGES TAKEN WITH R708 & R717 SET FOR MAX OUTPUT AT 4 ^


MULTIPLEX BOARD
00122883
POWER SUPPLY BOARD
00122924
DRIVER BOARD
00122919
HARMAN-KARDON
MODEL FIVE TWENTY
RUN = 335
SPEAKER SELECTOR
SYSTEM 1
SYSTEM 2
STEREO PHONES
BASS
HI-CUT

# INDEX

## A

## C



## D



## E



## F

## G



## H



## I

## J



## K



## L



## M



## N

## O



## P



## Q



## R

## S



## T

## V



## W

(continued from front flap)

tors represent the most comprehensive treatments of both topics available anywhere in the literature of the field.

An important section of the book covers noise and VHF phenomena to account for the important factor of the very high frequencies at which commercial FM receivers operate and the weak signal reception which is a major characteristic of this receiver. Each section of the work ends with a detailed summary of the material covered, and is augmented by sample problems which can be directly applied to the reader's work.

Throughout, the work places the greatest stress on the operational aspects of FM receivers, and in its completeness and coverage of the most recent developments in the field, represents the most useful reference tool for any professional in FM technology.

---

A. B. COOK is Staff Engineer at Perkin-Elmer Corp.

ALVIN LIFF is Administrator of Special Projects at the RCA Institute, Inc. In addition to the present volume, he is the author of two widely used laboratory manuals in his field.

PRENTICE-HALL, Inc.
Englewood Cliffs, New Jersey

968 • Printed in U.S. of America

## Other recommended books of interest . . .

### TRANSFORM CIRCUIT ANALYSIS FOR ENGINEERING AND TECHNOLOGY

by WILLIAM D. STANLEY, *Old Dominion College, Virginia*

Here is a book which details the fundamentals of transient circuit and systems analysis employing the LaPlace transform and pole-zero approach for analyzing and interpreting problems. Essential to engineers and technicians, this guide to the current transform methods is presented in a clearly illustrated step-by-step sequence, and is oriented to the engineering approach, thereby making only a minimum of advanced mathematics necessary to completely master the subject.

In its careful development of the material, the reader is carried from circuits through to basic systems, establishing a solid background for further work in electronics, control systems, and communications systems. Whether as a reference for working engineers and technicians, or as a self-instructional guide to the fundamentals of the subject, this book is a valuable working tool for anyone in electronic technology and engineering.

*Published 1967* — *368 pages*

### ELECTRONIC TROUBLESHOOTING: A Self-Instructional Programed Manual

by Members of the Staff, *Philco Technical Institute*

This book gives the vital know-how to troubleshoot rapidly and accurately to minimize equipment "downtime"; it guides you to fast, accurate troubleshooting by showing you how to apply logical thinking and a systematic approach to each problem.

The author explains in the Preface, ". . . This is not an ordinary text; it is a self-instructional programed manual. Although the pages are numbered consecutively, you will not study them in that order. After studying a particular information block, you check your understanding by selecting one of the several choices offered as possible answers to a question pertaining to that information block. You then turn to the indicated page and are told whether or not you made the correct selection. If you have chosen the correct answer, you go on with the learning process. If you have selected an incorrect answer, you are told why and asked to review the information before going any further. Part I outlines troubleshooting principles and practice; Part 2 provides actual troubleshooting problems."

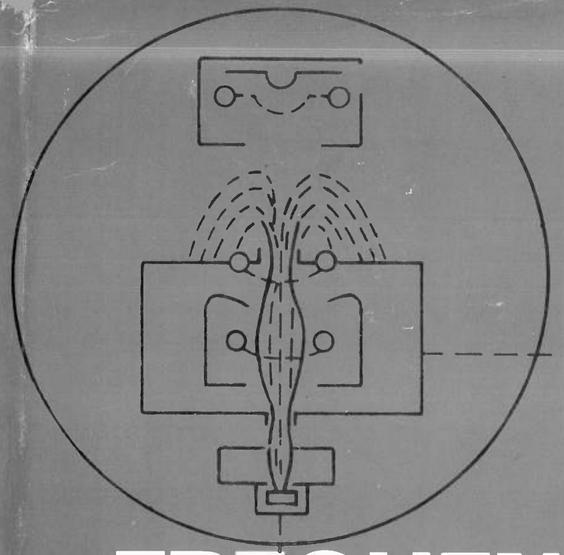
The major advance in this self-instructional programed manual is that you are never permitted to form erroneous ideas and/or arrive at the wrong conclusion. Following the procedure outlined in the Preface, you will acquire a real understanding of how to deal with any troubleshooting problem.

*Published 1966* — *276 pages*

PRENTICE-HALL, Inc.
Englewood Cliffs, New Jersey



FREQUENCY MODULATION RECEIVERS
A.B. Cook/A.A. Liff
PRENTICE-HALL



# FREQUENCY MODULATION RECEIVERS

## A.B. Cook/A.A. Liff

The most comprehensive and up-to-date treatment available on the operational aspects of wide band FM receivers. Covers all circuits, auxiliary circuits, and other major aspects of the field. An ideal reference for in-service professionals.

Providing one of the most comprehensive treatments on the subject of wide band FM receivers, this book offers an up-to-date reference to the entire subject of commercial Frequency Modulation broadcast receivers. It forms an ideal one-volume reference for in-service professionals—RF Engineers and Technicians, Communications Engineers, Broadcast Engineers, Service Technicians, as well as for those readers planning to enter any of these professions.

The book offers a complete discussion of the circuits which comprise an FM receiver, consequently covering all basic circuits such as the RF amplifier, mixer-oscillator, IF amplifiers, limiter, as well as all types of FM detectors. Major circuits are discussed in terms of vacuum tube, transistor, field effect transistors, and integrated circuit characteristics. Thus all areas of past and present practice are covered. Also discussed in detail are the various receiver auxiliary circuits such as tuning indicators, stereo multiplex, and squelch.

The two sections of this work covering FM detectors and tuning indica-

(continued on back flap)