FUNDAMENTALS OF RADIO

BY

FREDERICK EMMONS TERMAN

Professor of Electrical Engineering

Stanford University

WITH THE COLLABORATION OF

LT. F. W. MACDONALD, U. S. N.

FIRST EDITION

ELEVENTH IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK AND LONDON

1938

PRINTED IN THE UNITED STATES OF AMERICA

FUNDAMENTALS OF RADIO

Recommendations of the War Production Board for conserving paper and other critical materials have been observed in the manufacture of this book, and any variations from the format of earlier printings of the same book are the result of this compliance.

PREFACE

The purpose of "Fundamentals of Radio" is to present the basic principles of radio communication in a form suitable for use in an introductory radio course. This book is essentially an abridged version of the author's "Radio Engineering." It presents the subject with the same organization, the same viewpoint, and the same style, but the treatment is simplified. The length of the text has been nearly halved, and the problems have been increased in number and made primarily of a classroom type.

The simplification consists largely in confining the treatment more closely to fundamental principles. This policy has been followed in the belief that the most satisfactory method of presenting a subject in an elementary course is to concentrate on the fundamental concepts, and to avoid diverting attention from these by too much consideration of all the possible consequences, applications, and implications of these principles. By following this procedure the number of new ideas that the student must understand and organize in his own mind is kept at a minimum, and yet a solid foundation is laid for future study. In contrast with this, "Radio Engineering" strives for a complete comprehensive treatment of all phases of the subject, with extensive references to the significant literature. In comparing the two books, "Radio Engineering" can be thought of as being primarily a textbook for the more advanced student and as a reference book for the practicing engineer, while "Fundamentals of Radio" is primarily a textbook of more elementary character.

The chief prerequisite for study of the material in "Fundamentals of Radio" is an elementary understanding of alternating-current circuits, and in particular the conceptions associated with the terms leading, lagging, impedance, reactance, etc. A knowledge of complex quantities is not necessary.

The present book is provided with a comprehensive set of problems. There is, in fact, more than one problem per page of type, and care has been taken to distribute these problems uniformly over the text. An instructor using the book for class work will therefore find a number of problems directly applying to every text assignment made. These problems differ in many cases from those in "Radio Engineering" in that they are intended for use in connection with class assignments, and so are nearly all of the type that can be done overnight. In contrast with this, a considerable number of the problems in "Radio Engineering" represent projects that take several days of work to complete, and so are of the type that fit in well with home study or supervised individual work.

It is a pleasure to acknowledge the collaboration of Lieutenant F. W. MacDonald in the preparation of this work. Lieutenant MacDonald gave to this undertaking the benefit of his extensive experience in handling the introductory radio course given at the United States Naval Academy. In particular, he contributed a preliminary draft for the first ten chapters of the book, which has been closely followed in the final product. This draft was of great assistance in fixing a suitable level of difficulty for the presentation and in defining the scope of the treatment.

Valuable cooperation was rendered by Lt. Colonel C. L. Fenton of the United States Military Academy. In particular, Colonel Fenton was especially helpful in outlining the problems which he had encountered in conducting elementary radio courses, and in making available for the author's guidance the problems he has required of his students.

The author also wishes to acknowledge the assistance rendered by John R. Woodyard and Robert L. Sink, graduate students in electrical engineering at Stanford University. Mr. Woodyard proofread the preliminary and final drafts of the manuscript, as well as the page proof. He also made innumerable suggestions of value, and served as a most constructive critic. Mr. Sink rendered competent assistance in the preparation of the new figures, tables, etc., and made all the new calculations required in connection with the book.

Thanks are also due Gilfillan Bros., Inc., Zenith Radio Corp., RCA Mfg. Co., Heintz & Kaufman, Ltd., Collins Radio Co., J. W. Miller Co., E. F. Johnson Co., Cornell-Dubilier Corp., Aerovox Corp., Sprague Products Co., and Radio Condenser Co., for photographs and circuit diagrams made use of in the book.

FREDERICK EMMONS TERMAN.

STANFORD UNIVERSITY,
December, 1937.

CONTENTS

FUNDAMENTALS OF RADIO

CHAPTER I

THE FUNDAMENTAL COMPONENTS OF A RADIO SYSTEM

1. Radio Waves.—Radio communication is carried on by means of energy that travels from the transmitter to the receiver in the form of radio waves. These waves move with the velocity of light and represent electrical energy that has escaped into free space. The waves consist of magnetic and electrostatic fields at right angles to each other and also at right angles to the direction of travel, and if they were visible would appear as shown in Fig. 1.

The chief characteristics of a radio wave are the frequency (or wave length) and the intensity. The frequency represents the number of complete cycles of oscillations that the transmitter sends out per second, and so is the frequency of the alternating current producing the wave. The wave length is the distance in space occupied by one complete cycle of oscillation, as shown in Fig. 1. The wave length λ in meters and the frequency f in cycles are related by the equation

$$\lambda = \frac{300,000,000}{f} \tag{1}$$

where the quantity 300,000,000 is the velocity of light in meters per second. The frequency is ordinarily expressed in kilocycles, abbreviated kc, or in megacycles, abbreviated mc. A low-frequency wave is seen from Eq. (1) to have a long wave length, while a high frequency corresponds to a short wave length.

Radio waves differ from other electromagnetic waves, such as light, only in wave length (or frequency). The relationship between various types of electrical radiations is indicated by Table I.

TABLE I.—WAVE-LENGTH SPECTRUM OF ELECTRICAL PHENOMENA

Type	Typical Wave Lengths
Radio waves	2×10^{4} to 0.1 meter
Infra-red or heat waves	1×10^{-5} meter
Light waves	5×10^{-7} meter
Ultra-violet waves	3×10^{-8} meter
X-rays	1×10^{-10} meter
Gamma rays	1×10^{-12} meter
Cosmic rays	1×10^{-14} meter

The strength or intensity of a radio wave is measured in terms of the voltage stress produced by the electrostatic field of the wave, and is usually expressed in microvolts stress per meter. The strength in microvolts per meter is also exactly the same voltage that the magnetic flux of the wave induces in a conductor 1 meter long when sweeping across the conductor with the velocity of light.

The minimum field strength required to give satisfactory reception of a radio wave varies with the amount of interference that is present. Under very favorable conditions waves having a strength of less than 1 μv per meter will produce intelligible signals. Much greater field strengths are generally necessary, however, because of interfering waves generated by man-made and natural sources. Thus in rural areas experience has shown that it normally requires a field strength of the

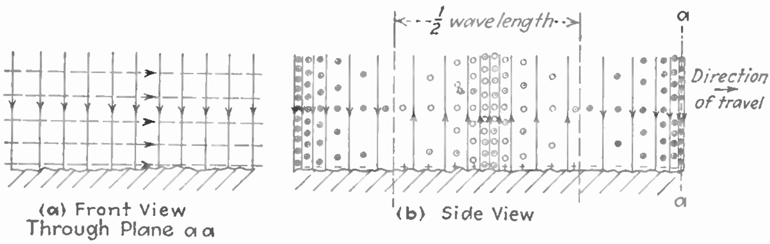


FIG. 1.—Front and side views of a vertically polarized wave. The solid lines represent electrostatic flux, and the dotted lines and circles indicate magnetic flux. This may be considered as an instantaneous view or "snapshot."

order of 100 μv per meter to give what the listener considers satisfactory service from a broadcast station, whereas in urban locations the man-made interference is so great that field strengths of 500 to 30,000 μv per meter are ordinarily needed to insure good reception at all times.

Propagation Characteristics of Radio Waves.—The strength of the radio wave reaching the receiver from a distant transmitter is affected by a number of factors. Most important of these are the spreading of the wave owing to distance, the absorption of energy by the earth, and the attenuation and refraction of the wave by the action of the ionized regions (or ionosphere) of the upper atmosphere. The effect of the earth and the ionized regions depends very greatly upon the frequency. Thus low-frequency radio waves such as 12 to 100 kc suffer very little attenuation other than that due to spreading, and the received signal strength does not differ greatly from day to night and season to season. In contrast with this, waves of broadcast frequencies (550 to 1500 kc) have such high attenuation in the daytime that only local stations can be heard, while at night the attenuation is frequently so low that very distant stations produce satisfactory signals. High-frequency waves, such as those in the range 6 to 30 mc, behave in a still different manner.

Here the ground very quickly absorbs the portion of the wave traveling along the earth's surface. Waves of such frequencies may, however, reach a distant point as a result of refraction (*i.e.*, bending) of energy earthward by the ionized region in the upper atmosphere. Finally, at very high frequencies the ground absorbs the portion of the wave traveling along the earth's surface, while the ionized regions are not capable of bending the wave path appreciably. Communication at these frequencies is hence possible only over distances so short that the earth's curvature permits a substantially straight-line path between transmitting and receiving points.

TABLE II.—CLASSIFICATION OF RADIO WAVES

Class	Frequency range, kilocycles	Wave-length range, meters	Outstanding characteristics	Principal uses
Low frequency....	Below 100	Over 3,000	Low attenuation at all times of day and of year	Long-distance trans-oceanic service requiring continuous operation
Medium frequency	100 to 1,500	3,000 to 200	Attenuation low at night and high in the daytime; greater in summer than winter	Range 100 to 500 kc used for marine communication, airplane radio, direction finding, etc. Range 550 to 1500 kc employed for broadcasting
Medium high frequency.	1,500 to 6,000	200 to 50	Attenuation low at night and moderate in the daytime	Moderate-distance communication of all types
High frequency...	6,000 to 30,000	50 to 10	Transmission depends solely upon the ionization in the upper atmosphere, and so varies greatly with the time of day and season. Attenuation extremely small under favorable conditions	Long-distance communication of all kinds; airplane radio
Very high frequency.	Above 30,000	Below 10	Waves travel in straight lines and are not reflected by ionized layers, so can only travel between points in sight of each other	Short-distance communication, television, two-way police radio, portable equipment, airplane landing beacons

As a result of the different propagation characteristics possessed by radio waves of different frequencies, each particular range of frequencies is best adapted for a particular type of communication service. The outstanding properties of the different classes of radio waves, as well as the uses for which each class has been found best suited, are tabulated in Table II.

The frequencies used in commercial radio communication range from about 15 kc as the lower limit to approximately 100,000 kc as an upper limit. Frequencies lower than about 15 kc are not used because of the difficulty of radiating energy at such low frequencies. Frequencies greater than 100,000 kc are difficult to generate, and so are not as yet used for commercial purposes, although extensive experimental work is being carried on at frequencies up to about 3,000,000 kc.

2. Radio Transmitters.—The essential components of a radio transmitter are a source of radio-frequency energy of the appropriate frequency, provision for controlling this energy in accordance with the information to be transmitted, and means for utilizing the energy to produce radio waves.

Any source of high-frequency energy can be used in a radio transmitter. During the history of radio various devices have been employed, such as the high-frequency alternator, the Poulsen arc, the oscillatory spark discharge, etc. However, all modern radio transmitters make use of vacuum tubes to produce the required power. This is because vacuum-tube oscillators and amplifiers are efficient, reliable, and very flexible. Over the range of frequencies used in commercial communication the power that can be obtained from vacuum tubes is of the order of tens to hundreds of kilowatts, with lesser powers obtainable at still higher frequencies.

Modulation.—The transmission of information by radio waves requires that means be provided by which the desired information can control the radio waves. In radio telegraphy this control is obtained by turning the transmitter on and off in accordance with the dots and dashes of the telegraph code, as illustrated in Fig. 2. In radio telephony the transmission is normally accomplished by varying the amplitude of the radio-frequency wave in accordance with the pressure of the sound wave being transmitted. Thus the sound wave shown at *d* in Fig. 2 would be transmitted from a radio-telephone station by causing the amplitude of the radiated wave to vary as shown at *e*. In picture transmission (including television) the amplitude of the wave radiated at any time is made proportional to the light intensity of the part of the picture that is being transmitted at that instant.

When the amplitude of the alternating-current wave is varied from time to time, the wave is said to be *modulated*. Thus the wave radiated from a radio-telephone station is modulated by sound waves, while during the transmission of a picture the modulation is in accordance with the light intensities of different portions of the picture, and in the case of radio telegraphy the modulation is by the telegraph code. Except in the case of telegraphy the modulation of the radio-frequency wave is usually accomplished by means of vacuum tubes that control the amplitude of the high-frequency energy in accordance with the information that is to be transmitted.

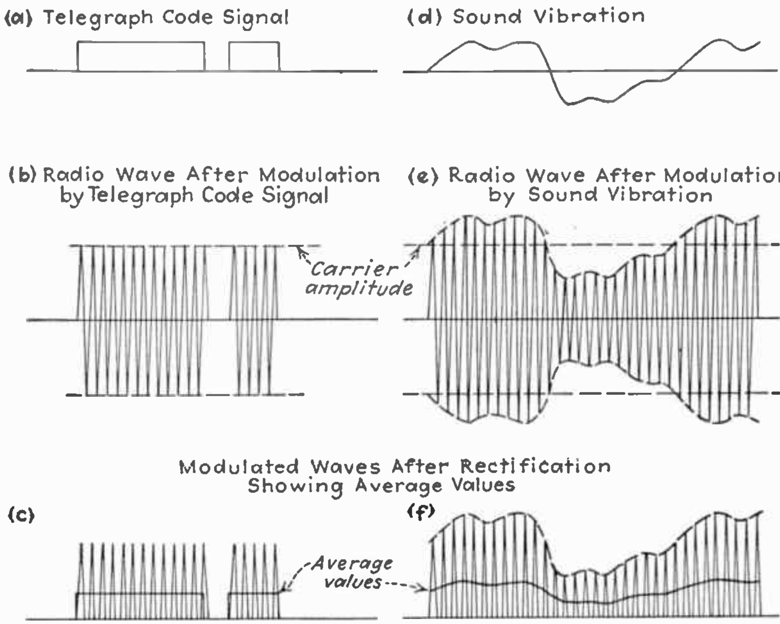


FIG. 2.—Diagram showing how a signal may be transmitted by modulating the amplitude of a radio wave and how the original signal may be recovered from the modulated wave by rectification.

3. Radiation of Electrical Energy.—Every electrical circuit carrying alternating current radiates a certain amount of electrical energy in the form of electromagnetic waves. The amount of energy thus radiated is extremely small, however, unless all the dimensions of the circuit approach the order of magnitude of a wave length. Thus a power line carrying 60-cycle current with 20-ft. spacing between conductors will radiate practically no energy because a wave length at 60 cycles is more than 3000 miles, and 20 ft. is negligible in comparison. On the other hand, a coil 20 ft. in diameter and carrying a 2000-kc current will radiate a considerable amount of energy because 20 ft. is comparable with the 150-meter wave length of the radio wave. The common radio antenna consisting of a system of overhead wires is essentially a condenser in which one plate is the ground while the other plate is the flat top. Such an arrangement will be a good radiator of electrical energy when the ratio of height to wave length is appreciable, that is, at least 1:100, and preferably greater.

It is apparent from these considerations that the size of radiator required is inversely proportional to the frequency (*i.e.*, directly proportional to the wave length). High-frequency waves can therefore be produced satisfactorily by a small radiator, while low-frequency waves require a high and large antenna system for effective radiation. The practical result of this situation is that the cost of a satisfactory antenna system increases as the frequency is lowered, and this sets a limit to the lowest frequency that it is practicable to employ.

4. Reception of Radio Signals.—In the reception of radio signals it is first necessary to abstract energy from the radio waves passing the receiving point. After this has been done, the radio receiver must separate the desired signal from other signals that may be present, and then reproduce the original information from the radio waves. In addition, arrangements are ordinarily provided for amplification of the received energy so that the output of the radio receiver can be greater than the energy abstracted from the wave.

Energy can be abstracted from a passing radio wave by means of an antenna system consisting of a wire oriented so that the magnetic flux of the wave cuts across the wire. The resulting induced voltage then acts against the impedance of the antenna circuit to produce a current. The energy represented by this induced current flowing in the antenna system is abstracted from the passing radio wave.

Since every wave passing the receiving antenna induces its own voltage in the antenna conductors it is necessary that the receiving equipment be capable of separating the desired signal from the unwanted signals that are also inducing voltages in the antenna. This separation is made on the basis of the difference in frequency between desired and undesired signals, and is carried out by the use of resonant circuits that are adjusted to discriminate very strongly in favor of the desired frequency. The ability to discriminate between radio waves of different frequencies is called *selectivity,* and the process of adjusting circuits to resonance with the frequency of a desired signal is spoken of as *tuning.*

Detection.—The process by which the information being transmitted is recovered from the radio-frequency currents present in the receiver is called *detection* (or *demodulation*). Where the information is transmitted by varying the amplitude of the radiated wave, detection is accomplished by rectifying the radio-frequency currents. The rectified current thus produced varies in accordance with the information originally modulated on the wave radiated from the transmitter. Thus when the modulated wave shown at e of Fig. 2 is rectified, the resulting current is shown at f and is seen to have an average value that varies in accordance with the amplitude of the original sound vibration. In the transmission of telegraph signals by radio, the rectified current reproduces the dots and dashes of the telegraph code as shown at Fig. 2c, and could be used to operate a telegraph sounder. When it is desired to receive the telegraph signals directly on a telephone receiver it is necessary to break up the dots and dashes at an audible rate in order to give a note that can be heard. Otherwise the telephone receiver would give forth a succession of unintelligible clicks.

Audio- and Radio-frequency Amplification.—The amount of energy that can be abstracted by the receiving antenna from a passing radio wave is ordinarily so small that it is customary to use amplification in the receiver. This amplification may take place before rectification, in which case it is termed *radio-frequency amplification,* or one may amplify the currents developed in the output of the detector, which is termed *audio-frequency amplification.* Ordinary radio receivers use both audio- and radio-frequency amplification.

The vacuum tube represents the only satisfactory method for amplifying the radio signals, and as a consequence is the basis of all modern radio receivers. Before the development of the vacuum tube it was necessary to depend entirely upon the energy absorbed from the passing waves by the receiving antenna. As a result of the small amplitude of this energy, the signals were always weak and radio reception from other than local stations was possible only in very quiet places.

5. Nature of a Modulated Wave.—The modulated wave that is sent out by most radio stations represents an oscillation of varying amplitude and so consists of a number of waves of different frequencies superimposed upon each other. The actual nature of a modulated wave can be deduced by writing down the equation of the wave and making a mathematical analysis of the result. Thus, in the case of the simple sine-wave modulations shown in Fig. 3, one has

$$e = E_0(1 + m \sin 2\pi f_s t) \sin 2\pi f t \tag{2}$$

where

e = instantaneous amplitude of wave
E_0 = average amplitude of envelope = carrier amplitude
m = degree of modulation
$= \dfrac{\text{crest alternating component of envelope variation}}{\text{average amplitude of envelope}}$
f_s = modulation frequency
= frequency at which envelope amplitude is varied
f = frequency of radio oscillation.

The degree of modulation m in Eq. (2) is sometimes expressed as a percentage. Thus $m = 0.50$ represents 50 per cent modulation.

Multiplying out the right-hand side of Eq. (2) gives

$$e = E_0 \sin 2\pi ft + mE_0 \sin 2\pi f_s t \sin 2\pi ft$$

By expanding the last term into functions of the sum and difference angles,[1] the equation of a wave with simple sine-wave modulation is

$$e = E_0 \sin 2\pi ft + \frac{mE_0}{2} \cos 2\pi(f - f_s)t - \frac{mE_0}{2} \cos 2\pi(f + f_s)t \qquad (3)$$

Equation (3) shows that the wave with sine-wave modulation actually consists of three separate waves. The first of these is represented by the term $E_0 \sin 2\pi ft$ and is called the carrier. Its amplitude is independent of the presence of modulation and is equal to the average amplitude of the wave. The two other components are alike as far as magnitude is concerned, but the frequency of one of them is less than the carrier frequency by an amount equal to the modulation frequency, while the frequency of the second is more than the carrier frequency by the same amount. These two components are called *side-band frequencies* and carry the information that is being transmitted by the modulated wave. The frequency by which the side bands differ from the carrier frequency represents the modulation frequency, whereas the amplitude of the side-band components compared with the amplitude of the carrier determines the degree of modulation, *i.e.*, the size of the amplitude variations that are impressed upon the radiated wave.

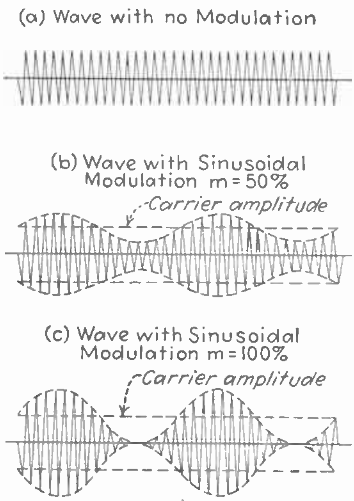


FIG. 3.—Waves having different degrees of simple sine-wave modulation.

When the modulation is more complex than the simple sine-wave amplitude variation of Fig. 3, the effect is to introduce additional side-band components. While the carrier wave is always the same, irrespective of the character of the modulation, there is a pair of side-band frequencies for each frequency component in the modulation. Thus, if the wave of a radio-telephone transmitter is modulated by a complex sound wave containing pitches of 1000 and 1500 cycles, the modulated

[1] The trigonometric formula for doing this is

$$\sin x \sin y = \tfrac{1}{2}[\cos (x - y) - \cos (x + y)]$$

wave will contain one pair of 1000-cycle side-band components and one pair of 1500-cycle side-band components. The amplitude of any side-band component is always one-half the amplitude of that particular frequency component which is contained in the modulation envelope.

Significance of the Side Bands.—The carrier and side-band frequencies are not a mathematical fiction, but have a real existence. This is evidenced by the fact that the various frequency components of a modulated wave can be separated from each other by suitable filter circuits.

The side-band frequencies may be considered as being generated as a result of varying the amplitude of the wave. They are present only when the amplitude is being varied, and their magnitude and frequency are determined by the character of the modulation. The carrier frequency, on the other hand, is independent of the modulation, being the same even when no modulation is present.

The information transmitted by the modulated wave is carried by the side-band components and not by the carrier, *i.e.*, the information is conveyed by the variations in the amplitude of the wave and not by the average amplitude. It is therefore desirable to put as much power into the side-band frequencies as is possible, which is equivalent to saying that the wave amplitude should be varied through the widest possible range. When the amplitude is carried completely to zero during the modulation cycle, the modulation is at a maximum, or 100 per cent, and the side bands contain the maximum amount of power possible. With sine-wave modulation such as shown in Fig. 3 this maximum side-band power is one-half the carrier power. With degrees of modulation less than 100 per cent the side bands will contain correspondingly less power.

It is apparent that the transmission of information requires the use of a band of frequencies rather than a single frequency. In speech and music there are important frequency components as high as 10,000 cycles, so that speech and music modulated upon a wave can produce side-band components extending to over 10,000 cycles on each side of the carrier frequency. A radio-telephone station therefore utilizes a frequency band about 20,000 cycles wide in transmitting high-quality signals. If this entire band is not transmitted equally well through space and by the circuits through which the modulated-wave currents must pass, then the side-band frequency components that are discriminated against will not be reproduced in the receiving equipment with proper amplitude. With telegraph signals the essential side bands are relatively narrow because the amplitude of the radio-frequency oscillation is varied only a few times a second, but a definite frequency band is still required. If some of the essential side-band components of the telegraph signal are not transmitted, the received dots and dashes tend to run together and become indistinguishable.

CHAPTER II

CIRCUIT ELEMENTS

6. Condensers.—A condenser consists of two electrodes separated by a dielectric. A voltage applied between these electrodes causes an electric charge to flow into them, with the resulting production in the dielectric of an electrostatic field that represents stored energy. The amount of energy stored in this way depends upon the voltage that is applied to the electrodes and upon the electrostatic capacitance (or capacity) of the combination. This capacitance is measured in farads, which is a capacitance of such a size that when 1 volt is applied between the electrodes a charge of 1 coulomb (equivalent to 1 amp. of current flowing for 1 sec.) will be stored. Because of its large size the farad is commonly divided into microfarads and micromicrofarads, abbreviated μf and $\mu\mu f$, respectively. Where the condenser electrodes are parallel plates having a constant spacing, the capacitance (neglecting edge effects) is given by the equation

$$\text{Capacitance in } \mu\mu\text{f} = 0.08842K\frac{A}{d} \tag{4}$$

where A is the area of the dielectric in square centimeters, d the spacing between plates in centimeters, and K a constant called the dielectric constant that depends upon the dielectric material and is substantially independent of frequency. Values of K for common dielectrics are given in Table III.

Losses in Condensers.—A perfect condenser when discharged gives up all the energy stored in it. This ideal is never completely realized, however. The result is that when an alternating voltage is applied to a condenser a certain amount of power is absorbed during each cycle. The power factor of the condenser current is a characteristic of the dielectric (assuming the resistance of the electrodes and leads is negligible) and is normally independent of the capacitance of the condenser, the frequency, or the applied voltage. While the power factor of the condenser is determined by the kind of dielectric employed, it is also affected by the conditions under which the dielectric operates. Thus the power factor always becomes higher as the temperature is raised and tends to be increased by absorbed moisture. Values of power factor for typical dielectrics are given in Table III.

TABLE III.—CHARACTERISTICS OF DIELECTRICS AT RADIO FREQUENCIES WITH NORMAL ROOM TEMPERATURE[1]

Material	Dielectric constant	Power factor
Air	1.00	0.000
Mica (electrical)	5 to 9	0.0001 to 0.0007
Hard rubber	3 to 5	0.006 to 0.014
Glass (electrical)	4.90 to 7.00	0.004 to 0.016
Bakelite derivatives	4.5 to 7.5	0.02 to 0.09
Celluloid	4.10	0.042
Fiber (dry)	4 to 6	0.02 to 0.09
Wood (without special preparation)		
Oak	3.3	0.039
Maple	4.4	0.033
Birch	5.2	0.065
Mycalex	8	0.002
Isolantite	6.1	0.0018
Porcelain (wet process)	6.5 to 7.0	0.006 to 0.008

[1] These data were compiled from various sources and represent typical values that can be expected.

The losses of a dielectric are sometimes expressed in terms of the angle by which the current flowing into the condenser fails to be 90° out of phase with the voltage. This angle is called the *phase angle* of the dielectric, and its value in radians is substantially equal to the power factor. Thus a power factor of 0.01 represents a phase angle of 0.01 radian or 0.573°.

The effect of the dielectric losses on the behavior of a circuit containing the condenser can be most conveniently expressed by considering the

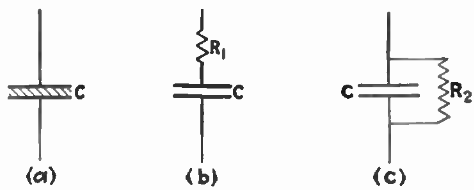


FIG. 4.—Representation of imperfect condenser by a perfect condenser of same capacitance, with series resistance, and by a perfect condenser with shunt resistance.

actual condenser to consist of a perfect condenser associated with a shunt resistance, as in Fig. 4*c*, or a series resistance as in Fig. 4*b*. The capacitance of the perfect condenser is made the same as the capacitance of the actual condenser, while the equivalent series or shunt resistance, as the case may be, is so selected that the power factor formed by the combination of perfect condenser and resistance is the same as the power factor of the actual condenser. The values of series and shunt resistances required are given by the following equations:

$$\text{Series resistance} = R_1 = \frac{\text{power factor}}{2\pi f C} \tag{5a}$$

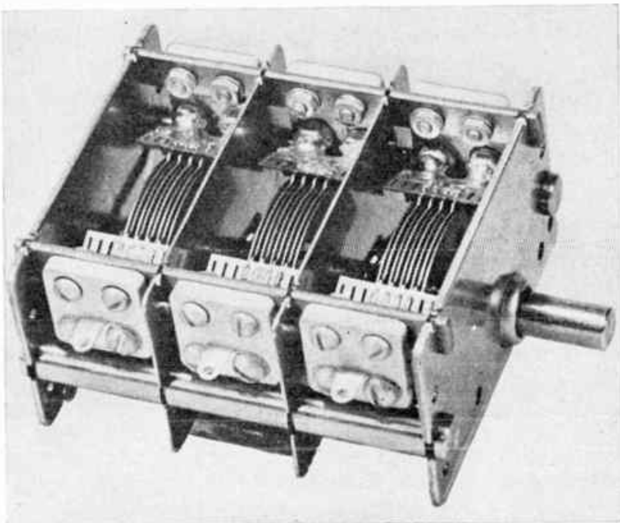
$$\text{Shunt resistance} = R_2 = \frac{1}{(2\pi f C) \times (\text{power factor})} \tag{5b}$$

Examination of these equations shows that the series and shunt resistances are inversely proportional to the frequency and the condenser capacitance.

Types of Condensers.—Paper, mica, and air are the dielectrics most frequently employed in condensers used in radio work. In addition, electrolytic condensers find use for certain purposes. Paper dielectric is inexpensive and gives a reasonably large capacitance in proportion to volume, but has the disadvantage of fairly high electrical losses. A good grade of mica has much lower losses than paper and so is particularly suitable for condensers where low losses are important, but is quite expensive. Photographs of typical paper and mica condensers are shown in Fig. 5.

Air is an almost perfect dielectric, having practically no losses if corona is avoided. The only losses of air condensers are then those arising from the resistance of the electrodes, and the dielectric losses in the mounting insulators. The result is a very low power factor when the condenser is properly made. The disadvantages of the air condenser are its low voltage rating and the large bulk required to obtain a reasonable capacitance. The chief use of air condensers is for tuning resonant circuits at radio frequencies, where a variable condenser having very low losses is required. Air condensers for this purpose employ a series of spaced fixed plates as one electrode and a series of similarly spaced rotating plates as the other electrode as shown in Fig. 5. The rotating plates mesh with the fixed plates without touching, thus giving a capacitance determined by the angle of rotation. Where several resonant circuits are to be tuned simultaneously, as in the case of broadcast receivers, it is customary to place all sets of rotating plates on a common shaft (see Fig. 5).

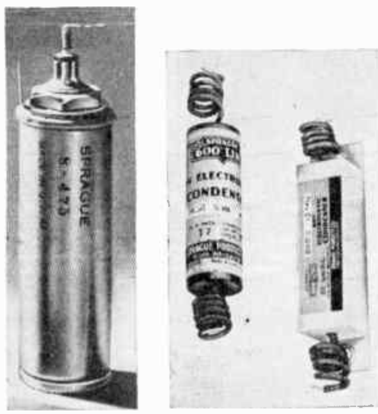
The electrolytic condenser makes use of the fact that aluminum (and certain other materials), when placed in a suitable solution and made the positive electrode, will form a thin insulating surface film that will withstand a considerable voltage and that at the same time will have a high electrostatic capacitance per unit area of film. Electrolytic condensers are characterized by a high capacitance in proportion to volume, and are the least expensive of all condensers. However, they have a much higher leakage current than do other condensers, and their power factor is high and dependent upon frequency. Furthermore, the capacitance, power factor, leakage, etc., change with age, applied voltage, and



Variable condenser

Small mica condenser

Large mica condenser



Electrolytic condensers

Paper condenser (cartridge type)

FIG. 5.—Photographs of various types of condensers commonly used in radio work.

temperature. Electrolytic condensers are widely used in radio work for filter and by-pass purposes. In these applications they are subjected to a direct-current voltage, and are then called upon to short-circuit any superimposed alternating potential. Under such conditions the losses and the exact capacitance are unimportant. Pictures of typical electrolytic condensers are given in Fig. 5.

Voltage Rating of Condensers.—The voltage that can be safely applied to condensers having solid dielectrics such as paper and mica depends upon the insulating strength of the dielectric used and upon the electrical losses in the dielectric. If the applied voltages exceed the dielectric strength, the dielectric will spark through and the condenser will be destroyed. However, if the losses in the condenser are sufficient to cause a moderate amount of heating, the allowable voltage will be something less than the dielectric strength of the material. This is because all dielectrics deteriorate rapidly when heated. As the condenser current and hence the losses are proportional to frequency, the voltage rating of a particular condenser will be highest on direct-current potentials, somewhat less at low frequencies, and increasingly lower as the frequency is increased. It is therefore very important that condensers which must withstand high radio-frequency voltages have low losses. The importance of losses in determining the voltage rating is illustrated by the ratings of a particular 0.001 μf mica condenser given in Table IV.

TABLE IV

Frequency, Kc	Rated Effective Voltage
Direct current	15,000
1	10,000
100	5000
300	3000
1000	1780
3000	605
10,000	178

The voltage limit of an air condenser is determined by either corona or spark-over, whichever appears first. When high voltages are to be handled, the plates must be widely spaced, and it is helpful to round off sharp corners in order to reduce the tendency for corona to form. Air condensers for high voltage service are very bulky in proportion to capacitance and so are used only when small capacitances are required.

The voltage limit of an electrolytic condenser depends upon the thickness of the insulating film at the positive aluminum electrode, and may be as high as 500 volts. If the allowable voltage is exceeded the film will puncture, but reforms when the excess voltage is removed. Electrolytic condensers are consequently not destroyed by momentary overvoltages as are paper and mica condensers.

7. Inductance.—Whenever a current flows in an electrical circuit, there is produced magnetic flux that links with (*i.e.*, encircles) the current. The number of flux linkages actually present with a given current is measured in terms of a property of the circuit called the inductance, and depends upon the arrangement of the circuit and the presence or absence of magnetic substances.

Inductance can be defined as the flux linkages per ampere of current producing the flux. That is,

$$\text{Inductance } L \text{ in henries} = \frac{\text{flux linkages}}{\text{current producing flux}} 10^{-8} \qquad (6)$$

A flux linkage represents one flux line encircling the circuit current once. Thus in Fig. 6 flux line *aa* contributes eight flux linkages toward the coil inductance because it circles the current flowing in the coil eight times. On the other hand, flux line *b* of the same coil contributes only one-half a flux linkage toward the coil inductance because this particular line encircles only one-half the coil current.

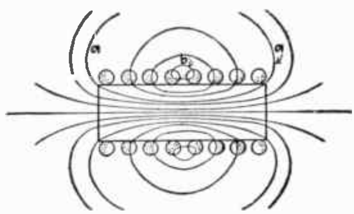


FIG. 6.—Flux and current distribution in a typical single-layer air-cored inductance coil. The current density at radio frequencies is indicated by the depth of shading.

The inductance of a coil is proportional to the square of the number of turns if the dimensions such as length, diameter, depth of winding, etc., are kept constant as the number of turns is altered. The reason for this behavior lies in the fact that, if the coil dimensions are kept constant, the amount of magnetic flux produced by a given coil current and the number of times each flux line links with the coil current are both proportional to the number of turns, so that the flux linkages per ampere are proportional to the square of the number of turns.

The inductance of coils having the same number of turns and the same shape is always proportional to the size (*i.e.*, to a linear dimension, such as length or radius) of the coil. Thus, if two coils have the same number of turns but one is twice as big as the other in every dimension (such as diameter, length, width, depth of winding, etc.), then the larger coil will have twice the inductance of the smaller one.

Formulas for calculating the inductance of coils are found in every radio or electrical handbook.

Incremental Inductance.—In many communication circuits there are iron-core coils that carry a large direct current upon which is superimposed a small alternating current. The inductance that such a coil offers to the superimposed alternating magnetism is termed the *incremental inductance*, *i.e.*, the inductance to an increment of magnetization.

Experiment shows that the incremental inductance increases as the superimposed alternating magnetization is increased up to the point where saturation begins, whereas increasing the direct-current magnetization reduces the incremental inductance. This is illustrated in Fig. 7.

Magnetic Materials.—Cores for such communication apparatus as audio-frequency transformers, power transformers, and filter reactors are most commonly of silicon-steel laminations. Special grades of such laminations have been developed which have lower losses and higher incremental permeability than ordinary silicon steel. In addition, a number of alloys having unusual magnetic properties are coming into increasing use in communication work. Chief among these are a series

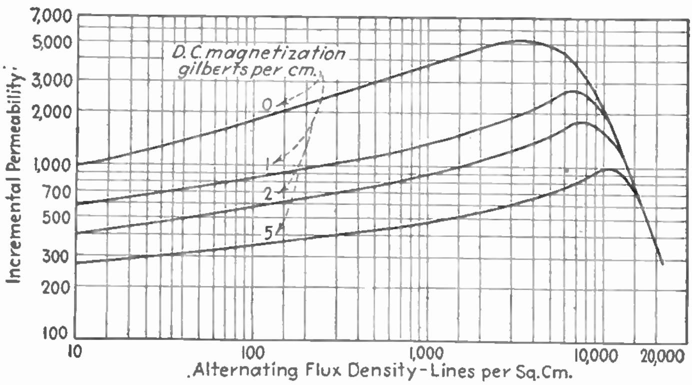


FIG. 7.—Curve giving incremental permeability of typical silicon steel as a function of alternating-current magnetization superimposed upon several values of direct-current magnetization, showing how the inductance decreases with increased direct-current and reduced alternating-current magnetization.

of nickel-iron alloys characterized by extremely high permeability and low hysteresis loss at low flux densities. Different compositions go under such names as permalloy, mu-metal, hypernik, nicoloi (or electric metal), etc. A related nickel-iron-cobalt alloy called *perminvar*, which has a moderately high permeability at low flux densities, is remarkable for a permeability that is practically constant for magnetomotive forces up to about two ampere turns per centimeter, coupled with an extremely low hysteresis loss. A wide variety of other alloys are available for other special uses, such as permanent magnets, high permeability at high flux densities, etc.

8. Mutual Inductance and Coefficient of Coupling.—When two inductance coils are so placed in relation to each other that flux lines produced by current in one of the coils link with the turns of the other coil as shown in Fig. 8*a*, the two inductances are said to be inductively coupled. The effects that this coupling produces can be expressed in terms of a property called the *mutual inductance*, that is defined by the relation:

$$\text{Mutual inductance } M \text{ in henries} = \frac{\begin{pmatrix}\text{flux linkages in second}\\ \text{coil produced by current}\\ \text{in first coil}\end{pmatrix}}{\text{current in first coil}} 10^{-8} \quad (7a)$$

$$= \frac{\begin{pmatrix}\text{flux linkages in first}\\ \text{coil produced by current}\\ \text{in second coil}\end{pmatrix}}{\text{current in second coil}} 10^{-8} \quad (7b)$$

Formulas (7*a*) and (7*b*) are equivalent and give the same value of mutual inductance. The flux linkages produced in the coil that has no current

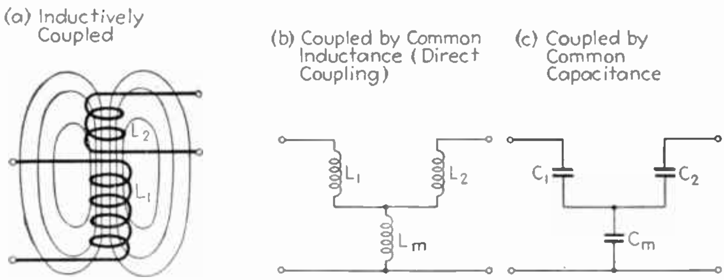


FIG. 8.—Several simple methods of coupling two circuits.

in it are counted just as though there was a current in this coil. Hence the number of times a flux line would encircle an imaginary coil current is the number of linkages contributed by this particular line.

When two coils of inductance L_1 and L_2, between which a mutual inductance M exists, are connected in series, the equivalent inductance of the combination is $L_1 + L_2 \pm 2M$. The term $2M$ takes into account the flux linkages in each coil due to the current in the other coil. These mutual linkages may add to or subtract from the self-linkages, depending upon the relative direction in which the current passes through the two coils. The maximum value of mutual inductance that can be obtained between two coils having inductances L_1 and L_2 is $\sqrt{L_1L_2}$. The ratio of mutual inductance actually present to this maximum possible value is called the *coefficient of coupling*. Hence

$$\text{Coefficient of coupling} = k = \frac{M}{\sqrt{L_1L_2}} \quad (8)$$

The coefficient of coupling is a convenient constant because it expresses the extent to which the two inductances are coupled, independently of the size of the inductances concerned. In air-core coils, as used in radio, a coupling coefficient of 0.5 is considered high and is said to represent "close" coupling, while coefficients of only a few hundredths represent "loose" coupling.

Any two circuits so arranged that energy can be transferred from one to the other are said to be coupled, even though this transfer of energy takes place by some means such as a condenser, resistance, or inductance common to the two circuits, rather than by the aid of a mutual inductance. Examples of various methods of coupling are shown in Fig. 8. *Any two circuits that are coupled by a common impedance have a coefficient of coupling that is equal to the ratio of the common impedance to the square root of the product of the total impedances of the same kind as the coupling impedances that are present in the two circuits.* That is,

$$\text{Coefficient of coupling} = k = \frac{Z_m}{\sqrt{Z_1 Z_2}} \tag{9}$$

where Z_m is the common impedance and Z_1 and Z_2 are the total impedances of the same kind as Z_m in the primary and secondary circuits respectively. Thus with case *b* in Fig. 8, where the coupling is furnished by the common inductance L_m, the total inductances of the two circuits are $L_1 + L_m$ and $L_2 + L_m$, and the coefficient of coupling is given by the equation

$$\text{Coefficient of coupling } k \text{ for Fig. } 8b = \frac{L_m}{\sqrt{(L_1 + L_m)(L_2 + L_m)}} \tag{10}$$

9. The Effective Resistance of Coils and Conductors at Radio Frequencies.—The effective resistance offered by conductors to radio frequencies is considerably more than the resistance measured with direct currents. This is the result of *skin effect*, which causes the current to be concentrated in certain parts of the conductor and leaves the remainder of the conductor to do little or nothing toward carrying the current. *As a result of this effect it is necessary to generalize the concept of conductor resistance when dealing with radio frequencies by considering the resistance to be that quantity which when multiplied by the square of the current will give the energy dissipated in the conductor.*

A simple example of skin effect is furnished by an isolated round wire. When a current is flowing through such a conductor, the magnetic flux that results is in the form of concentric circles as shown in Fig. 9. It is to be noted that some of this flux exists within the conductor and therefore links with, *i.e.*, encircles, current near the center of the conductor while not linking with current flowing near the surface. The result is that the inductance of the central part of the conductor is greater than the inductance of the part of the conductor near the surface. At radio frequencies the reactance of this extra inductance is sufficiently great to affect seriously the flow of current. Most of the current then flows along the surface of the conductor where the impedance is low, rather than near the center where the impedance is high. The center part of the conductor therefore does not carry its share of the current and the effective resistance is increased. The actual type of current distribution obtained in the case of a round wire is as shown in Fig. 9.

Skin effect occurs whenever some parts of a conductor have more flux linkages than other parts. The result of skin effect is to cause a redistribution of current over the conductor cross section of such a character that most of the current flows where it is encircled by the smallest number of flux lines. This is because those parts encircled by the fewest flux lines have the lowest inductance, and hence offer the least impedance to the current.

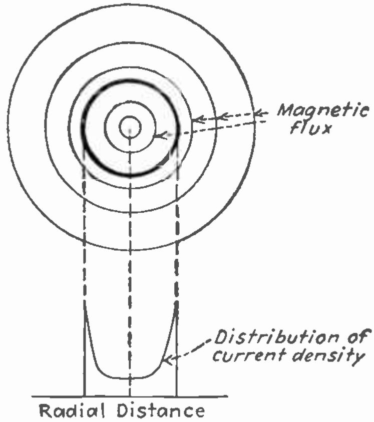


FIG. 9.—Isolated round conductor, showing flux paths and current distribution at radio frequencies. Note that the current density is highest for parts of the conductor encircled by the fewest flux lines.

The ratio that the effective alternating-current resistance bears to the direct-current resistance of a conductor increases with frequency and with the conductivity of the conductor material. This is because the higher frequency increases the reactance produced by the extra flux linkages, whereas a higher conductivity causes slight differences of inductance for different parts of the conductor to have more effect on the current distribution.

The non-uniformity of current distribution that results from skin effect can be reduced by employing a conductor consisting of a large number of small insulated wires connected in parallel at the terminals and thoroughly interwoven. If the interweaving is complete each conductor will on the average link with the same number of flux lines as every other conductor. This will give all the individual strands substantially the same inductance, and will therefore cause the current to distribute uniformly between the strands. A stranded cable of this type is called a *litz* conductor and is very effective in reducing the radio-frequency resistance at broadcast and lower frequencies.

Resistance of Coils at High Frequencies.—The same principles that govern the current distribution in an isolated conductor also determine the distribution of current in the conductors of a coil. That is, the current density is greatest in those parts of each coil conductor encircled by the smallest number of flux lines. The skin effect in coils is, however, much more complicated and much greater in magnitude than in isolated conductors because each turn of the coil produces flux that causes non-uniformity of current distribution in adjacent turns. The nature of the current distribution in the conductors of a typical radio coil is indicated by the shading in Fig. 6. This also shows the flux paths and brings out the relation between current density and flux linkages.

The losses in a coil are most conveniently expressed in terms of the ratio of the coil reactance ωL to the effective coil resistance R. This ratio approximates the reciprocal of the coil power factor. It is so important

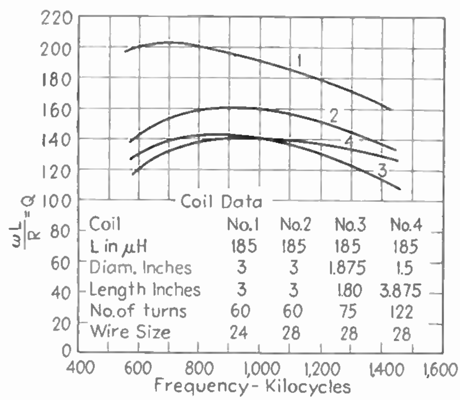


FIG. 10.—Curves showing Q as a function of frequency for a number of coils having the same inductance but different dimensions, different proportions, and different wire sizes.

in the theory of resonant circuits as to be considered a fundamental coil property, and is usually referred to by the symbol Q. That is,

$$Q = \frac{\text{coil reactance}}{\text{coil resistance}} = \frac{\omega L}{R} \tag{11}$$

The effective coil resistance R includes any dielectric loss that the coil may have.

It is convenient to express the characteristics of a coil in terms of the ratio Q, because this quantity is approximately constant for the same coil over its usual working range of frequencies. Furthermore, the value of Q for equally well-designed coils intended for use at different frequencies is approximately the same. That is, a value of Q denoting an efficient coil of a size suitable for use at 100 kc also represents the Q of an efficient coil for 1000-kc service. The values of Q actually obtained with coils used in receiving equipment range from thirty to several hundred, with still greater values frequently encountered in transmitter inductances.

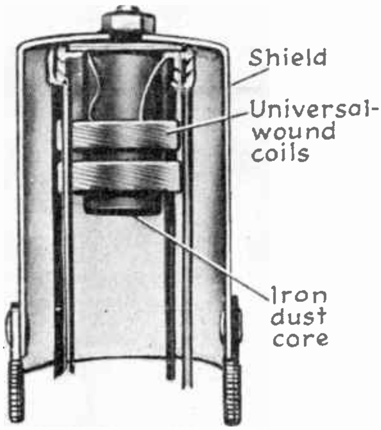
The actual value of Q depends in a complicated way upon the coil construction and the frequency. However, there are certain general trends that can be expected to hold. Thus the larger the physical size of the coil the higher tends to be the Q. The Q varies much more slowly than does the size, however, so very small coils are still reasonably efficient. Also, for any given inductance and physical size there is a best wire diameter that is often, although not always, the largest conductor that can be wound in the available space. For any given inductance and volume there is also a best shape, which for single layer solenoids is usually when the length of the coil is about the same as the diameter. The arrangement of the winding affects the Q, and it is often found that multilayer constructions are superior to single-layer coils having the same physical dimensions and inductance. The type of conductor is also important, as indicated below. The influence of some of these factors on coil Q is illustrated in Fig. 10.

10. Coils for Radio-frequency Circuits. *Coils for Tuned Circuits of Radio Receivers.*—Coils used in the resonant circuits of radio receivers must have low losses (high Q), must be physically small, and must have low distributed capacitance. The type of coil used to meet these requirements depends upon the frequency. Above 1500 kc, single-layer solenoids space wound with solid wire on a thin form are customary. At broadcast frequencies both single-layer solenoids and multilayer coils are used. Solid wire is employed where cost is an important factor although litz wire properly used will usually result in an increased value of Q at broadcast frequencies.

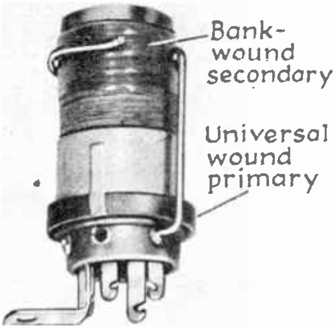
At frequencies below the broadcast range (below 550 kc) multilayer coils are best. Both solid and litz conductors are used, although litz is decidedly preferable in this frequency range. Coils for frequencies below 500 kc are often provided with magnetic cores of the type described below.

Typical coils used in commercial radio receivers are shown in Fig. 11.

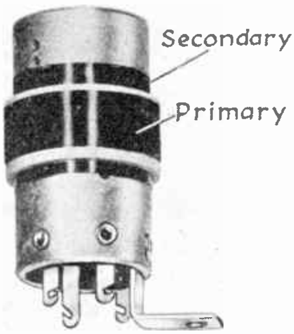
Coils for Tuned Circuits of Transmitters.—Coils used in the resonant circuits of transmitters differ from the corresponding coils for receivers in that the current that must be carried and the voltage that is developed across the terminals are much greater. The constructional details depend upon the frequency and the power involved. In general, however, the coils are provided with spaced turns in order to withstand the voltage satisfactorily, and ordinarily employ large conductors in order to carry the heavy currents with low losses. At high radio frequencies, where only a small inductance is required, solid wire space wound on a single-layer ceramic form is typical for low and medium powers, while self-supporting coils constructed from copper tubing are used for high power. Coils for broadcast frequencies generally employ either single-layer solenoid or pancake construction, and use solid wire, tubing, strip,



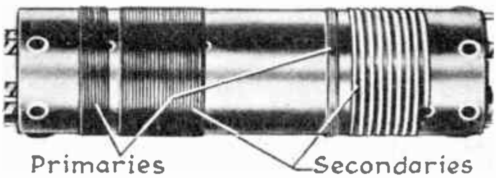
Universal-wound coil on iron dust core



Bank-wound coil

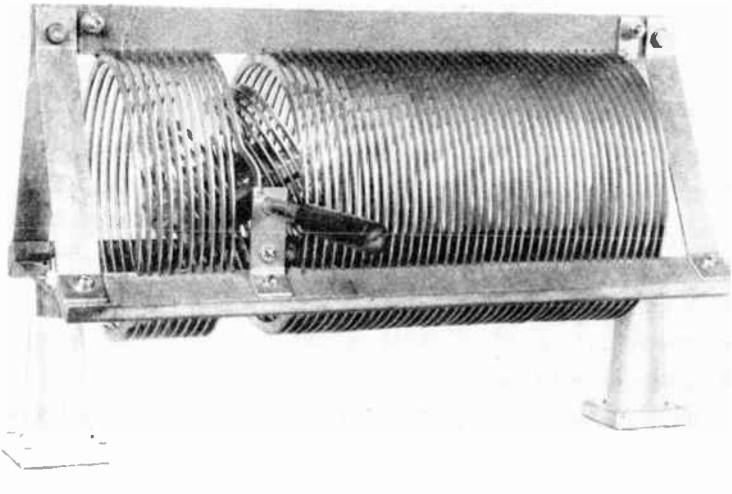


Single-layer solenoid

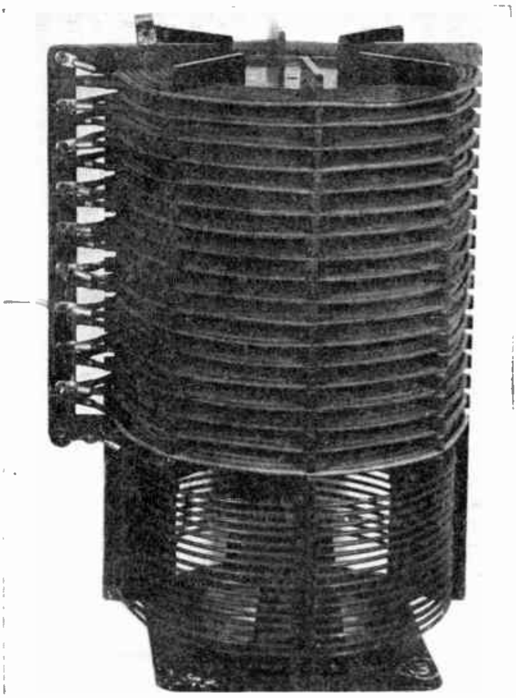


Two band coil assembly (single-layer solenoid)

FIG. 11.—Photographs of typical coils used in tuned circuits of radio receivers.



Broadcast transmitter inductance (edgewise-wound strip)



Large inductance for long-wave transmitter (multi-layer litz coil)

Short-wave coil (solid wire on ceramic form)

Short-wave coil (copper tubing)

FIG. 12.—Photographs of typical coils used in tuned circuits of radio transmitters.

or edgewise-wound strip. Coils for lower radio frequencies are usually of multilayer construction, preferably using litz wire.

Typical transmitter coils are shown in Fig. 12.

Distributed Capacitance of Coils with Particular Reference to Multilayer Coils.—In an inductance coil there are capacitances between adjacent turns, capacitances between turns that are not adjacent, and capacitances between terminal leads. In addition there can be capacitances to ground from each turn. Some of the different capacitances that may exist in a typical coil are shown in Fig. 13. Each of these capacitances stores a quantity of electrostatic energy that is determined by the capacitance and the fraction of the total coil voltage that appears across the turns involved. The total effect that the numerous small coil capacitances have can ordinarily be represented to a high degree of accuracy by assuming that these many capacitances can be replaced by a single condenser of appropriate size shunted across the coil terminals. This equivalent capacitance is called the *distributed capacitance* of the coil.

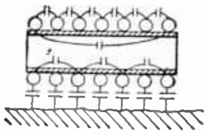


FIG. 13.—Some of the coil capacitances that contribute to the distributed capacitance of a single-layer coil.

In multilayer coils the distributed capacitance tends to be high because the layer arrangement of the winding causes turns from different parts of the winding to be located near each other. Thus in the two-layer winding shown at Fig. 14*a*, in which the turns are numbered in order, the first and last turns are adjacent, which greatly increases the shunting capacitance. The distributed capacitance of multilayer coils can be kept low by expedients such as illustrated in Fig. 14. Here the turns associated with the two ends of the coil are kept apart by a narrow winding form as at *c* and *d*, or by arranging the sequence of turns to form a bank winding as illustrated at *b* (also see Fig. 11). The "universal" (or honeycomb) coil illustrated in Figs. 11 and 15 is also often used for multilayer coils. This is a self-supporting winding with relatively few turns per layer, arranged so that the turns of adjacent layers have appreciable air separation except at the cross-over points.

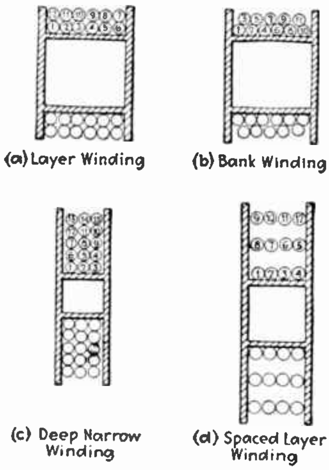


FIG. 14.—Several types of multilayer windings.

Radio-frequency Choke Coils.—The term *radio-frequency choke* is applied to coils that are designed to provide a high impedance over an appreciable frequency band. Such coils are used in receiver and transmitter circuits to carry direct current while preventing the flow of radio-frequency currents. Radio-frequency choke coils must have a high inductance so that very little radio-frequency current will flow through the winding. At the same time the coil must have low distributed capacitance so that there will be a minimum of radio-frequency current by-passed around the coil. Typical methods of constructing radio-frequency choke coils are illustrated in Fig. 15. The resistance of a

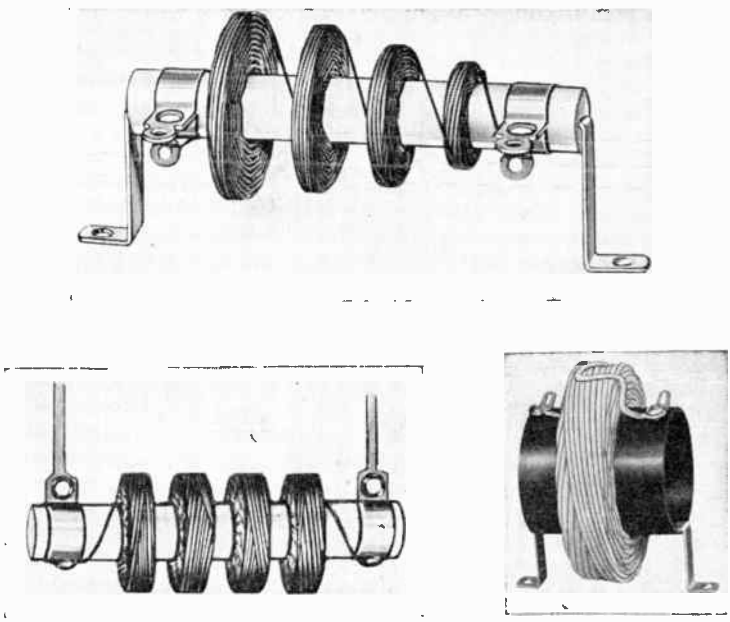


FIG. 15.—Photographs of different types of radio-frequency choke coils. All of these are of the "universal-wound" type.

radio-frequency choke coil is relatively unimportant, and the wire size is chosen on the basis of current-carrying capacity.

Radio-frequency Coils Having Magnetic Cores.—The usefulness of magnetic materials at high frequencies is limited by eddy-current losses. It has been found, however, that by suitably subdividing the core material the eddy-current losses can be reduced to low values even at radio frequencies. The required degree of subdivision is obtained by producing the magnetic material as a very fine dust that is mixed with an insulating binder and then formed under pressure to the desired shape. Coils employing such cores have reasonably low losses up to frequencies of 1000 kc, and are used to a considerable extent at frequencies below

500 kc. Such cores make it possible to reduce the physical dimensions while maintaining the coil Q at a reasonable value.

11. Shielding of Magnetic and Electrostatic Fields.—Circumstances often arise where it is necessary to confine electrostatic and magnetic fields to a limited space. Thus, in radio receivers having several coils associated with different tuned circuits, it is commonly necessary to shield these coils from each other to eliminate couplings that would transfer energy between the tuned circuits.

Electrostatic fields can be confined to a limited space by enclosing the space with a conducting shield upon which electrostatic flux lines terminate. If the shield material does not have excessive electrical resistivity the flux lines are virtually short-circuited, and the shielding is practically perfect.

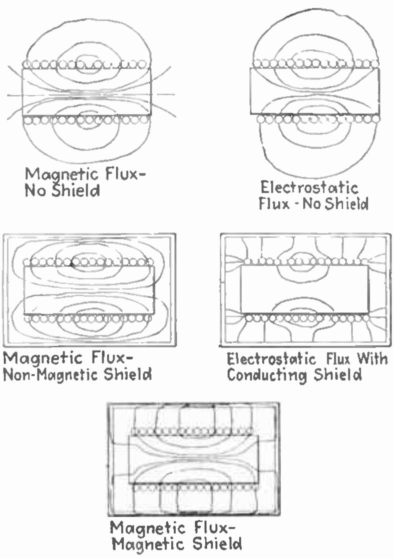


FIG. 16.—Paths of electrostatic and magnetic flux lines about the same coil with and without magnetic and non-magnetic shields.

High-frequency magnetic fields are ordinarily controlled by conducting shields. When alternating magnetic flux attempts to pass through such a shield, the magnetic flux induces a voltage in cutting through the conductor. This voltage produces eddy currents which in large measure prevent the magnetic flux from penetrating through the shield. If the resistance of the conducting shield to the eddy currents is made low by employing a thick shield of low-resistance material, the shielding is practically perfect.

Coils used for tuned circuits in radio receivers are commonly placed in aluminum or copper shield cans to obtain both magnetic and electrostatic shielding. This reduces the inductance of the coil by restricting the flux paths to the space within the shield as shown in Fig. 16 and thereby increasing the reluctance of the magnetic circuit. The shield likewise increases the effective coil resistance because the energy losses represented by the eddy currents in the shield are supplied by the coil. However, if the shield is large enough so that the spacing between coil and shield is everywhere at least equal to the coil radius, these effects on the coil properties are not serious.

Low frequency and direct-current magnetic fields require the use of a shield of magnetic material. Conducting shields are not effective because the induced voltage produced by the flux is zero in the case of direct fields and small at low frequencies. A magnetic shield acts as a short circuit to the flux lines as a result of the fact that its magnetic permeability is higher than that of the surrounding space. This is illustrated in Fig. 16. In order to be effective, magnetic shields must have high permeability at low flux densities and so should be made of permalloy or other high permeability alloy.

Problems

1. A variable condenser with air dielectric is to have a total capacitance of 150 $\mu\mu$f. If the spacing between plates is 0.07 cm, calculate the total area of dielectric required. If the rotating plates are semicircular with a radius of 3 cm, calculate the number of rotating plates required, assuming that both sides of every plate are active. Neglect the loss in area caused by the shaft.

2. A condenser employs a dielectric consisting of a glass plate having a thickness of 0.15 cm, a dielectric constant of 5, and a power factor of 0.008. Calculate the area of active dielectric required to obtain a capacitance of 0.002 μf.

3. In the condenser of Prob. 2, calculate:

a. The phase angle of the dielectric.

b. The equivalent series resistance at 1,000,000 cycles.

c. The power lost in the condenser when a potential of 250 volts effective of frequency 1,000,000 cycles is applied to the condenser.

4. A mica condenser with power factor 0.0005 has a capacitance of 0.001 μf.

a. What are its equivalent series and shunt resistances at frequencies of 1000, 100,000, and 10,000,000 cycles?

b. What are its phase angles at these same frequencies?

5. Derive an exact formula for the power factor of a condenser in terms of the phase angle, and demonstrate that when the phase angle is small the power factor is substantially equal to the phase angle expressed in radians.

6. A certain mica condenser is found by measurements at 500,000 cycles to have a series resistance of 0.35 ohm, and a capacitance of 510 $\mu\mu$f.

a. What is the power factor of the dielectric?

b. What would the series resistance be at 1,200,000 cycles?

7. A certain variable condenser having air dielectric with bakelite supports obtains 10 $\mu\mu$f of its capacitance through the bakelite dielectric having a power factor of 4 per cent and the remainder of its capacitance from the air, which has no losses.

a. What is the equivalent shunt resistance of the combination when the total capacitance is 100 $\mu\mu$f (90 $\mu\mu$f from air and 10 $\mu\mu$f from bakelite) and the frequency is 750,000 cycles?

b. What is the equivalent series resistance under the same conditions?

8. The condenser of Prob. 4 is able to stand an effective alternating-current potential of 2000 volts at very low frequencies without sparking through and is able to dissipate ½ watt of heat safely.

a. What is the highest frequency at which it is permissible to apply a potential of 2000 volts effective, assuming that the only losses are dielectric losses?

b. If the only losses are dielectric losses (electrode resistance neglected) what is the voltage rating at frequencies of 1, 100, and 10,000 kilocycles?

9. If in Fig. 6 the flux shown is produced by a current of 0.2 amp., estimate the coil inductance. (Assume that Fig. 6 gives a two-dimensional representation of the flux lines in the three-dimensional coil.)

10. Coil 1 in Fig. 10 is a single-layer solenoid with a 60-turn winding 3 in. long and 3 in. in diameter. How many turns would be required to obtain the same inductance if the coil were 2 in. in diameter and 2 in. long?

11. In an iron-core coil it is found that the introduction of a small air gap in the magnetic circuit of the core will increase the incremental inductance when there is appreciable direct-current magnetization present, but will reduce the incremental inductance if the direct-current magnetization is zero or very small. Explain.

12. Suggest a method by which mutual inductance might be measured, using a bridge capable of measuring only inductance.

13. Demonstrate that two coils which have their axes respectively parallel to and at right angles with the line joining the coil centers will have zero mutual inductance.

14. Derive a formula for the coefficient of coupling for the circuit of Fig. 8*c*.

15. Explain how skin effect makes the inductance of an isolated conductor slightly less at radio frequencies than at low frequencies.

16. Derive an equation giving the exact relation between the Q of a coil and the coil power factor, and from this calculate the error in the approximate relation: Power factor $= 1/Q$, when $Q = 50$.

17. In Fig. 10, calculate and plot: (*a*) the resistance as a function of frequency for the No. 24 wire case; (*b*) the ratio of alternating-current to direct-current resistance for the same coil.

18. Explain why the distributed capacitance of a coil is always increased by the wax or other coating used for protection against moisture.

19. Explain the reason that moisture in the insulation is very detrimental to the properties of a coil.

20. In transmitter inductances explain why copper tubing gives a resistance just as low as solid copper conductor of the same diameter.

21. Transmitter inductances wound with bare copper tubing are sometimes plated to prevent corrosion and improve the appearance. Discuss the relationship between resistivity and thickness of the plated layer and the radio-frequency resistance of the coil.

22. It is observed in a non-magnetic shield surrounding a solenoidal coil that the effectiveness of the shielding is not affected appreciably by a joint in the shield, provided this joint is in a plane perpendicular to the axis of the coil, but is very seriously reduced if the joint is in a plane that contains the axis of the coil. Explain.

23. Discuss the effect upon the distributed capacitance of a coil of: (*a*) a conducting shield, (*b*) a magnetic shield.

24. When low-frequency fields are shielded by a magnetic shield, what effect does the presence of the shield have on: (*a*) the coil inductance, (*b*) the effective coil resistance?

CHAPTER III

RESONANT CIRCUITS AND CIRCUIT ANALYSIS

12. Series Resonance.—When a constant voltage of varying frequency is applied to a circuit consisting of an inductance, a capacitance, and a resistance all in series, the current varies with frequency in the manner shown in Fig. 17. At low frequencies the reactance of the condenser is large and the reactance of the inductance small. The circuit

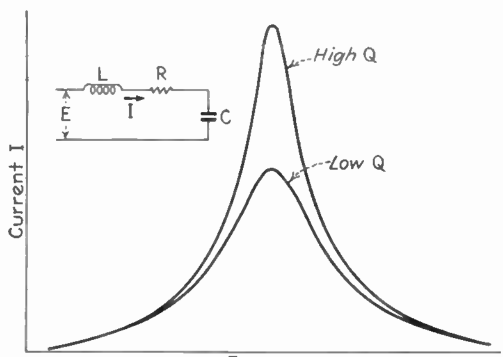


FIG. 17.—Series resonant circuit and typical resonance curves for different values of circuit Q.

then has a high impedance, and only a small current flows. Similarly at high frequencies, the reactance of the inductance is high and the reactance of the capacitance small, so that the impedance is again large and the current small. In between these extremes there is a frequency, called the *resonant frequency,* at which the inductive and capacitive reactances are exactly equal, and being of opposite sign consequently neutralize each other. At this resonant frequency there is accordingly only the resistance to oppose the flow of current, and if the resistance is low the current at resonance will be large.

The resonant frequency is the frequency at which the inductive and capacitive reactances are equal, and so is given by the relation

$$2\pi fL = \frac{1}{2\pi fC} \tag{12a}$$

or

$$\text{Resonant frequency} = \frac{1}{2\pi\sqrt{LC}} \tag{12b}$$

Equation (12*b*) shows that the resonant frequency depends only upon the product LC.

Analysis of Resonance Curves.—Curves such as shown in Fig. 17, giving the variation of current with frequency, are called resonance curves. The shape of these curves can be calculated by the usual method of analyzing alternating-current circuits. This analysis will now be carried out using the following symbols:

E = voltage applied to circuit
I = magnitude of current flowing in circuit
f = frequency in cycles
$\omega = 2\pi f$
$Q = \omega L/R$
R = effective series resistance of tuned circuit, including both coil resistance and equivalent series resistance of condenser.
L = inductance in henries
C = capacitance in farads
z = magnitude of impedance of series circuit
θ = phase angle of impedance of series circuit
$Z = z\angle\theta$ = vector impedance of circuit

The reactance formed by the inductance and capacitance of the series circuit is $\left(\omega L - \frac{1}{\omega C}\right)$. The impedance formed by this reactance in series with the resistance R of the circuit is therefore

$$\text{Circuit impedance } z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{13a}$$

$$\text{Phase angle } \theta \text{ of impedance} = \tan^{-1} \frac{\left(\omega L - \frac{1}{\omega C}\right)}{R} \tag{13b}$$

The current I that flows is the voltage divided by the impedance, or

$$I = \frac{E}{z} = \frac{E}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}} \tag{14}$$

This current lags behind the voltage by the angle θ given by Eq. (13*b*). For the special case of resonance, $\omega L = 1/\omega C$, and Eqs. (13*a*) and (14) become

$$\text{Series impedance at resonance} = R \tag{15a}$$

$$\text{Current at resonance} = \frac{E}{R} \tag{15b}$$

The current at resonance is in phase with the applied voltage because the impedance is resistive.

Shape of Resonance Curves. Universal Resonance Curve.—The sharpness of the resonance curve in the immediate vicinity of resonance is determined by the Q of the resonant circuit.[1] This is apparent from Fig. 17, and can be explained as follows. At resonance the current is inversely proportional to resistance, and so directly proportional to the circuit Q. At the same time, the current that flows for frequencies differing appreciably from the resonant frequency is substantially independent of the circuit resistance. This is because at such frequencies the reactance of the circuit is greater than the resistance and so controls the circuit impedance. The result is that increasing the circuit resistance (*i.e.*, lowering Q) flattens the resonance curve by reducing the response at resonance without affecting the response at frequencies appreciably different from resonance. This action is apparent in the curves of Fig. 17.

The resonance curve of a series circuit is for all practical purposes symmetrical about the resonant frequency, provided the circuit Q is at least moderately high.

The shape of the resonance curve can be conveniently expressed in terms of the universal resonance curve of Fig. 18. This curve gives the ratio of actual current to current at resonance, in terms of the circuit Q and the fractional deviation of the frequency from the resonant frequency. The universal resonance curve can be calculated from Eqs. (12), (14), and (15*b*).[2] It makes no approximations except the assumption that the circuit Q is constant over the range of frequencies involved. The universal resonance curve is particularly useful because it applies to all resonant circuits irrespective of circuit constants.

The use of the universal resonance curve in practical calculations is illustrated by the following examples:

Example 1.—It is desired to know how many cycles one must be off resonance to reduce the current to one-half the value at resonance when the circuit has a Q of 125 and is resonant at 1000 kc.

Reference to Fig. 18 shows that the response is reduced to 0.5 when $a = 0.86$. Hence

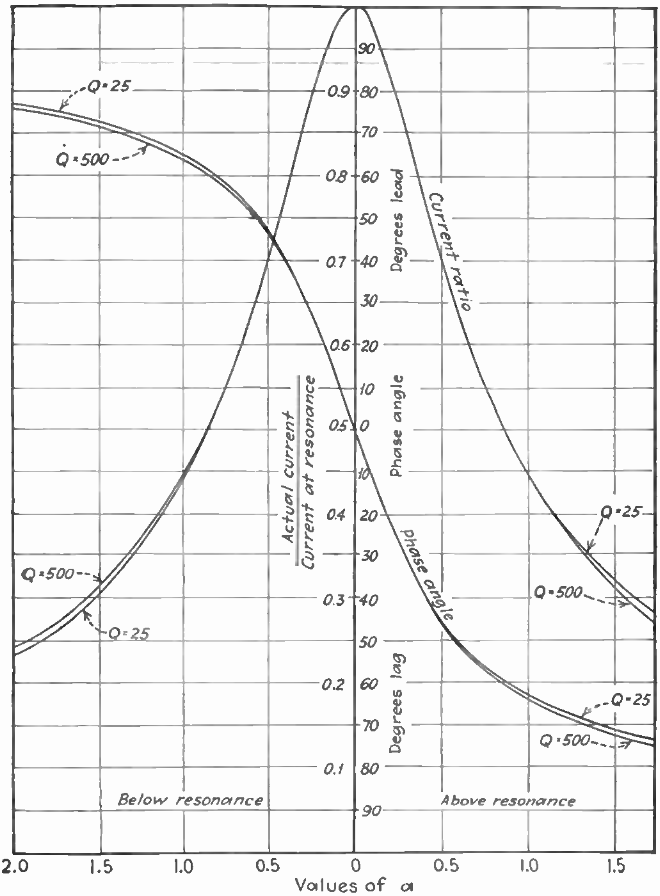
$$\text{Cycles off resonance} = \frac{0.86 \times 1000}{125} = 6.88 \text{ kc}$$

The phase angle of the current as obtained from the curve is 60° lagging, or leading, as the case may be.

[1] The term *Q of the circuit* used in this connection means $\omega L/R$, where R includes the series resistance of the condenser as well as the coil resistance. The circuit Q will accordingly be less than the coil Q, but the difference is normally small because condenser losses are ordinarily much less than coil losses.

[2] The actual derivation is given on p. 54 of the author's book "Radio Engineering," 2d ed.

Example 2.—With the same circuit as in the preceding example, it is desired to know what the response will be at a frequency 10,000 cycles below resonance.



$$\frac{\text{Cycles off resonance}}{\text{Resonant frequency}} = \frac{a}{Q} = \text{Fractional detuning}$$

$$\text{Cycles off resonance} = \frac{a}{Q} \times \text{resonant frequency}$$

$$a = Q\,\frac{\text{Cycles off resonance}}{\text{Resonant frequency}}$$

FIG. 18.—Universal resonance curve from which the exact ratio of actual current to current at resonance, as well as exact phase angle, can be determined for any series circuit in terms of the detuning from resonance. This curve can also be applied to the parallel resonant circuit by considering the vertical scale to represent the ratio of actual parallel impedance to the parallel impedance at resonance. When applied to parallel circuits, angles shown in the figure as leading are lagging, and vice versa.

To solve this problem it is first necessary to determine a.

$$a = (10/1000) \times 125 = 1.25$$

Reference to Fig. 18 shows that for $a = 1.25$ the response is reduced by a factor 0.36 and that the phase of the current is 68° leading.

The sharpness of the resonance curve in any particular case can be readily estimated by making use of the following rules that can be deduced from the equations of the resonant circuit.

Rule 1. *When the frequency of the applied voltage deviates from the resonant frequency by an amount that is* $1/2Q$ *times the resonant frequency, the current that flows is reduced to* 70.7 *per cent times the resonant current, and the current is* 45° *out of phase with the applied voltage.*

Rule 2. *When the frequency of the applied voltage deviates from the resonant frequency by an amount that is* $1/Q$ *of the resonant frequency, the current that flows is reduced to* 44.7 *per cent of the resonant current, and the current is* 63½° *out of phase with the applied voltage.*

Thus in the circuit considered in the above examples, the current would be reduced to 70.7 per cent of the value at resonance when the frequency is $\frac{1}{250}$ of 1000 kc, or 4000 cycles off resonance, and to 44.7 per cent of the resonant current for a frequency deviation of $\frac{1}{125}$ of 1000 kc, or 8000 cycles.

Voltage across Reactance of a Series Circuit.—The voltage developed across the inductance or capacitance of a series resonant circuit is equal to the product of current and the inductive (or capacitive) reactance. The resulting curve of voltage across the reactance as a function of frequency for constant applied voltage has substantially the same shape as the resonance curve. This is because the resonance phenomenon takes place in such a limited frequency range that the reactance across which the voltage is developed is substantially constant in the region about resonance. The product of current and reactance hence varies with frequency in practically the same manner as the current.

The voltages across the condenser and the inductance will both be very much greater than the applied voltage when the frequency is in the vicinity of resonance. This situation is possible because the voltages across the condenser and inductance are nearly 180° out of phase with each other and so add up to a value that is much smaller than either voltage alone. Since the current at resonance is E/R, the voltage across the inductance at resonance, which is ωL times the current, is

$$\text{Voltage across } L \text{ at resonance} = E\frac{\omega L}{R} = EQ \tag{16}$$

The voltage across the condenser also has this same value, since at resonance, $\omega L = 1/\omega C$. Equation (16) shows that *at resonance the voltage across the inductance (or condenser) is Q times the applied voltage (i.e., there is a resonant rise of voltage in the circuit amounting to Q times).* Since Q can be expected to have a value in the neighborhood of 100, a series resonant circuit will develop a high voltage even with small applied potentials.

Practical Calculation of Resonance Curves.—In the calculation of resonance curves, the use of Eqs. (13) and (14) involves practical difficulties. This is because the inductive and capacitive reactances ωL and $1/\omega C$, respectively, are both large, but in the vicinity of resonance very nearly equal to each other. Consequently any slight error in calculating either reactance, such as would result from the use of a slide rule, introduces a very large error in their difference, and hence in the calculated impedance and current. In view of this, the most satisfactory procedure for determining resonance curves is as follows: *First*, calculate the response at resonance by Eq. (15*b*) or (16). *Second*, use the working rules given above, together with the universal resonance curve, to obtain points on the resonance curve in the immediate vicinity of resonance. *Finally*, neglect the circuit resistance when calculating the response at frequencies farther off resonance than the range covered by the universal resonance curve. This procedure gives results at least as accurate as obtainable with four-place logarithms and involves much less work than using a slide rule.

13. Parallel Resonance.—When a resonant circuit is arranged so that the voltage is applied to the inductance and capacitance in parallel, the resulting impedance varies with frequency in the manner shown in Fig. 19. At very low frequencies the inductive branch draws a large lagging current while the leading current of the capacitive branch is small, resulting in a large line current and a low circuit impedance that is inductive. At high frequencies the inductance has a high reactance compared with the capacitance, resulting again in a large line current and a correspondingly low circuit impedance that is now capacitive. In between these two extremes there is a frequency at which the lagging current taken by the inductive branch and the leading current entering the capacitive branch are substantially equal, and, being nearly 180° out of phase, these currents tend to neutralize each other. The line current is then quite small, and the circuit impedance is high, as apparent in Fig. 19. This resonant impedance is much greater than the impedance of either the inductive or capacitive branch of the circuit, and represents one of the most valuable properties of the parallel resonant circuit.

The shape of the impedance curve of the parallel resonant circuit is practically identical with the shape of the resonance curve of the same circuit arranged for series resonance, provided only that the Q of the circuit is not too low. With values of Q ordinarily used in radio work the approximation is extremely close. The resonant frequency is hence the same for a given circuit irrespective of whether the connections are such as to give series or parallel resonance. The sharpness of the impedance curve also depends similarly upon the Q of the circuit. In particular, a low Q broadens the curve by reducing the impedance in the immediate vicinity of resonance while having very little effect on the impedance at frequencies appreciably different from resonance, as seen in Fig. 19.

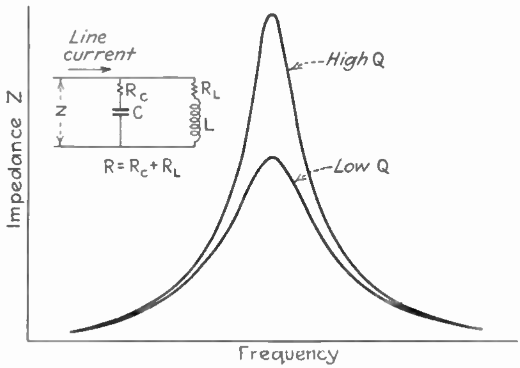


FIG. 19.—Parallel resonant circuit and typical curves of parallel impedance for different values of circuit Q. Note the similarity to Fig. 17.

Analysis of Parallel Resonance.—The following notation will be used in the analysis of parallel resonant circuits.

E = voltage applied to parallel circuit

$z_c = \sqrt{R_c^2 + (1/\omega C)^2}$ = magnitude of impedance of capacitive branch

$\theta_c = -\tan^{-1}\dfrac{(1/\omega C)}{R_c}$ = phase angle of capacitive branch

$Z_c = z_c\angle\theta_c$ = vector impedance of capacitive branch

$z_L = \sqrt{R_L^2 + (\omega L)^2}$ = magnitude of impedance of inductive branch

$\theta_L = \tan^{-1}\dfrac{\omega L}{R_L}$ = phase angle of inductive branch

$Z_L = z_L\angle\theta_L$ = vector impedance of inductive branch

$Z_s = Z_L + Z_c$ = series impedance formed by capacitive and inductive branches in series (vector value)

ω = 2π times frequency

ω_0 = 2π times resonant frequency

$Q = \omega L/R$

$R = R_c + R_L$ = total resistance of circuit considered as a series circuit

Line current = current flowing into circuit as result of the applied voltage E.

The quantities L, C, R_L, and R_c refer to the inductance, capacitance, and resistance components of the circuit as indicated in Fig. 19.

The resonant frequency of a parallel circuit is normally considered to be the frequency at which $\omega L = 1/\omega C$. With this definition the parallel resonant frequency of a tuned circuit is exactly the same as the series resonant frequency of the same circuit, and so is given by Eq. (12*b*).

The impedance of the parallel circuit is the impedance formed by the capacitive and inductive branches connected in parallel:[1]

$$\text{Parallel impedance} = \frac{Z_c Z_L}{Z_c + Z_L} = \frac{Z_c Z_L}{Z_s} \qquad (17)$$

It will be noted that the denominator of this expression is the impedance that the same circuit offers when connected in series, as given by Eqs. (13*a*) and (13*b*). Equation (17) is the fundamental equation of the parallel circuit and applies for all conditions, irrespective of the circuit Q, the frequency, or the division of resistance between the branches.

When the circuit Q is reasonably high, as is nearly always the case in actual practice, the exact expression of Eq. (17) can be simplified by neglecting the resistance components of the impedances Z_L and Z_c in the numerator. When this is done,[2]

$$\text{Parallel impedance} = \frac{(\omega_0 L)^2}{Z_s} \qquad (18)$$

At resonance the series impedance is simply the total resistance R. The parallel impedance then becomes a resistance having a magnitude

$$\left.\begin{array}{l}\text{Parallel impedance}\\ \text{at resonance}\end{array}\right\} = \frac{(\omega_0 L)^2}{R_s} = \omega_0 L Q \qquad (19)$$

The error involved in Eqs. (18) and (19) depends upon the Q of the circuit and the division of resistance between the branches. Under the most unfavorable condition, which is when all the resistance is in one branch, the magnitude of the parallel impedance as calculated by Eq. (19) has an error of 1 part in $2Q^2$, while the error in phase angle is $\tan^{-1}(1/Q)$. These errors are entirely negligible for ordinary calculations if the circuit Q is at all high. Thus for $Q = 100$ the error in magnitude is $\frac{1}{200}$ of 1 per cent, and the error in phase is 0.6°.

[1] Equation (17) is merely the mathematical statement of the rule that the impedance formed by two parallel branches is the product of the branch impedances divided by the sum of the branch impedances.

[2] This transformation is carried out as follows: If the resistance components are neglected, the product $Z_L Z_c$ becomes $\omega L/\omega C = L/C$. One can now eliminate the capacitance C in this expression by multiplying both numerator and denominator by ω_0 and then noting that $1/\omega_0 C = \omega_0 L$. That is,

$$Z_L Z_c = \frac{L}{C} = \frac{\omega_0 L}{\omega_0 C} = (\omega_0 L)^2$$

Examination of Eq. (18) shows that the impedance of a parallel circuit is equal to a constant divided by the series impedance of the same circuit. A consideration of Eq. (14) shows that the current in a series circuit is likewise equal to a constant (*i.e.*, the voltage) divided by the series impedance of the circuit. *Hence the resonance curve of impedance in a parallel circuit has exactly the same shape as the resonance curve of current for the same circuit connected for series resonance.* Consequently the universal resonance curve and the working rules that were given for estimating the sharpness of resonance of the series circuit also apply to the case of parallel resonance. The only difference is that the signs of the phase angles are now reversed, the phase of the current now being leading at frequencies higher than resonance and lagging at frequencies below resonance.

The line current that flows when a voltage is applied to a parallel circuit is relatively small at the resonant frequency because of the high circuit impedance. However, as the frequency departs from resonance

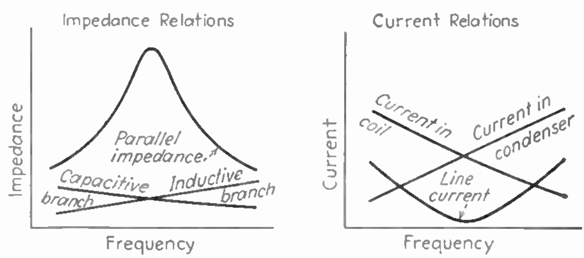


FIG. 20.—Curves showing line and branch currents and circuit impedance of a parallel circuit in the vicinity of resonance. The line current is not zero, where the branch currents have equal magnitude, because of the phase shifts produced by the circuit resistance.

the current becomes rapidly larger as the result of the lowered circuit impedance. The currents in the individual branches do not go through any resonance action, however. This is because the current in a branch is equal to the applied voltage divided by the branch impedance, and this impedance varies only slowly with frequency. The resonance phenomenon arises from the fact that at the resonant frequency both capacitive and inductive branches draw large currents, but these currents are largely reactive and add up to a very small resultant. This means that at resonance there is a large current circulating between the inductance and capacitance, with the line current being just large enough to supply the circuit losses. Inasmuch as the resistance of a tuned circuit is low, the energy losses and hence the line current will be correspondingly small at resonance. At frequencies off resonance, the line current increases because the reactive currents drawn by the capacitive and inductive branches are not equal, which makes it necessary for the line to supply considerable reactive current along with the power component of current.

These relations between line and branch currents are illustrated in Fig. 20.

Practical Calculation of Parallel Impedance Curves.—The proper procedure for calculating the impedance of a parallel resonant circuit is similar to that given above for calculating the resonance curve of a series circuit. The detailed steps are: *First*, calculate the impedance at resonance by Eq. (19). *Second*, use the working rules together with the universal resonance curve, to obtain points on the impedance curve in the immediate vicinity of resonance, keeping in mind that the sign of the phase angle of the line current for parallel impedance is reversed from the sign for series resonance. *Finally*, neglect the circuit resistance when calculating magnitude of impedance at frequencies farther off resonance than the range covered by the universal resonance curves.

Miscellaneous.—When the circuit Q is low, the approximations made in deriving Eq. (18) will introduce an error that depends on the division of resistance between the inductive and capacitive branches. If the resistances are divided so that $\omega L/R_L$ and $1/\omega CR_c$ are approximately equal, the error is at a minimum. The parallel impedance then very accurately follows the behavior discussed above. If, on the other hand, the resistances in the two branches are decidedly unequal, as is commonly the case since the losses are usually concentrated in the inductive branch, then Eq. (18) and the discussion based upon it are only approximately true, as mentioned above. The principal modification produced under such conditions is that the frequency at which the parallel circuit has maximum impedance is not the frequency for which it has unity power factor. With the resistance largely concentrated in one branch this behavior begins to be apparent when the circuit Q drops below 10. The exact behavior for low circuit Q's can be calculated from Eq. (17) by using the exact expressions for Z_L and Z_c in determining the numerator.

The distributed capacitance associated with an inductance coil is in shunt with the coil inductance and thus makes the coil equivalent to a parallel resonant circuit. The result is that the impedance of the coil as observed at the terminals varies as shown in Fig. 19 in the region where the inductance is in resonance with the distributed capacitance. In particular, *at frequencies greater than resonance the coil has a capacitive reactance and hence acts as a condenser.* At frequencies below resonance the presence of the distributed capacitance causes the apparent inductance and resistance of the coil, as measured between coil terminals, to be greater than the true values that would be obtained in the absence of distributed capacitance.

14. Inductively Coupled Circuits. *Theory.*—When mutual inductance exists between coils that are in separate circuits, these circuits are said to be inductively coupled. The effect of the mutual inductance is to make possible the transfer of energy from one circuit to the other by transformer action. That is, an alternating current flowing in one circuit produces magnetic flux which induces a voltage in the coupled circuit, resulting in induced currents and a transfer of energy from the first or primary circuit to the coupled or secondary circuit.

The behavior of inductively coupled circuits is somewhat complex, but can be readily calculated with the aid of the following rules:

Rule 1. *As far as the primary circuit is concerned, the effect of the presence of the coupled secondary circuit is exactly as though an impedance* $(\omega M)^2/Z_s$ *had been added in series with the primary,*[1] *where*

M = mutual inductance

$\omega = 2\pi f$

Z_s = series impedance of secondary circuit when considered by itself (vector value).

Rule 2. *The voltage induced in the secondary circuit by a primary current of* I_p *has a magnitude of* $\omega M I_p$ *and is* 90° *out of phase with the primary current.*

Rule 3. *The secondary current is exactly the same current that would flow if the induced voltage were applied in series with the secondary and the primary were absent.*[2]

Calculations of Coupled Circuits.—The three rules given above for determining the behavior of coupled circuits hold for all frequencies and for all types of primary and secondary circuits, both tuned and untuned. The procedure to follow in using these rules to compute the behavior of a coupled circuit is: *first,* determine the primary current with the aid of Rule 1; *second,* compute the voltage induced in the secondary, knowing the primary current and using Rule 2; and *finally,* calculate the secondary current from the induced voltage by means of Rule 3. The following set of formulas will enable these operations to be carried out systematically.

[1] This can be demonstrated as follows. The actual effect produced by the presence of the secondary is a back voltage induced in the primary by the secondary current. This secondary current is a result of the voltage induced in the secondary by the primary current I_p. The secondary induced voltage is $\omega M I_p$, and is 90° out of phase with I_p. The resulting secondary current is $\omega M I_p/Z_s$. This current then induces a voltage $[(\omega M)I_p/Z_s]\omega M = [(\omega M)^2/Z_s]I_p$ in the primary. This voltage is $90 + 90 = 180°$ out of phase with I_p and hence is a back voltage. The voltage drop corresponding to the back voltage is hence $(\omega M)^2/Z_s$ times the primary current, so that the effect of the secondary on the primary is equivalent to adding an impedance $(\omega M)^2/Z_s$ to the primary.

[2] It might be wondered why, although the secondary couples an impedance into the primary, the primary is not considered as coupling an impedance into the secondary. The reason is that the impedance coupled into the primary takes into account the voltage induced in the primary by the secondary current, as explained in the preceding footnote. The corresponding effect of the voltage induced in the secondary by the primary current is taken care of by Rule 3.

$$\left.\begin{array}{l}\text{Impedance coupled into primary circuit}\\ \text{by presence of the secondary}\end{array}\right\} = \frac{(\omega M)^2}{Z_s} \tag{20}$$

$$\text{Equivalent primary impedance} = Z_p + \frac{(\omega M)^2}{Z_s} \tag{21}$$

$$\text{Primary current} = I_p = \frac{E}{Z_p + \dfrac{(\omega M)^2}{Z_s}} \tag{22}$$

$$\text{Voltage induced in secondary} = \omega M I_p \underline{/\pm 90^\circ} \tag{23}$$

$$\text{Secondary current} = \frac{\text{Induced voltage}}{Z_s}$$

$$= \frac{\omega M I_p \underline{/\pm 90^\circ}}{Z_s} = \frac{\omega M E \underline{/\pm 90^\circ}}{Z_p Z_s + (\omega M)^2} \tag{24}$$

In these equations

M = mutual inductance between primary and secondary

$\omega = 2\pi f$

Z_s = series impedance of secondary circuit considered as though primary were removed (a vector)

Z_p = series impedance of primary circuit considered as though secondary were removed (a vector)

E = applied voltage.

The $\underline{/\pm 90^\circ}$ indicates either a lead or lag of 90° according to the sense of the mutual inductance. In making use of Eqs. (20) to (24) it is to be kept in mind that Z_p and Z_s are vector quantities.

Coupled Impedance.—The impedance $(\omega M)^2/Z_s$ that the presence of the secondary adds to the primary circuit is called the *coupled impedance.* This quantity is a vector having resistance and reactance components that are given by the expressions:

$$\left.\begin{array}{l}\text{Resistance component}\\ \text{of coupled impedance}\end{array}\right\} = \frac{(\omega M)^2 R_s}{R_s^2 + X_s^2} \tag{25a}$$

$$\left.\begin{array}{l}\text{Reactance component}\\ \text{of coupled impedance}\end{array}\right\} = -\frac{(\omega M)^2 X_s}{R_s^2 + X_s^2} \tag{25b}$$

Here R_s and X_s are the resistance and reactance components of the secondary impedance Z_s, with a positive reactance being inductive. The energy represented by the primary current flowing through the coupled resistance is the energy that is transferred to the secondary circuit. Similarly, the volt-amperes represented by the primary current flowing through the reactive component of the coupled impedance represents the reactive volt-amperes consumed by the secondary circuit.

Many of the important properties of a coupled circuit can be deduced by examining the nature of the coupled impedance. When the mutual inductance M is small, or when the secondary impedance Z_s is large, the impedance coupled into the primary by the presence of the secondary is small. The action in the primary circuit is then influenced only very slightly by the presence of the secondary. On the other hand, when the mutual inductance is large and the secondary impedance small, the coupled impedance will be large. The presence of the secondary will then have a very pronounced effect upon the behavior of the primary circuit. A particularly important case is when the secondary is a resonant circuit. Here the secondary impedance is small at resonance and comparatively large at frequencies appreciably different from resonance. Since the coupled impedance is inversely proportional to the secondary impedance, the effect produced by such a secondary on the primary circuit is large in the immediate vicinity of resonance, and small at other frequencies.

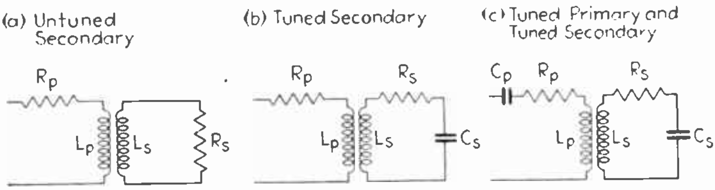


FIG. 21.—Common types of inductively coupled circuits.

15. Typical Examples of Inductively Coupled Circuits.—Some of the most commonly encountered and more important types of inductively coupled circuits, together with their principal properties, are considered below.

Coupled Circuit with Untuned Secondary Consisting of a Resistance and Inductance.—This circuit is illustrated in Fig. 21*a*, and is important because a metal mass such as a shield, metal panel, etc., near a coil acts as a secondary consisting of an inductance in series with a resistance. The impedance that such a secondary couples into the primary is seen from Eq. (25) to consist of a resistance and a capacitive reactance. The coupled resistance takes into account the energy losses in the metal mass, and increases the effective resistance of the primary coil. The coupled reactance neutralizes a portion of the inductive reactance of the coil, and hence is equivalent to reducing the effective inductance of the coil.[1]

Coupled Circuits with Untuned Primary and Tuned Secondary.—This circuit is shown on Fig. 21*b* and is of importance because it is the equivalent circuit of the tuned radio-frequency amplifier. As the circuit is ordinarily encountered the primary resistance R_p represents the plate resistance of a tube and is much larger than the reactance ωL_p of the

[1] It will be noted that these are the same effects arrived at by a different method in Sec. 11 for the case of a conducting shield surrounding a coil.

primary inductance. The most important characteristic of this circuit is the way in which the secondary current (or the voltage across the secondary condenser) varies with frequency for constant applied voltage. The behavior in a typical case is shown in Fig. 22 and is seen to result in a curve having a shape similar to that of a resonance curve. Close examination, however, shows that the Q corresponding to the curve of secondary current is lower than the actual Q of the secondary circuit.

Two Resonant Circuits Tuned to the Same Frequency and Coupled Together.—This circuit is illustrated in Fig. 21*c* and is the basic circuit of the band-pass filter commonly used in radio receivers. The most important characteristic of such a circuit is the variation of secondary

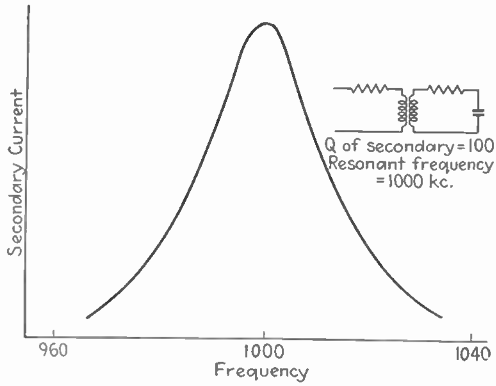


FIG. 22.—Curve of secondary current as a function of frequency for the circuit of Fig. 21*b* in the typical case where $R_p \gg \omega L_p$. This curve has the shape of a resonance curve corresponding to a Q lower than the actual secondary Q.

current (or voltage across the secondary condenser) as a function of frequency when a constant voltage is applied in series with the primary.

Typical results are shown in Fig. 23. When the coefficient of coupling k between primary and secondary is low the secondary current is quite small and the curve is very peaked ($k = 0.002$ in Fig. 23). As the coefficient of coupling is increased, the secondary current becomes greater and there is some reduction in the relative sharpness of the peak ($k = 0.005$ in Fig. 23). This effect continues with increase in coupling until the resistance that the secondary couples into the primary circuit at resonance is equal to the resistance of the primary. The coefficient of coupling required for this condition is termed the *critical coupling* and gives the maximum secondary current that it is possible to obtain. With critical coupling the curve of secondary current also tends to be flat at the very top with relatively steep sides ($k = 0.01$ in Fig. 23). When the coefficient of coupling is appreciably greater than the critical value, the secondary current at resonance becomes less than with critical coupling, and peaks of secondary current appear at frequencies on either side of resonance ($k = 0.015$, 0.03, and 0.05 in Fig. 23). The spacing between these peaks of secondary current is directly proportional to the coefficient of coupling. The current at these peaks is substantially the same as the peak current with critical coupling.

The reason for the behavior illustrated in Fig. 23 rests in the nature of the coupled impedance produced by a resonant secondary circuit. When the coupling is small the induced voltage and hence the secondary current will naturally be small. At the same time the shape of the secondary-current curve will be sharper than the resonance curve of the secondary circuit because of the fact that the induced voltage tends to be proportional to primary current and is therefore greatest at resonance.

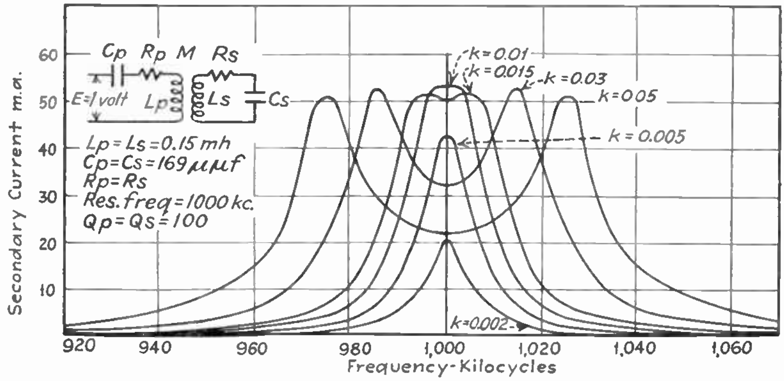


Fig. 23.—Curves for two circuits separately tuned to the same frequency and coupled together, showing variation of secondary current with frequency, for a number of coefficients of coupling.

On the other hand, when the coefficient of coupling is large, the coupled impedance at resonance is likewise large. This reduces the primary current at resonance to a small value and results in a low induced voltage and hence low secondary current. The humps that appear on the response curve with large coupling result from the fact that, with a tuned secondary circuit, the reactance coupled into the primary is inductive at frequencies below resonance and capacitive at frequencies above resonance. Since the sign of this coupled reactance is the opposite of the sign of the reactance of the primary circuit at the same frequencies, the presence of the secondary reduces the equivalent impedance that the primary offers to the applied voltage. This raises the primary current and hence increases the induced voltage at frequencies slightly off resonance. When the coupling exceeds the critical value this action is sufficient to introduce new resonant frequencies corresponding to the humps of secondary current.

The critical coefficient of coupling corresponds to the condition which gives the maximum possible transfer of energy to the secondary at the resonant frequency. This occurs when the coupled resistance at resonance equals the resistance of the primary circuit, *i.e.*, when $(\omega M)^2/R_s = R_p$. Hence for critical coupling

$$\omega M = \sqrt{R_p R_s} \tag{26a}$$

Introducing the ratios $Q_p = \omega L_p/R_p$, and $Q_s = \omega L_s/R_s$, and noting that $k = M/\sqrt{L_p L_s}$, gives

$$\text{Critical coupling } k = \frac{1}{\sqrt{Q_p Q_s}} \tag{26b}$$

The critical coefficient of coupling is usually very small. Thus if the circuit Q's are 100, the corresponding coefficient is only 0.01.

16. Band-pass Filters.—When two tuned circuits that are resonant at the same frequency are coupled with a coefficient of coupling slightly greater than the critical value, the secondary current (and hence the voltage across the secondary condenser) is substantially constant in the limited frequency range around the resonant frequency, but falls off very rapidly for frequencies appreciably removed from resonance. Such a characteristic is exhibited by the curve for $k = 0.015$ in Fig. 23. When this result is obtained one is said to have a *band-pass* filter. Such a band-pass characteristic is particularly desirable in handling modulated waves because it can be designed to give substantially the same response to all side-band frequencies contained in the wave. In contrast with this, ordinary resonant circuits have a rounded top and so discriminate against the higher side-band frequencies.

The most important properties of a band-pass filter are the width of the band of frequencies that is transmitted, and the uniformity of the response within this band. The width is given approximately by the relation

$$\frac{\text{Width of pass band}}{\text{Resonant frequency of tuned circuits}} = k \tag{27}$$

The uniformity of response within the pass band of frequencies depends upon the circuit Q in relation to the coefficient of coupling. If the circuit Q's are too high, pronounced double humps appear, and if too low the top of the response curve is rounded off instead of being flat (see Fig. 24). Experiments indicate that the best value of Q is about 50 per cent greater than that corresponding to critical coupling. Hence for uniform transmission within the pass band

$$\sqrt{Q_p Q_s} = \frac{1.5}{k} \tag{28}$$

where Q_p and Q_s are the Q's for the primary and secondary circuits, respectively.

In the design of band-pass filters, the proper procedure is first to select, with the aid of Eq. (27), a coefficient of coupling that will give the required band width. After this the appropriate circuit Q's for uniform transmission within the pass band are determined with the aid of Eq. (28).

From an examination of Eqs. (27) and (28) it is seen that, when a wide pass band is required, the coefficient of coupling must be large and

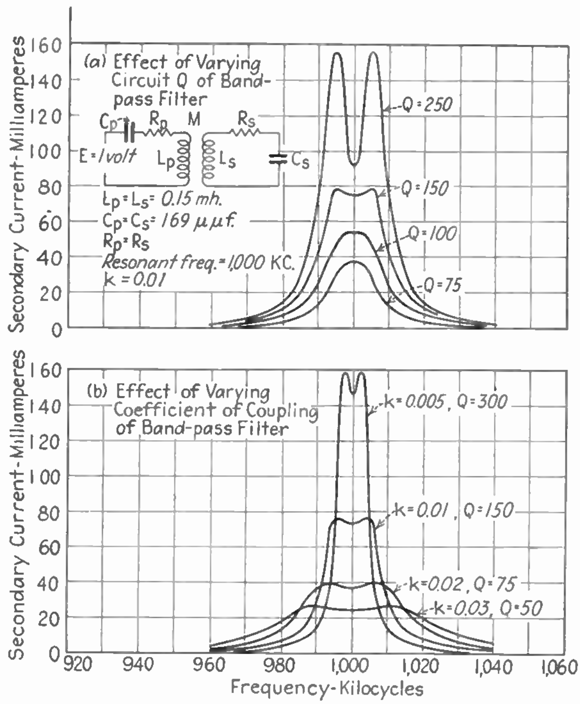


FIG. 24.—Characteristics of band-pass filter, showing (a) effect of circuit Q on uniformity of response within the pass band and (b) effect of coefficient of coupling upon the width of the pass band and the response within the pass band when the proper circuit Q is used.

the circuit Q's low. With low circuit Q's, the secondary current, and hence the response in terms of voltage developed across the secondary condenser, will be low. Hence the wider the pass band the less will be the output voltage. These properties are illustrated in Fig. 24, which shows the necessity of having the proper Q and also indicates how the band width and secondary response vary with coefficient of coupling when the circuit Q is always kept at the value given by Eq. (28).

Band-pass Filters with Parallel Excitation.—Band-pass circuits are very often arranged as in Fig. 25, where the voltage is applied in parallel with the primary through a high resistance R_p. This particular arrangement is especially important since it represents the equivalent circuit of the band-pass amplifier.

The circuit of Fig. 25*a* having parallel excitation can be reduced to the series-excited circuit of Fig. 25*b* by the method described in Sec. 18. The primary current that flows through the inductance of this equivalent circuit is exactly the same as the current through the inductance of the

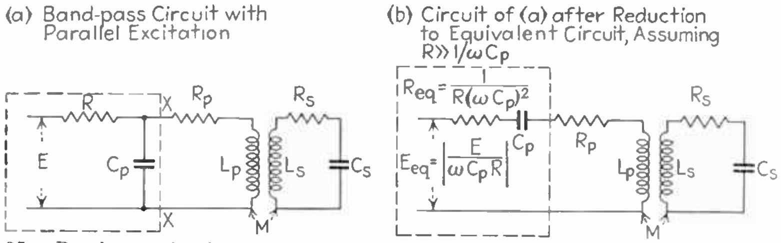


FIG. 25.—Band-pass circuit with parallel excitation together with equivalent series-excited circuit for the usual case where $R \gg 1/\omega C_p$.

actual circuit, and the secondary currents are likewise identical in the two cases. The only essential difference in the behavior of the band-pass circuit with parallel excitation, and the behavior of the same circuit with series excitation, is that the presence of the resistance R has an effect that is equivalent to reducing the Q of the primary circuit.

17. Miscellaneous Coupling Methods.—Energy can be transferred between two circuits by means other than simple inductive coupling such

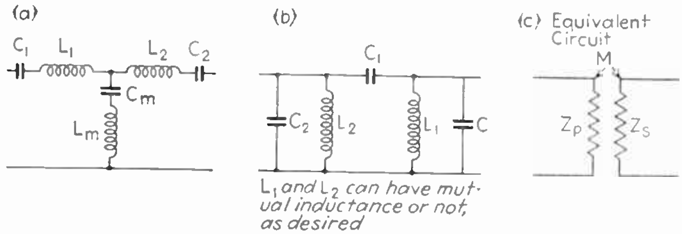


FIG. 26.—Circuits with complex coupling together with equivalent circuit that can be used to represent any coupled circuit.

as discussed in Secs. 15 and 16. Thus in Fig. 8*b* the transfer is accomplished by means of an inductance L_m common to the two circuits, giving what is commonly termed direct coupling. Likewise in Fig. 8*c* the energy transfer occurs as a result of a common capacitance C_m, giving capacitive coupling. It is also possible to combine capacitive and inductive couplings in various manners to give still more complicated methods of energy transfer. Examples of such complex coupling arrangements are shown in Figs. 26*a* and 26*b*.

These more complex coupling methods are often employed because they can be designed in such a manner that the energy transfer (or the equivalent coefficient of coupling) varies with frequency. Thus in the arrangement of Fig. 26*a* the coupling is zero at the frequency for which the mutual capacitance C_m and inductance L_m are in series resonance, while the coefficient of coupling is inductive and increases as the frequency becomes greater than the resonant frequency, and is capacitive and likewise increasing as the frequency becomes lower than resonance.

18. Thévenin's Theorem.—*According to Thévenin's theorem, any linear network containing one or more sources of voltage and having two terminals behaves, insofar as a load impedance connected across these terminals is concerned, as though the network and its generators were equivalent to a generator having an internal impedance Z and a generated voltage E, where E is the voltage that appears across the terminals upon open circuit, and*

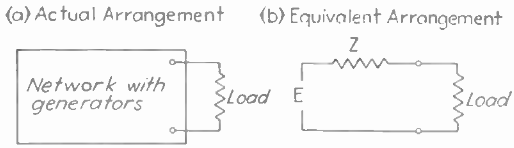


FIG. 27.—Schematic diagram illustrating how Thévenin's theorem can be used to simplify a complicated network containing generators.

Z is the impedance that is measured between the terminals when all sources of voltage in the network are short-circuited.[1] This theorem means that any network and its generators, represented schematically by the block in Fig. 27*a*, can be replaced by the equivalent circuit shown in Fig. 27*b*. The only limitation to the validity of Thévenin's theorem encountered in ordinary practice is that the circuit elements of the network must be linear, *i.e.*, the voltage developed must always be proportional to current.

Thévenin's theorem offers a very powerful means of simplifying networks as demonstrated by the following example.

Example.—Reduce the parallel-excited band-pass circuit of Fig. 25*a* to an equivalent series-excited circuit of the type illustrated in Fig. 23.

This is accomplished by opening the primary circuit of Fig. 25*a* at *xx* and replacing the network to the left of *xx* (*i.e.*, the network in the dotted rectangle) by an equivalent series circuit using Thévenin's theorem. With the source of voltage short-circuited, the impedance seen when looking to the left from *xx* is a condenser C_p shunted by a resistance R. This gives the impedance of the equivalent series circuit a magnitude

[1] When the sources of energy in the network are constant-current generators instead of constant-voltage generators, the internal impedance Z is the impedance that is measured between the terminals when all constant-current generators are *open-circuited.* This is due to the fact that a constant-current generator is equivalent to an infinite voltage source having an infinite internal impedance, so that short-circuiting the ultimate source of voltage of the constant-current generator still leaves an infinite impedance in the circuit.

$\dfrac{1/\omega C_p}{\sqrt{1 + (1/R\omega C_p)^2}}$, and a phase angle θ such that $\tan \theta = -1/\omega RC_p$. The voltage appearing across points xx on open circuit is the equivalent voltage that may be considered as acting in the equivalent series circuit. This voltage has a magnitude $E/(\omega C_p\sqrt{R^2 + (1/\omega C_p)^2})$, and a phase angle corresponding to the phase of the voltage across C_p as compared with the phase of E.

When $R >> 1/\omega C_p$, as is normally the case when the circuit of Fig. 25*a* is encountered in practice, the equivalent series voltage E_{eq} has a magnitude $E/\omega C_pR$, and acts in series with a capacitance C_p and a resistance $1/[R(\omega C_p)^2]$, as shown by the dotted rectangle in Fig. 25*b*.

19. Circuits with Distributed Constants.—When the inductance, capacitance, and resistance of a circuit are mixed together, rather than being separate lumps as in the case of the simple series and parallel circuits that have been considered, the circuit is said to have distributed constants. Examples of such circuits include telephone, telegraph, and power lines, as well as most types of radio antennas.

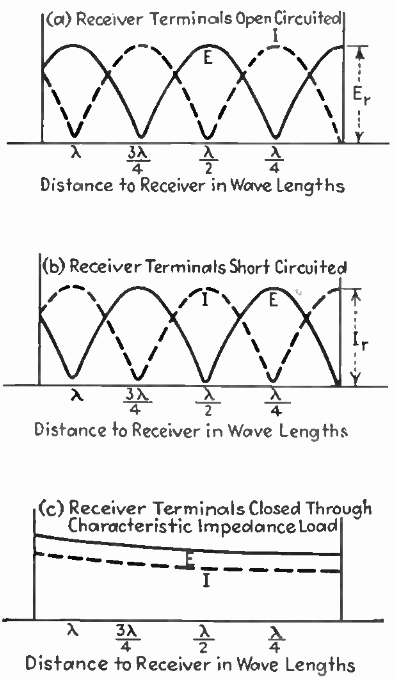


FIG. 28.—Typical voltage and current distributions for circuit with distributed constants, showing conditions existing with different receiver-end conditions.

Voltage and Current Distribution. When a potential is applied to one end of a transmission line, the resulting voltage and current distribution depend upon the load impedance and the length of the line. Results in typical cases are illustrated in Fig. 28, and can be most readily explained when the length of the transmission line is expressed in wave lengths. One wave length corresponds to a distance over which the voltage (or current) shifts in phase by exactly 360°. When the losses in the line are small, as is usually the case, then

$$\left.\begin{array}{l}\text{Distance along circuit corresponding}\\ \text{to one wave length}\end{array}\right\} = \frac{1}{f\sqrt{LC}} \tag{29}$$

where L and C are the series inductance and shunt capacitance, respectively, per unit length of circuit, and f is the frequency of the applied voltage. Where the circuit consists of one or more straight wires in space, as is the case with an antenna or a 60-cycle power line, the distance corresponding to one wave length found from Eq. (29) will always be almost exactly the same as the wave length of radio waves of the same frequency.

Examination of Fig. 28*a* shows that when the receiver terminals are open-circuited, the voltage and current go through cyclic variations as the distance from the receiver is increased.[1] The voltage is high at the receiver and at distances from the receiver that are even multiples of a quarter-wave length, and goes through minima at distances from the receiver that are an odd number of quarter-wave lengths from the receiver. The current distribution is the inverse of the voltage distribution. the current being a maximum where the voltage is minimum, and vice versa.

When the receiving-end terminals are short-circuited, the voltage and current again vary cyclically with distances to the receiver, as shown in Fig. 28*b*. However, the voltage and current distribution are seen to be interchanged from their behavior with an open-circuited receiver.

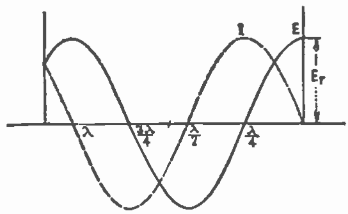


FIG. 29.—Schematic method of representing voltage and current distribution of Fig. 28*a*. This diagram indicates the 180° phase shift on opposite sides of the minima and is an accurate representation of the conditions existing along the circuit except where the voltage and current are shown going through zero.

When the load impedance at the receiving end of the line is a resistance equal to $\sqrt{L/C}$ ohms, the voltage and current distribution are as shown at Fig. 28*c*. The voltage and current now decrease exponentially as the receiver is approached, and there are no cyclic space fluctuations such as exhibited by the open- and short-circuited receiver cases. The load resistance $\sqrt{L/C}$ corresponding to the condition in Fig. 28*c* is termed the *characteristic impedance* and is an important characteristic of the line.

Load impedances other than open circuit, short circuit, and characteristic impedance also give a cyclic distribution. The exact character of this depends upon the load impedance, with the fluctuations becoming greater as the load departs farther from the characteristic impedance in either magnitude or phase.

When the resonance effects are large, as in the case of an open- or short-circuited receiver, the voltage and current at the minima are quite small, and the phase shifts through nearly 180° from one side of a minimum to the other side. In order to show this change of phase, the voltage and current distributions in circuits with distributed constants are frequently drawn as shown in Fig. 29. This corresponds to *a* of Fig. 28

[1] In a transmission line, the term "sending end" means the end of the line to which the generator is connected. The term "receiving end" is the opposite end, *i.e.*, the end at which the load is connected.

except that adjacent maxima are shown on opposite sides of the base line to indicate the reversal of phase. It will be noted that, although this method of representation gives the distribution along most of the line accurately, it fails to show the conditions actually existing at the minima. This is because the curves of Fig. 29 go through zero where the axis is crossed, whereas in reality the minima never reach zero. Also the 180° phase shift does not take place all at one point as indicated by Fig. 29. However, for many purposes these inaccuracies are unimportant, and the method of representation of Fig. 29 is often very convenient.

Transmission Lines as Resonant Circuits.—Transmission lines that are either open- or short-circuited at the receiving end have many of the properties of ordinary resonant circuits. Thus, if the line is open at the receiver and is an odd number of quarter-wave lengths long, the receiving voltage is much higher than the sending-end voltage, giving a resonant rise of voltage similar to that obtained with a series circuit. Also, if the line is an odd number of quarter-wave lengths long and is shorted at the receiver, the sending-end current is low and the line acts very much as a parallel resonant circuit when viewed from the sending end. When properly designed, resonant transmission lines have very high effective Q's at high frequencies, and at such frequencies are superior to tuned circuits consisting of an ordinary coil and condenser.

Problems

1. A variable condenser having a maximum capacitance of 350 $\mu\mu$f and a minimum capacitance of 20 $\mu\mu$f is used for tuning in a broadcast receiver.

a. What coil inductance is required to make the lowest resonant frequency 530 kc, assuming that the coil and circuit wiring contribute an additional capacitance of 20 $\mu\mu$f to the circuit capacitance?

b. Calculate the exact range of resonant frequencies that can be covered with the inductance selected.

2. It is desired to cover the short-wave bands by the use of additional coils to which the condenser of Prob. 1 can be switched. Assuming that the coil and wiring capacitance are the same as in Prob. 1, (*a*) determine the number of coils required to cover frequencies from 530 kc to 30,000 kc, (*b*) specify a suitable inductance value for each coil, and (*c*) calculate the exact tuning range for each coil chosen.

3. A particular coil of 150 μh inductance has $Q = 85$ at 1000 kc.

a. Calculate the condenser capacitance that is required for resonance at 1000 kc.

b. Calculate the circuit Q if the tuning condenser has a power factor of 0.0001.

c. Repeat (*b*) for a condenser with a power factor of 0.01.

4. A series circuit resonant at 800 kc has an inductance of 160 μh and a circuit Q of 75.

a. Calculate and plot the current that flows when 1 volt is applied to the circuit, carrying the curves up to 40 kc on each side of resonance. In making these calculations use the universal resonance curve in the range near resonance and neglect the circuit resistance when calculating points too far off resonance to be within the range of the universal resonance curve.

5. Assume that a series resonant circuit employs the coil of No. 24 wire in Fig. 10, and that the tuning condenser has negligible losses.

a. Calculate and plot the width of the frequency band, for which the tuned circuit response is at least 70.7 per cent of the response at resonance, as a function of resonant frequency from 550 to 1500 kc.

b. Discuss the results obtained in (*a*) with respect to the reception of broadcast signals having side-band frequencies extending up to 5000 cycles on each side of the carrier frequency. Consider both the uniformity of response to the different side-band frequencies, and the ability of the circuit to discriminate against undesired signals of other frequencies.

6. In a series circuit that is resonant at 1150 kc it is found that when the frequency differs from resonance by 15 kc, the current drops to 0.53 of the current at resonance for constant applied voltage. From this information determine the Q of the circuit.

7. A voltage of constant but unknown value is applied to a series circuit resonant at the frequency of this voltage. The circuit current is observed to be I_0. A known resistance R_1 is then added to the circuit, and it is found that, with the same applied voltage as before, the current is now reduced to I_1. Derive a formula for the circuit resistance in terms of I_0, I_1, and R_1.

8. Discuss the difference in magnitude of circuit Q desired when a resonant circuit is required to: (*a*) Discriminate as strongly as possible against all frequencies other than the resonant frequency, (*b*) respond to a modulated carrier wave with very little discrimination against the higher side-band frequencies.

9. In variable condensers used to tune the resonant circuits of radio receivers, it is customary to shape the plates so that the capacitance varies more slowly with angle of rotation at small capacitance settings than at high capacitance settings. Explain why this makes the resonant frequency more nearly linear with respect to angle of rotation than if semicircular plates were employed.

10. *a.* A tuned circuit having an inductance of 150 μh and a Q of 70 is adjusted to resonance at 1100 kc. If the circuit is connected for parallel resonance, calculate and plot the magnitude of the parallel impedance as a function of frequency up to 60 kc on each side of resonance. Use the universal resonance curve in the region about resonance and neglect the circuit resistance when calculating the impedance at frequencies too far off resonance to be within range of the universal resonance curve.

b. Repeat (*a*) for a circuit Q of 40.

11. Calculate and plot the resonant impedance when the coil of No. 24 wire in Fig. 10 is connected as a parallel resonant circuit, and the resonant frequency is varied from 550 kc to 1500 kc.

12. Using the same tuned circuit as in Prob. 10*a*, calculate and plot the following curves as a function of frequency, from 1075 to 1125 kc.

(*a*) Magnitude and phase angle of parallel impedance; (*b*) line current, and current in each branch, when the applied potential is 10 volts (assuming all the circuit resistance is in the inductive branch), and (*c*) reactance and resistance components of the impedance of (*a*).

13. A tuned circuit in a vacuum-tube power amplifier is required to have a parallel impedance of 6000 ohms with a Q of 12. If the resonant frequency is 3000 kc:

a. determine the inductance, capacitance, and resistance that the circuit must have.

b. determine the losses in the tuning condenser, assuming this to be a mica condenser with a power factor of 0.0004, when the *crest* voltage applied to the tuned circuit is 850 volts.

14. A particular coil has an inductance of 180 μh and a distributed capacitance of 10 $\mu\mu$f. Neglecting the coil resistance, calculate the inductive reactance across the coil terminals at frequencies of 500 kc and 2000 kc and from this determine the apparent inductance of the coil at these frequencies.

15. *a.* Prove that when a parallel resonant circuit is shunted by a high resistance, the impedance of the combination in the vicinity of resonance varies with frequency in a manner that is for all practical purposes still the same as the impedance variation of a parallel circuit.

b. If the shunting resistance in (*a*) is R_1, and the parallel resonant impedance of the unshunted circuit is R_0, prove that the shunt resistance R_1 reduces the equivalent Q of the circuit by the factor $R_1/(R_1 + R_0)$.

16. Demonstrate that when the secondary of a coupled circuit is a resonant circuit, the curve of coupled impedance as a function of frequency has the same shape as the impedance curve of a parallel resonant circuit if the frequency range being considered is so small that ωM may be considered as substantially constant.

17. Two identical coils, each having an inductance of 160 μh and a Q of 50, are coupled together with a coefficient of coupling of 0.05. If a potential of 1 volt at 500 kc is applied to the primary, calculate the voltage induced in the secondary when the secondary is open-circuited.

18. In the coupled arrangement of Prob. 17 the secondary coil is short-circuited. Calculate: (*a*) the coupled resistance and reactance at a frequency of 600 kc, (*b*) the total resistance and reactance of the primary circuit at 600 kc, including the effect of the secondary, and (*c*) the effective Q of the primary circuit, taking into account the coupled impedance.

19. The coil of No. 24 wire in Fig. 10 is coupled to a primary coil with a mutual inductance of 50 μh. If the secondary coil is tuned to resonance with a condenser having negligible loss, calculate and plot the coupled impedance at the resonant frequency of the secondary as this resonant frequency is varied from 550 kc to 1500 kc.

20. Two identical circuits resonant at 1000 kc, having $Q = 80$, and inductances of 140 μh, are coupled together.

a. Calculate the critical coefficient of coupling.

b. Calculate and plot the secondary current at the resonant frequency for 1 volt applied to the primary, as the mutual inductance is varied from zero to twice the critical value.

21. The coupling between the circuits of Prob. 20 is adjusted to make the coefficient of coupling have a value 0.02, and 1 volt is applied in series with the primary.

a. What will be the approximate frequencies at which the secondary current peaks will occur?

b. What will be the approximate height of these peaks of secondary current?

c. What will be the secondary current at the resonant frequency?

d. With the information obtained above, sketch the approximate shape of the secondary current curve as a function of frequency.

22. The two circuits of Prob. 20 are coupled with a mutual inductance of 3 μh.

a. Calculate and plot the resistance and reactance components of the coupled impedance up to 40 kc on each side of resonance.

b. Calculate and plot the resistance and reactance components of the primary circuit when the secondary is removed.

c. Add (*a*) and (*b*) to obtain the curve of total primary circuit resistance and reactance, and convert the results into curves giving the magnitude and phase of the primary impedance.

d. Calculate and plot magnitude of primary current and of secondary induced voltage.

e. Calculate and plot the secondary current curve.

Plot the curves for the various parts above each other.

23. Derive a formula for the coefficient of coupling of the circuit of Fig. 26*a*.

24. A particular band-pass filter to be used in the intermediate-frequency amplifier of a radio receiver must have a pass band 7000 cycles wide centering about a frequency of 456 kc. If the primary and secondary inductances are both 2 mh, specify the tuning capacitances, the proper coefficient of coupling, and the proper circuit Q.

25. Signals in the frequency range of 550 to 1500 kc are to be tuned by means of a band-pass filter. If the circuits are assumed to have $Q = 100$ over this frequency range, and the adjustment is such that $k = 0.015$ at 550 kc, discuss how the width and shape of the pass band will vary with resonant frequency when the tuning is obtained by varying the primary and secondary condensers simultaneously and when the coupling is: (*a*) inductive as shown at Fig. 8*a* or 8*b*, and (*b*) capacitive as shown at Fig. 8*c*. Illustrate the discussion with the aid of sketches showing types of response curves to be expected under various conditions.

26. *a.* Derive a formula for the secondary current of the circuit of Fig. 21*b*, by (1) replacing the network to the left of the terminals of the secondary condenser by an equivalent network and generator, using Thévenin's theorem, and (2), using this network to derive an equation for secondary current.

b. From the results of (*a*) show that the curve of secondary current as a function of frequency has substantially the same shape as a resonance curve. Discuss the factors that determine the effective Q corresponding to the secondary current curve, and also consider the factors that cause the peak of the curve of secondary current to occur at a frequency differing from the resonant frequency of the secondary.

CHAPTER IV

FUNDAMENTAL PROPERTIES OF VACUUM TUBES

20. Electron Tubes, Electrons, and Ions.—A vacuum tube includes a cathode capable of emitting electrons when heated, and an anode (often called the plate) that is normally operated at a positive potential and attracts these electrons. In most tubes there are also one or more additional electrodes for controlling the flow of electrons and influencing the characteristics. The cathode and other electrodes are enclosed in a gas-tight envelope, usually glass but sometimes metal. Most tubes are evacuated as completely as possible and are operated as high-vacuum devices. In some cases, however, small quantities of gas are intentionally introduced after the evacuation in order to modify the characteristics.

Tubes are classified as diodes, triodes, tetrodes, pentodes, etc., according to whether there are two, three, four, five, etc., electrodes present. Thus a tube with only cathode and anode is a diode, while the addition of a control electrode (a grid) converts it into a triode.

Electrons and Ions.—Electrons may be considered as minute negatively charged particles which are constituents of all matter. They have a mass of 9×10^{-28} gram, which is $\frac{1}{1840}$ that of a hydrogen atom, and a charge of 1.59×10^{-19} coulomb. They are also always identical irrespective of the source from which derived. Atoms are composed of one or more such electrons associated with a much heavier nucleus which has a positive charge equal to the number of electrons contained in the atom, so that an atom with its full quota of electrons is electrically neutral. The differences between chemical elements arise from differences in the nucleus and in the number of associated electrons, but not from variations in the character of the electron.

Positive ions represent atoms or molecules that have lost one or more electrons. Positive ions are hence charged bodies having the weight of the atom or molecule concerned, and a charge equal to that of the lost electrons. Unlike electrons, positive ions are not all alike and may differ in charge or weight, or both. They are much heavier than electrons and resemble the molecule or atom from which derived. Ions are designated according to their origin, such as mercury ions, hydrogen ions, etc.

Methods by Which Free Electrons and Ions May Be Obtained.—Electrons may be separated from matter in a number of ways. Thus at very high temperatures some electrons will escape from the surface of a solid, thereby giving free electrons by *thermionic emission*. This is the means by which the cathode of an ordinary vacuum tube produces free electrons. Electrons can also be obtained from solid materials as a result of the fact that rapidly moving electrons or ions will knock out electrons from a surface when striking with sufficient velocity. This process is termed *secondary electron emission* because it requires some primary source of electrons (or ions). Light in striking certain types of surfaces will likewise cause electrons to be emitted. Such emission is said to be caused by the *photoelectric effect*, and is the basis of the photoelectric cell. A swiftly moving particle (commonly an electron or ion) colliding with a gas molecule may knock one or more electrons out of the molecule and leave a positive ion. This is termed *ionization by collision*, and occurs in vacuum tubes in which gas is present.

The Motion of Electrons and Ions in an Electrostatic Field.—Electrostatic fields exert forces upon electrons and ions just as upon any other charged body. The electrons, being negatively charged, tend to travel toward the positive or anode electrode. The positive ions on the other hand travel in the opposite direction toward the negative or cathode electrode.

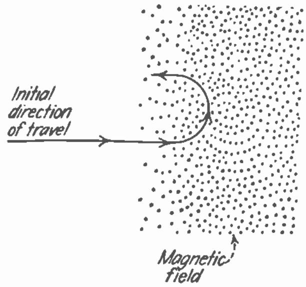


FIG. 30.—Path of electron that is projected with high velocity into a magnetic field. The moving electron is deflected in a direction that is at right angles to the magnetic field and at right angles to the direction in which the electron is traveling, at the moment.

In moving between two points in the electrostatic field, the energy received by an electron (or ion) equals the product of its charge and the difference in potential between these points. This energy is converted into kinetic energy of motion so that the velocity that a particular charged body possesses is commonly expressed in terms of volts. Thus when an electron is said to have a velocity of 10 volts this means the velocity that the electron would acquire in falling through a difference of potential of 10 volts. Since the velocity that a charged particle gains in falling through a difference of voltage is inversely proportional to the square root of the weight of the particle, the velocity represented by a given voltage will depend upon the weight of the charged body and will be much greater with electrons than ions, particularly the heavy ions. Thus a mercury ion having a charge equal to that of an electron will move less than one six-hundredth as fast as an electron in the same electrostatic field.

Effect of a Magnetic Field on a Moving Electron.—A magnetic field exerts a force on a moving electron because of the fact that a moving charge is an electrical current, and a magnetic field produces a force upon a current. This force produced by the magnetic field is at right angles to both the magnetic field and the line of current flow (*i.e.*, direction of travel of electron or ion), as shown in Fig. 30. The force is proportional to the charge of the moving particle, to the magnetic flux density, and to the component of the velocity that is at right angles to the magnetic field.

21. Thermionic Emission.—In order to escape from the surface of a conductor, an electron must overcome restraining forces acting at the surface of the metal. In the case of thermionic emission the energy required to do this must come from the kinetic energy possessed by the electron as a result of its motion within the conductor. The properties of matter are such that it is only at high temperatures, where the average kinetic energy possessed by the electrons is large, that an appreciable number will have sufficient kinetic energy to escape through the surface.

The process of electron emission from a solid substance is very similar to the evaporation of vapor from the surface of a liquid. In the case of the vapor the evaporated molecules represent molecules that obtained sufficient kinetic energy to overcome the restraining forces at the surface of the liquid, and the number of such molecules increases rapidly as the temperature is raised. The thermionic emission of electrons from hot bodies represents the same process, and may be considered as an evaporation of electrons in which the energy the electron must give up in escaping corresponds to the latent heat of vaporization of a liquid.

The number of electrons evaporated per unit area of emitting surface is related to the absolute temperature T of the emitting material and a quantity b that is a measure of the work an electron must perform in escaping through the surface, according to the equation

$$I = AT^2\epsilon^{-\frac{b}{T}} \tag{30}$$

where I is the electron current in amperes per square centimeter and A is a constant, the value of which may vary with the type of emitter. The temperature at which the electron current becomes appreciable is determined almost solely by the quantity b, which is accordingly the most important characteristic of an electron-emitting material. The value of A is of secondary importance, for the effects of wide variations in A can be compensated for by small temperature changes.

Electrons having more kinetic energy than that required to escape into the surrounding space will leave the metal with a velocity corresponding to their excess kinetic energy. This velocity is termed the *velocity of emission* and will reach as much as 1 volt for a few of the emitted electrons.

Power Required to Heat Electron Emitters.—The temperature required for electron emission can be obtained by arranging the cathode in the form of a filament which is brought to the desired temperature by passing current through it, or by forming the cathode into a cylinder that has an internal heater consisting of a tungsten filament. These are termed *filament*, and *heater-type*, cathodes, respectively.

Most of the power required to maintain the cathode at the proper temperature for electron emission represents energy sent out from the cathode in the form of radiant heat. In addition, there is a small loss by heat energy conducted away from the cathode through the support wires, and a further loss from the energy required to evaporate the electrons from the surface of the cathode. The heat radiated from a cathode that is at a considerably greater temperature than the surrounding objects is proportional to the fourth power of the absolute temperature. Hence, if other things are equal, the heating power will tend to be low when the temperature required for emission is low [*i.e.*, when the constant b in Eq. (30) is small].

22. Practical Emitters.—The properties desired in an electron emitter are high emission in proportion to heating power, and long life. These requirements are mutually conflicting, since a high emission in proportion to heating power calls for operation at the highest possible temperatures, and this reduces the life. The emitters that most satisfactorily meet the requirements of efficiency and life are tungsten, thoriated-tungsten, and oxide-coated cathodes.

Oxide-coated Emitter.—The oxide-coated emitter consists of a mixture of barium and strontium oxides coated on the surface of a metal such as platinum or nickel alloy. When properly prepared, such a surface will emit large numbers of electrons at temperatures of the order of 1150°K. This electron emission arises from a layer of alkaline-earth metal, *i.e.*, metallic barium and strontium, which forms on the surface of the oxide coating. The emission is maximum when this layer covers the entire surface of the oxide to a depth of approximately one molecule.

The high thermionic activity of oxide-coated cathodes appears to be caused by this surface layer becoming positively charged as a result of losing electrons. The surface then acts as a positively charged grid which covers the surface of the oxide coating and helps pull electrons from its surface.

The life of an oxide-coated emitter under favorable conditions is terminated by the exhaustion of the active material that maintains the emitting layer. Carefully made and properly operated tubes employing oxide-coated cathodes normally have a life of at least 5000 hr., and 20,000 hr. is not unusual. Under unfavorable conditions, or with improperly prepared cathodes, the life is much less. Thus if the cathode is operated at too high a temperature the electron emission in proportion to heating power is correspondingly increased, but the life is greatly shortened.

The presence of small traces of gas within a tube having an oxide-coated cathode produces an adverse effect on life. This gas is ionized as a result of collision with the electrons traveling to the anode, and the resulting positive ions bombard the cathode. This produces a mechanical disintegration of the cathode that becomes increasingly serious as the anode voltage of the tube is raised.

The oxide-coated cathode is the most efficient emitter of electrons that has been discovered. When operated under favorable conditions, *i.e.*, in tubes with low or moderate anode voltage and with a good vacuum, the life is also greater than with other emitters. The principal use of the oxide-coated emitter is in tubes where the anode potentials are small, *i.e.*, less than 500 volts, and where the total heat dissipated in the tube is small. The low anode voltage minimizes the effect of the residual gas, while low dissipation reduces the possibility of gas being released from the metal and glass parts of the tube during operation. Heater-type cathodes are always oxide-coated because this is the only emitter that will operate at the temperatures obtainable with indirect heating.

Tungsten.—Tungsten is a relatively poor emitter of electrons (*i.e.*, an electron must do considerable work to escape from a tungsten surface). Because of its refractory character, however, tungsten can be operated at very high temperatures. This, combined with its ruggedness, causes tungsten to be used where other emitters are not satisfactory, *i.e.*, where the anode voltage is very high and the vacuum is not so perfect as might be desired.

The life of a tungsten emitter is set by the evaporation of tungsten from the hot filament. The rate of evaporation increases rapidly with temperature, so that the life decreases as the temperature is raised. At the same time, raising the temperature increases the electron emission. The best compromise in this situation corresponds to a temperature that gives a life expectancy of 1000 to 3000 hr. This means a temperature of 2450 to 2600°K. The higher temperatures are used with larger diameter filaments, since then evaporation to a given depth reduces the diameter by a smaller percentage.

Thoriated-tungsten Emitters.—A thoriated-tungsten emitter consists of ordinary tungsten to which a small amount of thorium oxide and carbon has been added. When properly treated, a layer of thorium one molecule deep will be formed upon the surface of the emitter. This layer acts in much the same way as the corresponding film of barium on the surface of the oxide-coated emitter, and assists the electrons in escaping. Profuse emission is obtained at a temperature of about 1900°K., which is about 550 to 700° lower than the operating temperature of pure tungsten.

The mono-molecular film of thorium is formed by flashing the filament at a temperature of about 2700°K. for 1 or 2 min. and then glowing for some minutes at a temperature around 2150°K. The flashing raises the temperature to the point where the impregnated carbon reduces some of the thorium oxide to metallic thorium. The subsequent glowing allows this thorium to diffuse to the surface where it forms a layer one molecule deep that is the seat of the electron emission.

The best operating temperature for a thoriated tungsten filament is approximately 1900°K. At a higher temperature the thorium diffuses very rapidly to the surface where the excess is evaporated, thereby quickly exhausting the supply of this material. At lower temperatures, on the other hand, the thorium diffuses so slowly to the surface that thorium molecules are lost from the mono-molecular surface film through evaporation and positive-ion bombardment more rapidly than they are replaced by thorium from the interior of the cathode. At the proper operating temperature, the life of a thoriated-tungsten filament can be expected to be of the order of 5000 hr.

Cathode bombardment by positive ions has a very detrimental effect upon thoriated-tungsten emitters. Such bombardment strips the tungsten surface of the mono-molecular layer of thorium and thereby destroys the high emission. It is accordingly necessary that tubes employing thoriated-tungsten emitters be very thoroughly evacuated and so treated that a minimum of gas will be released during the life of the tube.

The detrimental effects of gas can be reduced by carbonizing the surface of the tungsten. This can be accomplished by glowing at about 1600°K. in a hydrocarbon vapor. The mono-molecular layer of thorium clings much more tenaciously to such a tungsten-carbide surface than it does to pure tungsten. This in turn permits operation at slightly higher cathode temperatures, thereby increasing the rate at which thorium diffuses to the surface to replace the thorium lost through positive-ion bombardment.

Thoriated-tungsten emitters are slightly less efficient than oxide-coated emitters, but have much higher efficiency than pure tungsten. At the same time, thoriated tungsten when carbonized is much more resistant to positive-ion bombardment than is the oxide-coated cathode, although less satisfactory from this point of view than pure tungsten. As a consequence, thoriated-tungsten emitters find their chief use in the larger air-cooled tubes, particularly those operating at anode potentials in excess of 500 volts.

23. Current Flow in a Two-electrode Tube. Space-charge Effects.—When an electron-emitting cathode is surrounded by a positive anode (*i.e.*, plate electrode) to form a two-electrode vacuum tube (or diode), the relation between the plate current (*i.e.*, the number of electrons collected by the anode) and the plate potential has the character shown in Fig. 31. When the plate is negative it repels the emitted electrons back into the cathode and the plate current is zero. On the other hand, at high positive potentials the plate attracts electrons as fast as they are emitted. The current is then given by Eq. (30) and is determined largely by the cathode temperature and is substantially independent of the electrode voltage. This condition is termed *voltage saturation.*[1]

When the plate potential is positive but low, the plate current is limited by the negative space charge produced by the electrons that are

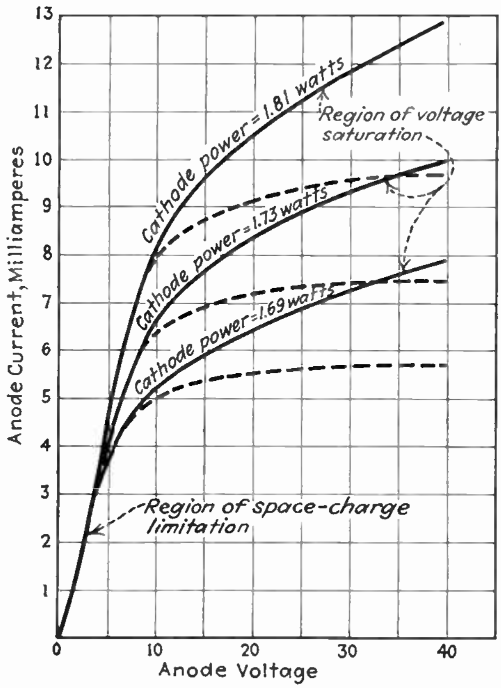


FIG. 31.—Anode current as a function of anode voltage in a two-electrode tube for three cathode temperatures. The solid lines are the characteristics actually obtained using an oxide-coated cathode, whereas the dotted lines show the type of curves given by tungsten and thoriated-tungsten cathodes.

in transit between cathode and plate. This is because the number of electrons in transit between electrodes at any instant cannot exceed the number that will produce a negative space charge which completely neutralizes the attraction of the positive plate upon the electrons just leaving the cathode. All electrons in excess of the number necessary

[1] The sharpness with which voltage-saturation effects appear differs greatly with the type of emitter. Thus the anode current with cathodes of tungsten or thoriated tungsten has a characteristic such as shown by the dotted lines in Fig. 31, in which the saturation effect is almost complete. On the other hand, with emitters of the oxide-coated type saturation takes place more gradually, as shown by the solid lines of Fig. 31.

to neutralize the effect of the plate voltage are repelled back into the cathode by the negative space charge of the electrons in transit. Where the plate current is limited in this way by *space charge,* the plate current is determined by the plate potential, and is substantially independent of the electron emission of the cathode.

When the cathode is an equipotential surface (heater cathode), the total plate current for positive plate voltages with space-charge limitation is given by the equation

$$\text{Plate current} = KE_b^{3/2} \tag{31}$$

Here K is a constant determined by the geometry of the tube, and E_b is the anode (plate) voltage with respect to the cathode.[1] For negative plate voltages the plate current is zero.

In filament-type tubes the voltage drop produced in the cathode by the heating current causes different parts of the filament to have different potentials with respect to the anode. The space-charge-limited current from each part of the filament is then proportional to the $3/2$ power of the voltage with respect to that part, but the total current is not exactly proportional to the $3/2$ power of the potential of the anode. When the anode potential is measured with reference to the negative end of the filament, as is customary, the anode current varies as a power of the anode potential that is $5/2$ at low anode potentials and decreases to $3/2$ when $E_b \gg E_f$.

The energy that is delivered to the tube by the source of anode voltage is first expended in accelerating the electrons traveling from cathode to anode and so is converted into kinetic energy. When these swiftly moving electrons strike the anode, this kinetic energy is then transformed into heat that must be radiated to the walls of the tube.

24. Action of the Grid.—The flow of electrons to the plate can be controlled by placing a screenlike electrode, or grid, between the cathode and plate. This gives a three-electrode or triode tube. The grid is normally operated at a negative potential with respect to the cathode and so attracts no electrons. However, the extent to which it is negative affects the electrostatic field in the vicinity of the cathode and so controls the number of electrons that pass between the grid wires and on to the plate.

The grid functions as an imperfect electrostatic shield that allows some but not all of the electrostatic flux from the anode to leak between

[1] When the anode voltage is low and very precise results are required, the voltage E_b appearing in Eq. (31) must be interpreted to mean the actual anode voltage plus a correction to take into account the contact potential existing between plate and cathode and also the effective velocity of emission of the electrons. Each of these corrections ordinarily amounts to less than 1 volt, and so can normally be neglected where the anode voltage is moderately high.

its wires. The commonest type of grid structure consists of a spiral helix of circular or elliptical cross section, but any arrangement can be

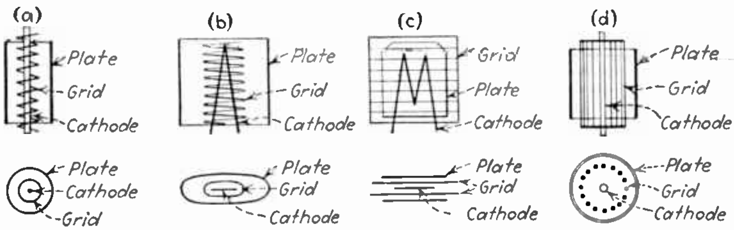


FIG. 32.—Grid, plate, and cathode structures of a number of typical tubes. It will be observed that in every case the grid is a screenlike electrode that affects the electrostatic field near the cathode while permitting electrons to flow to the plate.

employed in which the grid potential affects the electrostatic field at the cathode while allowing electrons to pass on to the plate. A number of typical grid structures used in commercial tubes are illustrated in Fig. 32.

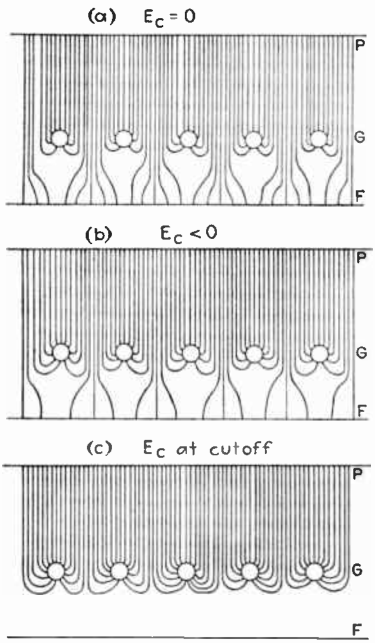


FIG. 33.—Schematic picture of electrostatic field produced between plate and cathode with different grid potentials, showing how the electrostatic field in the vicinity of the cathode can be controlled by the potential of the grid. These curves take into account only those fields produced by the electrode potentials and do not include the field developed by the space charge of electrons which is superimposed upon the fields shown.

The grid electrode controls the flow of electrons to the plate because the electrostatic field between the plate and cathode, and particularly the field near the cathode, is affected by the grid potential. This is shown in Fig. 33, which gives a schematic picture of the electrostatic field that exists between plate and cathode for several values of grid voltage (no space charge present). When the grid is at zero potential with respect to the cathode, the positive potential of the plate produces a stray electrostatic field near the cathode, which, although somewhat weaker than would be the case with the grid removed, is still not zero because the grid is not a perfect shield. As the grid is made negative, it produces an electrostatic field between cathode and grid that opposes the stray field produced by the plate potential and thereby weakens the electrostatic field in the vicinity of the cathode, as shown at *b* in Fig. 33. When the grid is made sufficiently negative, the stray electrostatic field produced at the cathode by the positive anode is entirely neutralized by the negative grid, as shown at c in Fig. 33. In this last case there is no electrostatic field to draw the emitted electrons away from the cathode, and the space current will be zero.

The number of electrons that reach the anode is determined almost solely by the electrostatic field near the cathode and is affected hardly at all by the field in the rest of the interelectrode space. This is true because the electrons near the cathode are moving very slowly compared with the electrons that have traveled some distance toward the plate. The volume density of electrons is hence large near the cathode and low in the remainder of the interelectrode space. The total space charge of the electrons in transit toward the plate is then made up largely of the electrons in the immediate vicinity of the cathode. Once an electron has traveled beyond this region, it reaches the plate so quickly as to contribute to the space charge for only a brief additional time interval. The result is that the space current in a three-electrode vacuum tube is for all practical purposes determined by the electrostatic field that the combined action of the grid and plate potentials produces near the cathode.

When the grid structure is symmetrical, it can be shown that in the absence of space charge the electrostatic field at the surface of the cathode is proportional to the quantity $\left(E_c + \frac{E_b}{\mu}\right)$, where E_c and E_b are the grid and anode (plate) voltages, respectively, with respect to the cathode, and where μ is a constant that is determined by the geometry of the tube and is independent of the grid and plate voltages. The constant μ is known as the *amplification factor* of the tube and is a measure of the relative effectiveness of grid and plate voltages in producing electrostatic fields at the surface of the cathode.

Quantitative Effect of Grid Potential on Space Current.—The space current in a perfectly symmetrical three-electrode tube varies with $\left(E_c + \frac{E_b}{\mu}\right)$ in exactly the same way that the space current in a two-electrode tube varies with the plate voltage. This is true because in both cases the current flow is determined by the electrostatic field near the cathode, and this field is in turn proportional to $\left(E_c + \frac{E_b}{\mu}\right)$ when a grid is present and to E_b when there is only a plate. With a perfectly symmetrical tube and no voltage drop in the cathode the space current is therefore proportional to $\left(E_c + \frac{E_b}{\mu}\right)^{3/2}$. When the grid is negative all this current goes to the plate, so that for positive values of $\left(E_c + \frac{E_b}{\mu}\right)$

and negative grid potentials

$$\text{Plate current} = K\left(E_c + \frac{E_b}{\mu}\right)^{3/2} \tag{32}$$

where K is a constant determined by the tube dimensions.[1] For negative values of $\left(E_c + \frac{E_b}{\mu}\right)$ the plate current is zero. It will be noted that this equation is analogous in all respects to Eq. (31), and that by interpreting $\left(E_c + \frac{E_b}{\mu}\right)$ to be the effective anode voltage they are identical.

25. Characteristic Curves of Triodes.—The most important characteristics of vacuum tubes with grid, plate, and cathode electrodes are the

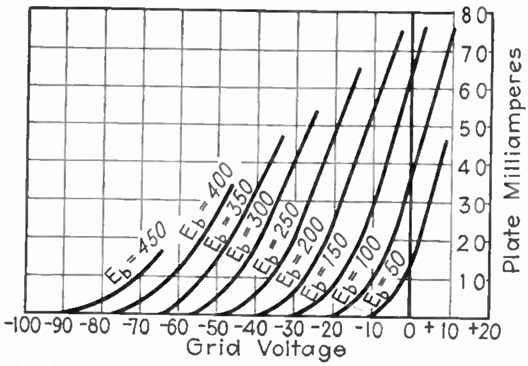


FIG. 34.—Relationship between grid voltage and plate current for several values of plate voltage in a typical three-electrode tube. Note that the principal effect of changing the plate voltage is to displace the curves without changing the shape.

relationships between: (1) plate current and plate voltage with constant grid voltage, and (2) plate current and grid voltage with constant plate voltage. Examples of such curves are shown in Figs. 34, 35, and 36.

It will be noted that the various curves of any one family are all of approximately the same shape. Furthermore, the curves of Fig. 34 have the same shape as the curves of Fig. 35, and this is the same shape as the part of Fig. 31 that is space-charge limited. These various properties of the characteristic curves of a three-electrode tube result from the fact that the plate current is determined only by $\left(E_c + \frac{E_b}{\mu}\right)$ and not by the particular combination of grid and plate voltages involved.

[1] For highest accuracy, the parenthesis on the right-hand side of Eq. (32) must be corrected for contact potentials and velocity of emission, exactly as Eq. (31). This correction ordinarily amounts to less than 1 volt, and so is unimportant unless the value of the parenthesis is small.

The range covered by Figs. 34 and 35 lies in the region where the anode current is limited by space charge. Figure 36 shows the situation that exists when the electron emission is sufficiently low to bring in voltage saturation. It is seen that the anode current is still a function of

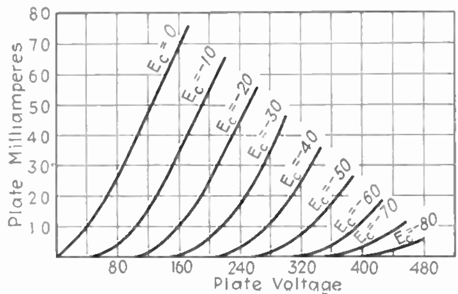


FIG. 35.—Relationship between plate voltage and plate current for several values of grid voltage for the same tube as in Fig. 34. Note that the principal effect of changing the grid voltage (provided the grid is at least slightly negative) is to displace the curves without changing the shape. Note also that these curves have the same shape as those of Fig. 34.

$\left(E_c + \frac{E_b}{\mu}\right)$ exactly as in Fig. 34, but the shape of the curves is now different as a result of voltage saturation. The curves of Figs. 34 and 35 would show similar saturation effects if extended to higher values of plate current.

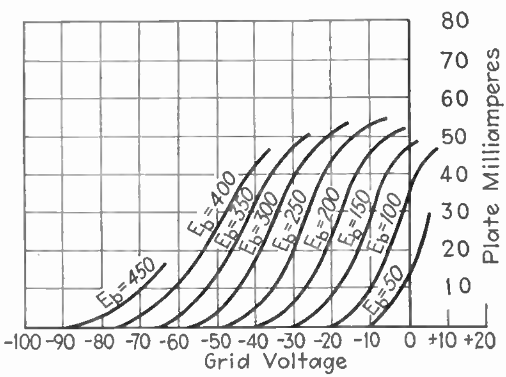


FIG. 36.—Grid-voltage–plate-current curves differing from those of Fig. 34 only in that the cathode temperature has been lowered to the point where voltage saturation begins to appear at the larger plate currents, causing the tops of the curves to bend over.

The plate current of a three-electrode tube becomes zero when $\left(E_c + \frac{E_b}{\mu}\right)$ is zero or negative. The condition $\left(E_c + \frac{E_b}{\mu}\right) = 0$ exists when the grid is just sufficiently negative to neutralize the attracting power of the plate at the cathode (see Fig. 33*c*). This condition is known as *cut-off* and corresponds to $E_c = -E_b/\mu$.

26. Constants of Triode Tubes.—The most important characteristics of a triode tube can be expressed in terms of three coefficients, or constants, termed the amplification factor μ, the dynamic plate resistance R_p (generally called plate resistance), and the mutual conductance G_m (also called transconductance). With the aid of these constants it is possible to make quantitative calculations of the tube performance under many conditions without resort to the complete characteristic curves.

Amplification Factor.—The amplification factor μ has already been defined in Sec. 24 as the ratio of the effectiveness of the grid and plate voltages in producing electrostatic fields at the cathode surface. It is determined by the geometry of the system comprising the grid, plate, and cathode electrodes, and its calculation in terms of the dimensions involved is a problem of pure electrostatics. The amplification factor depends primarily upon the grid structure and will be increased by anything that causes the grid to shield the cathode more completely from the plate. Thus larger grid wires or a closer spacing of the grid wires will increase the amplification factor. The amplification factor of ordinary three-electrode tubes ranges from about 3 as the minimum to about 100 as the practical maximum. The value in any particular case depends upon the purpose for which the tube was designed.

If the relative effects of the grid and plate voltages in producing electrostatic field at the cathode were the same for all parts of the cathode, the amplification factor μ would be absolutely independent of plate, grid, and filament voltages. In commercial tubes the necessity of supporting wires and the inevitable imperfections in construction result in dissymmetries that causes different parts of the tube to have somewhat different amplification factors. The over-all amplification factor of such a combination will vary with plate, grid, and filament voltages and will tend to become lower as cut-off is approached because, as the grid becomes more negative, those parts having the highest value of μ will reach cut-off first, leaving only the low μ parts of the tube contributing to the space current.

In the practical case where the amplification factor is not a pure geometrical constant, it is defined as the relative effectiveness of grid and plate voltages upon the plate current and so is expressed by the following mathematical relation:

$$\text{Amplification factor} = \mu = \frac{\partial i_p/\partial e_g}{\partial i_p/\partial e_p} \tag{33a}$$

The amplification factor can also be defined in terms of voltage increments Δe_g and Δe_p to the grid and plate potentials, respectively, that keep the plate current constant. That is,

$$\left.\begin{array}{l}\text{Amplification}\\ \text{factor}\end{array}\right\} = \mu = -\frac{\Delta e_p}{\Delta e_g}\bigg|_{i_p \text{ constant}} = -\frac{de_p}{de_g}\bigg|_{i_p \text{ constant}} \tag{33b}$$

Plate Resistance.—The plate resistance (sometimes called dynamic or a-c plate resistance) of a vacuum tube represents the resistance that the plate circuit offers to a small increment of plate voltage. Thus, when an increment of plate voltage Δe_p produces an increment in the plate current of Δi_p, the plate resistance is given by the relation

$$\text{Plate resistance} = R_p = \frac{\Delta e_p}{\Delta i_p} = \frac{\partial e_p}{\partial i_p} = \frac{de_p}{di_p}\bigg|_{e_g \text{constant}} \tag{34}$$

The plate resistance is therefore the reciprocal of the slope of the plate-current–plate-voltage characteristic shown in Fig. 35, and depends upon the grid and plate voltages at the operating point under consideration. It is important to remember that the plate resistance is *not equal to the ratio of total plate voltage to total plate current.*

In any particular tube the plate resistance depends primarily upon the plate current and only to a small extent upon the combination of grid and plate voltages used to produce this current. Furthermore, the plate resistance becomes progressively lower as the plate current is increased in the absence of saturation. This behavior is clearly apparent in Fig. 35.

The plate resistance of tubes differing only in grid structure decreases as the amplification factor is lowered. This is because the change in electrostatic field produced near the cathode by a given plate-voltage increment is inversely proportional to the amplification factor.

Mutual Conductance (or Transconductance).—The mutual conductance G_m (or, as it is often called, the transconductance) is defined as the rate of change of plate current with respect to a change in grid voltage. Thus, if the grid voltage is changed by Δe_g, the resulting plate-current change Δi_p is related to the mutual conductance by the equation

$$\Delta i_p = \Delta e_g G_m$$
$$G_m = \frac{\Delta i_p}{\Delta e_g} = \frac{\partial i_p}{\partial e_g} = \frac{di_p}{de_g}\bigg|_{e_p \text{ constant}} \tag{35a}$$

By combination of Eqs. (33), (34), and (35*a*) it is also found that the mutual conductance is the ratio of amplification factor to plate resistance. That is

$$G_m = \frac{\mu}{R_p} = \frac{\partial i_p}{\partial e_g} \tag{35b}$$

The mutual conductance has the dimension of a conductance, and is commonly expressed in micromhos, abbreviated μmho.

The mutual conductance is a rough indication of the design merit of a tube. This is because a low plate resistance and a high amplification factor are desired, and the mutual conductance measures the extent to which this feature is attained.

In a particular tube the mutual conductance depends primarily upon the plate current and to only a small extent upon the combination of grid and plate voltages used to produce this current. The mutual conductance also increases as the plate current is increased, provided there is a full space charge. Typical values of mutual conductance for normal operating conditions range from 500 to 5000 μmho.

27. Pentodes.—A pentode can be thought of as an ordinary triode tube to which two additional concentric grids have been added between cathode and plate. This gives a total of three grids, which are arranged

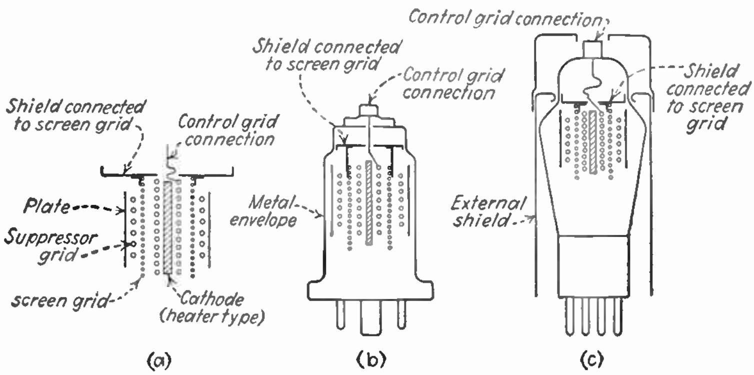


FIG. 37.—Schematic diagrams of typical pentode tubes showing the electrode arrangement and how the control grid is completely shielded from the plate in tubes intended for use at high frequencies. Audio-frequency power pentodes are similar, except that the shield is omitted.

as illustrated in Fig. 37. The inner grid is called the control grid and corresponds to the grid of a triode tube. The next grid is termed the screen grid, or screen, while the outer grid is called the suppressor. In normal operation the control grid is maintained negative with respect to the cathode, the screen grid is operated at a fixed positive potential, the suppressor is connected directly to the cathode, and the plate is operated at a positive potential.

The additional grids operated in this way modify the voltage and current relations existing within the tube in a way that is desirable for many purposes. The additional grids also provide electrostatic shielding between the anode and the control grid, thereby eliminating electrostatic coupling between circuits associated with the control grid and circuits associated with the anode. This shielding is particularly important in the case of radio-frequency amplifiers. Pentode tubes intended for such service realize practically perfect shielding by arrangements such as illustrated in Fig. 37.

Voltage and Current Relations in Pentode Tubes.—The nature of the voltage and current relations in a pentode tube can be derived by considering the potential distribution in the space between the plate and cathode. This distribution is shown in Fig. 38*a* for a typical pentode having a full space charge in the immediate vicinity of the cathode. The number of electrons drawn from the cathode under conditions of space-charge limitation is determined by the electrostatic field at the surface of the cathode, exactly as in the case of the triode. This field in pentodes depends upon the potential of the control and screen grids and the geometry of the tube. It is not affected appreciably by the plate potential, because the screen and suppressor effectively shield the cathode from electrostatic fields produced by the plate.

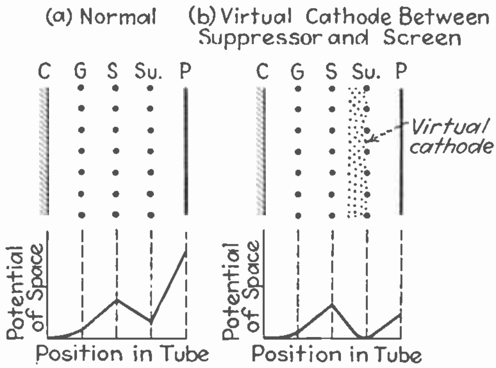


FIG. 38.—Potential distribution in pentode tubes with and without a virtual cathode between suppressor and screen.

The electrons drawn from the space charge pass between the control-grid wires and are accelerated to a high velocity as the screen grid is approached. At this high velocity the electrons travel in substantially straight lines, so very few except those which happen to be going directly toward the screen-grid wires are intercepted by the screen. The remaining electrons then pass through the screen grid and travel on toward the suppressor. As the suppressor is approached, the electrons slow down because of the retarding field between suppressor and screen, but if the plate is reasonably positive they will pass on through the spaces between the suppressor-grid wires and reach the plate. This is because the suppressor, being only an imperfect electrostatic shield, does not prevent the plate from attracting the electrons in the screen-suppressor space.

However, if the plate potential is very low, the electrons will not be attracted to the plate as fast as they approach the suppressor. A space charge, called a *virtual cathode*, then forms between the suppressor and screen as shown in Fig. 38*b*. As far as the suppressor and plate are concerned, this virtual cathode acts as a real cathode, and the combination of virtual cathode, suppressor, and plate then behaves as an ordinary triode tube. The plate current under such conditions tends to be independent of the control-grid and screen-grid potentials and to be determined solely by suppressor and plate voltages. The excess electrons arriving at the virtual cathode above those drawn off to the plate turn about and are collected by either the screen-grid or the cathode.

The total space current in the absence of a virtual cathode is determined by the electrostatic field at the surface of the cathode, and, for a perfectly symmetrical tube with an equipotential cathode, is given by the equation[1]

$$\text{Total space current} = I_b + I_{sg} = K\left(E_c + \frac{E_{sg}}{\mu_{sg}}\right)^{3/2} \qquad (36)$$

where

I_b and I_{sg} = plate and screen currents, respectively

K and μ_{sg} = constants determined by the tube construction

E_c and E_{sg} = control-grid and screen-grid potentials, respectively.

The plate voltage has practically no effect on the space current and so does not appear in the equation. It is to be noted that this equation is strictly analogous to Eq. (32) for triodes, the only difference being that the screen grid has taken the place of the plate in the triode equation.

The total space current given by Eq. (36) divides between the positive electrodes, *i.e.*, between the screen and the plate. With plate voltages that are sufficient to prevent the formation of a virtual cathode in front of the suppressor, the ratio of plate to screen currents is very nearly equal to the ratio of the area of the spaces between the wires of the screen-grid structure to the projected area of the wires themselves and is, to a first approximation, independent of the plate and screen potentials. The plate current then follows the same law as the total space current and so is given by an equation of the same form as Eq. (36). The plate current normally constitutes the major part of the space current and is commonly about 80 per cent of the total.

Characteristic Curves of Pentodes.—The actual voltage and current relations existing in a pentode tube can be shown by means of characteristic curves, of which those of Figs. 39 to 42 are typical. The total space current in the absence of a virtual cathode is seen from Eq. (36) to be independent of plate potential. The plate current is likewise proportional to the total space current under these conditions and varies with control grid and screen potentials in the same way that the plate current of a triode varies with grid and plate voltages.

[1] For highest accuracy the quantity inside the parentheses on the right-hand side of the equation must be corrected to take into account contact potential and velocity of emission, exactly as discussed in connection with Eq. (32) for the case of triodes.

At plate voltages so low or suppressor potentials sufficiently negative that an effective virtual cathode is formed between the screen and suppressor, the plate current tends to be independent of control-grid and

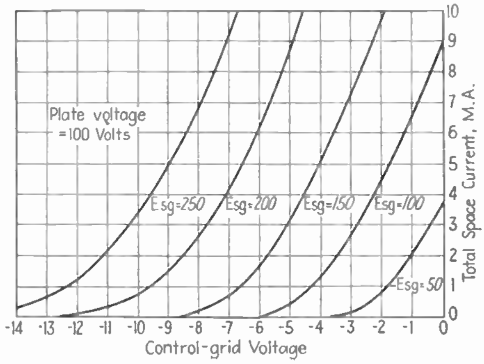


FIG. 39.—Curves showing total space current of a pentode as a function of control-grid potential for various screen potentials. Note the similarity of these curves to those of Fig. 34 for triodes.

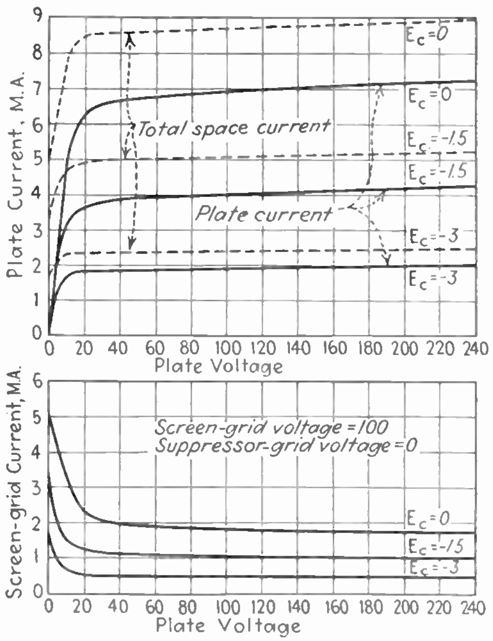


FIG. 40.—Curves showing plate and screen currents and total space current of a pentode as a function of plate voltage for various control-grid potentials. Note that when the plate potential is not too low the plate voltage has relatively little effect on the currents.

screen-grid potentials. The plate current then depends upon the plate voltage, as theory indicates should be the case. In the presence of a virtual cathode the part of the total space current that does not go to the

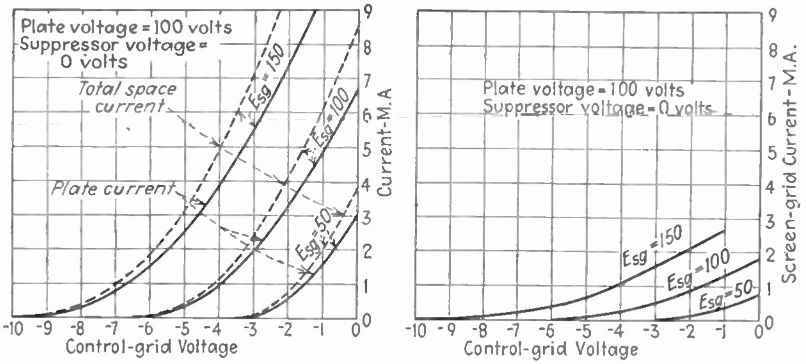


FIG. 41.—Curves showing plate and screen currents and total space current of a pentode as a function of control-grid voltage for various screen potentials. Note that all these curves are of the same general character as those of plate current of a triode (Fig. 34).

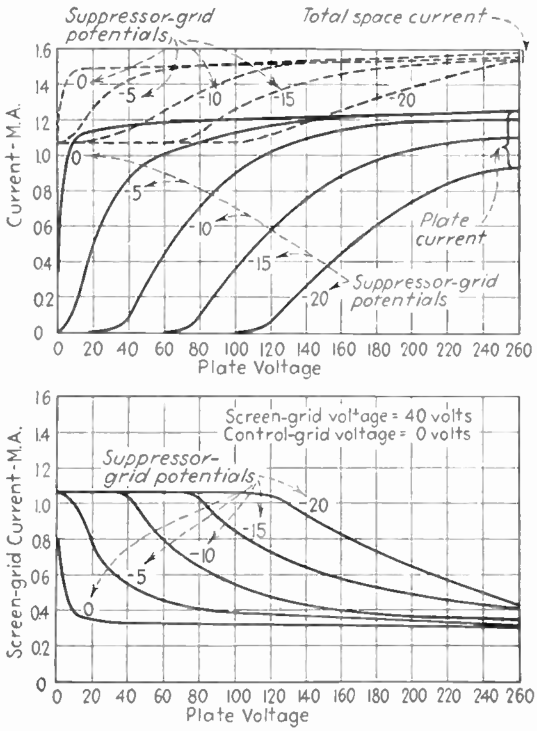


FIG. 42.—Curves showing plate and screen currents and total space current of a pentode as a function of plate voltage for various values of suppressor-grid potential. It will be noted that, for low plate voltages where a virtual cathode forms, the curves of plate current are somewhat like those of a triode as shown in Fig. 35.

plate is divided between the screen and cathode, so that the screen current increases and the total net space current decreases as the plate voltage is reduced. A virtual cathode can be formed even with high plate potentials, provided the suppressor grid is made sufficiently negative. This is illustrated in Fig. 42 and is made use of in certain circuits to give an additional means of controlling the plate current.

28. Screen-grid Tubes.—The screen-grid tube can be thought of as a pentode with the suppressor grid removed. This omission of the suppressor grid makes it possible for secondary electrons to flow between screen and plate. When the plate is at a lower potential than the screen

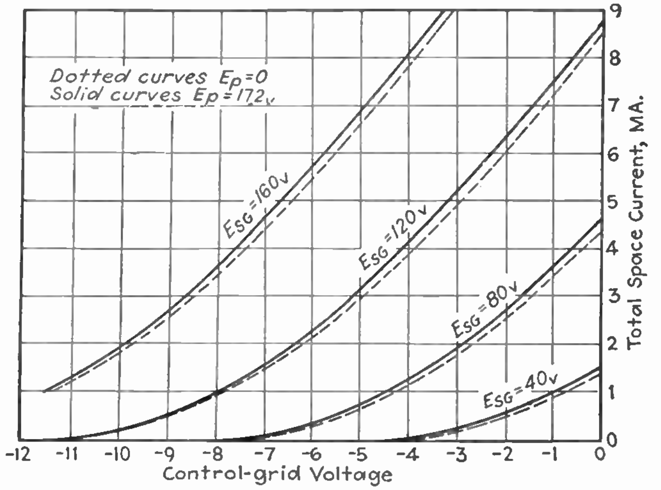


FIG. 43.—Typical curves showing the effect of electrode voltages on the total space current ($i_p + i_{sg}$) of a screen-grid tube. This total current is relatively independent of the plate potential and varies with control- and screen-grid potentials in exactly the same way as the plate current of a triode varies with grid and plate voltages, respectively.

grid, the secondary electrons produced at the surface of the plate by the impact of the primary electrons from the cathode will tend to be drawn to the screen. Similarly, when the plate is at a greater potential than the screen, the secondary electrons produced at the screen will tend to flow to the plate, while the secondary electrons produced at the plate will be attracted back to the plate. This interchange of secondary electrons between plate and screen is superimposed upon the flow of primary electrons from the cathode, and so makes the voltage and current relations of a screen-grid tube differ from those existing in a pentode. In the pentode there can be no such interchange of secondary electrons because the suppressor grid lowers the potential of the space between the screen and plate, as illustrated in Fig. 38, and makes each of these electrodes the most positive thing in its vicinity.

The number of secondary electrons produced at an electrode is proportional to the number of arriving primary electrons, increases as the voltage of the electrode becomes greater, and is particularly sensitive to the surface conditions. Secondary emission commonly becomes appreciable at potentials of 25 to 75 volts, and at these voltages it is not unusual for each primary electron to produce on the average one to two secondary electrons. With surfaces treated in such a way as to enhance secondary emission, as many as 10 secondary electrons may be produced for each primary electron, whereas surfaces prepared to resist secondary emission will on the average have only one secondary electron for perhaps 5 or 10 primary electrons.

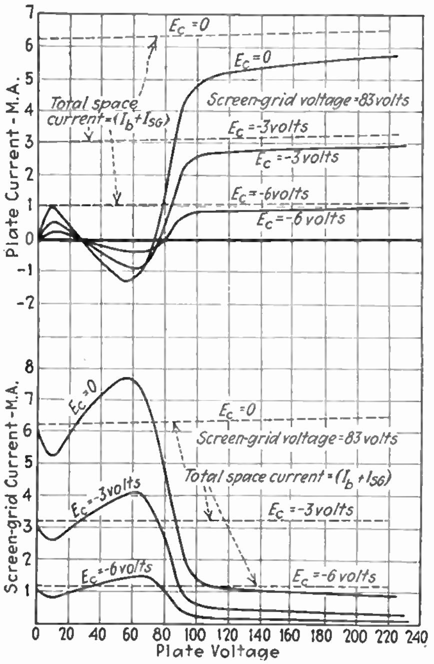


FIG. 44.—Variation of plate and screen-grid currents, and of total space current, with plate voltage (screen-grid voltage constant). It will be noted that changing the control-grid voltage alters the magnitude of the curves without changing their shape (*i.e.*, the control-grid potential affects the total space current but does not alter its division between the plate and screen grid).

Voltage and Current Relations in Screen-grid Tubes.—The voltage and current relations existing in a screen-grid tube can be expressed in terms of characteristic curves such as those given in Figs. 43 to 45. These curves are seen in many respects to be similar to those of pentodes given in Figs. 39 to 41. Thus the total space current of the screen-grid tube is determined primarily by the control-grid and screen-grid potentials and is substantially independent of the plate potential, exactly as in the pentode. This is a result of the fact that the screen grid serves as an electrostatic shield that prevents the plate from producing appreciable electrostatic field at the surface of the cathode. Furthermore, when the plate potential of a screen-grid tube is greater than the screen potential, the plate current is only slightly less than the total space current and is substantially independent of plate potential. The only essential difference between the characteristic curves of pentode and screen-grid tubes occurs when the plate potential is less than the screen voltage. Under these conditions secondary electrons, produced as a result of the primary electrons striking the surface of the plate, are attracted to the screen grid in large numbers. The plate current is thereby reduced to a value less than it would otherwise be.

The effects of secondary electron emission in a screen-grid tube can be understood by studying the way in which the plate and screen currents vary with plate potential, assuming that the screen and control-grid voltages are constant. Such characteristics are shown in Figs. 44 and 45. When the plate is more positive than the screen, the plate receives, in addition to the primary electrons emitted from the cathode, secondary electrons produced at the screen grid. The number of such secondary electrons received by the plate is relatively small, however, since the screen intercepts only a small fraction of the primary electrons and since the secondaries are produced on the cathode side of the screen and there-

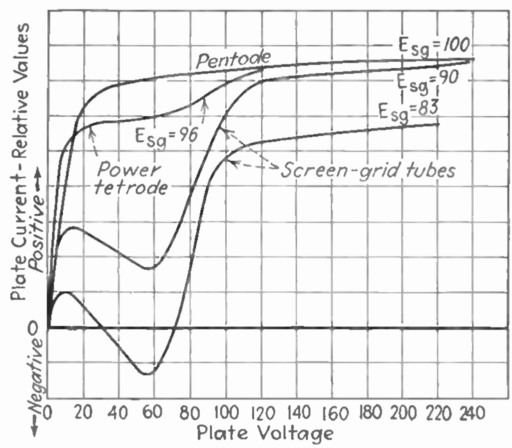


FIG. 45.—Characteristics of screen-grid and pentode tubes showing the effects of secondary emission. The power tetrode is a screen-grid tube in which the plate has a ribbed structure and is treated to reduce the production of secondary electrons.

fore are not under the direct influence of the plate. When the plate potential is reduced until it is less positive than the screen voltage, the situation changes suddenly. Secondary electrons produced at the surface of the plate are now attracted to the more positive screen, so that the actual plate current represents the difference between the number of primary electrons arriving and the number of secondary electrons lost. Very commonly each primary produces on the average more than one secondary electron, in which case the plate current reverses and becomes negative. A still further lowering of the plate potential does not change the number of primary electrons that strike the plate, but since the velocity of impact is now less, the number of secondaries is reduced, and the net plate current is increased accordingly. This increase in plate current with reduction in plate voltage continues until the plate potential becomes so low that a virtual cathode forms in front of the plate. Beyond this point the plate current tends to decrease with reduction in the plate potential.

During these variations of plate potential, the total space current ~~remains constant except at~~ very low plate potentials, when some of the electrons return to the cathode. The variations in screen current are consequently the inverse of the variations in plate current, because the screen tends to receive the current that does not go to the plate.

The actual shape of the plate-current–plate-voltage characteristic of a screen-grid tube for plate potentials less than the screen voltage is dependent upon the tendency for secondary electrons to be produced at the plate. This is illustrated in Fig. 45, which shows characteristics for three screen-grid tubes with different amounts of secondary emission at the plate. The corresponding characteristic of a pentode tube is also shown for comparison. It is seen that the addition of a suppressor grid gives the same characteristic as would be obtained with a screen-grid tube having no secondary emission at the plate.

Dynatron or Negative Resistance Characteristic of Screen-grid Tubes.—It will be noted from an examination of Figs. 44 and 45 that when there is appreciable secondary emission at the plate, there is a region where the plate current increases as the plate voltage is reduced. This represents a negative resistance characteristic, and the tube when used in this way as a negative resistance device is termed a *dynatron.*

29. Beam Tubes.—The beam tube is a special type of screen-grid tube in which the action of a suppressor grid is obtained by accentuating the space-charge effect of the electrons in transit in the space between screen and plate. The space charge required to do this can be obtained by making the distance between screen and plate large, so that many electrons will be in this space at any one instant, and by confining the electrons to a relatively narrow beam to increase their volume density. The resulting potential distribution in the plate-screen space under conditions of large space-charge effect is illustrated in Fig. 46*a*. It will be noted that this potential distribution is characterized by a minimum that persists until the plate potential is very low. This potential minimum arises as a result of the negative space charge being sufficient to bring the potential of the space below the potential of the screen and plate electrodes themselves. The presence of this potential minimum makes each electrode the most positive electrode in its vicinity and thereby prevents secondary electrons from being interchanged between screen and plate. As a consequence, the beam tube has characteristics that are essentially similar to those of a pentode tube.

In the beam tube, however, the transition from the condition where the plate current is independent of plate potential to the condition where the plate current depends on plate voltage is more abrupt than in the pentode and occurs at a lower voltage, as indicated in Fig. 46*b*. The gradual transition of the pentode characteristic arises from the fact that the effect of the suppressor grid is greater next to the grid wires than midway between them, thus giving rise to a form of variable-mu action (see Sec. 31). On the other hand, no corresponding dissymmetry exists in the beam tube because the space charge is uniformly distributed. The beam tube hence realizes what might be termed the ideal pentode characteristic, whereas the actual pentode tube fails to do so because of the non-uniform action of the suppressor grid.

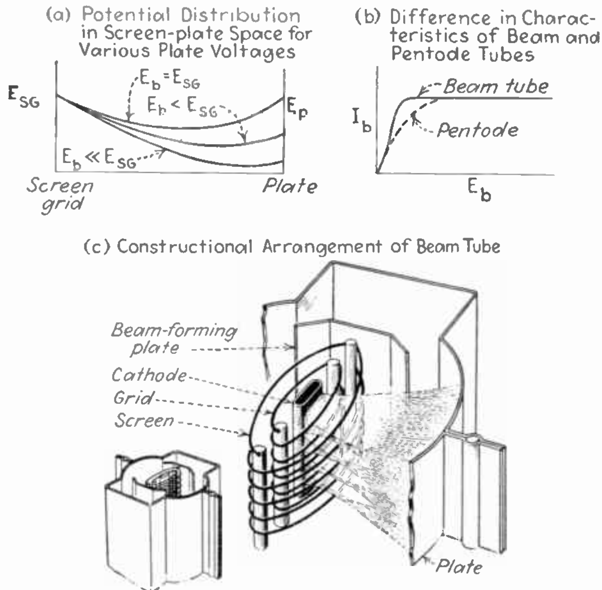


FIG. 46.—Characteristics and constructional features of beam power tube.

The construction of a practical beam tube is shown in Fig. 46*c*. The beam-forming plates are internally connected to the cathode and serve to concentrate the electrons in a beam as indicated. This increases the space-charge effect sufficiently so that, with the large screen-plate distance, the space charge in the screen-plate space will be enough to produce the required potential minimum. The beam-forming plates also serve to keep the electrons away from the edges of the control-grid structure where pronounced dissymmetry exists. The control-grid and screen-grid wires are preferably aligned so that the screen-grid wires are in the shadow cast by the control grid. This reduces to an unusually low value the fraction of the total space current intercepted by the screen.

30. Coefficients of Screen-grid, Beam, and Pentode Tubes.—The important coefficients of screen-grid, beam, and pentode tubes are the mutual conductance (or transconductance), plate resistance, and two amplification factors. The first amplification factor of importance is the constant μ_{sg} which appears in Eq. (36) and which represents the relative effectiveness of the control and screen grids in producing electrostatic field at the surface of the cathode. A definition of this constant in terms of the notation of Eq. (36) is

$$\mu_{sg} = -\frac{de_{sg}}{de_g}\bigg|_{i_p + i_{sg}\text{ constant}} \tag{37}$$

The amplification factor μ_{sg} is analogous to the amplification factor μ of triode tubes and can be termed the "cut-off amplification factor" since it determines the screen- and control-grid potentials giving plate current cut-off. Numerical values of μ_{sg} commonly encountered are in the range 6 to 30.

The second amplification factor of importance in beam, pentode, and screen-grid tubes measures the relative effectiveness of control-grid and plate potentials in controlling the plate current. This quantity is given the symbol μ, and is defined by the relation:

$$\mu = -\frac{de_p}{de_g}\bigg|_{i_p\text{ constant}} \tag{38}$$

where e_p and e_g are plate and control-grid potentials, respectively, and i_p is the plate current. The amplification factor μ is very high for operating conditions that make the plate current substantially independent of plate voltage. Values ranging from 100 to over 1000 are common, with the exact magnitude depending upon the electrode voltages and upon the extent to which the electron stream in the vicinity of the cathode is shielded from stray electrostatic effects of the plate electrode.

Plate Resistance.—The plate resistance of beam, pentode, and screen-grid tubes is defined in the same way as for triodes, *i.e.*, it is the resistance that the plate circuit offers to an increment of plate potential. Thus it is defined by the equation

$$\text{Plate resistance} = R_p = \frac{\partial e_p}{\partial i_p} \tag{39}$$

The plate resistance is seen to be the reciprocal of the slope of the plate-voltage–plate-current curve. Under operating conditions for which the plate current is substantially independent of plate potential the plate resistance is very high, normally exceeding 1 megohm in pentodes that have the plate completely shielded electrostatically from the electron stream in the vicinity of the cathode.

Mutual Conductance (or Transconductance).—The mutual conductance (or, as it is sometimes called, the transconductance) of beam, pentode, and screen-grid tubes is defined in the same way as in triodes and is given by the equation

$$\text{Mutual conductance} = G_m = \frac{\partial i_p}{\partial e_g} = \frac{\mu}{R_p} \tag{40}$$

where

i_p = plate current
e_g = control-grid voltage
μ = amplification factor given by Eq. (38)
R_p = plate resistance as given by Eq. (39).

The mutual conductance represents the rate of change of plate current with control-grid voltage. The mutual conductance is the most important single constant of screen-grid, beam, and pentode tubes when operated in the usual manner with sufficient plate voltage to make the plate current substantially independent of plate voltage.

Miscellaneous Constants. Screen-grid beam, and pentode tubes possess numerous other constants which may under special circumstances be useful in expressing properties of the tube. Thus each positive electrode has its own dynamic resistance defined in the same way as the plate resistance except that the expression is in terms of the voltage and current of the electrode involved. Likewise, each electrode possesses a mutual conductance or transconductance with respect to every other electrode. Finally, there are numerous amplification factors, each of which is defined in terms of the relative effectiveness of some particular pair of electrodes upon some current in the tube. Thus in the pentode tube one could define amplification factors giving the relative effectiveness of control-grid and suppressor-grid potentials upon the plate current, upon the screen-grid current, and upon the total space current $I_p + I_{sg}$.

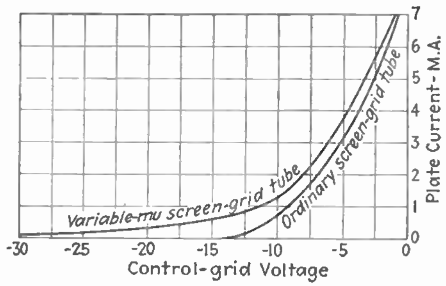


FIG. 47.—Characteristic curve of a typical variable-mu tube compared with the characteristic curve of a corresponding screen-grid tube of ordinary construction.

31. Variable-mu Tubes.—Variable-mu tubes (also called remote cut-off tubes and supercontrol tubes) are tubes in which the design has been modified in such a way as to cause the total space current of the tube to taper off at very negative control-grid potentials rather than to have a well-defined cut-off point. The characteristic curve of a typical variable-mu screen-grid tube is shown in Fig. 47, together with the corresponding characteristic of an ordinary sharp-cut-off screen-grid tube.

A variable-mu characteristic is obtained by using a non-uniform control-grid structure so that the amplification factor is different for different parts of the tube. Such an arrangement causes the various parts of the tube to reach cut-off with different grid-bias voltages, so that over-all cut-off comes gradually rather than abruptly. The usual method of obtaining the variable-mu action is illustrated in Fig. 48 and consists in varying the pitch of the control-grid structure.

The variable-mu principle may be applied to any tube. It is, however, normally used only in the small pentode and screen-grid tubes designed for voltage amplifier service. In such cases it is the amplification factor μ_{sg} that has the variable-mu characteristic.

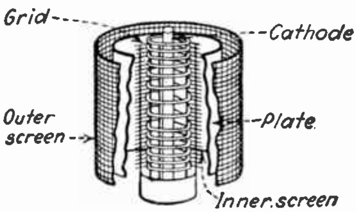


FIG. 48.—Cut-away illustration of a variable-mu screen-grid tube, showing the variable pitch of the control-grid structure that gives the variable-mu characteristic.

Variable-mu tubes are used where it is desired to control the amplification by varying the control-grid potential. The principal advantage of such tubes for this purpose over tubes having sharp cut-off is that by stretching out the part of the characteristic having low mutual conductance, the rate of change of curvature in this region, and hence the tendency to produce cross-talk and distortion, is greatly reduced.

32. Effect of Positive Control Grid.—Throughout the discussion given above for different types of tubes it has been assumed that the control grid is operated at a negative potential. If the control grid becomes positive, the total space current is still determined by the strength of the electrostatic field at the surface of the cathode, just as with the grid negative, but part of this current is diverted away from the other positive electrodes to the control grid. Hence Eqs. (32) and (36) become

For triodes:

$$I_b + I_c = K\left(E_c + \frac{E_p}{\mu}\right)^{3/2} \tag{41}$$

For screen-grid, beam, and pentode tubes:

$$I_b + I_{sg} + I_c = K\left(E_c + \frac{E_{sg}}{\mu_{sg}}\right)^{3/2} \tag{42}$$

where I_c is the control-grid current, and the remaining notation is the same as before.

The division of this total space current between the control grid and the remaining electrodes depends upon the type of tube and the electrode potentials. With triodes the grid current ordinarily increases as the control grid becomes more positive. The grid current is small, however, until the grid potential exceeds the plate voltage, when the control-grid current rises abruptly as a result of secondary electrons captured from the plate. The grid current also becomes greater the lower the plate potential. This is particularly true when the plate voltage is so low that there is a tendency for a virtual cathode to form in front of the plate. The characteristic curves of a typical triode tube in the positive-grid region are shown in Fig. 49.

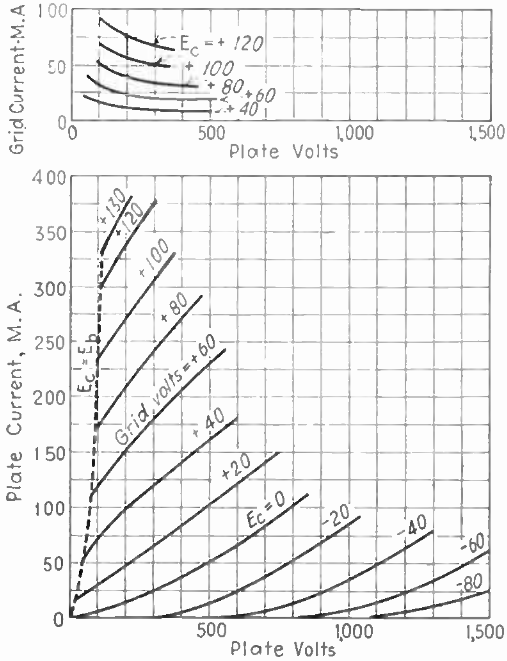


FIG. 49.—Characteristics of a small power triode (Type 800) in the positive control-grid region. It will be noted that the control-grid current tends to increase as the plate potential is reduced.

33. Effect of Gas upon Tube Characteristics.—Even very small traces of gas within a tube will modify the characteristics and behavior markedly. This is a result of positive ions produced by collision between electrons and the gas molecules. These positive ions are attracted toward the cathode and bombard it with sufficient intensity to damage the cathode unless the ion velocity is low. Some of the positive ions are also attracted to the negative grid, and cause power loss in the control-grid circuit. The positive ions furthermore produce a positive space charge that tends to neutralize the negative space charge of the electrons in the vicinity of the cathode. This positive space-charge action is considerable even when only small traces of gas are present, because the low velocity of the positive ions resulting from their large mass causes each positive ion to spend a considerable length of time in the tube before being collected by the cathode or grid.

Tubes in which gas effects are appreciable are commonly said to be "soft." When the amount of gas present is considerable (*i.e.*, when the tube is very soft) the ionization will produce a luminous glow in the gas.

Hot-cathode Gas Diodes.—When the gas pressure in a tube is of the order of 1 to 30 times 10^{-3} mm of mercury, as is the case when the tube contains mercury vapor in equilibrium with liquid mercury at room temperatures, the presence of the gas profoundly affects the characteristics. In the case of a diode, the plate current starts to increase with plate voltage in exactly the same way as in the high-vacuum tube of Fig. 31, but at some critical potential there is a sudden break and the current increases to the full cathode emission with little or no increase in the plate voltage, as shown in Fig. 50. This occurs when the plate voltage reaches the ionizing potential of the gas. The space-charge effect of the resulting positive ions is then sufficient to neutralize completely the space charge of the negative electrons around the cathode. The result is that the full emission current can be drawn to the plate with just enough plate potential to keep the ionization process functioning. This sort of characteristic finds practical application in hot-cathode mercury-vapor rectifier tubes, and is discussed in Sec. 86.

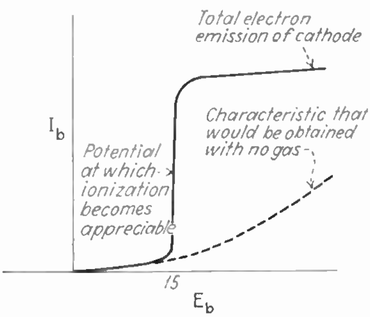


FIG. 50.—Characteristic of hot-cathode mercury-vapor diode showing how the space-charge limitation on the space current is removed as soon as ionization begins.

34. Constructional Features of Small Tubes.—The construction of typical small glass-envelope tubes is shown in Fig. 51. The electrode assembly is supported by wires held in the glass stem by means of the press. The individual electrodes, such as plate, grid, and cathode assembly, are separately fabricated and spot welded to these support wires. Proper spacings are maintained by means of punched mica disks. After this stem assembly has been completed, the glass bulb is sealed to the stem, evacuated, and based.

The metal-envelope tube differs from the glass tube primarily in that the outer envelope consists of a metal shell rather than a glass bulb. This change alters the manufacturing technique in many important respects, but the internal assembly of cathode, grids, and plate is essentially the same irrespective of the type of envelope employed. The principal features of a metal-envelope tube are shown in Fig. 52. Instead of being mounted upon the press, the various electrodes are supported from a metal "header," as shown. The lead-in wires are brought out through glass beads that are sealed to kovar eyelets, which in turn are welded to the header. Kovar is an alloy that has substantially the same coefficient of expansion as the glass bead and so makes possible a vacuum-tight joint. After the electrode structure has been completely assembled upon the header, the metal shell is welded to the header as indicated in Fig. 52, after which the tube is evacuated, sealed off, and based.

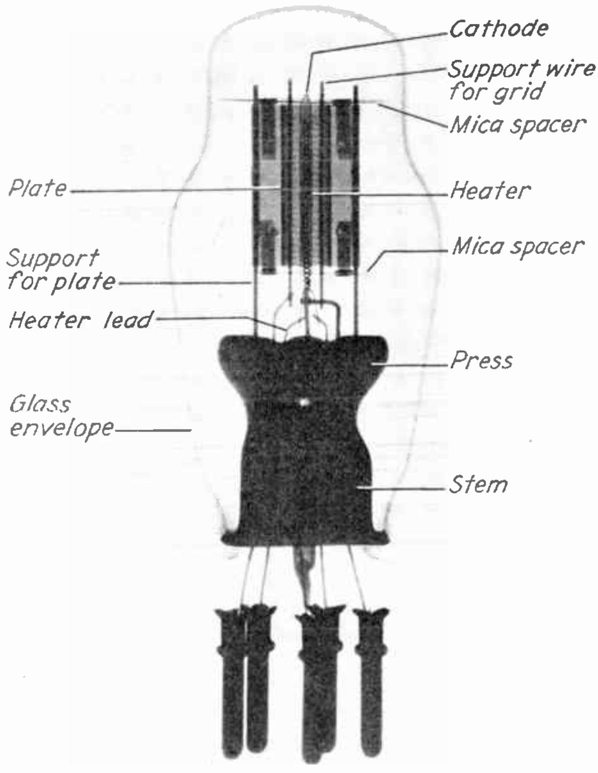


FIG. 51a.—An x-ray photograph of heater-type triode with glass envelope.

Evacuation of Tubes.—The exact details of the evacuation procedure depend upon the type of tube. In the small tubes commonly used in radio receivers a rough vacuum is obtained by means of a motor-driven pump. The final vacuum is then produced by volatilizing a small quantity of some substance, called a "getter," inside the tube to remove the residual gas either by chemical or by mechanical action. Magnesium is widely used for this purpose, but other materials such as barium, phosphorus, etc., can also be employed, as well as various mixtures. Just

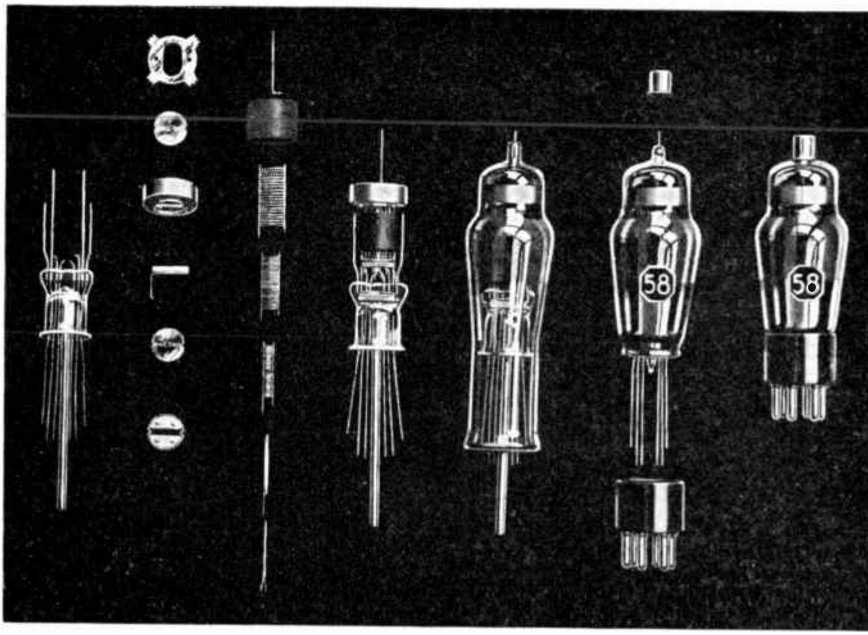


FIG. 51b.—Photographs showing constructional features of small glass-envelope pentode tube, having heater cathode and variable-mu control grid.

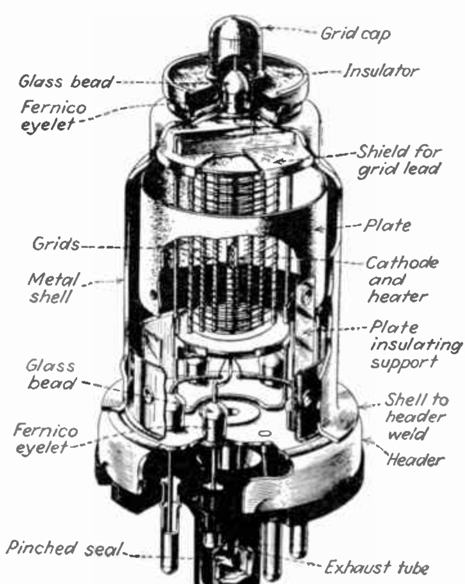


FIG. 52.—Cut-away drawing showing constructional features of a metal-envelope pentode tube.

before the getter is "flashed" the metal parts of the tube are brought to a red heat to drive out absorbed and adsorbed gas. With glass tubes this is usually accomplished with the aid of an induction furnace, while with metal tubes a gas flame playing on the metal shell is employed.

In larger tubes such as used in radio transmitters, and in some types of small tubes, notably many of those used in telephone work, the final vacuum is obtained with the aid of a molecular pump backed up by the motor-driven pump. The "getter" can then be omitted, making it possible to operate at glass temperatures that would volatilize the "getter" and destroy the vacuum. In such tubes the gas held by the glass parts is removed during evacuation by baking the entire tube at a temperature just below the softening point of glass.

TABLE V.—CHARACTERISTICS OF TYPICAL VACUUM TUBES USED FOR VOLTAGE AMPLIFICATION

Type	Equivalent tubes*	Heater or filament			Normal electrode rating					Properties		
		Type	Voltage	Current	Plate volts	Grid volts	Screen volts	Screen ma	Plate ma	G_m	R_p	μ
Triode Tubes												
56	76	Heater	2.5	1.00	250	−13.5	.....	...	5.0	1,450	9,500	13.8
30		Filament	2.0	0.06	180	−13.5	.....	...	3.1	900	10,300	9.3
2A6§	75§, 6Q7§†	Heater	2.5	0.8	250	− 2.0	.....	...	0.8	1,100	91,000	100
6C5†	....	Heater	6.3	0.3	250	− 8.0	.....	...	8.0	2,000	10,000	20
6F5†	....	Heater	6.3	0.3	250	− 2.0	.....	...	0.9	1,500	66,000	100
55§	85§, 27	Heater	2.5	1.0	250	−20.0	.....	...	8.0	1,100	7,500	8.3
Screen-grid Tubes												
24A	....	Heater	2.5	1.75	250	−3	90	1.7	4.0	1,050	0.6 meg.	630
32	....	Filament	2.0	0.06	180	−3	67.5	0.4	1.7	650	1.2 meg.	780
35‡	....	Heater	2.5	1.75	250	−3	90	2.5	6.5	1,050	0.4 meg.	420
Radio-frequency Pentode Tubes												
6C6	57, 6J7†, 77	Heater	6.3	0.3	250	−3	100	0.5	2.0	1,225	over 1.5 meg.	over 1,840
6D6‡	78‡, 6K7‡†, 58‡	Heater	6.3	0.3	250	−3	100	2.0	8.2	1,600	0.8 meg.	1,280
34‖‡	....	Filament	2.0	0.06	180	−3	67.5	1.0	2.8	620	1.0 meg.	620
39/44‖	....	Heater	6.3	0.3	250	−3	90	1.4	5.8	1,050	1.0 meg.	1,050
6B7§	2B7§	Heater	6.3	0.3	250	−3	125	2.3	9.0	1,125	0.65 meg.	730

* These tubes have similar (but not necessarily identical) Normal Electrode Ratings and Properties, and are suitable for performing the same functions.

† Metal envelope.

‡ Variable mu.

§ Tube with diode sections.

‖ Suppressor grid internally connected to cathode.

35. Table of Tube Characteristics.—The essential characteristics of different types of vacuum tubes in common use are given in Tables V, IX, X, XIV, and XV. The first table covers tubes used for voltage amplification and other similar purposes. These tubes may be classified as triodes, radio-frequency pentodes, and screen-grid tubes, and may be either of the sharp-cut-off or variable-mu type. Small power tubes such as are used in radio receivers are listed in Table IX (Chap. VI), and large power tubes, such as are used in radio transmitters, are covered in Table X (Chap. VI). Rectifier tubes are given in Tables XIV and XV (Chap. X).

36. The Mathematical Representation of Characteristic Curves of Tubes.—In carrying out the analysis of circuits involving vacuum tubes, it is often desirable to be able to express the characteristic curves of the tubes by means of mathematical expressions. The principal methods that have been employed to do this are the power-law method, and the power-series method.

Power-law Method of Expressing Tube Characteristics.—This method of representing tube characteristics has already been made use of in Eqs. (31), (32), (36), (41), and (42), which for the sake of convenience will be rewritten below in slightly modified form.

For diodes:

$$I_b = KE_b^{\alpha} \tag{43}$$

For triodes:

$$I_b + I_c = K\left(E_c + \frac{E_b}{\mu}\right)^{\alpha} \tag{44}$$

For screen-grid and pentode tubes:

$$I_b + I_{sg} + I_c = K\left(E_c + \frac{E_{sg}}{\mu_{sg}}\right)^{\alpha} \tag{45}$$

In pentodes and screen-grid tubes when the plate potential is sufficiently high to make the plate current substantially independent of plate voltage, the plate current is almost exactly proportional to the total space current, so that for these conditions one has

$$I_b = K\left(E_c + \frac{E_{sg}}{\mu_{sg}}\right)^{\alpha} \tag{46}$$

The notation in Eqs. (43) to (46) is the same as that used when they were first developed, with the addition that α is a constant, usually very close to $3/2$, but which may vary somewhat with electrode voltages when there are dissymmetries in the tube (*i.e.*, when μ and μ_{sg} are not true geometrical constants).

It is assumed in Eqs. (43) to (46) that the velocity with which the electrons are emitted from the cathode, and also the contact potentials in the grid and plate circuits, are negligibly small, and that there is no voltage drop in the filament. It is also assumed that there is a full space charge about the cathode and that an electron which has once been drawn out of this space charge will not return to it. Hence the equations do not necessarily hold when a virtual cathode is formed somewhere within the tube.

Power-series Method of Representing Characteristic Curves of Tubes.—In the power-series method, the tube characteristics are expressed in terms of a Taylor's series, or power series. The details of this method can be understood by applying it to the case of a triode with equipotential cathode and assuming that over a limited range about an operating point where the plate and grid voltages and plate current are E_b, E_c, and I_b, respectively, one can write

$$I_p = f\left(E_g + \frac{E_p}{\mu}\right) \tag{47}$$

where I_p is the change in plate current produced by changes E_g and E_p in grid and plate voltages, respectively, and μ is constant over the range of variation represented by E_g and E_p. Expansion of Eq. (47) into a Taylor's series gives

$$I_p = a_1\left(E_g + \frac{E_p}{\mu}\right) + a_2\left(E_g + \frac{E_p}{\mu}\right)^2 + a_3\left(E_g + \frac{E_p}{\mu}\right)^3 + \cdots \tag{48}$$

where

$$a_1 = G_m = \frac{\mu}{R_p}$$

$$a_2 = \frac{1}{2!}\frac{\partial G_m}{\partial E_g} = -\frac{\mu}{2R_p{}^2}\frac{\partial R_p}{\partial E_g}$$

$$a_3 = \frac{1}{3!}\left(\frac{\partial^2 G_m}{\partial E_g{}^2}\right)$$

Equation (48) expresses the characteristics of the vacuum tube about the operating point E_b, E_c, and I_b at which the a's are evaluated and is exact, provided the plate current is not cut off and is less than the saturation value, and provided μ may be considered constant over the range of E_g and E_p involved. In actual practice, the series converges so rapidly that one, two, or at most three terms are sufficient to explain many of the important aspects of tube behavior.

Problems

1. In a two-electrode tube, the anode potential is +200 volts and the anode current is 75 ma. How many electrons arrive at the anode each second?

2. When the tube in Prob. 1 consists of plane cathode and plate electrodes, it is found that when there is a strong magnetic field between the electrodes oriented so that the flux lines are parallel with the plane electrodes, the anode current becomes zero, but, if the flux lines are perpendicular to the plane electrodes, there is little or no effect on the plate current. Explain.

3. In a certain water-cooled tube the cathode consists of a tungsten filament 19.5 in. long, and 0.025 in. in diameter. Calculate and plot the electron emission as a function of temperature over the range 2200 to 2600°K., if the constants A and b in Eq. (30) are 60.2 and 52,400, respectively.

4. Discuss the economic factors involved in selecting the cathode operating temperature of a tube employing a tungsten cathode.

5. The useful life of a tungsten emitter is normally terminated by burnout of the filament. This is not the case with thoriated-tungsten and oxide-coated emitters, however. Explain the reasons for the difference.

6. In tubes employing thoriated-tungsten emitters it is found that accidental overloading of the tube may cause the filament emission to drop to a low value. The emission lost in this way can often be restored by operating the filament for some time at slightly above normal temperature. Explain, and also give the reasons that oxide-coated and tungsten emitters do not act in this manner.

7. How much power is dissipated at the anode in the tube of Prob. 1?

8. In a two-electrode tube it is found that at a plate voltage of +100 volts, the plate current with full space charge is 90 ma. What plate voltage will be required to produce a plate current of 45 ma?

9. When full space-charge conditions exist in a two-electrode tube only those electrons which have the highest velocities of emission reach the anode. Explain how this comes about.

10. In a two-electrode tube a sine wave alternating voltage instead of a direct-current voltage is applied to the plate. Sketch the resulting wave of plate current as a function of time.

11. Using the plate-current–plate-voltage characteristic curve of a two-electrode tube obtained from a tube manual (or assigned), check the extent to which Eq. (31) holds by plotting I_b versus E_b on logarithmic paper.

12. In a particular triode tube having an amplification factor of 8, the plate current is 10 ma when $E_b = 250$ volts and $E_c = -15$ volts. What will be the current when $E_b = 200$ volts and $E_c = -5$ volts, assuming that full space charge is maintained at all times?

13. Explain how one could plot an entire family of grid-voltage–plate-current curves such as shown in Fig. 34, knowing the amplification factor of the tube and having available one curve of the family.

14. A triode tube has an amplification factor of 13 and is operated at a plate potential of 275 volts. What grid potential is required to reduce the plate current barely to zero?

15. In a particular tube it is found that the increase in plate current resulting from 40 volts increase in plate potential can be eliminated by making the grid 3 volts more negative. What is the amplification factor of the tube?

16. What is the amplification factor of the tube of Figs. 34 and 35 for an operating point in the region $E_c = -20$, $E_b = 200$?

17. What is the plate resistance for the tube of Figs. 34 and 35 for the operating point $E_c = -20$, $E_b = 200$?

18. Describe how the mutual conductance of a tube at a particular operating point can be deduced from characteristic curves such as those of Fig. 34.

19. Two identical triode tubes are connected in parallel. How do the plate resistance, amplification factor, and mutual conductance of the combination compare with the values for the individual tubes?

20. It is desired to operate a direct-current relay in the plate circuit of a triode by the application of a small d-c voltage to the grid. Two tubes are available. One has an amplification factor of 3 and a plate resistance of 2100 ohms, the other has an amplification factor of 12 and a plate resistance of 9500 ohms. Which is preferable from the standpoint of sensitivity if the resistance of the relay is negligible?

21. What will be the change in plate current through the relay with each tube in Prob. 20 for a change in d-c grid voltage of 0.35 volt?

22. Show that the plate current of a triode with negative grid is proportional to $(E_b + \mu E_c)^{3/2}$.

23. Fill in the blanks in the following table:

	μ	E_b = 250 volts				E_b = 180 volts	
		R_p ohms	G_m µmhos	E_c volts	I_b ma	E_c volts	I_b ma
Triode 1	3.5		2,125	−50		−31.5	31
Triode 2	35	11,300		−5.0	6.0	−6.5	
Triode 3		9,500	1,450	−13.5	5.0		3.3
Triode 4			1,600	−22	12.6	−16.5	12.6

24. In the usual pentode tube, the suppressor grid has a rather coarse mesh. Discuss the effect of this on the tendency to form a virtual cathode as compared with a suppressor with fine mesh.

25. If the suppressor grid of a pentode were made positive so that it drew current, how would it be necessary to modify Eq. (36)?

26. Why is it that the suppressor-grid potential does not appear in Eq. (36)?

27. What would be the effect on the curves of Figs. 40 and 41 if the filament temperature were reduced to the point where saturation occurred at about 5 ma total space current? In answering this question include a sketch indicating the behavior when saturation exists.

28. *a.* In a screen-grid tube what would be the effect of treating the screen-grid surfaces in such a manner as to enhance greatly the secondary emission at the screen?

b. What would be the effect if the tube in (*a*) were a pentode?

29. Explain how the electrodes of a pentode tube could be connected so that the resulting tube would have characteristics corresponding to: (*a*) a diode, (*b*) a triode, and (*c*) a screen-grid tube.

30. If you were given a tube that might be either a pentode with the suppressor internally connected to the cathode, or a screen-grid tube, but you did not know which, what electrical tests could be made to determine which kind of tube it was?

31. Is it possible for a virtual cathode to form in the screen-anode space of a beam tube?

32. Evaluate the coefficients G_m, R_p, and μ of the pentode tube of Figs. 40 and 41 in the vicinity of $E_c = -1.5$, $E_{sg} = 100$, $E_b = 150$.

33. Evaluate μ_{sg} for the tubes of Figs. 41 and 43.

34. The amplification factor μ_{sg} that expresses the relative effectiveness of screen and control-grid potentials on the *total space* current of screen-grid, beam, and pentode tubes is a geometrical constant substantially independent of electrode voltages. The analogous amplification factor μ that takes into account the relative effectiveness of plate and control-grid potentials upon the *plate* current is not a geometrical constant, however, but rather depends upon the electrode potentials. Explain why this is.

35. Evaluate the negative resistance in the tube of Fig. 44 when $E_c = 0$, $E_{sg} = 83$, and $E_b = 40$.

36. There are no variable-mu triodes manufactured. Explain how to obtain a tube having the characteristics of a variable-mu triode by a proper connection of the electrodes of some standard tube that is manufactured.

37. Describe a construction other than that shown in Fig. 48 by which a variable-mu characteristic could be obtained.

38. Under some conditions it is found that the current drawn by a positive control grid is negative. Explain how this can happen in a high-vacuum tube and state the conditions favorable to the production of a negative control-grid current.

39. Tubes are commonly tested for the presence of gas by observing the grid current when the grid is negative and the plate (and also screen in tubes having a screen) is positive. Explain why this is a test for gas and why the grid current observed will be proportional to the space current in the tube and to the gas pressure.

40. In evacuating a tube it is essential that all metal and glass parts be heated during the exhaust period to temperatures higher than will be reached during normal operation. Explain the reason for this requirement.

41. *a.* If the plate current of a diode were exactly proportional to the plate voltage, what would be the value of the exponent α in Eq. (43)?

b. If the curve of plate current plotted as a function of anode voltage in a diode is a section of a parabola, what value would the exponent α in Eq. (43) have?

42. Write Eq. (48) for the following special cases, omitting the a coefficients that are zero: (1) when the plate current I_p is exactly proportional to $\left(E_g + \frac{E_p}{\mu}\right)$ in the vicinity of the operating point, and (2) when the plate current I_p in the vicinity of the operating point varies in such a way that the curve of I_p as a function of $\left(E_g + \frac{E_p}{\mu}\right)$ is a section of a parabola.

43. Write an equation analogous to Eq. (48) but giving the relation between a change E_p in the plate voltage and the resulting change I_p in the plate current for a diode about an operating point where the plate voltage and current are E_b and I_b, respectively.

CHAPTER V

VACUUM-TUBE AMPLIFIERS

37. Vacuum-tube Amplifiers.—A vacuum tube is able to function as an amplifier because of the fact that a voltage representing little or no energy applied to the grid of the tube is able to control a comparatively large plate current that represents appreciable energy in the plate circuit.

Classification of Amplifiers.—Amplifiers are classified in ways descriptive of their uses and properties. Thus a voltage amplifier is an amplifier arranged to develop the maximum possible amplified voltage. On the other hand, a power amplifier has as its objective the development of as much energy as possible without regard to voltage.

Another common basis of classifying amplifiers is according to the frequency to be amplified. This leads to the broad divisions known as audio-frequency, radio-frequency, video-frequency, and direct-current amplifiers. Audio-frequency amplifiers are intended for amplifying currents of audible frequencies, *i.e.*, from about 15 cycles per second to approximately 10,000 cycles. Frequencies higher than 10,000 to 15,000 cycles per second are considered as radio frequencies, while video frequencies are those contained in television signals, and commonly range from 10 cycles up to over 1,000,000 cycles.

Amplifiers, particularly power amplifiers, are also designated as Class A, Class AB, Class B, Class C, or linear amplifiers according to the operating conditions. The term Class A is applied to an amplifier adjusted so that the plate current flows continuously throughout the cycle of the applied voltage. This is the type of amplifier considered in this chapter. The remaining types of amplifiers are adjusted so that the plate current flows intermittently in a succession of pulses. Amplifiers of the latter types are considered in the next chapter.

38. Distortion in Amplifiers.—An ideal amplifier produces an output wave form that exactly duplicates the input wave form in all respects except magnitude. An actual amplifier can fall short of this ideal by failing to amplify the different frequency components of the input voltage equally well, by introducing new frequency components not present in the input voltage, or by making the relative phases of the different frequency components in the output differ from the relative phase relations existing in the input voltage. These effects are commonly referred to as frequency, non-linear (or amplitude), and phase distortion, respectively.

Frequency distortion limits the range of frequencies that a particular amplifier can handle satisfactorily and so is one of the most important factors that must be considered in connection with amplifier design. A typical example of frequency distortion is shown in Fig. 53*b*, where the high-frequency component of the original signal is discriminated against as a result of being amplified less than is the low-frequency component.

Non-linear distortion limits the output voltage or power that it is practicable to obtain from an amplifier by causing the output to contain frequency components not present in the input wave when the amplitude of the signal is large. An example of non-linear distortion is illustrated in Fig. 53*c*, where the negative half cycles are amplified less than the positive half cycles.

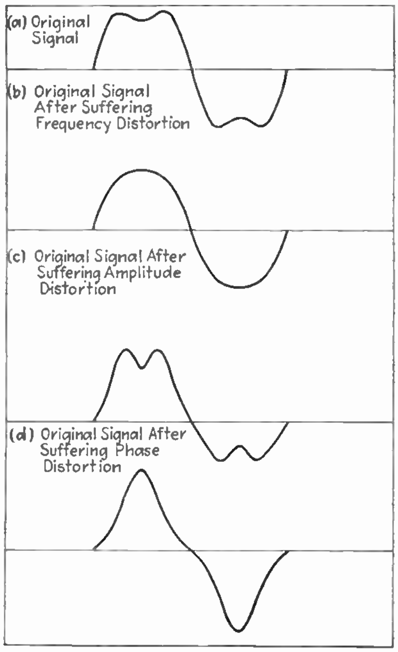


FIG. 53.—Series of waves showing the effects produced by frequency, non-linear, and phase distortion.

Phase distortion results whenever the different frequencies present in the input wave are not passed through the amplifier in the same amount of time. It can be shown that phase distortion exists if the phase shift of the amplifier between input and output when plotted as a function of frequency is not a straight line that passes through zero frequency at some integral multiple of π. When phase distortion exists the input and output wave shapes may appear quite different, as illustrated by Fig. 53*d*, even though the output contains the same frequency components in the same relative amplitude as the input. Under many conditions the phase distortion encountered in amplifiers is of no practical importance. This is particularly true when the ultimate amplified output is to be reproduced in the form of sound, since the ear is not sensitive to phase relations.

39. The Amplifier Circuit.—Typical amplifier circuits are shown in Fig. 54. The signal voltage that is to be amplified is applied to the control grid of the tube in series with a voltage E_c. This voltage has a polarity such as to make the grid negative with respect to the cathode in the absence of an applied signal, and is termed the grid-bias or C-voltage. The variations in the plate current caused by the signal voltage flow through an impedance Z_L that is in series with the plate-supply voltage. This impedance is commonly termed the load impedance, and the energy dissipated in it by the variations in plate current represent the amplified energy.

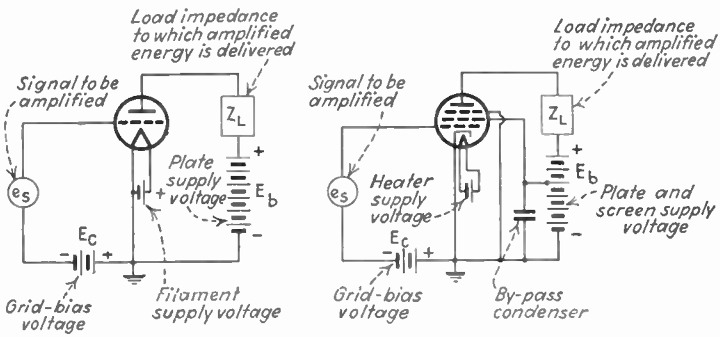


FIG. 54.—Basic circuits of vacuum-tube amplifiers employing triode and pentode tubes.

In practical amplifiers it is usually desirable to obtain the bias voltage from the plate-supply potential rather than from a separate battery. This can be accomplished as shown in Fig. 55, by connecting a resistance between the cathode and ground and by-passing this resistance with a

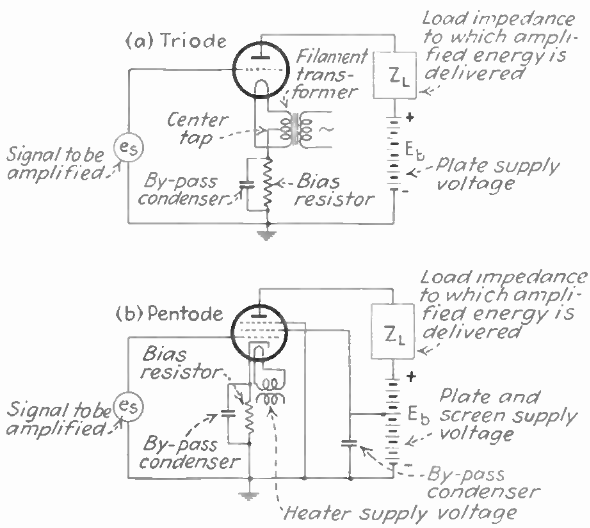


FIG. 55.—Vacuum-tube amplifiers similar to those of Fig. 49 except that the control grid is made negative with respect to the cathode by a self-bias resistance instead of a battery.

condenser large enough to be an effective short circuit to alternating currents of the lowest frequency to be amplified. Such an arrangement places the cathode at a positive potential with respect to the control grid and so gives the equivalent of a negative bias for the grid. The amount of bias obtained is the voltage drop of the *total* space current flowing through the bias resistance and is controlled by the amount of resistance. Further discussion of bias arrangements is given in Sec. 85.

40. Equivalent Circuit of Vacuum-tube Amplifier.—The variations produced in the plate current of a vacuum tube by the application of a signal voltage e_s to the control grid are exactly the same variations that would be produced in the plate current by a generator developing a voltage $-\mu e_s$ acting inside the tube from cathode toward the plate, where μ is the amplification factor as defined by Eq. (33) or (38). This voltage acts in a circuit consisting of the plate resistance of the tube in series with the load impedance. *The effect on the plate current of applying a signal voltage e_s to the grid is exactly as though the plate-cathode circuit of the tube were a generator developing a voltage $-\mu e_s$ and having an internal resistance equal to the plate resistance of the tube.* This leads to the equivalent circuit of the vacuum-tube amplifier shown in Fig. 56*b*, which is the basis of most amplifier designs and calculations.[1]

The plate current I_p resulting from the application of a signal potential E_s to the control grid, is found from the equivalent circuit to be:[2]

$$I_p = \frac{-\mu E_s}{R_p + Z_L} \tag{49a}$$

The voltage that this current develops across the load is

$$\text{Voltage across load} = I_p Z_L = \frac{-\mu E_s Z_L}{R_p + Z_L} \tag{49b}$$

The signs are such that a positive value for I_p means a current flowing in opposition to the steady direct current present in the plate circuit when no signal is applied.

It is sometimes convenient to rearrange Eqs. (49*a*) and (49*b*) by dividing both numerator and denominator of the right-hand side by R_p. Doing this gives

$$I_p = \frac{-\mu}{R_p} \frac{E_s}{1 + \frac{Z_L}{R_p}} = -G_m E_s \frac{R_p}{R_p + Z_L} \tag{50a}$$

$$\text{Voltage across load} = -G_m E_s \frac{R_p Z_L}{R_p + Z_L} \tag{50b}$$

[1] A rigorous proof of the equivalent amplifier circuit of the vacuum tube is given in Sec. 51.

[2] In Eq. (49*a*) and other equations to follow, capitals are used to indicate vector quantities, while small letters indicate either an instantaneous quantity, or a quantity that may be either instantaneous or vector as desired. Thus in Fig. 56, e_s can be either the instantaneous signal voltage or its vector, whichever is more convenient, while in Eq. (49*a*) E_s means the vector corresponding to the signal voltage.

Here $G_m = \mu/R_p$ is the mutual conductance of the tube. This form of the tube equations shows that the *effect of applying a signal voltage e_s to the control grid is also the same as though the tube generated a current $-e_sG_m$ flowing from plate toward the cathode through an impedance formed by the plate resistance in parallel with the load impedance.*

This leads to an alternative form of equivalent circuit, which is shown in Fig. 56*c* and which can be called *the constant-current generator* form of the equivalent circuit, as contrasted with *the constant-voltage generator* form shown in Fig. 56*b*. These two are equivalent, but the constant-voltage generator form is commonly most convenient to use when triodes are involved, while the constant-current generator form is preferable for pentode, beam, and screen-grid tubes (where the plate resistance is much higher than the load resistance).

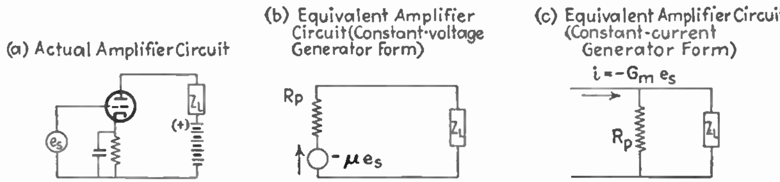


FIG. 56.—Equivalent circuits of the vacuum-tube amplifier. In the constant-voltage form the effect that is produced in the plate circuit by the signal e_s acting on the grid is taken into account by postulating that the plate circuit can be replaced by an equivalent generator of voltage $-\mu e_s$ acting from cathode toward the plate and having an internal resistance equal to the plate resistance R_p. In the constant-current form the tube is considered as generating a current $-G_me_s$ flowing from plate toward cathode through the impedance formed by the plate resistance of the tube in parallel with the load resistance.

The equivalent circuits of the amplifier give only those currents and those voltage drops that are produced as a result of the application of a signal voltage to the amplifier grid. The actual currents and potentials existing in the plate circuit are the sum of the currents and potentials developed in the equivalent circuit and those existing in the amplifier when no signal is applied.

The equivalent circuits give the exact performance of the vacuum-tube amplifier to the extent that the plate resistance R_p, the amplification factor μ, and the mutual conductance G_m, which are used in setting up the equivalent circuits, are constant over the range of variations produced in control-grid and plate voltages by the signal voltage. Hence, when the signal is small, the equivalent circuit is almost exactly correct because the changes produced by the signal are so small that the tube coefficients are substantially constant. As the signal voltage increases, the error involved in the equivalent circuit becomes larger. However, the effects resulting from variations in the tube coefficients are second-order effects even with rather large signal voltages, so that the equivalent circuit is found useful for most practical conditions.

41. Audio-frequency Voltage Amplifiers. Resistance Coupling.—Voltage amplification is commonly obtained by placing a high resistance in the plate circuit of the amplifier tube. Practical circuits for such resistance-coupled amplifiers are shown in Fig. 57*a*. Here R_c, commonly termed the coupling resistance, is the high resistance placed in the plate

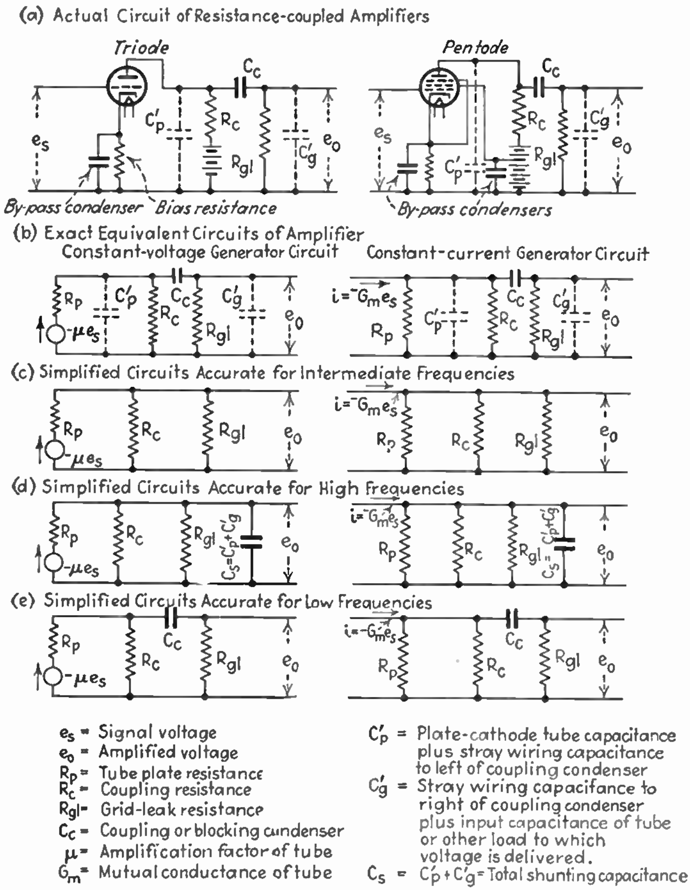


FIG. 57.—Circuits of resistance-coupled amplifiers together with equivalent circuits and simplifications of the equivalent circuits useful in making amplifier calculations. The constant-voltage generator circuit is used with the triode tube, and the constant-current generator form of the equivalent circuit is used for the pentode tube.

circuit. The alternating potential developed across this resistance by the amplified signal currents is separated from the direct-current voltage drop across the resistance by means of a grid-leak resistance R_{gl} and a coupling condenser C_c.

The most important property of the resistance-coupled amplifier is the way in which the amplification varies with frequency. Such a characteristic is shown in Fig. 58 for a representative case. This curve has as its distinguishing feature an amplification that is substantially constant over a wide range of frequencies, but which drops off at both very low and very high frequencies. The falling off at very low frequencies is a result of the fact that the high reactance which the coupling condenser C_c offers to low frequencies consumes some of the low-frequency voltage that would otherwise be developed across the grid leak. The reduction in amplification at high frequencies is caused by the tube and stray capacitances which shunt the coupling and grid-leak resistances. These capacitances have low enough reactance at high frequencies to lower the effective load impedance, with a consequent reduction in the voltage developed at the output.

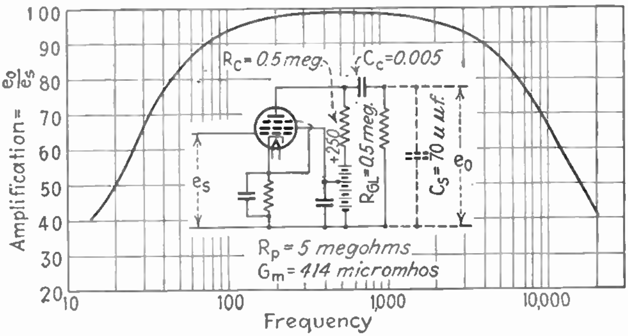


FIG. 58.—Variation of amplification with frequency in a typical resistance-coupled pentode amplifier.

Analysis and Calculation of Amplification Characteristic.—The characteristics of an amplifier are calculated by replacing the tube by its equivalent circuit and then analyzing the resulting electrical network. Either form of equivalent circuit can be employed, with the constant-voltage arrangement more convenient for triode tubes and the constant-current arrangement more satisfactory for pentode and other tubes having very high plate resistances.

Equivalent circuits of a resistance-coupled amplifier showing all essential circuit elements are given in Fig. 57*b* for both methods of representing the tube characteristics. The resulting circuits are quite complicated, but can be simplified by considering only a limited range of frequencies at a time. Thus in the middle range of frequencies the reactance of the coupling condenser C_c in an ordinary amplifier will be so small as to be the practical equivalent of a short circuit as compared with the grid-leak resistance, whereas the reactance of the shunting capacitances will still be so high as to be the practical equivalent of an open circuit. Under such conditions the equivalent circuits take the form shown in Fig. 57*c*. A very simple calculation based upon these circuits shows that the amplification in the middle range of frequencies is given by the equations:

For constant-voltage generator circuit (most suitable for triodes):

$$\left.\begin{array}{l}\text{Amplification in middle}\\ \text{range of frequencies}\end{array}\right\} = \frac{e_0}{e_s} = \mu\frac{R_L}{R_L + R_p} \tag{51a}$$

For constant-current generator (most suitable for pentodes):

$$\left.\begin{array}{l}\text{Amplification in middle}\\ \text{range of frequencies}\end{array}\right\} = \frac{e_0}{e_s} = G_m R_{eq} \tag{51b}$$

where

μ = amplification factor as defined by Eq. (33) or (38)

G_m = mutual conductance as defined by Eq. (35) or (40)

$R_L = \dfrac{R_c R_{gl}}{R_c + R_{gl}}$ = resistance formed by coupling resistance R_c and grid-leak resistance R_{gl} in parallel

$R_{eq} = \dfrac{R_c}{1 + \dfrac{R_c}{R_{gl}} + \dfrac{R_c}{R_p}}$ = equivalent resistance formed by plate resistance, grid-leak resistance, and coupling resistance, all in parallel.

At high frequencies it is necessary to take into account the effect of the capacitances shunting the coupling and grid-leak resistances. This leads to the equivalent circuit of Fig. 57*d*, which is applicable at high frequencies.

For this circuit[1]

$$\frac{\left.\begin{array}{l}\text{Actual amplification}\\ \text{at high frequencies}\end{array}\right\}}{\left.\begin{array}{l}\text{Amplification in}\\ \text{middle range}\end{array}\right\}} = \frac{1}{\sqrt{1 + (R_{eq}/X_s)^2}} \tag{52a}$$

[1] Equation (52*a*) is derived by applying Thévenin's theorem to the network to the left of the shunting capacity C_s in Fig. 57*d*. According to Thévenin's theorem this

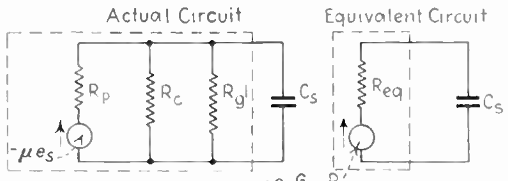


FIG. 59.—Simplification of the circuit of Fig. 57*d* by the use of Thévenin's theorem to replace the portion of the actual circuit shown in the dotted rectangle by the simple network in the dotted rectangle of the equivalent circuit.

network can be replaced by an equivalent generator in series with a resistance, as shown in Fig. 27. The voltage of the generator is the voltage appearing across the

where

$$X_s = \frac{1}{2\pi f C_s} = \text{reactance of total shunting capacitance } C_s$$

R_{eq} = resistance formed by plate, coupling, and grid-leak resistances, all in parallel.

The extent to which the amplification falls off at high frequencies is therefore determined by the ratio which the reactance of the shunting capacitance C_s bears to the equivalent resistance obtained by combining the plate resistance, coupling resistance, and grid-leak resistance in parallel. This loss of amplification at high frequencies can be estimated by the fact that at the frequency which makes the reactance of the shunting condenser C_s equal the equivalent resistance formed by R_p, R_c, and R_{gl} in parallel, the amplification drops to 70.7 per cent of its middle-frequency range value. The amount of falling off at other frequencies as calculated from Eq. (52*a*) is given in the universal amplification curve of Fig. 60.

At low frequencies the shunting capacitance C_s has such a high reactance as to be equivalent to an open circuit, but the reactance of the coupling condenser C_c becomes sufficient to cause a falling off in the amplification. The equivalent circuit under these conditions hence takes the form shown in Fig. 57*e*. A manipulation of the relations existing in this circuit shows that

$$\frac{\left.\begin{matrix}\text{Amplification at}\\ \text{low frequencies}\end{matrix}\right\}}{\left.\begin{matrix}\text{Amplification in}\\ \text{middle range}\end{matrix}\right\}} = \frac{1}{\sqrt{1 + (X_c/R)^2}} \tag{52b}$$

where

$$X_s = \frac{1}{2\pi f C_c} = \text{reactance of coupling condenser } C_c$$

$$R = R_{gl} + \frac{R_c R_p}{R_c + R_p} = R_{gl} + \frac{R_c}{1 + \dfrac{R_c}{R_p}} = \text{resistance formed by grid}$$

leak in series with the combination of plate and coupling resistances in parallel.

---

condenser terminals when the condenser is open-circuited, and so is the output voltage in the middle range of frequencies as given by Eqs. (51*a*) and (51*b*). The internal resistance of the generator is the resistance formed by plate, coupling, and grid-leak resistances, all in parallel, and so is R_{eq}. Referring to Fig. 59 it is apparent that the amplification at high frequencies is

$$\left.\begin{matrix}\text{High-frequency}\\ \text{amplification}\end{matrix}\right\} = G_m R_{eq} \frac{X_s}{\sqrt{X_s^2 + R_{eq}^2}} = G_m R_{eq} \frac{1}{\sqrt{1 + (R_{eq}/X_s)^2}}$$

Dividing this by Eq. (51*b*) gives Eq. (52*a*).

The extent to which the amplification falls off at low frequencies is therefore determined by the ratio of the reactance of the coupling condenser to the equivalent resistance obtained by combining the grid leak in series with the parallel combination of coupling resistance and plate resistance. This loss of amplification at low frequencies can be estimated by the fact that, at the frequency that makes the reactance of the coupling condenser equal the equivalent resistance R, the amplification falls to 70.7 per cent of its value in the middle range of frequencies. The amount of falling off at other frequencies as calculated from Eq. (52*b*) is given in the universal amplification curve of Fig. 60.

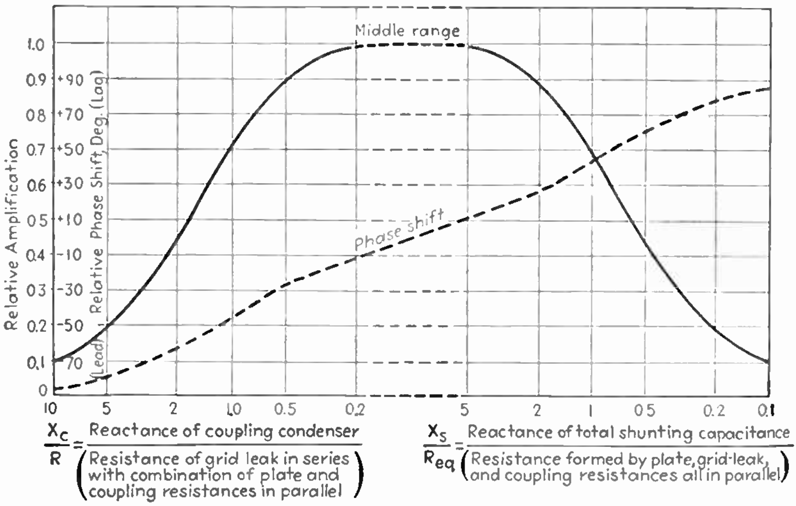


FIG. 60.—Universal amplification curve showing how the amplification of a resistance-coupled amplifier falls off at high and low frequencies.

Universal Amplification Curve.—The curve of Fig. 60, which gives the falling off in amplification at high and at low frequencies, can be thought of as a universal amplification curve because it applies to all resistance-coupled amplifiers.

The procedure for obtaining the amplification characteristic by using the universal amplification curve is: *first,* calculate the amplification in the middle-frequency range, using Eq. (51*a*) or (51*b*); *second,* calculate the frequencies that give X_c/R convenient values, and then use the universal amplification curve to get the low-frequency falling off; and *finally,* estimate the total shunting capacity C_s, calculate the frequencies at which X_s/R_{eq} has convenient values, and then get the high-frequency response with the aid of the universal amplification curve.

Example.—If the above procedure is applied to the circuit constants of Fig. 58, the amplification in the middle range of frequencies is found by substitution in Eqs.

(51a) and (51b) to be

$$\left.\begin{matrix}\text{Amplification in}\\ \text{middle range}\end{matrix}\right\} = G_mR_{eq} = 414 \times 10^{-6} \times 238{,}000 = 98.4$$

where

$$R_{eq} = \frac{500{,}000}{1 + \dfrac{500{,}000}{500{,}000} + \dfrac{500{,}000}{5{,}000{,}000}} = 238{,}000 \text{ ohms}$$

At low frequencies, referring to Eq. (52b), the amplification falls to 70.7 per cent of the middle-range value when the frequency is such as to make the reactance of the coupling condenser equal to $R = R_{gl} + \dfrac{R_c}{1 + \dfrac{R_c}{R_p}} = 955{,}000$ ohms. This is when $f = 1/2\pi CR = 33.4$ cycles. From Fig. 60 one finds that at frequencies of 66.8, 16.7, and 6.7 cycles (which are, respectively, 2, 0.5, and 0.2 times 33.4 cycles) the amplification is 0.90, 0.45, and 0.20 times the middle-frequency range value of 98.4, or 88, 44, and 20, respectively.

At high frequencies the amplification falls to 70.7 per cent of its middle-range value when the reactance of the shunting capacitance C_s equals R_{eq} [see Eqs. (51b) and (52a)]. This is at a frequency f such that

$$f = \frac{1}{2\pi C_s R_{eq}} = \frac{1}{2\pi \times 70 \times 10^{-12} \times 238{,}000} = 9{,}560 \text{ cycles}$$

By use of Fig. 60 it is found that at frequencies of 4,780, 19,120, and 47,800 cycles the amplification is 0.90, 0.45, and 0.20 times the middle-frequency range value of 98.4, or 88, 44, and 20, respectively.

Factors Affecting the Design and Performance of Resistance-coupled Amplifiers.—Both triode and pentode tubes are employed in resistance-coupled amplifiers, with the latter being generally preferred. Pentodes used for this purpose are ordinarily small tubes having a sharp cut-off (not a variable-mu characteristic). Triodes used in resistance-coupled amplifiers usually have a high amplification factor.

The coupling resistance R_c used with pentode resistance-coupled amplifiers intended for audio-frequency amplification is normally 250,000 to 500,000 ohms. Lower values reduce the equivalent load resistance R_{eq} in Eq. (51) and result in lower amplification, although there is a corresponding improvement in the high-frequency response. Higher values require that the plate current, and hence the mutual conductance, be excessively low. This is because an excessive fraction of the plate-supply voltage will otherwise be consumed as voltage drop in the very large coupling resistance. Triodes having amplification factors of the order of 100 call for coupling resistances that are about the same values as those used with pentodes, whereas with triodes having lower amplification factors the coupling resistance is reduced in proportion.

The grid-leak resistance should be as high as possible to minimize its shunting effect upon the coupling resistance. At the same time, the grid-leak resistance is in series with the grid circuit of the tube to which the amplified voltage is delivered and so must not exceed the maximum allowable value that can be placed in series with the grid.[1]

The coupling-condenser capacitance is chosen to give the desired low-frequency response with the grid-leak and coupling resistances employed. The low-frequency response will be better the larger the coupling condenser, and there is no difficulty in obtaining substantially constant response down to frequencies of a few cycles per second. It is undesirable, however, to make the low-frequency response any better than actually required, even if this can be done without difficulty. This is because the better the low-frequency response the greater will be the difficulty from regeneration at low frequencies, as discussed in Sec. 47.

The coupling condenser must have low leakage to direct-current potentials. This is because leakage in the coupling condenser allows a direct current to flow through the grid leak, and the resulting voltage drop in the grid leak biases the grid of the next tube positively. Low leakage is especially important when the circuit proportions are such as to give an unusually good low-frequency response, for then the capacitance and grid-leak resistance are both large.

The high-frequency response is determined by the equivalent load resistance, R_{eq} in Eq. (52*a*), and the shunting capacitances. With pentode tubes and high-mu triodes using coupling resistances of the order of 250,000 to 500,000 ohms, the high-frequency response will usually be satisfactory for high-fidelity audio-frequency systems. The only exception is when the amplified voltage is delivered to triode tubes which have high input capacitance.

The expected high-frequency response of a resistance-coupled amplifier can be calculated on the basis of an estimated shunt capacitance. This capacitance consists of that due to wiring, which can normally be kept to less than 10 $\mu\mu$f if care is taken, plus the output capacitance of the amplifier tube and the input capacitance of the tube to which the ampli-

[1] With any particular tube the resistance that it is safe to place in series with the grid is limited by the possibility of grid current from ionization of residual gas. The polarity of this grid current is such as to produce a voltage drop in the grid-circuit resistance that reduces the negative grid bias. If the grid-circuit resistance is too high the reduction in bias may increase the space current (and hence the ionization) to the point where the process becomes cumulative and destruction of the tube results.

With all small tubes, other than power tubes, the maximum resistance that can be safely placed in series with the grid is usually of the order 1 megohm. With power tubes the maximum allowable resistance varies with different types of tubes, and also depends on whether or not self-bias is used. With self-bias the allowable resistance is of the order of 250,000 to 1,000,000 ohms. With fixed bias the maximum permissible grid-circuit resistance is so low that resistance coupling to power tubes with fixed bias is usually not recommended.

fied voltage is delivered. This output capacitance commonly ranges from 3 to 12 $\mu\mu$f, depending upon the design of the tube, while the input capacitance will be about 4 to 7 $\mu\mu$f for pentodes and screen-grid tubes, and considerably larger for triodes.[1]

When the usual circuit proportions give an excessive falling off in amplification at high frequencies, it is necessary to reduce the coupling resistance in the case of pentode tubes. With triodes it is possible either to lower the coupling resistance or to use a tube having a lower plate resistance. It will be noted that the use of any of these expedients to improve the relative flatness of the amplification curve at high frequencies is accompanied by a reduction in the mid-frequency amplification.

The electrode voltages of a resistance-coupled amplifier must be carefully selected to insure proper operation. The control grid should be biased sufficiently negative to prevent the grid from drawing current at any time. This requires a bias about 1 volt greater than the crest signal voltage in order to overcome the velocity with which the electrons are emitted, and to take care of contact potentials. The plate-supply voltage is preferably as high as possible, but practical circumstances place a limit of 200 to 300 volts under ordinary conditions. In pentodes the screen-grid potential must be so chosen in relation to the grid bias that the voltage drop of the resulting plate current in the coupling resistance will not consume more than one-half to two-thirds of the plate-supply voltage. If the adjustment is such as to give a larger plate current, the drop in the coupling resistance is so great that the voltage left over for the plate of the tube will be very nearly zero. A virtual cathode then forms in front of the suppressor, and the amplifying action is largely lost. It is hence necessary to reduce the plate current of a pentode as the coupling resistance is increased, a point which inexperienced designers commonly overlook.

Practical Designs of Resistance-coupled Amplifiers.—A number of designs of resistance-coupled amplifiers suitable for audio-frequency service are given in Table VI. These indicate reasonable proportions for amplifiers employing pentodes and high-mu triodes.

[1] To a first approximation the input capacitance of a triode tube is given by the equation

$$\text{Input capacitance} = C_{gf} + C_{gp}(1 + A) \tag{53}$$

where C_{gf} is the grid-cathode capacitance, C_{gp} is the grid-plate capacitance, and A is the ratio of alternating voltage across the load impedance in the plate circuit of the output tube to the signal voltage applied to the control grid. A cannot exceed the amplification factor μ of the tube and will normally be at least one-half the amplification factor. Equation (53) is a simple form of Eq. (68) and is exact for the special case where the load is a resistance.

TABLE VI.—TYPICAL DESIGNS OF RESISTANCE-COUPLED AMPLIFIERS[1]

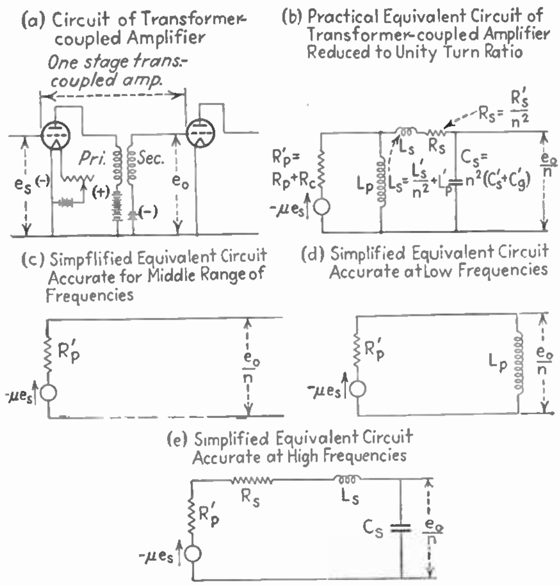
	Pentode 57,77, and 6C6 tubes						Triode section of 2A6 and 75 tubes					
Plate supply, volts	100		180		250		100		180		250	
Screen supply, volts	20	20	30	30	50	52						
Control-grid bias, volts	−1.05	−1.25	−1.30	−1.55	−2.1	−2.3	−1.10	−1.05	−1.30	−1.30	−1.35	−1.35
Cathode resistor, ohms	3,400	7,250	2,600	4,850	3,500	6,200	11,550	15,000	5,450	9,000	3,380	5,600
Plate resistor, megohms	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50	0.25	0.50
Grid resistor, megohms	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Plate current, milliamperes	0.23	0.13	0.38	0.24	0.48	0.30	0.09	0.07	0.24	0.14	0.40	0.24
Plate voltage as per cent of plate supply	42.5	35.0	47.2	33.3	52.0	40.0	77.5	65.0	66.6	61.1	60.0	52.0
Voltage amplification	54	53	92	93	100	110	36	37	56	55	59	58
Output voltage (peak volts) with only little distortion	22	23	44	48	65	65	17	17	36	34	41	40

[1] These designs represent recommendations of RCA Radiotron Company.

42. Audio-frequency Voltage Amplifiers. Transformer Coupling.—In the transformer-coupled amplifier the primary of a step-up transformer is connected in the plate circuit of the tube as shown in Fig. 61*a*. Compared with the resistance-coupled amplifier, this arrangement has the advantage of introducing negligible direct-current resistance in the grid circuit of the tube to which the amplified voltage is delivered. The transformer-coupled arrangement also adapts itself much more readily to the excitation of push-pull amplifiers than does resistance coupling.

The most important characteristics of a transformer-coupled amplifier are the amount of amplification and the way in which this amplification varies with frequency. A typical amplification curve is shown in Fig. 62. It has as its distinguishing features a relatively constant amplification in the middle range of frequencies, a falling off at low frequencies, and a falling off at high frequencies. The loss in amplification at low frequencies results from the low reactance that the primary has at low frequencies, whereas the behavior at high frequencies is a result of the action of the leakage inductance and effective secondary capacitance of the transformer. The useful frequency range depends largely upon the design of the transformer and can be readily made to cover voice frequencies. Difficulties are encountered in amplifying very low and very high frequencies, however, and resistance coupling is accordingly preferred where very wide frequency bands are involved.

Analysis of Transformer-coupled Amplifiers Used with Triode Tubes.—The behavior of a transformer-coupled amplifier can be determined by replacing the tube and transformer by equivalent electrical circuits and analyzing the resulting network. The transformer itself is a complicated combination of resistances, inductances and capacitances. For practical amplifier analysis it can, however, be replaced by the simple unity-turn-ratio network of Fig. 61*b*, in which the essential elements are the primary



Notation

R_p = plate resistance of tube
R_c = direct-current resistance of primary
L'_p = primary leakage inductance
L_p = incremental primary inductance
R'_s = resistance of secondary winding
L'_s = secondary leakage inductance
$L_s = \frac{L'_s}{n^2} + L'_p$ = total leakage inductance referred to primary side

C'_s = secondary distributed capacitance
C'_g = input capacitance of tube to which voltage e_o is delivered
n = ratio of secondary to primary turns
μ = amplification factor of tube
$C_s = n^2(C_s' + C_g')$

FIG. 61.—Actual circuit of a transformer-coupled amplifier, together with equivalent circuits.

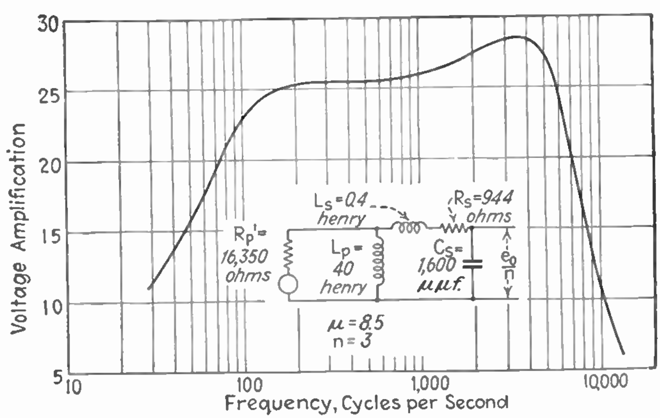


FIG. 62.—Variation of amplification with frequency in a typical transformer-coupled amplifier.

inductance, the total leakage inductance, the effective capacitance in shunt with the secondary, the turn ratio, and the winding resistances.[1]

In the middle range of frequencies the reactances of the primary inductance and of the secondary capacitance are so high as to be substantially open circuits. The equivalent circuit then reduces to the simplified form of Fig. 61*c*, from which

$$\left.\begin{array}{l}\text{Amplification in middle}\\ \text{range of frequencies}\end{array}\right\} = \frac{e_0}{e_s} = \mu n \tag{55}$$

At low frequencies the reactance of the primary inductance falls off, so that this circuit element can be no longer neglected. This leads to the equivalent circuit of Fig. 61*d*, which is applicable to low frequencies, and from which a simple analysis shows

$$\left.\begin{array}{l}\text{Amplification at low}\\ \text{frequencies}\end{array}\right\} = \frac{e_0}{e_s} = \mu n \frac{1}{[1 + (R_p'/\omega L_p)^2]^{1/2}} \tag{56}$$

Here $\omega L_p/R_p'$ is the ratio of the reactance of the primary inductance to the resistance formed by the plate resistance of the tube plus the direct-current resistance of the primary winding (*i.e.*, the effective plate resistance). Examination of this equation shows that the falling off in response at low frequencies is determined only by the ratio of primary reactance to effective plate resistance, *and that the response falls off to* 70.7 *per cent of its mid-range value* μn *at the frequency for which the reactance of the primary inductance equals the effective plate resistance* R_p'. The results of Eq. (56) are plotted in the low-frequency portion of Fig. 63, from which one can obtain the falling off at other low frequencies.

[1] For the greatest accuracy there should also be included a resistance R_e shunted across the primary inductance to represent eddy-current losses. This resistance is so high as to have only a small shunting effect, however, and so need not be considered unless extreme accuracy is required.

The effect of such a resistance shunted across the transformer can be readily allowed for by considering the resistance to be part of the tube and then applying Thévenin's theorem to the network consisting of tube and shunting resistance. Doing this shows that the effect of the eddy-current resistance R_e can be taken into account by using slightly reduced values of R_p' and μ in Fig. 61*b* and Eqs. (55), (56), and (57), according to the relations:

$$\left.\begin{array}{l}\text{Effective } \mu \text{ when taking}\\ \text{into account } R_e\end{array}\right\} = \frac{\text{actual } \mu \text{ of tube}}{1 + \dfrac{R_p'}{R_e}} \tag{54a}$$

$$\left.\begin{array}{l}\text{Effective } R_p' \text{ when taking}\\ \text{into account } R_e\end{array}\right\} = \frac{\text{actual value of } R_p'}{1 + \dfrac{R_p'}{R_e}} \tag{54b}$$

where R_p' is defined as in Fig. 61.

At high frequencies the reactance of the capacitance C_s in shunt with the secondary can no longer be neglected, and the effect of the leakage inductance upon the current drawn by this capacitance must also be considered. This leads to the equivalent circuit of Fig. 61*e*, which is applicable to high frequencies and is essentially a series resonant circuit having a high resistance. The behavior at high frequencies depends upon the resonant frequency of this series circuit, and the Q at resonance. An analysis shows that

$$\left.\begin{array}{l}\text{Amplification at}\\ \text{high frequencies}\end{array}\right\} = \frac{e_0}{e_s} = \mu n \frac{1/\omega C_s}{\sqrt{(R_p' + R_s)^2 + \left(\omega L_s - \dfrac{1}{\omega C_s}\right)^2}} \tag{57a}$$

$$= \mu n \frac{1}{\sqrt{(\gamma/Q_0)^2 + (\gamma^2 - 1)^2}} \tag{57b}$$

where

$$\gamma = \frac{\text{actual frequency}}{\text{series resonant frequency of } C_s \text{ and } L_s}$$

$Q_0 = \dfrac{\omega_0 L_s}{R_p' + R_s} =$ circuit Q at the frequency for which C_s and L_s are in series resonance.

The remaining symbols are as defined in Fig. 61.

The results of Eq. (57) are plotted in the high-frequency portion of Fig. 63. It is seen from this figure that the most uniform response in the high-frequency range is obtained when the Q of the series resonant circuit at the resonant frequency has a value of 0.85 to 0.90. Lower values of Q cause a falling off in the high-frequency response, while higher values introduce a high-frequency peak.

Universal Amplification Curve.—The curve of Fig. 63 can be thought of as a universal amplification curve since it applies to all transformers. The use of this curve in practical amplifier calculations is indicated by the following example.

Example.—Calculate the amplification curve of Fig. 62 from the circuit constants given in the illustration.

The amplification in the middle range of frequencies is found by Eq. (55) to be $\mu n = 8.5 \times 3 = 25.5$. At the low frequencies Eq. (56) shows that the amplification falls to 70.7 per cent of μn when $\omega L_p = R_p'$, or when $f = 16{,}350/2\pi \times 40 = 65.2$ cycles. From the universal amplification curve, Fig. 63, it is found that the amplification at frequencies of 0.2, 0.5, and 2.0 times this frequency is, respectively, 0.20, 0.45, and 0.90 times μn. Therefore the amplification at frequencies of 13.0, 32.6, 65.2, and 130.4 cycles is 5.1, 11.5, 18.0, and 22.0, respectively. For calculation of the high-frequency response, the series resonant frequency of L_s and C_s is found to be

$$\frac{1}{(2\pi\sqrt{L_s C_s})} = \frac{10^6}{(2\pi\sqrt{0.4 \times 1600})} = 6300 \text{ cycles.}$$

From Eq. (57), $Q_0 = \omega_0 L_s/(R_p' + R_s) = 2\pi 6300 \times 0.4/(16{,}350 + 944) = 0.914$, where $R_s = 8480/3^2 = 944$. Reference to Fig. 63 for this value of Q_0 shows that at 0.2, 0.5, 1.0, and 2.0 times the series resonant frequency, the amplification is 1.0, 1.10, 0.914, and 0.27 respectively, times μn. Therefore at frequencies of 1260, 3150, 6300, and 12,600 cycles, the amplification is 25.5, 28.1, 23.2, and 6.9 respectively. From these calculated values of amplification and the values previously obtained for low frequencies, a curve of amplification as a function of frequency can be plotted as shown in Fig. 62.

Discussion of Amplifier Characteristics.—The low-frequency response on a transformer-coupled amplifier is determined largely by the incremental inductance of the transformer primary when the direct-current

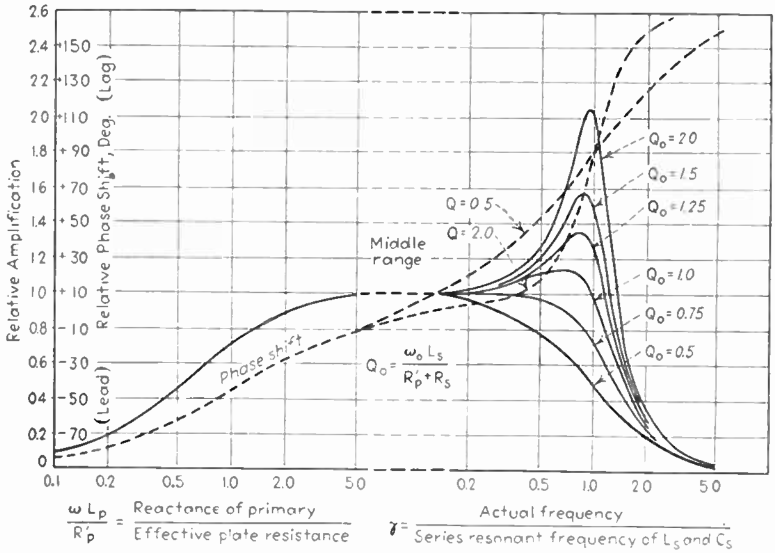


FIG. 63.—Universal amplification curve of transformer-coupled amplifiers.

saturation is that produced by the plate current of the tube. This means that for best low-frequency characteristics the transformer must have a large number of primary turns. This calls either for a transformer that is physically large or one in which most of the winding space is allotted to the primary and which therefore has a low turn ratio.

A good high-frequency response requires that the resonant frequency formed by the capacitance effectively shunting the secondary and the leakage inductance should be as high as possible. This condition can be realized by proper design of the windings and by keeping to a minimum the capacitance of the load connected across the transformer secondary. The capacitance of a tube connected across the transformer secondary is often quite important, since with triode tubes the input capacitance is commonly of the same order of magnitude as the distributed capacitance of the transformer. The best high-frequency response is obtained by employing a low turn ratio, which means sacrifice in amplification, and by keeping the physical size of the transformer small, which means poor low-frequency response. Actual transformers must therefore compromise between good high-frequency response, high gain, and good low-frequency response.

The frequency range of an audio transformer can be very materially extended by improving the core material. Better core material makes it possible to obtain a greater primary inductance without changing the windings in any respect, and hence without changing the leakage inductance and distributed capacitance. Conversely, better core material makes it possible to obtain the same primary inductance with a trans-

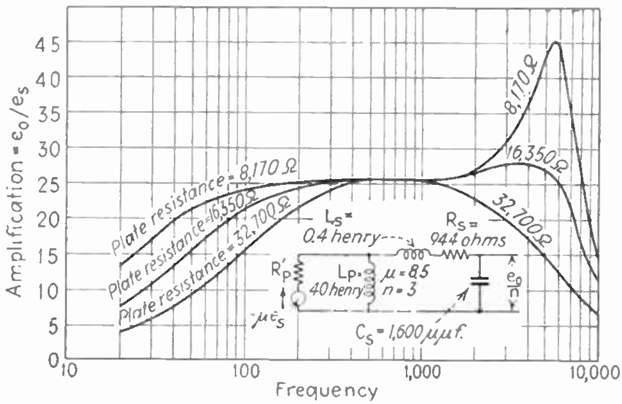


FIG. 64.—Effect of varying the effective plate resistance R_p in a transformer-coupled amplifier. Note that the best characteristic is obtained when the plate resistance is neither too high nor too low but is rather properly matched to the transformer.

former that is physically smaller, and which hence has a higher series resonant frequency.

There is always a proper plate resistance to use with any particular transformer. This is the resistance which makes the effective Q at the series resonant frequency of the secondary slightly less than unity. A lower plate resistance improves the low-frequency response but introduces a high-frequency peak, whereas a higher plate resistance makes both high- and low-frequency responses fall off more rapidly. This is shown in Fig. 64.

It is apparent from the above discussion that the tube and transformer must cooperate with each other to produce the desired result. Most audio-frequency transformers are designed to operate in the plate circuits of triodes having plate currents of 2 to 8 ma and plate resistances of the order 8000 to 15,000 ohms at the operating point. These conditions correspond to a general-purpose tube having an amplification factor of 8 to 20, and represent the situation that has been found to be most satisfactory with transformer coupling. The plate current is then low enough to prevent excessive direct-current saturation in the transformer primary, the plate resistance is low enough to insure the possibility of a satisfactory low-frequency response, and large output voltages can be developed without distortion.

Design of Transformer-coupled Amplifiers.—In practical transformer-coupled amplifiers, stock commercial transformers are ordinarily purchased. The manufacturer normally gives the proper plate resistance to use and the approximate frequency range that can be expected. About the only thing that the amplifier designer can do is hence select a transformer suitable for his purposes and match to the tube by adjusting the grid bias to get the optimum plate resistance.

In cases where specifications can be laid down for the transformer, the design procedure consists in selecting a tube, determining an incremental primary inductance that will give the desired low-frequency response when working with this plate resistance, specifying a series resonant frequency for the secondary that will give the desired high-frequency range, taking into account the fact that part of the capacitance C_s is supplied by the input capacitance to which the amplified voltage is applied, and specifying the leakage inductance of the transformer when reduced to unity turn ratio. The turn ratio is then made as high as possible, but the exact value will generally be fixed by the other requirements.

Determination of Transformer Characteristics.—The incremental inductance of the transformer primary can be determined by measuring at a moderately low frequency, using the appropriate direct-current saturation. The leakage inductance reduced to unity turn ratio is the inductive reactance between primary terminals observed at any convenient frequency, such as 1000 cycles, when the secondary terminals are short-circuited. The resonant frequency of the secondary can be obtained by applying a constant voltage of adjustable frequency to the primary of the transformer while observing the potential across the secondary with a vacuum-tube voltmeter or output tube. The series resonant frequency will then be the frequency for which the secondary voltage is maximum, and this together with a knowledge of the leakage inductance will make it possible to calculate the secondary distributed capacitance C_s'. The turn ratio can be measured by observing the secondary voltage by a vacuum-tube voltmeter when a known voltage of moderate frequency is applied to the primary. The effective shunting resistance R_e that takes into account eddy-current losses can be determined by measuring the effective impedance between primary terminals at 1000 cycles, with the secondary open. If the resulting impedance is expressed as a reactance shunted by a resistance, the resistance will represent R_e with an accuracy sufficient for all ordinary purposes.

Transformer-coupled Amplifiers Using Pentode Tubes.—While transformer-coupled amplifiers ordinarily employ triode tubes, it is possible to use pentode tubes, provided a resistance is shunted across the primary as shown in Fig. 65. This resistance should equal the plate resistance with which the transformer is designed to operate. When this is the case, the frequency-response characteristic will be the same for triode and pentode cases, and the amplifications will be in the ratio of the mutual conductances for the tubes involved.

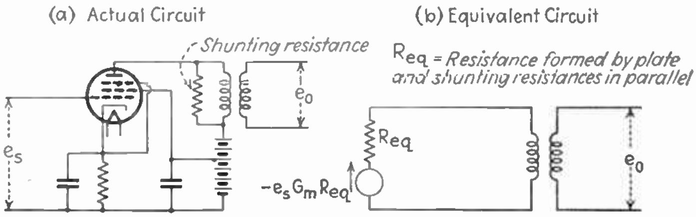


FIG. 65.—Circuit of transformer-coupled amplifier using a pentode tube. The resistance shunted across the primary of the transformer should equal the plate resistance that would be used if the tube were a triode.

43. Audio-frequency Voltage Amplifiers. Miscellaneous Coupling Methods.—While most audio-frequency voltage amplifiers employ either resistance or transformer coupling, other coupling methods are occasionally used. The most important of these are illustrated in Fig. 66, and include the resistance-inductance-coupled amplifier, resistance coupling with a grid choke, transformer coupling with shunt feed, impedance coupling, and coupling with input transformer.

The resistance-inductance-coupled amplifier is the same as the resistance-coupled amplifier except for the addition of a small inductance L_c in series with the coupling resistance, as shown in Fig. 66*a*. This inductance has no effect at low frequencies, but when properly proportioned will make the high-frequency response substantially constant up to higher frequencies than would otherwise be the case. Analysis shows that to obtain a flat response up to some particular frequency using pentode tubes, the coupling resistance should equal the reactance of the shunting capacitance at this frequency, and the inductance in series with the coupling resistance should at this same frequency have a reactance equal to half the coupling resistance. The grid-leak resistance should be much higher than the coupling resistance. With these proportions the amplification is substantially constant up to the desired frequency with negligible phase distortion. Amplifiers using resistance-inductance coupling find their chief use in television work.

Resistance-coupled amplifiers employing a grid choke are used when the amplified voltage is to be applied to a tube, such as a power tube, in which the maximum direct-current resistance that it is permissible to connect in series with the grid circuit is small. The low-frequency response of such an amplifier is normally determined by the inductance of the grid choke, the 70.7 per cent response point coming at the frequency where the inductive reactance of the choke is equal to the resistance formed by the plate and coupling resistances in parallel. It is possible, however, to modify the response at low frequencies by using a coupling condenser small enough to resonate with the grid choke at a moderately

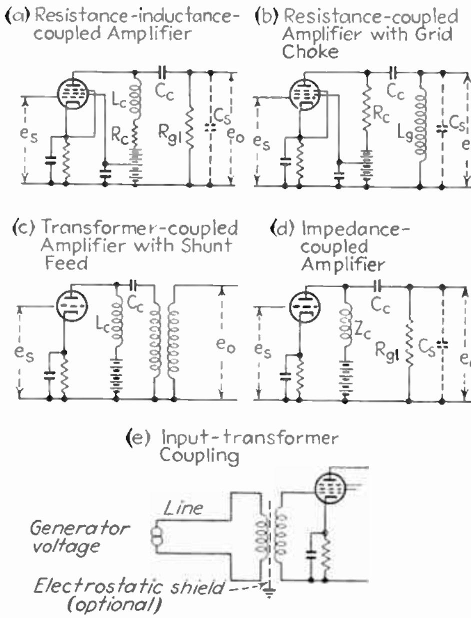


FIG. 66.—Miscellaneous coupling methods.

low frequency. In this way one may peak the response at moderately low frequencies.

The transformer-coupled amplifier with shunt feed, illustrated in Fig. 66*c*, is used where it is desired to avoid direct-current saturation in the primary winding of the transformer. Such an arrangement is frequently used when the transformer core is of permalloy or other similar alloy, and is also sometimes used with output transformers. The shunt-feed arrangement makes it possible to employ a core assembled in such a way as to have negligible air gap. This gives a correspondingly high primary inductance without affecting the leakage inductance or the distributed capacitance of the transformer. The shunt-feed choke, though in parallel with the primary of the transformer, can be made large enough to carry the d-c plate current and still have a high incremental inductance. At the same time the choke contributes nothing to the leakage inductance and secondary capacitance of the transformer, so does not affect the high-frequency response.

The impedance-coupled amplifier illustrated in Fig. 66*d* is occasionally employed. It has a frequency-response characteristic similar to that obtained with resistance coupling, with the falling off at low frequencies being caused by low reactance of the coupling inductance, and at high frequencies by the effect of the shunting capacitance.

An input transformer is used when it is desired to couple the grid of a tube to a transmission line or to a generator having a low internal impedance, as shown in Fig. 66*e*. Such an arrangement is essentially equivalent to an ordinary interstage coupling transformer as discussed above, with the internal impedance of the line (or generator) taking the place of the plate resistance of the tube. The only difference is that this generator resistance is commonly much smaller than the plate resistance of ordinary tubes.

The internal impedance of the generator supplying the power to an input transformer is customarily termed the *line impedance.* Thus an input transformer designed to operate with a microphone having an internal impedance of 4 ohms would be said to be designed to operate from a 4-ohm line, since a transmission line connecting the microphone to the transformer would appear to the transformer primary to have 4 ohms impedance.

44. Tuned Voltage Amplifiers.—In a tuned amplifier the load impedance is supplied by a resonant circuit, using parallel resonance to obtain the necessary high impedance. Such amplifiers find their principal use in the amplification of radio frequencies. Tuned radio-frequency voltage amplifiers practically always employ radio-frequency pentode tubes. These tubes provide complete electrostatic shielding between the plate and control-grid electrodes, and give more amplification than do other tubes that might be employed.

Analysis of Typical Tuned-amplifier Circuits.—Typical tuned-amplifier circuits are illustrated in Fig. 67. The simplest of these is the direct-coupled arrangement shown at Fig. 67*a* and will be considered first because it illustrates the principal properties of tuned amplifiers. The exact equivalent circuit of this amplifier is shown in Fig. 68*b*.[1] From this, the amplification can be written down at once as

[1] It will be noted that the wiring capacitance, the plate-cathode capacitance of the amplifier tube, and the capacitance between grid and ground of the tube to which the amplified voltage is delivered all assist in tuning the coil to parallel resonance in Fig. 68, but otherwise have no effect.

$$\text{Amplification} = \frac{E_0}{E_s} = G_m Z_L \tag{58}$$

where G_m is the mutual conductance of the amplifier tube and Z_L is the impedance formed by the tuned circuit, the grid-leak resistance, and the plate resistance, all in parallel.

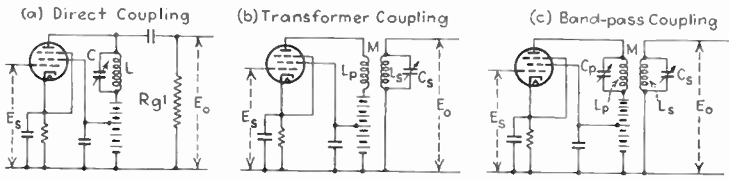


FIG. 67.—Typical tuned-amplifier circuits using pentode tubes.

The results in a typical case are shown in Fig. 67c in the frequency range around resonance. Examination of Fig. 67c and Eq. (58) shows that the amplification curve has the shape of a resonance curve, with the maximum amplification occurring at the frequency at which the tuned circuit is resonant when taking into account the stray shunting capacitances. At this frequency the parallel impedance of the resonant circuit is ωLQ, so that Eq. (58) becomes

$$\left.\begin{array}{l}\text{Amplification at resonance}\\ \text{for pentodes}\end{array}\right\} = G_m \frac{\omega LQ}{1 + \dfrac{\omega LQ}{R_p} + \dfrac{\omega LQ}{R_{gl}}} \tag{59a}$$

The notation is as illustrated in Fig. 68b, with Q being the actual Q of the tuned circuit. With ordinary circuit proportions the grid-leak resist-

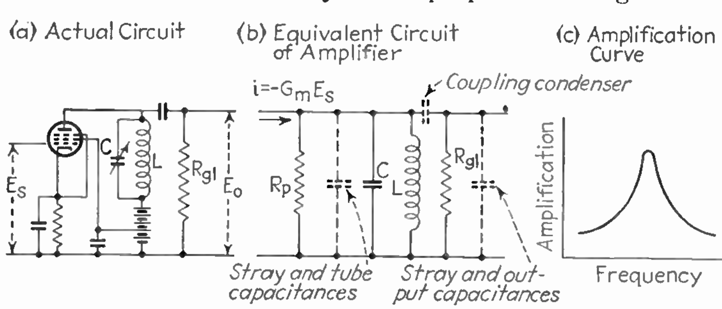


FIG. 68.—Equivalent circuit and amplification curve of a tuned amplifier with direct coupling. The amplification curve has the shape of a resonance curve but with an equivalent Q less than the actual Q of the tuned circuit.

ance is much higher than the parallel impedance of the resonant circuit. In the case of pentodes a further simplification is possible as a result of the fact that the plate resistance may be considered as substantially infinite. Hence with pentode tubes Eq. (59a) can to a good approximation be rewritten as

$$\left.\begin{array}{l}\text{Approximate amplification}\\ \text{at resonance}\end{array}\right\} = G_m\omega LQ \tag{59b}$$

The curve of amplification as a function of frequency will be found to have a shape corresponding to an effective Q lower than the actual Q of the resonant circuit. The ratio of this effective Q of the amplification curve to the actual Q of the tuned circuit depends upon the grid-leak and plate resistances associated with the tuned circuit. It is given by the equation[1]

$$\frac{\text{Effective } Q \text{ of amplification curve}}{\text{Actual } Q \text{ of tuned circuit}} = \frac{1}{1 + \dfrac{\omega LQ}{R_p} + \dfrac{\omega LQ}{R_{gl}}} \tag{60}$$

With pentode amplifiers the effective Q of the amplification curve is only slightly less than the Q of the tuned circuit. With triode tubes, however,

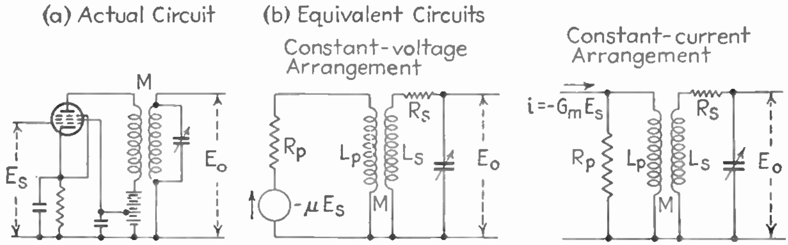


FIG. 69.—Actual and equivalent circuits of tuned amplifier employing transformer coupling

the low plate resistance makes the effective Q normally much lower than the actual Q of the tuned circuit.

The transformer-coupled amplifier of Fig. 67*b* behaves in much the same way as does the direct-coupled circuit of Fig. 67*a*, although differing in details of analysis. The equivalent circuit can for all practical purposes be reduced to that shown in Fig. 69*b*. This is a simple coupled circuit, and has already been mentioned in Sec. 15. Following the usual method of analyzing coupled circuits, one obtains

$$\left.\begin{array}{l}\text{Amplification at}\\ \text{resonance}\end{array}\right\} = G_m\frac{\omega MQ}{1 + \dfrac{(\omega M)^2/R_s}{R_p}} \tag{61a}$$

In the special case of pentode tubes the plate resistance R_p is so high that to a good approximation Eq. (61*a*) can be rewritten as

$$\left.\begin{array}{l}\text{Approximate amplification}\\ \text{at resonance for pentodes}\end{array}\right\} = G_m\omega MQ \tag{61b}$$

[1] This ratio is simply the ratio of the impedance formed by the combination of tuned circuit, grid-leak resistance, and plate resistance, all in parallel, to the impedance of the tuned circuit.

With triode tubes the maximum amplification is obtained when the coupled impedance $(\omega M)^2/R_s$ is equal to the plate resistance of the tube. With such tubes the use of transformer coupling provides a means of matching the tuned circuit to the tube in such a way as to give maximum possible amplification. In contrast with this, the amplification with pentodes becomes greater the higher the coupling, because the plate resistance is so high that a match with the load impedance can never be realized.

With transformer coupling the ratio of effective Q of the amplification curve to the actual Q of the tuned circuit is given by

$$\frac{\text{Effective } Q \text{ amplification curve}}{\text{Actual } Q \text{ of tuned circuit}} = \frac{1}{1 + \dfrac{(\omega M)^2/R_s}{R_p}} \tag{62}$$

With pentodes the plate resistance is so high that the effective Q of the amplification curve is practically the same as the Q of the tuned circuit. On the other hand, triode tubes have such a low plate resistance that the amplification curve normally has a Q that is appreciably lower than the actual Q of the tuned circuit.

Practical Calculation of Amplification Characteristic.—In calculating the amplification characteristic of a tuned amplifier, the first step is to obtain the amplification at the resonant frequency, using Eq. (59) or (61). Next, the effective Q of the amplification curve is calculated, using Eq. (60) or Eq. (62). Finally, the falling off in amplification at frequencies differing from resonance is determined by the aid of the universal resonance curve of Fig. 18.

Tuned Amplifiers with Complex Coupling.—When the resonant frequency of the tuned circuit is adjusted by varying the capacitance, the direct- and transformer-coupled arrangements of Fig. 67 give an amplification that is roughly proportional to the resonant frequency. This is evident when the equations of amplification at resonance are examined and it is remembered that Q is roughly independent of frequency. This behavior can be overcome by coupling the amplifier tube to the tuned circuit by a network that makes the equivalent coefficient of coupling become less as the frequency is increased.

An example of complex coupling to accomplish this result is given in Fig. 70. Here C_1 is a small coupling condenser, commonly only 2 or 3 $\mu\mu$f, and L_2 is a relatively large primary inductance not coupled to the secondary. This inductance has a shunting capacitance C_2 that resonates it to a frequency slightly below the lowest frequency to be amplified. In such a circuit the equivalent mutual inductance becomes less as the frequency is increased above the frequency at which L_2C_2 are in parallel resonance. With proper circuit proportions the effective coupling between the tube and the tuned circuit becomes just enough less as the frequency is raised to make the amplification substantially constant over an appreciable range of frequencies.

Band-pass Amplifiers.—A band-pass characteristic such as shown in Fig. 24 can be realized in an amplifier by the use of two resonant circuits tuned to the same frequency and coupled, as shown in Fig. 71*a*. This

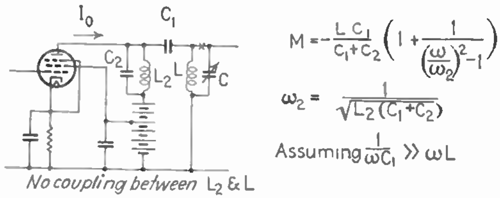


FIG. 70.—Tuned amplifier with complex coupling such that the coefficient of coupling becomes less as the frequency is increased.

arrangement is very desirable for the amplification of modulated waves because it can provide substantially constant amplification for all the essential side-band frequencies contained in the wave, while discriminating sharply against other frequencies. In contrast with this, a tuned amplifier employing a single resonant circuit has a rounded-off response curve, and so cannot discriminate against interfering frequencies just outside the desired frequency band without at the same time discriminating against the higher-frequency side-band components.

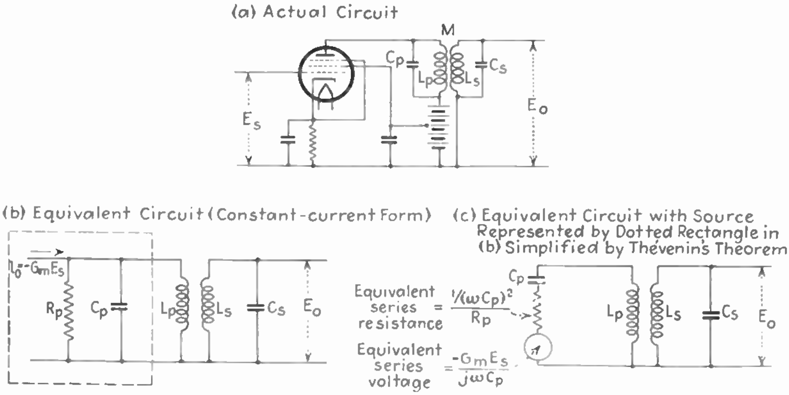


FIG. 71.—Actual circuit of band-pass amplifier together with equivalent circuit.

The equivalent circuit of a band-pass amplifier is shown in Fig. 71*b*. This can be reduced to a band-pass circuit with series voltage by the method described in connection with Fig. 25, with the result shown in Fig. 71*c*. This has already been discussed in Sec. 25 where it was shown that in the practical case with R_p much greater than the reactance of the primary condenser C_p, the equivalent series voltage is

$$\left.\begin{array}{l}\text{Equivalent voltage that can}\\ \text{be considered acting in series}\\ \text{with the primary}\end{array}\right\} = -E_s G_m \frac{1}{\omega C_p} \tag{63}$$

The amplification can be calculated from the equivalent circuit of Fig. 71c by following the method of coupled-circuit analysis discussed in Sec. 14. At the common resonant frequency of the two circuits, the amplification is

$$\left.\begin{array}{l}\text{Amplification at resonant}\\ \text{frequency}\end{array}\right\} = G_m k \frac{\omega_0 \sqrt{L_s L_p}}{k^2 + \dfrac{1}{Q_p Q_s}} \tag{64a}$$

where

G_m = mutual conductance of tube

k = coefficient of coupling between primary and secondary inductances

ω_0 = 2π times resonant frequency

Q_p = $\omega L_p/R_p$ for primary circuit, taking into account any equivalent series resistance that may be added by the plate resistance of the tube

Q_s = $\omega L_s/R_s$ for secondary circuit

L_p, L_s = primary and secondary inductances, respectively.

This amplification has its maximum possible value when the coefficient of coupling has the critical value $k = 1/\sqrt{Q_p Q_s}$, as defined by Eq. (26b). Substituting this value for critical coupling into Eq. (64a) gives

$$\left.\begin{array}{l}\text{Maximum possible}\\ \text{amplification}\end{array}\right\} = G_m \frac{\omega_0 \sqrt{L_p L_s} \sqrt{Q_p Q_s}}{2} \tag{64b}$$

Comparison of Eqs. (64a) and (64b) shows that, when the primary and secondary circuits are identical and the plate and grid-leak resistances are very high, the gain obtained with the band-pass arrangement is exactly half that which is obtained when a single tuned circuit is employed. With couplings greater than the critical value Eq. (64b) gives with satisfactory accuracy the amplification at the peaks of the double humped curve, while Eq. (64a) still gives the amplification at the common resonant frequency.

Selectivity and Selectivity Curves of Tuned Amplifiers.—The ability of a tuned amplifier to discriminate against interfering signals is often fully as important as the gain of the amplifier at resonance. The effectiveness of an amplifier in discriminating against frequencies off resonance can be most satisfactorily shown by plotting a curve of relative amplification in the manner shown in Fig. 72. This figure gives the relative input required to maintain the output constant as the frequency of the input is varied. The input required to produce the output at the frequency of maximum amplification is used as the reference value. Thus in the band-pass case in Fig. 72, it takes an input signal 25 times as strong to produce a given output when the carrier is 30 kc off resonance as it does when the carrier is exactly at resonance. The superiority of the band-pass amplifier in its ability to discriminate against frequencies appreciably off resonance, and at the same time respond uniformly to a band of frequencies about resonance, is clearly evident from Fig. 72.

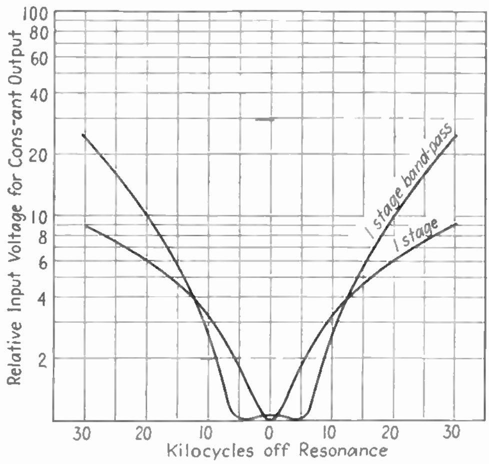


FIG. 72.—Selectivity curves of tuned amplifiers in which the ordinates show relative input voltages required at different frequencies to maintain the output voltage constant.

45. Noise, Hum, and Microphonic Effects.—All amplifiers give some output even when there is no input voltage. Such output is commonly referred to as hum, noise, or microphonic action depending upon its origin.

Hum.—The term hum is applied to voltages introduced into the amplifier by the action of power circuits. The chief sources of hum are stray electrostatic and magnetic fields, poor filtering in the rectifier-filter system for supplying anode power, and the use of alternating current for heating cathodes.

Hum from a cathode-heating current can be kept small by using properly designed heater-type tubes, as discussed in Sec. 84. Hum from the plate supply is readily controlled by the addition of a more adequate filter.

Electrostatic fields cause trouble with parts of the amplifier having a high impedance to ground, since any electrostatically induced current flowing to ground will produce a hum voltage that is proportional to the impedance between ground and the electrode upon which the field terminates. Trouble of this sort can be readily controlled by electrostatic shielding of tubes and critical parts of the circuit (particularly grid leads) and by proper circuit layout. It is also helpful to ground the metal chassis of the amplifier.

Magnetic fields are the most troublesome sources of hum. Such fields induce voltages in coupling transformers and sometimes also affect the flow of electrons in the tubes. Hum from magnetic fields is minimized by using resistance coupling in the low signal level part of the amplifier and by providing adequate spacing between the input parts of the amplifier and the power and filament transformers. Twisted filament leads are also advisable. When coupling transformers must be used at low-level points, as to match a low-impedance line to the first tube, the transformer is preferably of a type in which the windings are arranged to minimize hum pick-up and should also be provided with a magnetic shield. Residual hum pick-up can then be minimized by orientation of the transformer. Hum resulting from direct action of the magnetic field on the electron flow of the first tubes of an amplifier can be kept under control by using ample spacing, and by the use of tube shields of high-permeability magnetic material.

Hum problems are most severe in high-gain audio-frequency amplifiers, since here the hum voltages when once introduced into one of the earlier stages are efficiently amplified by the subsequent stages. Hum troubles are also present to some extent in tuned amplifiers, however, because the hum voltages superimposed upon the electrode voltages of the tube will cause the amplification of a radio-frequency stage to vary slightly at the hum frequency, thereby modulating the hum upon the radio-frequency voltages being amplified.

Microphonic Noise.—The term microphonic noise is applied to effects that result from mechanical vibration of the tube or other parts of the circuit. In audio-frequency amplifiers, these vibrations produce audio-frequency variations in the plate current that are then amplified. In radio-frequency amplifiers, vibrations modulate themselves up the high-frequency voltages being amplified.

Microphonic action can be minimized by mounting the tube or other part so that it is protected from mechanical vibrations transmitted either through the air or through the base. Individual tubes and tube types vary greatly in their tendency toward microphonic action, so that where this factor is a problem it is commonly desirable to select tubes individually for low microphonic noise.

Thermal Agitation Noise.—The random motion of free electrons in conductors produces minute voltages across the terminals of the conductor that are continually changing in character and amount. Since the motion of the electrons is due to temperature, these potentials are said to be the result of *thermal agitation*. The voltages produced by thermal agitation are of random character having energy uniformly distributed throughout the entire frequency spectrum.

The magnitude of the voltages produced by thermal agitation can be calculated by the formula

$$\left.\begin{array}{l}\text{Square of effective value of voltage}\\ \text{components lying between frequen-}\\ \text{cies } f_1 \text{ and } f_2\end{array}\right\} = E^2 = 4kT\int_{f_1}^{f_2} R\,df \qquad (65)$$

where

k = Boltzmann's constant = 1.374×10^{-23} joule per degree Kelvin
T = absolute temperature in degrees Kelvin
R = resistance component of impedance producing voltages of thermal agitation (a function of frequency), when the impedance is expressed as a resistance in series with a reactance
f = frequency.

In the special case where the resistance component of the impedance is constant over the range of frequencies from f_1 to f_2, Eq. (65) reduces to the much simpler form

$$E^2 = 4kTR(f_2 - f_1) \qquad (66)$$

The noise voltages arising from thermal agitation set an ultimate limit to the smallest potential that can be amplified without being lost in a background of noise. An idea as to the order of magnitude of this limit can be obtained by noting that with a resistance of ½ megohm at 300°K., the noise voltage according to Eq. (66) for a frequency band 5000 cycles wide is 6.4 μv. The noise level of most amplifiers is determined by the thermal agitation of the resistance in the grid circuit of the first amplifier tube.

Shot Effect and Related Phenomena.—The fact that the stream of electrons flowing from cathode to plate is made up of a series of particles rather than a continuous fluid causes minute irregularities in the plate current of the vacuum tube. This results in the possibility of noise of a kind termed *shot effect.*

The irregularities which shot effect give rise to depend upon the space charge that is present, and become small when there is a complete space charge. It is therefore important that the electron emission from the cathode be sufficient to produce a complete space charge. The extent to which the space charge is adequate can be readily observed by varying the cathode heating current and noting whether the space current changes appreciably.

In some types of cathodes there is a form of shot effect, termed flicker effect, that arises as a result of random changes of emission over small cathode areas. With oxide-coated cathodes the flicker effect is normally greater than the true shot effect.

Positive ions produced in the tube as a result of ionization of residual gas, or as a result of the occasional emission of positive ions by the cathode, also give rise to a modified form of shot effect. With tubes having a good vacuum the noise introduced in this way with ordinary space charges is of the same order of magnitude as the thermal-agitation noise occurring in the plate resistance of the tube.

Miscellaneous Sources of Noise.—In addition to the factors discussed above, there are a number of additional ways by which noise can be created. Thus poor contacts, leaky condensers, faulty resistances, etc., very commonly result in the introduction of noise.

Carbon resistors carrying direct current are also particularly troublesome sources of noise. Such noise arises from fluctuations in the contact resistance between adjacent granules and is similar in character to the "hiss" occurring in carbon microphones. The magnitude of the effect is so great that carbon resistors cannot be used as plate-coupling resistances in amplifiers that must amplify even moderately small voltages.

46. Input Admittance of Amplifiers and the Neutralization of Grid-plate Capacitance.—The input admittance of an amplifier is defined as the admittance that is seen when looking toward the control-grid electrode of the tube. This is the admittance across which the voltage to be amplified is applied.

With screen-grid and pentode tubes arranged to have complete electrostatic shielding between control-grid and plate electrodes, the input admittance is supplied by the sum of the grid-screen and grid-cathode capacitances. This is commonly 3 to 10 $\mu\mu$f and is independent of the conditions existing in the plate circuit of the amplifier.

In tubes having direct capacitance between control grid and plate, such as exists in triodes and also in screen-grid and pentode tubes having incomplete shielding, the input admittance is also influenced by the conditions in the plate circuit of the tube. This is because the amount of current that flows from the control-grid to the plate electrode through the direct capacitance between these electrodes is determined by the potential difference between grid and plate, and this potential difference obviously depends upon the amplified voltage developed across the load impedance in the plate circuit. Since the amplified voltage developed in the plate circuit is normally much greater than the signal voltage, it is apparent that the potential difference between grid and plate will be very high as compared with the signal voltage actually applied to the grid. This situation causes the current flowing from grid to plate to be unusually large. Direct capacitance between grid and plate is hence of very great importance in determining the input admittance.

When the load impedance in the plate circuit is a resistance, the amplified voltage developed in the plate circuit is exactly 180° out of phase with the signal voltage applied to the grid. Under these conditions the voltage difference between grid and plate is $E_s + E_p = E_s(1 + A)$, where A is the voltage amplification between the grid and plate electrodes of the tube (but does not include any transformer step-up). The current that flows from grid to plate is equal to this voltage divided by the capacitive reactance of the grid-plate capacitance and so is $E_s(1 + A)\omega C_{gp}$. Since this current is supplied by the signal voltage E_s, it represents an effective input capacitance of $(1 + A)C_{gp}$.

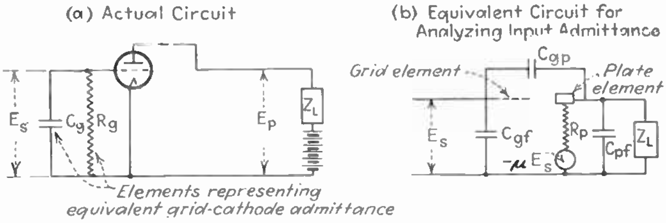


FIG. 73.—Equivalent circuit for analyzing amplifier input admittance. The input admittance is the ratio of current entering the grid electrode to the signal voltage applied to the grid and depends upon the load impedance in the plate circuit.

When the load impedance in the plate circuit is not a resistance, the analysis must be modified to take into account this fact. If the tube is considered as offering to the signal an input capacitance C_g shunted by a resistance R_g, as shown in Fig. 73, one then has

$$\text{Input resistance} = R_g = -\frac{1/\omega C_{gp}}{A \sin \theta} \tag{67}$$

$$\text{Input capacitance} = C_g = C_{gf} + C_{gp}(1 + A \cos \theta) \tag{68}$$

where

C_{gp} = grid-plate tube capacitance
C_{gf} = grid-cathode tube capacitance
A = ratio of voltage developed across load impedance in plate circuit to applied signal (*i.e.*, A is the amplification of tube alone, not taking into account any step-up of voltage in the load)
θ = angle by which voltage across load impedance leads equivalent voltage acting in plate circuit (θ positive for inductive load impedance).

The input resistance in Eq. (67) takes into account the fact that when θ is not zero, energy is transferred directly between grid and plate circuits through the grid-plate capacitance. When the plate load impedance has a capacitive component, the input resistance is positive, meaning that the signal voltage transfers energy directly to the plate circuit. On the other hand, when the plate load impedance has an inductive component, the input resistance is negative. The input circuit then receives some of the amplified energy present in the plate circuit. The magnitude of the input resistance decreases as the frequency becomes greater, because the higher the frequency the greater will be the current flowing through the grid-plate capacitance with a given potential difference. As a consequence, the effects of the input resistance become more important as the frequency is raised.

In audio-frequency amplifiers the input capacitance tends to spoil the high-frequency response. This is particularly the case with high-mu triode tubes, since the large value of A obtained with such tubes makes the input capacitance very high. The resistance component of the input admittance is ordinarily too high to be of much importance at audio frequencies.

In radio-frequency amplifiers the resistance component of the input admittance is particularly important. This is because the input resistance of the tube is in shunt with the tuned circuit supplying the signal voltage that is applied to the tube. In case the input resistance is positive, the effective Q of the tuned circuit will be greatly reduced. On the other hand, if the input resistance is negative, energy will be supplied to the tuned circuit, and oscillations can be expected. It is for this reason that radio-frequency voltage amplifiers ordinarily employ pentode or screen-grid tubes so designed as to give practically perfect electrostatic shielding between grid and plate electrodes. When triode tubes are employed for radio-frequency amplification, it is necessary to neutralize the effect of the energy transfer through the grid-plate capacitance.

Neutralization of the Input Admittance of a Vacuum-tube Amplifier.—The effect of the current that flows through the grid-plate capacitance of a triode tube can be neutralized by arrangements that provide for the flow of an equal and opposite current through an auxiliary (or neutralizing) condenser. The most common methods of doing this are the Neutrodyne and Rice circuits illustrated in Fig. 74. In the form of Neutrodyne circuit shown in Fig. 74, the inductance L_N is closely coupled to the coil L_P but is wound in the opposite direction. The voltage developed by L_N is accordingly proportional to the voltage developed in the plate circuit of the tube but is of opposite polarity. The neutralizing condenser C_N is then adjusted so that the current that flows through it will just neutralize the effect of the current flowing in the opposite direction through the grid-plate capacitance of the tube.

In the Rice system of neutralization the input circuit is split into two parts. By properly adjusting the capacitance of the neutralizing condenser, a voltage developed in the output between plate and ground sends currents into the two parts of the input circuit that cancel each other's effects.

47. Multistage Amplifiers with Particular Reference to Regeneration.—In a multistage amplifier it is possible for energy to be transferred between stages by the grid-plate tube capacitance as discussed above, by stray couplings, or by an impedance common to more than one stage. Such energy transfer is termed *regeneration,* and either modifies the amplification or when of sufficient magnitude will produce oscillations.

In resistance-coupled amplifiers such oscillations are of the multivibrator type (see Sec. 67) and have such a low frequency that when heard in a loud speaker they produce a characteristic "put-put" sound that is termed *motorboating.*

Regeneration in Audio-frequency Amplifiers.—The most important cause of regeneration in audio-frequency amplifiers is energy transfer

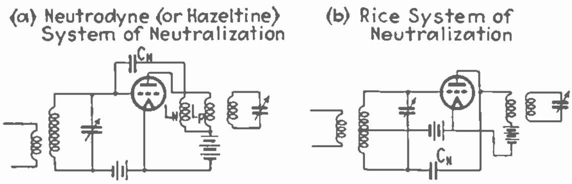


FIG. 74.—Typical circuit arrangements for neutralizing the effect of energy transfer between grid and plate circuits through the grid-plate tube capacitance by an equal and opposite energy transfer through a neutralizing condenser C_n.

between stages as a result of impedance in the source of plate voltage. Thus in Fig. 75*a*, amplified currents in the plate circuit of the final tube will produce a voltage drop in the internal impedance of the plate-supply system. This voltage then acts in the screen and plate circuits of the first tube, resulting in a transfer of energy that modifies the behavior.

The regeneration produced by a common plate impedance is of importance to the extent that the voltage transferred back to the plate and screen circuits of the first tube is of appreciable magnitude compared with the output voltage developed by the first stage in the absence of regeneration. It is not ordinarily necessary to consider regeneration between other tubes than the first and last because of the smaller difference in power levels involved. Regeneration arising from a common plate impedance is most troublesome in amplifiers having very high gain, because then the voltage drop produced in the common plate impedance is largest in proportion to the output voltage of the first tube.

Regeneration from a common plate impedance can be minimized in a number of ways. A large condenser in shunt with the power-supply system will reduce the common impedance, particularly at radio frequencies and at the higher audio frequencies. In addition, the plate and screen circuits of low-level stages can be isolated by the use of series impedances (resistances or inductances) and shunt condensers, as shown in Fig. 75b. Such arrangements are called filters, and serve to make the voltage actually transferred to the plate and screen circuits of a low-level stage less than the voltage developed across the common impedance. In order to be effective, the series impedance of such a filter must be

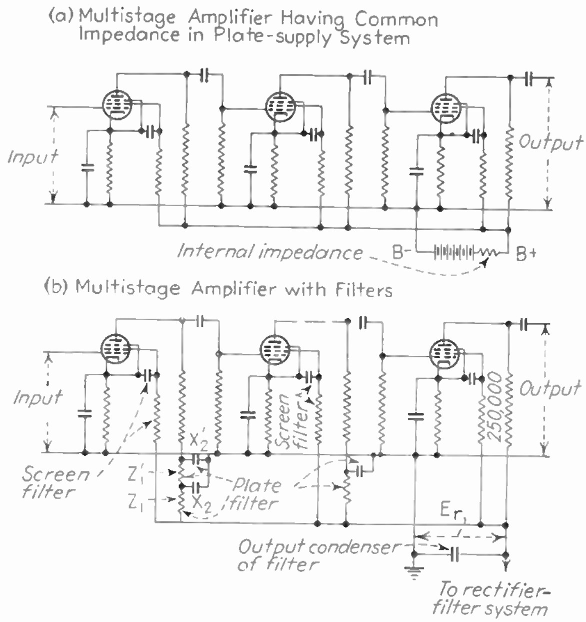


FIG. 75.—Multistage amplifier showing how the internal impedance of a common power-supply system can transfer energy from the final stage to earlier stages, together with same amplifier provided with filters for reducing this energy transfer.

much higher than the reactance of the shunt condenser at all frequencies for which the filter is to be effective. The high series impedance then prevents the voltage developed across the common impedance from sending appreciable current toward the input stage, while the low shunt impedance short-circuits substantially all of whatever current does flow. The result is that the voltage actually introduced into the plate and screen circuits of the input tube is very much less than the voltage developed across the common impedance. The reduction is given approximately by the equations[1]

[1] These equations assume that X_2 is small compared with Z_1 and the plate and coupling impedances of the input tube, and that Z_1 is much larger than the common impedance. Under these conditions a voltage E_r across the common impedance causes a current E_r/Z_1 to flow through Z_1 (Fig. 75b). The greater part of this current

$$\text{Reduction with one-stage filter} = \frac{X_2}{Z_1} \tag{69a}$$

$$\text{Reduction with two-stage filter} = \frac{X_2X_2'}{Z_1Z_1'} \tag{69b}$$

where Z_1 and Z_1' are the series impedances of the filter sections, and X_2 and X_2' are the reactances of the shunting condensers.

The effectiveness of filters and shunting condensers becomes less the lower the frequency, and trouble is practically always experienced in avoiding regeneration in high-gain amplifiers that have a good low-frequency response. This makes it desirable to design audio-frequency amplifiers so that the low-frequency response is no better than actually required. Even then it is often necessary to use a separate power-supply system for different parts of the amplifier.

The principles involved in designing a filter system are illustrated by the following example.

Example.—In the amplifier of Fig. 75*b* each stage gives a voltage gain of 60 in the middle-frequency range, and the filtering need not be effective below 40 cycles. The internal impedance of the plate-supply system is the reactance formed by an 8-μf shunting condenser. Assign values to the elements in the plate filter Fig. 75*b*, so that the voltage acting in the plate circuit of the first tube will not exceed 10 per cent of the voltage which this tube applies to the grid of the second tube.

On the assumption that the signal at the grid of the second tube is 1 mv, the voltage developed across the coupling resistance in the output of the final tube is

$$60 \times 60 \times 0.001 = 3.6 \text{ volts.}$$

The current through the coupling resistance is then $3.6/250{,}000 = 14.4\ \mu$a, and this flowing through 8 μf will at 40 cycles produce a voltage

$$14.4 \times 10^{-6}/(2\pi \times 40 \times 8 \times 10^{-6}) = 0.0072 \text{ volt.}$$

The filter for the first stage is then required to reduce this to 0.0001 volt, or by a factor of $1/72$. A variety of condenser-resistance combinations will do this satisfactorily. Thus substitution in Eq. (69*b*) shows that if $\frac{1}{2}$ μf shunting condensers are used, the series resistances for a two-stage filter must each be at least

$$\sqrt{72}/(2\pi \times 40 \times 0.5 \times 10^{-6}) = 68{,}000 \text{ ohms.}$$

It will be noted that little or no filter is required for the middle stage. This is because the voltage that this tube delivers to the grid at the final tube is 0.060 volt, which is quite large compared with the 0.0072 volt developed across the common impedance.

---

flows through X_2 and so produces a voltage E_rX_2/Z_1 across X_2. In the case of a one-stage filter, this is the voltage actually inserted in the plate circuit of the input tube, and is less than the voltage E_r by the factor X_2/Z_1. In the case of a two-stage filter, there is a further reduction of X_2'/Z_1', giving a total reduction of X_2X_2'/Z_1Z_1', as in Eq. (69*b*).

The screen-grid filter can ordinarily be taken care of by the voltage-dropping resistance in the screen circuit, provided the by-pass condenser from screen to cathode is of reasonable size. Thus in Fig. 75*b*, if a 1-megohm series resistance is required to drop the supply voltage to the value needed for the screen, and the screen-cathode condenser is 1 μf, then the reduction at 40 cycles as calculated from Eq. (69*a*) is 251, which is obviously adequate. However, with very high-gain amplifiers it is sometimes necessary to provide a two-stage filter in the screen circuit by dividing the voltage-dropping resistance into two parts, and using two by-pass condensers.

Audio-frequency amplifiers are also troubled by regeneration arising from electrostatic and magnetic couplings. Electrostatic coupling between unshielded tubes and leads, particularly grid leads of different stages, often causes trouble. This coupling usually results in a high audio-frequency oscillation, and the remedy consists in providing adequate shielding. Magnetic coupling between stages can arise when there is more than one coupling transformer employed. The usual remedy consists in a change in relative orientation of the transformers involved, or in increased spacing.

Regeneration in Radio-frequency Amplifiers.—At radio-frequencies a common cause of regeneration is electrostatic and magnetic coupling between tuned circuits of different stages. This cause of regeneration is ordinarily controlled by enclosing coils in copper or aluminum shielding cans, and by shielding the sections of the tuning condenser from each other. It is also necessary to employ screen-grid or pentode tubes, or to use neutralized triodes, in order to prevent energy transfer directly through the tube. Care must likewise be used in placing the wiring, particularly the grid and plate leads, so that these introduce no capacitive coupling between stages.

Impedances common to two or more stages are particularly troublesome causes of regeneration in radio-frequency amplifiers. This is because at the high frequencies involved a wire only a few inches long will often have sufficient reactance to provide an effective means of transferring energy. Unavoidable common impedances, such as those from common sources of electrode voltages, can be reduced to a low value by shunting with by-pass condensers, and in addition filters may be placed in the leads running to the individual amplifier stages. Indiscriminate use of the chassis as a return circuit often gives trouble, since the chassis, when used in this way, provides an impedance common to all the circuits.

48. Feedback Amplifiers.—In the feedback amplifier a certain amount of regeneration is deliberately introduced in such a way as to reduce the amplifier gain. By properly carrying out this operation it is possible to reduce the non-linear distortion and noise generated in the amplifier, to make the amplification substantially independent of electrode voltages and tube constants, and to reduce greatly the phase and frequency distortion.

The operation of a feedback amplifier can be understood by reference to the schematic diagram of Fig. 76. Here A represents an amplifier which has a gain A when used as an ordinary amplifier. Regeneration is introduced by superimposing on the amplifier input a fraction β of the output voltage E so that the actual input consists of a signal e_s plus the feedback voltage βE. The effective gain of the amplifier is then[1]

$$\left.\begin{matrix}\text{Gain, taking into}\\ \text{account feedback}\end{matrix}\right\} = \frac{A}{1 - A\beta} \tag{70}$$

In this equation the assumption as to signs is such that when the feedback voltage opposes the signal voltage, β is negative.

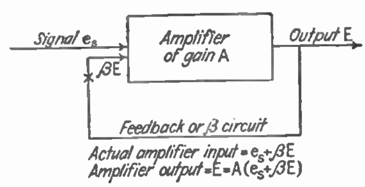


FIG. 76.—Schematic diagram of feedback amplifier.

The quantity $A\beta$ can be termed the feedback factor, and represents the amplitude of the voltage superimposed upon e_s compared with the actual voltage applied to the input terminals. Thus if $A\beta = 50$, then for each millivolt existing between the input terminals, the feedback voltage will be 50 mv. If the phase is such as to give negative feedback, a signal of 51 mv will then be required to produce 1 mv at the amplifier input terminals.

Examination of Eq. (70) shows that the amplification is reduced by the presence of negative feedback. Furthermore, when $A\beta \gg 1$, Eq. (70) reduces to

$$\left.\begin{matrix}\text{Amplification with}\\ \text{large feedback}\end{matrix}\right\} = \frac{-1}{\beta} \tag{71}$$

Expressed in words, Eq. (71) states that, when the feedback factor $A\beta$ is large, the effective amplification depends only upon the fraction β of the output voltage that is superimposed upon the amplifier input, and is *substantially independent of the gain actually produced by the amplifier itself.*

[1] Equation (70) is derived as follows: If E is the output voltage, then the feedback voltage is βE, and the actual input potential is $(e_s + \beta E)$. This input amplified A times must equal E, *i.e.*,

$$(e_s + \beta E)A = E$$

Equation (72) then follows by solving for E/e_s, which is the actual gain.

This remarkable behavior is a result of the fact that, when the feedback is large, the voltage actually applied to the amplifier input terminals represents a small difference between relatively large signal and feedback voltages. A moderate change in the amplification A therefore produces a large change in the difference between signal and feedback voltages, which tends to correct for the alteration in amplification. Thus in the amplifier considered above, where $A\beta = 50$, if the amplification A were halved by a change in design, it would then take 2 mv across the input terminals to deliver the same output as before. With β unchanged, the feedback voltage would still be 50 mv, so it would require a signal of $50 + 2 = 52$ mv instead of the previous 51 mv to produce the same output. Thus a 2 per cent change in effective over-all amplification has resulted when the gain A is altered by 50 per cent.

Inasmuch as the quantity β depends upon circuit elements such as resistances that are permanent, the amplification with large feedback is substantially independent of the tube characteristics and electrode voltages. Furthermore Eq. (71) shows that the amplification with large feedback is inversely proportional to β. Hence if the fraction β of the output voltage that is superimposed upon the input is obtained by a resistance network, the amplification will be substantially independent of frequency and will have negligible phase shift. On the other hand, if it is desired to have the amplification vary with frequency in some particular way, this can be readily accomplished by making the β (or feedback) circuit have the same transmission-loss characteristic as the desired gain characteristic.

The presence of negative feedback also reduces the non-linear distortion produced in the amplifier. This happens because the distortion components of the output voltage are fed back in such a polarity as to produce amplified distortion voltages that tend to cancel the distortion generated in the amplifier. It can be readily shown that as a result of feedback, the distortion is reduced according to the relation.

$$\left.\begin{matrix}\text{Distortion with}\\ \text{feedback}\end{matrix}\right\} = \frac{\left\{\begin{matrix}\text{Distortion in absence}\\ \text{of feedback}\end{matrix}\right.}{1 - A\beta} \tag{72}$$

If $A\beta$ is made large by employing a large amount of feedback, the result is a very great reduction in the ratio of distortion to desired output.

The signal-to-noise ratio is also improved by feedback under certain conditions. A comparison of the signal-to-noise ratios in amplifiers with and without feedback, when the same noise voltage is introduced some place in the amplifier and the signal outputs are the same, shows that

$$\frac{\left.\begin{array}{l}\text{Signal-to-noise ratio}\\ \text{with feedback}\end{array}\right\}}{\left.\begin{array}{l}\text{Signal-to-noise ratio}\\ \text{without feedback}\end{array}\right\}} = \frac{\alpha_f}{\alpha_0(1 - A\beta)} \tag{73}$$

where α_f and α_0 are the amplification between the place at which the noise is introduced and the output, with and without feedback, respectively. Examination of this relation shows that feedback will greatly reduce noises introduced in high-level parts of the amplifier, such as from a poorly filtered power supply in the plate circuit of the final tube. Feedback does not, however, help reduce noise introduced at very low power levels, such as thermal agitation, induced hum, microphonic noises, etc.

Feedback without Oscillations.—In order to realize the advantages of feedback, the amplifier and its feedback must be so arranged that oscillations do not occur. This can be accomplished by making the feedback voltage normally in phase opposition to the applied signal (*i.e.*, $A\beta$ negative and real) and by arranging the circuits so that there is no frequency where $A\beta$ is positive, real, and greater than unity.[1]

No trouble from oscillations need be expected when the feedback takes place between the plate and grid circuits of the same tube. Feedback can also be carried on from the plate circuit of one tube to the grid circuit of the next preceding tube without trouble except when a transformer-coupled stage is involved. However, with more than two stages, oscillations can generally be expected except when special design procedures are employed.

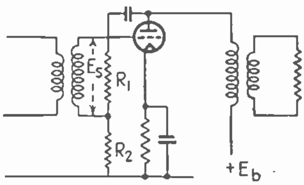


FIG. 77.—Typical feedback amplifier circuit in which a fraction $R_2/(R_1 + R_2)$ of the output voltage is superimposed on the signal voltage E_s.

The circuit of a practical feedback amplifier is illustrated in Fig. 77. Here a pair of voltage-dividing resistors R_1 and R_2 are used to superimpose a fraction $R_2/(R_1 + R_2)$ of the output voltage upon the input voltage. Other circuits, and general design procedures, are to be found in the literature.[2]

49. Behavior of Vacuum Tubes at Ultra-high Frequencies.—In all the discussion that has so far been devoted to amplification it has been assumed that the time required for an electron to travel from the cathode

[1] This is a sufficient condition for avoiding oscillations. There are circumstances, however, in which oscillations are not obtained even though this condition is not satisfied.

[2] In particular see F. E. Terman, Feedback Amplifier Design, *Electronics*, vol. 10, p. 12, January, 1937.

to the plate is negligible compared with the length of time represented by a cycle. At very high frequencies this assumption is no longer permissible.

The result of a finite time for the transit of the electrons modifies the tube characteristics in a number of important respects. In particular, the tube constants such as the amplification factor, mutual conductance (transconductance), and plate resistance become vector quantities with a magnitude and a phase that vary with frequency. The mutual conductance, for example, lags in phase because the finite transit time causes changes in the plate current to lag behind changes in grid voltage. The absolute magnitude of the mutual conductance is also modified slightly. The amplification factor tends to decrease in magnitude and to have an increasingly large phase angle as the frequency increases, while the plate resistance likewise depends upon the frequency.

Analysis shows that the usual equivalent circuits of the amplifier tube still hold, provided the amplification factor, mutual conductance, and plate resistance are given the appropriate values (including both magnitude and phase), which take into account the effect of the finite transit time.

The finite transit time also causes power to be consumed by the grid of the tube even when the grid is biased negatively and attracts no electrons. This comes about as a result of the interchange of energy between the signal voltage acting on the grid and the electrons traveling to the plate. The amount of power lost in this way can be expressed in terms of an equivalent resistance shunted between the grid and cathode electrodes of the tube. Analysis shows that this equivalent input resistance in the case of Class A amplifiers is inversely proportional to the square of the physical size of the tube and inversely proportional to the square of the frequency.

The grid loss occasioned by finite transit time is negligible with standard tubes at broadcast frequencies but becomes increasingly important at higher frequencies. Thus measurements upon a typical Type 57 pentode show that at 1000 kc the input resistance is approximately 21 megohms, but, since the input resistance is inversely proportional to the square of the frequency, at 30 mc it becomes approximately 23,000 ohms, and at 100 mc is only 2100 ohms! As a result, the maximum possible amplification obtainable from a 57 tube falls to less than unity at slightly above 100 mc.

50. Miscellaneous. *Methods of Volume Control.*—In all practical amplifiers it is necessary to provide some means of controlling the amplification so that the output voltage can be set at the desired level irrespective of the signal being amplified. The volume-control arrangement for accomplishing this result should be such that the gain setting has little or no effect upon the frequency response[1] and does not introduce amplitude distortion.

In audio-frequency amplifiers the gain is normally controlled in a resistance-coupled stage by the method shown in Fig. 78, where the grid leak is supplied by a high-resistance potentiometer. The only effect produced upon the amplifier characteristic by such a volume control is a slight improvement in high-frequency response at low volume settings.

The volume of tuned amplifiers using pentode tubes is practically always controlled by varying the mutual conductance of the amplifier tube, either by changing the control-grid bias or the screen-grid potential. This method of controlling volume does not affect the shape of the frequency-response characteristic, and it has the advantage of permitting the amplification of several stages to be controlled simultaneously. Variable-mu tubes are practically always employed when the gain of a tuned

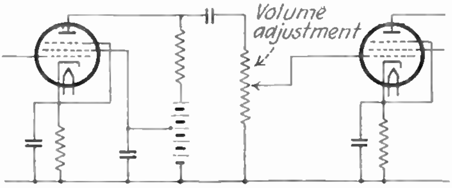


FIG. 78.—Volume-control arrangement normally used in resistance-coupled amplifiers.

amplifier is to be controlled, since these tubes eliminate amplitude distortion and cross-talk troubles at low volume settings (see Sec. 51).

Automatic Volume Control.—Circumstances sometimes arise where it is desired to maintain the output automatically constant irrespective of the input to the amplifier. Thus in a radio receiver it is desirable to maintain constant output when the radio wave being received fades through wide ranges in amplitude.

Automatic volume control of this type is accomplished by rectifying a portion of the amplified radio-frequency carrier voltage, passing the rectified d-c current through a resistance, and utilizing the voltage thereby developed across the resistance to control the volume in such a way as to tend to maintain the output constant. Detailed discussion of automatic-volume-control arrangements used in radio receivers is given in Sec. 99.

Equalization of Amplifier Gain.—Where several stages of amplification are employed in cascade, the way in which the amplification of an indi-

[1] The only exception to this is under conditions where it is desired to compensate for the fact that the sensitivity of the ear to high and low pitches relative to the middle range of frequencies is less for weak sounds than for loud sounds (see Sec. 135). This means that for proper reproduction at low output levels, frequency distortion tending to favor the high and low frequencies is desirable. Volume controls which provide this sort of characteristic are said to be "tone compensated."

vidual stage varies with frequency is not particularly important so long as the individual stages combine to give the desired over-all result. Thus an excessive falling off of the high-frequency response of one stage can be compensated for by introducing a corresponding high-frequency peak in another stage of amplification, as illustrated in Fig. 79. This possibility of compensating for the deficiencies in over-all frequency response of a number of stages by modifying the characteristics of only one of the stages simplifies greatly the problem of obtaining a desired result. In particular, it avoids the cost and trouble involved in attempting to make each individual stage perfect within itself.

Use of Decibels to Express Relative Amplification.—The variation of amplification with frequency in audio-frequency amplifiers is often expressed in decibels referred to some arbitrary level taken as zero deci-

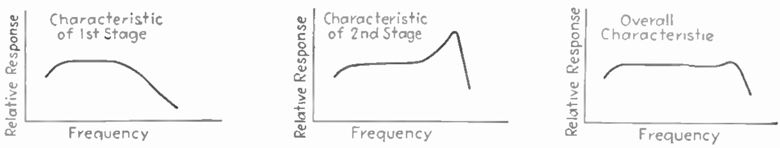


FIG. 79.—Curves illustrating how the deficiencies in the frequency response of one stage of an amplifier may be compensated for by the characteristic of another stage to give a uniform over-all response.

bels. The significance of such curves can be understood by considering what a decibel means. The decibel is a unit for expressing a power ratio, and is given by the relation

$$\text{Decibels} = \text{db} = 10 \log_{10} \frac{P_2}{P_1} \tag{74}$$

where P_1 and P_2 are the two powers being compared. The decibel has no other significance. If it is to be used in expressing relative amplification, it therefore signifies relative power output. Thus, if the output voltage varies with frequency as shown in Fig. 80*a*, one might replot this curve in decibels by assuming some arbitrary power as the standard. This might, for instance, be the power output obtained at 400 cycles. The power output at any other frequency is then proportional to $(E/E_{400})^2$ where E is the voltage output at the frequency in question, and E_{400} is the output voltage at 400 cycles. Since the power output under these conditions is proportional to the square of the voltage, one can rewrite Eq. (74) as follows for this particular case:

$$\text{db} = 10 \log_{10} \frac{P_2}{P_1} = 10 \log_{10} \left(\frac{E}{E_{400}}\right)^2 = 20 \log_{10} \left(\frac{E}{E_{400}}\right) \tag{75}$$

It is now possible to plot a curve giving relative amplification in terms of decibels, as is done at Fig. 80*b*. The significance of the decibel curve can be seen by considering a specific case. Thus the fact that the amplification in Fig. 80*b* is 5 db lower at 45 cycles than at 400 cycles means that the output power at 45 cycles is 0.316 times the power at 400 cycles.[1]

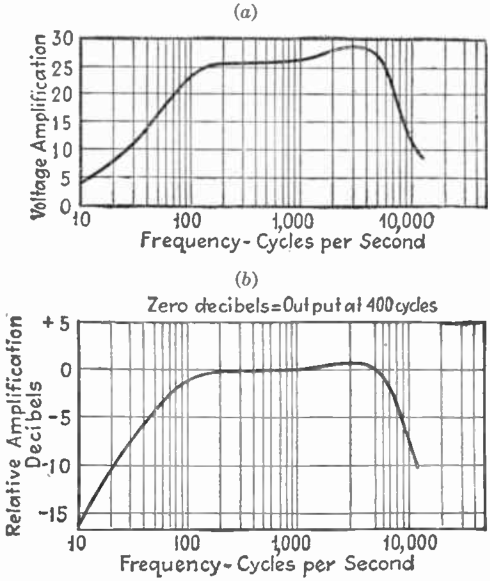


FIG. 80.—Illustration of how relative gain can be expressed in decibels.

51. Amplitude Distortion and Cross-modulation in Amplifiers.—In an ideal amplifier the output wave is exactly the same shape as the input signal. In actual amplifiers the output wave commonly contains harmonics and other components not contained in the signal, particularly when the signal voltage is large. The analysis of amplitude distorting

[1] When working with decibels it is desirable to memorize the following table of approximate equivalents.

TABLE VII

db	Approximate power ratio	Approximate voltage ratio for same resistance
0	1	1
− 1	0.8	0.9
− 3	0.5	0.7
−10	0.1	0.3
−20	0.01	0.1
−30	0.001	0.03

It will be noted that the 70.7 per cent point in amplifiers corresponds to a 3-db loss in amplification.

of this type can be accomplished for the case of a sine-wave signal and a resistance load by the aid of the dynamic characteristic as discussed in Sec. 52. A more comprehensive picture of the distortion in amplifiers, and one that is not limited to sinusoidal voltages and resistance loads, is, however, obtainable with the aid of the power-series method of expressing tube characteristics.

It was shown in Sec. 36 that, for electrode voltages such that the instantaneous plate current never became zero, the characteristics of a tube could be expressed in terms of the following power series:

$$I_p = G_m\left(E_g + \frac{E_p}{\mu}\right) + \frac{1}{2!}\frac{\partial G_m}{\partial E_g}\left(E_g + \frac{E_p}{\mu}\right)^2 + \frac{1}{3!}\frac{\partial^2 G_m}{\partial E_g^2}\left(E_g + \frac{E_p}{\mu}\right)^3 + \cdots \tag{76}$$

where

E_g = signal voltage applied to grid

E_p = voltage acting on plate as result of voltage drop of I_p in load impedance

G_m = mutual conductance at operating point

μ = amplification factor

I_p = plate current caused by the action of E_g and E_p.

Since the voltage E_p appearing in Eq. (76) represents the voltage arising from the current I_p in flowing through the load impedance Z, one can substitute the relation $E_p = -I_pZ$ in Eq. (76). When this is done and the series inverted to give an explicit expression for I_p, the result is[1]

[1] The transformation from Eq. (76) to Eq. (80) is carried out as follows: The substitution of $E_p = -ZI_p$ into Eq. (76) gives

$$I_p = G_m\left(E_g - \frac{I_pZ}{\mu}\right) + \frac{1}{2!}\frac{\partial G_m}{\partial E_g}\left(E_g - \frac{I_pZ}{\mu}\right)^2 + \cdots \tag{77}$$

This gives I_p as an implicit function of E_g, whereas what is wanted is an explicit solution for I_p of the form

$$I_p = \alpha_1 E_g + \alpha_2 E_g^2 + \alpha_3 E_g^3 + \cdots \tag{78}$$

The inversion from Eq. (77) to Eq. (78) is accomplished by substituting the value of I_p in the form of Eq. (78) for I_p in Eq. (77) and then equating the coefficients of like powers of E_g on both sides of the result to determine the values that the α's must have in Eq. (78) to satisfy Eq. (77). Doing this gives

$$\begin{aligned}
\alpha_1 &= \frac{\mu}{R_p + Z_1} \\
\alpha_2 &= \frac{\dfrac{1}{2!\mu G_m}\dfrac{\partial G_m}{\partial E_g}\left(\dfrac{\mu R_p}{R_p + Z_1}\right)^2}{R_p + Z_2} \\
\alpha_3 &= \frac{\dfrac{1}{3!\mu^2 G_m}\dfrac{\partial^2 G_m}{\partial E_g^2}\left(\dfrac{\mu R_p}{R_p + Z_1}\right)^3 - \dfrac{1}{\mu G_m}\dfrac{\partial G_m}{\partial E_g}\left(\dfrac{\mu R_p}{R_p + Z_1}\right)(\alpha_2 Z_2)}{R_p + Z_3}
\end{aligned} \tag{79}$$

Equation (80) now follows by substituting these values of α into Eq. (78).

$$I_p = \frac{\mu E_g}{R_p + Z_1} + \frac{1}{2!\mu G_m}\frac{\partial G_m}{\partial E_g}\frac{E_1^2}{R_p + Z_2} + \frac{\left(\frac{1}{3!\mu^2 G_m}\frac{\partial^2 G_m}{\partial E_g^2}E_1^3 - \frac{1}{\mu G_m}\frac{\partial G_m}{\partial E_g}E_1E_2\right)}{R_p + Z_3} + \cdots \quad (80)$$

where

R_p = plate resistance at the operating point

Z = load impedance. The subscripts 1, 2, and 3 denote the impedance offered to the first-, second-, and third-order components of the plate current. When any one order of plate current contains components of several frequencies, the appropriate value of Z must be used for each component

$E_1 = \mu E_g\left(\frac{R_p}{R_p + Z_1}\right)$ = voltage drop produced in plate resistance by first-order component of current

$E_2 = \left(\frac{Z_2}{R_p + Z_2}\right)\frac{1}{2!\mu G_m}\frac{\partial G_m}{\partial E_g}E_1^2$ = voltage drop produced across the load impedance Z_2 by the second-order component of current.

Study of Eq. (80) shows that the various components of the plate current may be considered as being produced by a series of equivalent

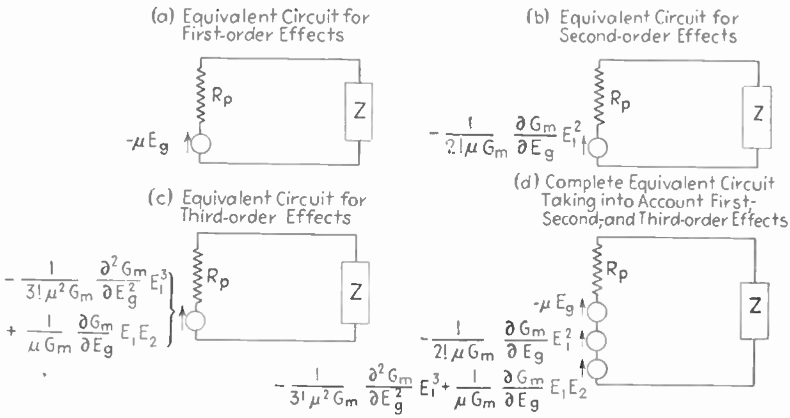


FIG. 81.—Equivalent circuits taking into account first-, second-, and third-order components of the plate current.

voltages acting in a circuit consisting of the plate resistance of the tube in series with the load impedance. If the positive direction is taken as from cathode toward plate (the opposite of the positive direction in Eq. (80)), these equivalent voltages and circuits are as shown in Fig. 81.

It will be noted that the first-order effects correspond to the equivalent circuit of the amplifier discussed in Sec. 40. This is now seen to be the first-order approximation that results when the characteristic curve of the tube is considered to be a straight line. The equivalent second-order

TABLE VIII.—FIRST-, SECOND-, AND THIRD-ORDER VOLTAGES IN TWO TYPICAL CASES

	Signal $= E_a \sin \omega t$	Signal $= E_a \sin \omega_a t + E_b \sin \omega_b t$
Equivalent first-order voltage	$-\mu E_a \sin \omega t$	$-\mu(E_a \sin \omega_a t + E_b \sin \omega_b t)$
Equivalent second-order voltage	$-\frac{\mu}{2!G_m}\frac{\partial G_m}{\partial E_g}E_a^2\left(\frac{1 - \cos 2\omega t}{2}\right)$	$-\frac{\mu}{2!G_m}\frac{\partial G_m}{\partial E_g}\Big(\frac{E_a^2 + E_b^2}{2} - \frac{E_a^2 \cos 2\omega_a t + E_b^2 \cos 2\omega_b t}{2}$ $+ E_aE_b \cos (\omega_a + \omega_b)t + E_aE_b \cos (\omega_a - \omega_b)t\Big)$
First part of equivalent third-order voltage[1]	$-\frac{\mu}{3!G_m}\frac{\partial^2 G_m}{\partial E_g^2}E_a^3\Big(\frac{3}{4} \sin \omega t$ $- \frac{1}{4} \sin 3\omega t\Big)$	$-\frac{\mu}{3!G_m}\frac{\partial^2 G_m}{\partial E_g^2}\Big\{\left(\frac{3}{4}E_a^3 + \frac{3E_aE_b^2}{2}\right) \sin \omega_a t$ $+\left(\frac{3}{4}E_b^3 + \frac{3E_a^2E_b}{2}\right) \sin \omega_b t - \frac{E_a^3 \sin 3\omega_a t + E_b^3 \sin 3\omega_b t}{4}$ $- \frac{3}{4}E_a^2E_b[\sin (2\omega_a + \omega_b)t + \sin (-2\omega_a + \omega_b)t]$ $- \frac{3}{4}E_aE_b^2[\sin (\omega_a + 2\omega_b)t + \sin (\omega_b - 2\omega_b)t]\Big\}$

[1] The second part of the third-order voltage is similar in character to the first-order part, although differing in detail.

voltage takes into account, to a first approximation, the error involved in the first-order equivalent circuit. The first- and second-order effects taken together assume that the characteristic curves of the tube are portions of a parabola. The third-order component of the equivalent voltage represents to a first approximation the error involved when only the first- and second-order components are taken into account.

Practical Application to the Analysis of Amplifier Distortion.—The nature of the distortion that occurs in Class A amplifiers can be determined by calculating the equivalent voltage acting in the plate circuit of the tube for the particular signal involved, using the equivalent circuit of Fig. 81*d*. The first-order component of this equivalent voltage is merely an enlarged replica of the applied signal and so needs no comment. The second-order component is obtained by writing down the voltage E_1 as a function of time, as determined from the results of the first-order term, and substituting this in the expression for the second-order voltage. The third-order equivalent voltage is similarly obtained by substituting appropriate expressions for the applied signal, for the first-order result E_1, and the second-order result E_2, all of which must be written as functions of time. This is done in Table VIII for signals consisting of a simple sine wave, and of two sine waves of different frequencies.

The first-order effects given in Table VIII represent the undistorted part of the amplifier output, and need no discussion. The second-order equivalent voltages are seen to give rise to second harmonics, to sum and difference frequencies, and to a d-c (or rectified) current.

The third-order component of the equivalent voltage produces third harmonics and third-order combination frequencies. There is also a third-order component having the same wave shape as the applied signal, but an amplitude proportional to the cube of the signal. This component combines with the first-order part of the output to introduce a lack of proportionality between the signal and the amplitude of the undistorted part of the output. Finally, when the applied signal consists of two frequencies there is a third-order portion of the output having the same frequency as one component of the signal, but an amplitude dependent upon the amplitude of the other component. This particular term gives rise to *cross-modulation*, since it makes the output on one frequency depend upon the presence or absence of other frequencies.

Problems

1. A Type 56 tube is to be used under conditions given in Table V, page 85. Determine the bias resistance required.

2. Determine the bias resistance required when a 6C6 tube is to be used at operating conditions given in Table V, page 85.

3. *a.* An alternating potential of 2 volts effective of a frequency of 1000 cycles is applied to the grid of a triode having $\mu = 14$, $R_p = 10{,}000$, and a resistance load of

12,000 ohms. Calculate (1) alternating current in the load, (2) alternating voltage across the load, (3) alternating power dissipated in the load, and (4) ratio of voltage across load to the voltage applied to the grid.

b. Repeat the above for a load consisting of a 2-henry inductance.

4. Repeat Prob. 3, using the constant-current generator form of the equivalent circuit.

5. One volt is applied to the grid of a 56 tube operating at the conditions listed in Table V, page 85. Calculate and plot a curve giving voltage developed across the load impedance as a function of load for resistance loads from 0 to 30,000 ohms.

6. Repeat Prob. 5 for the case of a 6C6 pentode operating under conditions listed in Table V, page 85.

7. Calculate and plot the amplification characteristic that would result if the amplifier of Fig. 58 had $R_{gl} = 1$ megohm, $C_c = 0.01$ μf, and delivered its output voltage to a tube having an input capacitance of 5 $\mu\mu$f (pentode tube). Assume that the output capacitance of the amplifier is 4 $\mu\mu$f and that the wiring capacitance is 10 $\mu\mu$f. Make use of the universal amplification curve in the calculation.

8. A resistance-coupled amplifier employs a pentode tube operated under conditions such that the mutual conductance is 300 μmho. If the coupling and grid-leak resistances are 500,000 and 1,000,000 ohms, respectively, and the coupling and estimated shunting capacitances are 0.005 μf and 35 $\mu\mu$f, respectively, calculate and plot the curve of amplification as a function of frequency. Use the universal amplification curve in the calculations, and assume the plate resistance is 3 megohms.

9. *a*. Design a resistance-coupled amplifier, using the pentode tube of Figs. 39 to 42, when the plate-supply potential is 275 volts and when only the audio-frequency range is to be covered. In this design specify suitable values for coupling resistance, grid-leak resistance, and coupling condenser. Select a grid-bias and screen voltage so that between one-third and one-half the plate-supply voltage will be actually applied to the plate after allowing for voltage drop in the coupling resistance. Also specify the grid-bias resistance and the capacitance of the bias by-pass condenser.

b. Calculate and plot the amplification as a function of frequency by making an estimate of the mutual conductance and plate resistance at the operating point, and assuming the total shunting capacitance to be 35 $\mu\mu$f.

10. *a*. Design a resistance-coupled amplifier to cover the frequency range 15 to 100,000 cycles with an amplification that is always at least 0.90 of the mid-frequency value. Use a 6C6 tube with a plate-supply potential voltage of 250 volts, and assume that the shunting capacitances amount to 15 $\mu\mu$f. Specify in the design the coupling and grid-leak resistance, the coupling-condenser capacitance, suitable grid-bias and screen voltages, and the proper bias resistance. Use plain resistance coupling rather than a resistance-inductance-coupled combination.

b. Calculate the amount of amplification that would be expected in the mid-frequency range, assuming reasonable values of plate resistance and mutual conductance for the operating point selected.

11. The direct-current leakage resistance of a condenser is inversely proportional to the capacitance of the condenser, so that the product of leakage resistance and capacitance is fixed for any given type of dielectric, although dependent upon the type of dielectric employed. Keeping this in mind, explain why: (*a*) trouble from direct-current leakage in the coupling condenser increases as the low-frequency response is made better, (*b*) the best low-frequency response that it is practicable to obtain is determined primarily by the character of the dielectric of the coupling condenser, and (*c*) rearranging the proportions of grid-leak resistance and coupling-condenser capacitance while maintaining the low-frequency response unchanged has very little effect on trouble from coupling-condenser leakage.

12. Using data from tube manuals, and assuming that the wiring adds a capacitance of 15 $\mu\mu$f, make an estimate of the total capacitance shunting the output of a resistance-coupled amplifier using a 6C6 pentode tube when the amplified voltage is delivered to: (1) another 6C6 tube, (2) a 47 tube, (3) a 75 tube operating as a resistance-coupled amplifier with a plate-supply potential of 250 volts, (4) a 56 tube operating as a resistance-coupled amplifier, and (5) a 2A3 tube.

13. It is stated in the text that if conditions in a pentode resistance-coupled amplifier are such that the direct-current voltage actually applied to the plate (after allowing for the voltage drop in the coupling resistance) is insufficient to prevent the formation of a virtual cathode in front of the suppressor, the amplification becomes small. Explain why this is the case.

14. A resistance-coupled amplifier is to be designed to use a pentode tube and to have a given high-frequency response characteristic. Explain why the amplification obtainable under these conditions first increases as the plate-supply potential is increased and then becomes constant with still further increases of plate-supply voltage.

15. Explain why in triode resistance-coupled amplifiers the amplification obtainable is roughly proportional to the amplification factor of the tube, while the high-frequency limit is generally greater the lower the amplification factor.

16. Derive Eq. (52*b*). In doing this start by using Thévenin's theorem to simplify the network on the tube side of the coupling condenser.

17. A certain transformer has the following characteristics:

Primary inductance at rated primary current	= 30 henries
Primary inductance with secondary shorted	= 0.15 henry
Ratio of transformation (voltage ratio at low frequencies)	= 3.0
Series resonant frequency of secondary (with no tube shunted across the secondary)	= 13,000 cycles
R_e so large as to be of no consequence	
Primary direct-current resistance	= 500 ohms
R_s' (actual value before reducing to unity-turn ratio)	= 10,000 ohms

This transformer is to be used with a 56 tube having an amplification factor of 14 and a plate resistance of 11,000 ohms. The input capacitance of the following tube is estimated as 50 $\mu\mu$f. Calculate and plot the way in which the amplification would be expected to vary with frequency. Make use of the universal amplification curve of Fig. 63 in this calculation.

18. What would be the best value of plate resistance to use with the transformer of Prob. 17?

19. Derive Eqs. (56) and (57).

20. Assume that in Fig. 62 the eddy-current resistance R_e that is actually in shunt with the transformer primary is 250,000 ohms. Calculate and plot the frequency-response characteristic obtained when R_e is taken into account and replot Fig. 62 on the same curve sheet to show the error that resulted from neglect of R_e in Fig. 62.

21. In a transformer-coupled amplifier it is found that the frequency-response characteristic at high frequencies is different when the amplified voltage is applied to the grid of a pentode tube than when it is applied to the grid of a triode tube. Explain.

22. A resistance shunted across the primary terminals of the transformer in a transformer-coupled amplifier tends to make the high-frequency response peaked, extends the region of substantially constant response to lower frequencies, and reduces the mid-frequency amplification. Explain, using Thévenin's theorem to simplify the network consisting of tube and shunting resistance that is on the tube side of the primary terminals.

23. A transformer-coupled amplifier is to cover the frequency range 80 to 8000 cycles with a response at least 70.7 per cent and not more than 110 per cent of the mid-frequency value. If the tube has a plate resistance of 10,000 ohms, specify the primary inductance, the leakage inductance reduced to unity-turn ratio, and the frequency of series resonance (when the load capacitance is connected across the secondary). Assuming that the turn ratio will be 3.5 and that the input capacitance of the tube to which the amplified voltage is delivered is 50 $\mu\mu$f, how much distributed capacitance should the secondary of the transformer have, and what is the series resonance frequency desired in the transformer when the secondary is not connected to the next tube?

24. A transformer is to couple a 500-ohm line to the grid of a tube. Assuming that the turn ratio is to be 15, specify the minimum allowable primary inductance, the maximum allowable leakage inductance (referred to the primary), and the desired series resonance frequency, if the response is not to fall to less than 70.7 per cent or to rise above 110 per cent of the mid-frequency response over the range 60 to 11,000 cycles.

25. A direct-coupled tuned amplifier uses coil No. 3 in Fig. 10. Assuming that the losses of the tuning capacitances are negligible, calculate and plot the amplification as a function of frequency up to 30 kc on each side of resonance for a resonant frequency of 1000 kc, when a 6D6 tube is used under conditions given in Table V, page 85, and the grid-leak resistance is 1 megohm.

26. Calculate and plot the amplification at resonance as a function of resonant frequency over the range 550 to 1500 kc for the amplifier of Prob. 25.

27. In Prob. 26, calculate and plot the frequency band for which the response is at least 70.7 per cent of the response at resonance, as a function of resonant frequency for the range 550 to 1500 kc.

28. The amplifier of Prob. 25 is changed to transformer coupling as shown in Fig. 67*b*. It is desired to obtain an amplification of 50 at 1000 kc. Assuming a coefficient of coupling of 0.50 between primary and secondary, calculate the required mutual inductance and primary inductance. Assume $[(\omega M)^2/R_s]/R_p \ll 1$.

29. In Prob. 28, if the amplification at resonance to be constant at 50 for the frequency range 550 to 1500 kc, calculate and plot the way in which the required mutual inductance must vary as a function of frequency. Assume $[(\omega M)^2/R_s]/R_p \ll 1$.

30. *a*. In a typical intermediate-frequency amplifier of the band-pass type shown in Fig. 71*a*, the primary and secondary coils each have inductances of 4 mh and are tuned to 260 kc. The coil Q's are 50 and the coupling is adjusted so that $k = 0.03$. If a 6D6 amplifier tube is used, calculate the amplification at the common resonant frequency, and at the two peaks of the response curve. Determine the band width between peaks, and then sketch the shape of the amplification curve in the vicinity of resonance.

b. Repeat (*a*) when $k = 0.05$.

31. Derive the following equations in the text: (*a*) Eq. (59*a*), (*b*) Eq. (61*a*), (*c*) Eq. (64*a*), and (*d*) Eq. (60). In the last case start by deriving the equation of secondary current, and from this deduce the equivalent Q of the secondary-current curve.

32. Explain why a transformer in a shield of permalloy or similar alloy reduces the hum pick-up as compared with a cast-iron shield.

33. What is the effective value of the noise voltage in the output of an amplifier having an over-all voltage gain of 100,000 times, if the input resistance across the first grid to filament is 100,000 ohms and the amplifier gives substantially constant gain over the band 40 to 8000 cycles and very little gain outside this range?

34. In a triode resistance-coupled amplifier:

$$\begin{aligned} \mu &= 100 \\ R_p &= 125{,}000 \text{ ohms} \\ R_C &= 250{,}000 \text{ ohms} \\ R_{GL} &= 500{,}000 \text{ ohms} \\ C_C &= 0.01\ \mu\text{f} \\ C_s &= 60\ \mu\mu\text{f} \\ C_{gp} &= 1.7\ \mu\mu\text{f} \\ C_{gf} &= 1.7\ \mu\mu\text{f}. \end{aligned}$$

Calculate and plot the input capacitance and input resistance as a function of frequency for the frequency range 1000 to 30,000 cycles.

35. Discuss the way in which the magnitude of the input capacitance and input resistance of an unneutralized tuned triode amplifier varies with frequency. Include a sketch showing type of variations to be expected.

36. *a.* Design filters for the first two stages in the example of Sec. 47 if the amplification per stage is increased to 110, and discuss the effect of the total amplification upon the tendency to regenerate.

b. Repeat (*a*) when the lowest frequency for which the filters are to be effective is 15 cycles, and the gain per stage is 60. Compare the results with those of the example in the text and explain why it is important that the low-frequency response be no better than actually required.

37. A three-stage resistance-coupled amplifier is to be constructed following the first column in Table VI, page 104. Design suitable filter systems for the plate leads of the first two stages so that the regeneration voltages acting in the plate circuits of the first two tubes will not, at frequencies exceeding 25 cycles, be greater than 10 per cent of the voltages which these tubes apply to the succeeding grids. The common plate impedance is an 8-μf condenser.

38. Regeneration in tuned radio-frequency amplifiers is normally most pronounced at the high-frequency end of the tuning range. Explain.

39. Devise a circuit for applying negative feedback to a two-stage resistance-coupled amplifier in which the feedback takes place from the output of the second stage to the input of the first stage.

40. In the feedback circuit of Fig. 77, the use of feedback makes the voltage developed across the load resistance fall off less at low frequencies than when there is no feedback, but does not reduce the falling off at high frequencies. Explain.

41. Sketch the circuit diagram of a volume control suitable for use with an impedance-coupled amplifier.

42. *a.* What would be the objections to varying the amplification in the resistance-coupled amplifier of Fig. 58 by varying the grid-leak resistance?

b. What would be the disadvantage of controlling the amplification of a transformer-coupled amplifier by varying a resistance shunted across the secondary?

43. *a.* Plot the amplification curve of Fig. 58 in decibels, assuming that the gain at 1000 cycles is taken as zero decibels.

b. What is the gain in decibels corresponding to zero decibels in (*a*)?

44. Analysis of the characteristic curves of a particular triode tube shows that at the operating point $\mu = 10$, $R_p = 50{,}000$, $G_m = 200$ μmho, $\partial G_m/\partial E_g = 10^{-5}$, and $\partial^2 G_m/\partial E_g^2 = -10^{-7}$. Calculate the exact currents that will flow in the plate circuit up to and including third-order effects, when an alternating signal potential of 2 volts crest value is applied to the grid of the tube and the load resistance is negligible.

45. The amount of second-harmonic distortion produced in an amplifier when the applied signal is a sine wave is often estimated roughly by noting the change in d-c plate current produced by the presence of the signal. Demonstrate with the aid of Table VIII that this rectified plate current is proportional to the amount of second harmonic present even when all effects up to and including the third order are taken into account.

46. An amplifier is overloaded so that the tube operates sufficiently far into the curved part of its characteristic to make effects up to and including the third order important. If the signal applied to the input consists of a 200-cycle wave and its sixth harmonic, 1200 cycles, list the frequencies of the components present in the output of the amplifier.

CHAPTER VI

POWER AMPLIFIERS

52. Class A Power Amplifiers Employing Triode Tubes.—In a power amplifier the object is normally to develop as much power output as possible without much regard to the signal voltage required to accomplish this result or the voltage at which this power is obtained. This is in contrast with the voltage amplifier, where the object is to increase a relatively small signal voltage as much as possible. In the Class A amplifier the tube is operated so that plate current flows at all times, and so the output voltage will have as nearly as possible the same wave shape as the signal voltage.

The problems involved in realizing the full possibilities of a triode tube as a Class A power amplifier can be understood with the aid of Fig. 82. Consider first the action in the circuit of Fig. 82*a*, in which the load resistance is connected directly in series with the plate circuit of the tube. The voltage actually at the plate of the tube in the absence of an applied signal has a value E_b, and this together with the grid bias E_c and corresponding plate current I_b determines an operating point such as indicated by O in Fig. 82*d*. An alternating signal voltage applied to the grid of the tube superimposed upon the grid bias then causes variations in the plate current and hence in the voltage drop in the load resistance. The voltage at the plate of the tube then varies likewise, since this voltage is the plate-supply potential minus the drop in the load resistance. In particular, at the positive peak of the signal the instantaneous plate current tends to be large, and the instantaneous plate voltage hence is lower than the voltage E_b at the operating point. Similarly, at the negative peak of the signal voltage the instantaneous plate current is small, and the instantaneous plate potential is larger than the voltage E_b at the operating point. The way in which the instantaneous operating condition varies when the signal voltage is applied is given by the line that is drawn through the operating point O and marked *dynamic characteristic* in Fig. 82*d*.

The dynamic characteristic gives the relation that exists between the input voltage wave and the output voltage wave, as shown in Fig. 82*d*. Thus with the operating point at $E_b = 160$ volts, $E_c = -30$ volts, and $I_b = 21.5$ ma, as shown, a signal having a crest amplitude of 30 volts varies the instantaneous plate current from a maximum value I_{max} of 38.5 ma to a minimum value I_{min} of 6.5 ma. At the same time the instantaneous plate potential varies from a minimum value E_{min} of about 95 volts to a maximum E_{max} of 217 volts, with the maximum coming when the instantaneous current is a minimum. The dynamic characteristic does not follow the static characteristic curve of the tube because

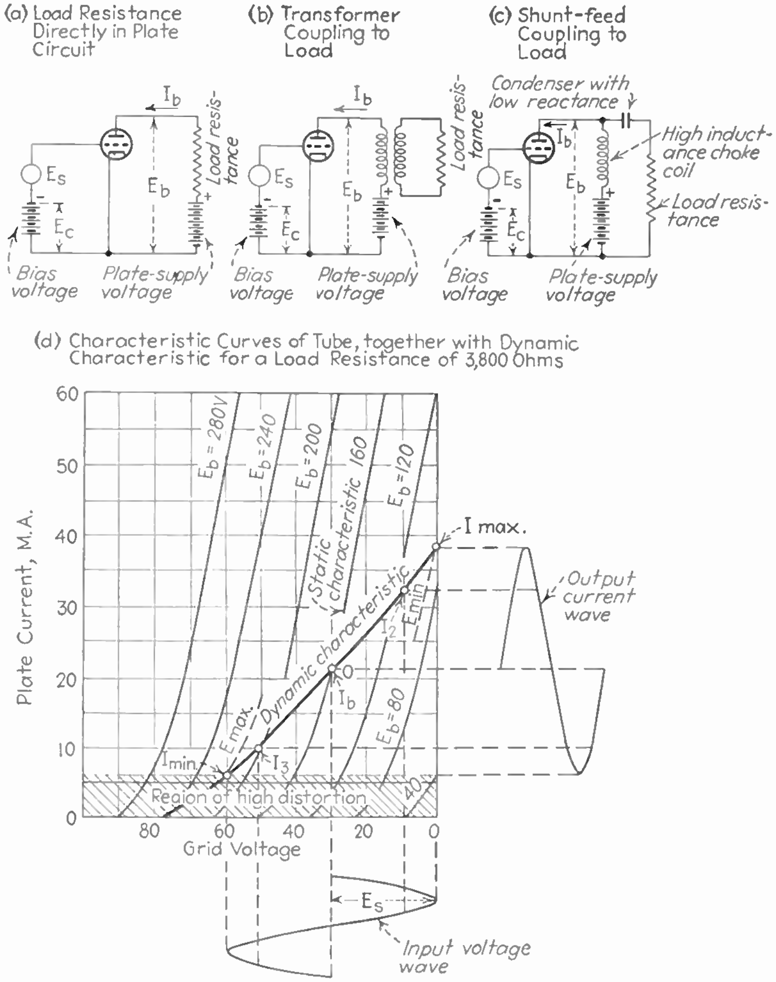


FIG. 82.—Circuits of power amplifier together with dynamic characteristic suitable for Class A operation without driving the grid positive, and resulting input and output waves.

of the variation of instantaneous plate voltage that is produced by the variations in signal voltage E_s applied to the grid.

The circuit of Fig. 82*a* has the disadvantage that a considerable amount of direct-current energy is dissipated in the load resistance. This can be avoided by coupling the load into the plate circuit with the aid of a transformer, as shown at Fig. 82*b*, or by use of the shunt-feed circuit of Fig. 82*c*. With these arrangements the plate-supply voltage differs from the operating voltage E_b only by the resistance drop in the choke or transformer, which is quite small. The action insofar as the dynamic characteristic is concerned is, however, essentially the same as with the arrangement of Fig. 82*a*. This is because of the action of the inductance through which the plate current must flow. Thus if in Fig. 82*b* the grid potential suddenly becomes sufficiently negative to reduce the plate current to a small value, this change in current induces a voltage across the primary inductance that adds to the plate-supply potential and increases the instantaneous plate potential above the operating value and hence above the plate-supply voltage. As long as the average or d-c plate current does not change when the signal voltage is applied, the same dynamic characteristic applies to circuits *a*, *b*, and *c* of Fig. 82, provided the operating point is the same.[1]

In order for the output wave in Fig. 82*d* to have the same shape as the signal, it is necessary to restrict the operating range to the portion of the dynamic characteristic that is substantially linear. The practical limiting factors are the excessive curvature of the dynamic characteristic at low plate currents and the fact that, if the grid is driven positive, the grid current that results tends to distort the input wave.

For proper operation it is necessary to have a careful balance between the operating point, signal voltage, and load resistance. Thus for a given operating point there is a particular load resistance that will utilize the amplifier possibilities to the best advantage. This is illustrated in Figs. 83 and 84. For the operating point used in Fig. 83, the load resistance giving optimum operation for the case of no grid current is marked *b*. With this load resistance the dynamic characteristic is such that a signal voltage that is just sufficient to bring the instantaneous grid potential to zero potential at the positive peaks of the signal cycle will at the negative peaks bring the instantaneous grid potential to the point where

[1] When the d-c current does change when a signal is applied, the dynamic characteristic of Fig. 82*d* is exactly correct only for circuit *a*. With circuits *b* and *c* any change in d-c current produces little or no change in d-c plate voltage, whereas the dynamic characteristic of Fig. 82*d* assumes that the change in d-c current flows through the load resistance. This results in an error which, however, is quite small unless the distortion of the output wave is large. Thus in Fig. 82, the d-c component of plate current for a signal of 30 volts crest is ½ ma, corresponding to a drop of only 1.9 volts in the 3800-ohm resistance.

The error arising from the change in plate current caused by applying a signal can be eliminated by drawing the dynamic characteristic as though the plate potential at the operating point was $E_b + R_L \Delta i$ instead of E_b, where E_b is the plate potential actually at the operating point, and Δi is the increase in plate current caused by the presence of the applied signal. Inasmuch as Δi depends on the way in which the dynamic characteristic is drawn, it is necessary to locate the dynamic characteristic by a trial-and-error process if an exact correction for Δi is to be obtained.

the dynamic characteristic begins to be excessively curved. The distortion of the output wave is then small. With a higher load resistance, corresponding to the dynamic characteristic marked *c*, the same signal will fail to reach the region of excessive curvature and the output power will be less. On the other hand, a low resistance, such as that corresponding to the dynamic characteristic marked *a*, causes operation to extend into the region of excessive curvature, even with signal voltages that do not drive the grid positive, and results in high distortion.

Operation so that the grid is driven positive at the peak of each signal cycle requires that the load resistance be increased to a value greater than

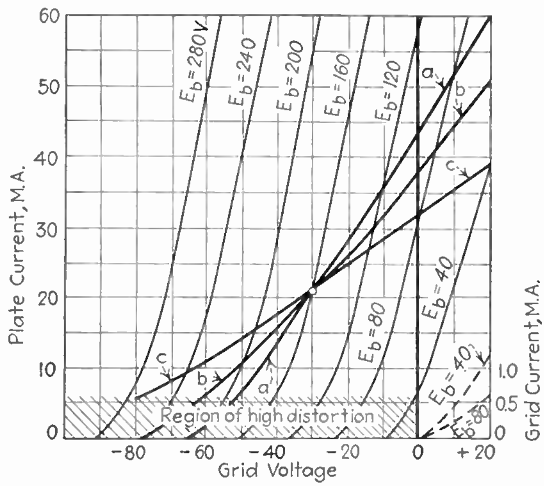


FIG. 83.—Characteristic curves of power amplifier together with dynamic characteristics for three different values of load resistance. The dotted lines show grid current. Note that if the grid is driven positive the proper load resistance is greater than negative-grid operation with the same operating point.

if the grid is maintained negative. This is apparent from Fig. 83, where, for a signal of 50 volts crest (grid driven 20 volts positive) shown, the optimum load resistance corresponds to dynamic characteristic *c*.

As the grid bias of the amplifier is made more negative without changing the operating plate voltage, the optimum load resistance becomes greater, as shown in Fig. 84. This is irrespective of whether or not the grid is driven positive, although for any given operating point, positive-grid operation calls for a higher load resistance than when the grid must be negative at all times.

Power Relations and Efficiency.—The alternating voltage developed across the load resistance in the plate circuit of a power amplifier swings the actual plate potential from a maximum value E_{max} to a minimum value E_{min}, as shown in Fig. 82*d*, while the current swings from a value I_{max} to I_{min}. The alternating components of voltage and current have peak amplitudes that are half these total swings. The power delivered to the load is equal to half the product of peak a-c voltage and peak a-c current, and so is given by the equation:

$$\text{Power output} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8} \tag{81}$$

The plate efficiency of a power amplifier is equal to the ratio of useful power delivered to the load as given by Eq. (81), to the d-c power sup-

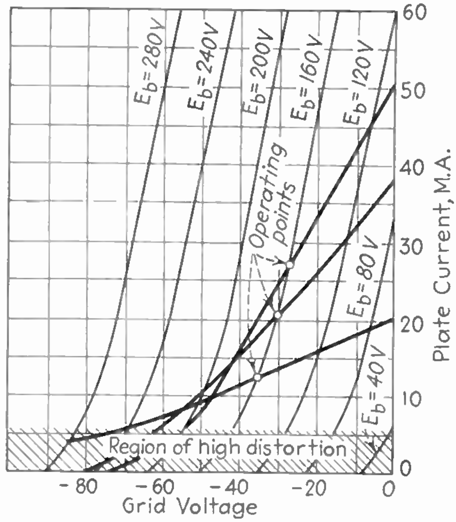


FIG. 84.—Characteristic curves of power amplifier together with dynamic characteristics for optimum load resistance for negative-grid operation with different operating points. Note that, as the operating grid bias becomes more negative, a higher load resistance is required if the operating plate potential is unchanged.

plied the plate circuit. The latter is equal to the product of plate battery voltage E_b and plate current I_b. The plate efficiency is hence

$$\text{Plate efficiency} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8E_bI_b} \tag{82}$$

This expression for efficiency is correct even when moderate distortion is present provided the d-c plate current I_b used in Eq. (82) is taken to mean the d-c plate current actually flowing when the signal is applied. That is to say, the d-c plate current used in Eq. (82) should include the plate current present in the absence of signal, plus whatever change in d-c current is produced by the non-linear or rectifying action of the tube.

The maximum efficiency that it is possible to obtain in a Class A amplifier is 50 per cent. This is because E_{min} and I_{min} can never be less than zero, while E_{max} and I_{max} cannot exceed twice the operating voltage and current, respectively. The actual efficiency is always less than 50 per cent and depends upon the ratio of E_{min}/E_b and I_{min}/I_b. These ratios depend upon operating conditions. When the operation is such that the grid is not driven positive, the efficiency is commonly 15 to 25 per cent with small power tubes such as used in radio receivers and may reach 40 per cent with tubes operated at higher plate voltage. Operation so that the grid is driven positive usually results in increased efficiency because it is then possible to draw the maximum plate current I_{max} through the tube at the positive peaks of the cycle with a smaller minimum plate potential E_{min} than is possible when the grid is not positive. The efficiencies realized under practical conditions with the grid driven positive normally reach 30 to 40 per cent.

Selection of Operating Conditions.—The basis of adjusting a power amplifier for best results varies with the type of operation desired. The principal factors to be considered are whether or not the grid is to be driven positive, the allowable plate dissipation of the tube, and the allowable plate voltage at the operating point.

The choice between operating a Class A amplifier with or without grid current depends upon the circumstances of the particular case. When the grid is not allowed to go positive, there is no distortion of the exciting voltage, less exciting power is required, and the amplifier is less difficult to design. At the same time, the plate efficiency and hence the output power are generally less than with positive-grid operation. The economics of the situation are such that allowing the grid to go positive becomes increasingly desirable as the power involved increases, because then the savings resulting from improved efficiency are of greater importance. In addition, the development of negative feedback has made it possible to compensate for the distortion introduced by grid current and thereby has reduced to some extent the advantages of negative-grid operation.

In adjusting a power amplifier for positive-grid operation the grid bias is selected so that the rated d-c plate current is obtained at normal plate voltage. The amount the grid is to be driven positive is then decided upon, and the proper load resistance for these conditions is selected.

When it is desired to operate without driving the grid positive, the maximum "undistorted" power output[1] is obtained when the load resistance approximates twice the plate resistance. This is the basis of adjustment normally used with power tubes operating at plate voltages of 250 to 300 volts and gives plate efficiencies of 15 to 20 per cent. Tubes intended for operation at higher plate potentials are usually so designed

[1] The term "*undistorted*" *power output* used in connection with audio amplifiers means that the distortion is small enough so as not to be excessive. In audio-frequency power amplifiers the allowable distortion is commonly arbitrarily taken as 5 per cent, which approximates the value that is just noticeable to the ear.

that, if the grid bias is the proper value for a load resistance twice the plate resistance, the allowable plate current is exceeded. With such tubes the negative grid bias must be increased until the rated plate current is obtained. The proper load resistance is then greater than twice the plate resistance. This raises the plate efficiency correspondingly, with values of 30 to 40 per cent often being possible.

The final determination of load resistance, irrespective of whether or not positive-grid operation is used, is preferably made by trying out various values of load resistance and selecting the one giving the most output with reasonable distortion. Graphical methods of doing this are given below.

Calculation of Distortion.—The dynamic characteristic of a Class A power amplifier gives the relation between instantaneous plate current and instantaneous signal voltage on the grid, for the case of a resistance load. The dynamic characteristic can hence be used to determine the wave shape of the output voltage as illustrated in Fig. 82*d*. In an ideal amplifier the dynamic characteristic would be a straight line over the operating range, so that the output wave would exactly reproduce the signal voltage. Actually, however, there is distortion because the characteristic curves of the tubes are not straight lines. Where the tube characteristics do not have an inflection point over the operating range, the principal distortion produced in the case of a sine-wave signal consists of a second-harmonic component, which is accompanied by an increase in the d-c plate current (commonly termed the "rectified" plate current). These second-harmonic and d-c components, as well as the fundamental component of the output, can be evaluated from a knowledge of the instantaneous plate current at the operating point, and at the maximum and minimum points on the dynamic characteristic, according to the relations[1]

$$\text{Direct-current component} = A_0 = \frac{I_{\text{max}} + I_{\text{min}} - 2I_b}{4} \tag{83a}$$

$$\text{Fundamental} = A_1 = \frac{I_{\text{max}} - I_{\text{min}}}{2} \tag{83b}$$

$$\text{Second harmonic} = A_2 = \frac{I_{\text{max}} + I_{\text{min}} - 2I_b}{4} \tag{83c}$$

$$\frac{\text{Second harmonic}}{\text{Fundamental}} = \frac{A_2}{A_1} = \frac{I_{\text{max}} + I_{\text{min}} - 2I_b}{2(I_{\text{max}} - I_{\text{min}})} \tag{83d}$$

[1] Equations (83) and (84) assume that the load offers the same resistance to the "rectified" plate current as to the alternating components. In the usual case the load is coupled to the plate circuit of the tube by an output transformer, and the resistance of the load to the rectified current is accordingly zero. Under such circumstances Eqs. (83) and (84) involve a slight error which is usually of a character such as to make the calculated distortion higher than the actual distortion, and which can if desired, be corrected for as indicated in the first footnote of this section.

$$\text{Power output} = \frac{(E_{max} - E_{min})(I_{max} - I_{min})}{8} \tag{83e}$$

The notation is indicated in Fig. 82*d*.

When third- and fourth-harmonic components can be expected in the output, as will be the case with pentodes and with badly overloaded triodes, one can determine the components by obtaining from the dynamic characteristic the instantaneous plate current when the applied voltage wave is at zero, maximum, and minimum and at 0.707 of the maximum and minimum values. The formulas then are

$$\left.\begin{matrix}\text{Direct-current}\\ \text{component}\end{matrix}\right\} = A_0 = \frac{\frac{1}{2}(I_{max} + I_{min}) + I_2 + I_3 - 3I_b}{4} \tag{84a}$$

$$\text{Fundamental} = A_1 = \frac{\sqrt{2}(I_2 - I_3) + I_{max} - I_{min}}{4} \tag{84b}$$

$$\text{Second harmonic} = A_2 = \frac{I_{max} + I_{min} - 2I_b}{4} \tag{84c}$$

$$\text{Third harmonic} = A_3 = \frac{I_{max} - I_{min} - 2A_1}{2} \tag{84d}$$

$$\text{Fourth harmonic} = A_4 = \frac{2A_0 - I_2 - I_3}{2} \tag{84e}$$

$$\text{Power output} = \frac{A_1{}^2 R_L}{2} \tag{84f}$$

The notation is indicated in Fig. 82*d*.

In using Eqs. (83) and (84) one needs only the plate current at certain critical parts of the cycle and does not require the complete dynamic characteristic. These critical points can be obtained most readily by the construction shown in Fig. 85, rather than by plotting out the entire characteristic as in Fig. 82*d*. Here a straight line, called the *load line*, is drawn through the operating point (E_b, I_b) with a slope such that it intersects the zero plate-current axis at a plate voltage $E_b + I_bR_L$, where R_L is the load resistance, as shown in Fig. 85. This load line is a dynamic characteristic of the type shown in Fig. 82, but plotted on an E_p-I_p coordinate system, as can be seen by comparing individual points. The load line of Fig. 85 can be thought of as showing how the voltage at the plate of the tube varies as variations in grid bias cause plate-current variations. Since the change in plate voltage is proportional to the change in plate current for a resistance load, the load line is straight. This is in contrast with the curvature obtained when the dynamic characteristic is plotted on the grid-voltage–plate-current coordinate system of Fig. 82*d*, which arises because the plate current is not exactly proportional to the grid voltage. The load line has a negative slope because the potential at the plate decreases as the plate current increases. The load line also intersects the zero plate-current axis as indicated above, because when the plate current is zero the change in plate current is $-I_b$, and this flowing through the load resistance R_L produces a drop $-I_bR_L$ that when subtracted from the operating voltage E_b makes the plate potential $E_b + I_bR_L$.

The use of load lines in determining power output and distortion for a particular load resistance is illustrated by the following example.

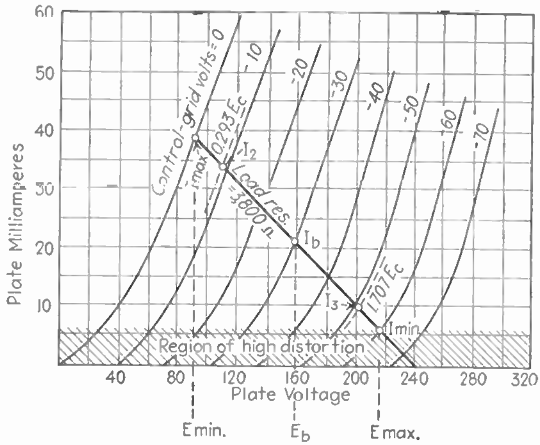


FIG. 85.—Load line drawn on plate-voltage–plate-current curves corresponding to dynamic characteristic of Fig. 82, showing critical points used in Eqs. (84) and (85) to determine distortion.

Example.—Obtain data for drawing the load line of Fig. 85 for the operating point shown and $R_L = 3800$ ohms, and from the load line calculate the power output, plate efficiency, and the distortion when the signal voltage is 30 volts crest (*i.e.*, when the signal has the maximum value that will not drive the grid positive).

The load line is obtained by drawing a straight line through the operating point $E_b = 160$ volts, $I_b = 21.5$ ma, and intersecting the zero plate-current line at a plate potential of $160 + 3800 \times 0.0215 = 242$ volts. This is the line shown in Fig. 85.

The distortion can be calculated from Eq. (83) because the tube is a triode not badly overloaded. Examination of Fig. 85 shows that when the instantaneous grid potential is zero (positive crest of signal) $I_{\max} = 38.5$ ma, and $E_{\min} = 95$ volts. Similarly, when the instantaneous grid potential is -60 volts (corresponding to the negative peak of the signal), $E_{\max} = 217$ volts, and $I_{\min} = 6.5$ ma. Substituting these values in Eq. (83) gives

$$\text{Power output} = \frac{(217 - 95)(0.0385 - 0.0065)}{8} = 0.48 \text{ watt}$$

$$\frac{\text{Second harmonic}}{\text{Fundamental}} = \frac{38.5 + 6.5 - 2 \times 21.5}{2(38.5 - 6.5)} = 0.031$$

$$\text{Direct-current component} = \frac{38.5 + 6.5 - 2 \times 21.5}{4} = 0.5 \text{ ma}$$

$$\text{Plate efficiency} = \frac{0.48}{160 \times (0.0215 + 0.0005)} \times 100 = 13.6 \text{ per cent}$$

In practical operation, if this distortion of 3.1 per cent is less than can be allowed, more "undistorted" power output can be obtained by using a lower load resistance. In such circumstances load lines are drawn for various resistances until the most favorable conditions are found.

Equations (83) and (84) are limited to the case of sine-wave signals and resistance loads. The behavior for this important case is particularly significant, since it indicates what can be expected under all conditions. However, when a more comprehensive picture of the amplifier behavior is desired, as for example the behavior when there is a complex signal wave, the analysis can be carried out by the power-series method discussed in Sec. 51.

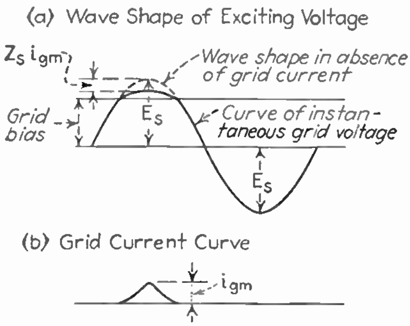


FIG. 86.—Wave shape of exciting voltage when grid is driven positive, showing flattening of the positive peaks resulting from grid current.

Distortion of the Signal Voltage Resulting from Grid Current.—When an amplifier is operated so that the grid goes positive at the positive peaks of the exciting voltage, the resulting grid current causes a drop in voltage that flattens off the positive peaks of the exciting voltage. The result is hence a distortion in the wave shape of the exciting voltage, as shown in Fig. 86. If i_{gm} represents the instantaneous grid current flowing at the peak of the cycle and Z_s represents the impedance that the grid of the tube sees when looking back toward the source of exciting voltage (*i.e.*, Z_s is the source impedance), then the actual signal voltage applied to the grid at the positive peak is less than it should be by an amount $Z_s i_{gm}$. It is therefore apparent that, in order to keep the distortion small, the voltage drop $Z_s i_{gm}$ must be made small compared with the crest exciting voltage E_s, by making the source impedance Z_s low or by not driving the grid far enough positive to draw much grid current, or both. Assuming that grid current flows for less than one-fourth the cycle, it can be found from Eq. (84) that the distortion of the driving voltage produced by the flattening of the positive peaks is

$$\begin{aligned}
\frac{\text{Second harmonic}}{\text{Fundamental}} &= \frac{Z_s i_{gm}}{4E_s - Z_s i_{gm}} \sim \frac{Z_s i_{gm}}{4E_s} \\
\frac{\text{Third harmonic}}{\text{Fundamental}} &= \frac{Z_s i_{gm}}{4E_s - Z_s i_{gm}} \sim \frac{Z_s i_{gm}}{4E_s} \\
\frac{\text{Fourth harmonic}}{\text{Fundamental}} &= \frac{Z_s i_{gm}}{8E_s - 2Z_s i_{gm}} \sim \frac{Z_s i_{gm}}{8E_s}
\end{aligned} \tag{85}$$

In these equations E_s is the crest exciting voltage that would be obtained if there were no flattening. The phase of the second harmonic is such as to tend to cancel the second harmonic generated by the curvature of the dynamic characteristic. An idea of the amount of flattening permissible can be gained by noting that, if the voltage drop produced by the grid current in flowing through the source impedance is 20 per cent of the crest signal voltage, the second- and third-harmonic components are each 5 per cent and the fourth is 2.5 per cent.

Tubes in which the grid is driven positive are preferably operated from transformer-coupled amplifiers having a low voltage step-up, or even a step-down. In this way the impedance Z_s of the source, as viewed by the grid of the power tube, will be low. The low transformation ratio reduces the voltage gain of the driving amplifier, and this together with the distortion of the exciting voltage is the price paid for the increased output obtained with positive-grid operation. Resistance coupling to a grid which is to be driven positive is not desirable (unless the grid leak is replaced by a choke as in Fig. 66*b*), because the grid current in flowing through the grid leak produces a bias on the grid of the power tube.

The proper use of negative feedback (see Sec. 48) is of considerable assistance in reducing the distortion of the driving voltage resulting from grid current.

Tubes for Class A Power Amplifiers.—The type of tube best suited for use as a Class A power amplifier differs from the tube preferred for voltage amplification. In particular, since the "undistorted" power output obtainable is commonly from 15 to 40 per cent of the direct-current power supplied to the anode electrode, it is apparent that the allowable anode dissipation must be proportional to the amount of output power desired. Also, in order to obtain the direct-current input power required to develop a reasonable output, it is necessary that the plate-supply voltage, or the rated d-c plate current, or both, be high.

In small power tubes, such as those used in radio receivers, the plate-supply voltage is limited by practical considerations of safety to about 250 to 300 volts. It is then necessary to employ tubes having a proportionately large plate current in order to obtain the necessary plate input power. In order that this large plate current may flow with a low plate voltage, it is necessary in the case of triode tubes to have a low amplification factor, commonly 2.5 to 5. In tubes designed for higher voltage operation, the plate current can be smaller in proportion to plate voltage, and amplification factors such as 8 to 20 are permissible.

The characteristics of typical Class A power-amplifier tubes of the sizes used in radio receivers and small public-address systems are tabulated in Table IX. The table includes triodes, double-grid tubes, triple-grid tubes, and pentodes. The double- and triple-grid tubes can be connected as Class A triode amplifiers by using the inner grid as the control electrode and connecting the remaining grid or grids to the plate. Pentode tubes are considered in the following section. Data on large

TABLE IX.—CHARACTERISTICS OF REPRESENTATIVE SMALL POWER TUBES

Type	Description	Cathode data			Use*	Typical operating conditions									
		E_f, volts	I_f, amp.	Type		E_p, volts	E_{sg}, volts	E_{cg},† volts	I_p, ma	I_{sg}, ma	R_p, ohms	μ	G_m, μ mho	Load resistance, ohms	Power output, watts
31	Triode	2	0.13	Filament	Class A	180	...	−30	12.3	...	3,600	3.8	1,050	5,700	0.375
45	Triode	2.5	1.5	Filament	Class A	250	...	−50	34	...	1,610	3.5	2,175	3,900	1.60
					Class AB, fixed bias	275	...	−68	28	...	....	.....	.....	3,200	18
					Class AB, self-bias	275	...	775Ω	72	...	....	.....	.....	5,060	12
2A3	Triode	2.5	2.5	Filament	Class A	250	...	−45	60	...	800	4.2	5,250	2,500	3.5
					Class AB, fixed bias	300	...	−62	80	...	....	.....	.....	3,000	15
					Class AB, self-bias	300	...	780Ω	80	...	....	.....	.....	5,000	10
19	Twin triode	2	0.26	Filament	Class B	135	...	0	10	...	....	.....	.....	10,000	2.1
53	Twin triode	2.5	2	Heater	Class B	300	...	0	35	...	....	35	.....	10,000	10
46	Dual grid	2.5	1.75	Filament	Class B	400	...	0	12	...	....	.....	.....	5,800	20
					Class A, triode	250	...	−33	22	...	2,380	5.6	2,350	6,400	1.25
49	Dual grid	2	0.120	Filament	Class A, triode	135	...	−20	6	...	4,175	4.7	1,125	11,000	0.170
					Class B	180	...	0	4	...	....	.....	.....	12,000	3.5
59	Triple grid	2.5	2	Heater	Class A, triode	250	...	−28	26	...	2,300	6	2,600	5,000	1.25
					Class A, pentode	250	250	−18	35	9	40,000	100	2,500	6,000	3
					Class B, triode	400	...	0	26	...	....	.....	.....	6,000	20
89	Triple grid	6.3	0.4	Heater	Class A, triode	250	...	−31	32	...	2,600	4.7	1,800	5,500	0.90
					Class A, pentode	250	250	−25	32	5.5	70,000	125	1,800	6,750	3.4
					Class B, triode	180	...	0	6	...	....	.....	.....	9,400	3.5

Type	Description	Cathode data			Use*	Typical operating conditions									
		E_f, volts	I_f, amp.	Type		E_p, volts	E_{sg}, volts	E_{cg},† volts	I_p, ma	I_{sg}, ma	R_p, ohms	μ	G_m, μ mho	Load resistance, ohms	Power output, watts
33	Pentode	2.0	0.26	Filament	Class A	180	180	−18	22	5	55,000	90	1,700	6,000	1.4
2A5	Pentode	2.5	1.75	Heater	Class A, triode	250	...	−20	31	...	2,700	6.2	2,300	3,000	0.65
					Class A, pentode	250	250	−16.5	34	6.5	80,000	190	2,350	7,000	3
					Class AB, triode fixed bias	350	...	−38	45	...	....	....	.....	6,000	18
					Class AB, triode self-bias	350	...	730Ω	50	...	....	.....	.....	10,000	15
6F6	Pentode	6.3	0.7	Heater	Class A, triode	250	...	−20	31	...	2,600	7	2,700	4,000	0.85
					Class A, pentode	315	315	−22	42	8	75,000	200	2,650	7,000	5
					Class AB, pentode fixed bias	375	250	−26	34	5	....	.....	.....	10,000	19
					Class AB, pentode self-bias	375	250	340Ω	54	8	....	.....	.....	10,000	19
6L6	Beam tube	6.3	0.9	Heater	Class A	375	250	−17.5	57	2.5	22,500	135	6,000	4,000	11.5
					Class AB_1 (no I_g)	400	300	−25	102	6	....	.....	.....	6,600	34
					Class AB_2 (with I_g)	400	300	−25	102	6	....	.....	.....	3,800	60

NOTE: Triode connection for pentode tubes with internally connected suppressor consists in connecting screen and plate together. The resulting characteristics are essentially those of a triode in spite of the suppressor grid.

* Where "Use" is Class AB or Class B, the output given is for two tubes and the load resistance given is the proper value from plate to plate, while the plate current is the zero-signal value for the two tubes.

† Where self-bias is used, the bias resistance is given.

TABLE X.—CHARACTERISTICS OF REPRESENTATIVE TRANSMITTING POWER TUBES
Triode Tubes

Type	Filament			Method of cooling	μ	Allowable plate loss, watts	Typical operating conditions as audio amplifier										Typical operating conditions as Class C amplifier							
	E_f, volts	I_f, amp.	Type				Use*	Plate voltage	Control-grid voltage	Peak exciting voltage per tube	Plate* current, ma	Mutual conductance, μ mhos	Plate resistance, ohms	Load resistance, ohms	Undistorted output, watts*	Plate efficiency, per cent	Plate voltage	Control-grid voltage	Peak signal voltage	Plate current, ma	Grid current, ma	Driving power, watts	Output, watts	Plate efficiency, per cent
204A	11	3.85	Thoriated	Air	23	250	Class B	3,000	−100	250	80	.....	.....	20,000	700	62.7	2,500	−200	440	250	30	15	450	72
207	22	52	Tungsten	Water	20	7,500	Class B	12,500	−575	1,150	400	.....	.....	10,000	22,500	64.4	12,000	−1,600	2,650	1,670	90	235	15,000	75
211	10	3.25	Thoriated	Air	12	100	Class A Class B	1,250 1,250	−80 −100	75 205	60 20	3,300	3,600	9,200 9,000	19.7 260	26.3 65	1,250	−225	375	150	18	7	130	69.3
800	7.5	3.25	Thoriated	Air	15	35	Class B	1,250	−70	150	30	.....	.....	21,000	106	65	1,000	−135	260	70	15	3	50	71.5
801	7.5	1.25	Thoriated	Air	8	20	Class A Class B	600 600	−55 −75	50 160	30 8	1,840	4,300	7,800 10,000	3.8 45	21.2 57.7	600	−150	260	65	15	4	25	64
848	22	52	Tungsten	Water	8	7,500	Class A Class B	8,000 12,500	−730 −1,560	800 2,080	900 400	....	....	5,200 10,000	2,000 22,000	27.8 62.8	10,000	−2,000	2,900	1,450	100	310	10,000	69
851	11	15.5	Thoriated	Air	20.5	600	Class A Class B	2,500 3,000	−92 −135	87 245	240 110	....	1,600	5,000 5,600	160 2,400	26.7 66.6	2,500	−250	450	900	100	45	1,700	75.5
852	10	3.25	Thoriated	Air	12	100	Class B	3,000	−250	390	14	.....	.....	41,000	320	66.5	3,000	−600	850	85	15	12	165	64.8
862	33	207	Tungsten	Water and forced air	48	100,000	Class B	12,000	0	1,000	3,000	.....	.....	1,800	90,000	57.5	18,000	−1,000	2,550	8,333	900	2,400	100,000	66.7
HK 354	5	10	Thoriated	Air	14	150	Class B	3,000	−205	315	65	.....	.....	22,000	665	71	3,500	−625	910	255	50	43	740	83
Eimac 35T	5	4	Thoriated	Air	30	35	Class B	1,250	−30	.....	.....	.....	.....	9,600	200	....	1,500	−120	.....	100	30	...	120	80

* Where "Use" is Class B, power output is for the two tubes in push-pull, load resistance from plate to plate, and plate current is zero-signal current for two tubes.

Screen-grid and Pentode Tubes

Type	Description	Filament			Allowable plate loss, watts	Allowable screen loss, watts	Connection	Typical operating conditions as Class C amplifier										
		E_f, volts	I_f, amp.	Type				Plate voltage	Control-grid voltage	Screen-grid voltage	Suppressor voltage	Plate, current ma	Screen-grid current, ma	Control-grid current, ma	Peak signal voltage	Driving power, watts	Output, watts	Plate efficiency
802	Pentode	6.3	0.9	Heater	10	6	Pentode Tetrode	500 500	−100 −60	200 100	0 ..	45 45	22 15	6 6	155 90	0.9 0.5	14 12	62.1 53.3
803	Pentode	10	5	Thoriated	125	30	Pentode	2,000	−90	500	40	160	45	12	175	2	210	65.6
804	Pentode	7.5	3	Thoriated	40	15	Pentode Tetrode	1,250 1,250	−100 −100	300 180	45 ..	92 92	27 23	7 8	150 160	0.9 1.2	80 80	69.5 69.5
860	Screen grid	10	3.25	Thoriated	100	10	....	3,000	−150	300	..	85	..	15	...	7	165	64.7
861	Screen grid	11	10	Thoriated	400	35	....	3,500	−250	500	..	300	..	40	725	30	700	66.7
865	Screen grid	7.5	2	Thoriated	15	3	....	750	−80	125	..	40	..	5.5	...	1	16	53.4

triode tubes such as are sometimes used as Class A amplifiers in radio transmitters are given in Table X, page 158.

53. Class A Power Amplifiers Using Pentode and Beam Tubes.—Class A power amplifiers using pentode and beam tubes require special consideration because the plate current of such tubes is substantially independent of plate voltage except at low plate voltages. As a consequence the dynamic characteristic is as shown in Fig. 87*a* and follows the characteristic curves of the tube irrespective of the load resistance except when the plate voltage becomes low. As compared with the corresponding characteristic of a triode shown in Fig. 82*d*, the dynamic characteristic of Fig. 87*a* has much greater curvature and develops an inflection point when the plate potential becomes so low that a virtual cathode forms in the tube.

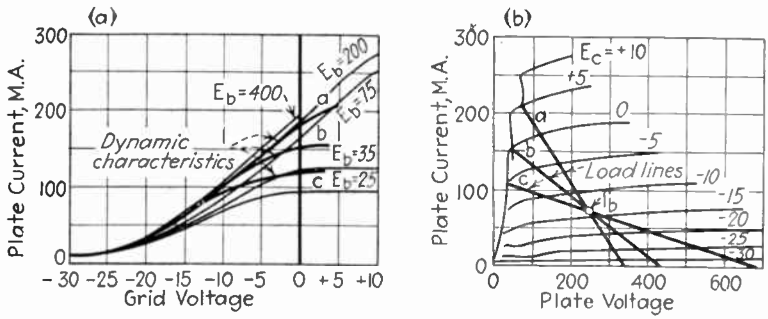


FIG. 87.—Dynamic characteristic of beam tube and resulting output waves for several values of load resistance, together with corresponding load lines. Note that the dynamic characteristics follow the characteristic curves of the tube except when the plate voltage becomes so low that a virtual cathode forms in the tube.

The voltages for the screen and control-grid electrodes of pentode and beam power amplifiers should be so chosen that a signal, which at the negative peak makes the instantaneous grid potential approach but not quite reach cut-off, will at the positive peaks bring the instantaneous grid potential to zero or the desired amount positive, as the case may be. At the same time the voltages must be such that the plate current at the operating point approaches the rated value.

The load resistance used in a pentode or beam power tube should be such that the minimum instantaneous plate potential reached at the positive crest of the exciting voltage is as low as possible without a virtual cathode forming. If this value of instantaneous potential is designated as E_{min}, then the crest alternating-current voltage developed across the load will be $(E_b - E_{min})$, and this must be produced by an alternating current having a crest amplitude $(I_{max} - I_b)$, where I_b is the plate current at the operating point and I_{max} is the instantaneous plate current at the positive crest of the cycle when the instantaneous plate potential is E_{min}. Hence

$$\left.\begin{array}{l}\text{Proper load resistance for pentode}\\ \text{or beam power amplifier}\end{array}\right\} = \frac{E_b - E_{min}}{I_{max} - I_b} \tag{86}$$

If the load resistance is too high, the peaks of the output wave are flattened, whereas if the load resistance is less, there is considerable loss in output.

Exact determination of the load impedance that gives the maximum output for a given allowable distortion can be carried out exactly as in the case of triodes. The only difference is that, since there is normally third-harmonic distortion in the output of pentode and beam tubes, Eq. (84) must always be used.

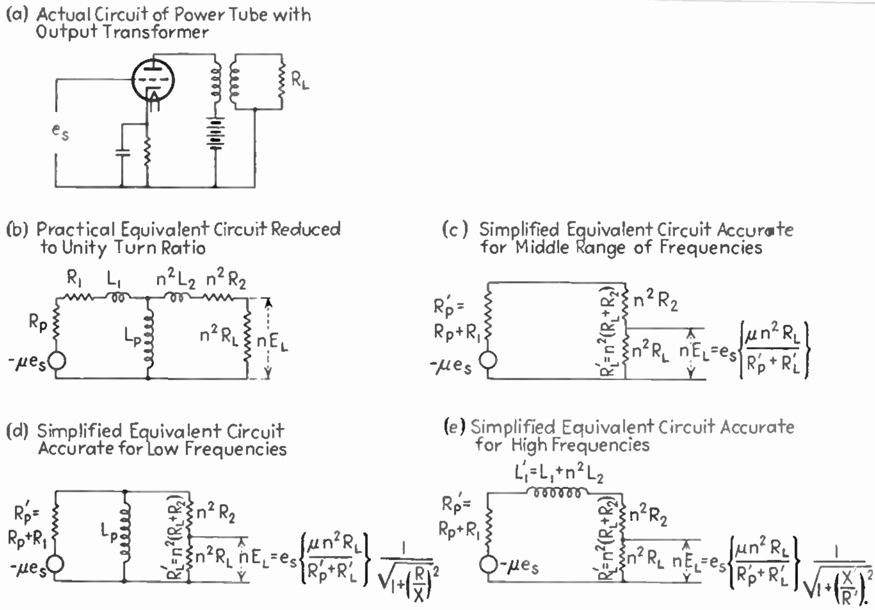
Compared with triode power amplifiers, pentode and beam tubes have the advantage that, with plate potentials below 500 volts, the plate efficiency, and hence the output power, are greater. Pentode and beam tubes also have the advantage of requiring a considerably smaller signal voltage to develop full output. The chief disadvantage of such tubes is that they have high distortion. This is in fact so bad that it is necessary to employ negative feedback if even reasonably low distortion is required. Beam tubes are superior to pentodes with regard to distortion and accordingly are rapidly displacing pentode tubes as power amplifiers. Characteristics of typical pentode and beam power tubes are given in Table IX.

54. Output Transformers for Class A Amplifiers.—The load impedance is usually coupled to the plate circuit of a Class A power amplifier by means of a transformer as shown in Figs. 82*b* and 88*a*. This arrangement avoids passing the d-c plate current through the load impedance and also makes it possible by the use of the proper turn ratio to make any load present the desired impedance to the tube.

The use of an output transformer causes the output of the amplifier to fall off at high and low frequencies in the manner shown in Fig. 89. The falling off at low frequencies is caused by the shunting action of the transformer primary inductance, whereas the falling off at high frequencies is caused by the voltage consumed in the leakage inductance.

Analysis of Frequency-response Characteristics.—The equivalent circuit of an output transformer reduced to unity-turn ratio is shown in Fig. 88*b*. In this circuit the tube and distributed coil capacitances have been neglected, as have the core losses, since all of these factors have relatively little effect in a properly constructed output transformer. The essential elements in the equivalent circuit are seen to be the primary inductance, the leakage inductance, the turn ratio, and the direct-current resistances of the windings.

At moderate frequencies the reactance of the primary inductance is so high as to have negligible shunting effect, while the reactance of the leakage inductance is so low as to produce very little voltage drop. The



μ = amplification factor of tube
R_p = plate resistance of tube (in a push-pull amplifier R_p is twice the plate resistance of one tube)
R_1 = d-c resistance of primary winding
R_p' = $R_p + R_1$ = effective plate resistance
R_2 = d-c resistance of secondary winding
R_L = load resistance
$R_L' = n^2(R_L + R_2)$ = effective load resistance reduced to unity turn ratio
$R = R_L'R_p'/(R_L' + R_p')$ = resistance formed by R_p' and R_L' in parallel
$R' = R_L' + R_p'$ = sum of effective load and effective plate resistances
n = step down voltage ratio = ratio of primary to secondary turns
L_p = primary inductance with appropriate d-c saturation
L_1 = leakage inductance of primary winding
L_2 = leakage inductance of secondary winding
$L_1' = L_1 + n^2L_2$ = total leakage inductance reduced to unity turn ratio
$X = \omega L_p$ = reactance of transformer primary inductance
$X' = \omega L_1'$ = reactance of transformer leakage inductance
e_s = input voltage
E_L = output voltage

FIG. 88.—Actual circuit of power amplifier with output transformer, together with equivalent circuits used in determining the frequency response.

equivalent circuit for all practical purposes then reduces to that shown in Fig. 88*c*, and the output voltage becomes

$$\left.\begin{array}{l}\text{Output voltage in middle-}\\ \text{frequency range}\end{array}\right\} = E_L = e_s\left(\frac{\mu n R_L}{R_p' + R_L'}\right) \qquad (87)$$

The notation is shown in Fig. 88.

At low frequencies it is not permissible to neglect the shunting reactance of the primary inductance, so that the equivalent circuit then takes the form shown in Fig. 88*d*. Analysis of the voltage and current relations of this circuit shows that

$$\left.\begin{matrix}\text{Output voltage at}\\ \text{low frequencies}\end{matrix}\right\} = \frac{\left\{\begin{matrix}\text{Output voltage in middle-}\\ \text{frequency range}\end{matrix}\right\}}{\sqrt{1 + (R/X)^2}} \tag{88}$$

Examination of this equation shows that the falling off in response at low frequencies is determined only by the ratio of the reactance X of the primary inductance to the resistance R formed by the effective load and the effective plate resistances in parallel. *The* 70.7 *per cent response*

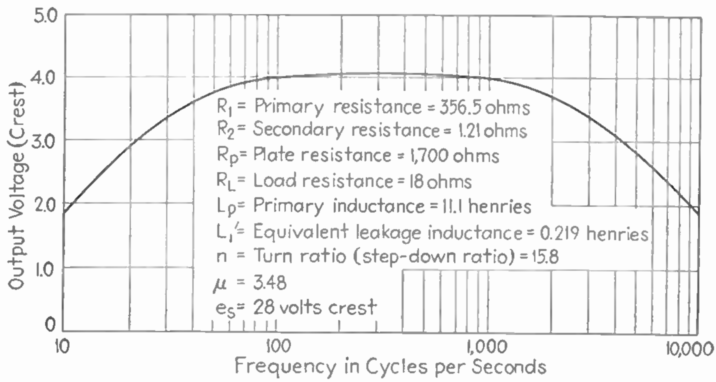


FIG. 89.—Way in which the amplification varies with frequency in a typical power amplifier with output transformer.

point occurs when the primary reactance equals this resistance, and the response at other frequencies can be obtained from Fig. 90.

At high frequencies the shunting effect of the primary inductance is negligible, but it is necessary to take into account the voltage drop in the leakage inductance. This leads to the equivalent circuit of Fig. 88*e*, which gives

$$\left.\begin{matrix}\text{Output voltage at}\\ \text{high frequencies}\end{matrix}\right\} = \frac{\left\{\begin{matrix}\text{Output voltage in middle-}\\ \text{frequency range}\end{matrix}\right\}}{\sqrt{1 + (X'/R')^2}} \tag{89}$$

An examination of this equation shows that the falling off in response at high frequencies depends upon the ratio of the reactance of the leakage inductance to the sum of the effective load and effective plate resistances. *The* 70.7 *per cent response point occurs when the frequency is such that the leakage reactance equals the sum of these resistances.* The response at other high frequencies can be obtained from Fig. 90.

The curve of Fig. 90 can be thought of as a universal amplification curve, since it gives the relative high- and low-frequency response for any output transformer. The practical use of this universal amplification curve of the output transformer can be understood by the following example.

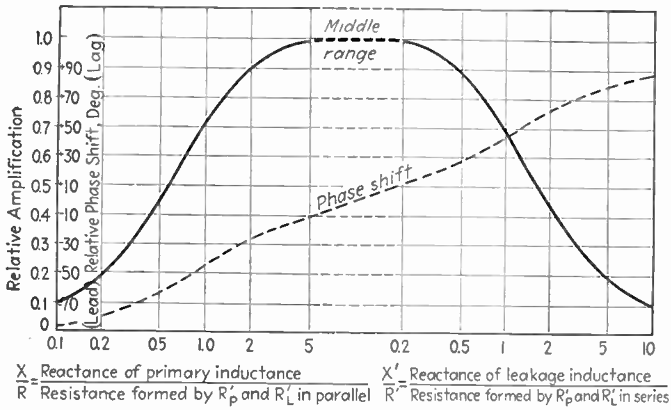


FIG. 90.—Universal amplification curve showing the way in which the amplification of an output transformer falls off at high and low frequencies.

Example.—Calculate the frequency-response curve of the output transformer of Fig. 89.

In the middle range of frequencies, Eq. (87) shows that the load voltage E_L is

$$E_L = 28\frac{3.48 \times 15.8 \times 18}{1700 + 356.5 + 15.8^2(18 + 1.21)} = 4.1 \text{ volts}$$

At low frequencies the output according to Eq. (88) falls to 70.7 per cent of this value when

$$2\pi f \times 11.1 = \frac{15.8^2(18 + 1.21) \times (1700 + 356.5)}{15.8^2(18 + 1.21) + (1700 + 356.5)} = 1435$$

from which $f = 20.6$ cycles. Amplification at other frequencies is obtained from Fig. 90, which shows that at 0.5, 1.0, 2, and 5, times the 70.7 per cent frequency of 20.6 cycles, the output voltage is 0.45, 0.707, 0.89, and 0.98 respectively times the mid-range value of 4.1 volts. Hence the output at 10.3, 20.6, 41.2, and 103 cycles is 1.85, 2.90, 3.65, and 4.02 volts respectively.

The high-frequency response is determined by Eq. (89), which shows that the output drops to 70.7 per cent of the mid-range value when $2\pi f \times 0.219 = 4800 + 2056$, or $f = 4{,}980$ cycles. Other high-frequency points are obtained from Fig. 90, which shows that at 0.2, 0.5, 1.0, and 2, times this frequency, the relative output voltage is 0.98, 0.89, 0.707, and 0.45 respectively. The actual output at 996, 2490, 4980, and 9960 cycles is therefore 4.02, 3.65, 2.90, and 1.85 volts.

These results are shown as a frequency-response curve in Fig. 89.

Examination of Eqs. (88) and (89), and also of Fig. 90, shows that for the best frequency response, the leakage inductance should be low and the primary inductance high. It is also seen that the use of a high load impedance improves the response at high frequencies, but makes the response at low frequencies worse.

Transformer Characteristics.—The turn ratio of the transformer should be such that the load impedance that the plate of the tube sees when looking toward the transformer primary is the value for proper amplifier operation. If n is the ratio of primary to secondary turns, a load impedance Z_L connected across the secondary appears to have a value of n^2Z_L when viewed from the primary of the transformer. This fact determines the proper value of n to match a given load to a particular tube requirement.

The primary inductance that is effective in an output transformer is the incremental inductance with the appropriate direct-current saturation, and is commonly measured at about 500 cycles. The transformer leakage inductance can be determined by measuring the inductance across the primary terminals when the secondary is short-circuited. The turn ratio is obtained by measuring the voltage ratio at a moderate frequency with the secondary open-circuited.

The power rating of an output transformer is determined by the current-carrying capacity of the winding and the voltage that may be developed across the primary without excessive flux densities in the core. When the flux density in the core is high, the magnetizing current becomes large and also has a wave shape that is seriously distorted. Since the magnetizing current flows through the plate resistance of the tube, a distorted wave of magnetizing current causes the voltage across the transformer primary likewise to be distorted. Distortion from core saturation is greatest at the lowest frequency to be amplified, because the flux density in the core is inversely proportional to frequency.

55. Push-pull Class A and Class AB Amplifiers.—In the push-pull amplifier two tubes are arranged as shown in Fig. 91. The grids are excited with equal voltages 180° out of phase, and the outputs of the two tubes are combined by means of an output transformer having a center tap. The advantages of the push-pull connection, assuming identical tubes, are:

1. No direct-current saturation in the core of the output transformer and hence higher incremental inductance with correspondingly better low-frequency response. (The direct currents in the two halves of the primary magnetize the core in opposite directions and so produce zero resultant magnetization.)

2. There is no current of signal frequency flowing through the source of plate power. This means the push-pull power amplifier produces no regeneration even when there is a plate impedance common to the power and other stages.

3. Hum effects produced by alternating-current voltages present in the source of plate power are greatly reduced because the hum currents flowing in the two halves of the primary balance each other out.

4. There is less distortion for the same power output per tube, or more power output per tube for the same distortion, as a result of cancellation of all even harmonics and even-order combination frequencies.

These advantages are so important that a push-pull arrangement using two small tubes is preferable to a single larger tube capable of developing the same total power output.

The reason for the elimination of the even harmonics in the push-pull amplifier can be understood by reference to Fig. 91. Assuming that the

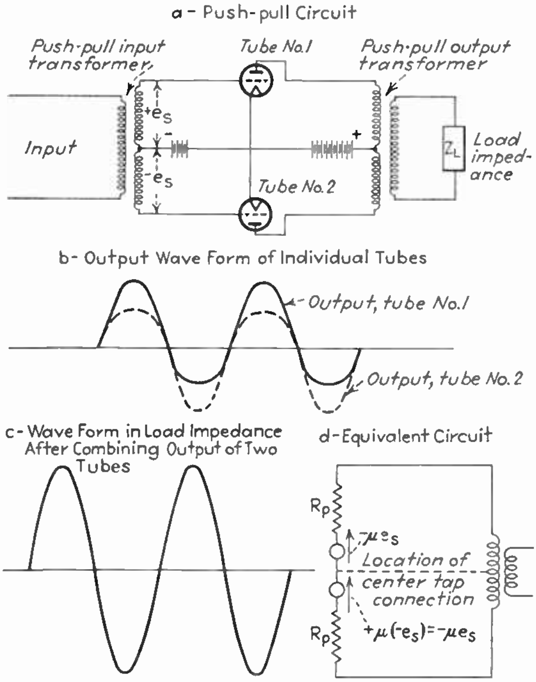


FIG. 91.—Circuit diagram of push-pull amplifier, together with equivalent circuit and wave shapes produced, showing how the push-pull connection makes the positive and negative halves of the output wave have the same shape even though this is not true of the outputs of the individual tubes. The result is that the output wave contains no even harmonics and hence suffers less distortion than do the outputs of the individual tubes.

amplifier is sufficiently overloaded so that some distortion occurs, the individual tubes develop output waves as shown in Fig. 91*b*. The sum of these waves, which represents the amplified output, is shown in Fig. 91*c*. It is seen that when the two tubes have identical characteristics, the positive and negative half cycles of the sum differ only in sign but not in shape. The output hence contains only odd harmonics. This similarity in the shapes of the positive and negative half cycles in the combined outputs results from the fact that at points one-half cycle apart the tubes have merely interchanged functions, *i.e.*, at the later time tube 2 is operating under exactly the same conditions as was tube 1 a half cycle earlier, and vice versa. A more complete analysis based upon the power-series method shows that not only are the even-order harmonics eliminated but also all even-order combination frequencies, particularly the sum and difference frequencies.

Class AB Push-pull Amplifiers.—The elimination of the even harmonics by the push-pull arrangement makes it possible to extend the operating range into the shaded region of Fig. 82*d*, which represents high distortion with single-tube operation. This can be carried to the point where the plate current is actually cut off for a small portion of each cycle without causing excessive distortion. This increases the power output obtainable, particularly when the grid is driven positive, and results in plate efficiencies that range from 40 to 50 per cent. A push-pull amplifier operating in this way is termed a Class AB amplifier.

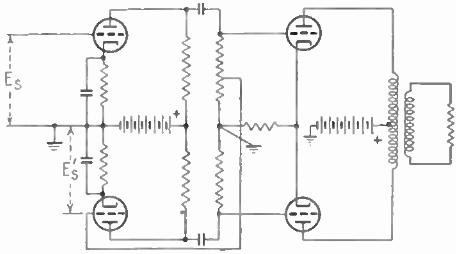


FIG. 92.—Push-pull amplifier excited with the aid of a phase-reversing tube.

The Class AB amplifier is characterized by a large "rectified" plate current when the full signal voltage is applied. This is an advantage, since it permits the direct-current power input to the plate to be increased in the presence of an applied voltage to a value exceeding the allowable direct-current plate power input with no signal. The large increase in plate current caused by the signal in a Class AB amplifier makes it impossible to realize the full possibilities of the tube when self-bias is used, however, since, if the bias keeps the plate current within an allowable value for no applied signal, the bias will be excessive when the "rectified" current flows.

The output transformer for a push-pull arrangement is an ordinary output transformer provided with a center-tapped primary and with the core assembled to give the smallest possible air gap. Examination of the equivalent circuit of Fig. 91*d* shows that the proper turn ratio is such that the plate-to-plate impedance presented by the transformer primary corresponds to a value that is twice the proper load impedance for a single-tube power amplifier. This is because, as far as the output is concerned, the two tubes of the push-pull arrangement are in series.

The excitation for a push-pull amplifier is ordinarily obtained from a transformer having a center-tapped secondary. Resistance coupling to a push-pull amplifier, however, can be employed by making use of an auxiliary phase-reversing tube as shown in Fig. 92. This tube is arranged to excite the second tube of the push-pull combination with a voltage that is equal in magnitude and opposite in phase to the exciting voltage of the first tube.

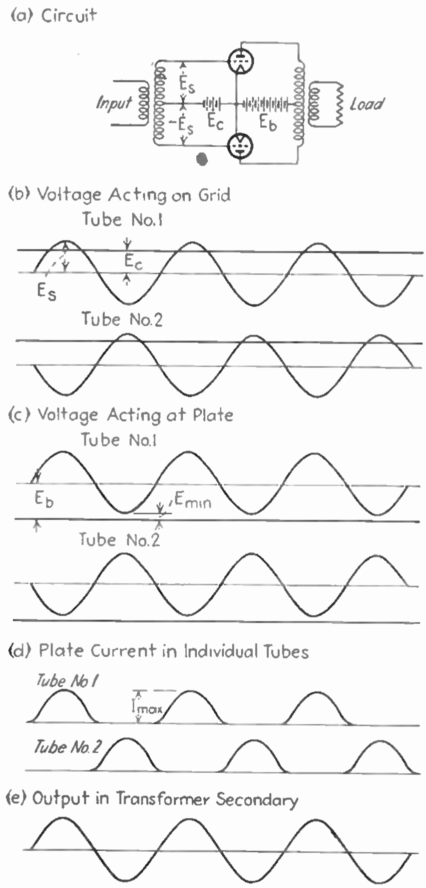


FIG. 93.—Circuit diagram of Class B amplifier together with oscillograms showing how the half sine-wave pulses of plate current in the individual tubes combine to produce the output.

56. Class B Audio-frequency Power Amplifiers.—The Class B audio-frequency amplifier is a push-pull amplifier in which the tubes are biased approximately to cut-off. Operated in this manner, one of the tubes amplifies the positive half cycles of the signal voltage while the other amplifies the negative half cycles, with the output transformer combining these in the load. Such an amplifier is characterized by a high plate efficiency. Also, when the grid is driven positive, as is normally the case, the output is also unusually high in proportion to the size of the tube.

Analysis of Class B Amplification. The fundamental factors involved in the operation of a Class B amplifier can be understood with the aid of oscillograms such as shown in Fig. 93. The potential applied to the grid of the tubes consists of a bias voltage approximating cut-off, upon which the alternating exciting voltage is superimposed, as shown at Fig. 93*b*. This causes pulses of plate current that are roughly half sine waves, as shown at Fig. 93*d*. When combined by the output transformer, these pulses produce a substantially sinusoidal current in the load resistance. The voltage actually at the plate of the tubes consists of the direct-current plate-supply potential minus the voltage drop across half the transformer primary. Since the voltage across the load is substantially sinusoidal, this drop is likewise sinusoidal, and the instantaneous plate potential varies as shown in Fig. 93c.

The peak amplitude of the plate-current pulse is the current that flows when the grid potential is at its positive peak value and the plate potential is at its minimum value E_{min}. In the usual case, when the grid is driven positive, the amplitude of the plate-current pulse is determined primarily by the amount the grid is driven positive. The minimum plate potential E_{min} is controlled mainly by the load resistance. This is because a high load resistance develops more voltage with a given pulse than does a low load resistance and so makes E_{min} smaller. In actual operation the load resistance should be such that E_{min} is small compared with the plate-supply voltage, to insure good plate efficiency, but is at the same time greater than the maximum positive potential reached by the grid in order to prevent the grid current from being excessive.

The quantitative relations existing in a Class B audio amplifier can be determined with an accuracy sufficient for ordinary purposes by assuming that the pulses of plate current are half sine waves. This is equivalent to assuming that the characteristic curves of the tube are substantially straight lines, and that the bias is adjusted to cut-off. Analysis based upon this simplifying assumption shows that[1]

$$\left.\begin{array}{l}\text{Proper load resistance}\\ \text{from plate to plate}\end{array}\right\} = R_L = 4\frac{E_b - E_{min}}{I_{max}} \tag{90}$$

$$\left.\begin{array}{l}\text{Power output from}\\ \text{two tubes}\end{array}\right\} = \frac{I_{max}(E_b - E_{min})}{2} \tag{91}$$

[1] These equations are derived as follows: Since the plate-current pulse of each individual tube flows through only one-half the transformer primary, the combined output of the two tubes is equivalent to an a-c current having a crest value $I_{max}/2$ flowing through the entire transformer primary. If R_L is the equivalent load resistance between primary terminals of the output transformer, the alternating drop produced between plate and cathode of a single tube is one-half the voltage drop of the current $I_{max}/2$ in the resistance R_L or $R_L I_{max}/4$. The minimum instantaneous plate potential is hence

$$E_{min} = E_b - \frac{R_L I_{max}}{4}$$

Solving this for R_L results in Eq. (90). The power output is one-half the square of the load current times the load resistance, or

$$\text{Power output} = \frac{R_L I_{max}^2}{8}$$

Equation (91) results when R_L is eliminated by the use of Eq. (90). The d-c plate current drawn by the individual tube assuming a sine-wave half cycle of current is I_{max}/π, so that the total d-c plate current of the two tubes is $2I_{max}/\pi$ and the power input is $2I_{max}E_b/\pi$. Dividing the output by this input gives the plate efficiency as in Eq. (92).

$$\text{Plate efficiency} = \frac{\pi}{4}\left(1 - \frac{E_{\min}}{E_b}\right) \tag{92}$$

Here I_{max} is the peak plate current of the individual tube, $E_{\min}$ is the minimum instantaneous plate potential reached during the cycle, and E_b is the plate-supply voltage. It will be noted that the maximum possible efficiency is $\pi/4$, or 78.5 per cent, and that the closeness with which the actual efficiency approaches this theoretical maximum is determined by the ratio $E_{\min}/E_b$.

Miscellaneous Considerations.—The non-linear distortion in the output of a Class B amplifier operating with the proper load resistance depends upon the curvature of the tube characteristics and upon the operating point. With actual tubes experiment shows that the distortion is small when the bias corresponds roughly to the bias that would be obtained if the main part of the grid-voltage–plate-current curve were projected to zero current as a straight line. This bias, which is commonly called *projected cut-off*, is somewhat less than actual cut-off. Such an operating point lowers the efficiency slightly because it results in a small d-c plate current in the absence of a signal but gives substantially distortionless amplification if the two tubes have identical characteristics.

The frequency response of a Class B audio amplifier is determined by the characteristics of the output transformer in much the same way as for a Class A power amplifier. There is a falling off at low frequencies as a result of low primary impedance at low frequencies, and a falling off at high frequencies, as a result of leakage inductance. It is hence desirable that the primary inductance be high and the leakage inductance small. The fact that the plate current in each half of the primary winding is not continuous also makes it very important that the leakage inductance from one-half the primary to the other half be as low as possible.

The same tubes employed for Class A and Class C power amplification can also be used for Class B amplifiers. Where only a relatively few watts are required and the plate-supply potential is limited to 300 or 400 volts, as in public-address systems and small radio transmitters, special tubes are often used for Class B amplifiers. These are designed with an amplification factor so high that projected cut-off corresponds approximately to zero grid bias, thereby greatly simplifying the bias problem.

The plate current of a Class B amplifier varies with the signal amplitude, so that for best results the plate-supply system should have good voltage regulation. This is particularly true when the operating point is not zero grid bias, since then variations in the supply voltage seriously alter the effective operating point.

The driving stage for a Class B amplifier must be designed to have a low internal impedance in order to minimize the distortion resulting from the grid current, as discussed in Sec. 52. The usual driving arrangement consists of a triode power amplifier having an output transformer with a low turn ratio, commonly a step-down.

Compared with Class A power amplifiers, the Class B arrangement has the advantage of higher plate efficiency and negligible power loss when no signal voltage is applied. It possesses the disadvantage, however, that the third-harmonic distortion is in general higher than that of a Class A amplifier. Furthermore, in order to insure low distortion, it is absolutely necessary that the desired operating conditions be very closely realized. In particular, the two tubes must be very closely balanced, since a difference of 10 per cent in the plate currents will produce 5 per cent second-harmonic distortion. It is also necessary that the load resistance have the proper value for all frequencies and that the exciting voltage applied to the grids have the proper amplitude. As a consequence, Class B audio amplifiers are extensively used in radio transmitters, where the equipment is operated under the continuous supervision of trained personnel, but they are not favored for radio receivers, small public-address systems, etc.

57. Class C Tuned Amplifiers.—The Class C tuned amplifier differs from an ordinary tuned amplifier in that the bias is made greater than the cut-off value corresponding to the plate-supply voltage. When a signal is applied to such an amplifier the plate current accordingly flows in pulses that last for less than half a cycle. The Class C amplifier is characterized by high plate efficiency and is used to develop radio-frequency power when it is not necessary for the output voltage to be proportional to the exciting voltage. The circuit of a typical Class C amplifier is shown in Fig. 94*a*.

Voltage, Current, and Power Relations.—The fundamental relations existing in a Class C amplifier can be understood from a study of oscillograms showing the voltage, current, and power relations of the tube. Considering first the voltage relations, it is seen from Fig. 94 that the voltage applied to the grid consists of the negative bias voltage E_c plus the alternating signal voltage E_s. Under normal conditions the signal voltage is sufficient to make the instantaneous grid potential positive at the positive crest of each cycle, so that the oscillogram of grid voltage is as shown in Fig. 94*c*.

The voltage applied to the plate of the tube consists of the plate-supply voltage E_b minus the voltage drop E_L developed across the tuned circuit. This drop is sinusoidal and is maximum when the grid potential is most positive. The result is that the instantaneous voltage acting on the plate electrode of the tube has the character shown in Fig. 94*b*.

The grid and plate currents that flow at any instant are the result of the combined action of the voltages acting on the grid and plate electrodes at that instant. The grid and plate current waves can accordingly be determined with the aid of characteristic curves of the tube. Oscillograms of current waves in a typical case are shown in Figs. 94*d* and 94*e*. The plate current flows in pulses that have a duration somewhat less than a half cycle, and the grid current flows in pulses of even shorter duration.

The power relations existing in the plate circuit of a Class C amplifier are shown in Fig. 94*f*. The power input at any instant is the product of

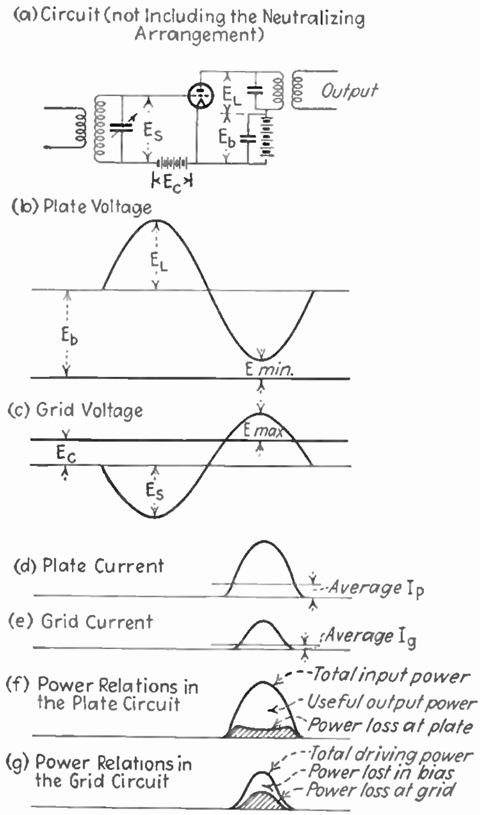


FIG. 94.—Voltage, current, and power relations of typical Class C amplifier.

plate-supply voltage and instantaneous plate current and so varies in the same manner as the plate-current pulse. The power lost at the plate at any instant is the product of the instantaneous plate voltage and plate current and so varies as indicated by the shaded area. The output power delivered to the tuned circuit is the difference between the input power and the loss at the plate.

The power that the signal or exciting voltage applied to the grid of a Class C amplifier must supply at any instant is equal to the product of instantaneous grid current and instantaneous exciting voltage and so varies according to the oscillogram of Fig. 94*g*. Since the grid current flows in a short pulse near the crest of the cycle of exciting voltage, the average driving power is very nearly equal to the product of crest exciting voltage and the d-c grid current. Part of this driving power is dissipated at the grid electrode of the tube, and the remainder is used up in charging the bias battery (or is dissipated in the grid-leak resistance when grid-leak bias is used), as shown in Fig. 94*g*.

The currents observed on direct-current instruments inserted in series with the grid and plate electrodes represent the average values of the pulses of grid and plate current, respectively. The average dissipation of energy at the plate of the tube is also the average of the oscillogram of plate loss.

The high efficiency of a Class C amplifier results from the fact that the plate current is allowed to flow only when the instantaneous potential on the plate is low compared with the plate-supply voltage. In this way, energy is supplied to the plate circuit of the amplifier only when most of the plate-supply voltage is used up as voltage drop across the tuned circuit and hence when most of the energy is delivered to the tuned circuit instead of being wasted at the plate. The efficiency is highest when the minimum plate potential E_{min} is low and when the fraction of the cycle during which plate current is allowed to flow is likewise small. Low minimum plate potential makes for high efficiency because it represents a low voltage drop in the tube. A short plate-current pulse improves the efficiency by restricting the flow of current to the period when the voltage drop across the tube is at or near the smallest value.

Fundamental Factors Involved in Class C Amplifiers.—In the practical adjustment of Class C amplifiers the desired characteristics are high output for a given plate loss, and low driving power. The proper procedure for obtaining the best possible results from a given tube can be understood by considering the fundamental factors that determine the plate current, efficiency, grid-driving power, etc. To begin with, the amplitude of the plate-current pulse is determined primarily by the maximum grid potential E_{max}. This is because the grid when positive is far more effective in drawing electrons from the space change than is the plate.

The minimum plate potential E_{min} is determined by the amplitude and duration of the pulses of plate current and by the impedance of the plate tuned circuit. This is because the minimum plate voltage E_{min} is the difference between the plate-supply voltage and the voltage drop produced by the pulses of plate current flowing through the tuned load circuit. Hence, a high load impedance, or large current pulses, develop a large voltage drop across the tuned circuit and make E_{min} small.

The fraction of the cycle during which the plate current flows is determined primarily by the grid bias, with a large bias corresponding to a short plate-current pulse. If the pulse is short the efficiency is improved but at the expense of increased driving power and reduced output. On the other hand, a long pulse gives large output with small driving power, but at a low efficiency. The usual compromise is a plate-current pulse lasting 120 to 140 electrical degrees, which corresponds to a bias of the order of 3 to 1½ times cut-off.

The plate efficiency is determined primarily by the ratio E_{min}/E_b of minimum plate potential to plate-supply voltage. For high efficiency the minimum plate potential E_{min} should accordingly be as small as possible. At the same time, the instantaneous plate potential should not be less than the maximum grid potential E_{max}, or the grid current, and hence driving power, will be excessive.

The grid driving power is determined by the exciting voltage and grid current. The exciting voltage required to drive the grid a given amount positive is roughly proportional to the bias. At the same time, the bias is determined by the length of the plate-current pulse, being greater as the length of pulse is shortened. The driving power hence increases as the length of the plate-current pulse is reduced and becomes excessive for pulses much shorter than 120 electrical degrees. The grid current is determined by the maximum grid potential E_{max} and the minimum plate voltage E_{min}. The current increases as the grid is driven more positive (*i.e.*, as E_{max} is increased) and as the minimum plate voltage E_{min} is reduced. If the grid becomes more positive than the plate, *i.e.*, if E_{max} is greater than E_{min}, the grid current will be excessive, since secondary electrons produced at the plate will then be attracted to the grid. Also, if E_{min} is made so low that a virtual cathode forms in the tube, the grid will capture electrons returning from the virtual cathode and so will draw an excessive current.

Circuit Arrangements.—The exact circuit details of a Class C amplifier may be arranged in a variety of ways. Thus the plate tuned circuit can be either series feed or shunt feed, as shown in Figs. 95*a* and 95*b*, respectively. Push-pull combinations such as Fig. 95*c* are also common, and because of their symmetrical character are superior to two tubes connected in parallel. With single tubes, neutralization can be accomplished by one of the systems shown in Fig. 74, while in push-pull amplifiers the cross-neutralization system of Fig. 95*c* is used.

The grid bias can be obtained from a fixed source such as a battery or generator, from a grid-leak–grid-condenser combination, or from a self-bias resistor, as shown at *c*, *a*, and *b* respectively, in Fig. 95. In the grid-leak method advantage is taken of the fact that the d-c grid current produces a negative bias when passed through a resistance in series with the grid circuit. The magnitude of the bias obtained is equal to the product of d-c grid current and grid-leak resistance, and for a given value of E_{max} it is controlled by the grid-leak resistance. The grid leak must be by-passed to radio-frequency voltages by a condenser appreciably larger than the input capacitance of the tube and also large enough to act as an effective by-pass for the resistance used. The grid-leak arrangement has the advantage of simplicity and the fact that it tends to be self-adjusting with respect to maximum grid potential E_{max}. Thus small changes in signal voltage, which would produce large changes in E_{max} with a fixed bias, do not do so with the grid leak. This is because any tendency to change E_{max} produces a large effect on the grid current, which tends to change the grid current and hence the grid bias in such a way as to maintain E_{max} nearly constant.

Self-bias arrangements employ a resistor between cathode and ground as shown in Fig. 95*b*. This makes the cathode positive with respect to the grid in the same manner as discussed in connection with Fig. 55.

Self-bias and fixed-bias arrangements have the advantage over grid-leak systems in that the bias is not lost when the exciting voltage is removed. Self-bias has the disadvantage, however, that the plate-supply potential must be greater than the required direct-current plate voltage by an amount equal to the bias.

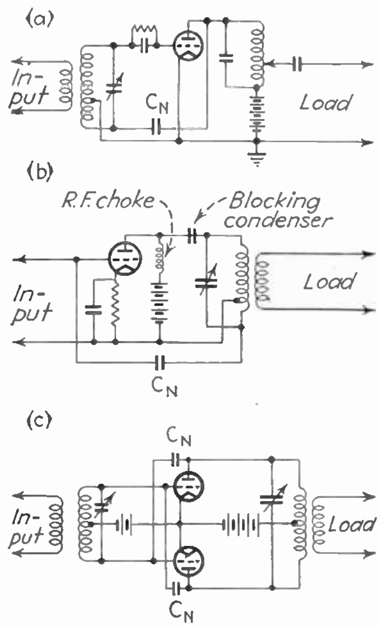


FIG. 95.—Typical circuit arrangements for Class C amplifiers.

Design of the Tank Circuit.—The tuned circuit connected between the cathode and plate of the Class C amplifier, commonly called the *tank circuit*, must supply the proper impedance and must not consume an undue proportion of the power output of the amplifier. The efficiency of the tank circuit is the fraction of the total power delivered to this circuit by the tube that is transferred to the load. This efficiency depends upon the actual loss resistance of the tuned circuit compared with the resistance that is coupled into the tuned circuit by the load. Thus, if the tank circuit in the absence of load has $Q = 100$, but has $Q = 10$ in the presence of the load, the efficiency of the tank circuit is $(100 - 10)/100 = 0.90$.

The proper effective Q of the tank circuit depends upon circumstances, but is commonly of the order of 10 to 30. If lower than 10, harmonic components of the plate-current pulses produce appreciable harmonic voltage drop across the tank circuit. On the other hand, if the Q of the tank circuit is high the discrimination against the harmonics contained in the plate-current pulses is improved, but there is a sacrifice in tank-circuit efficiency. Also, if the effective Q is less than about 10 the tank circuit will be difficult to tune properly, since the adjustment for maximum impedance will not be the adjustment giving a resistance impedance.

With the effective tank-circuit Q determined by the above considerations, the value ωL can be calculated from the voltage across the tank circuit and the power delivered to the circuit, using the relation

$$\omega L = \frac{E^2}{2PQ_{\text{eff}}} \tag{93}$$

where

P = power in watts delivered to tank circuit
ωL = reactance of inductive branch of tank circuit
Q_{eff} = effective Q of tank circuit, taking into account the effect of the coupled load resistance
E = crest alternating voltage across the tank-circuit inductance.

The voltage E across the tank circuit is such that the crest alternating potential developed between plate and cathode is about 80 to 90 per cent of the plate-supply voltage, while in the push-pull circuit of Fig. 96*c* the value of E will be about 1.6 to 1.8 times the plate-supply potential.

When the required ωL has been ascertained, the inductance and capacitance of the tank circuit can be calculated for any given frequency.

After the tank circuit has been given the proper inductance and capacitance, the desired effective Q is obtained by adjusting the coupling of the load. If this coupling is close, the resistance coupled into the tank circuit will be large and the effective Q will be low, and vice versa.

In shunt-feed circuits, as Fig. 95*b*, the radio-frequency choke is effectively in shunt with the tank circuit. The choke insulation must accordingly be able to stand the full voltage of the tank circuit, and the choke inductance should be large compared with the inductance of the tank circuit.

Practical Adjustment of Class C Amplifiers.—In adjusting a Class C amplifier one usually has available the manufacturer's data giving typical operating conditions for the tube used. Such information is included in Table X and commonly includes d-c plate and grid currents, power output, exciting voltage, grid bias, and driving power. The practical problem is to realize these conditions as nearly as possible using the d-c grid and plate currents (and possibly the output power) as guides.

The first step is to lay out the tank circuit, using Eq. (93) as explained above. Next, provision is made for obtaining the proper grid bias by use of a voltmeter in the case of fixed bias or by selecting a grid-leak or self-bias resistance that will develop the required bias with the expected grid or space current, as the case may be. The tank circuit is then brought to resonance at the desired frequency by applying exciting voltage at reduced plate potential and adjusting the tuning condenser until the d-c plate current is minimum.[1] Finally the full plate-supply potential is applied, and the exciting voltage and coupling between the load and tank circuit are adjusted until a combination is found that makes the d-c plate current, d-c grid current, and power output approximate as nearly as possible the expected values. In these final operations, it is desirable to retune the tank-circuit condenser for minimum plate current after each change in load coupling.

When carrying out the final adjustments it is helpful to keep in mind the fundamental factors involved, as discussed on page 173. In particular, it is to be noted that with given excitation conditions a reduction in the coupling between load and tank circuit increases the tank-circuit impedance and hence lowers E_{min}. For any given excitation condition, this load coupling should be as small as possible without causing the grid current to be excessive or the plate current to fall off seriously. If the load coupling arrived at in this way does not give the desired d-c plate current, then the amount the grid is driven positive should be changed by varying the excitation or altering the grid bias, or both. After this change in excitation conditions, the load coupling is readjusted as required.

It is sometimes found that the operating conditions recommended by the tube maker cannot be realized in all respects. This is particularly true of the grid current because of the erratic character of secondary emission at the control grid. When this is the case, the grid-leak resistance should be altered so that the desired grid bias is obtained with the actual grid current. Other variations in tube characteristics can ordinarily be taken care of by slight changes in the exciting voltage.

Class C Amplifiers Employing Screen-grid and Pentode Tubes.—Screen-grid and pentode tubes can be operated as Class C amplifiers by making the grid bias greater than the cut-off value corresponding to the screen-grid potential. The performance obtained is then similar to that of triode Class C amplifiers, but with the advantage that no neutralization

[1] Minimum plate current in a Class C amplifier employing triode tubes indicates maximum voltage drop in the load, and hence maximum load impedance. When the effective Q of the tank circuit exceeds 10, the adjustment for maximum impedance corresponds to resonance.

is required. The tubes are expensive in the larger air-cooled sizes, however, and are not available in water-cooled sizes.

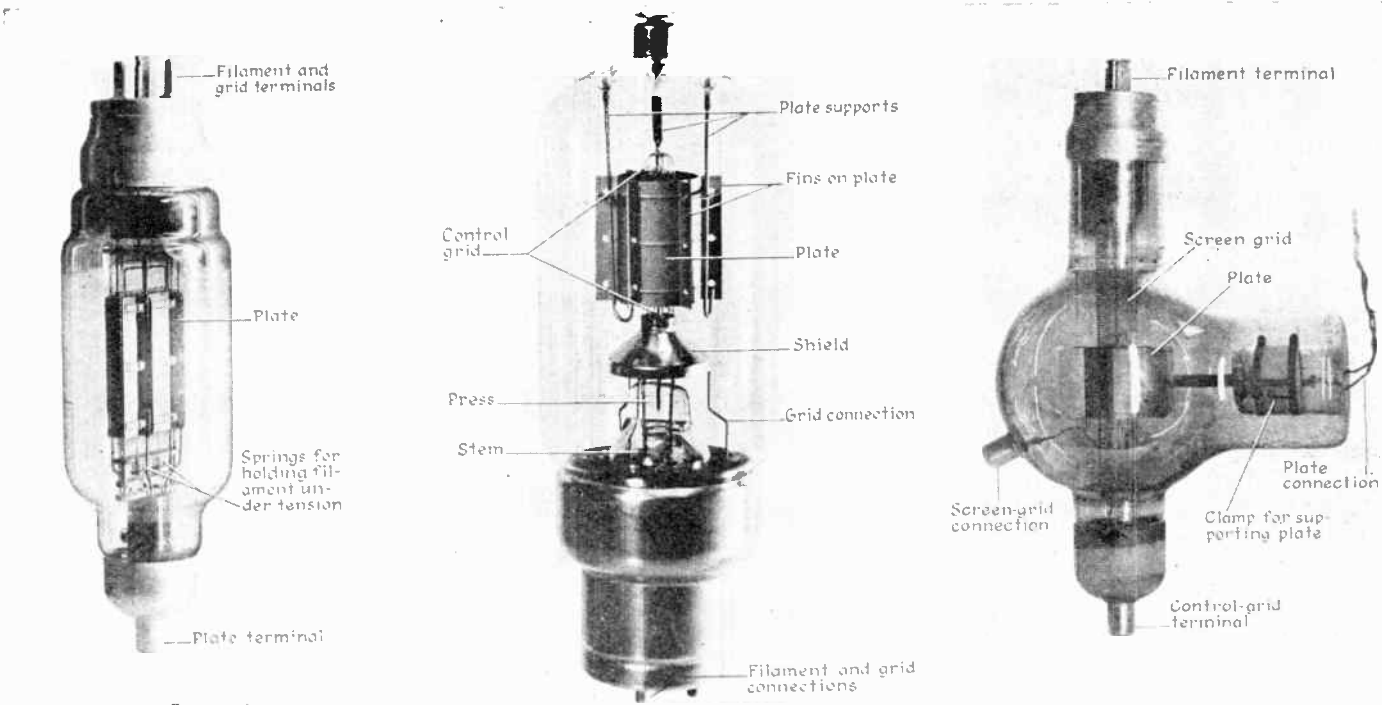
In adjusting screen-grid and pentode Class C amplifiers it is sometimes found that the plate potential has so little effect on the plate current that it is impossible to tune the tank circuit by adjusting for minimum plate current. Under such circumstances the tuning adjustment can be made for maximum current in the load.

58. Characteristics of Tubes Suitable for Use in Class C Amplifiers.—When the amount of power to be generated by a Class C amplifier does not exceed a few watts, it is possible to employ the ordinary small vacuum tubes commonly used in radio receivers. Larger tubes are, however, required for greater powers.

The amount of power that a tube can handle is determined by the plate voltage that may be applied to the tube with safety, by the electron emission of the cathode, and by the amount of power that can be dissipated within the tube without overheating. The large plate currents required in power tubes necessitate the use of correspondingly large cathodes. In order to withstand high anode voltages it is necessary to provide ample insulation between the plate and other electrodes, as illustrated in Figs. 96 and 97. In glass-envelope tubes the plate, grid, and cathode electrodes are commonly brought out through separate parts of the glass envelope (see Fig. 96) in order to provide the maximum possible insulation. The size of the plate electrodes and of the glass envelopes of air-cooled tubes is proportional to the power capacity. This is necessary in order that the heat generated within the tube may be dissipated safely. Even then the plates of air-cooled power tubes normally operate at a dull red heat.

Water-cooled Tubes.—The practical difficulty of radiating large amounts of energy from plates and through glass envelopes of reasonable size without excessive temperature rise, has made it impracticable to construct glass-envelope tubes having an allowable plate loss exceeding about 1500 watts. With ordinary efficiencies this means that the largest practical glass-envelope tubes have a power output rating of the order of a few kilowatts.

When more power is required, tubes having water-cooled anodes are employed. The construction of such a tube is illustrated in Fig. 97. Here the anode consists of a copper tube that is dropped into a jacket through which water is circulated. The copper plate serves as part of the envelope of the tube as well as acting as an anode. Because of the water cooling and the high thermal conductivity of copper, many kilowatts of energy can be dissipated by such an anode without an appreciable rise in temperature.



Type 849 Triode Tube. (Allowable plate dissipation = 400 watts)

Type HK-354 Triode Tube. (Allowable plate dissipation = 150 watts.)

Type 861 Screen-grid Tube. (Allowable plate dissipation = 400 watts.

FIG. 96.—Typical glass-envelope power tubes.

Emitters.—Thoriated-tungsten filaments are normally used in all air-cooled power tubes except the very smallest. This is because thoriated tungsten is an efficient emitter and stands up much better at high anode voltages than oxide-coated filaments. The latter are used only in very small tubes, particularly those operating at plate potentials of 500 volts or less. In water-cooled tubes the operating conditions with respect to positive-ion bombardment are so severe that only tungsten filaments can be used.

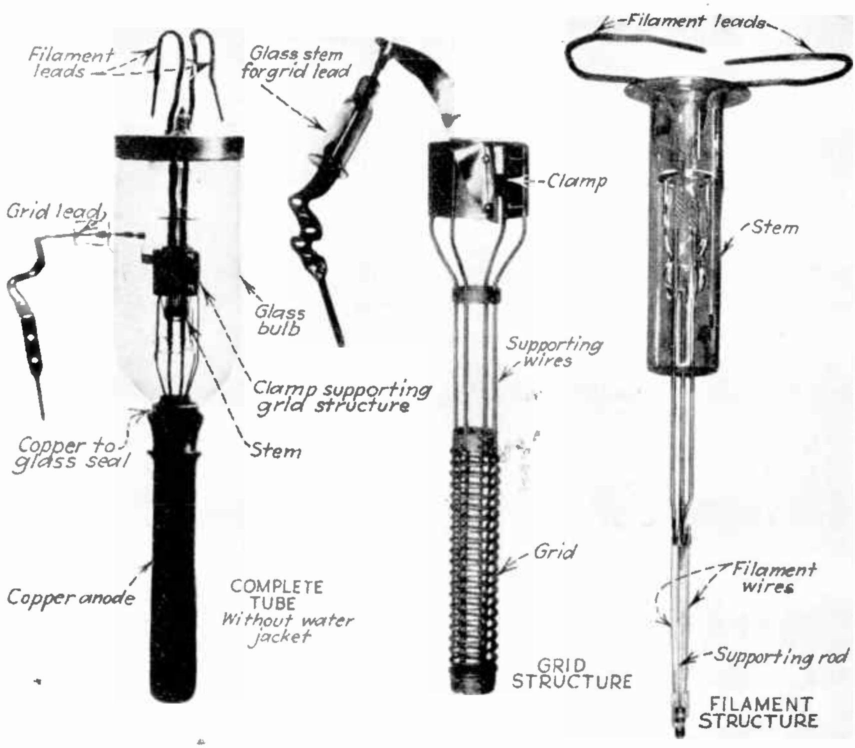


FIG. 97.—A water-cooled tube having a power rating of 20,000 watts output when acting as a Class C amplifier. The anode is made in the form of a copper cylinder which is cooled by immersion in a water jacket.

Construction of Power Tubes.—The plates of air-cooled power tubes are commonly of molybdenum, carbon, or tantalum. The last is the best material but is more expensive than the others. With water-cooled tubes the anodes are always made of copper.

Grid electrodes of air-cooled tubes are usually made of molybdenum, although tungsten and tantalum are sometimes employed. The grids of water-cooled tubes are invariably of tungsten because this is more refractory than any of the other available materials.

The evacuation of large tubes presents a very difficult problem. It is necessary not only to remove the air within the envelope but, what is much more difficult, to remove the gas that is occluded in the metal and glass parts. Otherwise there will be a slow release of gas, and what was originally a good vacuum will ultimately become a very poor one. For satisfactory results it is necessary to follow a systematic procedure of heating the glass and metal parts. The difficulty of obtaining a satisfactory vacuum can be understood when it is realized that water-cooled tubes require continuous pumping for approximately 24 hr. to remove the occluded gas.

Rating of Power Tubes.—Large power tubes are commonly rated on the basis of the allowable dissipation at the plate electrode under typical operating conditions. In addition, the manufacturers always give maximum allowable plate-supply voltage and maximum allowable d-c plate current. Values of d-c plate current, power output, d-c grid current, grid bias, driving power, peak exciting voltage, and efficiency are also usually specified for typical operating conditions. Characteristics of representative power tubes are given in Table X, page 158.

59. Linear Amplifiers.—The linear amplifier is a radio-frequency power amplifier adjusted so that the voltage developed across the load is proportional to the exciting voltage applied to the grid of the tube. Linear amplifiers are used to amplify modulated waves, since such amplifiers will not distort the modulation envelope.

Linear amplification of modulated waves is normally obtained by operating the tube with a fixed bias that approximates cut-off. The plate current then flows in pulses that are substantially half sine waves having an amplitude proportional to the exciting voltage.

Adjustment of Linear Amplifiers.—The linear amplifier is designed in much the same way as a Class C amplifier. The only special features involved are the fact that the grid bias is less than for Class C operation, and that attention must be paid to insuring a linear relation between exciting voltage and output voltage.

The linearity is determined by the operating grid bias, by the conditions existing at the crest of the modulation cycle, and by the regulation of the exciting voltage. Experiment shows that the best results are obtained when the grid bias is slightly less than cut-off. The optimum bias corresponds to "projected cut-off" as discussed in Sec. 56 in connection with Class B audio amplifiers.

The effect on the linearity of the conditions at the crest of the modulation cycle can be understood by considering the way the output voltage varies with exciting voltage. Typical relationships of this sort are shown in Fig. 98 for two load impedances. It is seen that as the exciting voltage is increased the output increases first almost linearly with excitation but ultimately flattens off. This flattening occurs when the alternating voltage developed between cathode and plate has a crest value only slightly less than the plate-supply potential, and is commonly termed *saturation*. The exciting voltage required to produce saturation is less the greater the tank-circuit impedance, since with more impedance it takes smaller plate-current pulses and hence less excitation to develop a given voltage across the impedance. Examination of Fig. 98 shows that the linearity between input and output voltages below saturation tends to be greater the higher the load impedance. At the same time, a high load impedance means less power, because the energy delivered to the tank circuit with a given voltage across the circuit is inversely proportional to the impedance.

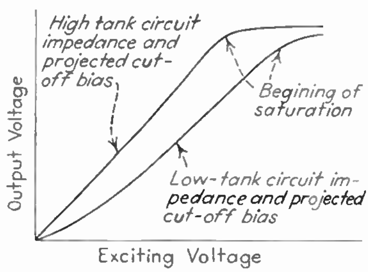


FIG. 98.—Relation of output voltage to exciting voltage in typical linear amplifier for two values of tank-circuit impedance. It will be noted that when the excitation is sufficient to make the alternating plate-cathode voltage have a crest value only slightly less than the direct-current plate voltage, saturation occurs, and further increase in excitation produces very little change in output.

These considerations indicate that the optimum operating conditions for a linear amplifier exist when the tank-circuit impedance is somewhat greater than would be used if power output were the only consideration, and when the exciting voltage is such that at the crest of the modulation cycle saturation conditions are not quite reached. In this way good linearity is obtained with a reasonable output.

Linearity is also influenced by the fact that the equivalent grid-circuit resistance varies with the exciting voltage. This places a variable load upon the exciter and tends to introduce distortion by reducing the exciting voltage at the peaks of modulation. Trouble of this sort can be minimized by not driving the grid of the linear amplifier any farther positive than necessary to develop a conservative output, and by employing an exciter of ample power capacity so that its voltage regulation will be good.

The linearity in any particular case can be obtained by observing the output voltage as a function of exciting voltage, using point-by-point measurements, a cathode-ray tube, or a modulation meter reading the minimum and maximum amplitudes of the modulated-wave envelope.

Since the pulses of plate current of the usual linear amplifier approximate half sine waves having an amplitude proportional to the exciting voltage, the d-c plate current of a linear amplifier will have a value that is approximately proportional to the average amplitude of the exciting voltage. The d-c plate current accordingly tends to be independent of the degree of modulation when the amplification is linear but otherwise will depend upon the degree of modulation. This fact is commonly used as a rough indication as to whether or not bad distortion is present in a linear amplifier.

Efficiency of Linear Amplifiers.—The plate efficiency of a linear amplifier, assuming cut-off bias and straight-line tube characteristics, is the same as in the Class B audio amplifier under the same assumptions. The efficiency accordingly cannot exceed 78.5 per cent, and in any particular case is given by Eq. (92). Practical efficiencies at the crest of the modula-

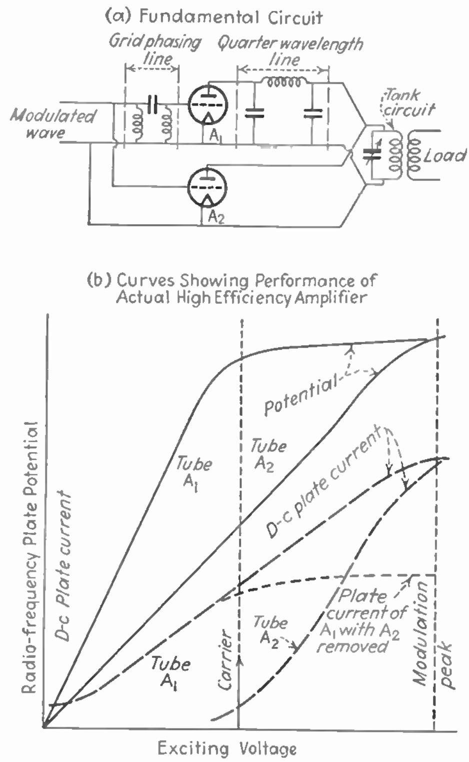


FIG. 99.—Schematic diagram of high-efficiency linear amplifier.

tion cycle are around 50 to 60 per cent. Since the efficiency is proportional to the exciting voltage, the efficiency for the unmodulated carrier is only half as great, or 25 to 30 per cent. The carrier power that can be developed by a linear amplifier will normally not exceed one-fourth the power that the same tube will develop when operating as a Class C amplifier.

High-efficiency Linear Amplifiers.—The poor efficiency of the linear amplifier can be avoided by the special form of linear amplifier shown schematically in Fig. 99*a*. Here the amplifier is divided into two parts with the output of one tube being delivered to the load through a quarter-wave transmission line as shown.[1] The same exciting voltage is applied to both parts of the amplifier, but A_2 is biased so that no plate current flows in this tube until the exciting voltage exceeds the carrier amplitude. On the other hand, A_1 is biased to projected cut-off and so operates as an ordinary linear amplifier. The impedance in the plate circuit of amplifier tube A_1 is adjusted so that, with the second tube inoperative, the output voltage just begins to flatten off with carrier excitation. The actual load that receives the output of the linear amplifier is coupled to the tank circuit of A_2. The output of A_1 is transmitted to this load through a quarter-wave-length line. The outputs of the two tubes are arranged to add in phase by employing a compensating phase shift in the grid circuit of A_1, such as shown in Fig. 99*a*. The circuits are designed so that if the proper plate-circuit impedance for tube A_1 with the other tube inoperative is called Z_L, then the characteristic impedance of the line in the plate circuit of A_1 is made $Z_L/2$, and the impedance that the tank circuit of tube A_2 connects across the receiving end of the line is made $Z_L/4$.

The operation can be explained as follows: For exciting voltages corresponding to carrier amplitude and less, tube A_2 is inoperative. A_1 then acts as a linear amplifier developing an output voltage in its plate circuit proportional to the exciting voltage. Under these conditions the voltage across the tank circuit of amplifier A_2 is half the voltage developed in the plate circuit of amplifier A_1. This is because of the action of the quarter-wave-length line.

At exciting voltages greater than the carrier, amplifier A_2 begins to supply output. From the point of view of the line and A_1 there is then more voltage developed across the tank circuit of A_2 in proportion to the power delivered by the line than was previously the case. Hence, as far as the line is concerned, the receiving-end load impedance of the line has been effectively increased. Now in a quarter-wave-length transmission line the impedance at the sending end of the line is equal to Z_0^2/Z_2, where Z_0 is the characteristic impedance of the line and Z_2 is the load impedance at the receiving end. Consequently, as amplifier A_2 delivers energy to the load, the impedance which the sending end of the line presents to the plate of A_1 is decreased. This enables A_1 to deliver more output to the load in spite of the fact that the alternating voltage in its

[1] In actual practice the quarter-wave-length line is replaced by a network that behaves as far as the sending and receiving terminals are concerned as a transmission line that is a quarter of a wave length long. This network may consist of series inductance and shunt condensers, as shown in the plate circuit of tube A_1 in Fig. 99*a*, or can be supplied by a series condenser and shunt inductances of the appropriate values, as in the grid circuit of tube A_1. The only difference between the two types of lines is that the phase shift is $+90°$ in one case and $-90°$ in the other.

plate circuit cannot increase appreciably. At the peak of the modulation cycle, A_1 and A_2 are delivering equal amounts of power to the load, and the alternating voltages across the two tank circuits are equal.

The detailed voltage and current relations in a typical high-efficiency linear amplifier are shown in Fig. 99*b*. It will be noted that, although the voltage developed in the plate circuit of A_1 flattens off at the carrier level, the d-c plate current (and hence the output power) keep on increasing up to the crest of the modulation cycle. If it were not for the tube A_2 and the impedance-inverting action of the quarter-wave line, the d-c plate current of A_1 would flatten off at the carrier level as shown by the dotted line. It will be noted that, although the plate current of A_2 does not become appreciable until the excitation exceeds carrier level, the d-c current increases twice as fast with further excitation as does the plate current of A_1 and so at the peak of the modulation cycle reaches approximately the same value.

The efficiency of the arrangement of Fig. 99*a* varies slightly with changes in percentage of modulation, but under practical circumstances averages 50 to 60 per cent, which is nearly twice the average efficiency of an ordinary linear amplifier. This high average efficiency comes about because A_1 operates most of the time with a relatively high alternating voltage in its plate circuit and hence with good efficiency. At the same time A_2 operates as a Class C amplifier which has good efficiency at the peak of the modulation cycle and ceases to deliver output under conditions for which the efficiency would be low.

60. Harmonic Generators.—By taking advantage of the fact that the pulses of plate current have appreciable harmonic content, a Class C amplifier can be used to generate output power that is a harmonic of the signal voltage applied to the control grid. It is merely necessary to tune the tank circuit to the desired harmonic and adjust the length of the plate-current pulse to a value that is favorable for generating the harmonic involved. It is to be noted that neutralization is unnecessary, since the grid and plate circuits are tuned to widely different frequencies. Harmonic generators of this character are frequently used in radio transmitters and for other communication purposes.

Oscillograms showing voltage, current, and power relations in a typical harmonic generator are shown in Fig. 100. Examination shows that these are almost identical with the corresponding oscillograms of Fig. 94 for a Class C amplifier. Hence most of the essential factors such as E_{max} and E_{min} are the same in the harmonic-generator case as with the Class C amplifier. The only difference is that since the harmonic content of the plate-current pulse depends upon the length of the pulses, this length must be carefully chosen in relation to the harmonic to be generated.

Under practical conditions the optimum pulse lengths are in the range given in Table XI.[1]

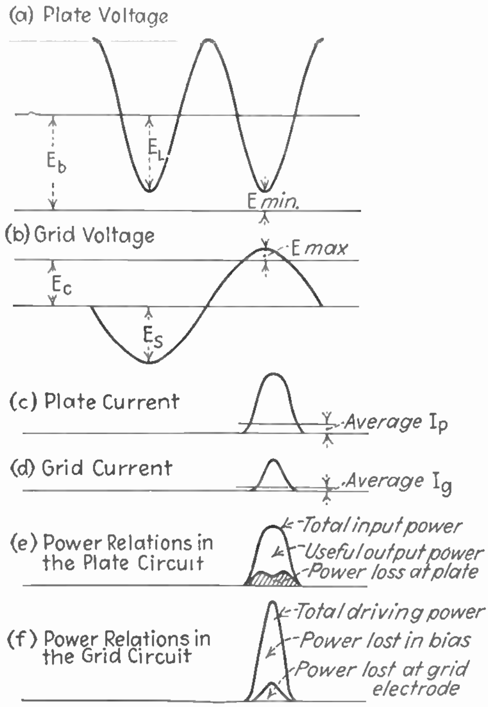


FIG. 100.—Voltage, current, and power relations of typical Class C harmonic generator. Note similarity to the curves of Fig. 94.

TABLE XI.—PLATE-CURRENT PULSE LENGTH AND POWER OUTPUT OF HARMONIC GENERATORS

Harmonic	Optimum length of pulse in electrical degrees at the fundamental frequency	Approximate power output, assuming normal Class C output is 1.0
2	90–120	0.65
3	80–120	0.40
4	70– 90	0.30
5	60– 72	0.25

[1] The grid bias required with a triode tube having an amplification factor μ and generating the nth harmonic is

$$\text{Grid bias} = \frac{E_b[1 - \cos(n\theta/2)] + E_{\min}\cos(n\theta/2)}{\mu[1 - \cos(\theta/2)]} + \frac{E_{\max}\cos(\theta/2)}{1 - \cos(\theta/2)} \qquad (94)$$

where E_b is the plate-supply voltage, θ is the length of the plate-current pulse in electrical degrees at the fundamental frequency, and $E_{\min}$ and $E_{\max}$ are minimum plate voltage and maximum grid voltage, respectively, as shown in Fig. 100.

The shorter the length of the current pulse in the case of a particular harmonic, the higher the plate efficiency is, but more bias, and hence more exciting voltage and greater driving power, are required. The plate efficiencies obtained in harmonic generation for conditions given in Table XI are somewhat less than those realized with Class C operation, but since the output obtained is likewise less, the plate dissipation is ordinarily no greater.

Harmonic generators are normally arranged to generate the second harmonic, although third- and fourth-harmonic generation is occasionally employed. Harmonics higher than the fourth are practically never used because, as the angle of flow becomes less than 90°, the grid bias, and hence the driving power, increase very rapidly and quickly become excessive.

The power output obtainable from a properly adjusted harmonic generator is almost inversely proportional to the order of the harmonic. The approximate ratio of the obtainable harmonic output to the output of the same tube, when acting as a Class C amplifier, is given in the last column of Table XI.

Problems

1. *a.* Describe a method of deriving a dynamic characteristic such as shown in Fig. 82, given the load resistance, the operating point, and a set of E_g-I_p characteristic curves.

b. Draw such a dynamic characteristic for an actual tube as assigned (or for the tube of Fig. 82, when the operating point is as shown and the load resistance is 6000 ohms).

2. A Class A power amplifier is to be operated without driving the grid positive. If the plate-supply voltage and grid bias at the operating point are kept constant, the proper load resistance will depend upon the amount of distortion that can be permitted. Explain how this comes about and in particular be certain to indicate whether the load resistance should be increased or decreased as the allowable distortion becomes less.

3. If the plate-supply voltage is increased, but the plate current at the operating point is kept constant, demonstrate that the proper load resistance for Class A amplifier operation with no grid current increases.

4. In a Class A amplifier, compare the power dissipated at the plate of the tube when a signal is applied to the grid, with the power dissipated when there is no signal.

5. *a.* Calculate the power output and plate efficiency (assuming the rectified current is zero) for: (1) the three dynamic characteristics in Fig. 84, assuming that in each case the signal is just sufficient to bring the operation to zero instantaneous grid voltage; and (2) curve *b* in Fig. 83 for negative-grid operation and curve *c* of the same figure for positive-grid operation when operation extends just to the edge of the region of high distortion.

b. Discuss the results with respect to factors affecting efficiency, power output, and exciting voltage required.

6. Using a 2A3 tube as a Class A amplifier with the conditions recommended for E_b = 250 volts (see Table IX), calculate the following when the maximum permissible signal (45 volts crest) is applied: (*a*) a-c plate current; (*b*) alternating-current voltage

developed across the load; (c) maximum and minimum instantaneous plate voltage and currents reached during the cycle (assuming zero distortion); (d) power dissipated in load; and (e) plate efficiency.

7. Using a set of I_p-E_p characteristic curves of an actual power tube, determine the following for an operating condition recommended as suitable for Class A power amplification: (a) maximum power output; (b) second-harmonic distortion; and (c) actual plate efficiency, taking into account the "rectified" plate current.

NOTE: The curves are preferably a full-page blueprinted reproduction of some actual tube. However, it is possible to use the curves of Fig. 35, using the operating point $E_b = 250$, $E_c = -28$ and $R_L = 5000$.

8. *a.* The tube of Fig. 83 is operated with a load resistance corresponding to dynamic characteristic *c*, and sufficient signal is applied to bring the operation just to the edge of the region of high distortion. Determine the maximum allowable source impedance of the driving voltage if the flattening of the positive peaks of the exciting voltage by the grid current is not to introduce more than 3 per cent second-harmonic distortion.

b. If the driver tube has a plate resistance of 10,000 ohms, determine the ratio of secondary to primary turns that the output transformer of the driving stage should have in order that the driver will offer the proper source impedance.

9. The grid of a single Class A power tube is excited from a tube having a plate resistance of 10,000 ohms, by a transformer in which the ratio of primary to secondary turns is 2. If the crest exciting voltage desired is 60 volts and the instantaneous grid current at the positive peak is 6 ma, calculate the second-, third-, and fourth-harmonic distortion produced in the exciting voltage by the grid current.

10. Sketch the shape of output wave that can be expected from a Class A power amplifier when distortion arises from (a) extending operation well into the high-distortion region of low plate current, and (b) positive-grid operation with a high source impedance.

11. A resistance-coupled voltage amplifier uses the pentode tube of Fig. 40. If the operating point is $E_c = -1.5$, $E_{sg} = 100$, $E_b = 100$, draw a load line for a load resistance of 30,000 ohms and determine the second-harmonic distortion in the output voltage when the applied signal is 1.5 volts crest.

12. Using a set of E_p-I_p characteristics of an actual pentode (or beam) tube, determine the following for a Class A operating condition recommended by the manufacturer: (a) maximum power output; (b) second- and third-harmonic distortion; and (c) actual plate efficiency, taking into account the "rectified" plate current.

NOTE: Curves are preferably a full-page blueprinted reproduction for some actual tube, but in the absence of such curves it is possible to use characteristics given in a tube manual.

13. *a.* Calculate the output power, the percentage of second and third harmonics, and the plate efficiency for the beam tube of Fig. 87*b* when the load resistance corresponds to the load lines *a*, *b*, and *c*, when the exciting potential is 15 volts crest.

b. Sketch the shape of the output wave for each load line in *a*, using Fig. 87*b*, and correlate the results with the corresponding dynamic characteristic curves of Fig. 87*a*.

14. The transformer of Fig. 89 is used with a load resistance of 25 ohms and a tube having a plate resistance of 2500 ohms. Calculate and plot the frequency response that can be expected.

15. A particular output transformer is to be used to couple the plate circuit of a tube having $R_p = 2000$ ohms, to a load of 600 ohms. The transformer must offer a load impedance of 4000 ohms to the tube and give a response that does not drop to less than 70.7 per cent of the mid-range value between 80 and 6000 cycles. Specify the required turn ratio, the largest permissible leakage inductance, and the lowest permissible primary inductance. Neglect resistance of windings.

16. Derive Eqs. (88) and (89).

17. Discuss the factors that determine the efficiency of an output transformer (*i.e.*, the ratio of power delivered to the load to the power delivered by the tube to the primary of the transformer).

18. An output transformer with the following constants is to be used with a push-pull amplifier employing Type 6F6 tubes operated as Class A triode amplifiers (not Class AB).

Primary inductance	18	henries
Leakage inductance (referred to primary)	0.20	henry
Step-down ratio	30	
Resistance, primary	400	ohms
Resistance, secondary	1.2	ohms

Determine the proper load resistance that should be used and calculate and plot the frequency response that can be expected with this load resistance.

19. An output transformer, designed for a push-pull Class A amplifier, will not give a correct impedance match to the load with a single output tube when one-half the center-tapped winding is used as a primary. Explain, and state where the tap should be placed.

20. A Class AB amplifier operates as a Class A amplifier for small signal voltages. Explain.

21. Check the operating conditions specified for Class B operation of the Type 800 tube in Table X. Do this by determining the crest alternating-current voltage across the load (and hence the minimum instantaneous plate potential) on the basis of the power output and load resistance given in the table. Then with the aid of the characteristic curves of Fig. 49, and Eqs. (91) and (92), determine power output and plate efficiency for the exciting conditions in the table and the calculated E_{min}, and compare with the results given in the table. Also calculate the load resistance by Eq. (90) and compare with the value in the table.

22. In a Class B amplifier, sketch the shape of the output wave that results when the load resistance is made excessively high and explain why this shape is obtained.

23. Design an audio-frequency amplifier that will deliver an "undistorted" output of 15 watts when operated from a microphone that has an internal impedance of 600 ohms and develops a voltage of approximately 50 μv on open circuit. The over-all frequency response should not vary more than 3 db over the frequency range of 80 to 8000 cycles. The load impedance to which the output is delivered is 10 ohms. The design includes selection of tubes, circuit layout, circuit constants, provision for volume control, specification of transformer constants, design of filter systems to prevent regeneration, etc., as well as calculation of expected frequency response. Make use of tube manuals, catalogs, etc., as required, and where complete data are not available, make reasonable estimates, just as the designer of commercial equipment would be required to do.

24. *a.* Make an approximate copy of the oscillograms of Fig. 94. Then show by dotted lines on the same curves the oscillograms that would result with the same tube if the length of the plate-current pulses were reduced by increasing the grid bias, while keeping E_{max}, E_{min}, and E_b constant. Point out the significant differences in the results of the two cases.

b. Repeat (*a*) but use the dotted lines to show the effects of making E_{min} larger, while keeping length of current pulses, and the maximum grid potential E_{max} the same. Assume that the change in grid bias resulting from the different value of E_{min} is negligibly small, as is the case in ordinary tubes in which the amplification factor is reasonably high.

25. In a Class C amplifier with fixed bias equal to twice cut-off, sketch a curve showing qualitatively ~~the way~~ in which the crest amplitude of the alternating voltage between plate and cathode of the tube will vary as a function of the alternating voltage applied to the grid. Do this for two values of tank-circuit impedance.

26. Derive a formula giving the length of the plate-current pulses of a Class C triode amplifier in electrical degrees, in terms of E_{max}, E_{min}, E_c, E_b, and the amplification factor of the tube.

27. Design a tank circuit suitable for use with a Type 852 tube operating at 7500 kc in the circuit of Fig. 95*a*, when the operating conditions given in Table X are to be realized. This design includes selection of a suitable effective Q and calculation of the tank-circuit inductance and capacitance.

28. At a certain point in the process of adjusting a particular Class C triode amplifier assume that it is found that the d-c plate current and output power are less than the desired value even though the grid bias is the proper value, and the plate efficiency is satisfactory. What readjustments should be made to remedy this situation?

29. In a Class C triode amplifier it is found that the output power and plate efficiency are satisfactory, but the driving power required is excessive. What adjustments can be made to remedy this situation?

30. *a.* If the coupling to the load in a properly adjusted Class C triode amplifier is removed, the d-c plate current will decrease to a small fraction of its original value, while the d-c grid current will increase somewhat. Explain the reasons for this.

b. If in (*a*) the removal of the load causes only a moderate decrease in d-c plate current (such as 50 per cent), what does this mean?

31. A Class C triode amplifier is operating under satisfactory conditions, using grid-leak bias. If the value of the grid-leak resistance is increased very considerably but no other adjustments made, what will be the effect on the grid bias, d-c plate current, grid driving power, and alternating voltage across the load? In each case explain the reasons for the answers given.

32. A Class C triode amplifier is operating under satisfactory conditions. The coupling between tank circuit and load is then increased (*i.e.*, the tank-circuit impedance is reduced). What will be the effect on the d-c grid and plate currents and the plate efficiency?

33. Derive Eq. (93).

34. The Type 852 tube of Table X is to be used as a linear amplifier. Estimate (1) the carrier power that can be expected, (2) the proper grid-bias voltage, (3) the exciting voltage at the crest of the modulation cycle, and (4) the approximate impedance the tank circuit should develop between plate and cathode.

35. *a.* Explain why a Class C amplifier cannot be used to amplify a modulated wave in place of a linear amplifier.

b. Explain why linear amplifiers are used for the amplification of modulated waves instead of Class A amplifiers.

36. In a high-efficiency linear amplifier sketch the following curves for one modulation cycle of a completely modulated wave: (1) modulation envelope, (2) envelope of voltage across the tank circuit of A_2, (3) envelope of voltage between plate and cathode of A_1, (4) plate current of A_2, and (5) plate current of A_1.

37. Calculate the grid bias required with a Type 800 tube for the generation of the second, third, fourth, and fifth harmonics, assuming $E_b = 1000$, $E_{min} = 150$, and $E_{max} = 120$. From the results tabulate the exciting voltages required for the different-order harmonics and discuss the results from the point of view of relative driving power.

38. Derive Eq. (94).

39. Tabulate relative values of impedance that it is desired for the tank circuit of a harmonic generator to develop between plate and cathode when generating the second, third, fourth, and fifth harmonics, assuming that the impedance for Class C operation is taken as unity.

CHAPTER VII

VACUUM-TUBE OSCILLATORS

61. Vacuum-tube Oscillator Circuits.—A vacuum tube is able to act as an oscillator because of its ability to amplify. Since the power consumed by the input of an amplifier tube is less than the amplified output, it is possible to make the amplifier supply its input. When this

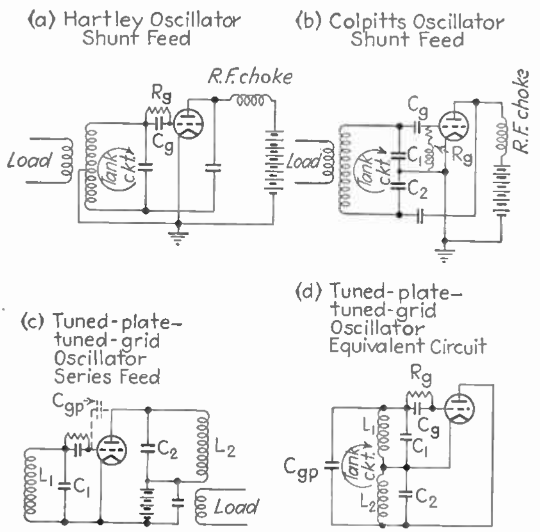


FIG. 101.—Typical oscillator circuits. In each case the frequency is determined by a resonant circuit, and the arrangement is such that the tube acts as an amplifier supplying its own input voltage.

is done, oscillations will be generated, and the tube acts as a power converter that changes the direct-current power supplied to the plate circuit into alternating-current energy in the amplifier output.

Any amplifier circuit that is arranged to supply its own input voltage in the proper magnitude and phase will generate oscillations. Many circuits can be used for this purpose, of which a number are shown in Fig. 101. In general, the voltage fed back from the output and applied to the grid of the tube must be approximately 180° out of phase with the voltage existing across the load impedance in the plate circuit of the amplifier and must have a magnitude sufficient to produce the output power necessary to develop the input voltage. In the Hartley and

Colpitts circuits this is accomplished by applying to the grid a portion of the voltage developed in the resonant circuit (*i.e.*, tank circuit). In the tuned-plate–tuned-grid circuit the energy transfer from plate to grid tuned circuits takes place through the grid-plate capacitance of the tube.

The frequency at which the oscillations occur is the frequency at which the voltage fed back from plate circuit to the grid is of exactly the proper phase to enable the tube to supply its own input. In oscillators associated in some way with a resonant circuit, as are all those of Fig. 101, the frequency of oscillation approximates very closely the resonant frequency of this tank circuit.[1]

62. Design and Adjustment of Power Oscillators.—In the usual case where the object of a power oscillator is to develop a large amount of power at good efficiency, the tube is adjusted to operate as a Class C amplifier. The power output, plate efficiency, etc., then depend upon the grid-bias voltage, the tank-circuit impedance, the maximum grid potential E_{max}, the minimum plate potential E_{min}, etc., exactly as in the case of a Class C amplifier. The only special factor involved is that the coupling between the grid and plate electrodes must be such that when the desired alternating voltage exists between plate and cathode electrodes (*i.e.*, crest voltage slightly less than the plate-supply voltage), the proper exciting voltage for Class C amplifier operation will be obtained.

A power oscillator normally uses grid-leak bias in order to insure that the oscillator will be self-starting. With grid leak, there is no bias when the electrode voltages are first applied, so that the tube then has a high mutual conductance and is in a condition to amplify. Any thermal-agitation voltage of the proper frequency will accordingly start the oscillations building up. On the other hand, when a fixed bias is employed, the tube will be biased beyond cut-off when the electrode voltages are applied, and oscillations cannot possibly start.

Circuit Design.—The first step in setting up a power oscillator is to design the tank circuit. In such circuits as the Hartley and Colpitts, it is to be noted that the potential developed across the tank circuit will ordinarily be greater than with the same tube employed as a Class C amplifier. Thus in Figs. 101*a* and 101*b* the total voltage developed across the tank circuit is the sum of the grid exciting voltage and the alternating voltage developed between plate and cathode. The actual tank-circuit design is carried out in the same manner as in the case of a Class C ampli-

[1] In the tuned-plate–tuned-grid oscillator the circuit effective in determining the frequency is the grid-plate tube capacitance resonating with the grid and plate tuned circuits, which are adjusted to offer an inductive reactance at the frequency of oscillation. This is apparent from the equivalent circuit of Fig. 101 and is discussed below in greater detail.

fier, using Eq. (93), except that the effective Q is ordinarily made somewhat greater than with Class C amplifiers in order to improve the frequency stability (see Sec. 63). Where shunt feed is employed, the considerations relating to the shunt-feed choke are the same as in the case of the corresponding Class C amplifier. That is, the choke must have a high reactance compared with the reactance of the coil or condenser it parallels.

After the tank circuit has been proportioned, the next step is to determine the approximate location of the connections from the tank circuit to the grid, plate, and cathode electrodes. Starting with the voltage to be expected across the tank circuit, the coupling to the plate must be such that the alternating voltage developed between plate and cathode will have a crest value just less than the plate-supply voltage. At the same time, the coupling to the grid must be such as to provide the exciting voltage corresponding to the Class C conditions desired. The circuit adjustments that are required to realize the necessary couplings can be determined by calculation, measurement, or by a trial-and-error process of adjustment.

The grid-leak resistance must be such as to develop the required grid bias with the grid current corresponding to the desired operating conditions. The grid condenser should have a capacitance somewhat larger than the input capacitance of the tube and should also have a much lower impedance to radio frequencies than does the grid-leak resistance. At the same time the grid condenser must not be too large, or intermittent oscillations may be produced, as discussed below.

The detailed procedure for laying out a power oscillator can be understood by the following example:

Example.—Design the circuits of a power oscillator using the shunt-feed Colpitts circuit of Fig. 101*b*, and a Type 800 tube at a plate potential of 1000 volts (see Table X). The frequency is 3000 kc, and the effective tank-circuit Q is to be 40.

The first step is the tank-circuit design. The total voltage across the tank circuit is the sum of the alternating plate-cathode voltage and the exciting voltage. The former will have a crest value of the order of 15 per cent less than the plate-supply voltage, or about 850 volts crest, while the latter is given in Table X as 260 volts. Hence from Eq. (93), with 50 watts output from Table X,

$$\omega L = \frac{(850 + 260)^2}{2 \times 50 \times 40} = 305 \text{ ohms}$$
$$L = 16.25\ \mu\text{h}$$

The capacitances C_1 and C_2 are determined by the fact that the sum of their reactances must be 305 ohms, and C_2 must develop a voltage drop that is 850/(850 + 260) of the total voltage across the tank circuit. Hence C_2 must have a reactance of

$$\frac{305 \times 850}{(850 + 260)} = 234 \text{ ohms,}$$

while the reactance of C_1 is $305 - 234 = 71$ ohms. The corresponding capacitances are 227 $\mu\mu$f, and 748 $\mu\mu$f.

The shunt-feed choke should have a reactance much greater than the reactance of the condenser that it shunts. A value 20 times the tank-circuit inductance is adequate, which calls for a choke inductance of at least 325 μh.

The grid-leak resistance must develop the bias of -135 volts called for in Table X, when the grid current has the expected value of 15 ma. The resistance must hence be $135/0.015 = 9000$ ohms.

The grid condenser is not critical. A value of the order of 250 $\mu\mu$f will be satisfactory, as this is considerably greater than the input capacitance of the tube, is an adequate short circuit at the frequency of oscillation, and is still small enough to minimize trouble from intermittent oscillations.

Practical Adjustment of Power Oscillators.—After the circuit of a power oscillator has been laid out as indicated above, with the grid, plate, and cathode connections located either by an exact measurement, calculation, or by an estimate, the electrode voltages are turned on, and the oscillator tank circuit adjusted to oscillate at the required frequency. The coupling between the load and tank circuit is then varied until the expected power input to the tube is obtained, after which the power output, grid bias, and plate efficiency are observed. If these are not satisfactory, the grid-leak resistance, the coupling of the grid and plate electrodes to the tank circuit, and the coupling of the load to the tank circuit are varied until the desired conditions are realized.

In making the final adjustments, it is helpful to keep in mind certain fundamental relations. Thus, with any adjustment giving reasonable plate efficiency, the current in the tank circuit assumes an amplitude such that the crest alternating voltage between plate and cathode is slightly less than the plate-supply potential. Hence, increasing the coupling between the grid-cathode electrodes and the tank circuit will increase the exciting voltage without producing an appreciable change in the voltage developed across the tank circuit or in the tank-circuit current. On the other hand, increasing the coupling between the tank circuit and the plate-cathode electrodes, with the coupling to the grid left unchanged, reduces the exciting voltage and reduces the tank-circuit current. This is because, with greater coupling to the plate, a smaller tank-circuit current will develop the required alternating plate potential. Altering the ratio of grid-cathode to plate-cathode voltage by varying the cathode connection to the tank circuit while leaving the plate and grid connections unchanged, changes both the exciting voltage and the tank-circuit current.

Varying the grid-leak resistance in a power oscillator produces, as its principal effect, changes in the amount the grid is driven positive. This is because a higher leak resistance develops the required bias with a proportionately smaller grid current and hence with a less positive grid.

Changes in the coupling between the load and the tank circuit have relatively little effect upon the amplitude of oscillations but a very great effect upon the d-c plate current. This can be understood by considering what happens when the resistance coupled into the tank circuit is reduced. At the first instant the oscillations tend to increase in amplitude. This makes the minimum plate potential E_{min} smaller, thereby decreasing the size of the plate-current pulses. Since the minimum plate potential E_{min}, and hence the amplitude of the plate-current pulses, are quite sensitive to small changes in the amplitude of the oscillations, the final result of reducing the load coupling is then a relatively large decrease in d-c plate current accompanied by a relatively small change in the amplitude of the oscillations.

Special Consideration Relating to Tuned-grid–Tuned-plate Oscillators.—The tuned-plate–tuned-grid oscillator can be reduced to the equivalent circuit of Fig. 101*d*. Examination of this circuit shows that oscillations can be obtained at a particular frequency by tuning both plate and grid tuned circuits to a slightly higher frequency than is to be generated. The resulting inductive reactance of the two tuned circuits then resonates with the grid-plate tube capacitance to form an oscillator circuit similar in many respects to the Hartley circuit.

The plate tuned circuit of the oscillator is designed by the use of Eq. (93), assuming that the entire oscillator output is delivered to this circuit and that the crest alternating voltage is only slightly less than the d-c plate voltage. The exact proportions of the grid tuned circuit are not particularly important, provided the proper tuning range is obtained.

In the practical adjustment of a tuned-plate–tuned-grid oscillator, the frequency is determined primarily by the tuning of the plate circuit, and the excitation is controlled by the adjustment of the grid tuned circuit. The tuning-up procedure accordingly consists in first making the plate circuit resonant at approximately the desired frequency and then varying the condenser of the grid tuned circuit and the coupling of the load to the plate tuned circuit until the grid current, plate current, and power output given by the tube manufacturer for Class C amplifier operation are approximated. The frequency of oscillation is then checked, and readjustments made as necessary.

Intermittent Operation.—When a grid leak is used to develop the grid bias, it is sometimes found that the oscillations are periodically interrupted. These interruptions may be at an audible rate or they may be at a radio frequency. They are the result of an excessively large time constant R_gC_g for the grid-leak–grid-condenser combination, and they can be cured either by decreasing the grid-leak resistance R_g or by reducing the grid-condenser capacitance C_g, or both.

The process involved in intermittent operation is as follows: The oscillations first build up in amplitude to the equilibrium condition. Any slight irregularity tending to reduce the amplitude of oscillations will then cause the oscillations to die out because, with the large time constant R_gC_g, the bias tends to remain constant, whereas, in order to prevent the oscillations from dying out, it is necessary for the bias to reduce as the amplitude of oscillations reduces. After cessation of oscillations the grid condenser gradually discharges through the grid leak, reducing the bias until the tube will again amplify. An oscillation thereupon builds up again to the equilibrium value, to repeat the process.

Synchronization of Vacuum-tube Oscillators.—Vacuum-tube oscillators have an inherent tendency to synchronize with any other oscillations of approximately the same frequency that may be present. The behavior of two oscillators loosely coupled together and generating frequencies that are not widely different illustrates what can be expected. If the two frequencies differ by only a small percentage, they are both shifted from their normal values in such a way as to reduce the difference. This attraction of the two frequencies becomes more pronounced as the difference between the normal oscillating frequencies is reduced, and finally becomes so great that the oscillators pull into synchronism. The extent to which the frequency of an oscillator can be shifted from its normal value by the presence of currents of a slightly different frequency will be greater as the amplitude of the injected current is increased and as the frequency stability of the oscillator is lowered. The natural tendency of vacuum-tube oscillators to synchronize means that parallel operation of several tubes in an oscillator circuit presents no difficulties.

Tubes for Power Oscillators.—The same tubes employed for triode Class C power amplifiers (see Sec. 58) are also used for power oscillators. For output powers up to a few kilowatts this means that air-cooled tubes are used, whereas water-cooled tubes are employed where greater power is required.

63. Frequency and Frequency Stability of Generated Oscillations.—The alternating current generated by a vacuum-tube oscillator has a frequency such that the voltage which the oscillations apply to the grid of the tube is of exactly the proper phase to produce the oscillations that supply this grid exciting voltage. This approximates the resonant frequency of the tank circuit, but the exact value is also usually influenced by such factors as the effective Q of the tank circuit, the electrode voltages of the tube, the load coupling, etc.

For many requirements it is essential that the generated frequency be as nearly constant as possible. The first step in achieving this is to maintain constant the resonant frequency of the tank circuit. Factors that can cause the resonant frequency to change are aging of coils, condensers, tubes, etc., variation of inductance and capacitance with temperature, variations in the interelectrode capacitance of the oscillator tube as a result of temperature effects or tube replacements, and variations in the reactance that the load couples into the tuned circuit. Temperature effects are particularly troublesome because the energy dissipated in the circuit causes progressive heating that is very difficult to compensate for. The temperature coefficient of a tuned circuit depends upon the construction, but is commonly in the range 10 to 100 parts in a million per degree centigrade.

Constancy of the resonant frequency of the tank circuit does not insure a stable frequency, however, since the frequency actually generated normally differs from the resonant frequency by an amount that depends upon the electrode voltages and the effective resistance of the tank circuit. Thus, consider the Colpitts circuit of Fig. 102*a*, and simplify by assuming that the shunt-feed choke has infinite reactance, that the grid condenser

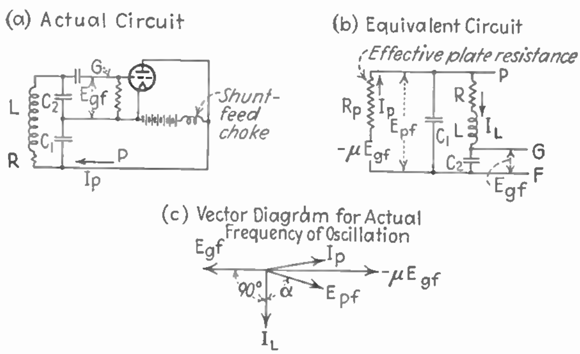


FIG. 102.—Circuit of Colpitts oscillator together with simplified circuit and vector diagram showing how tank-circuit resistance makes it necessary for the generated frequency to differ slightly from the resonant frequency of the tank circuit.

has negligible reactance, that the grid current is negligibly small, and that the grid-leak resistance is infinitely large. This gives the equivalent circuit of Fig. 102*b*. At the frequency of oscillation, the vector diagram giving the phase relations in this equivalent circuit is as shown at Fig. 102*c*, where E_{gf} is the exciting voltage applied to the grid, and $-\mu E_{gf}$ is the equivalent voltage acting in the plate circuit. Here the frequency of oscillation is slightly higher than the resonant frequency of the tank circuit, and the tank-circuit impedance offered to $-\mu E_{gf}$ is capacitive. The current I_p flowing into the tank circuit from the plate of the tube hence leads the voltage $-\mu E_{gf}$, while the resulting voltage drop E_{pf} across the condenser C_1 in the tank circuit lags $-\mu E_{gf}$, as shown in Fig. 102*c*. The current I_L through the inductive branch of the tank circuit lags behind E_{pf} by the angle α, which is less than 90° as a result of the resistance R representing the sum of coil resistance and coupled load resistance. The exciting voltage E_{gf} for this simplified case is the voltage drop produced by the current I_L flowing through the condenser C_2 and so lags 90° behind I_L. This voltage must be exactly 180° out of phase with the vector representing $-\mu E_{gf}$ in the diagram.

Study of Fig. 102*c* shows that it is necessary for the frequency of oscillation to differ from the resonant frequency of the tank circuit in order to produce a phase shift that will compensate for the effect on the phase relations produced by the resistance R. The amount of this frequency deviation from the resonant frequency will change (with a resulting variation in frequency) as the resistance of the tank circuit is varied, or as the electrode voltages of the tube alter the effective plate resistance. If, in addition, the effect of the grid current, grid-leak resistance, grid-condenser capacitance, shunt-feed choke, etc., is taken into account, the situation is more complicated, although the results are of the same general character.

The most important factor in obtaining a stable frequency is the use of a tank circuit having a high effective Q.[1] This is because the frequency change required to develop the phase shift that compensates for variations in load resistance, tube voltages, etc., is inversely proportional to the tank-circuit Q. The frequency stability is also helped by operating the tube so that the grid current is small and constant (as when a high grid-leak resistance is used) and by operating the oscillator with a conservative power output.

The frequency stability of an oscillator can be improved by arranging the circuits so that the frequency of the generated oscillations is exactly the same as the resonant frequency of the tank circuit. This can be done by placing a suitable reactance in series with the grid or plate electrodes to produce the phase shift that would otherwise have to be obtained by a shift in frequency. In the case of the Hartley and tuned-plate–tuned-grid circuits, the compensating reactance can be a capacitance in series with the grid and so can be supplied by adjusting the grid condenser to a suitable value. With the Colpitts circuit the compensating reactance can also be a capacitance in the grid lead, provided there is at the same time a small inductance in series with the plate electrode. The proper capacitance in any case can be readily determined experimentally by finding the grid-condenser setting for which there is the smallest change in frequency when the plate voltage is varied.

When frequency stability is important, the load impedance that is coupled into the tank circuit should be as nearly resistive as possible. This is because any reactance coupled into the tank circuit greatly alters the resonant frequency, whereas coupled resistance has only a slight

[1] A high tank-circuit Q corresponds to a low L/C ratio in the tank circuit, since Eq. (93) shows that the tank-circuit inductance L is inversely proportional to Q. Hence, the common statement that a low L/C ratio improves stability is the same as saying that a high Q is desired.

effect. Hence, by making the load resistive, variations in load have a minimum effect on the frequency.

Master-oscillator Power-amplifier and Electron-coupled Arrangements.—The difficulty of obtaining a large power output with maximum frequency stability has led to the wide use of master-oscillator power-amplifier arrangements. In these an oscillator designed to have a very stable frequency is used to excite a Class C amplifier. By operating the amplifier so that the driving power is small, the oscillator is effectively isolated from the load coupled into the tank circuit of the power amplifier and so operates under conditions favorable for frequency stability.

When a properly designed master-oscillator power-amplifier arrangement is operated so that the tank circuit is either at constant temperature or is temperature compensated to have the same resonant frequency over

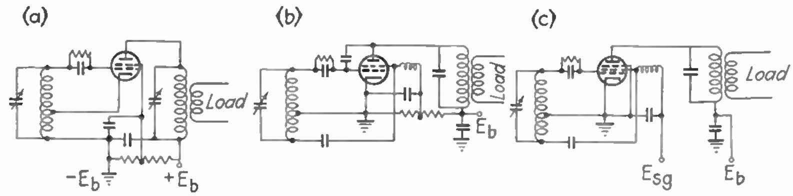


FIG. 103.—Electron-coupled oscillator circuits.

ordinary temperature ranges, a frequency stability of the order of 20 to 100 parts per million can readily be maintained under commercial conditions.

The electron-coupled oscillator consists of a screen-grid tube in an arrangement such as shown in Fig. 103*a*. Here the cathode, control grid, and screen grid are operated as a triode oscillator with the screen serving as the anode. Only a few electrons are intercepted by the screen, but the circuits can be so designed that these are enough to maintain the oscillations. The remaining electrons, which represent most of the space current, go on to the plate and through the load impedance that is connected in series with the plate electrode. This arrangement is called an *electron-coupled oscillator,* because the oscillator and output circuits are coupled by the electron stream. By operating the screen grid at ground potential, as shown in Fig. 103*a*, or neutralizing as in Fig. 103*b*, capacitive coupling between the output and oscillator sections of the tube is minimized. The load and oscillator sections of the tube are then effectively isolated, and variations of the load impedance will have negligible effect on the frequency of oscillations, provided only that the instantaneous plate voltage is never low enough to permit the formation of a virtual cathode between screen and plate. The result is hence equivalent to a master oscillator and power amplifier combined in one tube.

Experiment shows that increasing the plate voltage of an electron-coupled oscillator causes the frequency to change in one direction, whereas increasing the screen voltage makes the frequency vary in the opposite way. Hence by obtaining the screen potential from a voltage divider, as shown in Fig. 103, and by locating the screen tap at the proper point (as determined by trial) it is possible to balance the two actions against each other and make the frequency independent of the plate-supply voltage.

It is also possible to use electron coupling with pentode tubes, as shown in Fig. 103*c*. This has the advantage of permitting operation with the cathode at ground potential with no neutralization. However, it is ordinarily no longer possible to find a ratio of screen to plate voltage that will make the frequency independent of the plate-supply voltage.

Because of the way in which it combines simplicity, isolation of the output from the frequency-controlling part of the circuit, and a frequency independent of plate-supply voltage, the electron-coupled oscillator finds general favor where frequency stability is important. Although not so stable in frequency as is the crystal oscillator, because of temperature effects, the electron-coupled oscillator permits the frequency to be continuously varied and so is used under conditions where a crystal oscillator cannot be employed.

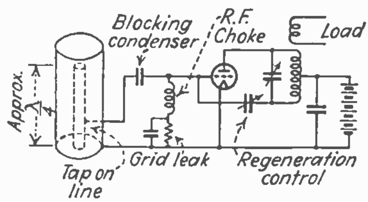


FIG. 104.—Circuit of practical vacuum-tube oscillator employing resonant line to control frequency.

Oscillators Employing Resonant Lines.—The frequency of oscillators operating at very high frequencies is often controlled by a tuned circuit supplied by a resonant transmission line. This is possible because, as explained in Sec. 19, a quarter-wave transmission line short-circuited at the receiving terminals acts as a parallel resonant circuit.

A practical oscillator circuit utilizing a resonant transmission line to control the frequency is shown in Fig. 104. This is a tuned-plate–tuned-grid type of oscillator with the grid tuned circuit supplied by the high-Q resonant line, and the plate tuned circuit supplied by a conventional coil and condenser arrangement to which the load is coupled. For maximum frequency stability, the resonant line must be loosely coupled to the grid of the tube by connecting the grid to the line at a point relatively close to the shorted end, and the adjustments are such that the resonant line has most of the reactive energy and so tends to control the frequency.

The resonant-line oscillator is able to generate large amounts of power with good frequency stability, and without the use of a power amplifier. At the same time, resonant-line oscillators are suitable only when the wave length is small enough for the lines to be of reasonable physical size (*i.e.*, when the frequency is very high).

64. Crystal Oscillators.—The frequency of an oscillator can be maintained very accurately at a fixed value by using a piezo-electric quartz crystal as the frequency-controlling element. Such a crystal acts as a mechanically resonant circuit, with the piezo-electric effect providing a connection between the mechanical vibrations and the electrical circuits.

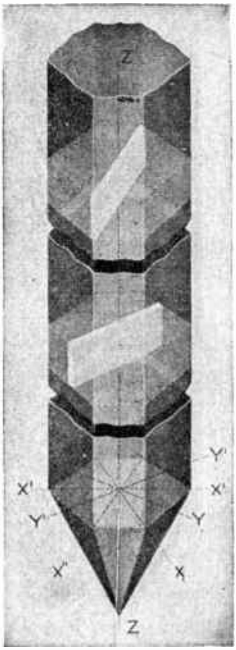


FIG. 105.—Illustration showing the natural quartz crystal and the relation of the electric or X, the mechanical or Y, and the optical or Z axes to the crystal structure. The upper section shows a Y (or 30-deg.) cut plate while the plate in the center section is X (or Curie) cut. The third Y axis, $Y''Y''$, is not shown because the perspective of the drawing makes it coincide with the ZZ axis.

Piezo-electricity.—Piezo-electric properties are exhibited by a number of crystalline substances, such as quartz, Rochelle salt, tourmaline, etc. Such crystals are characterized by the fact that when an electrostatic stress is applied in certain directions a mechanical stress is produced in other directions, and vice versa.

The most active piezo-electric substance is Rochelle salt, which is used in piezo-electric microphones and loud speakers. Quartz, though exhibiting the piezo-electric effect to a much smaller degree than Rochelle salt, is employed for frequency control in oscillators because of its permanence, low temperature coefficient, and low frictional losses. Tourmaline is similar to quartz, but is not used because of its high cost.

Piezo-electric quartz crystals have a hexagonal cross section with pointed ends as shown in Fig. 105. The axis joining the points at the ends of the crystal is known as the optical axis, and electrostatic stresses applied in this direction produce no piezo-electric effect. The three axes X', X'', and X''' passing through the corners of the hexagon that forms a section perpendicular to the optical axis are known as the electrical axes, while the three axes, Y', Y'', and Y''', which are perpendicular to the faces of the crystal, are the mechanical axes.

A mechanical stress in the direction of a Y axis produces an electrostatic stress in the direction of the X axis that is at right angles to the Y axis involved. The polarity of the electrostatic stress depends upon whether the mechanical strain is a compression or tension. Conversely, when an electrostatic stress is applied in the direction of an electrical axis, mechanical strains are produced along the mechanical axis at right angles to the electrical axis. This interconnection between mechanical and electrical properties is exhibited by practically all sections cut from a piezo-electric crystal. Thus, if mechanical forces are applied across the faces of a crystal section having its flat sides perpendicular to a Y axis, as in Fig. 106a (which is known as the Y or 30-degree cut), electrostatic

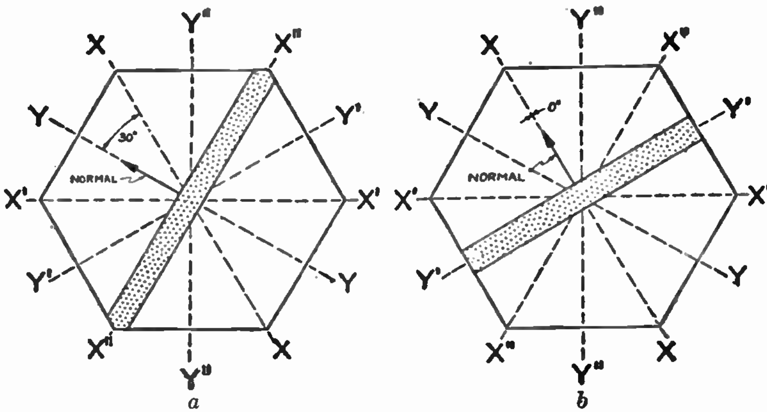


FIG. 106.—Cross sections of the quartz crystal shown in Fig. 105 taken in planes perpendicular to the optical axis ZZ. The plate at a is a Y cut (also called 30-deg. cut), because its face is perpendicular to a Y (*i.e.*, mechanical) axis, while the plate at b is an X cut plate (Curie cut), because its face is perpendicular to an X (*i.e.*, electric) axis.

stresses will be developed because the forces in such a crystal have a component across one of the Y axes.

Equivalent Electrical Circuit of Quartz Crystal.—When an alternating voltage is applied across a quartz crystal in such a direction that there is a component of electrostatic stress in the direction of an electrical axis, alternating mechanical stresses will be produced in the direction of the Y (or mechanical) axis which is perpendicular to the X axis involved. These stresses will cause the crystal to vibrate, and, if the frequency of the applied alternating voltage approximates a frequency at which mechanical resonance can exist in the crystal, the amplitude of the vibrations will be very large. In the vicinity of such a resonant frequency the current that is drawn by the crystal as a result of the vibrations is exactly the same current that would be drawn by a series circuit composed of resistance, inductance, and capacitance. In addition to this current representing the vibrational characteristics of the piezo-electric crystal, there is also a component of leading current resulting from the electrostatic capacitance of the quartz considered as an ordinary dielectric.

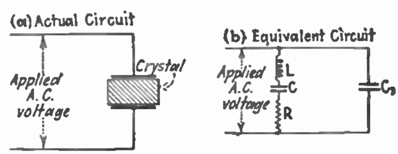


FIG. 107.—Equivalent electrical network that represents the effect which a vibrating quartz crystal has on the electrical circuits associated with it.

The action of a quartz crystal, as far as the associated electric circuits are concerned, can consequently be replaced by the equivalent electrical network shown in Fig. 107. Here C_1 represents the electrostatic capacitance between the crystal electrodes when the crystal is not vibrating, and the series combination L, C, and R represents the electrical equivalent of the vibrational characteristics of the material. The inductance L is the electrical equivalent of the crystal mass that is effective in the vibration, C is the electrical equivalent of the effective mechanical compliance, and R represents the electrical equivalent of the coefficient of friction. The frequency at which L and C are in series resonance is also the frequency of mechanical resonance. The electrical energy drawn by the equivalent L-C-R series circuit represents energy that the electrical circuit supplies to maintain the crystal vibrations. An electrical circuit involving a piezo-electric crystal can therefore be analyzed by replacing the crystal with its equivalent electrical network and then determining the behavior of the resulting circuit.

The magnitudes of L, C, R, and C_1 that enter into the equivalent electrical network of the vibrating quartz crystal depend upon the way in which the crystal is cut, the size of the crystal, and the type of vibration involved.

Values for a typical case are shown in Table XII. It will be noted that the quartz-crystal resonator is characterized by an extremely high Q and also by an unusually high ratio of equivalent inductance to equivalent capacitance.

TABLE XII.—CHARACTERISTICS OF QUARTZ CRYSTAL

Dimensions	thickness	0.636 cm
	width	3.33 cm
	length	2.75 cm

Resonant frequency (thickness vibration) = 474 kc

Equivalent electrical characteristics:

$$L = 3.66 \text{ henries}$$
$$C = 0.031\ \mu\mu\text{f}$$
$$C_1 = 5.76\ \mu\mu\text{f}$$
$$R = 4518 \text{ ohms}$$
$$Q = 2294.$$

Crystal-oscillator Circuits.—A quartz crystal can be used to control the frequency of an oscillator by so locating the crystal in the oscillator circuit that the equivalent electrical network of the crystal takes the place of a tuned circuit that would normally control the frequency. Typical arrangements of this type are shown in Fig. 108.

The circuit of Fig. 108*a* is seen to be a tuned-plate–tuned-grid arrangement, with the grid tuned circuit supplied by the crystal. In such an arrangement the amplitude of oscillations is determined by the grid-plate capacitance of the tube and the tuning of the plate resonant circuit. In some cases the triode tube of Fig. 108*a* is replaced by a pentode, as shown in Fig. 108*b*. The coupling between grid and plate circuits is then provided either by the residual grid-plate capacitance present in audio-frequency power pentodes or by a small external capacitance in the case of radio-frequency pentodes. In either case, it is possible to make the coupling much less than when triodes are used.

Another type of circuit is shown in Fig. 108*c*, where the crystal supplies the coupling between the grid and plate circuits of a radio-frequency pentode. The amplitude of oscillations in this arrangement is determined by the size of the condenser C_2 in series with the crystal, and the capacitance C_3 that supplies the plate-return to the cathode for radio frequencies. The circuit of Fig. 108*d* is an electron-coupled crystal

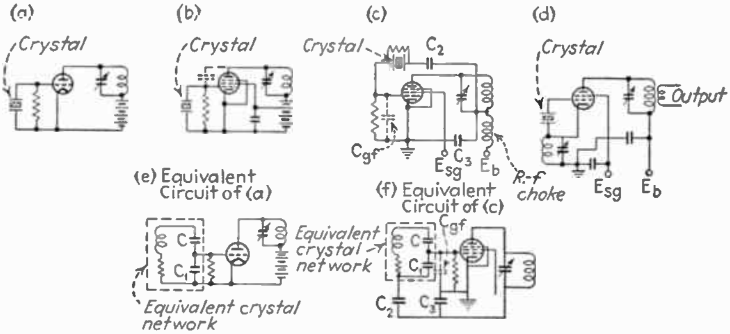


FIG. 108.—Typical crystal-oscillator circuits together with several equivalent circuits in which the crystal is represented by an electrical network.

oscillator, in which the control grid, cathode, and screen grid, function as in an ordinary crystal-oscillator circuit. The output is developed in the plate circuit, however, and is coupled to the oscillator section only through the electron stream, thereby providing isolation. In some cases the output tuned circuit is tuned to a harmonic of the crystal frequency, thereby combining crystal oscillator and harmonic generator in one tube.

The frequency of oscillation obtained with the arrangements of Fig. 108 depends slightly upon the circuit adjustments and can be varied over a very small range by such an expedient as placing a variable condenser in shunt with the crystal. The control over the crystal frequency obtainable in this way is small, however, because the small ratio of C/C_1 in the equivalent crystal circuit means that the coupling between the crystal resonator and the tube circuits is so low that the latter have very little effect.

Crystal Cuts and Modes of Oscillation.—The crystals used in oscillators are commonly in the form of plates cut from the natural crystal with orientations known as *X*, *Y*, *AT*, and *V* cuts.

The relationship between the X-cut plate and the crystal axis is shown in Fig. 105. Such a plate has two principal modes of oscillation. One of these is determined by the thickness of the crystal in the direction of the electrical axis and involves standing waves in this direction. The other frequency is determined by the width of the crystal measured along the Y axis and involves standing waves in the width direction. The temperature coefficient of the resonant frequency of both modes of vibration is approximately -20 parts in a million per degree centigrade. The circuits of Fig. 108 can be made to excite either the width or thickness vibration at will, according to the resonant frequency of the plate tuned circuit.

The relationship between a Y-cut plate and the crystal axes is shown in Fig. 105. Such a crystal has two principal resonant frequencies corresponding to those obtained with the X cut. The temperature coefficient of the width vibration of a Y-cut crystal is approximately -20 parts in a million per degree centigrade, and the thickness vibration has a temperature coefficient that may vary from -20 to $+100$ parts in a million per degree centigrade, with the exact value depending upon the operating temperature and ratio of width to thickness.

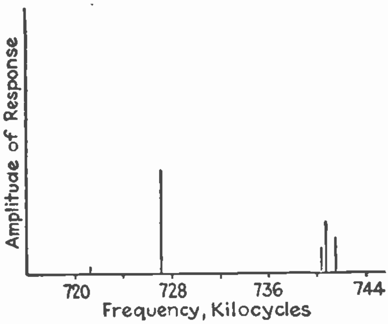


FIG. 109.—Typical frequency spectrum of thin quartz plate showing how several closely spaced resonant frequencies sometimes occur. The length of the vertical lines indicates in a rough way the relative tendency to oscillate at the different resonant frequencies.

Both X- and Y-cut crystals possess additional modes of vibration. These resonances are in some cases produced by harmonics of the width vibration and in other cases are the result of flexural and torsional vibrations and their harmonics. The resonances are more numerous the thinner the crystal, and they cause very thin crystals to have a complicated frequency spectrum such as is illustrated in Fig. 109. The increasing complexity of this frequency spectrum as the plate becomes thinner sets a practical limit to the highest frequency that it is feasible to generate in a crystal oscillator using X- and Y-cut crystal plates.

A crystal cut out of a plane rotated about the X axis, as shown in Fig. 110, has substantially zero temperature coefficient for the thickness frequency and also has a very simple frequency spectrum in the vicinity of the thickness resonant frequency. This type of crystal plate is known as the AT cut.

The AT-cut crystal is preferred to X and Y cuts for high-frequency operation. This is because the low temperature coefficient of the AT cut makes it possible to obtain constant frequency without keeping the crystal at constant temperature. Also, the very simple frequency spectrum of the *AT* cut makes it possible to use thinner crystals, and hence to go to higher frequencies, than is feasible with *X*- and *Y*-cut crystals.

The *V*-cut crystal is an orientation about the *Z* and *X′* axis. It is possible in this way to obtain a piezo-electrically active crystal having a zero temperature coefficient for the thickness vibration, and such crystals are used to a considerable extent at high frequencies.

At the lower frequencies, *X*- or *Y*-cut crystals employing width vibrations are used. The temperature coefficient of these cuts is such that the crystal must be maintained at constant temperature if the frequency stability is to be high.

Crystal Mountings.—The crystal holder must be arranged to add as little damping as possible to the crystal vibrations, and yet it should, when possible, hold the crystal rigidly in position. *X*- and *Y*-cut plates in which the width vibration is utilized can be clamped in the middle of the width direction, since this is a nodal point for the vibration. *AT*- and *Y*-cut crystals in which the thickness vibration is utilized can be clamped by spring pressure between electrodes, which are preferably cupped so that the pressure is exerted at only a few points, such as the corners of a square crystal or the periphery of circular crystals. With *X*-cut crystals using the thickness vibration, it is usually desirable to allow an air gap between the upper electrode and the crystal because clamping introduces appreciable damping.

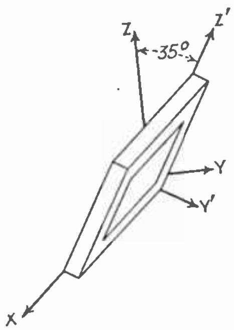


FIG. 110.—Diagram showing how crystal plane is rotated about *X* axis to obtain the *AT* cut.

The variation of the resonant frequency with temperature in *X*- and *Y*-cut crystals makes it necessary to maintain such crystals at a constant temperature if the frequency stability is to be high. This is done by placing the crystal in an electrically heated oven maintained at constant temperature by means of a thermostat. With the *AT* and *V* cuts the temperature coefficient of frequency is so low that temperature control can normally be dispensed with.

Frequency Stability of Crystal Oscillators.—The frequency of a crystal oscillator depends primarily upon the resonant frequency of the crystal, but is affected to a small extent by the constants of the associated electrical circuits. When the crystal is maintained at constant temperature, or when a zero-temperature-coefficient cut is employed, a stability of 10 to 30 parts in a million can readily be obtained over long periods of time under commercial conditions if the voltages applied to the oscillator tube are reasonably constant. When all possible precautions and refinements are taken, including such measures as maintaining the electrical circuits and tube, as well as the crystal, at constant temperature, frequency stabilities as high as 1 part in 10 million or better can be obtained.

The frequency of a crystal oscillator is determined primarily by the crystal dimensions. Hence, in order to change the frequency, it is usually necessary to employ another crystal or to regrind the original crystal if the new frequency is higher. Small changes in the frequency can be made, without grinding the crystal, by varying a capacitance shunted across the crystal holder and, in the case of X- and Y-cut crystals, by varying the crystal temperature.

Power Obtainable from Crystal Oscillators.—The power obtainable from a crystal oscillator is limited at high frequencies by heating of the crystal, and at low frequencies by the strains that the vibrations set up in the crystal structure and that will crack the crystal if the vibrations are too intense. A crystal having a large area will develop more power than a small crystal at the same resonant frequency, but the extent to which one can go in this direction is limited by the difficulty of grinding large crystals sufficiently true.

An ordinary crystal is capable of developing sufficient voltage at high frequencies to excite a tube generating approximately 50 watts of power. However, working the crystal this hard results in appreciable dissipation of power within the crystal, which causes the temperature to rise. Since X- and Y-cut crystals have a high temperature coefficient, it is very desirable to operate such crystals with relatively light load, the common practice being to control only a few watts of power. With the AT and V cuts the operation of the crystal lightly loaded is not so important, since the temperature can be allowed to rise appreciably without change in the frequency.

65. Parasitic Oscillations.—The term *parasitic* is applied to any undesired oscillation occurring in an amplifier or osciliator. Parasitic oscillations absorb power that would otherwise represent useful output and also give rise to distortion in linear amplifiers, modulated oscillators and amplifiers, and audio amplifiers.

Examples of Parasitic Oscillations.—Parasitic oscillations result from the fact that, when the tube capacitances, lead inductances, shunt-feed chokes, etc., are all taken into account, it is usually found that several possible modes of oscillation can exist in addition to the desired type of operation. The nature of these parasitic circuits and the means for preventing them from giving rise to oscillations can be understood by considering several typical examples. Consider first the simple Class C amplifier circuit of Fig. 111*a*. At very high frequencies this reduces to the circuit of Fig. 111*b*, which is a tuned-plate–tuned-grid oscillator circuit with the grid and plate tuning condensers supplied by the grid-filament and plate-filament tube capacitances, respectively. The grid inductance L_1 and plate inductance L_2 are supplied by the inductance of the leads between these two electrodes and ground through the tuning capacitances C_1 and C_2, which can be considered as short circuits at the parasitic frequency. It will be noted that the neutralization is not effective, because the coil L_p does not participate in the parasitic oscillation. Instead, the neutralizing condenser in series with the inductance of its leads merely forms a shunt across L_1 to ground through C_2, modifying the resonant frequency of the grid tuned circuit. The circuit shown at Fig. 111*b* will oscillate, provided the plate tuned circuit offers inductive reactance at the resonant frequency of the grid tuned circuit.

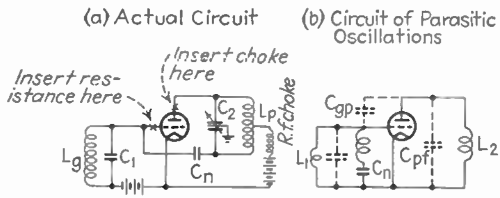


FIG. 111.—Typical Class C amplifier circuit showing how high-frequency parasitic oscillations are possible.

The remedy for such parasitic oscillations is to insert either a resistance in series with the grid or a small choke coil in series with the plate, as shown in Fig. 111*a*. Such a resistance is directly in series with the grid tuned circuit and so reduces the tendency to oscillate at the parasitic frequency. The plate choke reduces the resonant frequency of the plate tuned circuit, causing it to offer a capacitive reactance at the resonant frequency of the grid tuned circuit, and thereby produces a non-oscillatory condition. Neither grid resistance nor plate choke has appreciable effect on the desired mode of operation, because they are not in series with the resonant circuits effective for normal operation.

Next, consider the push-pull circuit of Fig. 112*a*. An arrangement of this sort is commonly troubled with a high-frequency parasitic oscillation. At this high frequency, the capacitances C_1, C_1', C_2, and C_2' are effectively short circuits. This places the two tubes in parallel and leads to the equivalent parasitic circuit of Fig. 112*b*, which is a tuned-plate–tuned-grid arrangement in which the neutralizing capacitance is effectively in parallel with the grid-plate tube capacitance and so increases the tendency to oscillate. The inductance effective in the grid tuned circuit is the inductances of the leads from grids to ground through the capacitances C_1 or C_1'. The plate-circuit inductance is similarly the lead inductance, and the tuning capacitances are supplied by the tubes. This type of parasitic oscillation can be eliminated by inserting resistances in the grid circuits next to the grids, or by detuning the plate parasitic circuit by the use of small choke coils inserted in the plate leads next to the plates.

The circuit shown in Fig. 112*a* may also develop low-frequency parasitic oscillations in which the grid and plate chokes participate. At these low frequencies the grid and plate tuning inductances L_g and L_p can be considered as short circuits, giving an equivalent circuit having the form shown at Fig. 112*c*. This is seen again to be a tuned-plate–tuned-grid circuit, with the grid and plate chokes now supplying the

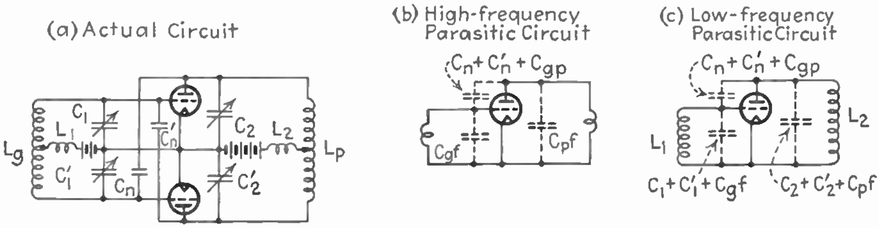


FIG. 112.—Neutralized push-pull amplifier, showing how the tubes can operate in parallel to produce a high-frequency parasitic oscillation and also a low-frequency parasitic oscillation.

tuning inductances. Since these chokes are large and are tuned by a considerable capacitance, the frequency of the corresponding parasitic oscillation is quite low. The remedy consists either in arranging the circuit proportions so that the plate tuned circuit is resonant to a lower frequency than the grid tuned circuit or, better yet, in eliminating one of the chokes.

When tubes are connected in parallel as in Fig. 113*a*, parasitic oscillations in which the two tubes operate in series can almost always be expected. Thus at high frequencies the lead inductances from grid to grid and plate to plate, in conjunction with the interelectrode capacitances of the tube, form a tuned-grid–tuned-plate circuit, as illustrated in Fig. 113*b*. This may produce parasitic oscillations unless grid resistors or detuning plate inductances are employed. The neutralization has no effect because the neutralizing condenser is between points that are at ground potential for the parasitic oscillation.

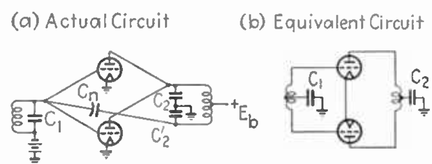


FIG. 113.—Amplifier with parallel tubes showing how the tubes may act in push-pull to produce a high-frequency parasitic oscillation.

Miscellaneous Comments on Methods of Investigating and Eliminating Parasitic Oscillations.—Parasitic oscillations can be expected as a matter of course in any new design of power amplifier or power oscillator involving large tubes. The simplest method of investigating the presence of parasitic oscillations in a Class C amplifier is to remove the exciting voltage, make the grid bias small or even zero, and operate with a plate voltage lowered sufficiently to keep within the rated dissipation of the tube. This insures a high mutual conductance and gives conditions favorable for the excitation of most types of parasitics. These oscillations can then be searched for by means of a neon lamp on a stick. When the lamp is brought near a point of high voltage, it will glow, thus making it possible to determine what parts of the circuit are involved. After a parasitic oscillation has been thus located and its frequency determined, the equivalent circuit can be deduced and remedial means devised. It is commonly found that, upon the elimination of one parasitic oscillation, another oscillation of a different type will appear, and that upon the elimination of this still other parasitic oscillations may start up.

Screen-grid and pentode power amplifiers are practically immune from most types of parasitic oscillations, provided that the screen grid is really at ground potential, since no energy transfer through the tube is then possible. However, the circuit from the screen-grid electrode through the by-pass condenser to cathode has inductance, and, if this circuit is not made as short as possible, the voltage drop in this inductance will be sufficient to prevent the screen from being at the cathode potential. Feedback through the tube with the possibility of parasitic action will then be present.

Troubles from parasitic oscillations can be minimized by using simple circuits with short leads. It is often helpful in this connection to make the grid leads as short as possible and then use somewhat longer plate leads, since the extra inductance of the plate leads then gives much the same effect as a small plate choke. Neutralization systems that are symmetrical and so maintain the neutralization over a wide frequency band are also helpful. It is also desirable to employ inductive coupling to the input and output circuits of the tube rather than capacitive or tapped inductance coupling. Shunt-feed chokes should be avoided when feasible and should not be employed in both grid and plate circuits of the same tube.

66. Oscillators Employing Electron Oscillations. *Barkhausen Oscillators.*—Frequencies higher than those obtainable by the use of conventional oscillators can be generated by means of electron oscillators, either of the Barkhausen or of the magnetron type. The nature of the Barkhausen electron oscillator can be understood by reference to Fig. 114, which shows a triode operated with the grid at a high positive potential and the plate at a slight negative potential. Electrons emitted from the cathode under these circumstances are attracted toward the grid, but most of them pass through the spaces between the grid wires into the grid-plate space, where they slow down and ultimately stop just before reaching the plate. The electrons are then drawn back toward the grid, but, if they are not captured by the grid wires, they will pass on into the grid-cathode space and slow down upon approaching the filament. This oscillation about the grid may be repeated over and over before the electron is ultimately removed from the tube by a chance impact against the grid wires. Several such electron oscillations about the positive grid are illustrated schematically in Fig. 114*b*.

The frequency of the electron oscillation is determined primarily by the grid potential and the dimensions of the tube, but space charges, the external tuned circuit, the negative potential of the plate, etc., may alter the frequency by perhaps 30 to 50 per cent. With cylindrical electrodes

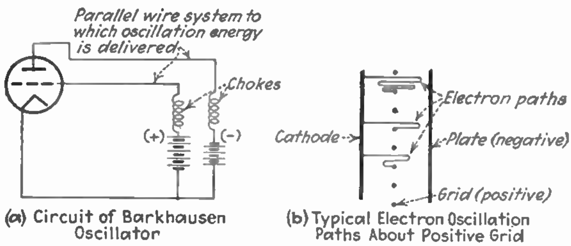


FIG. 114.—Circuit for generating Barkhausen oscillations, together with sketch showing how electron oscillations take place about a positive grid.

and equal grid-cathode and grid-plate spacings, the wave length is given approximately by the relation

$$\text{Wave length in centimeters} = \frac{670d}{\sqrt{E_g}} \tag{95}$$

where d is the diameter of the plate in centimeters and E_g is the grid voltage. It is thus apparent that the higher the frequency to be generated, the greater must be the grid potential and the smaller the dimensions of the tube.

The efficiency of an electron oscillator is low because of the inherent nature of the generating process, with values of 2 or 3 per cent being typical under favorable conditions. The output power obtainable is also low because of the low efficiency and because, with the small dimensions required to generate high frequencies, the allowable power loss at the grid (which must absorb most of the tube loss) is relatively small. For best output and efficiency it is necessary to adjust carefully the grid voltage, negative plate voltage, and cathode temperature in relation to the resonant frequency of the external circuit.

The resonant circuits used at the extremely high frequencies that can be generated by means of electron oscillations ordinarily consist of resonant lines arranged in the form of parallel bars having a length such as to give resonance at the desired frequency.

Electron Oscillators Employing Magnetron Tubes.—Electron oscillations can also be produced by employing a split-anode magnetron tube in the circuit of Fig. 115*a*. Such a tube consists of an axial filament surrounded by a cylindrical anode that is split into two halves as shown in the figure. When the tube is placed in an axial direct-current magnetic field of suitable strength, the electrons that are attracted to the plate are deflected by the magnetic field in such a manner as to follow a curved path, as shown in Fig. 115*b*. This gives rise to electron oscillations having a frequency corresponding to the time required by an electron to make one complete revolution.

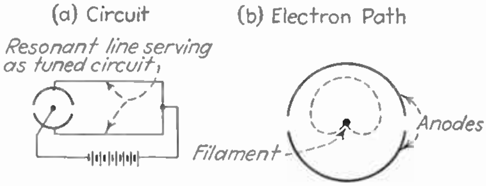


FIG. 115.—Circuit of electron oscillator of split-anode magnetron type together with electron path. The tube is placed in a magnetic field that is approximately axial.

The magnetic field required to give best results in an electron oscillator of the magnetron type is that which makes the electrons just miss the plate surface. This value is given approximately by the relation

$$\text{Magnetic field in lines per square centimeter} = \frac{6.72}{r}\sqrt{E} \tag{96}$$

where r is the anode radius in centimeters and E is the anode voltage. When this magnetic field is employed, the time of flight of an electron is such that the wave length of the resulting oscillation is given approximately by the equation

$$\text{Wave length in centimeters} = \frac{1930r}{\sqrt{E}} \tag{97}$$

Examination of Eq. (97) shows that, in order to obtain very high frequencies, the anode must have a small radius and high electrode voltages must be employed.

The efficiency of electron oscillators of the split-anode magnetron type is relatively low, values of 5 to 10 per cent being typical under favorable conditions. The power output obtainable is likewise low, particularly at high frequencies, because of the poor efficiency and the fact that the small physical dimensions required to generate energy at high frequencies limit the permissible power dissipation. The best efficiency and power output are obtained when the anode voltage, the filament voltage, and the strength of the magnetic field are properly adjusted in relation to the resonant frequency of the external circuit, and the tube is tilted slightly with respect to the magnetic field.

As compared with the Barkhausen oscillator, the split-anode magnetron is capable of operating satisfactorily at higher frequencies because of its simpler structure and great ability to dissipate energy. The magnetron also has higher efficiency and hence more output.

67. The Multivibrator (or Relaxation Oscillator).—The multivibrator is a two-stage resistance-coupled amplifier in which the voltage developed

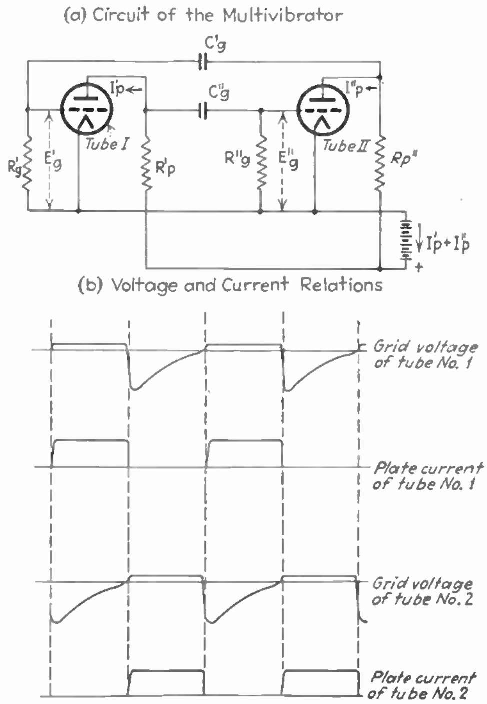


FIG. 116.—Circuit of multivibrator together with oscillograms showing the way in which the instantaneous grid potential and plate currents vary during the cycle of operation.

by the output of the second tube is applied to the input of the first tube as shown in Fig. 116. Such an arrangement will oscillate because each tube produces a phase shift of 180°, thereby causing the output of the second tube to supply to the first tube an input voltage that has exactly the right phase to sustain oscillations.

The operation of the multivibrator can be understood by reference to the oscillograms shown in Fig. 116*b*. Oscillations are started by an infinitesimal voltage at the grid of one of the tubes, say a positive potential on the grid of tube 1. This voltage is amplified by the two tubes and reappears at the grid of the first tube to be reamplified. This action takes place almost instantly and is repeated over and over, so that the grid potential of tube 1 rises suddenly to a positive value, while the grid potential of tube 2 just as suddenly becomes more negative than cut-off. The amplification then ceases, and for the moment one tube is drawing a heavy plate current while the other tube takes no plate current. This situation is not permanent, however, because the leakage through the grid-leak resistances gradually brings the grid potentials back toward normal. When this leakage has reached the point where amplification is just on the verge of being possible, some infinitesimal voltage will change the potentials enough to start the amplification process in the reverse direction, *i.e.*, the grid of tube 1 will suddenly become negative and the grid of tube 2 positive. This action is exactly the same as the initial action except that the relative functions of the two tubes have been interchanged. Next the potentials on the two grids gradually die away as a result of the action of the grid leaks, just as before, and finally reach a point at which the cycle repeats.

The frequency of the multivibrator oscillation is determined primarily by the grid-leak resistance and grid-condenser capacitance, but it is also influenced by the remaining circuit constants, the tube characteristics, and the electrode voltages. The multivibrator can be adjusted to generate frequencies ranging anywhere from perhaps 1 cycle per minute to the highest frequency at which satisfactory resistance-coupled amplification is possible. The usefulness of the multivibrator arises from the fact that the wave that is generated is very rich in harmonics, and from the fact that the frequency of oscillation is readily controlled. Thus injection of an alternating voltage into the multivibrator circuit tends to cause the multivibrator to adjust itself to a frequency that is m/n of the injected frequency, where m and n are integers. It is possible in this way to use the multivibrator to reduce frequency (*i.e.*, to generate a subharmonic of the injected frequency), and it is entirely practicable to maintain the control rigidly when n is as large as 10 and m is 1.

Problems

1. In the Hartley and Colpitts oscillator circuits of Fig. 101, follow through the detailed phase relations and show that conditions necessary for oscillation can be realized.

2. In the Colpitts circuit of Fig. 101 explain why it is necessary for the grid-leak resistance to go from grid to cathode as shown, instead of being in series with the grid as in the other circuits.

3. When the electrode voltages are applied to a power oscillator, the oscillations start with an infinitesimal amplitude and build up to an equilibrium value. Explain what stops this building-up process from continuing indefinitely.

4. Design the circuits for a power oscillator using a shunt-feed Hartley circuit, and the Type 852 tube (Table X, page 158) at a plate potential of 3000 volts. The frequency is 5000 kc, a tank-circuit Q of 30 is desired, and the grid and plate connec-

tions are to be made to the two ends of the tank circuit. The design includes determination of tank-circuit inductance and capacitance, grid-leak resistance, a suitable grid-condenser capacitance, a suitable inductance for the shunt-feed choke, and approximate location of cathode tap to the tank circuit. Neglect the shunt-feed choke when determining tank-circuit proportions.

5. In a properly adjusted oscillator, increasing the load resistance normally coupled into the tank circuit will increase the d-c plate current very greatly, while causing the d-c grid current and the a-c current in the tank circuit to decrease slightly. Explain.

6. In a power oscillator it is desired to increase the efficiency of operation by reducing the fraction of the cycle during which plate current flows. What adjustments would be made?

7. In a Hartley oscillator adjusted for good efficiency it is found that when the resonant frequency of the tank circuit is changed by varying the tank-circuit capacitance while leaving all other adjustments fixed, the tank-circuit current is inversely proportional to frequency. Explain why this is.

8. A tuned circuit that has its resonant frequency varied by means of a variable condenser is loosely coupled to the tank circuit of an oscillator. The condenser adjustment that makes the resonant frequency of the circuit the same as the frequency of oscillations is indicated by a slight jump in the d-c plate current of the oscillator. Explain.

9. *a.* What sort of behavior in a power oscillator would lead one to suspect that the minimum plate potential was too large?

b. What adjustments could be made in the connections between tank circuit and tube that would reduce the minimum plate potential?

10. It is sometimes said that the frequency stability of an oscillator is improved by increasing the reactive energy circulating in the tank circuit in proportion to the power output. Demonstrate that this is equivalent to saying that an increase in the tank-circuit Q improves the frequency stability.

11. In a master-oscillator power-amplifier arrangement, explain how changes in load conditions in the plate circuit of the power amplifier can affect the oscillator to some extent as long as the grid of the amplifier is driven positive.

12. In an electron-coupled oscillator explain why the frequency ceases to be independent of the load impedance in the plate circuit when this impedance is great enough to make the minimum plate voltage very low.

13. *a.* By means of the equivalent electrical circuit of the quartz crystal, demonstrate that when a crystal is used in the oscillator circuit of Fig. 108*a* the frequency of oscillation can be varied slightly by means of a variable condenser in shunt with the crystal.

b. Calculate the change produced in resonant frequency of the crystal circuit LCC_1, in Fig. 107*b* when a capacitance of 10 $\mu\mu$f is shunted across the crystal of Table XII.

14. In the crystal oscillator of Fig. 108*a* it is found that the following behavior is observed as the capacitance of the plate tuned circuit is increased from its minimum. First there are no oscillations, followed by oscillations that gradually increase in intensity and reach a maximum just before resonance with the crystal oscillations is reached. The oscillations then suddenly stop as the plate tuned circuit becomes resonant at the oscillation frequency and do not reappear with further increase in capacitance. Explain this behavior.

15. Explain how parasitic oscillations can exist in a triode amplifier employing the Rice system of neutralization (see Fig. 74*b*).

16. Is it possible for parasitic oscillations to exist in a tuned-grid–tuned-plate oscillator? Explain.

17. In a particular power amplifier it is found that parasitic oscillations can be stopped by making the plate lead long, whereas a long grid lead makes the parasitics worse. Explain.

18. Explain qualitatively the physical reason why the wave length of a Barkhausen oscillation varies with d and E_g as given by Eq. (95).

19. In an electron oscillator of the magnetron type, explain why the axial length of the tube in the direction of the filament has no effect on the frequency of oscillation even though a longer tube has more interelectrode capacitance.

20. In a multivibrator oscillator the frequency is reduced by making the grid-leak resistance or the coupling-condenser capacitance greater. Explain the reason for this behavior.

CHAPTER VIII

MODULATION

68. Waves with Amplitude Modulation.—In all the commonly used systems of radio communication the information is transmitted by varying the amplitude of the radiated waves, as explained in Sec. 5 and illustrated in Figs. 2 and 3. Communication carried on in this way is said to take place by means of *amplitude modulation*, and the equation of the envelope of the modulated wave represents the equation of the information actually transmitted. If the envelope variations of the modulated wave exactly reproduce the original variations in sound pressure, light intensity, etc., these variations are transmitted without distortion. Otherwise the information contained in the modulation envelope is not exactly the same as the original.

Analysis of Modulated Wave.—It was shown in Sec. 3 that a modulated wave consists of a carrier wave and a number of side-band components. The carrier wave represents the average amplitude of the envelope. Since this is the same irrespective of the presence or absence of modulation, the carrier wave transmits no information. The information is carried by the side-band frequencies, which can be thought of as being generated by the process of varying the amplitude of the radio-frequency wave.

The nature of a modulated wave can be determined by writing down the equation of the envelope. The average value that results is the carrier, and each sinusoidal component of the envelope gives rise to a pair of side-band frequencies that are respectively greater and less than the carrier frequency by the corresponding envelope frequency. The amplitude of each side-band component is one-half the corresponding envelope component. Thus a modulated wave having an envelope given by the equation

$$\text{Envelope amplitude} = 100 + 50 \sin 2\pi f_1 + 20 \sin 2\pi f_2$$

consists of a carrier wave having a crest value of 100 volts, a side band having a crest amplitude of 25 volts and a frequency that is f_1 cycles greater than the carrier frequency, a companion side band of the same amplitude but of frequency f_1 cycles less than the carrier frequency, and a second pair of side-band components, each of 10 volts amplitude and having frequencies f_2 cycles more and f_2 cycles less than the carrier frequency.

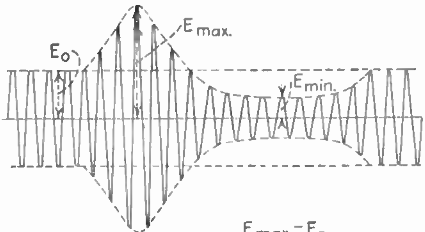
Side Bands Required in Telegraph and Telephone Transmission.—In the transmission of telegraph signals by the Continental Morse Code it is possible to operate telegraph relays provided each side band has a width of 0.131 cycle per letter transmitted per minute.[1] Thus transmission at the rate of 100 letters per minute can be carried on using side bands which extend only 13.1 cycles on each side of the carrier frequency.

In the transmission of speech and music of good quality, side-band components extending from 5000 to 8000 cycles on each side of the carrier frequency must be employed. Such a band provides for the transmission of audio-frequency sounds having pitches up to 5000 to 8000 cycles. While frequencies up to approximately 15,000 cycles are audible, these higher pitch sounds are not essential for reasonably satisfactory results. Understandable speech requires the reproduction of all frequencies from about 250 to 2700 cycles, or side-band frequencies ranging from 250 to 2700 cycles above and below the carrier frequency.

Degree of Modulation.—The extent of the amplitude variations in a modulated wave is expressed in terms of the degree of modulation. For sinusoidal variation, as illustrated by Fig. 3,

$$\left.\begin{matrix}\text{Degree of}\\ \text{modulation}\end{matrix}\right\} = m = \frac{\left\{\begin{matrix}\text{average envelope}\\ \text{amplitude}\end{matrix}\right\} - \left\{\begin{matrix}\text{minimum envelope}\\ \text{amplitude}\end{matrix}\right\}}{\text{average envelope amplitude}} \tag{98}$$

When the envelope variation is not sinusoidal, it is necessary to define the degree of modulation separately for the peaks and troughs of the envelope, according to the equations given in Fig. 117.



$$\text{Positive-peak modulation} = \frac{E_{max.} - E_0}{E_0}$$

$$\text{Negative-peak (or trough) modulation} = \frac{E_0 - E_{min.}}{E_0}$$

FIG. 117.—Unsymmetrically modulated wave and equations giving peak and trough modulation.

The degree of modulation is also sometimes expressed as a percentage. Thus $m = 0.50$ corresponds to 50 per cent modulation. When the degree of modulation is 1.0 (or 100 per cent) as in Fig. 3*c*, the amplitude variations carry the envelope amplitude to zero during the troughs of the modulation cycle. The modulation is then complete, *i.e.*, the envelope variations have their maximum possible value.

69. Plate-modulated Class C Amplifiers.—A plate-modulated amplifier consists of an ordinary Class C amplifier in which the plate voltage is varied in accordance with the desired modulation envelope. When properly adjusted, such an arrangement will develop an output voltage almost exactly proportional to the plate voltage and so will reproduce the

[1] This figure represents the minimum side band that can be used. Under ordinary conditions a band width several times as large is required.

desired modulation. The variation in plate-supply voltage is obtained by coupling a power amplifier tube, or *modulator*, to the plate circuit of the Class C tube, as illustrated in Fig. 118.

Factors Influencing Distortion in Plate-modulated Class C Amplifiers.—The radio-frequency amplifier of a plate-modulated arrangement is designed, adjusted, and operated in much the same manner as an ordinary Class C amplifier. Substantially distortionless modulation is obtained by adjusting the exciting voltage and tank-circuit impedance so that at the crest of the modulation cycle, when the voltage applied to the plate is twice the direct-current plate voltage, the minimum instantaneous

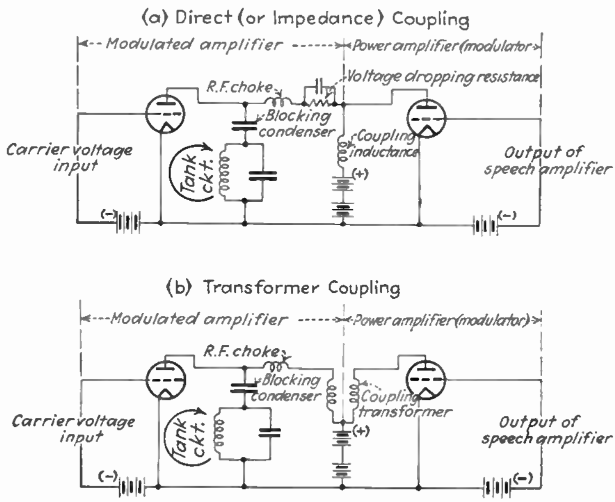


FIG. 118.—Circuit diagrams for plate-modulated Class C amplifiers. For the sake of simplicity the neutralizing arrangement for the modulated amplifier is omitted.

plate potential E_{min}, reached during the cycle of radio-frequency voltage, will be small. During the remainder of the modulation cycle the exciting voltage will then be more than adequate to make the minimum plate potential small, and the radio-frequency voltage developed in the output will follow the plate voltage very closely, as shown by curve *b* in Fig. 119*a*. It will be noted that this adjustment corresponds to somewhat greater exciting voltage in proportion to load impedance than would be used if the tube were adjusted as a simple Class C amplifier for a plate potential corresponding to the direct-current plate voltage. If this were not so, there would be a tendency for the output to flatten off at the peaks of the modulation cycle, as shown by curve *c* of Fig. 119*a*.

The linearity of modulation can be improved somewhat by exciting the Class C tube by a radio-frequency voltage that has rather poor regulation. In this way the reduction in grid current that occurs at the crest of the modulation cycle because of increased minimum plate voltage causes the exciting voltage to increase just when more excitation is needed to help carry the peaks and prevent a flattening off such as exhibited by curve c in Fig. 119a. The use of a grid-leak bias for the Class C tube

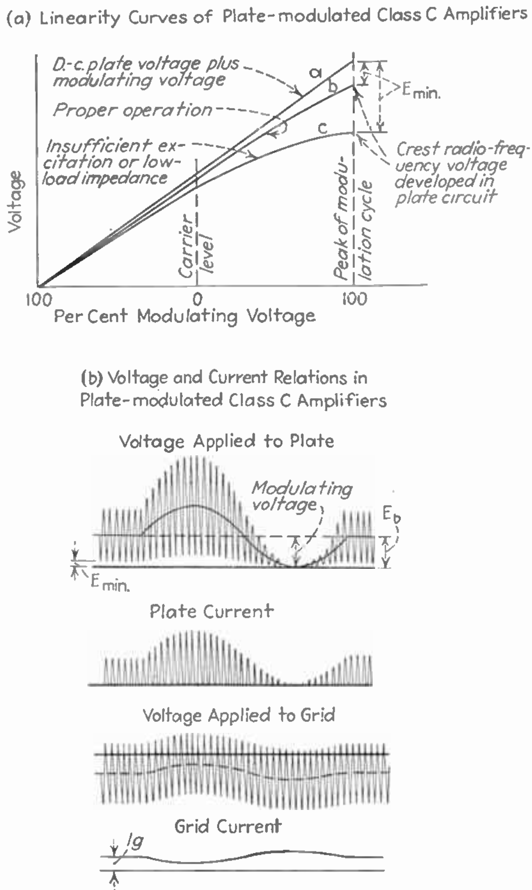


FIG. 119.—Linearity curves and voltage and current relations of plate-modulated Class C amplifier. Note that to prevent flattening of the output at the modulation peaks the adjustment should be such that the minimum instantaneous plate potential E_{min} is small at the crest of modulation. The voltage and current relations show that the increase of exciting voltage at the peaks of modulation caused by poor regulation, and the simultaneous reduction of grid bias that occurs with grid-leak bias, both help to prevent flattening of the modulation peaks.

contributes to the same end by reducing the bias at the modulation peaks and hence increasing the amount the grid is driven positive at this part of the modulation cycle.

Oscillograms showing the voltage and current relations in a properly adjusted plate-modulated Class C amplifier are shown in Fig. 119b.

Power and Modulator Considerations.—When a plate-modulated Class C amplifier is properly adjusted, the plate current of the modulated tube is almost exactly proportional to the voltage that is applied to the plate. To the extent that this is true, the Class C tube offers to the modulating voltage a load resistance that is equal to the ratio E_b/I_b, where E_b is the plate-supply voltage and I_b is the d-c plate current in the absence of modulation. The degree of modulation is determined by the alternating voltage developed by the modulator across this load impedance. For 100 per cent modulation the crest modulating voltage must equal the plate-supply voltage, and this makes the modulating power equal to $(E_bI_b)/2$, or exactly one-half the direct-current power which the source of direct-current plate voltage is called upon to supply. For lesser degrees of modulation the modulating power will be proportional to the square of the degree of modulation. *The power relations that exist in a plate-modulated Class C amplifier can hence be summarized by stating that the power required to generate the carrier wave is supplied from the direct-current plate voltage supply, whereas the power required to generate the side-band components of the modulated wave must be supplied from the output of the modulator.*

The power that the modulator must deliver becomes large in the case of high-power modulated Class C amplifiers. Thus if a 10-kw carrier is to be modulated 100 per cent and the plate efficiency is 66.7 per cent, the direct-current plate power required is 15 kw, and the modulator power output is half of this or 7.5 kw. If the degree of modulation is less than 100 per cent, the demand made upon the modulator is correspondingly less, but so is the side-band energy generated.

The modulator may be a Class A, Class AB, or Class B power amplifier. Transformer coupling, as illustrated in Fig. 118*b* is used with Class AB, Class B, and push-pull Class A power amplifiers, whereas either transformer coupling or the direct-coupled circuit of Fig. 118*a* may be employed when a single-ended Class A modulator is to be used. If 100 per cent modulation is to be obtained with direct coupling, it is necessary to place a voltage-dropping resistance in series with the plate of the Class C tube so that the direct-current plate potential of this tube will be only 60 to 80 per cent of the voltage at the plate of the modulator. This is because the crest voltage that can be developed in the plate circuit of a Class A power tube does not exceed 60 to 80 per cent of the plate-supply voltage. The voltage-dropping resistance must be by-passed by a condenser having negligible reactance at the modulating frequency.

The plate efficiency of the Class C tube in a plate-modulated amplifier is the same as in any other Class C amplifier and so is of the order of 60 to 80 per cent. The plate efficiency of the modulator depends upon the type of power amplifier employed and will range from perhaps 30 per cent in the case of Class A operation to perhaps 60 per cent with Class B.

With Class A and Class AB modulators, the low plate efficiency makes it necessary to install more tube capacity for the modulator than for the Class C amplifier. Even with Class B modulators it is necessary to provide nearly as much allowable plate dissipation in the modulator tubes as in the Class C amplifier.

Tubes and Tube Operating Conditions.—Plate-modulated amplifiers use the same triode tubes commonly employed in triode Class C amplifiers. In selecting operating conditions for the tubes it is necessary to keep in mind, however, that the plate potential, plate current, and plate dissipation are greater at the crest of the modulation cycle than for unmodulated conditions. Hence, the rated operating conditions based upon unmodulated conditions must necessarily correspond to rather conservative operation for the same tube used as a simple Class C amplifier.

Pentode and screen-grid tubes are sometimes plate modulated, although the linearity of modulation tends to be poor. When such tubes are employed it is helpful to modulate the screen voltage as well as the plate voltage.

Practical Design and Adjustment of Plate-modulated Amplifiers.—In laying out a plate-modulated Class C amplifier one normally has available the manufacturer's data on typical operating conditions for the unmodulated conditions of the Class C tube, as given in Table XIII. This simplifies the problem of determining what tube should be used to develop the required carrier, and makes it possible to design the tank circuit and other circuit details, following the procedure outlined for Class C amplifiers in Sec. 57. A knowledge of the approximate operating conditions for the Class C tube also gives the modulating power required and the load impedance to which the modulator must deliver this power. This makes it possible to select suitable modulator tubes and to design the coupling system between the Class C amplifier and the modulator.

In placing a plate-modulated Class C amplifier in actual operation, the first step consists in adjusting the Class C tube so that it realizes the operating conditions specified by the tube manufacturer as suitable for the unmodulated intervals. This is done exactly as in the case of an ordinary Class C amplifier. A modulating voltage is then applied and the resulting linearity of modulation determined either by a cathode-ray tube as described below, or by a modulation meter. If the modulation fails to be linear up to complete modulation as a result of flattening of the positive peaks, the probable cause is either insufficient radio-frequency exciting voltage for the Class C tube, or a tank circuit with too low an impedance, as explained in connection with curve *c* of Fig. 119*a*. With proper adjustment of the Class C tube, the only other causes of non-linearity are distortion in the modulator tube, and feedback of radio-frequency energy from the modulated output of the Class C tube to a lower-level point where there is no modulation. Feedback gives trouble because the modulation of the output wave causes the amount of feedback

TABLE XIII.—CHARACTERISTICS OF REPRESENTATIVE TUBES USED IN PLATE-MODULATED CLASS C SERVICE

Type	Filament			Method of cooling	μ	Allowable plate loss, watts	Typical operating conditions for unmodulated intervals						
	E_f volts	I_f amp.	Type				Plate voltage	Control-grid voltage	Peak signal voltage	Plate current, ma	Grid current, ma	Driving power, watts	Output, watts
800	7.5	3.25	Thoriated	Air	15	23	1000	− 200	325	70	15	4	50
211	10	3.25	Thoriated	Air	12	67	1000	− 260	410	150	35	14	100
204A	11	3.85	Thoriated	Air	23	167	2000	− 250	500	250	35	20	350
							1500	− 200	450	250	35	20	225
851	11	15.5	Thoriated	Air	20.5	500	2000	− 300	525	850	125	65	1250
							1500	− 250	475	900	150	75	900
207	22	52	Tungsten	Water	20	6600	6000	−1200	1860	760	150	280	3500
							10,000	−2000	2660	750	70	185	6000
862	33	207	Tungsten	Water and forced air.	48	50,000	8000	− 700	1700	4000	1000	1700	24,000
							12,000	− 800	2000	5000	1000	2000	45,000

to vary during the modulation cycle. Distortion in the modulator tube may affect either the positive or negative peaks of modulation, or both. Modulator trouble can be checked by observing the wave shape of the voltage applied to the plate of the Class C tube, using a cathode-ray oscillograph.

70. Grid-modulated Class C Amplifiers.—In this method of modulation the output of a Class C amplifier is controlled by varying the grid bias of the tube. A suitable circuit arrangement is shown in Fig. 120*a* and consists of an ordinary Class C amplifier in which the effective bias voltage consists of a direct-current component upon which is superimposed the alternating modulating voltage. With proper circuit proportions the radio-frequency output voltage of such an arrangement can be made to vary almost linearly with changes in the effective bias.

For proper operation of a grid-modulated amplifier, the conditions at the crest of the modulation cycle should correspond to ordinary Class C operation, with a minimum instantaneous plate potential E_{min} preferably somewhat smaller than customary in Class C amplifiers. The only other special consideration is that it is desirable to operate with less grid current than is customary with Class C amplifier operation. Grid current tends to make the radio-frequency exciting voltage drop off at the peaks of modulation and also flattens the positive peaks of the alternating-current modulating voltage. These actions cooperate to flatten the positive peaks of modulation and represent the principal cause of nonlinearity in the modulation process, provided the conditions at the crest of the modulation cycle make the instantaneous minimum plate potential E_{min} small.[1]

In laying out a grid-modulated Class C amplifier, a tube is chosen such that the normal Class C amplifier rating is at least four times the desired carrier power. This is so that the tube will be able to handle the crest of the modulation cycle, when the power output is four times the carrier power. The tube is then adjusted for conditions at the crest of the modulation cycle as described in Sec. 57, in the same manner as any other Class C amplifier, using data supplied by the tube manufacturer as a guide. In doing this it is desirable to favor operation with the grid driven no farther positive than absolutely necessary, and to have the highest tank-circuit impedance possible with reasonable power output and low grid current.

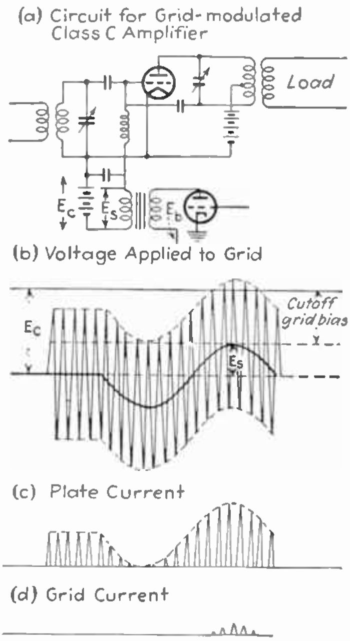


FIG. 120.—Circuit of grid-modulated Class C amplifier, together with oscillograms showing details of operation.

The crest modulating voltage and direct-current grid bias can be determined from the grid bias corresponding to the crest of the modulation cycle and the grid bias required to make the radio-frequency output voltage approach zero. The latter can be determined experimentally by making the grid more negative without changing the excitation, until the

[1] In cases where low distortion is very important, as in broadcast transmitters, the grid is sometimes not driven positive at all. This improves the linearity of modulation but results in very low power output from the tube in comparison with the output obtained when the grid is driven moderately positive.

output drops to zero, or can be calculated.[1] The proper direct-current bias is the average of the bias voltages for these two conditions, and the crest modulating voltage is half the difference of the two bias voltages.

Final adjustment of a grid-modulated Class C amplifier is obtained by applying the proper alternating-current modulating voltage and observing the linearity of the resulting modulation, using a modulation meter or cathode-ray tube. If the output voltage varies with modulation as shown at *b* in Fig. 121, either the minimum plate potential E_{min} at the crest of the modulation cycles is too large, in which case the effective impedance offered by the tank-circuit impedance should be increased, or there is distortion in the source of modulating voltage. On the other hand, flattening of the positive peaks of modulation, as indicated by *c* in Fig. 121, is ordinarily caused by grid current at the peak of the modula-

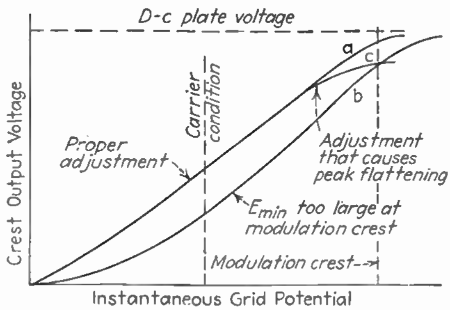


FIG. 121.—Types of non-linearity that can occur in a grid-modulated amplifier.

tion cycle reducing the exciting voltage and flattening the positive peak of the modulating voltage. The remedy for this is either operation so there is less grid current at the modulation peaks or better regulation of radio-frequency exciting voltage and modulating voltage, or both.

The grid-modulated Class C amplifier requires very little modulating power but has the disadvantage of low average plate efficiency. This is because the crest alternating voltage between cathode and plate must have a value less than half the plate-supply voltage during unmodulated intervals if this alternating voltage is still to be less than the plate-supply potential at the crest of the modulation cycle. As a consequence, the plate efficiency during unmodulated conditions is only half as great as for an ordinary Class C amplifier, or between 30 and 40 per cent. The

[1] The bias required to cut off the output with triode tubes can be calculated by the formula

$$\text{Bias voltage for reducing the output of triode to zero}\} = \left(\frac{E_b}{\mu} + E_s\right)$$

where E_s is the crest exciting voltage, E_b is the plate-supply potential, and μ is the amplification factor of the tube.

linearity of a grid-modulated Class C amplifier is also poorer than for the plate-modulated Class C amplifier unless the adjustment is such that very little output power is obtained.

High-efficiency Grid-modulated Amplifier.—The poor plate efficiency of the ordinary grid-modulated amplifier can be overcome by the arrangement of Fig. 122.[1] This is similar to the high-efficiency linear amplifier circuit of Fig. 99 except that the tubes are now excited by an unmodulated carrier wave upon which modulating voltages are superimposed. Amplifier tube A_1 is operated as an ordinary grid-modulated tube, adjusted so that with no modulation and the second tube inoperative the minimum plate potential will be small. The second tube A_2 is biased so that it delivers output only on the positive half cycles of modulation. The modulating voltage is applied to the grids of the two tubes in the same

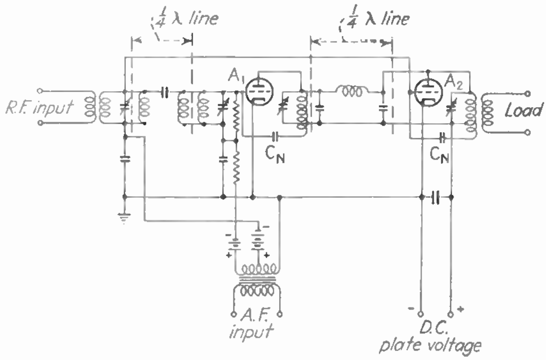


FIG. 122.—Schematic circuit diagram of high-efficiency grid-modulated amplifier.

phase. With high-mu triodes, screen-grid tubes, and pentodes, the same modulating voltage is required for both grids, whereas with triodes having moderate or low amplification factors the grid of tube A_2 requires less modulating voltage than does tube A_1. In any case, the adjustments should be such that at the positive crests of the modulation cycle the two tubes are developing substantially equal outputs.

The efficiency of the arrangement of Fig. 122 is high for the same reason that the linear amplifier of Fig. 99 has high efficiency. Under practical conditions the efficiency that can be realized is 60 to 80 per cent, which is nearly twice that for an ordinary grid-modulated amplifier and about 10 per cent higher than obtainable for the high-efficiency amplifier of Fig. 99.

71. Miscellaneous Modulation Methods. *Suppressor-grid Modulation.*—This method of modulation makes use of a pentode tube adjusted to operate as a Class C amplifier. Modulation is accomplished by apply-

[1] This system of high-efficiency grid modulation was worked out by the author with the aid of Mr. John R. Woodyard, graduate student at Stanford University.

ing to the suppressor grid the modulating voltage superimposed upon a suitable negative bias, as illustrated in Fig. 123. This arrangement takes advantage of the fact that as the suppressor grid is made negative, a virtual cathode forms in front of the suppressor. The combination of virtual cathode, suppressor, and plate then functions as a triode, as illustrated in Fig. 42. The result is that variation of the suppressor voltage will control the radio-frequency output in much the same manner that grid modulation is accomplished.

The suppressor-grid-modulated amplifier is similar to the control-grid-modulated amplifier with respect to modulating power required, efficiency, and output power obtainable. The linearity of modulation is usually somewhat poorer, however.

The high-efficiency modulation system of Fig. 122 can be readily adapted to suppressor-grid modulation by making obvious modifications.

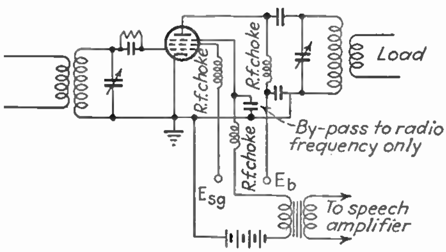


FIG. 123.—Circuit of suppressor-grid-modulated amplifier.

This avoids the poor efficiency and small power output of suppressor-grid modulation, while retaining the advantage of low modulating power.

The van der Bijl Modulated Amplifier.—This modulated amplifier consists of an ordinary power amplifier, to the grid of which are applied a small radio-frequency carrier voltage and a large modulating voltage. Because of the curvature of the plate-current–grid-voltage characteristic, the amplification of the small carrier voltage depends upon the amplitude of the modulating voltage, thus causing the amplified output to be the desired modulated wave. The detailed mechanism by which the modulation is produced is shown in Fig. 124.

Plate-modulated Oscillator.—In this method of modulation advantage is taken of the fact that, in an oscillator adjusted to operate at good efficiency, the alternating voltage developed across the tank circuit is almost exactly proportional to the plate potential as explained in Sec. 62 and illustrated in Fig. 125. It is hence possible to modulate the output of such an oscillator by superimposing the modulating voltage upon the plate-supply potential as in the case of plate-modulated Class C amplifiers.

The power relations existing in a plate-modulated oscillator are similar to those of the plate-modulated Class C amplifier. The source of d-c plate voltage is hence called upon to supply the power for generating the carrier, and the modulator must deliver sufficient power to generate the

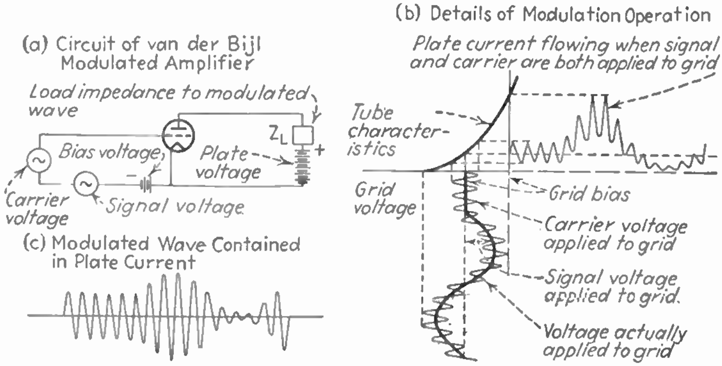


FIG. 124.—Circuit of van der Bijl type of modulated amplifier, together with oscillograms showing details of operation. The curvature of the grid-voltage–plate-current tube characteristic causes the amplification of the carrier to depend upon the grid potential, which in turn varies in accordance with the modulating voltage.

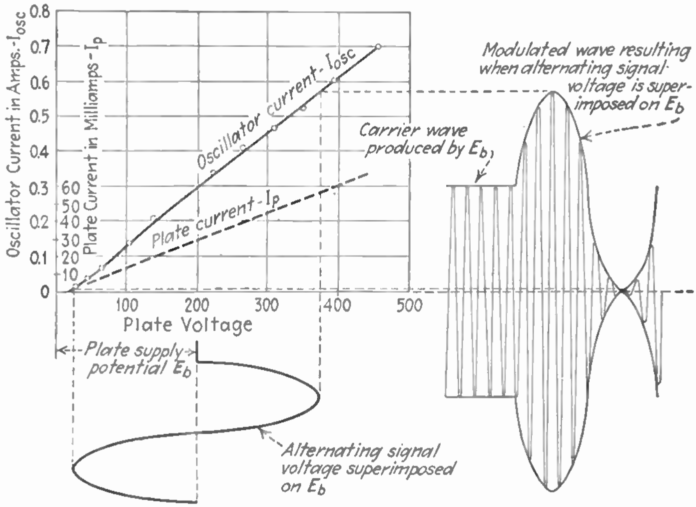


FIG. 125.—Plate and resonant-circuit currents as a function of plate-supply voltage in a typical oscillator having grid leak and condenser, and adjusted for efficient operation. The tank-circuit current is seen to be almost exactly proportional to the plate voltage, so that the oscillations can be modulated by varying the plate-supply voltage about the average value E_b as shown in figure.

side-band energy. The plate efficiency is the same as that of an ordinary oscillator and so commonly ranges from 60 to 80 per cent.

The linearity of a properly adjusted plate-modulated oscillator is practically perfect. The only special precaution required to obtain good linearity, other than arranging the tube to operate at high plate efficiency, is to employ a small enough grid condenser so that the grid bias is capable of following the modulation at the higher modulation frequencies. What distortion occurs is usually produced by the modulator.

The chief limitation of the plate-modulated oscillator arises from the fact that the frequency generated by ordinary oscillators depends somewhat upon the plate-supply voltage. The carrier frequency generated by a plate-modulated oscillator therefore tends to vary with the modulation, introducing frequency modulation (see Sec. 75). As a result, the plate-modulated oscillator, which was once the universally used method of modulation, is now employed only under special circumstances.

72. Carrier-suppression and Single-side-band Systems of Communication.—The carrier component of a modulated wave is not affected in any way by the presence or absence of modulation and so contains no part of the information being transmitted. The carrier wave can therefore be suppressed at the transmitter by some arrangement such as the

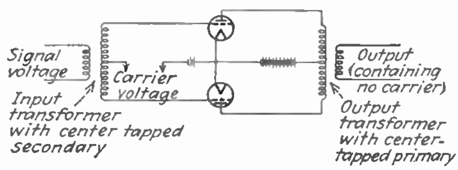


FIG. 126.—Balanced-modulator circuit arranged to suppress the carrier wave from the output without alternating the side bands.

balanced modulation circuit shown in Fig. 126. Here the carrier voltage is applied to the two van der Bijl modulated amplifiers in the same phase as shown, while the modulating voltage is applied in opposite phase to the two grids by means of the center-tapped transformer. The outputs in the plate circuits of the two tubes are combined through a transformer with a center-tapped primary in such a way that voltages applied to the two grids in the same phase cancel each other in the output, while voltages applied to the two grids 180° out of phase will add in the output. The result is that the carrier voltage, which is applied to the two tubes in the same phase, does not appear in the secondary of the output transformer, whereas the side-band components, which are produced in opposite phase, are added to give an output that is a modulated wave from which the carrier component has been removed.

Since each side band taken alone contains all the information present in a modulated wave, it is possible to carry on communication by transmitting only a single side band and by suppressing the carrier and other side band at the transmitter. The single side band is obtained by first suppressing the carrier by some such arrangement as shown in Fig. 126 and then passing the resulting side bands through filter circuits that are sufficiently selective to transmit one side band while suppressing the other.

The single-side-band system of communication is able to transmit a given signal with a frequency band only half as wide as that required by a modulated wave consisting of two side bands and a carrier, and also saves over two-thirds in power because of the suppression of the carrier. Single-side-band transmission is extensively used in carrier-current communication over wire lines, but the difficulty of producing large amounts of single-side-band power at radio frequencies, and the difficulty of receiving the signals, have prevented single-side-band transmission from being standard practice in radio work.

73. The Use of Feedback in Modulation Systems.—The linearity of any modulation system can be greatly improved by making use of a

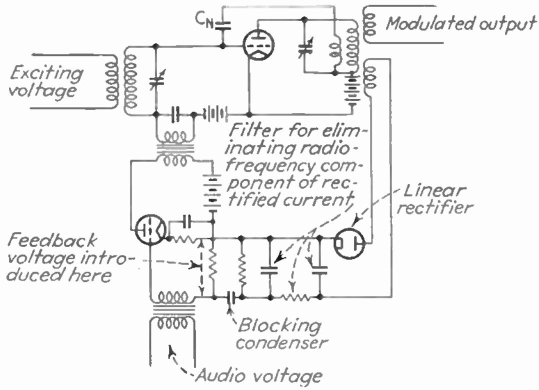


Fig. 127.—Negative feedback as applied to a grid-modulated amplifier.

modified form of negative feedback. This is illustrated in Fig. 127 as applied to a grid-modulated amplifier. Here a portion of the modulated output is rectified in order to obtain a pulsating current that reproduces the modulation envelope.[1] The audio-frequency component of this rectified current is then separated from the direct-current component and fed back into the audio amplifier with a polarity that, in the middle-frequency range, is negative with respect to the audio signal being amplified. The resulting effect is similar to that obtained by the use of negative feedback in audio-frequency amplifiers. That is, the non-linear and frequency distortion and hum of the modulation envelope are greatly reduced, particularly if the feedback is large.

The use of negative feedback in this manner makes it possible to adjust a modulated amplifier on the basis of good efficiency, large output, etc., while allowing the linearity of modulation to be a secondary factor. The

[1] The rectifier must be designed according to the principles discussed in Sec. 77 in order to avoid distortion in the rectified output.

feedback is then utilized to eliminate whatever non-linear distortion would otherwise be present in the modulation envelope.

In the application of feedback to modulation systems it is necessary to keep in mind that the radio-frequency circuits shift the phase of the modulation envelope in the same manner that audio-frequency circuits shift the phase of the audio-frequency currents. It is hence advisable to feed back from the output of the modulated amplifier to the input of the modulator stage. Under these conditions little trouble with oscillations need be expected. However, when attempts are made to include an additional radio-frequency or audio-frequency stage in the feedback loop, there is considerable likelihood of oscillation unless the feedback is quite small or unless special design procedures are devised.

74. Experimental Determination of Modulation Linearity.—The linearity of modulation that is actually obtained under practical working conditions can be determined by modulating with a sine-wave voltage and observing the resulting modulation envelope by means of a cathode-ray tube, or by measuring the degree of modulation at the positive and negative peaks with a modulation meter. The cathode-ray tube is less accurate but, because of its availability, is ordinarily used for the general run of adjustments.

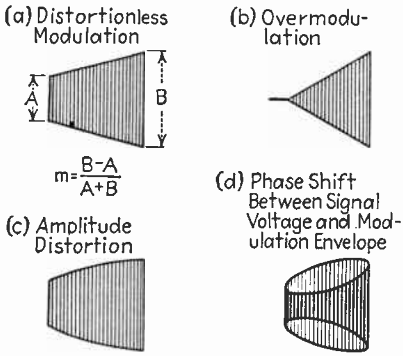


FIG. 128.—Patterns obtained under various conditions from a cathode-ray tube in which the modulated wave is applied to the vertical deflector plates and the signal voltage to the horizontal deflectors.

The simplest method of using a cathode-ray tube is to apply the modulating voltage to the horizontal deflecting plates and the modulated output voltage to the vertical deflectors. If the modulation is distortionless and without phase shift, the resulting pattern is a trapezoid with straight sides, as indicated in Fig. 128*a*, with the degree of modulation determined by the ratio of the vertical sides, as indicated in the figure. Overmodulation, *i.e.*, modulation to the point where the output is reduced to zero for an appreciable part of the cycle, is indicated by the pattern shown at Fig. 128*b*. Amplitude distortion causes the sides of the trapezoid to become curved as in Fig. 128*c*, with the nature of the curvature usually indicative of the source of the distortion. A shaded ellipse as in Fig. 128*d* indicates a phase shift between the modulation envelope and the modulating voltage and does not necessarily mean distortion.

In the absence of a modulation meter or cathode-ray tube, a rough idea as to the linearity of the modulation may be obtained by taking advantage of the fact that, when the modulation is distortionless, the d-c plate current of most types of modulators is substantially independent of the degree of modulation, whereas distortion is usually accompanied by a change in the d-c current when modulation is applied. The percentage of modulation can also be estimated for sinusoidal modulation by taking advantage of the fact that the effective value of a sinusoidally modulated wave is proportional to $\sqrt{1 + \frac{m^2}{2}}$ and so with complete modulation is 1.225 times the unmodulated value when distortion is absent.

75. Frequency and Phase Modulation.—Frequency modulation is produced by varying the frequency of the radio-frequency wave in accordance with the information to be transmitted, while maintaining the amplitude constant. The extent of the frequency variation in such a wave is made proportional to the amplitude of the modulating voltage, whereas the rate of frequency variation, *i.e.*, the number of times the frequency is changed between the minimum and maximum values per second, corresponds to the modulation frequency in amplitude modulation. Thus, if a 500-cycle wave is to be transmitted by frequency modulation of a 1,000,000-cycle carrier wave, this could be done by varying the transmitted frequency between 1,000,010 and 999,990 cycles, 500 times a second. If the modulating frequency is increased to 1000 cycles, the carrier frequency will be varied between the same two limits 1000 times a second. Furthermore, a modulating wave of twice the intensity will be transmitted by varying the carrier frequency through twice the frequency range, *i.e.*, from 1,000,020 to 999,980 cycles in the above case. The appearance of a frequency-modulated wave is shown in Fig. 129*b*.

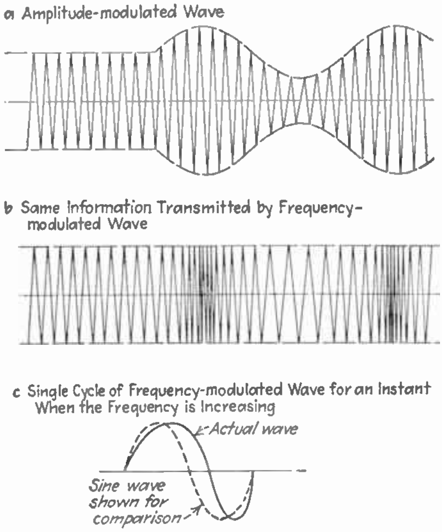


FIG. 129.—Character of waves produced by frequency modulation, together with large-scale reproduction of a single cycle, showing how the wave shapes are not sinusoidal.

Analysis of the Frequency-modulated Wave.—A superficial examination of frequency modulation might lead one to believe that information could be transmitted in this way with an extremely narrow frequency band, since in the case cited above it appears that only 20 cycles band width is required to transmit the 500-cycle wave. This is not correct, however, because the variation in the frequency prevents the individual cycles from being exactly sinusoidal in shape. This is illustrated in Fig. 129c, where it is apparent that, since the changing frequency causes the time required to complete one quarter cycle to differ from the time required by the next quarter cycle, the actual wave is not a sinusoidal oscillation. In fact, exact analysis shows that the frequency-modulated wave not only contains the same side-band frequencies as does the corresponding amplitude-modulated wave, but also has higher-order side bands that differ from the carrier frequency by integral multiples of the modulation frequency. Thus, when a carrier wave of frequency f is frequency modulated at a rate of F cycles per second, the resultant wave contains components having frequencies of $f \pm F$, $f \pm 2F$, $f \pm 3F$, etc.

The ratio of the amplitudes of the various side-band components to the amplitude of the unmodulated carrier is determined by the modulation index m_f, which is defined as

$$\left.\begin{matrix}\text{Modulation}\\ \text{index } m_f\end{matrix}\right\} = \frac{\left\{\begin{matrix}\text{Variation in carrier frequency away}\\ \text{from average carrier frequency}\end{matrix}\right\}}{\text{Modulating frequency}} \tag{99}$$

The amplitudes of the first-, second-, and third-order side-band components, *i.e.*, the side bands that differ from the carrier frequency by F, $2F$, and $3F$, are given in Fig. 130. These curves show that, when the modulation index m_f is less than 1, *i.e.*, when the range through which the frequency is varied is less than the audio frequency, the amplitude of the first-order side band is approximately proportional to the modulation index, while the higher-order side bands are comparatively small.[1] When the modulation index exceeds unity, *i.e.*, when the range through which the radio frequency is varied is greater than the audio frequency, the second- and other higher-order components become of importance, while the carrier amplitude drops rapidly and may even be zero.

Analysis of Phase Modulation.—In phase modulation the amplitude of the wave is maintained constant while the phase of the radio-frequency output is varied in accordance with the information to be transmitted. Phase modulation is similar to frequency modulation, since a changing phase corresponds to a varying frequency. The principal difference is in the mechanism by which the modulation is produced. With phase modulation, the modulation index is equal to the phase shift in radians and is commonly designated by the symbol m_p.

[1] The question naturally arises as to why frequency modulation under such conditions is not the same as amplitude modulation, since the modulated waves contain substantially the same frequency components in the two cases. The difference lies in the fact that the carrier phase with respect to the side-band phase differs by 90°. Hence a frequency-modulated wave in which the second- and still higher-order side bands are negligible can be converted into an amplitude-modulated wave by shifting the carrier phase 90° with respect to the side bands.

A phase-modulated wave with sinusoidal modulation contains the same side-band components as does a frequency-modulated wave, and, if the modulation indices in the two cases are the same, the relative amplitudes of these different components will also be the same. The relative amplitude of the carrier and the first three side bands can hence be obtained from Fig. 130 for any given modulation index. As long as the modulation index is less than unity, (*i.e.*, phase shift less than 57.3°), only the first-order side-bands are of appreciable magnitude, but each additional 57.3° of phase shift will add another pair of important side-band components.

Combinations of Phase, Amplitude, and Frequency Modulation.—Frequency and phase modulation are often combined with amplitude modulation as undesirable by-products. For example, in the plate-modulated oscillator the output possesses both frequency and amplitude modulation, as has already been explained. For this reason modulated oscillators are practically never used in radio communication.

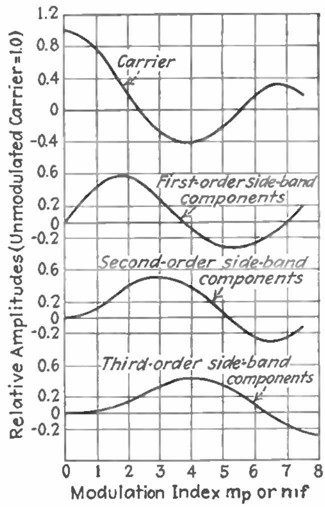


FIG. 130.—Amplitudes of components of a frequency- or phase-modulated wave assuming that the amplitude of the unmodulated wave is 1.0. In the case of the side bands the amplitude shown is the amplitude of the individual side-band component and not of the pair of companion side bands taken together.

Combined phase and amplitude modulation can occur in a number of ways. Thus, if the tank circuit of a modulated amplifier or linear amplifier is not tuned exactly to resonance, there will be a phase shift that will vary with the modulation. Another important cause of combined amplitude and phase modulation is energy transfer between the wave after modulation, and the unmodulated wave before modulation.

Phase and frequency modulation that occurs as a by-product of amplitude modulation is very undesirable in radio transmitters. This is because such modulation produces high-order side-band frequencies, which represent energy radiated upon adjacent frequency bands and may interfere with other communications. Phase modulation is particularly bad in this respect, because the modulation index m_p of phase modulation is independent of the frequency of modulation, whereas with frequency modulation the index m_f is inversely proportional to the modulation frequency and so tends to be low when the modulation frequency is high enough to make the second- and third-order side bands lie in adjacent channels.

Problems

1. The equation of a modulated wave is

$$e = 25(1 + 0.7 \cos 5000t - 0.3 \cos 10{,}000t) \sin 5 \times 10^6 t$$

a. What frequency components does the modulated wave consist of, and what is the amplitude of each?

b. Sketch the modulation envelope and evaluate the degree of modulation for the peaks and troughs.

2. Write the equation of a 100-volt carrier wave of 1000 kc when modulated 40 per cent at 400 cycles.

3. Calculate the total band width to which a radio receiver should respond for satisfactory reception of (*a*) ordinary broadcast signals, (*b*) perfect reproduction of speech and music, (*c*) telegraph signals sent at 150 letters per minute under average conditions, and (*d*) radio signals that represent radio extensions of wire telephone systems.

4. Calculate (*a*) the number of telegraph stations transmitting at a speed of 25 words per minute, 5 letters per word, which, theoretically, could operate simultaneously in the present broadcast band (550 kc to 1500 kc), (*b*) the number of ordinary broadcast stations reproducing speech frequencies up to 8000 cycles that could be accommodated in the broadcast band, and (*c*) the number of single-side-band telephone conversations for the same band.

5. Assuming that a man talks at 120 words per minute and that the average word has 5 letters, find the relative amounts of information that can be transmitted in a given frequency band by telegraphy and by telephony.

6. Draw the circuit of a plate-modulated Class C amplifier similar to Fig. 118*b* but using a Class B modulator and a neutralized push-pull Class C amplifier.

7. A plate-modulated amplifier is to deliver a carrier output of 100 watts. Select a suitable Class C tube from Table XIII and specify the modulator power that will be required for complete modulation and the load impedance to which the modulator delivers its output.

8. In a direct-coupled arrangement as shown in Fig. 118*a*, the modulator consists of 3 Type 211 tubes (see Table X, page 158) connected in parallel and acting as Class A amplifiers for a single Type 211 Class C tube. What are the proper plate voltage and plate current for the Class C tube to operate at if the modulator is to function at full output and give complete modulation, and how many ohms should the voltage dropping resistance have?

9. A plate-modulated amplifier is to employ a Type 204-A tube operated to give a carrier output of 350 watts, as in Table XIII.

a. Design a suitable tank circuit.

b. Select a tube to provide the required exciting power with a moderate margin of safety.

c. Determine the modulating power required for complete modulation and the load impedance to which this power is delivered.

d. Select suitable modulator tubes from Table X, page 158, to operate as Class B amplifiers, and determine the proper turn ratio of the output transformer used to couple the modulator to the plate of the Class C tube.

10. If the plate efficiency of a plate-modulated Class C amplifier is substantially constant throughout the modulation cycle, what must be the ratio of actual plate dissipation during unmodulated periods to the allowable plate dissipation, if the allowable dissipation is not to be exceeded with complete modulation.

11. Draw the circuit of a grid-modulated Class C amplifier using push-pull Class C tubes and draw a push-pull modulator that is transformer coupled to the grid of the Class C tube. Include provision for neutralization of the Class C tubes.

12. A Type 211 tube is to be used as a grid-modulated Class C amplifier.

a. Estimate the carrier output power that can be developed.

b. Determine the proper direct-current bias voltage and the peak modulating voltage that will be required for complete modulation.

c. From the grid current that flows at the crest of the modulation cycle determine the maximum internal impedance that the source of modulating voltage can have without introducing more than 5 per cent second-harmonic distortion in the modulating voltage, and, from this, design a Class A push-pull modulator. The design of modulator includes selection of tubes and tube operating conditions and specification of the proper turn ratio for the output transformer.

13. A grid-modulated Class C amplifier is adjusted so that at the crest of the modulation cycle the plate efficiency is 70 per cent, falling to 35 per cent during unmodulated intervals. Compare the relative plate dissipations for the two modulation conditions.

14. In a grid-modulated Class C amplifier the average plate dissipation is less when the output is modulated than during unmodulated periods. What is the reason for this?

15. In a suppressor-grid-modulated amplifier, the screen current is greater at the peaks of the modulation cycle than at the troughs. Explain.

16. By making reasonable assumptions as to plate efficiencies involved, estimate the total direct-current input power required by modulator and modulated tubes when the carrier power is 1000 watts and the modulation is zero and 100 per cent, for (*a*) plate-modulated Class C amplifier, Class A modulator; (*b*) plate-modulated Class C amplifier, Class B audio modulators; (*c*) grid-modulated amplifier; (*d*) suppressor-grid-modulated amplifier; (*e*) high efficiency grid-modulated Class C amplifier.

17. Draw a circuit of a plate-modulated oscillator analogous to the circuit of Fig. 118*b* for a plate-modulated Class C amplifier.

18. Would it be possible for a balanced modulator to employ plate-modulated Class C amplifiers?

19. Draw a circuit in which feedback is applied to a plate-modulated Class C amplifier.

20. Describe a means by which a frequency-modulated wave could be produced.

21. A wave is frequency modulated at an audio rate of 5000 cycles per second. Plot the relative amplitude of the carrier, and first-, second-, and third-order side-band components as a function of the amount the frequency is varied away from the carrier up to 15,000 cycles.

CHAPTER IX

VACUUM-TUBE DETECTORS

76. Detection of Radio Signals.—Detection, or *demodulation* as it is sometimes called, is the process of reproducing the transmitted information from the modulated radio-frequency wave. Since all systems of radio communication in practical use transmit information by varying the amplitude of the radiated wave, the detection process must ordinarily produce currents that vary in accordance with the modulation envelope. This is done by rectifying the modulated wave.

An ideal detector reproduces in its output the exact information modulated upon the radio wave. If the detector fails to do this, distortion results. Thus the detector output may include frequencies that were not contained in the original modulation, thereby giving rise to amplitude distortion. The detector may also discriminate between different modulation frequencies and so introduce frequency distortion. Finally, a detector may reproduce the different components of the original modulation in altered phase relations, resulting in phase distortion.

77. Diode Detectors.—A simple diode detector is shown in Fig. 131*a*. Here a two-electrode tube, or diode, is used as a rectifier, with the resistance R serving as the load to which the rectified output is delivered. This resistance is shunted by a condenser C large enough to act as a short circuit to radio-frequency voltages, but small enough to be substantially an open circuit to rectified currents of modulation frequencies.

When a modulated radio-frequency voltage is applied to the diode detector, the action is as illustrated in Fig. 131. The potential existing between the plate and cathode electrodes is the radio-frequency wave minus the voltage E_1 developed across the load resistance R by the rectified current, as shown at Fig. 131*c*. It is seen from Fig. 131*d* that the plate becomes positive with respect to the cathode for a brief period each cycle, and that during the time the plate is positive a pulse of current flows. With the usual circuit proportions the voltage E_1 developed across the load by the pulses of plate-current is only slightly less than the crest amplitude of the radio-frequency wave, so that when the wave is modulated the voltage developed across the diode load resistance R varies likewise and accordingly reproduces the modulation envelope.

Efficiency, Input Resistance and Linearity of Simple Diode Detectors Having a Resistance Load.—The ratio of voltage developed across the load resistance R to crest amplitude of the radio-frequency signal voltage is termed the efficiency of detection. This efficiency commonly ranges from 80 to 95 per cent if the amplitude of the applied signal exceeds a few volts. The exact value of efficiency depends upon the ratio of load resistance R to the resistance of the diode tube, and increases as this ratio is made greater. This is because a high load impedance requires less rectified current to develop a given output voltage and hence reduces the amount by which the plate must go positive each cycle. A low diode resistance leads to the same result, since this gives more current for the same positive voltage on the plate.

The efficiency of rectification can be measured by applying an unmodulated alternating voltage of known crest amplitude to the diode and measuring the resulting rectified current flowing through the resistance R. From this one can calculate the voltage developed across R and hence the efficiency of rectification. Measurements of efficiency of rectification can be made at audio frequencies, provided the condenser C has a low reactance compared with the resistance R at the frequency used.

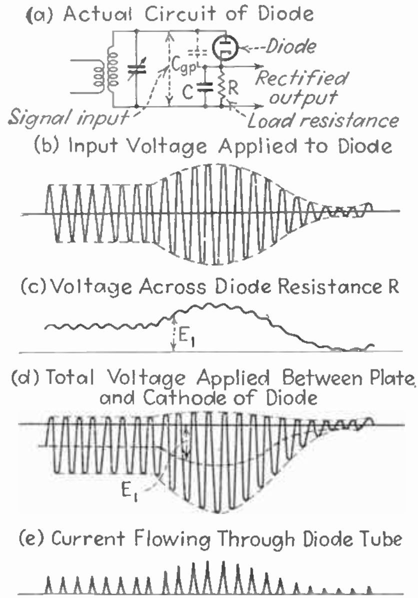


FIG. 131.—Circuit of simple diode detector with oscillograms illustrating details of operation.

A diode detector with its load consumes energy from the applied signal as a result of the current flowing through the diode tube. For the simple diode of Fig. 131a, the equivalent resistance that the diode plus load offers to the applied radio-frequency voltage is to a high degree of accuracy given by the equation[1]

[1] This is shown as follows: The diode current flows only when the signal voltage is at or near its crest value, as is clearly shown in Fig. 131. The power absorbed by the detector input is accordingly only very slightly less than the product of the crest radio-frequency voltage and the average diode current. Since the average current is equal to $\eta E/R$, where the output voltage across R is ηE, and E is the crest value of radio-frequency voltage, one can write

$$\text{Power loss} = \frac{\eta E^2}{R} = \frac{(\text{effective signal})^2}{(R/2\eta)}$$

The denominator of this last term represents the equivalent input resistance to the signal, which is, accordingly, $R/2\eta$.

$$\text{Input resistance} = \frac{R}{2\eta} \tag{100}$$

where η is the efficiency of rectification.

In an ideal detector the voltage E_1 is exactly proportional to the amplitude of the applied radio-frequency voltage. That is, the efficiency of rectification is independent of the applied voltage. In practical detectors this is only approximately true, because at low amplitudes the diode plate resistance increases, causing the rectification efficiency to drop. Trouble from this source may be minimized by using a high load resistance and making the carrier amplitude applied to the diode reasonably large. The rectification efficiency will then be high at all times even if not constant, and the rectified voltage developed across the load resistance will follow the modulation envelope quite closely.

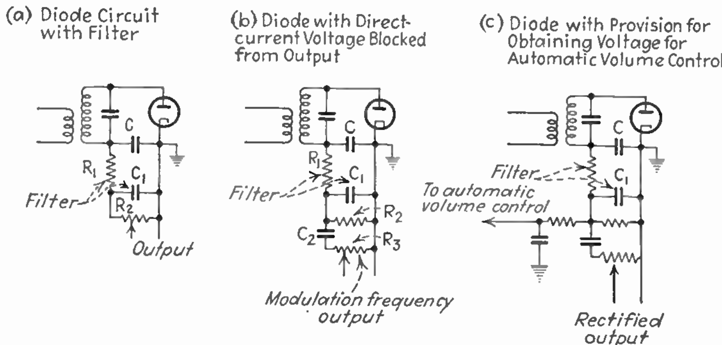


FIG. 132.—Practical diode detector circuits.

Practical Diode Detector Circuits and Complex Load Impedances.—In practical detectors it is generally found necessary to provide a filter such as shown in Fig. 132*a* in order to prevent radio-frequency voltages from reaching the output. In this filter, condensers C and C_1 are commonly equal, whereas R_2 may be several times R_1. A filter reduces the useful rectified output by the factor $R_2/(R_1 + R_2)$ as a result of the voltage drop in R_1.

A modulation-frequency voltage, separated from the direct-current voltage developed across the diode load resistance, can be obtained by the use of a resistance-condenser combination such as illustrated in Fig. 132*b*. It is also common practice to use the direct-current voltage developed across the diode load resistance for automatic volume control by means of an arrangement such as illustrated in Fig. 132*c* for separating this d-c voltage from the alternating modulation-frequency voltage.

A study of Figs. 132*b* and 132*c* shows that these arrangements do not provide a simple resistance load to the rectified current as does the arrangement of Fig. 131*a*. Thus in Fig. 132*b*, if the blocking condenser

C_2 is an effective by-pass to the modulation frequencies, as it is supposed to be, then the resistance R_3 is in shunt with the resistance R_2 to the modulation-frequency components of the rectified current, but not to the direct-current component. The diode load circuit in Fig. 132*b* hence offers a different impedance to the modulation-frequency components of the rectified current than to the direct-current component. A similar situation exists in the arrangement of Fig. 132*c*.

In all the diode detector circuits, even including the simple circuit of Fig. 131*a*, it is to be noted that the condenser C that serves as a radio-frequency by-pass to the diode load circuit will also act as a partial by-pass for very high modulation frequencies. Under these conditions the impedance of the diode load will have a reactive component and will also be reduced in magnitude.

Effect of a Complex Load Impedance on Non-linear Distortion.—A consideration of the action taking place in a diode detector shows that when the detection efficiency is high, as is normally the case, the rectified current assumes whatever value is required to make the voltage developed across the load impedance have a value only slightly less than the amplitude of the modulation envelope. This means that the average value I_0 of the rectified current must be such that $I_0 \times R_0 = \eta E_0$, where R_0 is the direct-current resistance offered by the diode load to the rectified current η is the efficiency of rectification, and E_0 is the average or carrier amplitude of the applied radio-frequency voltage. At the same time, the modulation-frequency component I_m of the rectified current must be such that when this current flows through the impedance Z_m offered by the diode load circuit to the modulation frequency, one has $I_m \times Z_m = \eta m E_0$.

From these relations it follows that as the ratio Z_m/R_0 of alternating-current to direct-current load impedance is made less than unity, the ratio of the modulation-frequency component of the rectified current to the direct-current component is increased, as shown in Fig. 133*a*. However, the crest value of the modulation-frequency component of the rectified current cannot exceed the direct-current component because the rectifier tube will transmit current in only one direction. Consequently, the maximum degree of modulation that the applied radio-frequency voltage can possess, and still make it possible for the modulation-frequency component of the rectified output to reproduce the modulation envelope, has the value

$$\left.\begin{array}{l}\text{Maximum allowable}\\ \text{degree of modulation}\\ m \text{ for distortionless}\\ \text{rectification}\end{array}\right\} = \left|\frac{Z_m}{R_0}\right| = \frac{\left\{\begin{array}{l}\text{Magnitude of the impedance}\\ \text{of diode load to modulation}\\ \text{frequency}\end{array}\right.}{\left\{\begin{array}{l}\text{Resistance of the diode load}\\ \text{circuit to direct current}\end{array}\right.} \qquad (101)$$

If the degree of modulation of the applied signal exceeds this value the rectified output will be distorted.

The nature of the non-linear distortion that results when the modulation of the applied signal is greater than $|Z_m/R_0|$ depends upon the magnitude and phase angle of the complex impedance Z_m/R_0. In circuits such as those of Figs. 132*b* and 132*c*, the impedance Z_m offered to moderate modulation frequencies is a resistance R_m appreciably less than the resistance R_0 to direct current.[1] When the degree of modulation equals R_m/R_0

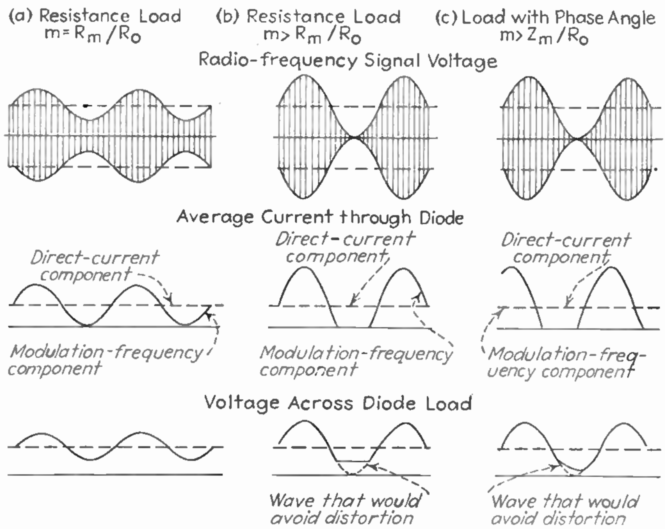


FIG. 133.—Action of diode detector having complex load impedance, showing negative peak and diagonal clipping.

the rectified current then varies through the maximum range that it can and still reproduce the envelope of the applied radio-frequency signal (see Fig. 133*a*). If the degree of modulation exceeds this value, as shown in Fig. 133*b*, the result is a clipping of the negative peaks in the rectified output voltage as shown. This introduces non-linear distortion.[2]

At high modulation frequencies the ratio Z_m/R_0 is not only less than unity but also has a phase angle. This is because of the shunting effect

[1] By moderate modulation frequencies is meant those frequencies high enough for the blocking condenser in the rectifier load circuit to be a short circuit, and yet low enough so that the condenser across the input to the diode load has negligible shunting effect.

[2] When the amount of clipping is small, the total r.m.s. distortion that results is roughly:

$$\left.\begin{array}{l}\text{Approximate}\\ \text{r.m.s. dis-}\\ \text{tortion}\end{array}\right\} = \frac{[\text{Actual modulation}] - [\text{Modulation allowed by Eq. (101)}]}{2 \times (\text{actual modulation})} \qquad (102)$$

of the condenser C at the higher modulation frequencies. Under these conditions the situation is as illustrated in Fig. 133c. It will be noted that the modulation-frequency variations in the rectified current are no longer in phase with the modulation envelope. This must be so, since, if the voltage developed across the diode load circuit is to be only slightly less than the envelope amplitude of the applied radio-frequency voltage, the rectified current cannot be in phase with the modulation envelope when the load impedance has a reactive component. If Z_m is reactive and the modulation of the applied signal exceeds $|Z_m/R_0|$, as is the case in Fig. 133c, the negative half cycles of the output wave are clipped diagonally as shown.

Input Resistance and Demodulation of Applied Signal with Complex Diode Load Impedance.—When the ratio $|Z_m/R_0|$ is less than unity, the diode offers a lower input impedance to the side bands than to the carrier. This is because under these conditions the modulation-frequency variations in the rectified current are proportionately greater than the amplitude variations in the modulation envelope (see Fig. 133a).

The input resistance that the diode offers to the carrier voltage has already been shown in Eq. (100) to be $R_0/2\eta$, where η is the efficiency of rectification. In a similar fashion, the input impedance that the diode offers to the side-band components of the applied signal has a magnitude

$$\left.\begin{array}{l}\text{Input impedance of diode}\\ \text{to side-band frequencies}\end{array}\right\} = \frac{Z_m}{2\eta} \tag{103}$$

The phase angle of the input impedance to the upper (or sum-frequency) side band is the same as the phase angle of Z_m to the modulation frequency, while the phase angle of the input impedance to the lower (or difference-frequency) side band has the same magnitude as the phase angle of Z_m but is of opposite sign. At moderate modulation frequencies the impedance Z_m that the diode load circuit offers to the modulation frequency is usually a resistance lower in value than the resistance offered to direct current. At the high modulation frequencies the impedance Z_m has a reactive component because of the shunting effect of the condenser C, which causes the input impedance that the diode offers the tuned input circuit to contain a reactive component.

The input impedance of the diode detector makes the radio-frequency voltage appearing at the terminals of the diode less than if the diode plus load consumed no energy. The amount of this reduction is determined by the input impedance of the diode in relation to the equivalent generator impedance of the source of applied voltage. Where $|Z_m/R_0|$ is less than unity, the lower input impedance that the diode offers to the side bands causes these to be reduced proportionately more than the carrier. The result is that the degree of modulation of the radio-frequency voltage actually applied to the input terminals of the diode is less than the degree of modulation of the original radio-frequency signal, in the ratio[1]

$$\frac{\left.\begin{array}{l}\text{Degree of modulation of}\\ \text{diode input voltage}\end{array}\right\}}{\left.\begin{array}{l}\text{Actual modulation of}\\ \text{original radio-frequency}\\ \text{wave}\end{array}\right\}} = \left|\frac{Z_m\left(Z_s + \dfrac{R_0}{2\eta}\right)}{R_0\left(Z_s' + \dfrac{Z_m}{2\eta}\right)}\right| \tag{104}$$

where the equivalent generator impedances of the applied voltage to carrier and side-band frequencies are Z_s and Z_s', respectively.[2] The

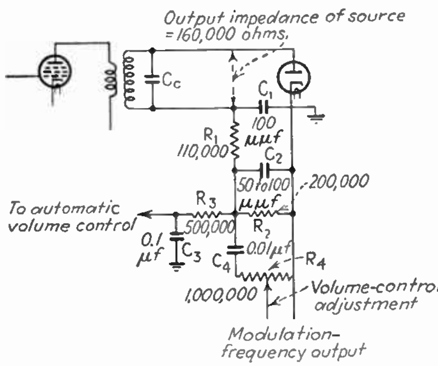


FIG. 134.—Typical design of diode detector in which provision is made for an automatic-volume-control voltage and for obtaining the modulation-frequency output free of direct-current voltage.

lowering of the degree of modulation of the applied signal that results when $|Z_m/R_0|$ is less than unity tends to reduce the distortion that would otherwise occur when the degree of modulation of the radio-frequency signal is high.

Practical Diode Detectors.—The circuit details of a practical diode detector are shown in Fig. 134. Here a modulation-frequency output voltage free of rectified direct-current potentials is obtained with the aid

[1] This is derived as follows: To the carrier the diode offers a load resistance $R_0/2\eta$, so that the presence of the diode reduces the carrier voltage across the tuned input circuit by the factor $\dfrac{(R_0/2\eta)}{\left(Z_s + \dfrac{R_0}{2\eta}\right)}$. Since the input impedance to the side bands is $Z_m/2\eta$, the presence of the diode reduces these by the factor $\dfrac{(Z_m/2\eta)}{\left(Z_s' + \dfrac{Z_m}{2\eta}\right)}$. The ratio of these two reduction factors gives the factor by which the modulation is altered, and leads at once to Eq. (104).

[2] In Eq. (104) it is assumed that the impedance Z_s of the source to the side bands has the same magnitude for the lower as for the upper side band, but that the phase angles are opposite in the two cases.

of a filter R_4C_4, while a direct-current voltage proportional to the carrier amplitude of the applied radio-frequency wave and free of modulation-frequency components is obtained for the automatic-volume-control system with the aid of the filter R_3C_3. A filter $R_1C_1C_2$ prevents radio-frequency voltages from reaching the modulation-frequency output. The resistance R_1 in this filter is normally made rather high in spite of the loss of output which results, since then Z_m/R_0 is more nearly unity.

The procedure to be followed in analyzing a diode detector circuit is illustrated by the following example.

Example.—Calculate maximum degree of modulation that the applied signal can have in Fig. 134 and still avoid negative peak clipping at moderate modulation frequencies. Also determine the approximate distortion that results when the original signal voltage is modulated 100 per cent. Assume that the efficiency of rectification is 0.90.

The direct-current load resistance R_0 is 110,000 + 200,000 = 310,000 ohms. The load impedance to moderate modulation frequencies is less, however, because R_3 and R_4 are effectively in shunt with R_2 to alternating currents. Hence

$$Z_m = 110{,}000 + \frac{0.2 \times 0.5 \times 1.0 \times 10^{18}}{(0.2 \times 0.5 + 0.2 \times 1.0 + 0.5 \times 1.0)10^{12}} = 235{,}000 \text{ ohms.}$$

Accordingly, $Z_m/R_0 = 235{,}000/310{,}000 = 0.76$. This represents the maximum degree of modulation that the applied radio-frequency voltage can have and still avoid negative peak clipping. From Eq. (104), the actual degree of modulation of the original signal required to make the modulation of the signal at the diode terminals 0.76 is

$$0.76\frac{310{,}000\left(160{,}000 + \dfrac{235{,}000}{2 \times 0.9}\right)}{235{,}000\left(160{,}000 + \dfrac{310{,}000}{2 \times 0.9}\right)} = 0.88$$

When the original signal is completely modulated the degree of modulation at the diode input terminal is $1.0 \times 0.76/0.88 = 0.86$. Substitution in Eq. (102) then gives

$$\text{Approximate r.m.s. distortion} = \frac{0.86 - 0.76}{2 \times 0.86} = 0.058.$$

Tubes for Diode Detectors.—The tube used as a diode detector should have a low plate resistance. The diodes employed in actual practice include small, especially designed two-electrode tubes, diode sections build into triode and pentode tubes, and also ordinary triodes converted into diodes by connecting the grid and plate electrodes together.

78. Plate Detectors.—In the plate detector (also called bias detector and anode detector), advantage is taken of the fact that, when an amplifier tube is biased approximately to cut-off and a relatively large alternating voltage is applied to the grid, there will be a pulse of plate current during each positive half cycle of the applied voltage and no plate current during the negative half cycles, as illustrated in Fig. 135. The amplitude of these pulses of plate current is proportional to the alternating voltage applied to the grid so that, if this input is a modulated wave, the average value of the plate current will vary in accordance with the modulation envelope. Modulation-frequency output is obtained by coupling to the plate circuit as in an ordinary audio-frequency amplifier, using resistance, transformer, or other form of coupling.

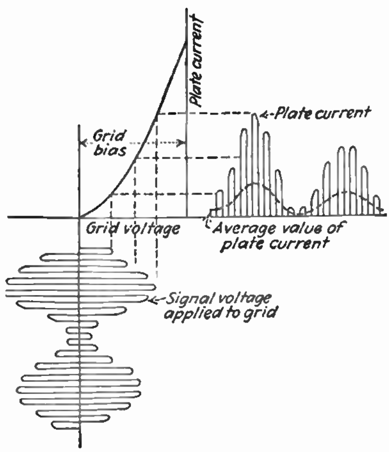


FIG. 135.—Details of action taking place in plate rectifier showing how a modulated wave applied to a grid that is biased to cut-off will cause plate-current impulses having an average value that varies in accordance with the modulation envelope.

The plate detector has the advantage over the diode in that the input resistance is infinite if the grid is not allowed to go positive. At the same time, the plate detector can handle only a limited applied voltage and also does not produce a rectified voltage of the proper polarity for automatic-volume-control purposes. As a result, diode detection is preferred to plate detection under most circumstances.

Analysis of Plate Detectors.—The behavior of the modulation-frequency currents in the plate circuit of a plate detector is exactly as though these were produced by a generator acting inside the tube. This leads to the equivalent circuit of Fig. 136, which is of the same form as the equivalent circuit of the amplifier shown in Fig. 56. The only difference between the equivalent circuit of the plate detector and the equivalent circuit of the amplifier is that with the detector the equivalent generator voltage E_r

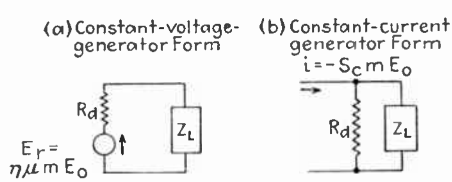


FIG. 136.—Equivalent plate circuits of plate detector. Note the similarity to the equivalent amplifier circuits of Fig. 56.

and the equivalent internal resistance R_d are now determined by different factors.

The equivalent modulation-frequency voltage E_r that can be considered as acting in the equivalent circuit of the plate detector is given by the equation

$$E_r = \eta\mu m E_0 \tag{105}$$

where

m = degree of modulation of the applied signal
E_0 = carrier amplitude of the applied signal
μ = amplification factor of tube
η = efficiency of rectification.

It will be noted that the modulation-frequency envelope actually acting in the plate circuit is $\mu m E_0$, and that the equivalent rectified voltage E_r is this envelope multiplied by the factor η to take into account the incompleteness of the conversion from radio-frequency envelope to modulation-frequency voltage. In practical plate detectors the efficiency of rectification is of the order of 0.8 to 0.9 provided the applied carrier is not too small.

The detector plate resistance R_d in Fig. 136 is the equivalent dynamic resistance of the plate circuit of the tube measured when a radio-frequency carrier is applied. This equivalent detector plate resistance is normally two or three times the plate resistance of the same tube when operated as an ordinary amplifier.

The ratio $\mu/R_d = S_c$ corresponds to the mutual conductance of amplifiers and is termed the *conversion transconductance*. The conversion transconductance is the fundamental constant of the pentode plate detector. This is because in such tubes the detector plate resistance R_d is substantially infinite, so that the modulation-frequency voltage developed across the load impedance Z_L in the plate circuit is for all practical purposes given by the equation

$$\left.\begin{array}{l}\text{Modulation-frequency output}\\ \text{voltage of pentode plate}\\ \text{detector}\end{array}\right\} = S_c m E_0 Z_L \tag{106}$$

The value of S_c for plate detection is normally 0.3 to 0.4 times the mutual conductance of the same tube when operated as an ordinary amplifier.

Design of Anode Detectors.—The load impedance placed in series with the plate of the anode detector is designed exactly as in the case of voltage amplifiers, except that the equivalent detector plate circuit of Fig. 136 is used instead of the equivalent amplifier circuit of Fig. 56. Either resistance or transformer coupling is normally used, and the only special feature involved is that a filter must be provided as shown in Fig. 137 to prevent radio-frequency voltages from reaching the output, while providing a low-impedance path in the plate circuit to radio-frequency currents. The condensers in this filter must be small, for they are effectively in shunt with the load impedance and so tend to reduce the response at high modulation frequencies, particularly when resistance coupling is used.

The tubes used for audio-frequency voltage amplification are also suitable for anode detection. Both triode and pentode tubes are satisfactory, although the latter are preferred because of their greater gain. The grid bias should approximate the cut-off value and so depends upon the screen potential in pentodes and upon the plate potential in triodes.

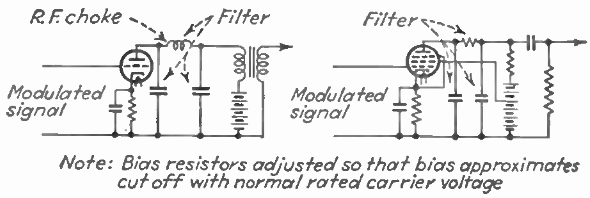


FIG. 137.—Typical plate-detector circuits.

When a bias resistor is used, its value should be chosen so that the necessary bias is obtained *in the presence of the rated carrier input voltage.*

79. Heterodyne Detection.—When two signals of different frequencies are superimposed, the envelope of the resulting oscillation varies in amplitude at a frequency that is equal to the difference between the frequencies of the two a-c currents, and swings through an amplitude range determined by the amplitude of the smaller of the two voltages, as is shown in Fig. 138. This result is obtained because at one moment the two waves will be in phase and so will add together, whereas a short time

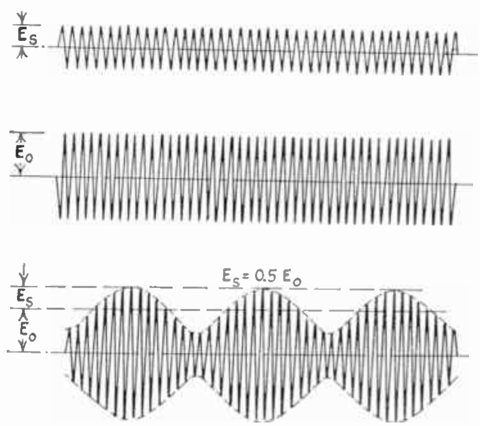


FIG. 138.—Typical heterodyne waves, showing how the combining of two waves of slightly different frequencies results in a wave that pulsates in amplitude at the difference frequency of the component waves.

later the higher frequency wave will be one-half cycle ahead of the other wave and so will combine with it in phase opposition. The rate at which the amplitude of the envelope varies is called the *beat frequency* (or the difference frequency), and the production of such beats by combining two waves is known as *heterodyning.* Since rectification of such a heterodyne signal gives a rectified current that varies in amplitude at the beat frequency, *heterodyne action gives a means of changing the frequency of an a-c current.*

The procedure for changing the frequency of an unmodulated signal by heterodyne action is to superimpose a local oscillation having a frequency that differs from that of the signal by the desired frequency. The local oscillation may have either a higher or lower frequency than does the signal, since it is only the difference that is important. The combined wave is then applied to a detector, and the desired beat frequency will be contained in the rectified output. If the wave that is to have its frequency changed is modulated, the amplitude of the beats that are produced by the superposition of the local oscillation will vary in accordance with the amplitude of the modulated wave, and the final result of the heterodyne operation is to change the frequency of the carrier wave to the beat frequency without disturbing the character of the modulation.

The heterodyne principle of frequency changing has a number of important applications in radio communication. Thus it can be used to change the frequency of a telegraph signal to an audible frequency such as 1000 cycles, which can be used to actuate a telephone receiver. This result is accomplished by making the difference between the signal and local oscillation frequencies a suitable audio frequency, and is known as heterodyne code reception.

Another application of the heterodyne principle is in the superheterodyne type of radio-frequency amplification, in which the heterodyne principle is used to change the carrier frequency of the incoming radio wave to a predetermined and readily amplifiable radio frequency, termed the *intermediate frequency*, at which the amplification takes place. In this way the frequency of the signal is changed to fit the amplifier, rather than requiring that the amplifier be adjusted to fit the signal.

80. Converters for Superheterodyne Receivers.—The superheterodyne receiver is so widely used that special tubes and methods have been devised for performing the frequency-changing operation involved. The heterodyne detector for such purposes, commonly referred to as *first detector*, *converter*, or *mixer*, is required to develop a difference frequency, ordinarily in the range 75 to 500 kc, by combining the incoming radio wave with a local oscillation differing in frequency by the desired amount. The principal arrangements employed to do this in practice are plate detection, the 6L7 pentagrid mixer tube, and the pentagrid converter.

The plate-detector type of mixer is illustrated in Fig. 139, and is an ordinary pentode plate detector biased approximately to cut-off, with provision for introducing a locally generated voltage in the tube circuits. The output load impedance is a tuned circuit resonant at the difference frequency to be produced, and is designed exactly as in an ordinary tuned amplifier. The local oscillator may be of any type and can be coupled into the mixer tube in any convenient way, but preferably in such a manner that the voltage introduced is approximately constant as the frequency of the oscillator is varied to accommodate signals of different frequencies.

The difference-frequency output of a pentode mixer tube is calculated in exactly the same way as is the amplification of a tuned amplifier. It is merely necessary to use the conversion transconductance in place of the mutual conductance, and the detector plate resistance in place of the ordinary plate resistance. Since the plate resistance of a pentode mixer tube is practically infinite, the output obtained is consequently given by Eq. (106) and is about one-third as great as that obtained from an ordinary tuned amplifier because of the fact that the conversion transconductance is about one-third the mutual conductance.

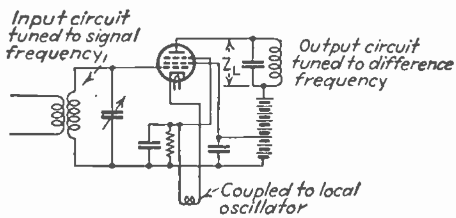


FIG. 139.—Circuit of typical plate-detector type of mixer tube.

The plate-detector type of mixer has the disadvantage of introducing coupling between the local oscillator and the circuit tuned to the incoming signal. Under such circumstances adjustment of the input tuned circuit will affect the oscillator frequency, and strong signals having frequencies differing only slightly from the local-oscillator frequency will tend to make the latter synchronize automatically with the interfering signal (see Sec. 62). As a consequence, the plate-detector type of mixer, although perfectly satisfactory for broadcast and lower frequencies, has very important shortcomings when used at radio frequencies so high that the percentage difference between signal and local oscillator frequencies is small.

The Pentagrid Mixer Tube (or 6L7).—The pentagrid (or hexode) mixer tube, or 6L7, isolates the local-oscillator and signal-frequency circuits by the expedient of utilizing the local oscillator to grid modulate the amplified signal-frequency currents. The pentagrid mixer tube contains five grids connected as shown in Fig. 140. The first or inner grid G_1 is the normal control grid, designed to have a variable-mu characteristic, and has the signal voltage applied to it. The next grid G_2 is an ordinary screen grid, whereas the third grid G_3 is a suppressor grid that is used to modulate the electron stream produced by the control and screen grids

G_1 and G_2. The next grid G_4 is a screen grid, whereas G_5 is a suppressor grid that is connected to the cathode. It will be noted that this arrangement is essentially a suppressor-grid-modulated amplifier, modified by the addition of grids G_4 and G_5 in order to make the plate resistance of the tube similar to that obtained with an ordinary pentode even when a virtual cathode forms between grids G_2 and G_3. The arrangement is capable of producing a difference frequency, since, when the local oscillator is modulated upon the incoming signal, the lower side band has a frequency that is the difference between the signal carrier frequency and the local-oscillator frequency.

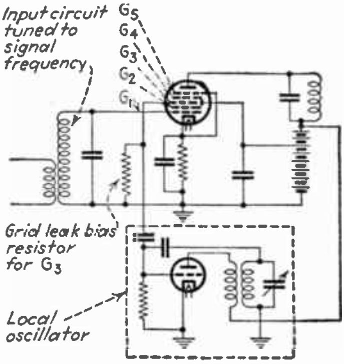


FIG. 140.—Circuit of typical 6L7 mixer tube.

The difference-frequency output obtained with a pentagrid mixer tube is calculated exactly as in the case of the pentode plate detector, and so is given by Eq. (106). The conversion transconductance, and hence the amplification, obtained with a hexode mixer tube are of the same order of magnitude as with other types of converters.

The Pentagrid Converter.—The pentagrid converter is a combined oscillator and detector tube. A typical circuit arrangement is shown in Fig. 141, where the cathode, first grid G_1, and second grid G_2 function as an ordinary triode oscillator with grid G_2 being the anode electrode. This oscillator serves to control the flow of electrons from the cathode, causing the current to travel toward the plate in pulses occurring at the peak of each cycle of the oscillator. Most of the electrons in the pulses drawn from the cathode pass through the spaces between the wires of G_2, pass the screen-grid G_3, which is for the purpose of providing an electrostatic shield, and come to rest in front of the grid G_4, where a virtual cathode is formed. The signal voltage is applied to G_4 and so controls the number of electrons that the plate is able to draw from this virtual cathode. Grid G_5 is a screen grid that serves to make the plate current substantially independent of plate voltage, thereby giving the tube a high plate resistance corresponding to that obtained with a screen-grid tube. The virtual cathode in front of G_4 forms with each pulse of space current and then disappears between

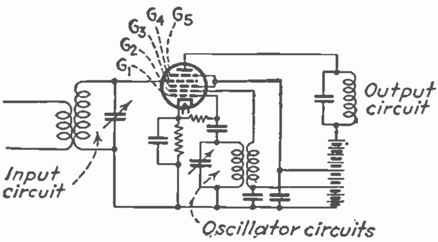


FIG. 141.—Circuit of typical pentagrid converter tube.

pulses. As a consequence the current actually arriving at the plate is modulated by both oscillator and signal voltages, giving a result equivalent to modulating the signal upon the oscillator frequency, thereby developing a difference frequency as one side band.

The difference-frequency output of a pentagrid converter is calculated in exactly the same manner as the output of the pentode plate detector and the hexode mixer. The conversion transconductance, and hence the amplification, are likewise of the same order of magnitude as obtained with other types of converters.

The pentagrid converter has the merit of simplicity in that it avoids the necessity of a separate oscillator tube, and also eliminates much wiring that would otherwise be required. Its usefulness is limited, however, by capacitive coupling between the signal grid G_4 and the space charge of the virtual cathode formed between this grid and G_3. This space charge pulsates at the oscillator frequency and induces oscillator-frequency currents in the signal circuits. This represents a form of coupling between signal circuits and oscillator that becomes more serious as the signal frequency is increased. The result is that the pentagrid converter, although satisfactory for signals of broadcast and lower frequencies, tends to be unsatisfactory at high frequencies.

81. Oscillating Detectors.—The oscillating detector consists of an ordinary amplifier with sufficient coupling between the plate circuit and the tuned input circuit to cause oscillations Typical circuit arrangements for oscillating detectors are shown in Fig. 142. These all consist of a simple oscillator circuit with grid-leak bias and means of controlling the coupling between the output of the tube and the parallel resonant input circuit that is connected between grid and ground, and that controls the frequency of oscillations.

When the adjustment is such that oscillations are barely able to exist, the amplitude of the oscillations is very sensitive to any additional voltage induced in the tuned input circuit. When the phase of the oscillations of such an induced signal is such as to add to the local oscillations, the amplitude of the voltage developed across the tuned input circuit increases greatly, whereas if the induced signal subtracts from the oscillations, the result is a marked decrease in amplitude. An induced signal having a frequency slightly different from that of the oscillations acts as though its phase were continually changing from addition to subtraction at a rate corresponding to the difference frequency. Hence when such a signal is induced in the input circuits of the oscillating detector, it causes the amplitude of the oscillations to pulsate at the difference frequency. This causes variations in the bias voltage developed by the grid leak, which in turn causes the current in the plate circuit to vary at the difference frequency.

The oscillating detector is used extensively in the reception of radiotelegraph signals. This is accomplished by adjusting the local oscillations to a frequency that differs by approximately 1000 cycles from the frequency of the incoming signal. In this way the frequency of the incoming signal is converted to 1000 cycles, and the telegraph characters are heard as dots and dashes having a pitch of 1000 cycles.

The oscillating detector has the advantage of simplicity and great sensitivity. It is easy to adjust because a signal is tuned in by merely varying the resonant frequency of the oscillator tuned circuit until an audible difference frequency appears. Because of the mechanism of operation, the sensitivity for weak signals is tremendous, a one-tube receiver consisting of an oscillating detector delivering as much output as can be obtained with perhaps two stages of good tuned radio-frequency amplification followed by a detector with a separate local oscillator.

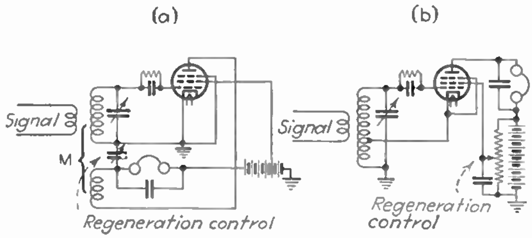


FIG. 142.—Typical circuits for regenerative and oscillating detectors. The telephone receivers indicated in the figure can be replaced by an amplifier when further amplification is desired.

The oscillating detector has two important limitations, however. In the first place, it is always possible to find two signal frequencies, one higher and one lower than the oscillator frequency, that will give the same frequency difference. In the second place, the oscillating detector cannot be used when the difference frequency is an appreciable percentage of the signal frequency, as for example in the superheterodyne receiver. When this percentage is not small, the tuned circuit that controls the frequency of oscillations is so far out of resonance with the incoming signal as to give very poor response.

82. Miscellaneous. *Square-law Detectors.*—The term *square-law* is applied to a detector which produces a rectified current that is proportional to the square of the applied signal voltage. This is in contrast with such detectors as the diode considered in Sec. 77, which is called a *linear detector,* in which the usual objective is to produce a rectified current directly proportional to the applied signal.

Square-law action results whenever a signal voltage is applied to a circuit in which the curve giving the relationship between current and voltage may be considered as a section of a parabola. Thus, when a vacuum tube is adjusted so that the operating point is on a curved part of the characteristic, as shown in Fig. 143, and a small signal is applied, the positive half cycles will be amplified more than the negative half cycles. This results in an increase in the plate current that is proportional to the square of the applied signal.

Square-law detectors find their principal use in vacuum-tube voltmeters (see Sec. 83). Square-law action is not desirable in the rectification of modulated waves, because the non-linear relationship between signal amplitude and rectified current results in distortion of the modulation envelope.

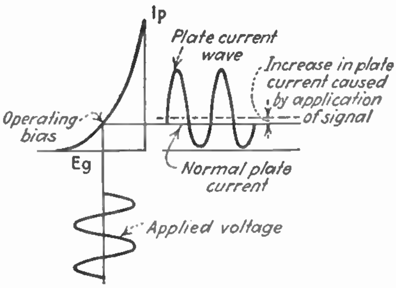


FIG. 143.—Vacuum tube adjusted to operate as a square-law device.

Every type of rectifier becomes a square-law device if the applied signal is sufficiently small. This is because in every rectifier the transition between the conducting and non-conducting condition is gradual rather than abrupt.

Grid-leak Detection.—This method of detection makes use of the non-linear relation existing between grid current and grid voltage in a tube operated at zero grid bias. A typical grid-leak detection circuit is shown in Fig. 144 and is roughly equivalent to a diode detector combined with a one-stage audio-frequency amplifier. The grid circuit in this arrangement acts exactly as a diode detector circuit, with the grid electrode taking the place of the diode plate, and the grid-leak–grid-condenser combination R_gC_g corresponding to the diode load impedance RC of

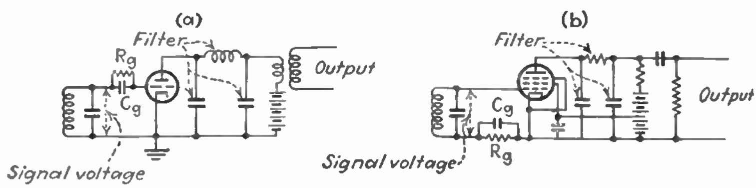


FIG. 144.—Circuits of grid-leak detector.

Fig. 131. When a signal voltage is applied to such a circuit, rectified current is produced that develops a voltage drop across the grid-leak resistance by the same mechanism involved in the diode detector. Since the voltage drop across the grid-leak resistance developed by the rectified current represents a potential difference existing between grid and cathode, this is amplified in the plate circuit to develop the useful output.

Compared with the diode detector, the grid-leak arrangement has the disadvantage of being able to handle only a limited signal voltage. This is because the tube operates with zero bias in the absence of signal, which limits the permissible plate-supply voltage. Also the radio-frequency signal voltage is applied to the input of the tube along with the rectified voltage developed across the grid leak and so tends to overload the tube. As a result, a tube with a separate diode is normally preferred to the grid-leak arrangement.

A grid-leak detector is a very satisfactory square-law detector of small signal voltages, and at one time it was universally used in radio receivers. However, with the development of satisfactory radio-frequency amplifiers, the need of a sensitive weak-signal detector ceased to exist.

Regenerative Detectors.—In the regenerative detector the amplitude of the applied signal voltage is increased by means of regeneration obtained by utilizing the amplifying action of the detector tube. The usual circuit arrangements are the same as those used with the oscillating detector, except that the coupling between the plate circuit and the tuned input circuit is reduced to the point where oscillations just cease to exist. In Fig. 142 grid-leak detection is indicated, although it is possible to employ plate detection. The effect of regeneration is equivalent to reducing the effective resistance of the tuned input circuit. Regeneration hence increases the voltage that a given signal will apply to the grid of the detector.

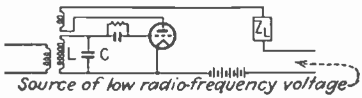


FIG. 145.—Simple superregenerative circuit.

Regeneration represents an inexpensive means of obtaining radio-frequency amplification. It has a number of disadvantages, however, and so with the development of satisfactory radio-frequency amplifiers has fallen into disfavor. Among these disadvantages are the excessive discrimination against the higher side-band frequencies that result from the reduction in effective resistance of the tuned input circuit, the critical character of the adjustments required to obtain an appreciable amount of regeneration, and the squeals that result when the regeneration is accidentally allowed to become sufficient to make the detector oscillate and heterodyne with incoming signals.

Superregenerative Detectors.—A superregenerative detector is a regenerative detector in which the regeneration is varied from an oscillatory to a non-oscillatory condition at a low radio-frequency rate. During the oscillatory interval oscillations build up, only to be subsequently suppressed or "quenched." The action is such that with proper adjustment an applied signal is amplified enormously before detection. A typical superregenerative circuit is shown in Fig. 145 and consists of a tube arranged to regenerate in the manner shown in Fig. 142*c* but supplied with a plate voltage that is a low radio frequency, such as 25 kc. Oscillations then build up during the half cycles when the plate is positive but die out (*i.e.*, are quenched) during the time the plate is negative.

The average amplitude of the oscillations existing across the tuned input circuit of the detector depends upon the signal voltage acting in this circuit. Hence an output current varying in accordance with the envelope of the applied signal is obtained by rectifying the voltage existing across the tuned input circuit.

A properly adjusted superregenerative detector gives very large amplification of weak signals. At the same time, the superregenerative detector is critical to adjust, produces a background hiss of considerable intensity in the absence of a signal, and has poor selectivity because of the side-band frequencies that must be accommodated as a result of the modulating action of the quenching voltage. The principal use of the superregenerative detector is in the reception of signals having a frequency so high that ordinary methods of amplification are not satisfactory.

83. Vacuum-tube Voltmeters.—A detector can be used as a voltmeter by making use of the rectified direct current to measure the voltage

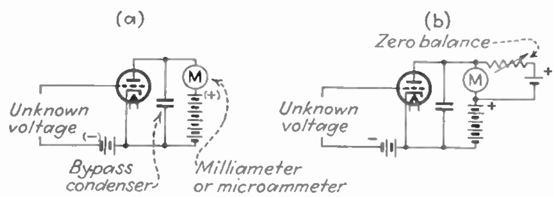


FIG. 146.—Typical vacuum-tube voltmeter circuits.

applied to the detector. The resulting vacuum-tube voltmeter is one of the most useful measuring devices available for audio and radio frequencies. It consumes little or no power from the voltage being measured, and when properly designed it can be calibrated at a low frequency, such as 60 cycles, and used at any radio or audio frequency up to ultra-high frequencies.

Circuit arrangements for two typical vacuum-tube voltmeters are illustrated in Fig. 146. The arrangement *a* is an ordinary anode detector in which the change in d-c plate current is used as a measure of the voltage applied to the grid. The circuit shown at *b* is a modification of *a* in which the residual d-c current that is present in the absence of signal voltage is balanced out of the meter by the arrangement shown, thereby enabling the indicating instrument to read the increment of plate current directly.

The relationship between the rectified current produced by a vacuum-tube voltmeter and the applied voltage depends upon the adjustment. Thus in Fig. 146*a* or 146*b*, a grid bias less than cut-off combined with a signal that is not too large, gives square-law action as shown in Fig. 143. On the other hand, if the bias is adjusted to cut-off as in Fig. 135, the negative half cycles are suppressed, and the response to the positive half cycles is proportional to the square of the applied voltage when this voltage is moderately large, but tends to be linearly proportional to the positive half cycles if the signal is large. This gives half-wave square-law, and linear action, respectively. Finally, if the grid bias is greater than cut-off, the rectified output tends to be determined by the peak amplitude of the signal.

Problems

1. In a diode detector an unmodulated voltage of 10 volts effective is applied. The load resistance is 500,000 ohms, and a microammeter shows that the rectified current in this resistance is 23 μa.

a. What is the efficiency of detection?

b. What is the input resistance of the detector?

2. Explain the detailed mechanism whereby the input resistance of a diode detector of good efficiency increases with load resistance R but is nearly independent of the diode plate resistance.

3. Explain why the amount the anode goes positive at the peak of each cycle in Fig. 131*d* will be more when (*a*) the diode plate resistance is increased and (*b*) the diode load resistance is decreased.

4. In the diode detector of Fig. 131*a*, $R = 250{,}000$ ohms, $C = 100$ $\mu\mu f$, and $\eta = 0.85$. Calculate and plot, as a function of modulation frequency up to 15,000 cycles, the highest degree of modulation that the applied modulated wave can have without introducing distortion.

5. In a diode detector, is the maximum permissible degree of modulation that the original signal can have without negative peak clipping increased or decreased by making the internal impedance of the source of the modulated wave high. Explain.

6. In a diode detector the input circuit has a resonant impedance of 160,000 ohms. Assuming an efficiency of rectification of 0.90, calculate and plot maximum permissible degree of modulation of original induced signal without negative peak clipping, as a function of the ratio of alternating-current to direct-current impedances of the diode load for ratios between 1 and 0.5, and for direct-current load resistances of 150,000, 300,000, and 600,000 ohms. Also calculate and plot the approximate r.m.s. distortion if the original induced signal is completely modulated.

7. Assume that a 50 per cent modulated carrier wave having an amplitude of 15 volts effective is applied to the diode detector of Fig. 134, and the efficiency of rectification is 0.90. Calculate (*a*) direct-current voltage developed for automatic-volume-control purposes, (*b*) modulation-frequency voltage developed across R_4 at moderate modulation frequencies, (*c*) radio-frequency voltage appearing across R_4 when the radio-frequency voltage across C_1 is 1 volt, and the frequency is 1000 kc.

8. In the detector circuit of Fig. 134, calculate the impedance the diode load circuit offers to the rectified current at modulation frequencies of 5000 and 10,000 cycles (assuming $C_1 = C_2 = 100$ $\mu\mu f$) and from this calculate the maximum degree of modulation the radio-frequency voltage applied to the diode input terminals can have at these modulation frequencies without distortion occurring in the rectified output.

9. Assume that the circuit of Fig. 134 is changed by making the resistance R_1 zero. With these new circuit conditions, calculate the r.m.s. distortion produced as a result

of negative peak clipping when the original carrier is fully modulated and compare with the distortion calculated in the example in the text.

10. In a diode detector, the reading of a direct-current microammeter located in series with the load resistance will not be affected by the presence or absence of modulation provided there is no clipping of the negative peaks. However, the microammeter reading will increase whenever there is clipping. Explain why this is the case.

11. Design a plate detector using a 56 tube with $E_b = 250$ volts and having a transformer-coupled output. In this design specify grid bias and maximum allowable carrier amplitude that can be handled, while allowing for complete modulation; estimate detector plate resistance and from this specify proper primary inductance of plate transformer for 70.7 per cent response at 60 cycles; and estimate maximum audio output voltage obtainable with reasonable detector efficiency and a transformer turn ratio of 2.

12. *a.* In a 6C6 pentode tube with $E_{sg} = 100$ volts, the control-grid bias for cut-off is -6 volts. When a carrier of 3 volts crest is applied, the d-c plate current has an average (or direct-current) value of approximately 0.5 ma. From this information, design a resistance-coupled plate detector using a 6C6 tube with $E_{sg} = 100$ volts and $E_b = 250$ volts. In the design specify coupling, grid-leak, and bias resistors and blocking-condenser capacitance.

b. Make a reasonable estimate as to conversion transconductance, calculate the mid-frequency amplification, and compare with the gain of a resistance-coupled amplifier using the same tube and plate-supply voltage, as given in Table VI, Chap. V.

13. Tubes with variable-mu characteristics are commonly employed in converters of the plate-detection type but are not recommended as ordinary plate detectors of modulated waves. Explain the reasons for this.

14. Would a pentagrid mixer tube operate if the signal were applied to grid G_3 and the local oscillator to grid G_1? Explain.

15. In an oscillating detector to which there is applied a reasonably strong signal, it is found that, as the resonant frequency of the oscillator tuned circuit is varied about the frequency of the signal, the pitch of the beat note first decreases until a low value such as 400 cycles is reached, then ceases until the oscillator circuit is tuned slightly to the other side of resonance, whereupon the low pitch reappears and becomes higher as the adjustment process continues. Explain the silence in the region where the oscillator tuned circuit is approximately in resonance with the signal.

16. Two sine waves having equations $e_1 = E_1 \sin (\omega_1 t + \Phi_1)$, and $e_2 = E_2 \sin (\omega_2 t + \Phi_2)$ are applied to a square-law detector. By the method of analysis given in Sec. 51 determine the various components of current that appear in the detector output.

17. In the vacuum-tube voltmeter circuits of Fig. 146*a* and 146*b*, what would be the result of omitting the by-pass condenser in the plate circuit?

18. *a.* Explain how a diode detector could be used as a vacuum-tube voltmeter by the addition of a microammeter.

b. Explain why the voltmeter in *a* would be a peak voltmeter.

CHAPTER X

SOURCES OF POWER FOR OPERATING VACUUM TUBES

84. Cathode Heating Power.—The most frequently used sources of power for heating the cathodes of tubes are commercial lighting circuits (both alternating and direct current), direct-current generators, storage batteries, and dry cells and other types of primary batteries. Commercial alternating-current power is used wherever possible because of its economy and simplicity. Direct-current generators are sometimes used in radio-telephone transmitters to eliminate the possibility of alternating-current hum. Storage batteries are employed in automobile and airplane

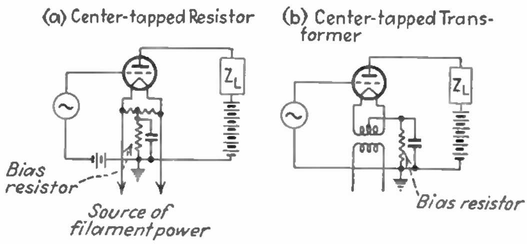


FIG. 147.—Methods of connecting the grid and plate return leads of filament tubes when alternating filament current is used.

radio equipment, and where very low hum is important. Primary batteries, such as dry cells and "air-cell" batteries, are used in portable equipment and where commercial power sources are not available.

Alternating-current Hum.—When 60-cycle a-c current is used to heat the cathode of a tube, there is always the possibility of introducing 60-cycle and 120-cycle components into the plate current, and also of modulating at these frequencies any signal voltages being amplified. These effects give rise to what is commonly termed *alternating-current hum,* since the result is a low-pitch hum appearing in the loud-speaker.

In filament tubes 60-cycle hum can be eliminated by bringing the grid and plate return leads to a point that has the same potential as the center of the filament. This can be done either by means of a center-tapped resistance across the filament as shown in Fig. 147*a*, or by a center-tap on the filament transformer as in Fig. 147*b*. However, there still remains a residual 120-cycle hum resulting from the magnetic field produced by the filament current and from the fact that the filament is not an equipotential surface.[1]

The use of heater-type tubes reduces the hum to a low value. There is, however, still a residual hum resulting from such influences as the magnetic field of the heater current, electrostatic fields from unshielded portions of the heater and heater leads, etc.

85. Grid-bias Voltage.—The grid bias for voltage-amplifier tubes is ordinarily derived from the plate-supply voltage by means of a self-bias resistance as illustrated in Figs. 55 and 148. This arrangement produces a bias by making the cathode positive with respect to the grid. The bias voltage is equal to the product of bias resistance and total space current and is accordingly controlled by the amount of resistance. In the case of pentode, screen-grid, and beam tubes, it is necessary to keep in mind that the total space current includes the screen current.

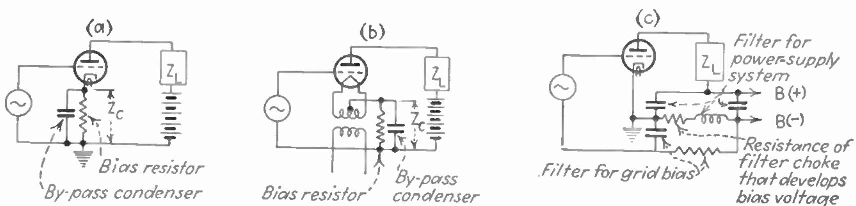


FIG. 148.—Methods of using the plate-supply voltage to make the grid negative with respect to the cathode.

The bias resistance must be by-passed with a condenser such that the amplified currents flowing in the plate circuit will produce negligible voltage between cathode and ground. Any voltage developed across the bias resistance by the amplified currents is superimposed upon the applied signal and so modifies the amplification. *In order for the by-passing of the self-bias resistance to be adequate, the voltage developed across the bias impedance by the amplified plate current must be small compared with the signal voltage being amplified.*[2] There is no difficulty in meeting this requirement at radio and the higher audio frequencies, but at low audio frequencies the capacitance required tends to become large.

[1] It might be thought that the use of a-c current to heat a filament would introduce hum as a result of temperature variations of the cathode. Actually, however, the ratio of heat energy stored in a filament to the rate of heat energy radiated is such that, with 60-cycle heating current, temperature fluctuations are so small as to introduce negligible hum.

[2] If the by-pass condenser is omitted, the voltage developed across the bias resistance by the amplified plate current opposes the signal voltage, and introduces negative feedback, such as discussed in Sec. 48. This method of obtaining negative feedback is often employed in resistance-coupled amplifiers in preference to an arrangement of the type shown in Fig. 77.

The grid-bias voltage is sometimes obtained from the resistance drop in the plate filter inductance. This can be done with the circuit arrangement of Fig. 148*c*, in which the filter choke is placed in the negative lead of the plate-supply system, and a resistance-condenser combination used to prevent ripple voltages developed across the filter choke from reaching the grid of the tube. Such an arrangement is often used with power-amplifier tubes, because it requires less plate-supply potential than do self-bias arrangements.

The grid bias for very large tubes, such as water-cooled tubes, is obtained in a variety of ways. Direct-current generators, self-bias resistances, rectifier-filter systems, and grid-leak arrangements are all employed to some extent.

Battery bias, which was once universally employed in all vacuum-tube amplifiers, is now used only in laboratory equipment or where voltages for plate and cathode are also obtained from batteries.

86. Rectifiers for Supplying Anode Power.—The anode power required by vacuum tubes is ordinarily obtained by rectifying commercial 60-cycle power and using a filter system to convert the rectified output into substantially pure d-c current. The rectifiers for this purpose are usually two-electrode tubes, of either the high-vacuum type or hot-cathode mercury-vapor type, according to the circumstances.

High-vacuum Thermionic Rectifiers.—The high-vacuum thermionic rectifier consists of a vacuum tube containing an electron-emitting cathode surrounded by an anode or plate electrode. Such a two-element tube acts as a rectifier because it will pass current only when the plate is positive with respect to the cathode and so, when placed in series with an alternating supply voltage, will permit current to flow in only one direction.

The important characteristics of the high-vacuum thermionic tube are the allowable peak plate current and the maximum allowable peak inverse voltage. The peak plate current represents the maximum electron emission that the cathode can be counted upon to supply during the useful life of the tube while maintaining a full space charge. Since the rectifier never allows current to flow for more than half the time, the average plate current, *i.e.*, the d-c output current, will never exceed one-half the peak plate current, and may be less. The maximum allowable peak inverse voltage is the largest negative voltage that may be applied to the plate with safety, and it determines the direct-current voltage that can be obtained from the rectifier tube. The exact relationship between direct-current output voltage and the actual inverse voltage depends upon the rectifier circuit employed, but in general the inverse voltage will be at least as great as the direct-current voltage, and in certain rectifier connections will be π times as great.

High-vacuum thermionic rectifier tubes are constructed in much the same way as the corresponding power tubes. In fact, some types of rectifier tubes are merely standard filament-type three-electrode tubes with the grid omitted. The principal exception to this is in the case of small rectifiers used in supplying anode power for radio receivers, where the low inverse voltages encountered permit special constructions that place the plate very close to the filament. The characteristics of a number of representative high-vacuum thermionic rectifiers are shown in Table XIV.

TABLE XIV.—CHARACTERISTICS OF TYPICAL HIGH-VACUUM THERMIONIC RECTIFIER TUBES

Rating				Filament data			
Type	Maximum allowable peak plate current, milli-amperes (approx.)	Maximum safe inverse voltage	Voltage drop with one-half peak current	Volts	Amperes	Watts	Type
80*	250	1,400	62	5	2	10	Oxide filament
5Z3*	500	1,400	61	5	3	15	Oxide filament
83V*	400	1,100	22	5	2	10	Heater
84*	120	1,000	17	6.3	0.5	3.2	Heater
81	340	2,000	120	7.5	1.25	9.4	Oxide filament
217A	600	3,500	210	10	3.25	32.5	Thoriated filament
836	1,000†	5,000	110	2.5	5	12.5	Heater
214‡	7,500	50,000	2,000	22.0	52.0	1,144	Tungsten filament

* These tubes have two cathodes and two anodes, and so are essentially two half-wave rectifiers in one envelope. The allowable plate current and voltage drop are given for a single anode, but the filament data are for both filaments.

† The maximum allowable *average* plate current in this tube is 0.25 amp.

‡ This is a water-cooled tube now becoming obsolete.

The Hot-cathode Mercury-vapor Rectifier.—The hot-cathode mercury-vapor rectifier is essentially a high-vacuum thermionic rectifier that contains mercury vapor in equilibrium with liquid mercury. The presence of the mercury vapor makes the characteristics very different from those of a high-vacuum rectifier tube, as explained in Sec. 33. When the plate potential of a hot-cathode mercury-vapor tube reaches 15 to 20 volts, ionization of the mercury takes place and the tube is able to draw the full electron emission of the cathode without further increase in plate potential. This is because the positive ions produced by the collisions of the electrons with the mercury vapor drift back toward the cathode and neutralize the negative space charge that would otherwise surround the cathode. The full space current can then be attracted to the plate with a plate potential that is merely sufficient to maintain the ionization.

Since the positive ions eventually fall into the cathode, it might be thought that the cathode life of hot-cathode mercury-vapor tubes would be very short. Experiments have shown, however, that if the plate is never allowed to become more than 22 volts positive with respect to the cathode, the positive-ion bombardment of the cathode produces no injurious effect.

The important characteristics of the hot-cathode mercury-vapor rectifier are the maximum allowable peak plate current, the maximum permissible average anode current, and the maximum safe inverse plate voltage. The peak plate current is determined by the electron emission that can be obtained from the filament and is unaffected by the presence of the mercury vapor. The safe average current is determined by the allowable heating of the plate. The maximum safe inverse plate voltage is the sparking voltage through the low-pressure mercury vapor and is less than for a corresponding high-vacuum tube.

The hot-cathode mercury-vapor tube must be operated with much more care than high-vacuum rectifiers. Thus the cathode of such a tube should be brought to full operating temperature before the plate voltage is applied, for otherwise the voltage drop in the tube during the warming-up process will exceed the cathode-disintegration value, and the cathode will be permanently damaged. It is also necessary to avoid short circuits and momentary overloads because permanent damage to the cathode will result if the voltage drop exceeds 22 volts even for only a short time.

The distinctive constructional features of a hot-cathode mercury-vapor tube are the relatively small envelope, relatively small plate, and an oxide-coated cathode having a filament voltage never over 5 volts. The small envelope and plate can be used even in tubes having the highest anode current ratings, because of the low power loss in the tube. The plate is usually in the form of a cup fitting over the cathode as in Fig. 149, because this reduces the tendency to flash back and also shields the plate-cathode space from external electrostatic fields. Oxide-coated cathodes are used because of their high efficiency. Low filament voltages are necessary because the total voltage drop in the tube is normally only about 15 volts, and the filament potential should be small compared with this.

Compared with the high-vacuum rectifier, the hot-cathode mercury-vapor tube has the advantage of higher efficiency, better voltage regulation, lower filament power, and low first cost. At the same time, the mercury-vapor tube has a limited inverse-voltage rating, displays a

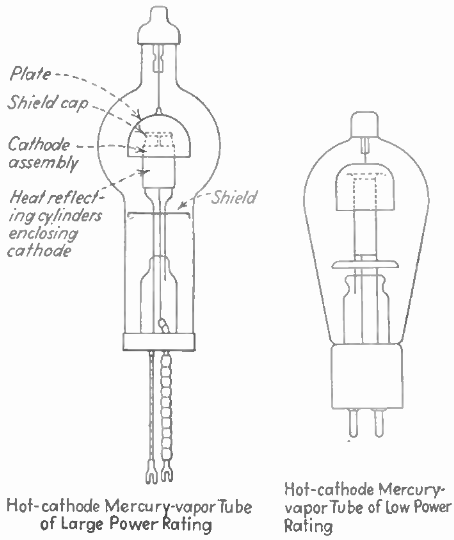


FIG. 149.—Typical hot-cathode mercury-vapor tubes.

tendency to flash back, produces radio-frequency transients, and will suffer damage to the cathode as a result of momentary overloads. When these considerations are balanced together, the result is that the hot-cathode mercury-vapor tube is found to be best for use with transmitters,

TABLE XV.—CHARACTERISTICS OF TYPICAL HOT-CATHODE MERCURY-VAPOR TUBES

Type	Rating			Filament data		
	Maximum allowable peak plate current, amperes	Maximum allowable average plate current, amperes	Maximum safe inverse voltage	Volts	Amperes	Watts
866	1.0	0.250	7,500	2.5	5	12.5
866A	1.0	0.250	10,000	2.5	5	12.5
872	5.0	1.25	7,500	5	10	50
872A	5.0	1.25	10,000	5	6.75	33.75
869A	10.0	2.5	20,000	5	18	90
857	40.0	10	22,000	5	30	150
870	450	75	16,000	5	65	325

where the large powers involved make economic considerations important, but not in radio receivers where ruggedness and freedom from radio-frequency transients are fundamental considerations.

Characteristics of typical hot-cathode mercury-vapor rectifier tubes are given in Table XV.

87. Rectifier Circuits.—Rectifiers operating from a single-phase source of power ordinarily employ either the center-tapped or bridge circuits shown in Fig. 150. Both these arrangements give full-wave rectification and deliver a voltage such as shown in Fig. 150*c* to a resistance load. The center-tapped circuit is usually preferred with tube rectifiers because then only two tubes are required, whereas the bridge circuit is normally used with copper-oxide rectifiers.

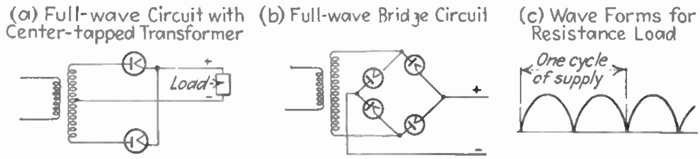


FIG. 150.—Rectifier circuits for operation with single-phase power sources together with wave form of voltage developed across a resistance load.

Polyphase Circuits.—When a polyphase source of alternating power is employed, the number of possible rectifier connections is almost unlimited, although only relatively few of these are of practical importance. The polyphase rectifier circuits most commonly used in radio work with three-phase power sources are shown in Fig. 151, and develop voltages across a resistance load that have the wave forms indicated in the figure.

In the three-phase half-wave rectifier the current is carried by the anode that is the most positive at that moment. Each of the three tubes accordingly carries the current one-third of the time, and the output voltage pulsates at three times the supply frequency as shown in Fig. 151*a*. In order to avoid direct-current saturation of the core of the transformer, it is necessary to employ a three-phase transformer rather than three single-phase transformers.

The circuit of Fig. 151*b* is essentially two three-phase half-wave rectifiers connected in parallel but with the polarity of each secondary in the second transformer reversed from the polarity of the corresponding secondary of the first transformer. In this way the output is relatively steady, and what pulsations there are have a frequency that is six times the supply frequency, as shown. The neutrals of the two three-phase half-wave rectifiers are connected together through a center-tapped inductance (often called an interphase reactor, or a balance coil) as shown in Fig. 151*b*, so that each three-phase rectifier operates independently. In this way there are always two tubes to carry current at any one time, whereas if the reactor were omitted only one tube at a time would carry current, and the output would be reduced accordingly.

The three-phase full-wave rectifier circuit shown at Fig. 151*c* gives the same output wave as does the double three-phase half-wave rectifier of Fig. 151*b* but differs in that the tubes are arranged so that full-wave rectification is obtained through each leg of the secondary winding. This circuit requires only one three-phase secondary and no interphase

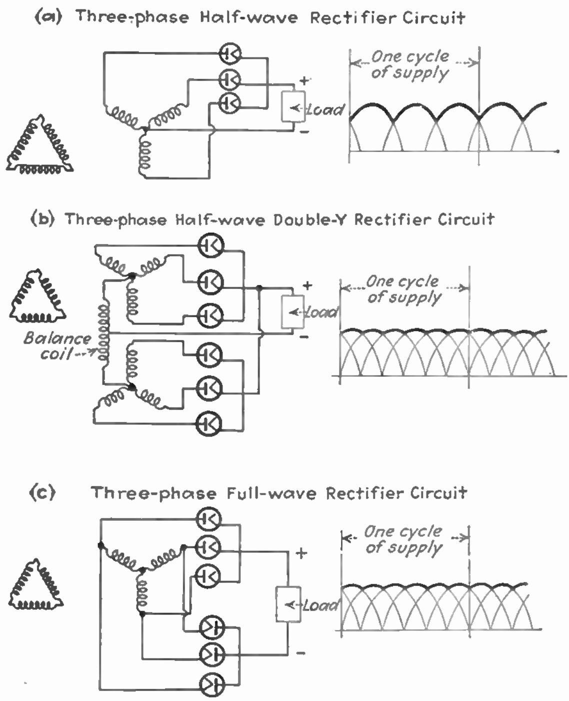


FIG. 151.—Rectifier circuits for operation with three-phase power sources together with wave forms of voltage developed across resistance load.

reactor, but the filament transformer for the rectifier tubes must have four separate secondaries.

Polyphase rectifiers are used where the direct-current power required is in the order of 1 kw or more. Compared with the single-phase circuits, the polyphase rectifiers develop an output voltage wave that is much closer to a steady direct-current potential, and the more desirable polyphase circuits give a higher output voltage in proportion to the peak inverse voltage and also utilize the possibilities of the transformer more effectively.

88. Filter Circuits Having a Series-inductance Input.—The pulsating voltage delivered by the rectifier output can be smoothed into a steady direct-current voltage suitable for applying to the anode circuit of a vacuum tube by being passed through an electrical network called a

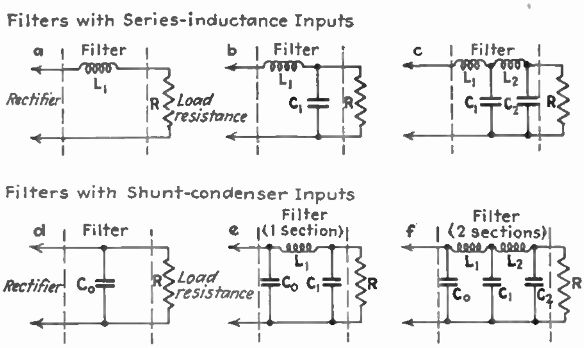


FIG. 152.—The filter circuits most commonly employed to smooth the pulsating rectifier output into a steady direct-current voltage.

filter, which ordinarily consists of series inductances and shunt condensers. The most commonly used filter circuits are those shown in Fig. 152 and may be divided into two general classes according to whether the input consists of a series inductance or a shunt condenser.

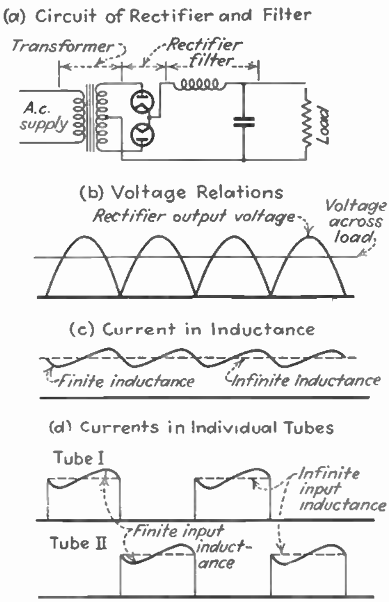


FIG. 153.—Oscillograms showing action taking place in filter having a series-inductance input, when supplied power from a single-phase full-wave rectifier. These curves are idealized in that they neglect transformer leakage reactance, tube drop, and the effect of energy losses in the filter inductances and condensers.

The action that takes place in a properly adjusted filter having a series-inductance input can be understood from an examination of the oscillograms of Fig. 153 for the case of a full-wave single-phase rectifier delivering its output to the filter of Fig. 152*b*. When the input inductance is infinite, the current through the inductance is constant and is carried by the anode that has a positive applied voltage. As the applied alternating-current voltage being rectified passes through zero, the current suddenly transfers from one anode to the other, giving square anode current waves as shown by the dotted lines in Fig. 153*d*. When the input inductance is finite and not too small, the situation is as shown by the solid lines of Fig. 153. The current through the input inductance then tends to increase when the output voltage of the rectifier exceeds the average or direct-current value, and to decrease when the rectifier output voltage is less than the direct-current value, as shown by Fig. 153*c*. This causes the current through the individual anodes to be modified, as shown by Fig. 153*d*. If the input inductance is too small, the current decreases to zero during a portion of the time between the peaks of rectifier output voltage, and the conditions then correspond to a condenser-input filter system, as discussed in the next section.

The current through the input inductance consists of a direct-current component upon which is superimposed an alternating component. The first shunt condenser following the input inductance tends to short-circuit this alternating component, and hence develops a potential across its terminals that is almost a pure direct-current voltage. However, if the residual fluctuations in voltage are greater than can be tolerated, an additional filter section as shown in Fig. 152*c* can be added to give additional reduction of the alternating voltage appearing in the output.

Analysis of Voltage Delivered by Rectifier to the Filter.—The action taking place in a filter having an input inductance of adequate size to maintain a continuous flow of current from the rectifier can be calculated with an accuracy sufficient for most practical purposes by considering that the rectifier applies to the filter input a voltage having the wave shape shown by the idealized curves of Figs. 150 and 151. This neglects the leakage reactance and resistance of the supply transformer, and the voltage drop in the rectifier, but is justified because both these factors are merely modifying influences in rectifier-filter systems of the type used in supplying anode power.

The idealized output wave of the rectifier may be considered as consisting of a direct-current component upon which are superimposed alternating-current voltages termed *ripple* voltages. Thus, in the case of the full-wave single-phase rectifier, the output wave has the equation.

Output voltage of single-phase full-wave rectifier =

$$\frac{2E}{\pi}\left(1 - \frac{2}{3}\cos 2\omega t - \frac{2}{15}\cos 4\omega t - \frac{2}{35}\cos 6\omega t \cdots\right) \qquad (107)$$

where E represents the crest value of the alternating-current voltage applied to the rectifier tube, and ω is the angular velocity ($2\pi f$) of the supply frequency. In this case the direct-current component of the output wave is $2/\pi$ times the crest value of the alternating-current wave. The lowest frequency component of ripple in the output is twice the supply frequency and has a magnitude that is two-thirds the direct-current component, and the remaining ripple components are harmonics of this lowest frequency component.

Table XVI gives analyses for the waves delivered by the rectifier connections of Figs. 150 and 151. It will be observed that in the three-phase half-wave rectifier the lowest ripple frequency is three times the frequency of the power supply, whereas in the three-phase full-wave rectifier it is six times that of the power supply. In all cases the amplitude of the ripple components diminishes rapidly as the order of the harmonic is increased. The ripple voltages are much less for the three-phase half-wave rectifier than for the single-phase connection and are still less for the arrangements of Figs. 151*b* and 151*c*.

Calculation of Direct-current and Alternating-current Components of Filter Output.—If the voltage drop in the rectifier, and the resistance and leakage reactance of the supply transformer are neglected, the direct-current voltage applied to the filter input is related to the alternating-current transformer voltage as indicated by line *a* of Table XVI. The regulation of the output voltage is then the regulation that would result from the filter resistances. Actually, the resistance and the leakage reactance of the supply transformer and the voltage drop in the rectifier tube slightly reduce the output voltage and make the regulation appreciably poorer, although Table XVI still gives the magnitude of the direct-current output voltage delivered to the filter input with an accuracy sufficient for most purposes.

The alternating-current voltage that appears across the output of the filter is the potential developed when the alternating-current ripple voltages given in Table XVI are applied to the filter input. Since the smoothing action of the filter results from the fact that the series inductance of the filter chokes out these alternating-current voltages, whereas the shunt condensers tend to short-circuit them, the output condenser must have a reactance that is low compared with the load resistance, and each inductance must have a high reactance compared with the reactance of the condenser that immediately follows it. Furthermore, the input inductance must also have sufficient reactance in relation to load resistance to satisfy Eq. (109) if current is to flow into the filter throughout the cycle.

An exact determination of the alternating-current voltage that appears across the output of the filter involves considerable labor because of the complicated electrical network involved, but for most purposes it is permissible to simplify the analysis by assuming that the reactance of each condenser is small compared with the reactance of the inductances immediately preceding and following the condenser, and that the reactance of the output condenser is small compared with the load resistance. The fraction of the alternating-current ripple voltage applied to the filter input that reaches the filter output is then given by the following equations.[1]

[1] The derivation of relations, such as those given by Eq. (108*a*), is as follows: The a-c current that flows in L_1 as a result of an applied voltage E is $E/\omega L_1$ if the

For the filter of Fig. 152c:

$$\frac{\text{Alternating-current voltage across load}}{\text{Alternating-current voltage applied to input}} = \frac{1}{\omega^4 L_1 L_2 C_1 C_2} \qquad (108a)$$

For the filter of Fig. 152b:

$$\frac{\text{Alternating-current voltage across load}}{\text{Alternating-current voltage applied to input}} = \frac{1}{\omega^2 L_1 C_1} \qquad (108b)$$

TABLE XVI.—CHARACTERISTICS OF RECTIFIERS OPERATED WITH A FILTER SYSTEM HAVING A SERIES INPUT INDUCTANCE

	Rectifier circuit				
	Single-phase, full-wave, center-tapped connection	Single-phase, full-wave bridge	Three-phase, half-wave	Double three-phase, half-wave	Three-phase, full-wave
Voltage Relations (Direct-current component of output voltage taken as 1.0):					
a. R.m.s. value of transformer secondary voltage (per leg)	1.11*	1.11	0.855	0.855	0.428
b. Maximum inverse voltage	3.14	1.57	2.09	2.09	1.05
c. Lowest frequency in rectifier output (F = frequency of power supply)	$2F$	$2F$	$3F$	$6F$	$6F$
d. Peak value of first three alternating-current components of rectifier output					
Ripple frequency	0.667	0.667	0.250	0.057†	0.057
Second harmonic of ripple frequency	0.133	0.133	0.057	0.014	0.014
Third harmonic of ripple frequency	0.057	0.057	0.025	0.006	0.006
Current Relations:					
e. $\frac{\text{Average anode current}}{\text{Peak anode current}}$	0.500	0.500	0.333	0.333	0.333
f. $\frac{\text{Average current per anode}}{\text{Direct-current load current}}$	0.500	0.500	0.333	0.167	0.333
Transformer Utilization Factors:					
g. Primary	0.900	0.900	0.827	0.955	0.955
h. Secondary	0.637	0.900	0.675	0.675	0.955

NOTE: This table assumes that the input inductance is sufficiently large to maintain the output current substantially constant, and neglects the effects of voltage drop in the rectifier and the transformers.

* Secondary voltage on one side of center tap.

† The principal component of the voltage across the balance coil has a frequency of $3F$ and a peak amplitude of 0.250.

reactance of C_1 is small compared with that of L_1. Practically all this current flows through C_1 because of the much higher reactance of the path through L_2, so that the voltage developed across C_1 is $E/\omega^2 L_1 C_1$. This potential is then applied to L_2 and C_2, and, if the reactance of C_2 is low compared with the load resistance, a repetition of the method used to get the voltage across C_1 shows that the alternating-current potential across C_2 is $E/(\omega^4 L_1 L_2 C_1 C_2)$, and Eq. (108$a$) follows at once.

For the filter of Fig. 152*a*:

$$\frac{\text{Alternating-current voltage across load}}{\text{Alternating-current voltage applied to input}} = \frac{R}{\sqrt{R^2 + (\omega L_1)^2}} \qquad (108c)$$

In these equations ω is the angular velocity ($2\pi f$) corresponding to the frequency of the component involved, and the alternating-current voltage applied to the input is given by Table XVI for different rectifier connections.

An examination of Eqs. (108) shows that the filtering action increases very rapidly with the number of filter elements, *i.e.*, the number of inductances and capacitances. The filter is also seen to be more effective the higher the frequency, and this, coupled with the fact that the largest component of the ripple voltage in the rectifier output is always the one having the lowest frequency, makes it permissible to neglect all frequency components in the rectifier output except the fundamental in ordinary calculations of the ripple voltage that appears across the load.

Factors Involved in the Design of Rectifier-filter Systems.—The input inductance of a filter should be of sufficient size to maintain a continuous flow of current from the rectifier. With polyphase circuits any inductance adequate for filtering will meet this requirement. However, with single-phase full-wave rectifiers, a continuous flow of current requires that[1]

$$L_1 \geqslant \frac{R_{\text{eff}}}{1130} \qquad (109)$$

where L_1 is the input inductance, R_{eff} is the effective load resistance, *i.e.*, the actual load resistance plus direct-current resistances of the filter inductances, and the supply frequency is 60 cycles. It is seen that, for a given inductance L_1, Eq. (109) will fail to be satisfied when the load resistance exceeds a critical value that is proportional to the input inductance.

When Eq. (109) is not satisfied, the action becomes similar to that existing with shunt-condenser filter systems, as described in the next section. The voltage across the filter input then depends on the load resistance, thereby causing the voltage regulation to be poor, and also producing other changes in behavior.

[1] In order to insure a continuous flow of current in the rectifier, the crest value of the alternating current flowing through the input inductance must not exceed the average or direct current. The direct current is equal to the direct-current voltage in the rectifier output divided by R_{eff}, and the crest alternating current in the first inductance is very nearly equal to the fundamental ripple-frequency voltage contained in the rectifier output divided by the reactance ωL_1 of the input inductance to this lowest frequency component of the ripple voltage. Equation (109) follows when this is applied to the single-phase case in Table XVI.

When the load resistance varies through a wide range it is often necessary to shunt a resistance, commonly called a bleeder resistance, across the output so that Eq. (109) will still be satisfied at light loads. Even then, with a bleeder drawing a reasonable proportion of the total current, the input inductance required will normally be larger than needed for suppressing ripple. An economical way of meeting this situation is to make the air gap of the input inductance small so that the incremental inductance is high with small currents. With large currents the incremental inductance drops, but Eq. (109) is still satisfied, and the filtering, though not so perfect as with a larger input inductance, is still acceptable. An input inductance adjusted to operate in this way is called a *swinging choke*.

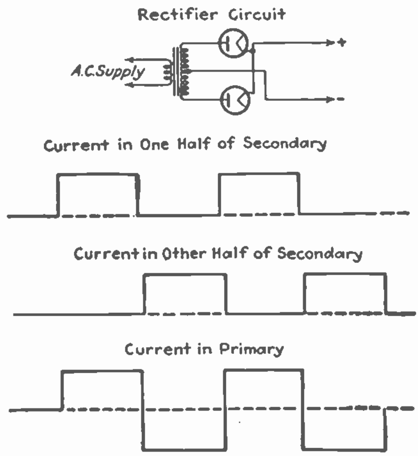


FIG. 154.—Wave shapes of current in primary and secondary windings of a center-tapped transformer supplying a single-phase full-wave rectifier operating into a filter with large series-input inductance.

The ratio of average to peak anode current depends primarily upon the rectifier connection and upon the size of the input inductance. Table XVI gives the results for the common circuits when the input inductance is infinite. When the input inductance is finite, the peak anode current is a little greater as a result of the variations in the inductance current, as seen in Fig. 153.

The maximum inverse voltage that the rectifier tube will be called upon to withstand depends upon the rectifier connections, and may vary from only slightly more than the direct-current output up to π times this potential. Values for the commonly used circuits are given in Table XVI.

A transformer used to supply power to a rectifier will normally run hotter than when delivering the same amount of energy to a resistance load because of the irregularly shaped current waves drawn by the rectifier. Thus in the case of the full-wave single-phase rectifier the current waves with a large input inductance have the shape shown in Fig. 154. The ratio of direct-current power output to the normal alternating-current rating for the same transformer copper losses is called the *utilization factor* of the transformer and depends upon the rectifier connections. Table XVI gives the utilization factor of the primary and secondary windings for some of the more commonly used rectifier connections.

By making use of Eqs. (108) and Table XVI, it is possible to design rectifier-filter combinations having an input inductance and to calculate everything about the performance except the voltage regulation[1] with an accuracy that is sufficient for most practical purposes. The method by which this is done is illustrated by the following example:

Example.—It is desired to design a three-phase half-wave rectifier-filter system to operate from a 60-cycle power supply and to deliver a direct-current output of 2500 volts and 0.4 amp. with a ripple that must not exceed 2 per cent.

If the direct-current resistance of the filter inductances is neglected, the rectifier must deliver a direct-current output voltage of 2500 volts, and Table XVI shows that the r.m.s. voltage which each secondary leg must develop is $2500 \times 0.855 = 2135$ volts. Since the utilization factors of the primary and secondary, as given by Table XVI, are 0.827 and 0.675, respectively, each leg of the primary must have a rating of

$$\frac{2500 \times 0.4}{(3 \times 0.827)} = 403 \text{ watts},$$

and each leg of the secondary a rating of $2500 \times 0.4/(3 \times 0.675) = 493$ watts. Tentative calculation based on Eq. (108*b*) shows that the filter of Fig. 152*b* with $C_1 = 1.0$ μf and $L_1 = 9.8$ henries will keep the ripple voltage down to 2 per cent and will be generally satisfactory. Reference to Table XVI shows that the peak anode current would be 0.4 amp for infinite input inductance, and the maximum inverse voltage that each rectifier must stand is $2500 \times 2.09 = 5225$ volts. Type 866 mercury-vapor tubes (see Table XV) will meet these requirements. In actual practice the secondary voltage of the transformer would be made greater than 2135 volts by perhaps 10 per cent to compensate for the loss of voltage caused by the resistance of the filter inductance, the voltage drop in the rectifier, the leakage reactance of the transformer, and the transformer resistance.

89. Filter Circuits Having a Shunt-condenser Input.—When the input to the filter is a shunt condenser, the action is somewhat different from that which takes place when a series inductance is used, as is apparent from the oscillograms of Fig. 155. Each time the crest alternating-current voltage of the transformer is applied to one of the rectifier anodes, the input condenser charges up to just slightly less than this peak voltage and then discharges into the first inductance until another rectifier anode reaches a peak potential, when the condenser is charged again. The rectifier current flows only a small part of the time, since during most of the cycle the condenser is more positive than all the anodes. During the discharge period the condenser voltage drops off nearly linearly because of the fact that the first filter inductance draws a substantially constant current from the input condenser. The result is that the input condenser applies a saw-toothed voltage wave to the first inductance.

[1] Methods have also been devised for calculating the voltage regulation, but these are so complicated as to be beyond the scope of this book. For further information see D. C. Prince and F. B. Vogdes, "Mercury-arc Rectifiers and Their Circuits," McGraw-Hill Book Company, Inc., New York.

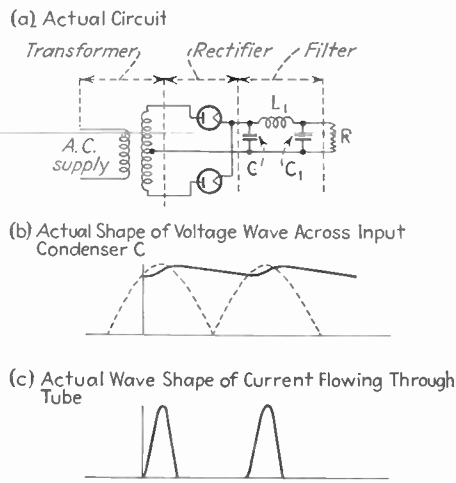


FIG. 155.—Circuit of condenser-input filter system together with oscillograms showing operation under typical conditions.

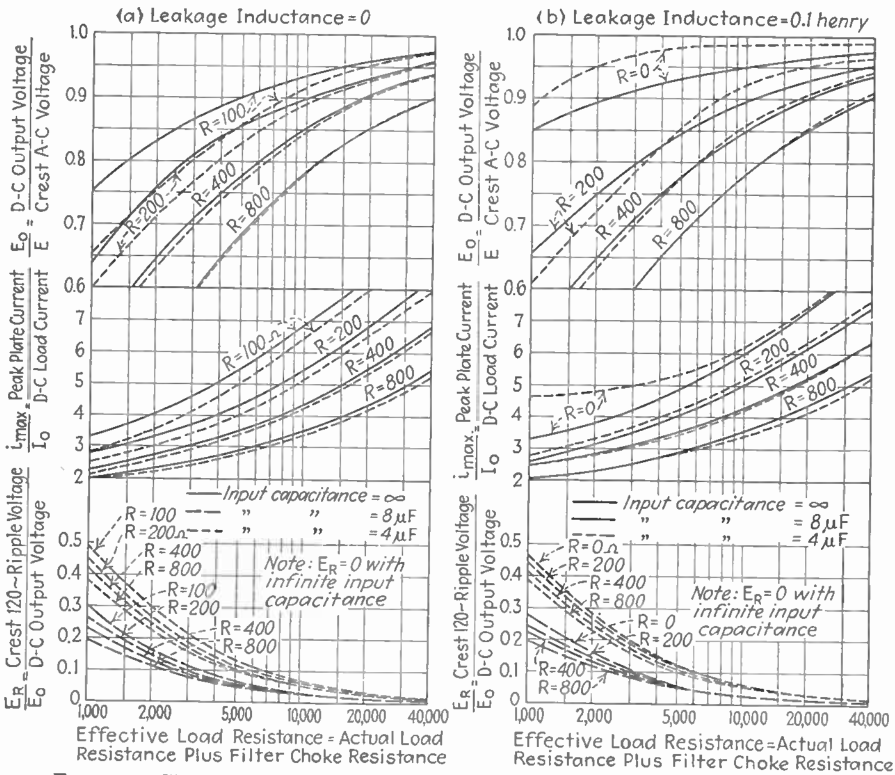


FIG. 156.—Charts giving performance of condenser-input filter systems for 60-cycle power source and full-wave rectifier. In these curves R is tube resistance plus transformer resistance of one half of secondary with primary shorted, and L is leakage inductance referred to half secondary with primary shorted.

This first inductance and the remainder of the filter then act to prevent the ripple component of voltage across the input condenser from reaching the load, in accordance with Eqs. (108).

An analysis of the voltage and current relations existing in the condenser-input filter system gives the results shown in Fig. 156,[1] in which the direct-current voltage across the input condenser, the peak space current through the rectifier tube, and the ripple voltage are expressed as a function of load resistance for various parameters, assuming a 60-cycle power source and the full-wave rectifier circuit of Fig. 155*a*.

The use of the curves of Fig. 156 in calculating the performance of condenser-input filter systems can be understood by the following example:

Example.—A load of 4000 ohms is to be supplied by direct current from a condenser-input rectifier-filter system. The power transformer develops a potential of 300 volts r.m.s. on each half of the secondary. The effective leakage inductance and resistance are 40 mh and 125 ohms, respectively, and a Type 80 rectifier tube is used. The input condenser is 4 μf, and is followed by a single-section filter consisting of a 10-h choke with 400 ohms resistance and an 8-μf condenser across the output.

The first step in the solution is the determination of the effective load and source resistances. The former is the actual load resistance plus the choke resistance, or 4000 + 400 = 4400 ohms. The average resistance of the rectifier tube is, from Table XIV, approximately 62/0.125 = 498 ohms, so that the source resistance is 125 + 498 = 623 ohms. Reference to Fig. 156 shows that, for these values of source and load resistances and input capacity, one has

	For $L = 0$	For $L = 0.1$	For $L = 0.040$
E_0/E	0.73	0.73	0.73
I_{max}/I_0	2.9	3.00	2.95
E_R/E_0	0.12	0.115	0.117

The last column is obtained by interpolation between the first two. From this column one has:

D-c voltage across input condenser $= 0.73 \times 300\sqrt{2} = 309$

D-c voltage across actual load $= 309 \times 4000/4400 = 281$

D-c current in load $= 309/4400 = 0.0703$ amp.

Peak space current $= 0.0703 \times 2.95 = 0.207$ amp.

Ripple voltage across input condenser $= 0.117 \times 309 = 36$ volts crest

Ripple voltage across load [from Eq. (108*b*)]

$$= \frac{36}{(754)^2 \times 10 \times 8 \times 10^{-6}} = 0.80 \text{ volt crest}$$

Discussion of Properties.—Examination of Fig. 156 shows that the direct-current voltage developed across the input of a shunt-condenser

[1] For details of the analysis see p. 492 of the author's book "Radio Engineering," 2d ed.

filter system depends primarily upon the load resistance. When this resistance is high the direct-current voltage approaches the crest amplitude of the applied alternating voltage that is being rectified. When the load resistance is reduced, the increased rate of discharge of the input condenser lowers the average voltage across the condenser, and if the resistance is very low the voltage will approach the value obtained with inductance input.

The ripple voltage developed across the input condenser of the filter is approximately inversely proportional to the input-condenser capacitance, and becomes less as the load impedance is increased, as is seen from Fig. 156. This is because a high load resistance, or a large input condenser, reduces the extent that the input-condenser voltage drops off during the period that this condenser discharges into the filter system. As the load resistance is varied from infinity to a relatively low value, the ripple voltage varies from zero to a value approaching that obtained with inductance input.

The ratio of peak to average current when a condenser-input filter system is employed is always greater than with inductance input, because of the fact that the current flows in the form of pulses, rather than continuously. The ratio of peak to average current with shunt-condenser input becomes greater the higher the load resistance, although the actual magnitude of peak current decreases as the load resistance increases. Because of the very peaked character of the current waves obtained when the rectifier delivers its output to a shunt-condenser system, the utilization factor of the transformer is relatively poor.

Compared with the series-inductance filter system, the shunt-condenser arrangement is seen from the above factors to give more direct-current output voltage from a given transformer, and to have less ripple voltage across the input of the filter. At the same time, however, the shunt-condenser system has much poorer voltage regulation, a higher peak plate current in proportion to d-c load current, and a poorer transformer utilization factor.

Filters with shunt-condenser inputs are generally employed in radio receivers, public-address systems, etc., when the amount of direct-current power required is small. An inductance-input arrangement is used whenever good regulation of the direct-current voltage is important, as in Class B amplifiers. Inductance-input systems are also used when the amount of power involved is large, because the higher utilization factor and lower peak current result in important savings in tube and transformer costs. Inductance-input systems are ordinarily employed in polyphase rectifiers.

90. Filter Systems. Miscellaneous Comments.—The condensers used in filters must be capable of *continuously* withstanding a direct-current voltage that is equal to the maximum direct-current voltage that can be delivered by the rectifier output. Ordinarily a single condenser should be used to withstand the entire voltage, rather than several condensers in series. When condensers are in series, the direct-current voltage stress divides between them in proportion to their leakage resistances rather than to their dielectric strength, and the leakage resistances are variable and uncertain.

The inductance coils used in the filter must have laminated iron cores with an air gap that is sufficient to prevent the direct current from saturating the iron. The inductance that is effective to the a-c currents, *i.e.*, the incremental inductance, will vary with the superimposed d-c current as discussed in Sec. 7, and will normally be greatly lowered by the presence of the d-c current.

In cases where it is necessary for the ripple voltage to be very small, as at the anode of the first tube of a high-gain audio-frequency amplifier, it is customary to use a resistance-condenser filter in the circuit of the individual tube, in addition to the regular inductance-condenser filter of Fig. 152. This is illustrated in the amplifier of Fig. 75, where the filters that are shown for the purpose of avoiding coupling from a common plate impedance also serve as resistance-condenser filters to reduce the ripple voltage applied to the plates and screens of the first two tubes. The reduction in ripple produced by an individual stage of resistance-condenser filtering is given by the equation

$$\frac{\text{Alternating-current voltage across load}\}}{\text{Alternating-current voltage applied to input}\}} = \frac{1}{R\omega C} \tag{110}$$

where it is assumed $R \gg 1/\omega C$, and the notation is

R = series resistance of filter

C = shunting capacitance of filter

ω = 2π times the ripple frequency involved.

91. Miscellaneous Supply Systems. *Vibrator Systems.*—Where the only available power is d-c current at low voltage, such as a storage battery in the case of automobile radio receivers, direct-current plate power may be obtained by using a vibrating contact to change the d-c current from the battery into a-c current. This is then stepped up in voltage by a transformer and rectifier to produce high-voltage direct-current power. A typical vibrator circuit is shown in Fig. 157, in which the vibrator operates on the same principle as the ordinary buzzer. The condenser C_1 is for the purpose of minimizing sparking at the contacts, while C_2 and the radio-frequency chokes are to isolate radio-frequency transients that are produced by the vibrator. The output voltage of the vibrator transformer can be rectified by a tube or by means of additional contacts mounted upon the vibrator armature, as in Fig. 157.

The vibrator represents an inexpensive means of obtaining plate power from a storage battery, but has the disadvantage of a limited life and of requiring a relatively complicated filtering and shielding system to prevent surges from producing interference in the radio receiver.

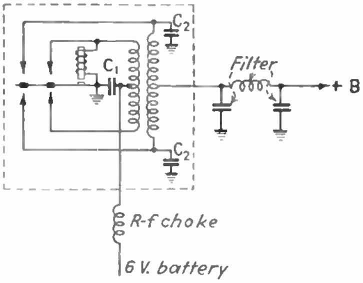


FIG. 157.—Circuit diagram of vibrator power supply.

Dynamotor.—A dynamotor is a combined motor and generator having two or more separate armature windings and a common field circuit. One of the armature windings is operated from a low-voltage direct-current power source, commonly a storage battery, to give motor operation, and the other windings then serve as generators to produce the desired direct-current voltages.

Dynamotors using permanent-magnet fields and operating from a storage battery are often used for supplying the anode power required by automobile radio receivers. Compared with the vibrator systems

FIG. 158.—Interrupted continuous waves generated by applying an alternating-current voltage directly to the plate of an oscillator.

the dynamotor has the advantage of longer life and less trouble from radio-frequency surges, but is considerably more expensive.

Self-rectifying Circuits.—If an alternating voltage is applied directly to the plate of an oscillator or power amplifier in place of the usual direct-current potential, the tube will operate on the positive half cycles of the applied voltage and so will generate wave trains having the character shown in Fig. 158. These are known as *interrupted continuous waves* (abbreviated I.C.W.) and are sometimes used in radio telegraphy. Another type of self-rectifying circuit is shown in Fig. 159, in which the two plates of a two-tube oscillator are supplied with alternating-current voltages that are 180° out of phase, and a large inductance is inserted in the common filament return

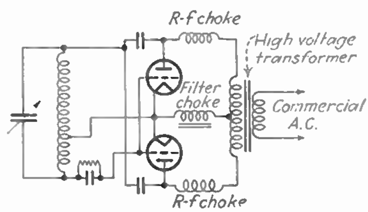


FIG. 159.—Self-rectifying oscillator circuit capable of generating oscillations of substantially constant amplitude if the filter choke is very large. Similar circuit arrangements may be applied to power amplifiers.

lead. When the inductance is large, the total current drawn by the two tubes is substantially constant, and the output tends to be constant.

Problems

1. Explain why alternating-current hum from a-c filament current is a much more serious problem in audio-frequency amplifiers than in radio-frequency amplifiers.

2. What bias resistance is required when a Type 6C6 tube is to be operated at conditions corresponding to those given in Table V, page 85?

3. Determine a suitable by-pass condenser for a resistance-coupled amplifier using a 6C6 tube and designed according to the fourth column of Table VI, Chap. V, assuming that the regeneration from this bias impedance is not to affect the amplification seriously for frequencies above 60 cycles.

4. Discuss the practicability of using a self-bias arrangement consisting of a resistor between cathode and ground for (*a*) a Class C amplifier, (*b*) a linear amplifier, and (*c*) a Class B audio amplifier.

5. In a push-pull Class A audio amplifier, explain why it is that no by-pass condenser is needed in shunt with the bias resistor between cathodes and ground.

6. Explain why it is customary to design a rectifier tube with a maximum permissible plate dissipation that is less than the plate dissipation that would result if the allowable peak current were drawn continuously through the tube by applying a direct-current anode voltage.

7. Redraw the circuit of Fig. 151*c*, lettering the secondary windings *A*, *B*, *C*, and numbering the tubes from 1 to 6. Then draw the voltage waves developed by each secondary and indicate on each voltage wave the portions of the cycle during which the corresponding secondary carries current and also the number of the tube that is involved each time a secondary carries current.

8. Check lines *a*, *b*, and *c* in the columns of Table XVI for the single-phase full-wave center-tapped and bridge circuits.

9. A particular single-phase full-wave rectifier-filter system with inductance input is required to develop 150 ma at 1000 volts. What is the minimum allowable input inductance, neglecting the resistance in the filter chokes?

10. Two Type 46 tubes are to be used in a push-pull circuit as Class B audio amplifiers with a plate potential of 400 volts. The plate current drawn by the combination with no applied signal is 12 ma and with full signal is 100 ma.

a. Determine the bleeder resistance that must be placed across the filter output if the total direct current that the rectifier tubes supply is not to vary over a range exceeding 4 to 1.

b. What is the minimum allowable input inductance with no applied signal, and with full signal assuming bleeder is used?

c. Assuming that the input inductance has the minimum allowable value for no signal conditions and reduces to one-third this value at full signal, what size condenser must be used in the circuit of Fig. 152*b* to make the ripple less than 1 per cent at full signal, assuming 60 cycle full-wave circuit?

d. What will be the ripple at no signal with the conditions as in *c*?

11. Draw curves similar to those of Fig. 154, but for the case of the full-wave bridge circuit of Fig. 150*b* with (1) infinite input inductance, and (2) finite input inductance.

12. Draw curves corresponding to those of Fig. 154 showing the current waves flowing through each tube in the circuit of Fig. 151*a* for (*a*) infinite input inductance and (*b*) finite input inductance.

13. Design a power-supply system for operating a push-pull Class C amplifier using Type 852 tubes (allowable plate loss 100 watts per tube) and operating at $E_b = 3000$

volts. The maximum allowable ripple voltage is 2 per cent and the available power is 60 cycle, single phase. Use hot-cathode mercury-vapor rectifier tubes.

14. The power supply for the final power amplifier of a high power radio transmitter employs six Type 869-A hot-cathode mercury-vapor tubes in a three-phase full-wave circuit operating from 60 cycles. The power supply is rated at 18,000 volts and 4 amp., and has a single-section inductance-input filter consisting of 0.5 henry and 3.0 μf. Neglecting the resistance of the filter inductance and the leakage reactance and resistance of the power transformer, calculate (*a*) approximate peak current through the rectifier tubes, (*b*) peak inverse voltage, (*c*) secondary voltage of power transformer, (*d*) kva rating of transformer primary and secondary windings, (*e*) amount of 360-cycle and 720-cycle ripple voltage appearing in the output.

15. In the power-supply system of Prob. 14, determine total power loss in the six rectifier tubes, including both filament power and plate dissipation, and from this calculate the rectifier efficiency. In evaluating the plate loss assume a tube drop of 12 volts.

16. Draw oscillograms corresponding to those of Fig. 155*b* and 155*c* for two load resistances, one high and one low.

17. *a.* In a particular condenser-input rectifier-filter system the power transformer develops a potential of 375 r.m.s. volts across half the secondary, while the effective transformer resistance and leakage inductance referred to one-half the secondary are respectively 92 ohms and 28 *mh*. If a 4-μf input condenser is employed with a 5Z3 tube, and if the effective load resistance is 2500 ohms, determine direct-current voltage and ripple voltage across the input condenser, and the peak anode current.

b. Design a filter system to follow the input condenser that will make the ripple in the filter output less than 0.01 volt.

18. In the example of Sec. 89, calculate the performance when the shunt input capacitance is 8μf, and infinity, tabulate the results along with the results of the example, and discuss briefly the effect of input-condenser capacitance.

19. Derive Eq. (110).

20. In a vibrator supply system such as shown in Fig. 157 sketch oscillograms showing, (*a*) voltage across each half of the primary, (*b*) voltage across the secondary winding, and (*c*) current flowing in each half of the primary.

21. Describe the operation of the self-rectifying circuit of Fig. 159, with particular reference to the action of the choke in the common lead. In this discussion sketch oscillograms illustrating the general character of the plate voltage and d-c plate current in each tube, and the current in the choke.

CHAPTER XI

RADIO TRANSMITTERS

92. Radio Transmitters. *General Considerations.*—A radio transmitter is essentially a device for producing radio-frequency energy that is controlled by the information to be transmitted. A transmitter accordingly represents a combination of oscillator, power amplifiers, harmonic generators, modulator, power-supply systems, etc., which will best achieve the desired result. Inasmuch as the various elements that go to make the completed transmitter have already been individually considered, the present chapter is devoted to a discussion of typical combinations used in practice.

The electrical design of practically all radio transmitters is dominated by the need of maintaining the transmitted frequency as nearly constant as possible over long periods of time. Thus in broadcast work two or more transmitters are commonly assigned the same carrier frequency, and in order to minimize the resulting interference it is essential that the carrier frequencies be as nearly as possible the same. In short-wave work, the side bands required are so narrow that the frequency separation between transmitting stations is determined primarily by the frequency stability of the carriers. For example, telegraph transmission at a speed of 100 letters per minute (hand keying) could, according to Sec. 68, be carried on with a side band 13 cycles wide. With perfect frequency stability this would mean that two such transmitters could theoretically operate on frequencies differing by less than 30 cycles. A very limited space in the frequency spectrum would then accommodate a large number of transmitters. Actually, however, it may be impractical to maintain the frequency closer than 0.05 per cent to the desired value under all conditions. At 20 mc this would require a spacing of 20,000 cycles. The number of transmitters that can be accommodated by a given section of the frequency spectrum in the short-wave range is hence largely determined by the frequency stability that can be maintained, rather than by the side-band requirements.

Transmitters are normally mounted in a metal cabinet provided with metal shelves and compartments that shield the different parts of the transmitter from each other and at the same time provide mechanical support. The controls and meters are mounted on the front panel and arranged to be at ground potential. The back, sides, and occasionally the front of the cabinet are either removable or hinged in order to make the transmitter accessible for repair or inspection. Typical examples of this type of construction are found in Figs. 162 to 164 and Fig. 168.

93. Telegraph Transmitters. *Short-wave Transmitters.*—Short-wave telegraph transmitters are ordinarily designed around some form of oscillator having high frequency stability. Crystal oscillators are usually employed, although master-oscillator power-amplifier arrangements, as well as electron-coupled and resonant-line oscillators are used to a limited extent.

When a quartz crystal is used to control the frequency, the oscillator power output obtainable under commercial conditions is only a few watts, and the highest practical crystal frequency is of the order of 5 mc. To obtain greater power Class C amplifiers are required, and higher fre-

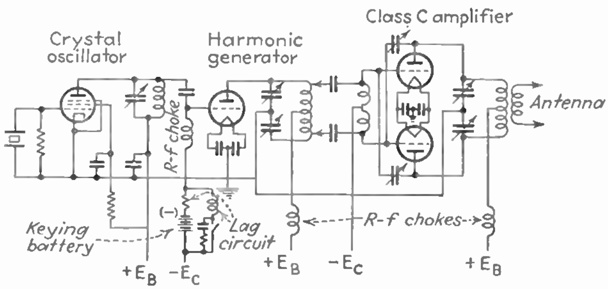


FIG. 160.—Simple short-wave crystal-controlled telegraph transmitter showing keying system provided with time-lag circuit.

quencies necessitate the use of harmonic generators. The circuit of a simple crystal-controlled short-wave transmitter is illustrated in Fig. 160. Here the crystal oscillator employs the circuit of Fig. 108*a* and is followed by a harmonic generator that drives a push-pull Class C amplifier. This represents a typical amateur design for a transmitter developing an output of the order of 50 to 200 watts of power.

The schematic circuit of a commercial transmitter is shown in Fig. 161. This is designed to operate at any frequency in the range 1.5 to 20 mc and delivers an output of 350 watts. Provision is also included for modulating the final power amplifier to give phone transmission with 250 watts carrier, but this has been omitted from the circuit diagram for the sake of simplicity. The transmitter of Fig. 161 consists of a crystal oscillator employing the circuit of Fig. 108*c*, followed by three intermediate or buffer power amplifiers, and a final Class C output amplifier. The intermediate power amplifiers amplify the crystal output sufficiently to excite the final amplifier, function in some cases as harmonic generators, and isolate the crystal oscillator from the output circuits of the transmitter. The crystal operates in the frequency range 1.5 to 5 mc, and for

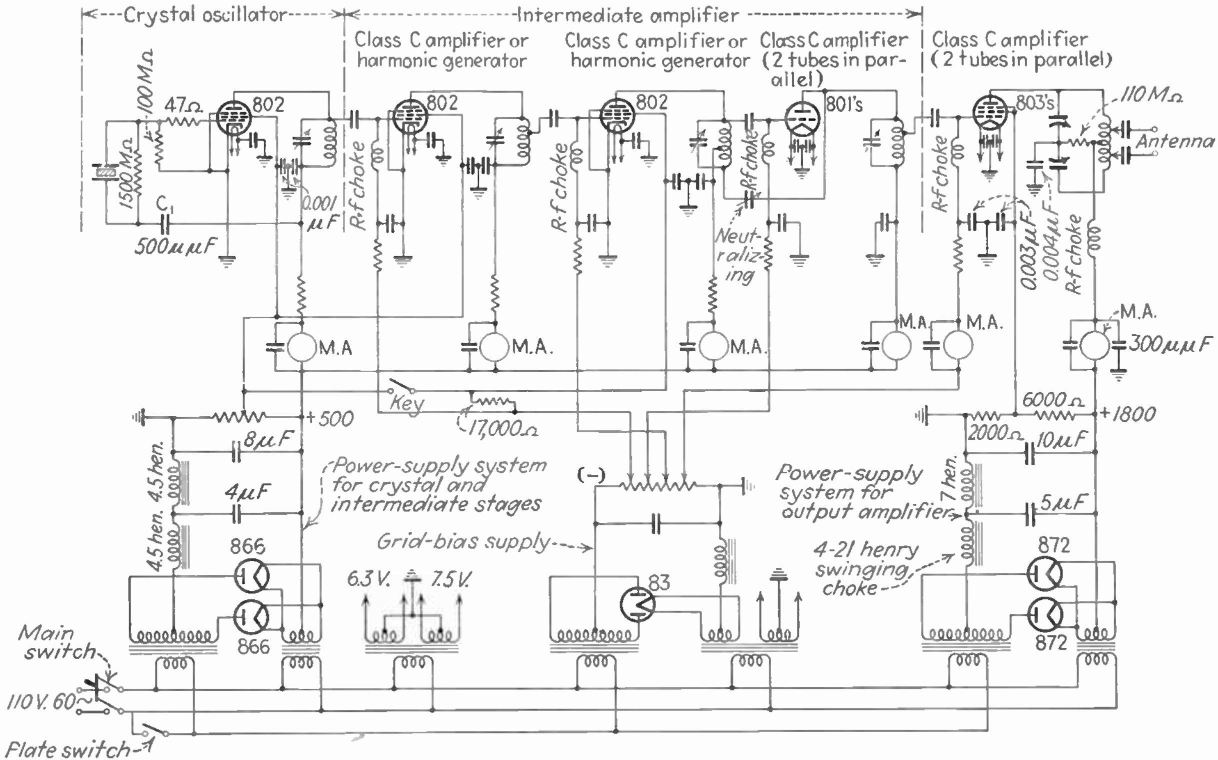


FIG 161.—Schematic circuit diagram of 350-watt commercial short-wave radio-telegraph transmitter.

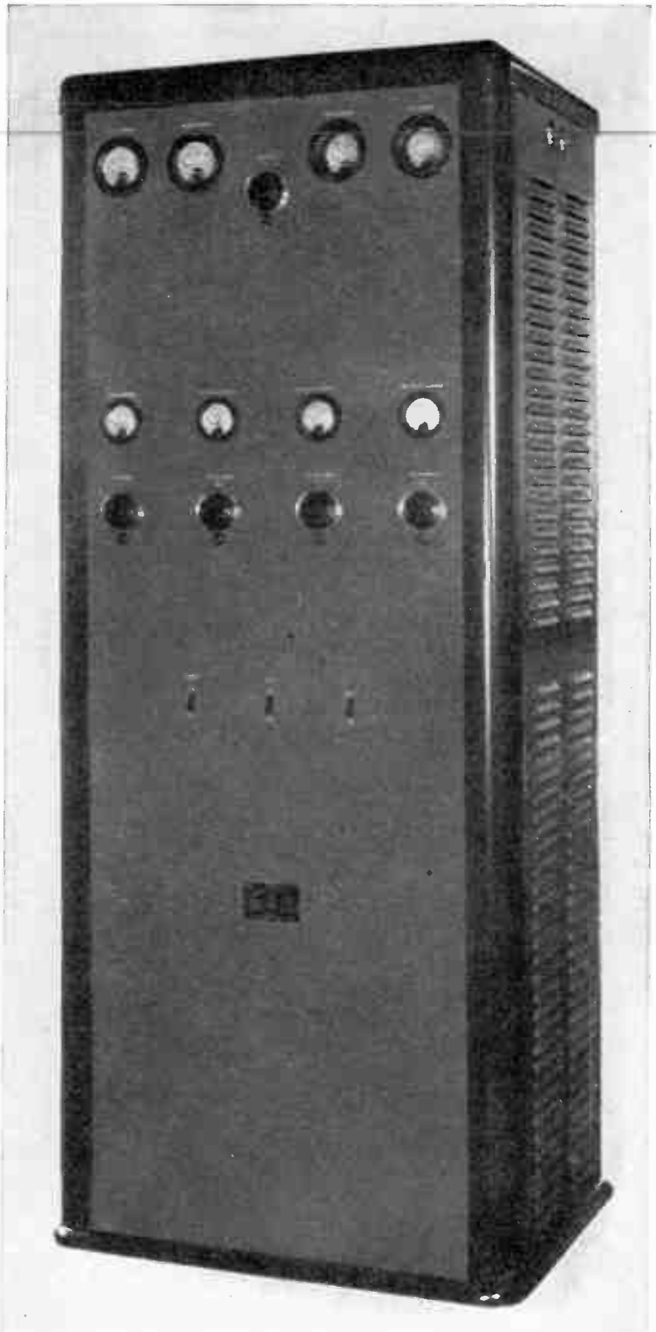


Fig. 162.—Front view of cabinet housing transmitter of Fig. 161, showing controls and meters.

Fig. 163.—Rear view of cabinet of Fig. 162 with back opened showing the metal-shelf and tray construction and how the essential leads of each compartment are brought to a terminal strip.

frequencies in this range all intermediate stages operate as Class C amplifiers. For frequencies in the range 5 to 10 mc, the first buffer amplifier acts as a frequency doubler, whereas in the range 10 to 20 mc the first two buffer tubes act as frequency doublers. The various resonant circuits are adjusted to the appropriate frequencies by means of variable condensers, plug-in coils, and in some instances, by means of supplementary condensers that can be added by link connections.

The general method of construction of this transmitter is illustrated in Figs. 162, 163, and 164. This shows the metal cabinet arranged with shelves and trays typical of present-day methods. Such arrangement provides effective shielding and at the same time makes the wiring, coils, and essential circuit points accessible.

The chief difference between the transmitters of Figs. 160 and 161 is that the isolation between input and output stages of the latter is greater. If amplification were the only consideration, the transmitter of Fig. 161 could dispense with at least one and possibly two of the buffer amplifiers. If this were done, however, keying would produce a certain amount of frequency shift that would produce adjacent channel interference.

When more power is desired, arrangements such as those of Figs. 160 and 161 can be used to excite large air-cooled tubes that will give outputs up to several kilowatts. These in turn can be used to operate Class C amplifiers employing water-cooled tubes to develop powers of the order 20 kw and up.

Although the arrangements illustrated in Figs. 160 and 161 are typical of equipment in actual use, the circuit details can be modified in many respects. Thus the frequency range in which the crystal operates may be different, necessitating different combinations of harmonic generators. It is also possible to vary the resonant frequency of the tuned circuits by tapped coils controlled by switches instead of using plug-in coils. In low-frequency transmitters fixed condensers are often used in the tank circuits, with the resonant frequency being varied by shorting out turns of the inductance coils for coarse adjustments and rotating a short-circuited ring for fine adjustments. The method of neutralization, tube types employed to accomplish a given purpose, and so on, will also vary greatly with the individual design.

In circumstances where it is necessary to obtain a continuous frequency range instead of having available only certain predetermined frequencies, the crystal oscillator can be replaced by an electron-coupled oscillator. The frequency of this oscillator can be adjusted either to the frequency to be transmitted, or to one-half, one-third, one-fourth, etc., of the required frequency. The electron-coupled oscillator is followed by harmonic generators and Class C amplifiers as required to give

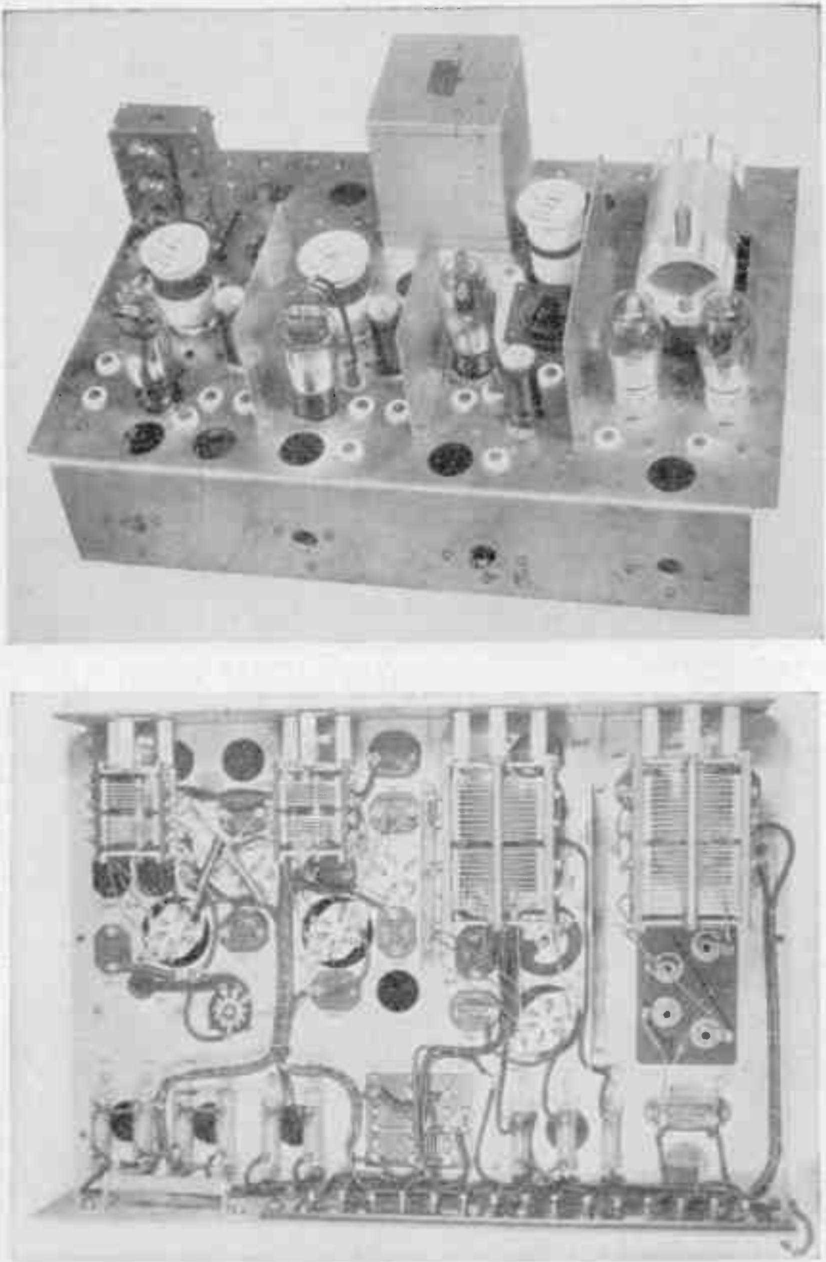


FIG. 164.—Top and bottom views of tray containing crystal oscillator and the three intermediate or buffer stages. Note the plug-in coils and the shields between stages as well as the method of mounting small parts and the cabling of wiring.

the desired frequency and power and to provide adequate isolation of the oscillator from the output circuits.

Long-wave Telegraph Transmitters.—Telegraph transmitters operating at frequencies below the broadcast band (*i.e.*, below 550 kc) are commonly of the master-oscillator power-amplifier type. This is permissible because the frequency stability required to keep within a given number of cycles of the assigned frequency is directly proportional to the carrier frequency and so is low at low frequencies. The use of a master-oscillator arrangement in place of crystal control has the advantage of simplifying the transmitter layout and makes it easier to change frequency.

An example of a commercial marine transmitter for operation in the frequency range 500 to 375 kc is shown in Fig. 165. Here a single Type 211 tube functions as a master-oscillator in a modified Colpitts circuit, while three similar tubes in parallel serve as a power amplifier to develop an output of 300 watts. By neutralizing the power amplifier, using an oscillator with a high Q tank circuit, and operating the power amplifier so that the driving power required is small, the frequency stability is adequate for marine service at long waves.

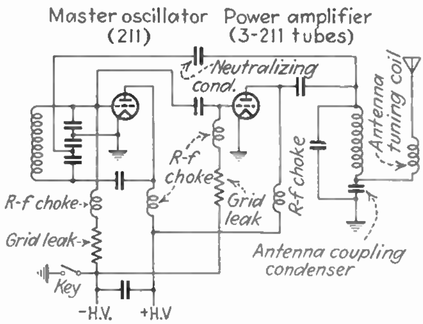


FIG. 165.—Commercial master-oscillator–power-amplifier radio-telegraph transmitter capable of delivering 300 watts output in the frequency range 375 to 500 kc.

94. Keying of Telegraph Transmitters.—The output of a crystal-controlled telegraph transmitter is ordinarily turned on and off in accordance with the characters of the telegraph code by making one of the stages following the crystal oscillator inoperative. The crystal oscillator itself and usually one or more buffer stages are allowed to operate continuously. This improves the frequency stability, particularly if the crystal and buffers are operated from a separate power supply. The proper stage for keying is a compromise between the desirability of having only small currents and voltages to handle and the desirability of having as many buffer tubes as possible between the crystal and the point of keying.

Master-oscillator–power-amplifier arrangements such as those of Fig. 165 are normally keyed by turning both master oscillator and power amplifier on and off. It might be thought desirable to operate the oscillator continuously and key only the amplifier. However, the fact that all tubes commonly operate from the same power-supply system, and that no buffer tubes are provided, largely nullifies the advantages that would result from continuous operation of the oscillator.

The actual details of the keying system can vary greatly. In Fig. 160 the harmonic generator is made inoperative by arranging matters so that a large negative bias is applied to the grid when the key is up. In the transmitter of Fig. 161, depressing the key removes a negative bias from the screen grid of the second intermediate tube and applies the normal positive screen voltage. In Fig. 165 the key operates by opening both the plate and the grid return leads to the cathode. Still other arrangements include placing the key between the center-tap cathode connection and ground, and keying directly in the plate-supply system either in the primary of the power transformer or in the output of the filter system.

Keying Troubles.—A proper keying system gives clean-cut dots and dashes having constant carrier frequency and produces a minimum of interference. The most common keying trouble encountered is the production of thumps, termed *key clicks,* in near-by radio receivers as the key makes and breaks contact. Such key clicks are caused by high-order side-band frequencies produced by the sudden starting and stopping of oscillations, as shown in Fig. 166*a*. In a hypothetical case where the transmitted energy is assumed to be turned on and off instantly, the resulting wave contains side-band components extending all the way from zero to infinite frequency. Although the amount of energy at any frequency is extremely small, it is still sufficient to produce interference on near-by radio receivers irrespective of their tuning.

The simplest remedy for key clicks is to provide a time delay or lag circuit in association with the key so that the oscillations start and stop gradually rather than abruptly. Such an arrangement is illustrated in Fig. 160, and should be so proportioned that the change in amplitude is slow enough to prevent the generation of high-order side bands, but at the same time is not so gradual that the dots and dashes become indistinct (see Figs. 166*b* and 166*c*). It is also helpful to key at a point such that the keyed output must pass through several tuned circuits before being radiated, since these circuits when properly shielded will tend to suppress high-order side bands.

Other keying troubles that are sometimes encountered include frequency (or phase) shift, gradual drift of the transmitted frequency, and distortion of the dots and dashes by transients in the power-supply system. Frequency or phase shift is the result of insufficient isolation of the oscillator from the amplifier stages that carry keyed currents, and produces spurious side bands. Frequency drift is particularly common in master-oscillator–power-amplifier arrangements, and is caused by gradual temperature rise during operation. Transients occur in the power-supply system as a result of the variation in load produced by keying. These transients are particularly troublesome when the keying is done in the primary of the power transformer, in which case there is a tendency for the dots and dashes to run together and become indistinct, as in Fig. 166*c*.

95. Radio-telephone Transmitters.—In radio-telephone transmitters the output is modulated in accordance with the sound wave to be transmitted instead of being simply turned on and off by means of a key. The usual radio-telephone transmitter consists of a stable oscillator, usually a crystal oscillator, followed by a chain of amplifiers, one of which is modulated by one of the methods described in Chap. VIII. The modulation may take place in the final power amplifier, which is referred to as *high-level* modulation, or may take place in one of the lower power stages, which is termed *low-level* modulation. In the latter case it is necessary that all the amplifiers following the modulated stage be linear amplifiers in order to avoid distortion of the modulation.

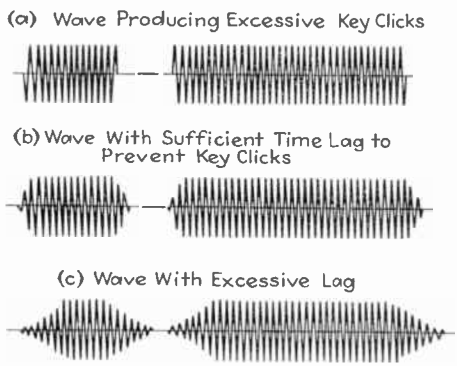


FIG. 166.—Output waves obtained with different keying conditions.

Broadcast Transmitters.—All broadcast transmitters employ a crystal oscillator to generate the frequency to be transmitted. This is followed by one or more Class C buffer amplifiers, after which comes a modulated amplifier, followed in some cases by one or more stages of linear amplification.

Schematic diagrams of typical broadcast transmitters are shown in Figs. 167 and 169. The transmitter of Fig. 167 is an RCA 1-kw model employing high-level modulation. The tube sequence in the radio-frequency chain is as follows:

5-watt crystal oscillator.
7½-watt screen-grid Class C amplifier.
100-watt neutralized triode Class C amplifier.
Two 100-watt triodes operated as neutralized push-pull Class C amplifiers.
Four 250-watt triodes connected in parallel push-pull, and plate modulated.

The audio-frequency system consists of a three-stage amplifier, the final stage of which is a Class B audio amplifier capable of delivering sufficient audio power (about 800 watts) to modulate completely the output of the

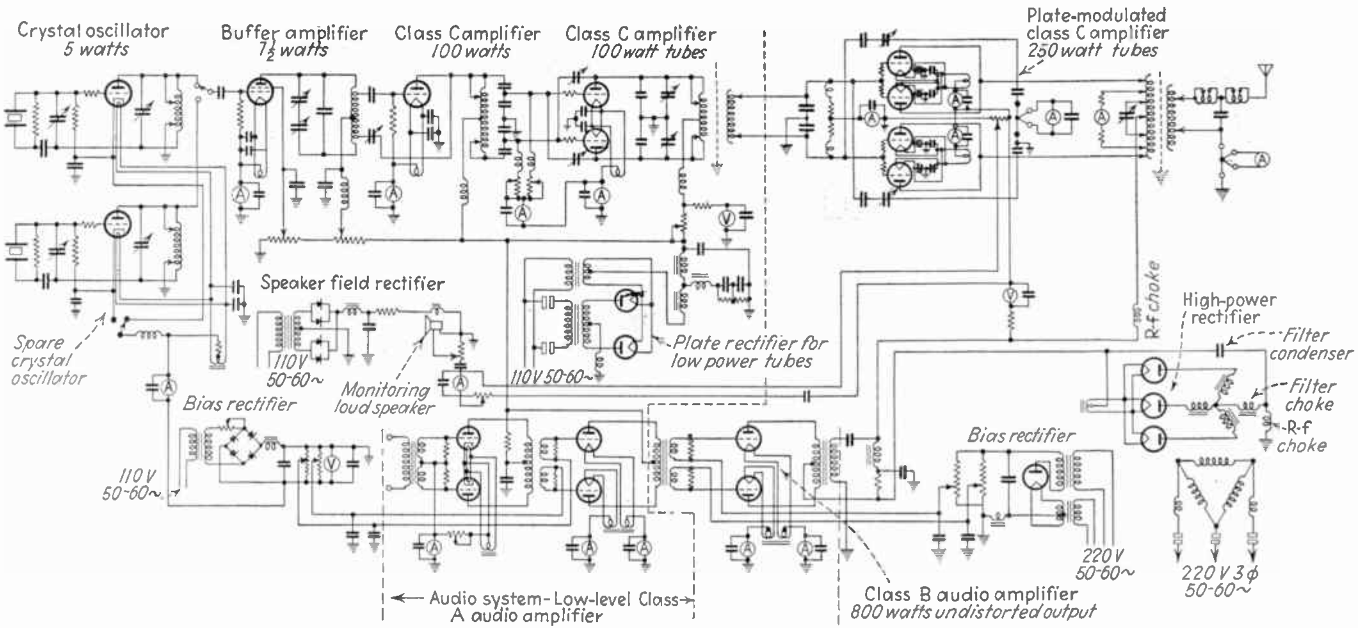


FIG. 167.—Simplified schematic circuit diagram of RCA 1-kw. broadcast transmitter employing high-level modulation.

final radio-frequency power amplifier. Photographs of this transmitter are shown in Fig. 168.

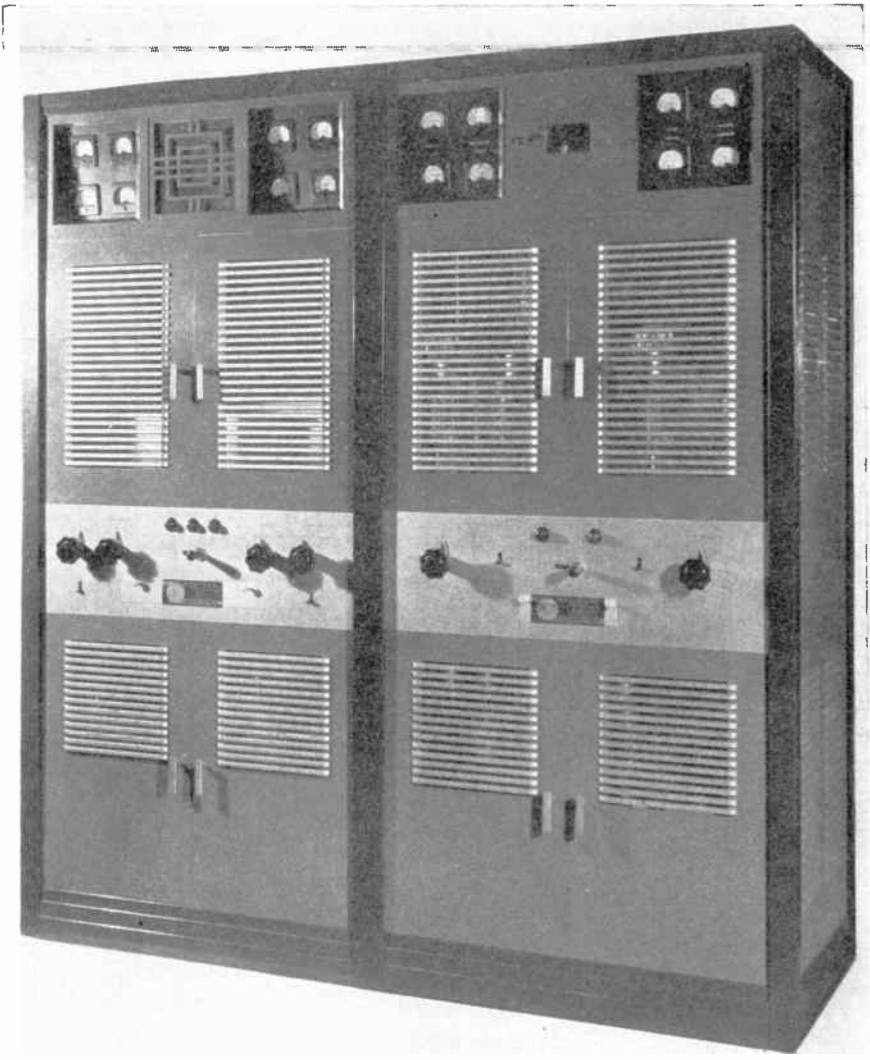


Fig. 168a.—Front view of RCA 1-kw broadcast transmitter of Fig. 167.

The schematic circuit diagram of a 5-kw Western Electric transmitter employing low-level grid modulation is shown in Fig. 169. The sequence of tubes in the radio-frequency chain is as follows:

5-watt crystal oscillator.
5-watt buffer amplifier.

Two 5-watt tubes connected in parallel as Class C amplifier.
Two 250-watt tubes connected in push-pull and grid modulated.
Two 10-kw water-cooled tubes as push-pull linear amplifier.

The audio-frequency system consists of a two-stage amplifier, the second stage being a single triode capable of developing an undistorted power

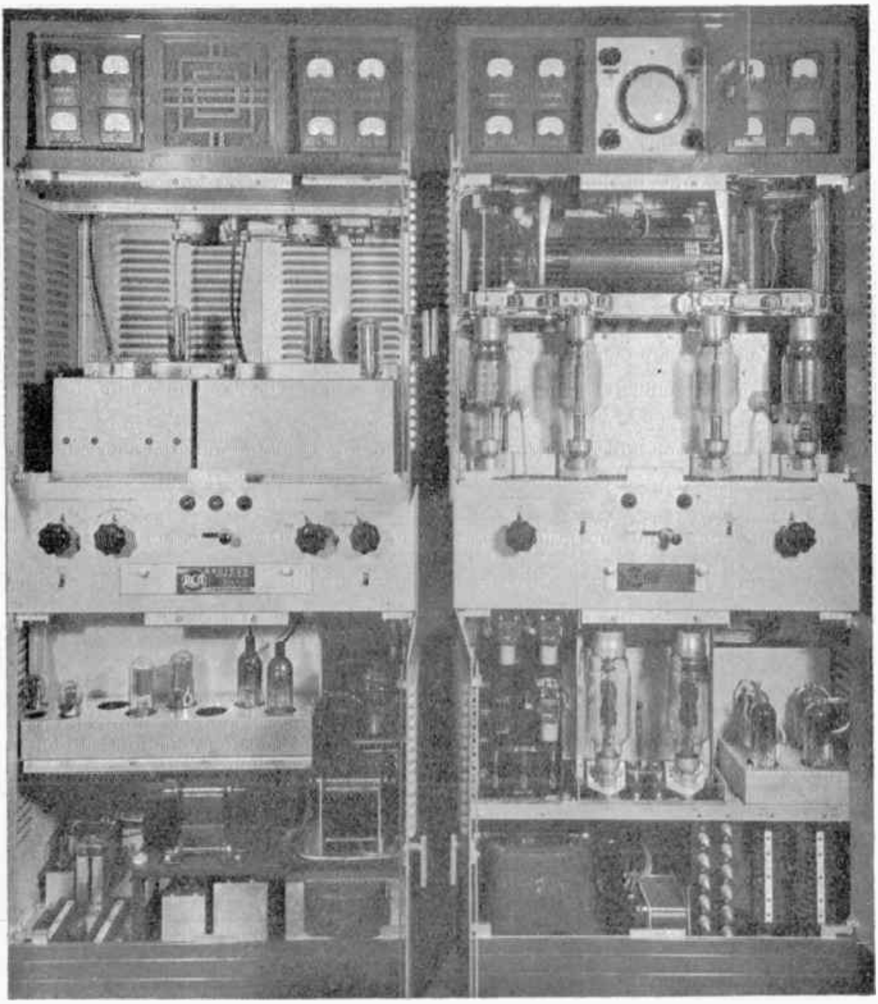


FIG. 168b.—Front of RCA 1-kw broadcast transmitter with front panels removed.

output of 10 watts when operating as a Class A amplifier. When the linear amplifier is omitted and the output of the grid-modulated stage is delivered directly to the antenna, the result is a 100-watt transmitter suitable for low-power broadcast stations.

In broadcast transmitters of very high power, as 10 kw and more, the power consumption and installed tube capacity represent a major item of expense. The most economical arrangement from these points of view is a high-level system in which the output stage employs the high-effi-

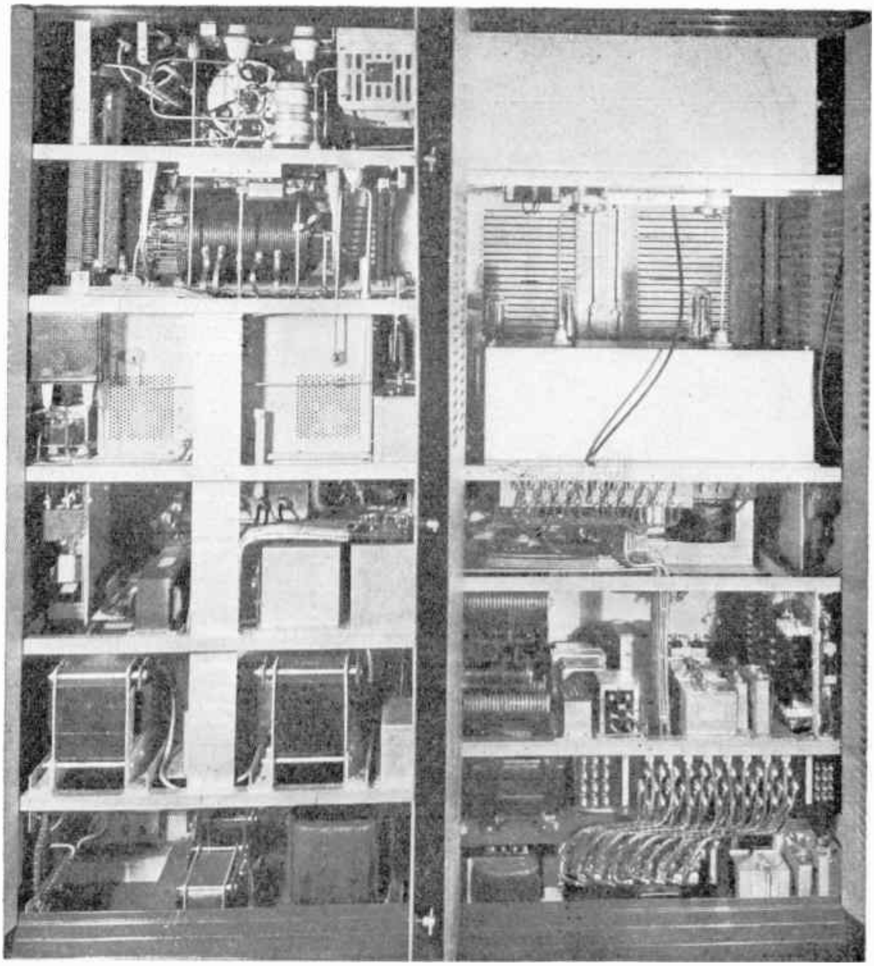


FIG. 168c.—Rear of RCA 1-kw broadcast transmitter with rear covers removed.

ciency grid-modulation arrangement of Fig. 122. This is too new for its commercial possibilities to have been evaluated, however. High-power transmitters actually being placed in commercial service employ the high-efficiency linear amplifier of Fig. 99 when low-level modulation is used, whereas corresponding installations of the high-level type employ Class B audio modulators as in Fig. 167.

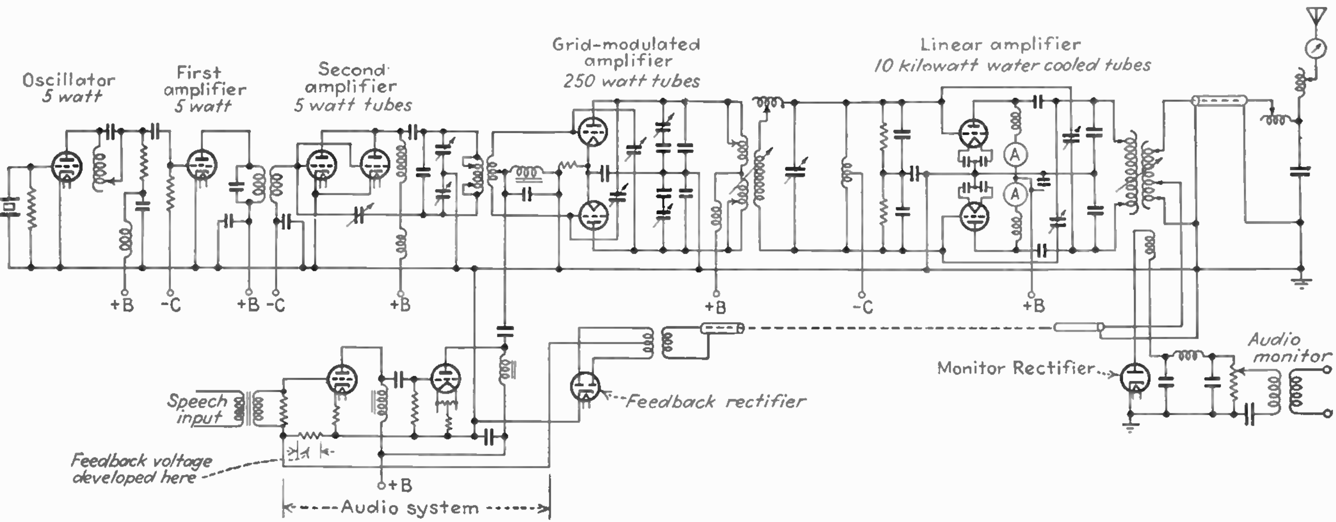


FIG. 169.—Simplified schematic circuit diagram of Western Electric 5-kw broadcast transmitter employing low-level grid modulation.

Short-wave Radio-telephone Transmitters.—Radio-telephone transmitters for operation at high frequencies are similar to broadcast transmitters, except that harmonic generators are often required between the crystal oscillator and the modulated stage in order to produce the required carrier frequency. Many short-wave radio-telephone transmitters, notably those used for police, aircraft, and amateur purposes, do not have to meet the quality and frequency stability requirements of broadcast work. This permits circuit arrangements having less buffer action and makes it permissible to use modulation systems, such as suppressor-grid modulation, that have relatively high distortion.

96. Radio Transmitters. Miscellaneous. *Transmitters for Ultra-high Frequencies.*—The commercial use of frequencies greater than 30 mc is such a recent development that transmitters for this frequency range have not yet become standardized. Some transmitters for operation in the range 30 to 100 mc employ crystal oscillators and differ from ordinary short-wave transmitters only in the use of additional harmonic generators. In other cases, master-oscillator—power-amplifier arrangements are employed, with the master oscillator having its frequency stabilized by means of a resonant transmission line designed to have a very high Q. Self oscillators, particularly those of the resonant-line type, are also used to some extent.

Special tubes designed to have low transit time are often used at frequencies greater than 30 mc and are required if appreciable power is to be obtained at frequencies in the range 100 to 1000 mc. At still greater frequencies some form of electron oscillator is required.

Monitoring Systems.—In order to insure satisfactory performance of a radio transmitter it is necessary to monitor the transmitted signals systematically. In broadcast stations this is accomplished by providing means for observing the carrier frequency and the degree and linearity of the modulation. A portion of the modulated output wave is also rectified and supplied to a loud-speaker, so that the performance can be continuously observed aurally.

Negative Feedback in Radio Transmitters.—Amplitude distortion, frequency distortion, and hum in a radio-telephone transmitter can be reduced by the application of a modified form of negative feedback, as illustrated in Fig. 169. It will be observed that this is essentially the same arrangement as discussed in Sec. 73 in connection with modulation systems, except that the linear amplifiers following the modulated amplifier are included in the feedback loop. The application of negative feedback to transmitters tends to correct for imperfections in the modulator, modulated amplifier, and linear amplifiers, and so improves the performance obtainable.

Suppression of Harmonics.—Since Class C and linear amplifiers generate harmonics because of the distorted shape of the plate-current pulses, it is necessary to prevent these harmonics from being radiated. Tank circuits having a high effective Q are of help in this regard, and additional suppression can be obtained, when required, by a filter network between tank circuit and antenna. The amount of discrimination against harmonics that is required is proportional to the carrier power, since, for a given allowable harmonic radiation, the permissible percentage of harmonic is inversely proportional to the carrier power.

Adjacent Channel Interference.—Radio-telephone transmitters sometimes produce interference on frequencies that are outside the normal side-band range and yet are relatively close to the carrier frequency. Such *adjacent channel interference* is caused by high-order side bands resulting either from distortion of the modulation envelope or from frequency or phase modulation of the transmitter. Thus, if the highest modulation frequency applied to a transmitter is 5000 cycles, second-order side bands extend to 10,000 cycles on either side of the carrier, fifth-order side bands extend to 25,000 cycles, etc. These extended side bands interfere with the local reception of signals from distant transmitters by producing a babel often referred to as "monkey chatter."

The chief causes of adjacent channel transmission are overmodulation and phase (or frequency) modulation. Overmodulation occurs when the modulating voltage is greater than required to give 100 per cent modulation. This produces distortion in the modulation envelope that introduces high-order side-band components. Such adjacent channel interference can be minimized by carefully monitoring the transmitter so that overmodulation occurs only infrequently.

Phase and frequency modulation is the result of poor transmitter design, or improper adjustment, and is particularly troublesome on short-wave telephone and telegraph transmitters. Adjacent channel interference of this type can be avoided by the use of properly designed and shielded buffer stages between the crystal oscillator and the point where the keying or modulation takes place, and by carefully tuning all tank circuits exactly to resonance.

Interrupted Continuous Waves.—In telegraph transmitters it is sometimes desired to interrupt the transmitted dots and dashes at an audio rate, giving what is termed I.C.W. (interrupted continuous waves). This is commonly accomplished by the use of a self-rectifying circuit using 500-cycle power, or by the use of some form of motor-driven interrupter or chopper. Interrupted continuous waves can be received upon ordinary radio receivers without heterodyne action. They also have the advantage in short-wave work that they do not fade out so readily as an unmodulated wave (see Sec. 104).

Controlled-carrier Transmitters.—The average plate power loss of a linear amplifier handling a voice-modulated wave can be greatly decreased by reducing the carrier amplitude during intervals when the modulation is small or zero, and transmitting the full carrier amplitude only when the modulating voltage is maximum. The required variations in carrier amplitude can be obtained by using an auxiliary control tube, or a saturated reactor, to vary the carrier amplitude in accordance with the average speech level. The circuits must be arranged so that this control action is rapid enough to follow the normal syllabic variations in speech power but not sufficiently rapid to follow the lower speech frequencies.

The use of controlled carrier reduces very considerably the average exciting voltage, and hence average plate loss, of a linear amplifier. This is because the average amplitude of speech is much less than the peak amplitude. A linear amplifier handling a controlled-carrier wave can hence be adjusted to deliver more peak power output without exceeding the allowable plate loss than if an ordinary carrier wave were applied. At the same time, the use of a controlled carrier introduces distortion because the carrier amplitude cannot be changed with sufficient rapidity to accommodate very sudden peaks of modulation. The use of controlled carrier also introduces complications when the receiver is provided with automatic volume control. As a result, the use of controlled carrier is limited to special applications in amateur, aircraft, and portable equipment where economy of cost and weight is more important than high quality.

Problems

1. Answer the following questions concerning the telegraph transmitter of Fig. 161.

a. What system of neutralization is used in the final intermediate amplifier?

b. Make a detail drawing of the keying system and explain how it works.

c. Explain why trouble is not experienced from key clicks in spite of the fact that there is no lag circuit associated with the key.

d. How much ripple voltage does the power-supply system for the final power amplifier develop in its output?

2. Make a detail drawing of the oscillator in the transmitter of Fig. 165. Number each coil, condenser, and resistance and explain its function in the circuit.

3. Design a telegraph transmitter delivering 150 watts output at a frequency of 7500 kc and having crystal control, with the crystal operating at half the transmitted frequency. The design includes a circuit layout, choice of tubes, and rough determination of tube-operating conditions (such as preliminary estimates of grid bias, driving power, d-c plate current, plate efficiency, etc., for each stage). In this problem make full use of tube data given in Table X and provide at least 100 per cent more driving power than estimates indicate is necessary.

4. *a.* Make sketches of the system used in the transmitter of Fig. 167 to couple the modulator tube to the modulated amplifier.

b. Explain the factors that determine the extent to which the high and low audio frequencies fall off as a result of imperfect coupling.

5. In the transmitter of Fig. 167, draw to a large scale the circuit of the output stage (the part to the right of the dotted line), number each circuit element, and explain its purpose. Do not include the power-supply system.

6. In the transmitter of Fig. 167 determine values for the filter inductance and condenser such that the hum voltage modulated upon the transmitter output will not produce a degree of modulation exceeding 0.001.

7. Design a crystal-controlled transmitter for operation at 1600 kc and capable of developing 50 watts of completely modulated carrier, with high-level plate modulation. The design includes circuit for radio-frequency and modulator stages, selection of tubes using Tables IX, X, and XIII, specification of circuit constants, and rough determination of such tube-operating conditions as expected output, d-c plate current, grid driving power, grid bias, etc. In this problem provide at least 100 per cent more driving power for each stage than estimates indicate will be needed.

8. *a.* Make a detail drawing of the power-supply system for the output amplifier in Fig. 167.

b. Make an estimate of the direct-current power that this system must supply.

9. Make a detail drawing of the feedback system used in the transmitter of Fig. 169 and explain how it operates.

10. In the transmitter of Fig. 169, the high-frequency side bands tend to be discriminated against by the tuned circuits. Explain why it is not possible to equalize the frequency characteristic of the audio-frequency amplifier to correct for this without causing adjacent channel interference when the degree of modulation approaches 100 per cent.

11. Explain why problems of adjacent channel interference become more important as the transmitter power increases.

12. Assuming reasonable values of plate efficiency, estimate the total plate losses and the total power input (including modulator and radio-frequency tubes) required for a broadcast transmitter developing a 50-kw carrier when (*a*) unmodulated and (*b*) completely modulated, for the following systems: (1) high-level plate modulation using Class A audio system, (2) high-level plate modulation employing Class B audio system, (3) low-level modulation followed by conventional linear amplifier, and (4) low-level modulation followed by Doherty high-efficiency linear amplifier.

In making these estimates neglect all low-level radio-frequency stages and all audio stages except the last stage of high-level systems.

CHAPTER XII

RADIO RECEIVERS

97. Characteristics of Broadcast Receivers.—The most important characteristics of a receiver for radio-telephone signals are sensitivity, selectivity, and fidelity. The sensitivity is expressed in terms of the carrier voltage, modulated 30 per cent at 400 cycles, that is required to develop an output of 50 mw in a non-inductive resistance substituted for the loud-speaker or other output device. A curve showing the sensitivity of a typical broadcast receiver as a function of carrier frequency is shown in Fig. 170*a*.

Selectivity is the property that enables a radio receiver to discriminate between radio signals of different carrier frequencies. Selectivity is expressed in the form of curves, such as those of Fig. 170*b*, which show the amount by which the signal input must be increased in order to maintain the output constant as the carrier frequency is varied away from the frequency to which the receiver is tuned. These curves therefore indicate the extent to which interfering signals are discriminated against.

Fidelity represents the extent to which the receiver reproduces the different modulation frequencies without frequency distortion. The fidelity of a radio receiver is expressed in curves, such as those of Fig. 170*c*, which give the variation in audio-frequency output voltage as the modulation frequency of the signal is varied. In order to facilitate comparison, the output is expressed in terms of the ratio of actual output to the output obtained when the modulation frequency is 400 cycles.

Measurement of Receiver Characteristics.—The characteristics of a radio receiver are measured by using an artificial signal to represent the voltage that is induced in the receiving antenna. This artificial signal is applied to the input terminals through a network, or "artificial antenna," that simulates the impedance of the actual antenna with which the receiver is to be used.[1] In making this test the receiver output is determined by substituting a resistance load of the proper value for the loud-

[1] The accepted standard artificial antenna for the broadcast band (550 to 1500 kc) consists of a capacitance of 200 $\mu\mu$f, a self-inductance of 20 μh, and a resistance of 25 ohms, all in series. For other cases the artificial antenna depends upon the antenna system with which the set is intended to be used. Thus receivers operating in conjunction with a non-resonant transmission line would be tested with an artificial antenna consisting of a resistance corresponding to the characteristic impedance of the transmission line.

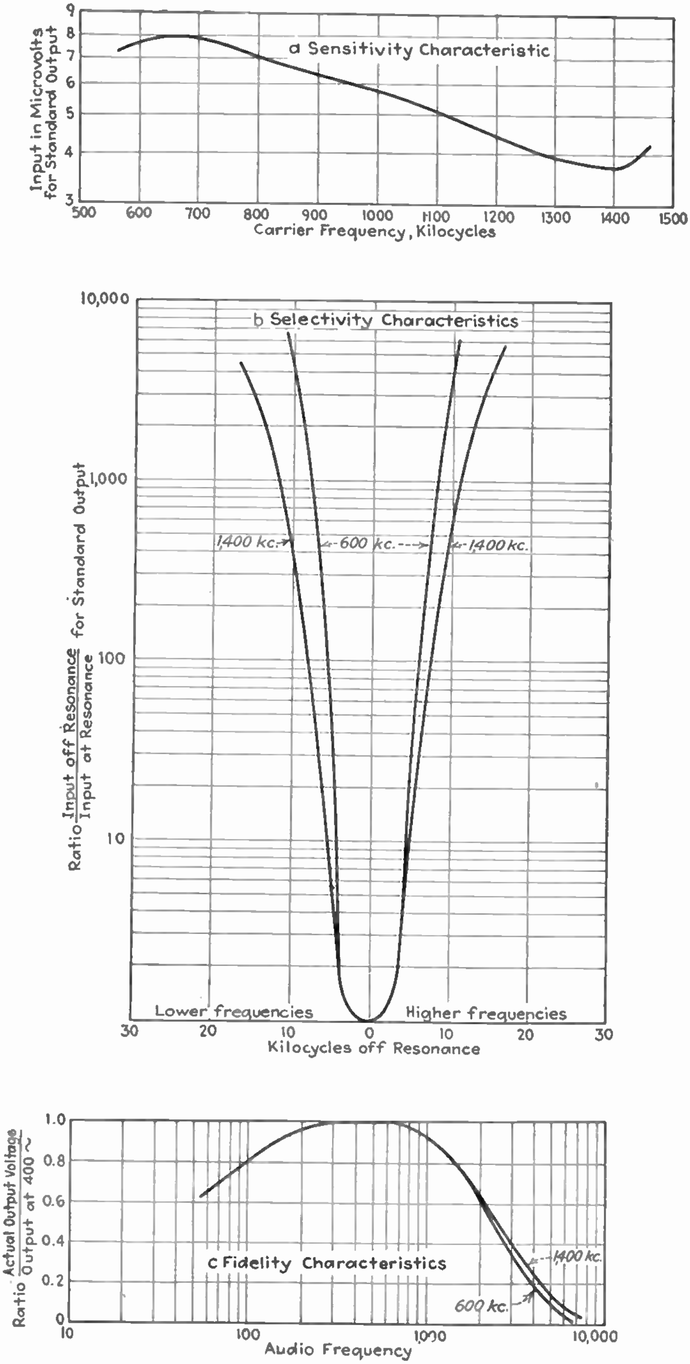


FIG. 170.—Typical sensitivity, selectivity, and fidelity curves of superheterodyne receiver.

speaker and measuring the audio-frequency power in this resistance. The experimental setup is illustrated in Fig. 171.

The equipment for producing the artificial signal is called a *standard signal generator*. It consists of a thoroughly shielded oscillator coupled to an attenuating system that is capable of producing known voltages from about 1 μv up to perhaps 200,000 μv.

98. Typical Broadcast Receivers.—Practically all broadcast receivers are of the superheterodyne type and accordingly have the schematic

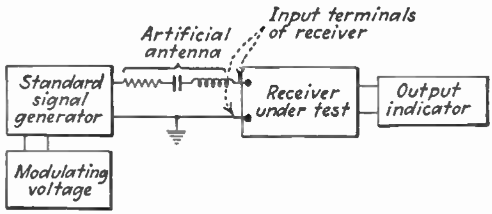


FIG. 171.—Schematic arrangement of equipment for making measurements of receiver performance.

layout indicated in Fig. 172. The radio-frequency section is tuned to the signal frequency and delivers a signal voltage at the grid of the first detector. The first detector (also called *mixer* and *converter*), together with its associated oscillator, converts the incoming oscillations to a fixed predetermined *intermediate frequency* by use of the heterodyne principle. The intermediate-frequency section between the two detectors is tuned to this predetermined difference frequency and delivers an intermediate-frequency voltage to the second detector, usually a diode, which recovers the modulation envelope from the wave by rectification.

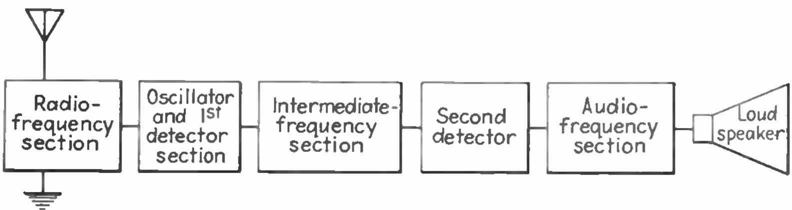


FIG. 172.—Schematic diagram of superheterodyne receiver.

The resulting audio frequency is then amplified by the audio-frequency system and delivered to the loud-speaker for reproduction.

Although all superheterodyne receivers follow the general scheme outlined in Fig. 172, individual receivers differ greatly in detail. Thus the radio-frequency section may include one or more stages of amplification, or may be a simple tuned circuit between the antenna and first detector. The first detector and oscillator section may involve any of the arrangements discussed in Sec. 80. Likewise, the intermediate-frequency section may consist of a simple tuned circuit between first and second detectors or may include one or two stages of amplification. The details of the second detector and audio-frequency amplifier may also be varied.

All receivers also offer some combination of special features such as automatic volume control, tuning indicators, tone control, quieting arrangements, automatic frequency control, etc. The number and exact nature of these incorporated in a particular receiver vary greatly with the price and are affected by merchandising styles.

Examples of Broadcast Receivers.—The circuit combinations involved in ordinary broadcast receivers can be understood by considering the details of several typical models.

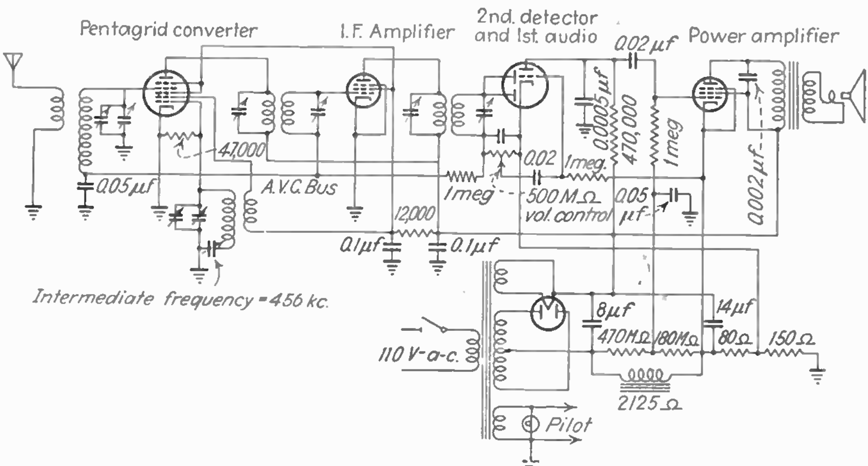


FIG. 173.—Circuit diagram of typical low-priced broadcast receiver.

The receiver of Fig. 173 is a representative low-priced superheterodyne receiver designed for covering the frequency band 540 to 1750 kc. Five tubes are employed, consisting of a pentagrid converter, intermediate-frequency amplifier, diode detector and resistance-coupled triode voltage amplifier combined in one envelope, and a pentode Class A power amplifier, together with the rectifier tube.

The receiver of Fig. 174 is a medium-priced all-wave receiver. There are three tuning ranges, namely, 550 to 1730 kc, 1800 to 6000 kc, and 6000–20,000 kc. The seven tubes include a pentagrid converter, intermediate-frequency amplifier, diode second detector, resistance-coupled audio amplifier, pentode Class A power amplifier, rectifier, and tuning indicator. In the broadcast band a band-pass circuit is employed between antenna and first detector.

The circuit arrangement of a moderately high-priced receiver is shown in Fig. 175. This is an eleven-tube model covering the frequency range 540 to 22,000 kc in three bands. The tube line-up, exclusive of the rectifier, is a radio-frequency amplifier, pentagrid mixer tube with separate oscillator, two stages of intermediate-frequency amplification, a

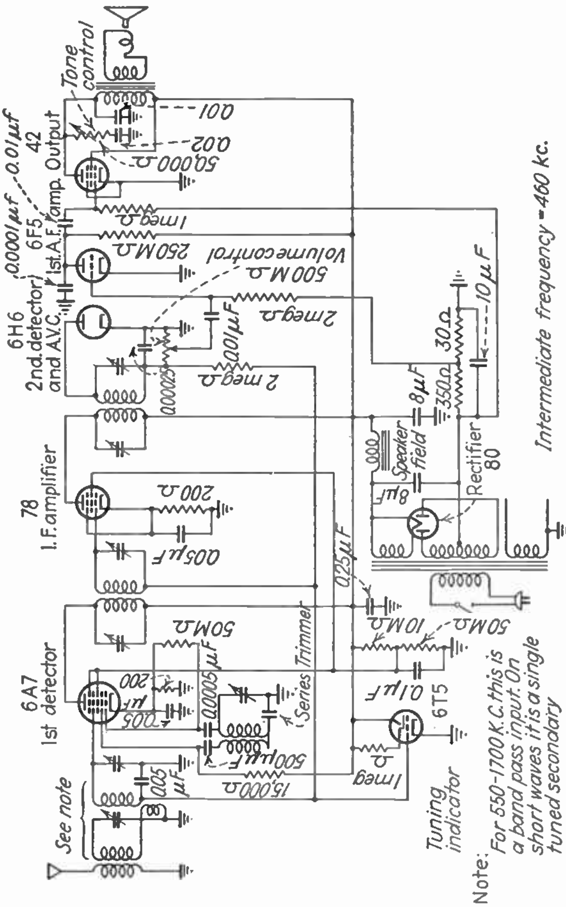


FIG. 174.—Circuit diagram of typical medium-priced broadcast receiver.

double-diode second detector, an audio-frequency amplifier and phase inverter combined in one envelope, and a push-pull pentode output stage. In addition, there is a control tube for the automatic-frequency-control system, and a tuning indicator tube.

99. Broadcast Receivers. Miscellaneous Features. *Automatic Volume Control.*—Practically all broadcast receivers are provided with some form of automatic volume control (abbreviated A.V.C.) to maintain the carrier voltage at the detector substantially constant irrespective of the

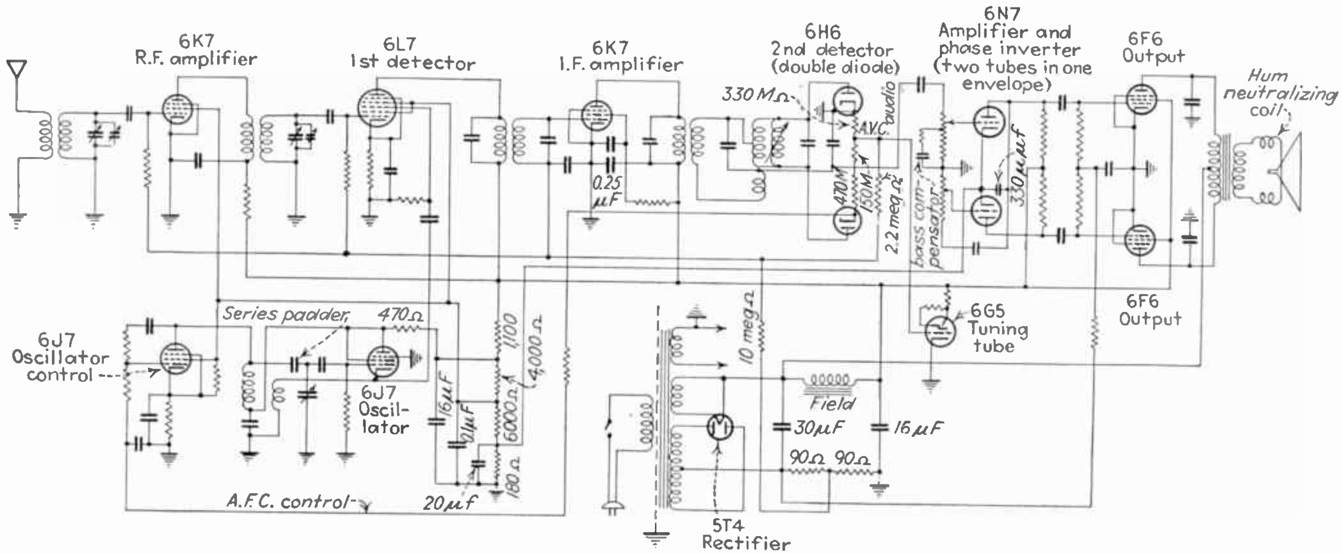


FIG. 175.—Circuit diagram of typical moderately high-priced all-wave broadcast receiver with automatic frequency control.

signal strength. This is accomplished by biasing the grids of the radio-frequency, intermediate-frequency, and the converter tubes negatively with a direct-current voltage derived by rectifying the carrier. An increase in the signal hence increases this negative bias and thereby tends to counteract the increased signal by reducing the amplification, while, if the signal becomes weaker, the automatic-volume-control bias is correspondingly less, and the gain of the controlled tubes increases.

The actual details of automatic-volume-control systems vary greatly. A simple diode detector can be used for both detection of the signal and for producing an automatic-volume-control bias by arrangements as illustrated in Figs. 134, 173, and 174. In other cases a double-diode detector is used, with one diode serving as the ordinary detector and the other developing the automatic-volume-control bias.

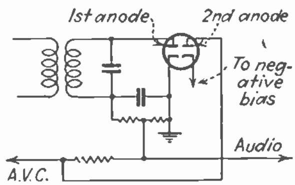


FIG. 176.—Method of obtaining delayed automatic volume control.

Automatic-volume-control systems are often arranged so that no A.V.C. voltage is developed until the carrier voltage at the detector exceeds a predetermined value. This is termed *delayed automatic volume control*, and has the advantage that the A.V.C. system does not begin to reduce the amplification until the carrier voltage at the detector reaches the desired value. A typical method of obtaining delayed A.V.C. is illustrated in Fig. 176, in which the second anode of a double-diode tube is connected to the A.V.C. line but has its cathode biased negative by the amount of delay desired. In such an arrangement the A.V.C. bus assumes the potential of the cathode of the second half of the diode tube until there is sufficient carrier to produce an A.V.C. voltage greater than the bias of the second cathode. Beyond this point the second anode is inoperative, because it becomes more negative than its cathode, and the A.V.C. system functions in a normal manner.

Tuning Indicators.—In receivers provided with automatic volume control, difficulty is encountered in tuning the receiver because the automatic-volume-control system tends to maintain the output constant even when the receiver is not tuned exactly to resonance. At the same time, with slight mistuning the carrier is on the side of the response curve of the receiver, and considerable distortion results. Consequently various devices have been developed for providing assistance in tuning the receiver.

The simplest form of tuning indicator is a direct-current meter through which flows the rectified output of the second detector. The tuning is then adjusted for a maximum deflection of the instrument. Another common arrangement employs a light valve actuated by the d-c current of the detector to vary the width of a shadow band.

A more recent development in tuning indicators is a special miniature cathode-ray tube in which the size of the luminous area is determined by the negative bias applied to a control electrode. By deriving this bias from the automatic-volume-control system the size of the luminous area serves as an aid to tuning. Such an arrangement is used in the receivers of Figs. 174 and 175.

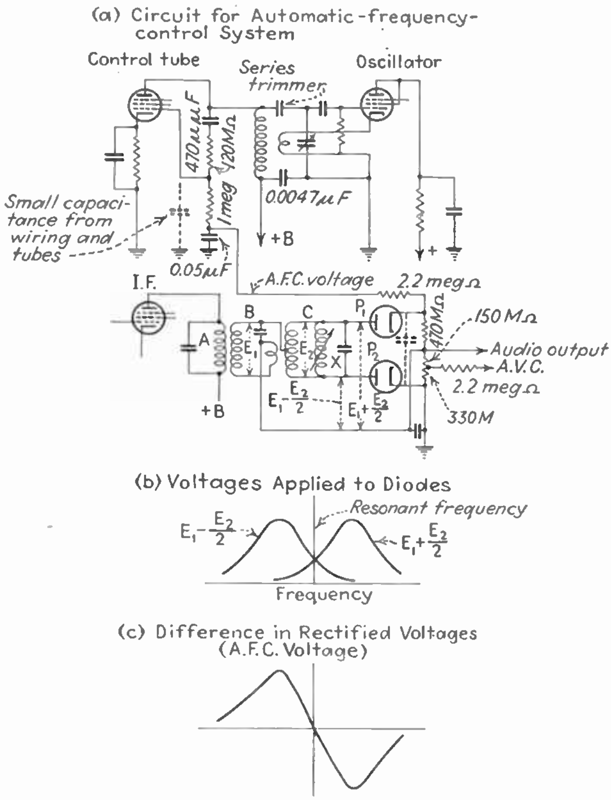


FIG. 177.—Details of automatic-frequency-control system used in the receiver of Fig. 175

Automatic-frequency-control Systems.—The difficulty of accurately tuning a receiver having automatic volume control, as well as troubles caused by drifts in the frequency of the beating oscillator when receiving short-wave signals, can be eliminated by an arrangement that will automatically shift the beating oscillator frequency so as to produce an intermediate frequency of exactly the proper value, provided the tuning is at least approximately correct. This is termed automatic frequency control (abbreviated A.F.C.) and is employed in the receiver of Fig. 175.

The details of the automatic-frequency-control arrangement of the receiver of Fig. 175 are shown in Fig. 177*a*. Here *A* and *B* represent the primary and secondary tuned circuits of an ordinary band-pass intermediate-frequency amplifier, and C is a third tuned circuit that is coupled to B. Coil X in this third circuit is for the purpose of varying the resonant-frequency of circuit C by adjusting the position of a dust core with which coil X is provided. The connections to the double-diode detector are such that anode P_1 has applied to it the voltage developed across circuit B plus half the voltage across circuit C, and anode P_2 receives the voltage developed across circuit B minus half the voltage across circuit C. Because of the phase relationship that exists in two coupled circuits tuned to the same frequency, these diode voltages vary with frequency as shown in Fig. 177*b*. The difference between the direct-current voltages developed in the outputs of the two diode detectors accordingly varies with frequency as shown at Fig. 177*c*. This difference voltage developed by the two diodes is then used to control the frequency of the local oscillator of the superheterodyne receiver so that this oscillator frequency will be shifted in a direction that tends to reduce the difference between the actual intermediate frequency obtained and the resonant frequency of the tuned circuit.

The usual method of controlling the oscillator frequency is illustrated in Fig. 177*a*. This involves a pentode tube, the plate circuit of which is in parallel with the tuned circuit of the oscillator. The grid of the pentode is excited from the oscillator with a voltage 90° out of phase with the voltage existing across the oscillator tuned circuit, so that the current drawn by the pentode tube is 90° out of phase with the radio-frequency voltage across the oscillator circuit. The control tube consequently acts as a reactance in parallel with the tuned circuit of the oscillator. The magnitude of the reactance, and hence the oscillator frequency, is determined by the mutual conductance, which in turn is controlled by the difference voltage developed by the two diode plates. In this way the local oscillator frequency will be automatically shifted in such a way as to correct for a moderate amount of mistuning, or for the effects of ordinary temperature variations.

Bass Compensation.—The characteristics of the human ear are such that, when sounds are reproduced at lower than normal volume levels, the low notes appear to be abnormally weak, whereas, when the sound is reproduced at greater than normal level, the low notes appear to be abnormally loud.[1] In order to correct for this, the manual volume control is often arranged so that at low levels the intensity of the low notes is not reduced so much as is the volume of the higher pitched sounds. A simple example of such a bass-compensated volume control is found in the receiver of Fig. 175, where at low volume settings the section of the potentiometer in use is shunted to ground through a resistance and

[1] Further discussion of the characteristics of the ear is to be found in Sec. 135.

capacitance combination that has a lower impedance to moderate and high frequencies than to low frequencies, and so discriminates against the former.

Tone Control.—Most receivers provide a *tone control* so that the listener may discriminate against the higher audio frequencies. The tone control is usually some form of resistance-capacitance combination, with the arrangements shown in the receivers of Figs. 174 and 175 being typical.

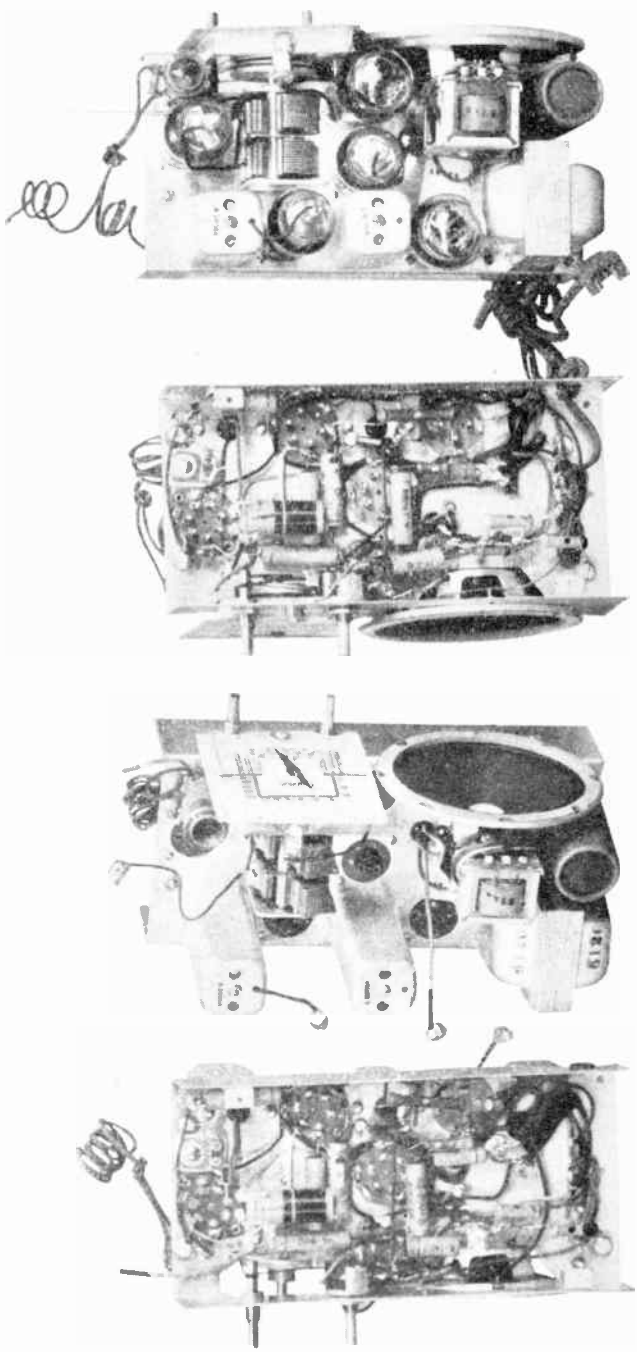
Quieting Systems.—In tuning a sensitive receiver provided with automatic volume control, the noise output between stations is high because, when no signal is being received, the A.V.C. system increases the sensitivity of the receiver to the maximum possible value. Arrangements for eliminating this inter-station noise are variously known as Q circuits, quieting systems, squelch circuits, and codans.[1] Various means can be employed to accomplish this result. Thus a possible arrangement makes use of an auxiliary tube arranged to bias the grid of the first audio tube beyond cut-off unless the grid bias on the auxiliary tube approaches or exceeds cut-off. By using the A.V.C. system to bias the auxiliary tube, it is then possible to make the receiver inoperative until a carrier of predetermined amplitude is present.

100. Broadcast Receivers. Construction and Design Considerations. Broadcast receivers are mounted on a chassis of sheet metal bent or drawn in the form of an inverted tray. This is punched to receive sockets, coils, transformers, etc., which are usually held in place either by rivets or by self-tapping screws. This chassis is then slipped into a cabinet and is connected to the loud-speaker by means of a flexible cord terminated in a plug.

Photographs showing the construction of typical receivers are given in Figs. 178 and 179. The tubes, transformers, coil assemblies, gang tuning condenser, electrolytic condensers, and other bulky items are usually mounted on the top of the chassis, while small parts such as by-pass condensers, resistors, etc., are placed below along with most of the wiring. The arrangement of parts, and also the location of wires carrying radio-frequency and low-level audio-frequency currents, are carefully worked out to avoid troubles from hum, regeneration, etc. The remainder of the wiring is frequently run at random as shown in Fig. 178*b* in order to reduce the cost of production.

Cost is of prime importance in the design and construction of broadcast receivers. Carbon-rod resistors with pigtail leads, and tubular by-pass condensers with cardboard covering and pigtail leads, are employed wherever possible. These are wired directly into the circuits

[1] The term *codan* is coined from the first letters of the words in the phrase "carrier-operated device, anti-noise."



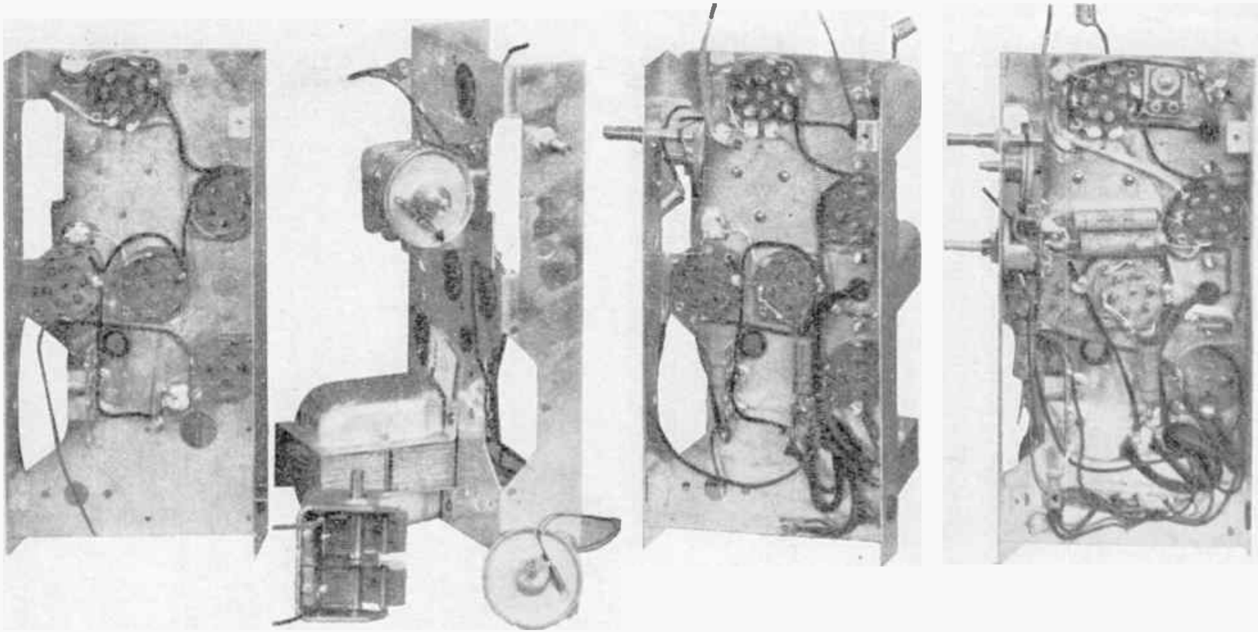


Fig. 178.—Photographs showing successive stages in the assembly of a broadcast receiver. (*Courtesy of Gilfillan Bros.*)

so as to be self-supporting, or are arranged on mounting strips, according to the circumstances. Chokes, coils, and transformers are avoided wherever possible in order to reduce cost. The filter system for supplying the voltage to the power stage is nearly always a single-section shunt input-condenser arrangement employing electrolytic condensers together with a choke provided by the field of the loud-speaker. Additional filtering for the remaining tubes is obtained by resistance-condenser combinations as seen in Figs. 173 to 175.

Coil and Band-switching Arrangements.—The radio-frequency coils of broadcast receivers are very small and are usually, although not always, mounted in a shield can. The tuning coils for the regular broadcast band are either single layer or bank wound, preferably of litz wire, while coils for the high-frequency bands of all-wave receivers are nearly always a single layer of solid wire.

Intermediate-frequency coils are commonly arranged in pairs with the primary and secondaries separately tuned to provide a band-pass effect, with the coupling usually slightly more than the critical value. Universal-wound coils are always employed, and some form of iron-dust core is very common. Adjustment of the resonant frequency of the tuned circuits is obtained either by an adjustable condenser or, in the case of some iron-core types, by a screw-driver adjustment of the core position.

Examples of coil assemblies used for the radio-frequency, oscillator, and intermediate-frequency tuned circuits of receivers are shown in Fig. 180. The arrangement of coils on a typical receiver chassis is shown in Fig. 179.

In all-wave receivers it is necessary to provide means for switching the radio-frequency and oscillator coils so that the inductances of the tuned circuits can be changed for the different frequency bands. The coil switch for this purpose must provide low resistance contacts and introduce negligible coupling between the tuned circuits of different stages. The physical arrangement of switch and coils should also be such that the connections to the switch are as short and direct as possible. The switch contacts are customarily arranged so that unused coils are short circuited. This is so that the distributed capacitance will not make the unused coils resonant in other frequency bands, and thereby introduce the possibility of coupling appreciable impedance into the coils in use.

Some General Design Considerations.—The actual electrical design of broadcast receivers depends to a considerable extent upon the particular sales features that are being emphasized at the time. There are, however, certain general considerations governing the design. The critical factors are the desired audio-frequency power output and sensitivity, and the carrier amplitude at the detector for normal operation. The audio

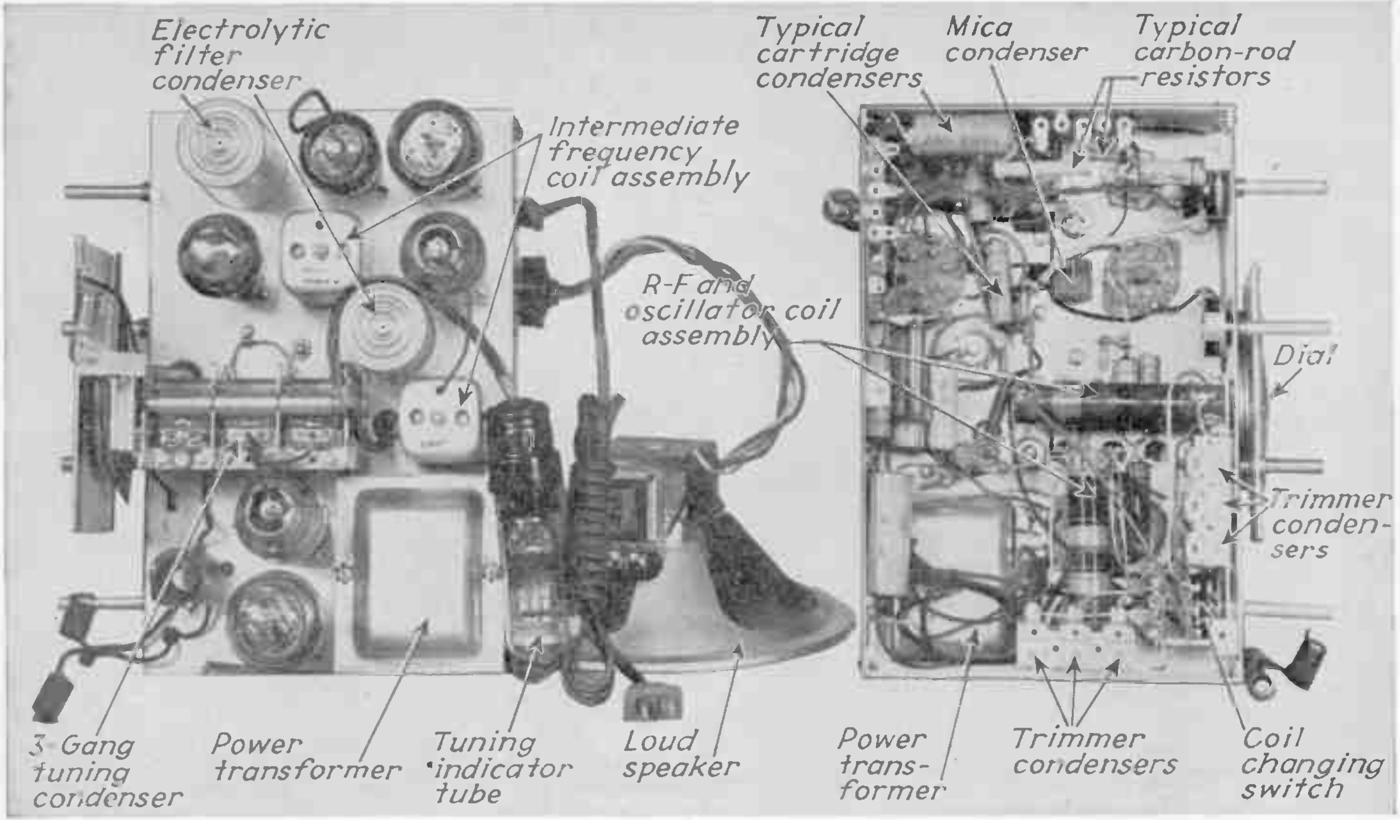
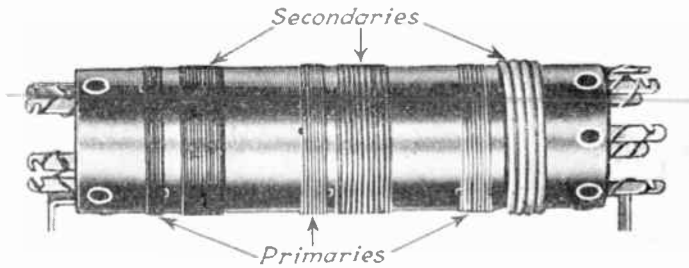
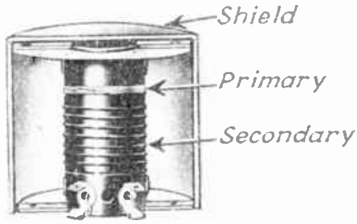


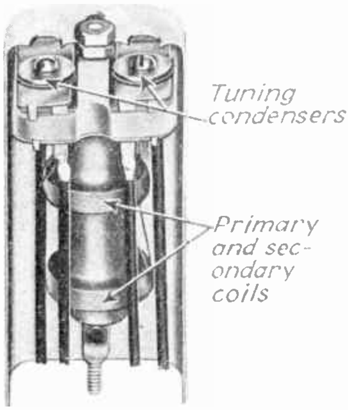
Fig. 179.—Photographs showing views of chassis for receiver of Fig. 174. (*Courtesy of Gilfillan Bros.*)

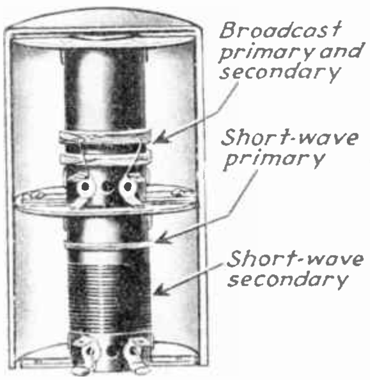


Three-band radio-frequency coil assembly.



One-band radio-frequency coil assembly.





Intermediate-frequency coil assembly. Two-band radio-frequency coil assembly.

FIG. 180.—Typical coil assemblies for broadcast receivers. In the case of intermediate-frequency coils, note the adjustable condensers mounted in the shield cans and used for controlling the resonant frequency.

power tubes and plate-supply voltage are selected to develop the required output. The minimum permissible audio-frequency amplification between detector and power tubes is then the gain that will provide full excitation for the power tubes when the carrier voltage at the detector is completely modulated and has an amplitude equal to the normal detector level. Knowing the audio-frequency amplification, it is then a simple matter to calculate the carrier voltage modulated 30 per cent that must be applied to the detector to develop the standard output of 50 mw. The total voltage step-up between antenna and second detector must then be such that, with a voltage corresponding to the required sensitivity acting in the antenna circuit, the required carrier level will be developed at the second detector. From this it is possible to estimate the number of stages of amplification required on the basis of gains to be expected according to Table XVII.

TABLE XVII.—TYPICAL VOLTAGE GAINS IN BROADCAST RECEIVERS

	Gain
Radio-frequency section:	
Antenna to grid of first tube[1]	3– 6
Radio-frequency amplifier, broadcast band	20– 40
Radio-frequency amplifier, short-wave bands	5– 25
Intermediate-frequency amplifier section:[2]	
Converter grid to first intermediate-frequency tube	40– 60
Interstage intermediate frequency	50–180
Intermediate frequency to diode	50–125

[1] In automobile receivers designed to operate with a particular antenna of specified constants, it is possible to make the antenna-to-grid step-up 20 to 50.

[2] These values are for air-core coils for 460 kc. For air-core coils at 260 kc the gains will be perhaps 30 per cent higher. With iron-dust cores the gains will be 20 to 60 per cent higher than with air cores.

The actual gains in any particular case depend upon the size and design of the coils and the associated coupling systems. In the radio-frequency section the gain varies with the tuning, normally being greatest at the high-frequency end of each tuning range. In the broadcast band (550 to 1500 kc) the performance of both antenna and interstage parts can be improved by using a complex coupling arrangement with a high-inductance primary resonated at a frequency just below the broadcast band, as illustrated in Fig. 70. Such arrangements cannot be used in the short-wave bands, however, because the primaries would resonate with signals in the next frequency band. Short-wave radio-frequency coils hence ordinarily employ a simple coupling system consisting of a low-inductance primary. The antenna-to-first-grid gain is necessarily low in ordinary broadcast receivers because these receivers must be designed so that the first tuned circuit will maintain its alignment when antennas of widely different constants are used, and to do this the antenna coupling must be small.

Spurious Responses in Superheterodyne Receivers.—Superheterodyne receivers are troubled with a variety of spurious responses. These often arise from the fact that there are two possible signal frequencies that will form the required intermediate frequency at any dial setting. Thus with an intermediate frequency of 50 kc, such as was employed in the early superheterodynes, the local oscillator would be adjusted to 1050 kc to receive a 1000-kc signal. Another signal of 1100 kc would, however, also produce a difference frequency equal to the intermediate frequency, and so might be simultaneously heard in the receiver output. This undesired signal differing in frequency from the desired carrier by twice the intermediate frequency is termed the *image signal.*

Response to the image signal can be eliminated by preventing it from reaching the grid of the converter tube. This requires that the radio-frequency section of the receiver have an adequate number of tuned circuits, and that the intermediate frequency be high enough to enable these circuits to suppress the image frequency. In receivers covering only the regular broadcast band, the intermediate frequency is commonly about 262.5 kc, which makes the image frequency differ from the desired frequency by 525 kc. In all-wave receivers a higher intermediate frequency is employed, usually about 460 kc. Even then the image suppression is none too good in the high-frequency bands, as a 920-kc difference between desired and image frequencies then represents only a small percentage of the desired carrier frequency.

Whistles can also appear at certain places on the dial as a result of intermediate-frequency harmonics produced by the second detector getting coupled into the radio-frequency input circuits. Thus, if the intermediate frequency is 452 kc and the signal being received is 900 kc, the second harmonic of the intermediate frequency is 904 kc. If this gets back into the input circuits, it will give a 4000-cycle beat note with the desired signal. The remedy for trouble of this sort is adequate by-passing of the circuits of the second detector, together with shielding and proper placement of circuit elements.

Cross-talk.—Cross-talk (or cross-modulation) is the name given to interaction between radio signals in which the modulation of an undesired signal is transferred to a carrier wave on some other frequency. The following example represents a typical form of cross-modulation: The receiver is tuned to a powerful local station—the "desired" signal—which is so strong as to require a low-gain condition of the automatic volume control. At the same time there is another powerful local station—the "unwanted" signal—operating on a frequency not greatly different from that of the station being received. During the interval in which the desired station is sending out an unmodulated carrier wave, the modulation of the unwanted signal will be heard. However, if the desired station ceases to radiate its carrier wave, the interfering signals from the unwanted station disappear. This type of cross-modulation is a result of third-order curvature in the tube characteristics, as explained in Sec. 51.

Cross-modulation troubles, once a major design factor, have been virtually eliminated by the development of the variable-mu tube. Such tubes have very little third-order curvature, irrespective of the grid bias, and so can have their amplification controlled by variation of the bias without at the same time introducing cross-talk. This is in contrast with sharp cut-off tubes, which have a comparatively large third-order effect near cut-off.

Alternating-current Hum.—An alternating-current power-line hum often appears in the output of radio receivers energized by alternating current. The chief causes of such hum are ripple in the output of the rectifier-filter system and in the field of the loud-speaker, hum pick-up by the input circuits of the first audio-frequency tube, and power-line disturbances that are coupled into the receiver through the power transformer. Hum from rectifier ripple and from speaker field can be minimized by providing the filter choke with a hum-bucking coil that is connected in series with the voice coil of the loud-speaker, as in Fig. 175. By suitable design, the ripple voltage induced in this bucking coil will neutralize a considerable part of the hum voltage in the amplifier output. Hum pick-up by the input circuits to the audio-frequency system can be prevented by shielding the grid lead, or by careful placement with respect to wires carrying alternating currents.

Noise.—A sensitive radio receiver, operated with the antenna disconnected and adjusted for maximum volume, produces a continuous hiss in the output. In a properly designed receiver provided with means for preventing the entrance of power-line noise, this hiss arises primarily from thermal-agitation voltages developed across the first tuned circuit in the radio receiver. The adequacy of the design from the point of view of noise can therefore be checked by short-circuiting this input circuit and noting whether or not the hiss is appreciably reduced in intensity. The thermal-agitation voltages developed in the input resonant circuit set an ultimate limit to the receiver sensitivity that can be usefully employed. In the case of broadcast and other radio-telephone receivers this maximum usable sensitivity lies between 1 and 10μv.

Connecting an antenna to the receiver input terminals ordinarily introduces noise in the form of crackles, hiss, etc., resulting from natural and man-made electrical disturbances. Under conditions favorable for long-distance reception this additional noise collected by the antenna should exceed the hiss generated within the radio receiver. If not, either the antenna has inadequate energy pick-up, or the coupling arrangement between receiver and antenna does not have the proper efficiency.

Noise of man-made origin can often be reduced by careful attention to details. Thus in the case of a home receiver, such noise is commonly propagated along power lines and so can be reduced by the proper location of the receiving antenna, and by employing shielded power transformers. In some cases it is also helpful to place a radio-frequency filter in the power line where it enters the receiver chassis.

The noise problem is particularly difficult when a sensitive receiver is operated in the immediate vicinity of an internal-combustion engine using spark ignition, as in the case of automobile and aircraft radio receivers. Receivers for use under such conditions are normally mounted in a metal case to provide shielding, and all leads (except the antenna lead-in) entering this case are provided with radio-frequency filters. These precautions, combined with proper shielding of the ignition system and its associated wiring, together with careful location of the antenna, and the use of a shielded antenna lead-in wire, will reduce the ignition noise down to the same order of magnitude as the noise level from other sources.

Acoustic Feedback.—When the loud-speaker operates in the immediate vicinity of its radio receiver, there is always the possibility that sound vibrations transmitted to the receiver through the air or the receiver cabinet will modulate the carrier wave being received. Under unfavorable circumstances the sound produced in the loud-speaker by this modulation will exceed the sound causing the original modulation. When this is the case, the process continues in a cumulative manner, and a sustained audio-frequency howl results. This is termed *acoustic feedback*, and is particularly troublesome in all-wave receivers.

The parts of a receiver most likely to introduce acoustic feedback are the plates of variable condensers, and the radio-frequency and oscillator coils. Tubes are also sometimes involved, although in recent years great improvements have been made in the development of tubes that are relatively immune to mechanical vibration.

The remedies for acoustic feedback include the mounting of the receiver chassis on rubber feet to prevent mechanical vibrations being transmitted to the receiver from the cabinet, the use of rigid construction, and in some cases the mounting of the tuning condenser and coils on a separate sub-assembly that is supported on rubber. In the console models the chassis can also be placed on the top of a shelf that helps protect the receiver against sound vibrations produced by a loud-speaker mounted below the shelf.

101. Alignment.—All modern broadcast receivers are tuned by means of a single control. This means that the resonant frequency of all tuned circuits in the radio-frequency section must vary together, and at the same time the oscillator frequency must always differ from the frequency of the radio-frequency section by exactly the intermediate frequency. As a consequence, the provisions for making the circuits of the receiver align, or "track," are of fundamental importance.

The usual procedure is to use gang tuning condensers in which the various sections are as nearly identical as possible, and then to attain the necessary tracking by the use of trimmer condensers and by proper coil inductances. In a receiver covering only the regular broadcast band, the radio-frequency stages are aligned at the high-frequency end of the band by means of adjustable shunt padding condensers, as illustrated in Fig. 181*a*. At the low-frequency end of the band exact alignment is obtained by bending the end plates of the condensers. When care is

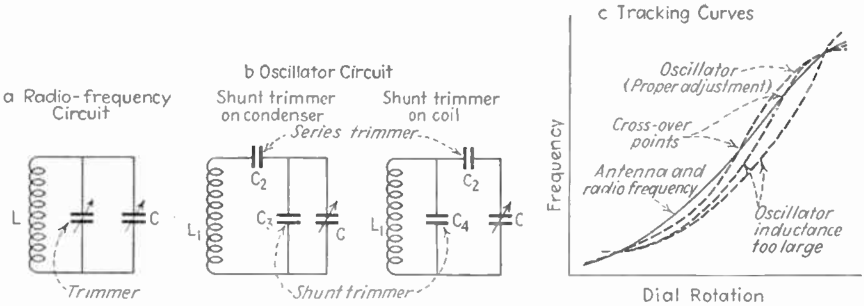


FIG. 181.—Trimmer systems for circuits of superheterodyne radio receiver together with schematic curves showing kind of tracking obtained for typical cases. The solid curve in *c* represents perfect tracking with antenna and radio-frequency circuits, while the dotted curves give the actual tracking obtained under various conditions.

taken to insure that the coils and condensers are initially very nearly alike, satisfactory tracking will then be obtained over the entire band, except possibly for the antenna circuit, where the wide variety of antenna constants encountered makes perfect alignment under all conditions impossible.

The oscillator of a superheterodyne receiver is ordinarily tuned by a condenser that is identical with the condenser gangs used for the radio-frequency circuits. The required difference frequency is then obtained by using series and parallel trimmer condensers as shown in Fig. 181*b*, together with the proper choice of tuning inductance. Analysis of the circuits of Fig. 181*a* and 181*b* shows that with suitable values for C_2, C_3 or C_4, and L_1/L it is possible to make the difference in the resonant frequency of the radio-frequency and oscillator circuits exactly equal to an assigned value at three settings of the tuning condenser. Furthermore, if two of these frequencies of perfect tracking are near the extremes of the tuning range, and the third is located near the middle, then the actual tracking will be almost perfect throughout the tuning range, as shown in Fig. 181*c*.

Tracking Problem in All-wave Receivers.—In all-wave receivers the tracking problem is complicated by the fact that it is not possible to bend the end plates of a condenser gang to line up the low-frequency end of one band without thereby throwing out the adjustment for all other bands. The usual arrangement in all-wave receivers is to employ separate trimmers associated with each coil. In the radio-frequency section each coil has its individual shunt trimmer that is used to align the high-frequency end of the tuning band for which the coil is used. Accurate control of coil inductance, maximum condenser capacitance, stray capacitance, etc., is then ordinarily depended upon to insure proper lining up at the low-frequency end of each tuning range.[1]

Each oscillator coil of an all-wave receiver is provided with individual series and shunt trimming condensers, following one of the arrangements illustrated in Fig. 181*b*. The shunt trimmers are individually adjusted so that the oscillator frequency tracks properly at the high-frequency end of each band. The series trimmers are usually fixed except for the broadcast band, and careful manufacturing control of the coil and condenser constants, or adjustment of the oscillator coil inductance, is depended upon to obtain alignment at the low-frequency end of each band.

Experimental Procedure for Aligning Receivers.—The alignment of a receiver can be readily carried out experimentally by using a test oscillator and some form of output indicator.[2] The first step is to line up the intermediate-frequency amplifier. This is done by setting the test oscillator to the intermediate frequency and adjusting the resonant frequency of each intermediate-frequency stage, one after the other, starting with the tuned circuits immediately preceding the second detector and working back toward the first detector. During this process the test oscillator is always connected to the grid of the tube immediately preceding the tuned circuits that are under adjustment at the moment.

The next step is to align the radio-frequency and oscillator circuits for the broadcast band. The receiver dial is turned to a point somewhere near the high-frequency end of the band, and the test oscillator set to the

[1] In some cases, particularly in the high-frequency bands where a small change in the inductance of the lead wires represents a considerable percentage change in the total inductance of the tuned circuit, it is necessary or desirable to adjust the circuit inductance individually for each coil by varying the position of an end turn of the coil, or by adjustment of the lead wires.

[2] The test oscillator is very often provided with a frequency wobbler or frequency modulator that automatically varies the frequency of the test oscillator over a range of 20 to 40 kc at a rate of 20 to 60 times per second. By then observing the output on a cathode-ray tube, it is possible to determine the adjustment giving the most satisfactory shape of response curve, as well as the adjustment for maximum output for a single frequency.

frequency indicated on the dial. The test oscillator is next connected to the antenna input and adjusted so that only a small output is obtained. The shunt trimming condensers on the radio-frequency stages are then adjusted until the output is maximum, after which the shunt trimmer on the oscillator condenser is likewise varied to give maximum output.[1] The receiver dial and test oscillator are then set to a point near the low-frequency end of the band, and the series padding condenser of the oscillator is adjusted for maximum response. If the receiver covers only the regular broadcast band, it is also permissible to improve the alignment of the radio-frequency stages by bending the end plates of the condensers. Finally, receiver and test oscillator are reset to the original high frequency, and the shunt trimmer condenser on the oscillator is checked to make sure that the change in series padding condenser has not affected the required shunting condenser. The receiver is now perfectly aligned at the high-frequency end of the band, and the oscillator tracks at both high- and low-frequency ends and also at an in-between frequency determined by the oscillator inductance that the manufacturer placed in the receiver.

In all-wave receivers this process is repeated for each band, except that for the higher frequency bands the series trimmer on the oscillator is usually non-adjustable.

102. Receivers for Telegraph Signals.—Receivers for handling telegraph signals differ from broadcast receivers in that some means must be provided for interrupting the dots and dashes at an audible rate so that reception can be obtained with a telephone receiver.[2] This is accomplished by the heterodyne principle, the procedure being to combine the incoming signal with a locally generated oscillation differing in frequency by about 1000 cycles. Rectification of the combination gives the telegraph code characters with a frequency of about 1000 cycles, as explained in Sec. 79.

A simple telegraph receiver is shown in Fig. 182 and consists of an oscillating detector of the type discussed in Sec. 81, followed by one stage of audio-frequency amplification. Such an arrangement possesses remarkable sensitivity because of the very large regenerative amplification

[1] All oscillator adjustments are preferably made while rocking the receiver tuning dial slightly in order to take care of any slight interaction between the oscillator and radio-frequency stages.

[2] The only exceptions to this are when the signals are produced by spark transmitters, or are in the form of interrupted continuous waves, or are to be recorded on a tape recorder. Spark and I.C.W. signals are interrupted at the transmitter and so can be received on an ordinary broadcast receiver. In the case of tape recording it is possible to operate the inking mechanism without producing an audible tone, although most of the commercial systems of tape recording involve an audible note that is rectified to operate the inking mechanism.

obtained with a properly adjusted oscillating detector. In some cases the receiver of Fig. 182 is modified by placing a radio-frequency amplifier tube ahead of the oscillating detector, using either a tuned or untuned coupling between the antenna and the grid of this tube. Such a tube is primarily for the purpose of making the local oscillator frequency independent of the receiving antenna and for preventing the radiation of local oscillator energy, but it also adds somewhat to the sensitivity.

Any ordinary superheterodyne receiver can be converted into a telegraph receiver by the addition of a local oscillator for the purpose of heterodyning with the intermediate frequency. This oscillator is detuned approximately 1000 cycles from the intermediate frequency and is loosely coupled into the circuits of the second detector, thereby causing the latter to develop a beat note of approximately 1000 cycles with the incoming carrier wave. If the superheterodyne receiver is provided with automatic volume control, this must be disconnected in order to prevent

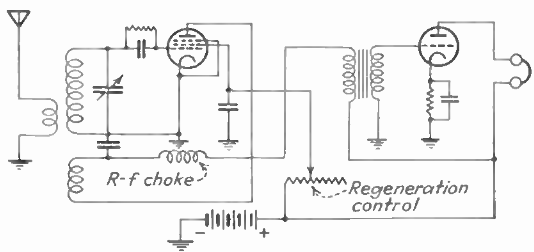


FIG. 182.—Simple telegraph receiver consisting of an oscillating detector followed by one stage of audio-frequency amplification.

the local oscillator from affecting the receiver sensitivity, and a manual volume control must be substituted.

Single-signal Receiver.—With the ordinary heterodyne radio-telegraph receiver there are two possible signal frequencies that will give the same difference frequency. Thus if the desired beat note is 1000 cycles, and the beating oscillator is adjusted to a frequency 1000 cycles less than the desired signal, then another signal having a frequency 2000 cycles less than the desired signal will also give a beat note of 1000 cycles. Image-signal interference of this type can be eliminated by using a superheterodyne receiver in which the intermediate-frequency amplifier possesses such great selectivity that signals differing by only a few hundred cycles from the desired signal are prevented from reaching the second detector. Such an arrangement is termed a *single-signal receiver*.

The usual method of obtaining the high selectivity required to achieve the single-signal effect is to employ a quartz crystal somewhere in the intermediate-frequency amplifier. A typical arrangement is illustrated in Fig. 183*a*. Here the crystal is used as the coupling arrangement between two conventional intermediate-frequency tuned circuits. By replacing the crystal with its equivalent electrical circuit it is seen that this gives a coefficient of coupling between the two circuits that varies with frequency, being very low at the series resonant frequency of the crystal, and very high at frequencies appreciably different. The resulting response curve has the character shown in Fig. 183*b*. The width of the peak of high response can be controlled somewhat by the tuning of the resonant circuit applying voltage to the crystal. Also the location of the frequency of very low response can be adjusted to eliminate the image signal by means of the "rejector" control that varies the equivalent unneutralized capacitance between the plates of the crystal holder.

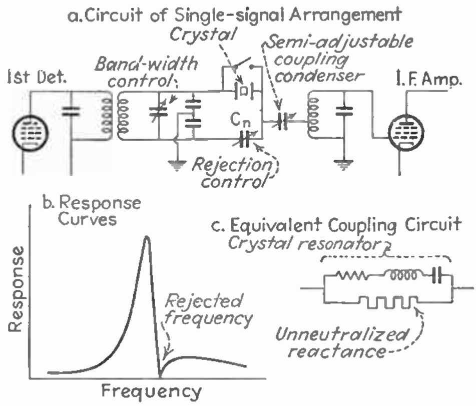


Fig. 183.—Crystal coupling arrangement used in single-signal receivers together with response curve illustrating how the image interference can be suppressed.

A single-signal receiver such as illustrated in Fig. 183 can be converted into an ordinary superheterodyne receiver for the reception of radio-telephone signals by merely short-circuiting the crystal.

103. Miscellaneous Types of Receivers. *Radio-telephone Receivers for Other than Broadcast Purposes.*—Telephone receivers for other than broadcast purposes are in the main similar to broadcast receivers, but with modifications of the tuning range and of the design details. Thus many airplane receivers are similar to automobile radio receivers in the general method of construction, in the use of remote-control tuning, and in that a storage battery is the source of power. The chief differences are that the airplane receiver requires somewhat sturdier construction, covers only the airplane bands, and is arranged to operate a telephone receiver instead of a loud-speaker. In some cases airplane receivers are arranged so that only two frequencies can be received, one being the short-wave day frequency, and the other the short-wave night frequency, assigned to the air line involved. Superheterodyne receivers for such two-band operation sometimes employ crystal oscillators for the frequency-changing process, with a separate crystal for each frequency.

Receivers for long-wave telephone signals may consist of two or three stages of radio-frequency amplification, followed by a detector and audio amplifier, or may be of the superheterodyne type having an intermediate frequency that is either quite low or higher than the highest frequency to be received.

Triple-detection Receivers.—The ordinary superheterodyne receiver tends to give relatively poor image suppression at short waves unless an excessive number of tuned circuits is employed in the radio-frequency section. This difficulty can be overcome by employing a triple-detection receiver as illustrated schematically in Fig. 184. The first intermediate frequency of such a receiver is made high enough to provide adequate

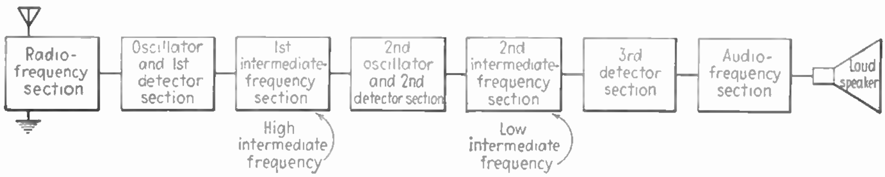


FIG. 184.—Schematic diagram of triple-detection receiver.

image suppression with one or two resonant circuits in the radio-frequency section. The second intermediate frequency is commonly the same as in broadcast receivers and makes it possible to discriminate against signals differing only slightly from the desired frequency. The second oscillator of a triple-detection receiver operates at a fixed frequency determined by the two intermediate frequencies, so that the only circuits that must be adjusted for the incoming signal are the radio-frequency and first-oscillator sections.

Receivers for Ultra-high-frequency Signals.—The most satisfactory receiver for signals in the wave-length range of ½ to 10 meters is the superheterodyne. The radio-frequency and oscillator stages of such an arrangement preferably employ "acorn" tubes, particularly if the frequency is less than 5 meters. The intermediate frequency should be relatively high in order to give good image suppression, and the intermediate-frequency amplifier should respond to a relatively wide frequency band, so that the signals will not be lost if there is a moderate amount of frequency drift in the transmitter or the beating oscillator.

Superregenerative detectors are often employed in receivers operating at wave lengths below 10 meters. The simplest arrangement of this type consists of a superregenerative detector followed by one or two stages of audio-frequency amplification. In more elaborate arrangements one or two tuned radio-frequency stages are placed between the antenna and the superregenerative detector in order to increase selectivity and prevent radiation from the receiver. Superregenerative receivers are remarkably sensitive and have the advantage of producing nearly as much audio-frequency output with very weak signals as with strong signals, so that there is an inherent tendency to discriminate against ignition and other similar interference that is in the form of pulses of high intensity but short duration. The chief disadvantage of the superregenerative arrangement is that a strong hiss is present except when a carrier wave is being received.

Reception of Frequency-modulated Signals.—Signals that are frequency (or phase) modulated can be received by discriminating against one side band so that the carrier and remaining side band heterodyne together to give a beat note corresponding to the original modulation.[1] Any ordinary receiver can accordingly be used to receive frequency-modulated signals by detuning the receiver so that the carrier is on the side of the response curve as illustrated in Fig. 185.

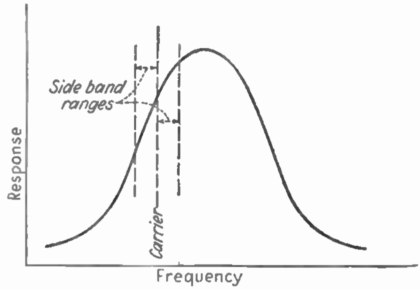


FIG. 185.—Response curve of receiver, showing how one side band of a frequency-modulated wave can be discriminated against by detuning the receiver slightly.

104. Receiving Systems for Minimizing Fading.—Short-wave signals received from a distant transmitter practically always vary in strength (*i.e.*, fade) in a random manner. Night signals from broadcast transmitters not too close to the receiving point likewise fade. The variation in signal strength commonly encountered with fading signals is such as seriously to impair the reception, and furthermore severe fading of radiotelephone signals is usually accompanied by marked quality distortion.

Automatic volume control is an obvious means of minimizing the effects of fading and is used in nearly all broadcast receivers. This is not a complete solution of the problem, however, because the signals sometimes fade out so completely as to be lost in the background noise. Also, as the signals fade in and out, the sensitivity of the receiver changes, resulting in a continual variation in the background noise. Furthermore, automatic volume control does not prevent the quality distortion of radiotelephone signals that usually accompanies fading.

Diversity Receiving Systems.—The most effective means of eliminating fading troubles consists in taking advantage of the experimentally

[1] The reception of frequency-modulated signals by detuning the receiver is sometimes explained on the basis that, as the frequency of the signal is varied, the response of the radio-frequency circuits varies likewise, thus converting the frequency-modulated wave into an amplitude-modulated wave. This view is only partially correct, however, since the frequency-modulated wave contains carrier and side-band components as explained in Sec. 75.

observed fact that signals induced in antennas spaced 10 wave lengths or more apart fade independently. Hence by combining the outputs of three suitably spaced receiving antennas, the probability of the signals fading out completely on all three antennas at the same instant is almost zero. Such an arrangement is known as a *diversity receiving system* employing space diversity and is commonly used on transoceanic radio-telephone and radio-telegraph circuits. In such a diversity receiving system, each antenna is provided with a separate receiver, and the combination takes place after detection has occurred. In order that the noise level in the combined output will not be excessive, it is customary to operate all receivers from a common automatic-volume-control system so that the sensitivity of all three receivers is controlled by the particular receiver in which the signal happens at the moment to be strongest. In this way a receiver that at the moment is getting very little signal will not as a consequence have its gain increased and so bring in excessive noise.

Fading can also be minimized by taking advantage of the fact that signals transmitted on frequencies differing by some 500 cycles or more will ordinarily fade independently. Arrangements that minimize fading troubles in this way are said to employ frequency diversity. In telegraph transmitters frequency diversity can be realized by employing interrupted continuous waves (I.C.W.), or by modulating the transmitted wave at a frequency of the order of 500 to 1000 cycles. In either case, the amplitude variations of the transmitted wave introduce side-band frequencies and cause the telegraph message to be transmitted simultaneously on a number of frequencies. The result is reduced probability of a complete fade-out at the receiver.

Problems

1. Discuss the parts of a superheterodyne receiver that can cause the selectivity, fidelity, and sensitivity to vary with the frequency being received.

2. Make a detail drawing of the means used to obtain the bias for the first audio tube and the power tube of the receiver of Fig. 173, number each circuit element, explain the function of each numbered part, and outline means of calculating the bias voltages.

3. Make a detail drawing of the A.V.C. and manual volume-control system used in the receiver of Fig. 174, number each circuit element and explain its function.

4. Make an estimate of the total direct-current voltage and current that the output of the power-supply system must supply in the receiver of Fig. 174, making use of data from tube manuals.

5. Make a detail drawing of the converter system used in the receiver of Fig. 174, number each circuit element, and explain the function of each numbered element.

6. In the receiver of Fig. 175, what determines the bias on the radio-frequency and intermediate-frequency tubes when there is no carrier and hence no bias from the A.V.C. system?

7. Make a detail drawing of the means used to obtain screen voltages for the pentode tubes (excluding the 6F6 tubes) in Fig. 175, number each circuit element in the diagram, and explain the purpose of each numbered part.

8. Compare the methods used to obtain the screen voltages for the intermediate-frequency amplifier tubes in the receivers of Figs. 173 and 174.

9. Assuming that the speaker field of the receiver of Fig. 174 has an incremental inductance of 8 henries, calculate the ratio of alternating-current voltage in the plate supply of the output tube to the alternating-current ripple voltage appearing across the input condenser of the rectifier-filter system.

10. Describe the method by which the tone control of the receiver in Fig. 174 operates.

11. In the automatic-frequency-control system of Fig. 177, describe the arrangement by which the grid of the control tube receives an exciting voltage 90° out of phase with the voltage acting in the plate.

12. Estimate the sensitivity of the receiver of Fig. 174 by making reasonable estimates of the gain per stage.

13. Plan the tube line-up for a radio receiver that will have a sensitivity of 3 μv and is capable of developing 15 watts of undistorted output power. Indicate tube types to be used, and the stage gains.

14. If a broadcast receiver is designed so that it discriminates very effectively against carrier frequencies only slightly different from the desired carrier, it is found that the fidelity at high audio frequencies is poor. Explain.

15. A receiver covering the frequency range 530 to 1650 kc has an intermediate frequency of 460 kc. List the conditions and frequencies for which whistles or spurious responses are most likely to occur.

16. A receiver that has a tendency toward acoustic feedback also tends to have excessive microphonic noise when used under conditions where mechanical vibration is present. Explain.

17. What effect can improper alignment in a receiver be expected to have on: (*a*) sensitivity, (*b*) fidelity; and (*c*) selectivity?

18. Work out the circuit of a telegraph receiver consisting of one stage of tuned radio-frequency amplification, an oscillating detector, and a two-stage audio-frequency amplifier.

19. When a single-signal receiver is converted into an ordinary receiver by short-circuiting the crystal, the noise level in the output increases appreciably. Explain.

20. When an ordinary receiver is used to receive a frequency-modulated signal by tuning so that the carrier is on the side of the resonance curve, it is found that there are two carrier frequencies for which the reception of frequency-modulated waves is equally efficient. Explain how these two responses occur, and suggest a means whereby the receiver can be modified to eliminate or greatly reduce the reception of this second carrier.

21. It is desired to use automatic frequency control in a triple detection receiver to take care of slight mistunings that might accidentally occur. Should the automatic frequency control be applied to the first oscillator, the second oscillator, or both oscillators?

CHAPTER XIII

PROPAGATION OF RADIO WAVES

105. Ground-wave Propagation.—The energy radiated from an antenna located near the earth's surface can be conveniently divided into a *sky wave* and a *ground wave.* The sky wave represents the energy radiated at an angle above the horizontal, and its ultimate behavior is determined by the ionized regions in the upper atmosphere, as discussed in Sec. 106. The ground wave, on the other hand, represents the energy traveling along the surface of the earth, and its behavior depends upon the character of the surface involved, and the frequency.

The ground wave glides over the surface of the earth as shown in Fig. 1. The wave is vertically polarized because a horizontally polarized component in the immediate vicinity of the earth has its electrostatic field short-circuited by the ground. The ground wave is accompanied by charges induced in the earth as indicated in Fig. 1. These moving charges constitute a current, and, since the earth offers resistance to the flow of current, there is a dissipation of energy in the earth that represents energy absorbed from the ground wave. The portion of the wave in contact with the earth is therefore being continuously wiped out, only to be replenished at least in part by diffraction of energy downward from the portions of the ground wave immediately above the earth.

Sommerfeld Analysis of Ground-wave Propagation.—The actual strength of the wave at the surface of the ground for a *flat* earth is

$$\left.\begin{array}{l}\text{Ground-wave field strength}\\ \text{in millivolts per meter}\end{array}\right\} = \frac{300\sqrt{P}}{d}\gamma A \qquad (111)$$

where

A = factor taking into account the ground loss
d = distance to antenna in kilometers (1 mile = 1.609 km)
P = radiated power in kilowatts
γ = factor taking into account directivity of transmitting antenna.

The factor γ is unity when the radiated field is proportional to the cosine of the angle above the horizontal (*i.e.*, $\gamma = 1$ for short vertical antenna). For other antenna directivities γ will equal the ratio of actual field strength radiated along the horizontal by the antenna to the field strength that would be obtained for the same radiated power with the cosine law directivity.

The factor A in Eq. (111) takes into account the effect of ground loss and depends in a relatively complicated way upon the conductivity and dielectric constant of the earth, the frequency, and the distance to the transmitter. The relationships involved are presented in graphical form in Fig. 186, which is based upon a modification of the original analysis made by A. Sommerfeld. In Fig. 186 the reduction factor A is expressed in terms of two auxiliary variables, the numerical distance p, and the phase constant b. These auxiliary constants p and b are determined by the frequency, distance, and characteristics of the earth when considered as a conductor of radio-frequency currents.

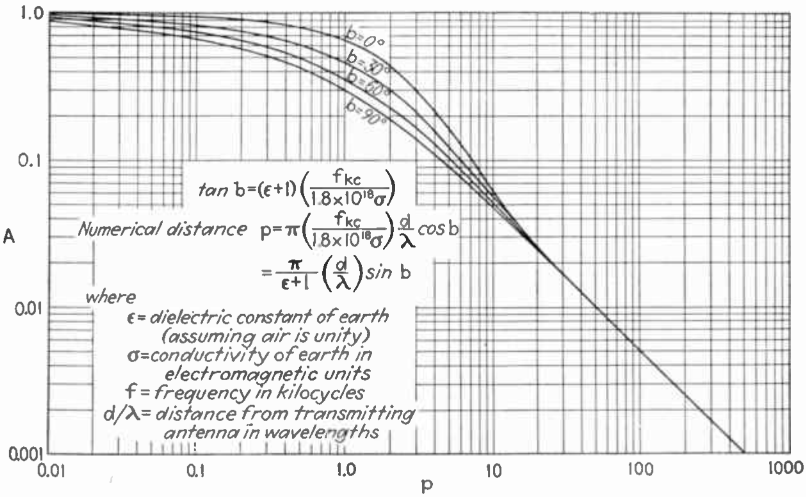


FIG. 186.—Sommerfeld reduction factor A as a function of the numerical distance p for various values of phase constant b.

Examination of Fig. 186 shows that the reduction factor A depends primarily upon the numerical distance p. When p is less than unity the factor A differs only slightly from unity. The losses in the earth then have little effect on the strength of the wave, which is accordingly nearly inversely proportional to distance. However, as the numerical distance p becomes greater than unity, the factor A decreases rapidly, until, for $p > 10$, the factor is almost exactly equal to $1/2p$. Inasmuch as the numerical distance p is proportional to actual physical distance, this means that when $p > 10$, the factor A is inversely proportional to distance, making the field strength of the ground wave inversely proportional to the *square of the distance*.

The use of Fig. 186 in practical radio problems is illustrated by the following example.

Example.—A police radio transmitter operating at a frequency of 1690 kc is required to provide a ground wave having a strength of at least 0.5 mv per meter at a distance of 10 miles (16 km). The transmitting antenna is expected to have an efficiency of 50 per cent, *i.e.*, radiates 50 per cent of the energy actually delivered to it, and produces a radiated field that is proportional to the cosine of the angle of elevation. The ground is such that a conductivity of 5×10^{-14} e.m.u. and a dielectric constant of 15 can be expected. Determine the transmitter power required.

The first step in the solution is to evaluate the factor A for the conditions of the problem. Reference to the equations in Fig. 186 gives

$$\begin{aligned} \tan b &= (15 + 1)\frac{1690}{1.8 \times 10^{18} \times 5 \times 10^{-14}} = 0.301 \\ b &= 16.7^\circ \\ p &= \pi\left(\frac{1690}{1.8 \times 10^{18} \times 5 \times 10^{-14}}\right)\frac{16 \times 1690 \times 10^{3}}{3 \times 10^{5}}\,0.957 \\ &= 5.1 \end{aligned}$$

From Fig. 186,

$$A = 0.15$$

Substitution in Eq. (111) then gives

$$0.5 = \frac{300\sqrt{P} \times 1}{16}\,0.15$$

from which

$$P = 0.0315 \text{ kw}$$

With an antenna efficiency of 50 per cent the transmitter must deliver 63 watts to the antenna.

Curves showing ground-wave attenuation as a function of distance for different wave lengths and earth conductivities are given in Fig. 187 for the case of a transmitting antenna radiating 1 kw of power and having a directional characteristic such that the strength of the radiated field is proportional to the cosine of the angle of elevation above the horizontal. For other powers the field strength will be directly proportional to the square root of the power, and with other directivities the field strength must be multiplied by a factor giving the relative horizontal field strength for 1 kw radiated in the actual case compared with the field strength for the assumed directivity. Thus a 50-kw broadcast transmitter, employing an antenna that increases the field strength along the horizontal to 1.41 times the field strength obtained with the assumed cosine law, would make the radiated field $1.41\sqrt{50} = 10$ times the value obtained from Fig. 187. For purposes of comparison, inverse-distance curves corresponding to zero earth losses and a flat earth are shown.

The curves of Fig. 187 have been corrected to take into account the curvature of the earth, which Eq. (111) and Fig. 186 neglect by assuming a plane surface. This correction is small at distances up to a few hundred miles, but thereafter makes the ground wave somewhat weaker than calculated with the aid of Eq. (111).

Discussion of Ground-wave Propagation.—The conductivity and dielectric constant effective in the equations of Fig. 186 are the average values for the earth extending down to depth of a few feet to over a hundred feet, according to the frequency. The dielectric constant and conductivity can be most satisfactorily derived from data on ground-wave propagation taken over the desired path at a convenient and not too different frequency. Knowing the frequency, distance, and rate of attenuation, it is possible to work backward and find the earth constants required to account for the results. The values so obtained can then

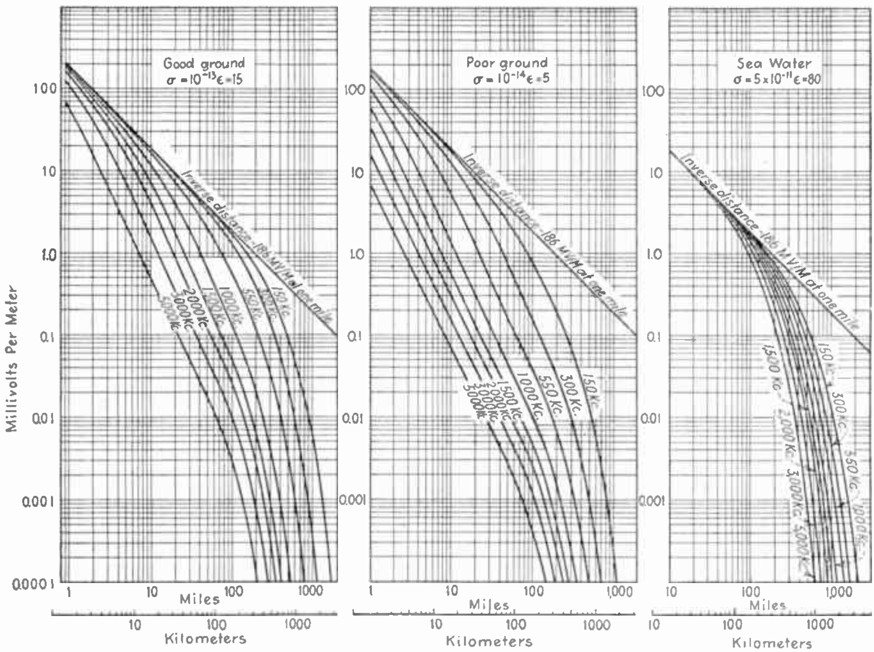


FIG. 187.—Strength of ground wave as a function of distance, frequency, and ground conductivity for 1 kw of radiated power from an antenna producing a radiated field proportional to the cosine of the vertical angle. (For more complete curves see *Proc. I.R.E.*, vol. 21, p. 1419, October, 1933.)

be used to determine the ground-wave propagation of waves of other not too different frequencies over the same general path.

The earth conductivity varies greatly under different conditions, ranging from about 30×10^{-14} electromagnetic units for moist loam to values of the order of 10^{-14} for dry sandy and rocky soil. The conductivity of salt water depends on the temperature and the saline content, and averages about 500×10^{-14} electromagnetic units. The dielectric constant of earth commonly ranges from 5 to 30, with the low values tending to go with dry soils of poor conductivity and the larger values with moist conducting earth. The dielectric constant of water is approximately 80.

At broadcast and lower frequencies, which is where the practical importance of ground-wave signals is greatest, the earth impedance is primarily resistive for the usual ground conductivities. Under such conditions the reduction factor A in Eq. (111) is very sensitive to the earth conductivity and the frequency but is almost independent of the dielectric constant. Strong ground-wave signals at considerable distance are then obtained only when the earth conductivity is high, or the frequency is low. In particular, it is to be noted that a salt-water path is tremendously superior even to good earth, and this is in turn very much better than poorly conducting earth. An idea of the importance of conductivity and frequency in determining ground-wave signal strength can be gained by study of Fig. 187.

106. The Ionosphere and Its Effect upon the Sky Wave.—After leaving the transmitting antenna, the sky wave travels through space

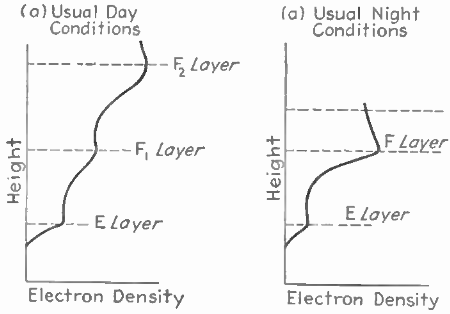


FIG. 188.—Schematic diagram illustrating the variation of electron density with height above earth under typical conditions.

with a field strength inversely proportional to distance, and ultimately reaches the ionized region in the outer portion of the earth's atmosphere. Here, if conditions are favorable, the path of the wave will be bent earthward. In this way it is possible for waves to travel around the curvature of the earth and reach points much too distant to receive appreciable ground wave.

Nature of the Ionosphere.—The term ionosphere designates the ionized region that exists in the outer portion of the earth's atmosphere. This is also sometimes called the Kennelly-Heaviside layer, after the two scientists who independently and almost simultaneously suggested the existence of such a region. The ionosphere results from ionization of the outer atmosphere by solar radiation, and consists of a mixture of free electrons, positive ions, and negative ions, in a rarefied gas.

The effect that the ionized region has on radio waves can be thought of as being caused by the free electrons, and is determined by the distribution of the electron density with height. This distribution has the character shown schematically in Fig. 188 and is normally characterized by several distinct maxima, or "layers." In the daytime there are normally three such maxima of progressively increasing electron density, as at Fig. 188*a*. On the other hand, at night the middle or F_1 layer normally fades out, and the upper or F_2 layer descends to occupy approximately the same location as the daytime F_1 layer, to give the result shown in Fig. 188*b*.

Average heights of these layers under typical conditions are shown in Fig. 189. The lowest or E layer maintains a height of the order of 100 to 120 km with little seasonal or diurnal change. The F_1 layer normally occurs at 200 to 250 km, and likewise shows only a small seasonal and diurnal variation. On the other hand, the height of the F_2 layer varies greatly with the season and time of day, with typical heights being of the order of 250 to 350 km. The night F (or F_2) layer ordinarily has a height around 250 km, and like the F_1 layer, exhibits only moderate seasonal and diurnal fluctuations.

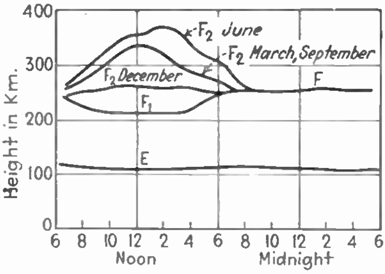


FIG. 189.—Average height of the ionized regions at Washington, D. C. during 1933–1934.

The maximum electron density in the layers varies with time of day and season as shown in Fig. 190. In the E and F_1 layers the maximum density is seen to follow a regular diurnal and seasonal cycle, being great-

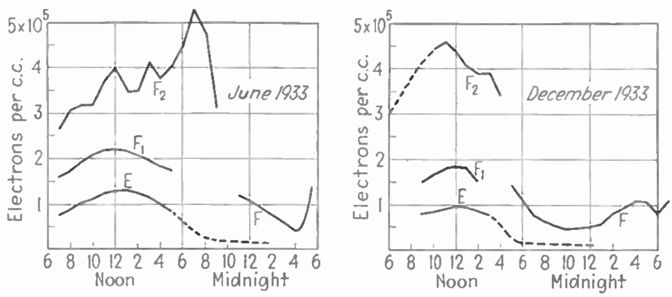


FIG. 190.—Average electron density of the various ionized layers at Washington, D. C., during 1933 (quiet part of sunspot cycle).

est at noon and tapering off at both earlier and later hours, and also being somewhat greater in summer than in winter. The maximum electron density in the F_2 layer shows a much larger diurnal and seasonal variation, and also shows a much larger random fluctuation from day to day. It will be noted that all the maximum electron densities decrease greatly during the night, presumably as a result of recombination in the absence of ionizing solar radiation.

In addition to seasonal effects, there is a long-time variation in the electron density of the entire ionosphere that is associated with the 11-year sunspot cycle. This is illustrated in Fig. 191, which shows marked changes in the electron density of the F_2 layer in successive years with distinct though lesser differences in the lower layers. The greatest electron densities occur during the most active sunspot periods.

The heights of the various layers are influenced very little by the sunspot cycle, although the F_1 layer sometimes disappears.

Mechanism by Which Ionosphere Affects Radio Waves.—The effect that an ionized region has on a radio wave can be understood by considering the behavior of a single ion or electron when under the influence of a passing radio wave. Take first the case of an electron in a vacuum with no magnetic field present other than the weak magnetic field of the wave.

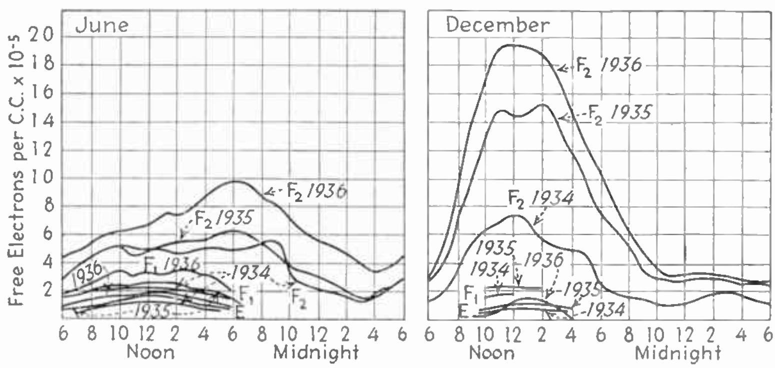


FIG. 191.—Electron densities of ionosphere on successive years, as observed at Washington, D. C.

The wave's electrostatic field exerts forces on the electron that cause the electron to vibrate along a path parallel with the flux lines of the wave. The kinetic energy of this vibration then represents energy abstracted from the radio wave. Since a moving charge is an electrical current, the vibrating electron acts as a small antenna which reradiates the energy it has acquired from the wave. However, the velocity with which the electron moves is 90° out of phase with the electrostatic field of the wave. There is, consequently, a phase difference between the reradiated field and the field of the passing wave. Because of this the resultant direction of energy flow is altered in such a manner as to cause the wave path to bend away from the regions of high electron density toward regions of lower density. A wave entering the ionosphere hence tends to be bent back toward the earth. Under favorable conditions the wave will return to the earth, as shown in Fig. 192.

The refraction (*i.e.*, bending) that the wave path suffers in the ionosphere is greater the lower the frequency. This is because the average velocity with which the electrons vibrate when acted upon by the passing wave is inversely proportional to frequency. Consequently the lower the frequency the greater will be the energy that is transferred from the wave to the electrons, to be ultimately reradiated and produce bending of the wave path.

The action of the ionosphere is also influenced to some extent by the earth's magnetic field. Thus at high frequencies the electrons, instead

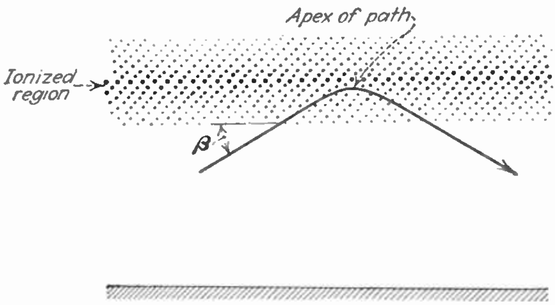


Fig. 192.—Diagram illustrating refraction of wave by ionosphere and also indicating notation used in Eq. (113).

of vibrating along a linear path as would be the case in the absence of the earth's magnetic field, follow an elliptical path as shown in Fig. 193, provided there is a component of the earth's magnetic field at right angles to the electrostatic field of the wave. This is due to the fact that such a magnetic field exerts a force on the moving electron that is at right angles to the direction of motion, as explained in Sec. 20. As a consequence of this elliptical path, the electron possesses a component of velocity that is at right angles to the direction of the electrostatic field of the wave. A certain amount of the energy reradiated by the electron then has a plane of polarization that is at right angles to the polarization of the passing radio wave. Consequently a wave that travels in the ionosphere will have its resultant plane of polarization changed.

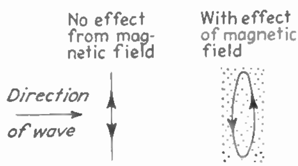


Fig. 193.—Paths followed by vibrating electron with and without the influence of the earth's magnetic field. The magnetic field is perpendicular to the plane of the paper.

A quantitative analysis shows that the presence of the magnetic field, in addition to affecting the polarization, also causes the wave to be split into two components termed the *ordinary* and *extraordinary* rays, which follow different paths, travel with different velocities, and suffer different attenuations. As a consequence, a wave that has passed through an ionized region will have both vertically and horizontally polarized components irrespective of the polarization of the waves radiated from the transmitter. Furthermore, the vertically and horizontally polarized components will in general not be in the same phase, *i.e.*, the wave will be elliptically polarized.

Refraction of Sky Waves to Earth by the Ionosphere.—The conditions required to return a sky wave to earth can be expressed in terms of the refractive index of the ionized region. The refractive index is determined primarily by the frequency and the electron density, although affected somewhat by the earth's magnetic field and by collisions between the vibrating electrons and gas molecules. If the effects produced by the magnetic field and the collisions are neglected, then

$$\text{Refractive index } \mu = \sqrt{1 - \frac{81N}{f^2}} \tag{112}$$

where

N = electron density in electrons per cubic centimeter

f = frequency in kilocycles.

Examination of Eq. (112) shows that the electrons in the ionosphere cause the refractive index to be less than the free-space value of unity by an amount that increases with the electron density and decreases with the frequency.

A wave entering the ionosphere with an angle of incidence β and returned to earth, as shown in Fig. 192, will, according to Snell's law, penetrate the ionized region to a point where the refractive index has a value μ_0 given by the equation

$$\left.\begin{matrix}\text{Refractive index} \\ \text{at apex of path}\end{matrix}\right\} = \mu_0 = \cos\beta \tag{113}$$

An examination of Eqs. (112) and (113) shows that as the frequency of the wave is increased while the angle of incidence is kept constant, the depth of penetration into the ionized region will increase. This is to be expected, since the higher the frequency the less is the effect of the electrons on the wave path, and hence the greater is the electron density required to return the wave to earth. The situation is accordingly as shown in Fig. 194. A low-frequency wave penetrates only slightly, as indicated by path *a*, because at low frequencies it takes only a small electron density to produce a low refractive index. As the frequency is raised the penetration becomes progressively greater (path *b*), until finally Eq. (113) can be satisfied only when the electron density corresponds to the maximum density of the lowest or *E* layer of the ionosphere. Under such conditions the wave penetrates to the middle of the *E* layer before being returned to earth, as shown by path *c* in Fig. 194. If the frequency is still higher, the wave passes through the *E* layer and enters the next higher layer, penetrating to a point where Eq. (113) is satisfied and then returning to earth as shown by path *d* in Fig. 194. However, if the frequency is so high that the maximum electron density in the higher layer is insufficient to satisfy Eq. (113), then the wave passes

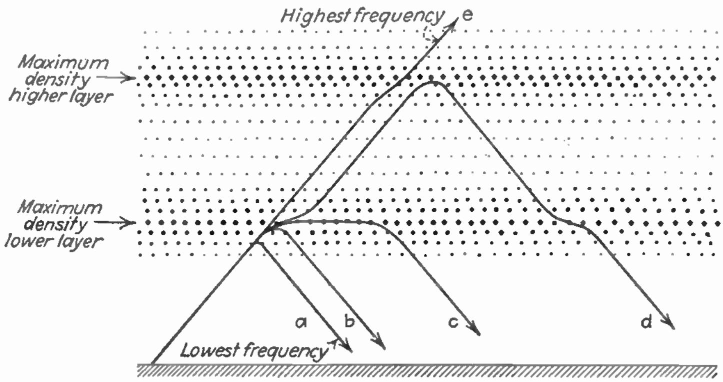


FIG. 194.—Wave paths followed in ionosphere for constant angle of incidence but progressively higher frequency.

through both layers, as indicated by path *e*, and will not return to earth unless there is a higher layer of still greater electron density.

A study of Eqs. (112) and (113) also shows that waves of a given frequency are returned to earth more readily by the ionosphere the more glancing (*i.e.*, the smaller) the angle of incidence. This is to be expected,

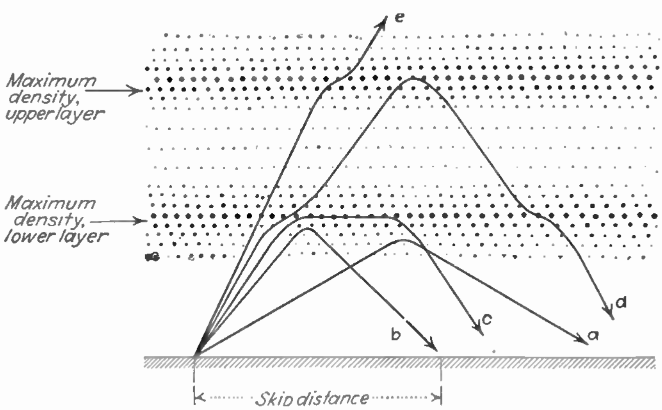


FIG. 195.—Wave paths followed in ionosphere for short waves of constant frequency but different angles of incidence, illustrating skip phenomenon.

since the angle through which the wave path must be bent becomes less as the angle of incidence is reduced. The situation that exists under typical conditions is accordingly as shown in Fig. 195. When the angle of incidence is small, as in the case of path *a*, the penetration of the ionosphere is very slight. However, as the angle of incidence increases, the electron density required to satisfy Eq. (113) becomes greater, and the wave follows a path such as *b* in Fig. 195. With still further increase in the angle of incidence a point is finally reached where Eq. (113) can be satisfied only when the electron density corresponds to the maximum density of the lowest or *E* layer. The wave is then barely returned to earth by this layer, as shown by path *c* in Fig. 195. With further approach to vertical incidence the wave penetrates the *E* layer and enters the next higher layer. Here it will be returned to earth from a point where the electron density is sufficient to satisfy Eq. (113), as shown by path *d*, or will pass on through to higher regions, as in path *e*, when the angle of incidence is too great for Eq. (113) to be satisfied even by the maximum electron density in the layer.

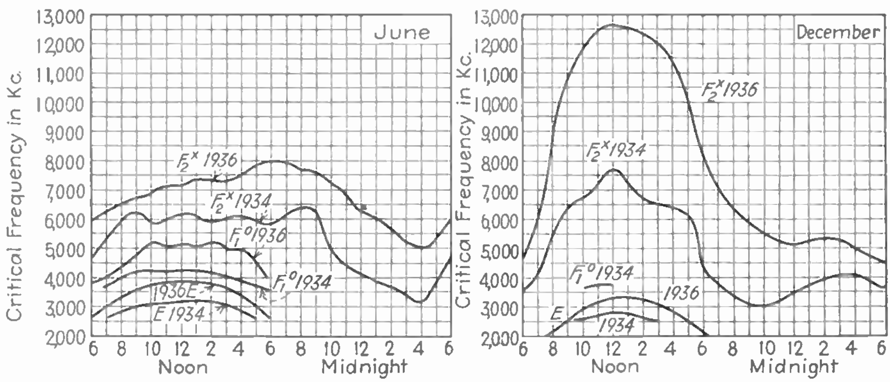


FIG. 196.—Critical frequencies of ionosphere under typical conditions. (Superscripts *x* and *o* denote extraordinary and ordinary rays, respectively.)

Critical Frequencies.—For each layer of the ionosphere there is a frequency, called the *critical frequency*, for which the refractive index is zero at the point of maximum electron density. *When the frequency of the sky wave is less than the critical value, the layer returns the wave to earth for all angles of incidence, even including normal incidence.* Each layer of the ionosphere has its own critical frequency corresponding to its maximum electron density, and as the electron density goes through diurnal and seasonal changes, the critical frequency does likewise. Since the F_2 (or F) layer has the greatest electron density, the critical frequency of this layer determines whether or not sky waves of any particular frequency will fail to be refracted back to earth when the radiation strikes the ionosphere with normal incidence. Critical frequencies of the various layers under typical conditions are shown in Fig. 196. It will be noted that these show diurnal, seasonal, and year-to-year changes, with the critical frequency of the F_2 layer being particularly variable.

Skip Phenomena and Skip Distance.—When the frequency exceeds the critical value of the F_2 (or F) layer, the ionosphere is not able to return waves back to earth in the immediate vicinity of the transmitter, because this would require that waves with vertical incidence be returned to earth. However, it may still be possible to refract waves to earth when the angle of incidence is somewhat glancing. Such waves will not return to earth until some distance from the transmitter, thus causing the sky wave to skip over the immediate vicinity of the transmitter. This is illustrated in Fig. 195, where path b represents the sky wave that returns to earth nearest the transmitter. The distance from the transmitter to the point where the first sky ray returns is called the *skip distance* and is indicated in Fig. 195. When the skip distance is appreciable, the ground wave will die out completely in much less than the skip distance. The skip region then contains a zone of silence where no signals are received, even though strong signals are obtained at greater distances.

The skip distance for a particular ionosphere layer can be calculated with fairly good accuracy by assuming that the wave undergoes a mirror-like reflection in the ionosphere at a height corresponding to the point of maximum electron density. When the skip distance is small enough so that the curvature of the earth can be neglected, this gives[1]

$$\left.\begin{array}{l}\text{Approximate skip distance}\\ \text{for flat earth}\end{array}\right\} = \frac{2\mu h}{\sqrt{1-\mu^2}} \tag{114a}$$

where h = height of layer involved and μ = refractive index at the point of maximum electron density of the layer involved.

The particular layer determining the skip distance observed at the earth is the layer having the shortest skip distance. In making skip-distance calculations it is accordingly necessary to make computations for each layer present that can return the wave to earth, and then choose the smallest result obtained. The layer that is effective will depend upon the frequency, the skip distance, and the relative heights and electron densities involved.

[1] When the skip distance is so large that the earth's curvature cannot be neglected Eq. (114*a*) takes the form

$$\mu^2 = \frac{\sin^2 (s/2r)}{\sin^2\left(\frac{s}{2r}\right) + \left[1 + \frac{h}{r} - \cos\left(\frac{s}{2r}\right)\right]^2} \tag{114b}$$

where s is the skip distance, r is the earth's radius, and μ is defined as in Eq. (114*a*). When using Eq. (114*b*) to determine the skip distance corresponding to a given frequency, the simplest procedure is to calculate the refractive index (and from this the frequency) corresponding to several values of assumed skip distance. These results can then be plotted and interpolation made for the frequency of interest.

Highest Frequency Returned to Earth.—The highest frequency that is returned to earth is the frequency for which it is barely possible for the maximum electron density of the F_2 (or F) layer to satisfy Eq. (113) when the angle of incidence at the ionosphere has the smallest value possible in view of the curvature of the earth. This highest frequency depends upon the height and electron density of the F_2 layer, so shows pronounced diurnal, seasonal, and year-to-year variations. Typical daytime values are of the order of 20 to 35 mc, with night values roughly half as great.

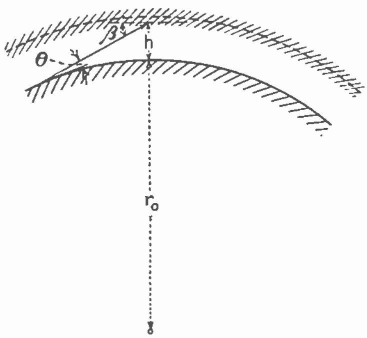


FIG. 197.—Diagram illustrating notation of Eq. (115).

In calculating the behavior of a wave in the ionosphere, it is to be noted that, with small angles of incidence, the angle between the wave path and the surface of the earth differs from the angle of incidence at the ionosphere because of the curvature to the earth. The relationship can be readily shown to be

$$\cos \beta = \frac{\cos \theta}{1 + \dfrac{h}{r_0}} \tag{115}$$

where

β = angle of incidence of wave at lower edge of the ionized layer (see Fig. 197)

θ = angle above horizontal at which wave leaves surface of earth (see Fig. 197)

h = height of the lower edge of the ionosphere above the earth

r_0 = radius of the earth.

The angles β and θ will be virtually the same except for grazing incidence. In the limit where θ is zero, it will be found that the smallest possible angle of incidence at the ionosphere is of the order of 10 to 15 deg., with the exact value depending upon the height of the layer above earth.

Effect of Earth's Magnetic Field on Path of Radio Wave.—It has already been pointed out that the earth's magnetic field causes the ionosphere to rotate the plane of polarization of the sky wave. In addition, the magnetic field causes the ionosphere to have two indices of refraction. One of these is the same as the refractive index given by Eq. (112) for the case of no magnetic field and results in a portion of the wave following a path commonly designated as the *ordinary ray*. The other refractive index is slightly lower than the refractive index as given by Eq. (112)

and results in a second portion of the wave following a path termed the *extraordinary ray*.

The extraordinary ray is either relatively weak or non-existent in the case of waves refracted by the E and F_1 layers. Waves refracted from the F_2 layer ordinarily show a strong extraordinary ray, however, and since the extraordinary ray has a higher critical frequency than the ordinary ray, the F_2 extraordinary ray determines the highest frequency that is returned to earth with normal incidence, and also the highest frequency that can be returned under any condition. The difference in the critical frequencies of the ordinary and extraordinary rays depends upon the horizontal component of the earth's magnetic field, and is about 800 kc for the F_2 layer in the latitude of Washington, D. C.

Attenuation of Waves in the Ionosphere.—A wave passing through an ionized region suffers a loss of energy because the vibrating electrons will from time to time collide with gas molecules. When this occurs, the kinetic energy that the electron has acquired from the passing wave is lost insofar as the wave is concerned. The amount of energy thus absorbed depends upon the gas pressure (*i.e.*, upon the probability of the electron colliding with a gas molecule), and upon the average energy that the electron acquired in its vibration (*i.e.*, upon the energy lost per collision). Because of the decrease of atmospheric pressure with height, most of the loss of energy that a wave suffers as a result of the ionosphere takes place in the E layer, particularly the lower edge of the E layer some distance below the point of maximum electron density. Once the wave has penetrated well into the ionized region the attenuation is small in spite of the high electron density, because there is then so little gas present for the vibrating electrons to collide with.

Theoretical analysis shows that, under conditions where most of the absorption takes place in regions well below the apex of the path followed by the wave in the ionosphere, the reduction in field strength caused by absorption is inversely proportional to the square of the frequency and directly proportional to the cosine of the angle of incidence with which the wave enters the ionosphere. This is the situation that exists for frequencies high enough to pass through the absorbing lower edge of the E layer (*i.e.*, frequencies appreciably greater than about 1500 kc), and means that for such frequencies the absorption tends to decrease as the frequency is raised, and as the wave enters the ionosphere with more nearly vertical incidence.

When the apex of the wave path is in a region where there is appreciable absorption, then the total attenuation suffered by the wave tends to be independent of the frequency and even to decrease at very low frequencies. This is because a low-frequency wave penetrates less deeply into the absorbing region, thus more or less making up for the fact that the rate of absorption is greater the lower the frequency.

The total attenuation along any wave path in the ionosphere is especially sensitive to the ionization in the lower part of the E layer. As a result, the absorption tends to be higher during the day than at night, since during the day the electrons are present at lower levels than at night. Solar disturbances that increase the electron density below the main part of the E layer also increase the attenuation very greatly.

107. Reflection of Sky Wave by the Ground.—When a sky wave that has been refracted earthward by the ionosphere strikes the ground, it

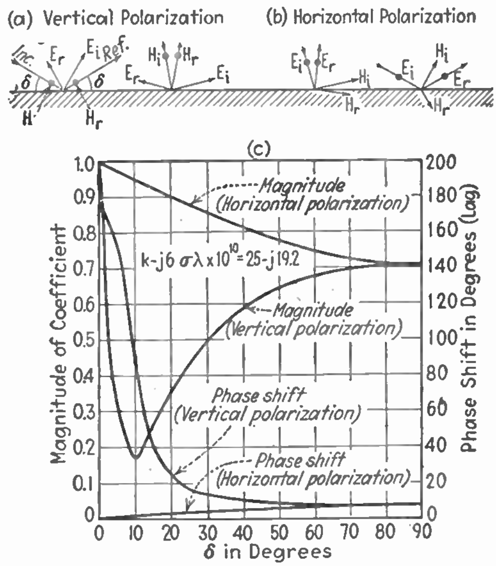


FIG. 198.—Reflection coefficient as a function of angle of incidence together with diagrams illustrating assumed positive polarities for Eq. (116) for the case of a perfect earth. (E and H denote electrostatic and magnetic flux, respectively, and subscripts i and r denote incident and reflected components.)

is reflected. The angle of reflection is equal to the angle of incidence, but because of the fact that the earth is not a perfect conductor the reflected wave is smaller in amplitude than the incident wave and is shifted in phase. The exact behavior depends upon the angle of incidence, the polarization of the wave, the frequency, and the properties of the earth. With *plane waves* the vector ratio of reflected wave to incident wave is

Component of vertical polarization (*Fig.* 198*a*)

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = \frac{\epsilon \sin \delta - \sqrt{\epsilon - \cos^2 \delta}}{\epsilon \sin \delta + \sqrt{\epsilon - \cos^2 \delta}} \tag{116a}$$

Component of horizontal polarization (Fig. 198b)

$$\frac{\text{Reflected wave}}{\text{Incident wave}} = \frac{\sqrt{\epsilon - \cos^2 \delta} - \sin \delta}{\sqrt{\epsilon - \cos^2 \delta} + \sin \delta} \tag{116b}$$

where

δ = angle of incidence as in Fig. 198*a* and *b*.
$\epsilon = \sqrt{k - j6\sigma\lambda \times 10^{10}}$
k = dielectric constant of earth (e.s.u.)
σ = earth conductivity (e.m.u.)
λ = wave length in centimeters
$j = \sqrt{-1}$.

The quantity on the right-hand side of Eqs. (116*a*) and (116*b*) represents the reflection coefficient. The absolute magnitude of this coefficient gives the ratio of reflected-wave amplitude to incident-wave amplitude, and the phase angle of the reflection coefficient is the amount by which the phase of the reflected wave differs from the phase that would exist for a perfect earth. The phase for the latter case is illustrated in Figs. 198*a* and 198*b*.

The phase and magnitude of the reflection coefficient as a function of angle of incidence are shown in Fig. 198*c* for a typical short-wave case. It will be noted that for vertical polarization there is a particular angle at which the reflection coefficient passes through a minimum, and that for more glancing incidence the phase angle of the reflected wave approaches 180° (*i.e.*, is reversed). This minimum corresponds to the Brewster angle encountered in optics, and decreases as the earth conductivity is increased.

108. Propagation of Low-frequency Radio Waves (15 to 550 Kc).— The propagation of low-frequency waves is characterized by a relatively low ground-wave attenuation and by the fact that the sky wave is refracted back to earth after only a very slight penetration into the ionosphere and with little absorption. Furthermore, the ground acts as a practically perfect reflector for low-frequency sky waves that are refracted to earth by the ionosphere. As a result of this situation, low-frequency waves reaching distant points act very much as though they were propagated in the space between two concentric reflecting spherical shells, representing the earth and the lower edge of the ionosphere. The attenuation under such conditions is that caused by spreading plus the added loss occurring at the earth's surface and at the edge of the ionosphere. At the very lowest radio frequencies these losses are small. However, since the ionosphere conditions go through diurnal and seasonal variations, the strength of very low-frequency radio waves received from a distant transmitter will fluctuate slightly, being greatest at night and in the winter. This is illustrated by the curve for 17,300 cycles in Fig. 199.

As the frequency is increased, the ground-wave attenuation becomes greater, and it is necessary to depend more upon the sky wave for communication to distant points. Furthermore, as the frequency is increased the loss at the lower edge of the ionosphere tends to be much greater in the daytime and in the summer than during the night and winter. Hence the strength of signals that have traveled a considerable distance is much more variable as the frequency increases, as shown in Fig. 199.

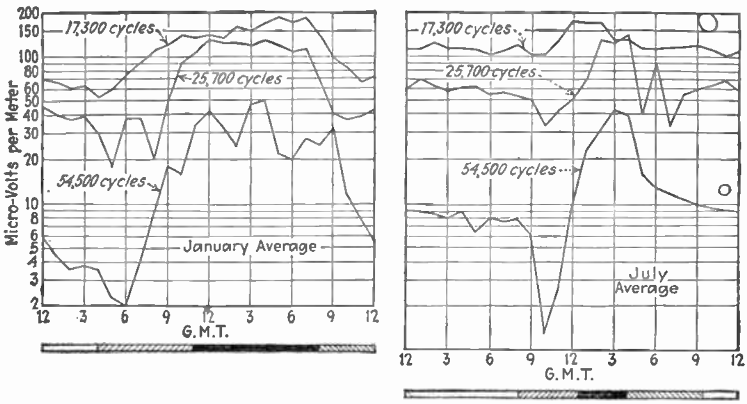


FIG. 199.—Curves showing average diurnal variation in strength of long-wave signals of different frequencies propagated across the north Atlantic during midwinter and midsummer months. (Note that signal strengths at different frequencies cannot be compared because the radiated power was not the same at all frequencies.) The solid and clear strips at the bottom of the figure indicate periods when the entire transmission path is in darkness and light, respectively, and the shaded strips indicate part of the path in darkness and part in light.

This trend with increase in frequency continues, until at 500 kc the sky wave is almost completely absorbed in the day, although it suffers relatively little absorption at night, particularly winter nights. Under such conditions transmission in the daytime takes place by means of the ground wave, and the range is limited to a moderate value. On the other hand, at night the low sky-wave absorption enables sky-wave signals to reach distances far beyond the ground-wave range.

Low-frequency radio signals commonly show a very sharp reduction in field strength during the sunset period, as indicated in Fig. 199. This is a result of transition from day to night conditions in the ionosphere.

109. Propagation of Waves of Broadcast Frequencies (Frequency Range 550 to 1500 Kc).—At broadcast frequencies the ground-wave attenuation is appreciably greater than at lower frequencies, and increases rapidly as the high-frequency end of the band is approached. Sky waves of all broadcast frequencies are almost completely attenuated during the day, but the night attenuation of the sky wave varies greatly with the conditions in the ionosphere, sometimes being moderately high, and at other times, particularly during winter nights, becoming very low.

Day Conditions.—During the day the signals received from a broadcast station are the result of ground-wave propagation. The factors determining the strength of the signals are hence the transmitter power, the frequency, and the earth conductivity, as discussed in Sec. 105. The dielectric constant of the earth is only of secondary importance, because at broadcast frequencies the earth impedance is primarily resistive.

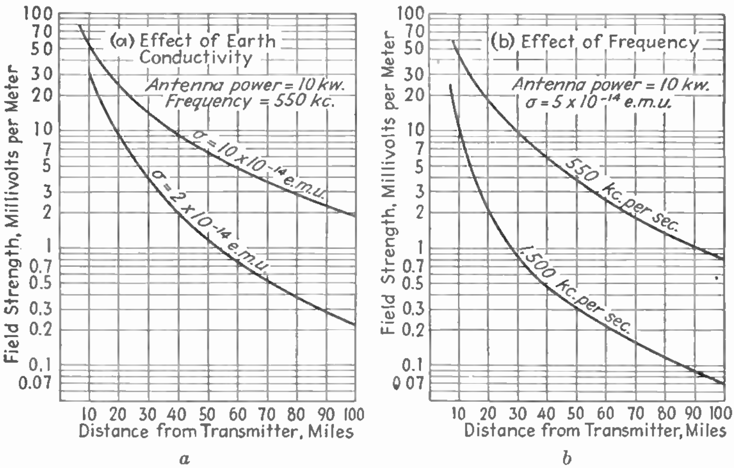


FIG. 200.—Ground-wave strength of broadcast signal as a function of distance for different earth conductivities and different frequencies. The antenna is assumed to radiate a field proportional to the cosine of the angle of elevation.

The influence of frequency and ground conductivity upon ground-wave propagation is shown in Fig. 200. This has been calculated from Eq. (111) and shows that when it is desired to have a strong ground wave at considerable distance from the transmitter the frequency should be as low as possible. It is also apparent that even with the lowest broadcast frequency it is impracticable to obtain strong ground-wave signals at appreciable distances if the ground conductivity is unusually low.

The region about the transmitter in the daytime can be conveniently divided into primary and secondary service areas. The primary service area represents the region where the ground wave has sufficient strength to override all ordinary interference, either naturally or man made. The strength of ground wave required to do this ranges from 5 to 30 mv per meter in metropolitan areas, to perhaps 0.5 mv per meter in many rural regions.

The secondary service area represents the region where the signals have fair strength but are not always sufficient to give perfect reception. Field strengths suitable for daytime secondary coverage depend upon conditions, and may be as low as 0.5 mv per meter in urban and suburban areas, and perhaps 0.05 mv per meter in rural districts.

Night Conditions.—The situation that exists at night is illustrated in detail in Fig. 201, where it is seen that there are three distinct zones. First, near the transmitter the sky wave is relatively weak compared with the ground wave, and the latter predominates. Second, at somewhat greater distances from the transmitter, the ground wave becomes attenuated, whereas the sky wave becomes stronger, until finally the

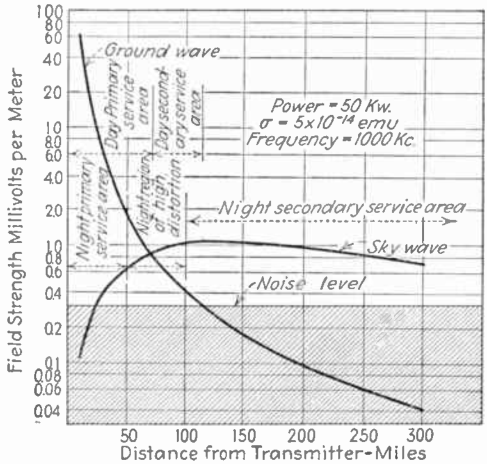


FIG. 201.—Diagram illustrating different types of coverage obtained from a high-power broadcast station during the day and night periods. The top of the shaded area represents the lowest field strength that completely overrides the noise level.

ground and sky waves become of approximately equal strength. Third, at still greater distances the sky wave tends to become still stronger and to maintain a relatively high and constant signal strength up to considerable distances. If the transmitter power is large, fair night secondary coverage is accordingly obtained over a large area that does not receive usable signals during the day.

This behavior of the sky wave is a result of two factors. In the first place, because of the height of the ionosphere, the total distance that the refracted sky wave must travel is almost the same when the sky wave returns to a point 100 km from the transmitter, as when returning to the immediate vicinity of the transmitter. In the second place, broadcast transmitting antennas radiate more energy the lower the vertical angle.

The region in Fig. 201 where the ground and sky waves have approximately equal intensity is of special importance. Here the resultant signal is the vector sum of two waves that have traveled along different paths. Since the difference in path lengths when measured in wave lengths changes rapidly with variation in frequency, different side-band frequencies will combine differently, with some adding and others subtracting. The result is frequency distortion and received signals of poor quality. Furthermore, slight changes in the ionosphere, such as continually take place, can easily vary the difference in path lengths by a half wave length and hence change an addition of the two waves to a subtraction, or vice versa. This causes the carrier, and also each individual side-band component, to fade in and out more or less independently, resulting in what is termed *selective fading*. As a result of this situation, the signals received in the region where the ground and sky waves have approximately equal intensity are of such poor quality as to have little or no entertainment value.

At distances from the transmitter so great that only the sky wave is received, there is also ordinarily a certain amount of fading and quality distortion. This is the result of two or more sky waves following different paths in traveling to the receiver. The fading and distortion under such conditions are much less severe, however, than when a ground wave is also involved.

Factors Determining Broadcast Coverage.—The factors determining the coverage that a particular broadcast transmitter renders are the transmitter power, the frequency, the earth conductivity, the directional characteristic of the transmitting antenna, and the noise level. The effect of transmitter power is shown in Fig. 202*a*. It is seen that, since an increase in power produces a corresponding increase in the strength of both sky and ground waves, more power gives better daytime coverage and better night signals to distant listeners but does not remove the high distortion region. With large power this region of high distortion is commonly within the daytime primary service area of the ground wave. Under such conditions many listeners receiving satisfactory signals in the daytime find the signals unusable at night because of selective fading, with the result that the primary service area is less at night than in the daytime.

The intensity of the interfering noise level has much the same effect upon reception as does the transmitter power. This is because it is the ratio of signal strength to noise that is important, rather than the absolute intensity of either. A decrease in the noise level therefore is equivalent to increasing the power.

The effect of the ground-wave attenuation is shown in Fig. 202*b*. High attenuation, such as produced either by low-conductivity earth or

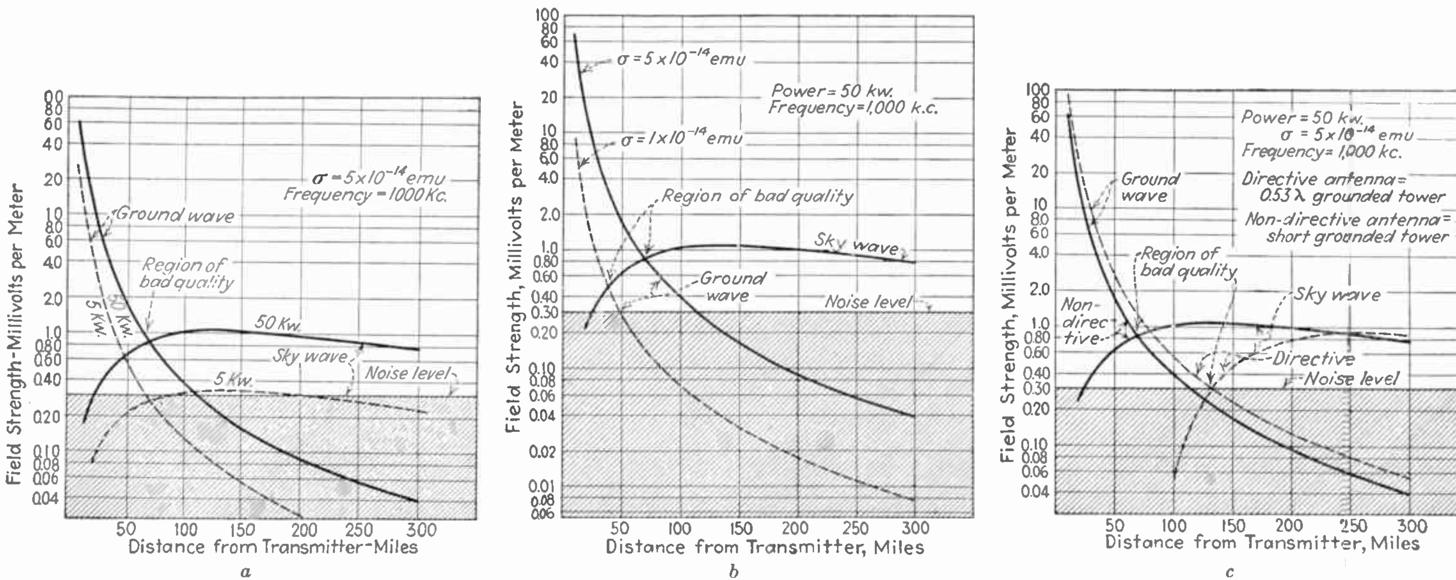


FIG. 202.—Effect of various factors on broadcast coverage. The solid line in each case is for the case illustrated in Fig. 201.

high frequency, reduces the daytime coverage very greatly and also brings the night high-distortion region closer to the transmitter.

A change in the directional characteristics of the transmitting antenna in such a manner as to cause most of the energy to be radiated at low angles above the horizon, and relatively little at high angles, produces results shown in Fig. 202*c*. The ground wave and also the distant sky wave are now stronger than before because of the increased concentration of radiation along the horizontal. At the same time, the sky wave returned to earth at moderate distances from the transmitter is reduced. The result is that increased concentration of the radiated energy along the horizontal increases the primary service area during the day, gives better night coverage to distant listeners, and moves the region of high distortion farther away from the transmitter. When the transmitter power is sufficiently great so that in the absence of increased directivity the night high-distortion region lies in the day service area, the benefits obtained by using an antenna that concentrates the radiation along the horizontal are very great.

Calculation of Broadcast Signal Strength.—The strength of the ground wave can be calculated by the methods discussed in Sec. 105. When the ground is reasonably homogeneous, the resulting accuracy is entirely satisfactory.

The strength of the sky-wave signals produced by a broadcast transmitter is determined by the transmitter power, the directional characteristics of the transmitting antenna, the height of the refracting layer, and the absorption occurring in the ionosphere. In calculating the strength of the sky wave at broadcast frequencies, it is ordinarily assumed that the refraction is equivalent to a reflection from a mirrorlike layer having a height of 100 km (the E layer), that there is no energy loss except in the ionosphere (*i.e.*, field strength inversely proportional to distance except for ionosphere loss), and that the reflected wave has a field strength that is 20 per cent of the incident wave. This last is admittedly a rather rough approximation but represents the best that can be done in view of the rather incomplete knowledge concerning the action of the ionosphere at broadcast frequencies.

Calculations of sky-wave field strength can be facilitated by the use of Fig. 203. This gives the vertical angle at which the radiation leaves the transmitter, and the ratio of sky-wave to ground-wave path lengths, as a function of the distance from the transmitter at which the sky wave returns to earth, assuming a 100-km layer height.

Detailed procedure for calculating the field strengths to be expected at broadcast frequencies is illustrated by the following example.

Example.—It is desired to determine the strength of the ground and sky waves received from a particular broadcast transmitter 100 km (60 miles) distant. The transmitter operates at a frequency of 1000 kc and delivers 10 kw to the antenna. The antenna radiates 75 per cent of the energy supplied to it, and has a directional characteristic such that the strength of the radiated field is proportional to the cosine of the angle above the earth. Propagation data obtained in the same region for waves of other frequencies indicate that the ground conductivity is 0.5×10^{-13} e.m.u., and that the dielectric constant is 15.

From Fig. 186,

$$\tan b = 16 \frac{1000}{1.8 \times 10^{18} \times 0.5 \times 10^{-13}} = 0.178$$

Hence $b = 10.1°$, and $\cos b = 0.984$. Also from Fig. 186

$$p = \pi \frac{1000}{1.8 \times 10^{18} \times 0.5 \times 10^{-13}} \frac{100{,}000}{300} 0.984 = 11.45$$

$$A = 0.053$$

The ground-wave strength is hence

$$\left.\begin{matrix}\text{Ground-wave}\\ \text{strength}\end{matrix}\right\} = \frac{300\sqrt{7.5}}{100} 0.053 = 0.44 \text{ mv per meter}$$

From Fig. 203 the sky wave leaves the transmitter at an angle of 64 deg. and the actual distance traveled by the sky wave is 220 km. Therefore the strength of the sky wave is $(300\sqrt{7.5}/220) \cos 64$ deg. $\times 0.20 = 0.33$ mv per meter.

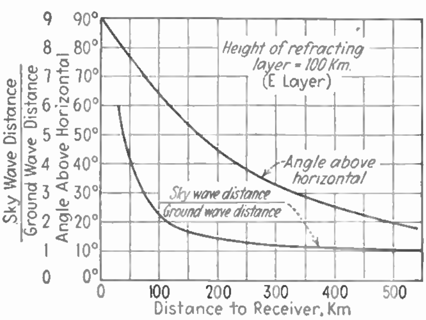


FIG. 203.—Relation between angle at which wave leaves the earth and the distance from transmitter at which return takes place and also the factor by which the ground-wave distance must be multiplied to give the distance the sky wave travels. These curves take into account the curvature of the earth and assume only slight penetration of the ionized E layer.

110. Propagation Characteristics of Short Waves (Frequency Range 1500 to 30,000 Kc).

—At frequencies above 1500 kc the ground wave attenuates so rapidly as to be of no importance except for transmission over very short distances.[1] Short-wave communication therefore ordinarily depends upon the ability of the ionosphere to refract the high-frequency sky wave back to earth at the receiving point without excessive attenuation.

Optimum Transmission Frequency.—The first requirement for short-wave communication is that the skip distance be less than the distance to the receiver. This determines the highest frequency that can possibly be used for transmission. Because of variations in the height and electron

[1] The only cases where ground-wave propagation of short waves is of importance is in connection with such things as police radio, where the distances are often very small, and in moderate-distance transmission over water, where the ground-wave attenuation is very low because of the high conductivity of the water. Frequencies in the range 1500 to 3500 kc are used for ground-wave coverage of this type.

density of the ionosphere, this maximum frequency varies with time of day, season, etc., being highest in the day and during the active part of the sunspot cycle. The maximum frequency also depends upon the distance. Thus at very short distances the frequency must not exceed the critical frequency, while at very great distances it can approach the highest frequency that is returned to earth.

The frequency that places the receiver just outside the skip distance, *i.e.*, the highest frequency that is returned to earth at the receiving point, is also ordinarily the frequency for which the ionosphere produces the least attenuation. This arises from the fact that short-wave signals penetrate far enough into the E or higher layers so that most of the absorption of energy takes place well below the apex of the path followed by the wave. Under such conditions the energy loss that accompanies passage through the layer decreases with increase in frequency, as discussed in Sec. 106.

In practical short-wave communication it is desirable to employ a frequency somewhat less than the highest possible frequency that might conceivably be employed. Although some loss of signal strength results when this is done, there is then little likelihood of having the signals skip over the receiver as a result of slight changes in the ionosphere such as occur from hour to hour and from day to day. As a consequence, the lower frequency gives on the average more dependable service and also does not require that the frequency be changed so often.

Since the electron density has diurnal, seasonal, and year-to-year variations, the optimum frequency for transmission changes accordingly. The result is that it is normally desirable to use at least two frequencies to maintain continuous communication between two points with the aid of high-frequency radio signals. One of these is a relatively low frequency for use at night, while the other is about twice as high and is employed in the daytime. The optimum frequency for long-distance communication during times of minimum sunspot-cycle activity is approximately 20 mc in the daytime, and 10 mc at night, with the values increasing somewhat during the active part of the sunspot cycle.

For short-distance sky-wave communication the frequency must be less than the critical frequency, and cannot exceed the critical value appreciably even for distances up to 400 to 500 miles. Examination of Fig. 196 shows that the optimum day and night frequencies for short-distance communication are of the order of 6000 and 3000 kc, respectively, for quiet parts of the sunspot cycle, and become greater when the sunspots are more numerous.

Wave Paths and Desirable Antenna Directivity.—Under practical conditions where the transmitted frequency is somewhat less than the maximum that could be employed, there are usually several possible paths that the signal may follow when traveling to the receiver. This is illustrated in Fig. 204*a*. Here radiation leaving the transmitting antenna at a very small angle above the horizontal follows a path such as *a* and reaches the receiving location in a single hop. Energy radiated at higher vertical angles follows paths such as *b* and *c* in Fig. 204*a*, which involve more than one round trip between earth and ionosphere.

The particular path having the lowest total attenuation, and hence the path followed by most of the energy reaching the receiver, is a compromise between conflicting factors. Thus the wave must traverse the absorbing region at the lower edge of the ionosphere twice for each step, and if there are very many steps the wave is hence highly attenuated. On the other hand, a single-step path such as *a* passes through the absorbing region the minimum possible number of times, but does so with a very

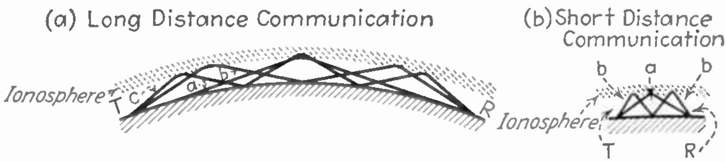


FIG. 204.—Typical wave paths by which energy might travel from transmitter to receiver. For the sake of simplicity only a single ionosphere layer is shown as being effective in determining the paths.

glancing angle, which means a comparatively long time spent in the absorbing region. Experiments with short-wave signals under conditions corresponding to those existing in transoceanic communication indicate that for long-distance communication the paths having the least attenuation leave the transmitting antenna at an angle of the order of 10 to 25 deg. above the horizontal. In the case of transatlantic communication this corresponds to two to three steps under the usual conditions (paths *b* and *c* of Fig. 204*a*).

At short and moderate distances, a single-step path such as *a* in Fig. 204*b* corresponds to energy radiated at an appreciable angle above the horizontal. The attenuation along such a path is lower than for paths having two or more steps as *b* in Fig. 204*b*, so that short-distance communication tends to take place along a single-step path, utilizing energy radiated at the appropriate vertical angle.

The particular ionosphere layer that is effective in returning to earth waves leaving the transmitter at different vertical angles depends upon the frequency, the vertical angle, and the heights and electron densities of the various layers. Under some practical conditions most of the energy reaching the receiver will have been refracted by the *E* layer, while at other times the path having the least attenuation will penetrate through the *E* layer and involve a reflection by a higher layer. In still other circumstances it is possible for two waves leaving the transmitter with different vertical angles to be refracted by different layers and reach the receiver with the same number of steps.[1]

In long-distance short-wave communication it is desirable to use an antenna system that will concentrate the radiated energy at an angle of about 10 to 20 deg. above the horizontal. This sends the energy in directions where it is most likely to reach the receiver. At short and moderate distances the optimum angle is usually somewhat higher and increases as the distance is reduced. When short-wave communication is to be carried on between two fixed points, it is also desirable to concentrate the radiation in the horizontal plane as well as in the vertical, since all energy radiated in horizontal directions other than toward the receiver is wasted.

Character of Received Signals.—Short-wave signals received from a distant transmitter ordinarily represent the vector sum of several sky waves that have traveled over paths of different lengths. This is because of the splitting of the wave in the ionosphere as a result of the earth's magnetic field, and because of paths having different number of steps or paths refracted from different layers. The small variations that are continually taking place in the ionosphere accordingly cause the resultant signal to fade more or less continuously. Since this fading depends upon the frequency, the different side-band frequencies contained in a modulated wave may fade independently, giving selective fading with consequent distortion, as discussed in Sec. 109.

Short-wave signals observed at the receiver have both vertically and horizontally polarized components which bear no apparent relation to the

[1] In investigating the paths followed by waves leaving the transmitter at different vertical angles it is sometimes convenient to make use of the relation

$$\left.\begin{array}{l}\text{Distance in kilometers from}\\ \text{transmitter to point where}\\ \text{wave returns to earth}\end{array}\right\} = 222(\beta - \theta) \tag{117a}$$

where θ is the angle above the horizontal at which the wave leaves the antenna and β is the angle of incidence at the ionosphere as given by Eq. (115). Both θ and β are expressed in degrees. In the special case where the distance involved is so small that the curvature of the earth can be neglected, Eq. (117*a*) reduces to

$$\left.\begin{array}{l}\text{Distance in kilometers from}\\ \text{transmitter to point where}\\ \text{wave returns to earth}\end{array}\right\} = \frac{2h}{\tan\theta} \tag{117b}$$

where h is the height of the ionosphere in kilometers.

Equations (117) assume that the ionosphere produces a mirrorlike reflection at a height corresponding to the height of the maximum electron density of the layer involved. As a result of the sharply defined character of the layers this gives reasonably accurate results except when the frequency is so high as to be barely returned by the layer.

polarization of the transmitting antenna, which fade independently of each other, and which ordinarily are not in phase. This situation is the result of the earth's magnetic field, which causes a wave refracted by the ionosphere to have its plane of polarization rotated, and which also splits the wave into several components that travel along different paths with different velocities.

111. Propagation of Ultra-high-frequency Waves.—Frequencies greater than 25 to 40 mc in daytime, and over 10 to 20 mc at night, pass through the ionosphere without being refracted to earth even when the angle of incidence has the lowest value that can be realized in view of the curvature of the earth. At the same time, the ground wave is very rapidly attenuated at these frequencies. The use of such ultra-high radio frequencies for communication hence requires that the waves pass through the space above the earth from an elevated transmitting antenna to an elevated receiving antenna.

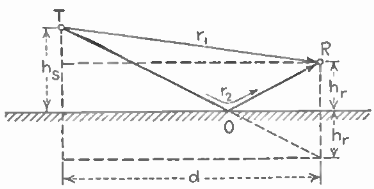


FIG. 205.—Diagram showing the direct and indirect paths by which energy may travel from transmitter to receiver. For the sake of clarity the antenna heights have been greatly exaggerated in comparison with the distance.

When the distance between transmitter and receiver is small enough for the earth to be considered flat, the resulting situation is as shown in Fig. 205. Here waves may reach the receiver either by a direct path between transmitting and receiving antennas, or by a route involving reflection from the surface of the ground. These two waves tend to cancel each other at the receiver provided the distance is much greater than the antenna heights, since then the earth reflection is with almost grazing incidence and consequently takes place with reversal in phase [see Eq. (116) and Fig. 198]. The cancellation is not complete, however, since the indirect path is slightly longer than the direct path. When the distance from transmitter to receiver is large compared with the height of transmitting and receiving antennas, the field strength that results is[1]

[1] This can be shown as follows: Referring to Fig. 205, it is seen from the dotted construction that $r_1^2 = (h_s - h_r)^2 + d^2$, and $r_2^2 = (h_s + h_r)^2 + d^2$. For $d \gg (h_s + h_r)$, one can then write

$$r_1 = d + \frac{(h_s - h_r)^2}{2d}, \qquad r_2 = d + \frac{(h_s + h_r)^2}{2d}$$

Consequently the difference in path lengths is

$$r_2 - r_1 = \frac{(h_s + h_r)^2 - (h_s - h_r)^2}{2d} = \frac{2h_s h_r}{d}$$

The corresponding phase difference caused by the path difference is

$$\text{Field at receiver} = \frac{4\pi h_s h_r E_o}{\lambda d^2} \tag{118}$$

where E_o is the strength of the direct wave at a distance $d = 1$ from the transmitter, and dimensions and distances are measured in the same units.

If the earth were removed, the field strength at the receiver would be that due to the direct wave alone, which is E_o/d, and so would be inversely proportional to distance.[1] The presence of the earth hence reduces the field strength below this free-space value by the factor $4\pi h_s h_r/\lambda d$, and makes the field strength inversely proportional to the square of the distance.

Effect of Earth Curvature.—When the distance between transmitter and receiver is considerable, it is necessary to take into account the curvature of the earth. This curvature reduces the received field strength below the values that would be obtained for a flat earth, but experience shows that the correction required is so small as to be negligible as long as a straight-line path exists between transmitter and receiver. If a straight-line path does not exist, it is still possible, however, for energy to reach the receiver as a result of refraction by the earth's atmosphere.

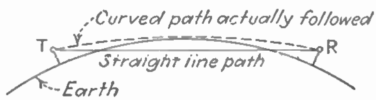


FIG. 206.—Diagram illustrating how the refraction in the earth's atmosphere permits direct-ray transmission at ultra-high frequencies even when the straight-line path is intercepted by the earth's curvature.

The refraction by the earth's atmosphere at ultra-high frequencies comes about because the variation of atmospheric pressure, temperature, and moisture content with height causes the refractive index of the atmosphere to decrease with elevation. This tends to bend the waves back toward the earth in much the same manner as does the ionosphere. The amount of curvature that results varies with the atmospheric conditions, but on the average it is equivalent to assuming that the earth's radius is increased by 25 to 35 per cent. As a consequence of this refraction, it is possible to obtain direct-ray propagation of energy between transmitter and receiver under conditions where a straight-line path falls slightly below the earth's surface, as in Fig. 206.

$$2\pi \cdot \frac{2h_s h_r}{\lambda d} = \frac{4\pi h_s h_r}{\lambda d} \text{ radians.}$$

It is because of this phase angle that the direct and indirect rays fail to cancel, so that the resultant of the two waves is 2 sin $(2\pi h_s h_r/\lambda d)$ times the amplitude E_o/d of one of the waves. When the angle is so small that the sine of the angle equals the angle in radians, the result is Eq. (118).

[1] When the transmitting antenna radiates a field proportional to the cosine of the angle of elevation, then $E_o = 300\sqrt{P}$ mv per meter, where E_o is in the direction of maximum radiation, P is the transmitter power in kilowatts, and d is in kilometers.

The range of the direct rays depends upon the heights h_s and h_r of the transmitting and receiving antennas, respectively, and the effective radius of the earth, according to the formula

$$\left.\begin{array}{l}\text{Maximum possible distance for}\\\text{direct-ray transmission}\end{array}\right\} = 1.225k(\sqrt{h_s} + \sqrt{h_r}) \qquad (119)$$

where the antenna heights are in feet, the distance is in miles, and k represents the ratio of effective to actual earth radius and takes into account the refraction in the earth's atmosphere. The result of Eq. (119) is illustrated in Fig. 207 for $k = 1.33$, which shows that direct-ray transmission over considerable distances is possible only when the antennas are located at high elevations, such as upon mountain peaks. Even then the range is comparatively modest compared with that obtainable with waves of lower frequency where refraction from the ionosphere can be utilized.

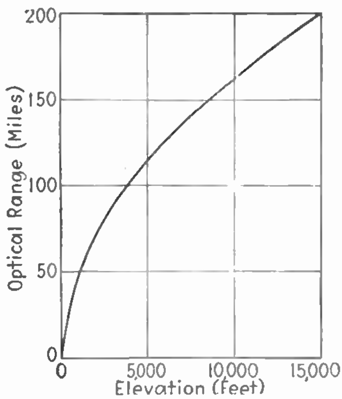


FIG. 207.—Maximum possible optical range between an elevated point and the surface of the earth, assuming that atmosphere refraction increases the effective value of the earth's radius by a factor of 1.33. The maximum possible optical range between two elevated points is the sum of the ranges as obtained in the above diagram for the two heights involved.

Miscellaneous Considerations.—The transmission paths at ultra-high frequency are extremely stable as long as a straight-line path is possible between transmitting and receiving points. However, when the transmission distance is so great that it is necessary to depend upon the refraction in the earth's atmosphere to obtain direct-ray transmission to the receiver, fading frequently occurs as a result of variations in the earth's atmosphere changing the amount of refraction that the wave suffers. This fading becomes very severe when the transmission distance approaches or exceeds the maximum possible distance of direct-ray transmission.

112. Miscellaneous. *Relation of Solar Activity to the Propagation of Radio Waves.*—The fact that the propagation of all except the very shortest radio waves depends to a marked extent upon the ionosphere results in radio-wave propagation being influenced by such meteorological factors as sunspot cycles, magnetic storms, etc. The effect of the 11-year sunspot cycle upon the electron density in the various ionosphere layers has already been discussed in Sec. 106.

Magnetic storms have a very pronounced effect upon radio signals.[1] Thus a severe magnetic storm ordinarily makes the short-wave circuits

[1] The term "magnetic storm" refers to an abnormal condition characterized by

across the north Atlantic completely inoperative for a period of several days, and causes subnormal field strengths for a week or more. The adverse effect is much less, however, when the entire transmission path is nowhere near either pole.

During a magnetic storm the electron density in the F_2 layer is decreased, and the layer height becomes greater. The result is that the optimum frequency for transmission is increased. At the same time there is a marked increase in absorption, presumably as a result of increased ionization in the lower edge of the E layer.

At the very lowest radio frequencies, such as 15 to 100 kc, magnetic storms increase the strength of the day signals slightly, decrease the strength of the night signals to approximately the day level, and cause the sunset drop in signal intensity to disappear. Hence at very low frequencies magnetic storms are, if anything, helpful to long-distance communication, rather than harmful as in the case of short waves.

High-frequency sky-wave signals sometimes experience sudden "fade-outs" over the entire illuminated half of the earth. These last from a few minutes to one hour and are caused by a sudden increase in the absorption resulting from increased ionization in the lower part of the E layer. The source of this ionization has been traced to bursts of radiation from a sudden solar eruption.

Use of Radio Waves to Investigate the Ionosphere.—Since the way in which radio signals propagate is dependent on the conditions in the ionosphere, it is possible, by working backward from observed propagation characteristics of radio signals, to obtain information regarding the probable nature of the ionized layer. Indirectly this also gives data on the composition of the upper atmosphere.

The usual method of making such investigations consists in transmitting a short-wave train lasting about 10^{-4} sec. and taking an oscillogram record of the signal as received at a point within range of the ground wave. Since the wave that reaches the receiver after refraction by the ionosphere must travel an appreciably longer path to reach the receiver than does the ground wave, there will be a time interval between the pulses arriving over the two routes. The length of this time interval is a measure of the difference in path lengths, and can be used to calculate the height of the layer. The received records ordinarily show a number of returned pulses with different time delays, as indicated in Fig. 208. These may be the result of double refraction produced as a consequence of the earth's magnetic field, or they may be due to multiple reflections such as two round trips between the earth and the ionosphere.

---

rapid and excessive fluctuations of the earth's magnetic field. Magnetic storms are frequently associated with sunspot activity.

The layer heights, obtained from pulse signals by making calculations on the basis of the velocity of light, are *virtual* or apparent heights. These are always greater than the actual height reached by the wave because the wave travels with a velocity less than the velocity of light when in an ionized region. Because the layers of the ionosphere are fairly sharply defined, the difference between the virtual and actual height is small, however, unless the wave passes through a region where the electron density is such that the wave is near a critical frequency. Under these conditions the slowing down of the wave by the ionization is very great, and the virtual height becomes correspondingly high. The term *layer height* as ordinarily used without further qualification is customarily taken to mean the virtual height for a frequency differing sufficiently from the critical frequency so that the virtual and actual heights for the wave path are substantially the same.

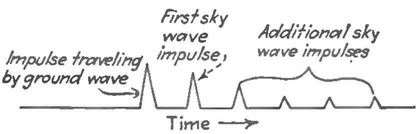


FIG. 208.—Typical oscillogram of received signal when transmitter within ground-wave range sends out a short impulse.

The most effective means of utilizing the pulse method in investigating the ionosphere is to vary the transmitter frequency, either continuously or in steps, and to obtain the variation of apparent or virtual height as a function of frequency. An example of a simple record of this type is

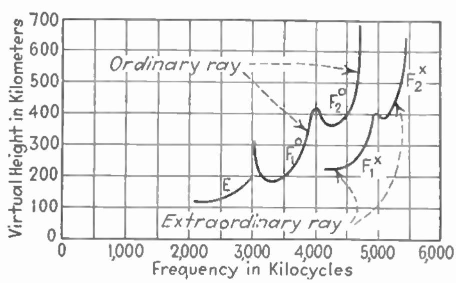


FIG. 209.—Typical curves of virtual height as a function of frequency as obtained in ionosphere investigations.

illustrated in Fig. 209. This shows that at a frequency of about 2000 kc reflections were returned from a layer having a height about 110 km (the E layer). As the frequency was increased the apparent layer height first increased gradually, and then suddenly jumped to about 300 km at 3000 kc. This jump takes place at the critical frequency for the E layer. The critical frequency for the ordinary ray in the F_1 layer and for the ordinary and extraordinary rays in the F_2 layer in Fig. 209 are 4000, 4700, and 5500 kc, respectively, and the heights of the F_1 and F_2 layers are approximately 190 and 365 km, respectively.

Static.—Radio waves generated by natural causes are referred to as static, and produce the familiar clicks, rumblings, crashes, etc., sometimes heard in all radio receivers. Static normally has its origin in thunderstorms and similar natural electrical disturbances, and is in the form of impulses, the energy of which is distributed throughout the range of useful radio frequencies. Since static is fundamentally a radio signal containing frequency components distributed over a wide range of frequencies, the static within any frequency range is propagated over the earth in the same way as ordinary radio signals of the same frequency. Thus static impulses travel great distances under favorable conditions, arrive at a receiving point from a definite direction, and possess diurnal and seasonal variations in intensity as a result of corresponding variations in wave propagation.

At low radio frequencies the static intensity is high because most of the energy of a static impulse is concentrated on the lower radio frequencies, and because radio waves propagate great distances with small attenuation at low frequencies. The intensity of long-wave static becomes greater as the frequency is reduced and in northern latitudes is greater at night and in summer than in the daytime and winter, respectively.

At moderate frequencies, such as those in the broadcast range, a large fraction of the static observed during the day is of local origin, because of the poor propagation of such frequencies during daylight hours. At night, however, the lower attenuation causes static impulses of distant origin to be heard, with the result that the noise level is ordinarily greater at night than in the daytime.

The static intensity commonly observed at short waves, *i.e.*, frequencies from 1500 to 30,000 kc, is much less than at lower frequencies, and during a good part of the time is of the same order of magnitude as the noise level of a typical radio receiver. A large fraction of short-wave static is ordinarily produced by rather distant sources. As a result, the static received on a particular frequency tends to be greatest when the conditions are favorable for long-distance reception on that frequency, and vice versa.

Very little static is found on frequencies too high to be refracted from the ionosphere. At these frequencies the range of the signals is so limited that all static must be of relatively local origin. Most of the noise observed at ultra-high frequencies is in fact produced by electrical apparatus, ignition systems, etc.

The interference that static produces in a radio receiver can be minimized by designing the receiver to respond to the narrowest frequency band that will accommodate the desired signal. This is because the continuous distribution of static energy over the frequency spectrum makes the static reaching the receiver output proportional to the width of the *a*. Frequency 550 kc, with (1) sea-water earth, (2) $\sigma = 10^{-13}$ e.m.u., $\epsilon = 20$, and (3) $\sigma = 0.2 \times 10^{-13}$ e.m.u., and $\epsilon = 5$.

b. Same earth conditions as *a* but for a frequency of 1500 kc.

Tabulate the results.

5. A series of field strength measurements about a broadcast station operating at 900 kc shows that at a distance of 20 miles the strength of the ground wave is 0.25 of the value calculated on the basis of zero ground losses. Deduce the earth conductivity for this case, assuming a reasonable value for the dielectric constant.

6. Describe how conditions in the ionosphere south of the equator could be expected to differ from those for the northern hemisphere as illustrated in Figs. 189 and 190.

7. Positive and negative ions in the ionosphere vibrate under the influence of a radio wave in much the same manner as do electrons. However, the amount of refraction produced by ions is relatively much less than that caused by electrons in proportion to density. Explain.

8. The earth's magnetic field has no effect upon a wave passing through the ionosphere provided the wave is so oriented that the electrostatic lines of force of the wave are parallel with the magnetic flux lines of the earth. Explain.

9. Calculate and plot the refractive index of the F_1 layer as a function of frequency up to the frequency for which $\mu = 0$, assuming that the electron density is 2.2×10^5 electrons per cu. cm (corresponding to summer noon at Washington, D. C. as in Fig. 190).

10. From the data given in Fig. 190 calculate and plot as a function of time of day the highest frequency that on the average could be used for short-distance communication (corresponding to a sky wave striking the ionosphere with nearly vertical incidence) at Washington, D. C., in June, 1933, during the hours 8 A.M. to 6 P.M.

11. Calculate and plot the skip distance as a function of frequency up to distances of 2000 km, for ionosphere conditions corresponding to those obtained at noon in June, 1933, at Washington, D. C. (see Figs. 189 and 190). Indicate on the skip-distance curve the layer effective in determining the skip distance for each portion.

12. Derive Eq. (114*a*).

13. What is the skip distance that an amateur transmitter operating at 7200 kc can expect when ionosphere conditions correspond to those at Washington, D. C., around noon in June, 1933 (Figs. 189 and 190)?

14. Calculate and plot skip distance as a function of time of day from 7 A.M. to 9 P.M. for a frequency of 14,000 kc when the ionosphere conditions correspond to those existing at Washington, D. C., in June, 1933 (see Figs. 189 and 190).

15. Calculate the highest frequency that will be returned to earth under any conditions when the ionosphere conditions correspond to those at Washington in June, 1933 (see Figs. 189 and 190), when the time of day is (*a*) noon, (*b*) midnight, and (*c*) 3 A.M.

16. A short-wave transmitter operates at a frequency of 5000 kc, and the ionosphere conditions correspond to noon in June, 1933, at Washington, D. C. (see Fig. 190). For what angles of incidence at the ionosphere will the waves entering the ionosphere be returned to earth by: (*a*) the *E* layer, (*b*) the F_1 layer, and (*c*) the F_2 layer?

17. Determine the transmitter power (assuming that the radiated field is proportional to the cosine of the angle of elevation) required to produce a ground-wave field strength of 1 mv per meter at a distance of 200 km when the soil conductivity is 10^{-13} e.m.u. and the dielectric constant is 15, for frequencies of 550 and 1500 kc. From the results discuss the economic feasibility of obtaining large ground-wave coverage at the different broadcast frequencies.

18. Determine the transmitter power (assuming that the radiated field is proportional to the cosine of the angle of elevation) required to produce a ground-wave field strength of 1 mv per meter at 200 km when the frequency is 550 kc and the earth conductivity is 10^{-11} (sea water), 10^{-13} (good soil), and 10^{-14} (poor earth).

19. A proposed broadcast transmitter is to deliver 50 kw of power to an antenna that has a directional characteristic as shown in Fig. *A*. The field strength along the ground at a distance of 1 mile (which is so close that ground losses can be neglected) is expected to be 1750 mv per meter. The transmitter frequency is to be 800 kc.

a. Plot a curve of ground-wave field strength as a function of distance up to 500 km for an earth conductivity of 1×10^{-13}.

b. Calculate sky-wave field strength as a function of distance up to 500 km by making reasonable assumptions as to the ionosphere action, and plot the results upon the same curve sheet as used for (*a*).

c. Discuss the resulting coverage for both day and night, including consideration of such factors as day and night primary and secondary service areas, location of high-distortion area, etc.

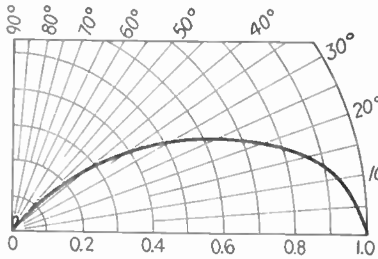


FIG. *A*.

20. It is found experimentally that the location of the high-distortion region about a broadcast transmitter changes somewhat during the course of a single night and also varies from day to day and season to season. Explain.

21. The optimum directivity for a broadcast antenna places the nighttime high-distortion region at the outer edge of the secondary service area rendered by the ground wave. Explain why this optimum directivity depends upon (*a*) transmitter power, (*b*) transmitter frequency, and (*c*) earth conductivity.

22. Calculate and plot the frequency giving the strongest possible received signals, as a function of distance, for ionosphere conditions corresponding to noon at Washington, D. C., in June, 1933 (see Figs. 189 and 190).

23. It is desired to carry on continuous communication between two points only a moderate distance apart and to use not more than two frequencies to cover the 24-hr. period. If the ionosphere conditions correspond to those in Washington, D. C., in December, 1933 (see Figs. 189 and 190), suggest two suitable values of frequency, making the frequencies 25 per cent lower than the curves of averages would call for to take care of variations about the average, and indicate the hours during which each frequency would be used.

24. Twenty-four-hour communication is to be carried on between points that are a great distance apart, using only two frequencies. It is desired that all waves leaving the earth at vertical angles up to 25 deg. above the horizontal be returned to earth. Suggest two frequencies that would be suitable when the ionosphere conditions correspond to those in Washington, D. C., December, 1933 (see Figs. 189 and 190), making the frequencies 25 per cent lower than called for by the curves to take care of variations in ionosphere conditions about the average, and indicate the hours during the day when each frequency would be used.

25. In short-distance communication using short waves, the night frequency will be returned to earth close to the transmitter during the day as well as at night. Explain why it is customary to use a different frequency for day conditions, instead of using the night frequency at all times.

26. Show that the data given in Figs. 189 and 190 are consistent with the fact that the optimum frequency for long-distance short-wave communication is much lower at night than in the day but only slightly different in winter than in summer.

27. Short-wave communication is to be carried on at noon between two points 200 km (125 miles) apart. Determine the vertical angle at which the transmitted energy should be concentrated, assuming that the ionosphere conditions correspond to those for December in Figs. 189 and 190 and that the transmitter frequency is (*a*) 2750 kc and (*b*) 4000 kc.

28. Plot curves, showing distance from transmitter at which waves return to earth, as a function of the angle above horizontal at which radiation leaves the transmitting antenna, when the ionosphere conditions correspond to those at noon in Washington, D. C., in June, 1933. Plot separate curves for each layer.

29. The antenna for a proposed television transmitter is to be located at the top of a building 750 ft. high. (*a*) Over what distance is direct-ray coverage possible, assuming that the average receiving antenna will have a height of 30 ft.? (*b*) Over what distance is a straight-line path possible, assuming the same receiving antenna?

30. An ultra-high-frequency transmitter operating at a wave length of 5 meters with a power of 100 watts is to be used for communication between two points 30 miles apart. The height of the transmitter antenna is 400 ft. (*a*) Determine the minimum height of the receiving antenna for which direct-ray reception is possible. (*b*) Estimate the field strength received by a receiving antenna 400 ft. high, assuming the transmitting antenna radiates a field proportional to the cosine of the angle of elevation.

31. What transmitter power is required to deliver a 0.050 mv per meter signal at 5 meters when the transmitting and receiving antennas are both 50 ft. high, the distance is 10 miles, and the radiated field is proportional to the cosine of the angle of elevation?

32. When pulse signals are being returned from the F_1 layer, the virtual height is greater when the frequency is just barely high enough to permit penetration of the E layer than when a higher frequency is used, even though the waves actually penetrate farther up into the F_1 layer the higher the frequency. Explain.

CHAPTER XIV

ANTENNAS

114. Fundamental Laws of Radiation.—An understanding of the mechanism by which energy is radiated from a circuit, and the derivation of equations for expressing this radiation quantitatively, involve conceptions that are unfamiliar to engineers. The mathematical formulas that express the results of such an analysis are, however, quite simple and understandable. Thus the strength of the field radiated from an elementary length of wire δl carrying a current I (see Fig. 210) is given by the formula

$$\begin{aligned} \mathcal{E} &= \frac{60\pi}{d\lambda}(\delta l)I \cos \omega\left(t - \frac{d}{c}\right) \cos \theta \\ &= \frac{60\pi f}{dc}(\delta l)I \cos \omega\left(t - \frac{d}{c}\right) \cos \theta \quad (120) \end{aligned}$$

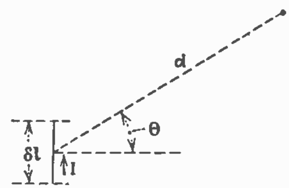


FIG. 210.—Elementary doublet consisting of a length of wire δl carrying a current I.

where

$\mathcal{E}$ = the strength of the wave in volts per meter
δl = the length of wire from which the radiation takes place, measured in the same units as λ
$I \cos (\omega t + 90°)$ = current flowing in the wire in amperes
d = distance from P to the antenna in meters (assumed large compared with λ)
θ = angle of elevation of point at which field is desired with respect to a plane perpendicular to the conductor δl
f = frequency of current in cycles per second
$\omega = 2\pi f$
t = time in seconds
c = velocity of light = 3×10^8 meters per second
λ = wave length corresponding to frequency f.

Since the radiated field is proportional to $\cos \theta$, the field is maximum in a plane perpendicular to the axis of the elementary antenna and zero in the direction of the axis. The phase of the radiated field depends upon the phase of the antenna current I and upon the distance. The wave front lies in a plane perpendicular to a line drawn toward the antenna, and the polarization of the wave is the same as that of the antenna, *i.e.*, the antenna and an electrostatic flux line of the wave lie in the same plane. The magnetic field associated with the wave is perpendicular to the electrostatic flux and is related to it by the equation

$$\mathcal{E} = 300H \tag{121}$$

where H is in lines per square centimeter, and $\mathcal{E}$ is in volts per centimeter.

The field radiated from a complete antenna is found by adding up the separate fields produced by the elementary lengths of the radiator, taking into account phase relations and polarization in making this addition. The process of determining the field radiated by an antenna system is accordingly a matter of mathematical or graphical integration, based upon the configuration of the antenna conductors and the distribution of current flowing in them.

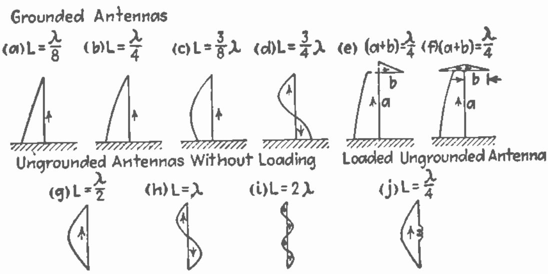


FIG. 211.—Current distribution in typical antennas. In each case the current has a sinusoidal distribution and is zero at the open ends.

Current Distribution in an Antenna.—An antenna is a circuit with distributed constants and so has a current distribution of the type discussed in Sec. 19. The current is zero at the open ends of the antenna and also approaches zero at all points that are an exact multiple of a half wave length distant from an open end. At the same time the current is maximum at points that are odd quarter wave lengths distant from the open ends. The distribution of current with length is substantially sinusoidal except near the current minima. Because of this, and the fact that the currents on adjacent sides of a minima are almost 180° out of phase, it is customary to show the current distribution of a radio antenna by means of sine waves, as illustrated in Figs. 29 and 211. While it has already been pointed out in Sec. 19 that this is not strictly correct, the resulting error involved when calculating radiated fields is ordinarily negligible. In this sinusoidal distribution of current, one cycle occupies a distance along the wire corresponding to one wave length of a radio wave in space. Examples of current distribution in a number of typical antennas are shown in Fig. 211. The current in each case follows a sinusoidal law and is zero at the open ends. When the lower end of the antenna is grounded, as in a to f, or when the length of an ungrounded antenna is not an exact multiple of a half wave length, as at j, the current distribution is made up of sections of sine waves, as shown.

Example of Antenna Calculation.—The calculation of total field radiated from an antenna is illustrated by the following example.

Example.—It is desired to calculate the radiated field produced by an antenna a half wave length long carrying a current I_o at the center of the antenna and located in space remote from ground.

Taking the mid-point of the antenna as the reference ($x = 0$ at mid-point), then the current distribution is given by the equation

$$i = I_o\left(\cos\frac{2\pi x}{\lambda}\right)[\cos(\omega t + 90°)] \tag{122}$$

Referring to Fig. 212a, the field radiated from an element dx of the antenna in the direction P has from Eq. (120) a magnitude $d\mathcal{E}$ given by

$$d\mathcal{E} = \frac{60\pi}{d\lambda}(dx)I_o\cos\frac{2\pi x}{\lambda}\cos\theta \tag{123a}$$

Considering the field radiated from the mid-point ($x = 0$) as the reference phase, then the radiation from an element dx at a distance x from the mid-point reaches

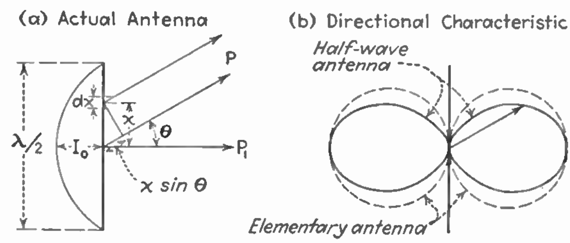


FIG. 212.—Half-wave antenna in space, and directional characteristic, with directional characteristic of elementary antenna shown for comparison.

a distant point P sooner than does the radiation from the mid-point, because the distance is less by $x \sin\theta$, which corresponds to $(2\pi x/\lambda)\sin\theta$ radians advance in phase. The in-phase and quadrature components of $d\mathcal{E}$ are hence

$$\left.\begin{matrix}\text{In-phase com-}\\ \text{ponent of } d\mathcal{E}\end{matrix}\right\} = \frac{60\pi}{d\lambda}I_o(dx)\left(\cos\frac{2\pi x}{\lambda}\right)\cos\theta\cos\left(\frac{2\pi x}{\lambda}\sin\theta\right) \tag{123b}$$

$$\left.\begin{matrix}\text{Quadrature com-}\\ \text{ponent of } d\mathcal{E}\end{matrix}\right\} = \frac{60\pi}{d\lambda}(dx)I_o\left(\cos\frac{2\pi x}{\lambda}\right)\cos\theta\sin\left(\frac{2\pi x}{\lambda}\sin\theta\right) \tag{123c}$$

The total field resulting is obtained by integrating Eqs. (123b) and (123c) from $x = -\frac{\lambda}{4}$ to $x = \frac{\lambda}{4}$. Accordingly,

$$\left.\begin{array}{l}\text{In-phase com-}\\ \text{ponent of } \mathcal{E}\end{array}\right\} = \frac{60\pi}{d\lambda} I_o \cos\theta \int_{-\frac{\lambda}{4}}^{\frac{\lambda}{4}} \cos\left(\frac{2\pi x}{\lambda}\right) \cos\left(\frac{2\pi x}{\lambda}\sin\theta\right) dx$$

$$= \frac{60\pi}{d\lambda} I_o \cos\theta \left[\frac{\sin\left(\frac{2\pi x}{\lambda}(1-\sin\theta)\right)}{\frac{4\pi}{\lambda}(1-\sin\theta)} + \frac{\sin\left(\frac{2\pi x}{\lambda}(1+\sin\theta)\right)}{\frac{4\pi}{\lambda}(1+\sin\theta)} \right]_{-\frac{\lambda}{4}}^{\frac{\lambda}{4}}$$

$$= \frac{15}{d} I_o \cos\theta \left[\frac{2\sin\left(\frac{\pi}{2}(1-\sin\theta)\right)}{(1-\sin\theta)} + \frac{2\sin\left(\frac{\pi}{2}(1+\sin\theta)\right)}{(1+\sin\theta)} \right]$$

$$= \frac{30}{d} I_o \cos\theta \left[\frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{1-\sin\theta} + \frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{1+\sin\theta} \right]$$

$$= \frac{60}{d} I_o \frac{\cos\left(\frac{\pi}{2}\sin\theta\right)}{\cos\theta} \qquad (124)$$

The corresponding integration of Eq. (123c) shows the quadrature component to be zero, so that Eq. (124) gives the total field.

Equation (124) is shown graphically in Fig. 112*b*, in which the length of the radius vector in any particular direction is proportional to the strength of the field radiated in that direction.

Examination of Fig. 212*b* shows that whereas the field radiated from an individual elementary length of the antenna is proportional to cos θ, as shown by the dotted lines, the total field from the half-wave antenna is more sharply concentrated at right angles to the axis of the wire. The reason for this is that, although radiation from all parts of the antenna adds in phase to a distant observer in the direction P_1, an observer in another direction such as P is closer to one end of the half-wave antenna than to the other end, so that the radiation from the different parts of the half-wave antenna will then more or less cancel.

115. Fundamental Properties of Receiving Antennas and Reciprocal Relations Existing between Transmitting and Receiving Properties.—A receiving antenna abstracts energy from a passing radio wave as a result of the voltages that the magnetic flux of the wave induces in the antenna. These induced voltages are distributed along the entire length of the antenna and have a value which per meter of antenna length is $\mathcal{E}\cos\psi\cos\theta$, where $\mathcal{E}$ is the field strength of the wave in volts per meter, ψ is the angle between the plane of polarization and the wire in which the voltage is induced, and θ is the angle between the wave front and the direction of the antenna wire. It will be observed that the quantity $\mathcal{E}\cos\psi\cos\theta$ is the component of the electric field that is parallel to the antenna.

Useful output is obtained from an antenna by means of a series load impedance Z_L, as shown in Fig. 213. By Thévenin's theorem it is possible to replace the antenna with its distributed constants and distributed induced voltage by a generator of voltage E with internal impedance Z_a as shown in Fig. 213. The equivalent generator voltage E represents the voltage appearing across the load terminals when the load impedance Z_L is infinite, and the impedance Z_a is the impedance that the load sees when looking toward the antenna. The ratio of generator voltage E to field strength $\mathcal{E}$ is termed the *effective height* of the receiving antenna. The current that flows in this equivalent circuit is exactly the same as the current that flows in the corresponding part of the real antenna.

The energy absorbed by the load is maximum when the resistance of the load equals the resistance component of the antenna impedance, and the reactance of the load is equal in magnitude but opposite in sign to the reactive component of the equivalent antenna impedance Z_a.

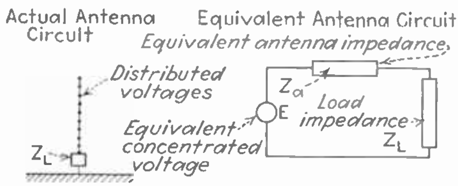


FIG. 213.—Actual receiving antenna with load impedance Z_L and distributed induced voltages together with equivalent antenna circuit in which the distributed voltages are replaced by a lumped voltage E and the distributed antenna impedance by an equivalent concentrated impedance Z_a.

Reciprocal Relations between Receiving and Radiating Properties, and the Rayleigh-Carson Reciprocity Theorem.—Many properties of an antenna are the same when radiating as when receiving. These reciprocal relations between the transmitting and receiving properties of an antenna are incorporated in reciprocal theorems, the most important of which was discovered by Rayleigh and extended to include radio communication by John R. Carson. It is to the effect that, *if a voltage E inserted at point A in antenna* 1 *causes a current I to flow at point B in antenna* 2, *then the voltage E applied at point B in antenna* 2 *will produce the same current I (both in magnitude and phase) at point A in antenna* 1. The Rayleigh-Carson theorem fails to be true only when the propagation of the radio waves is appreciably affected by an ionized medium in the presence of a magnetic field, and so holds for all conditions except short-wave transmission over long distances. Even then, it is to be expected that *on the average* the theorem will still apply, even when it cannot be depended upon to be exactly correct at every instant.

An important consequence of the reciprocity theorem is that the directional characteristics of an antenna are the same when it is abstracting energy from passing waves as when it is radiating its own field.

116. Effect of the Ground. Image Antennas.—When an antenna is near the ground, energy radiated toward the earth is reflected as shown in Fig. 214, so that the total field in any direction represents the sum of a direct wave plus a reflected wave. In the case of a perfect earth (infinite conductivity) the reflection is complete and can be taken into account by replacing the ground by an image antenna as illustrated in Figs. 214 and 215. The fields produced by the joint action of the actual antenna and its image are the same fields that actually exist in the space above earth in the presence of the ground.

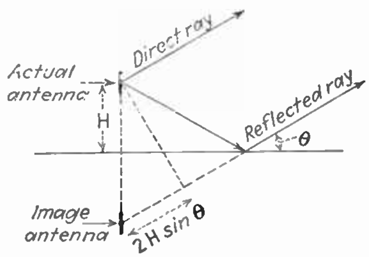


FIG. 214.—Diagram illustrating how the wave reflected from the earth can be considered to have been produced by an image antenna.

Examples of image antennas for a number of cases are given in Fig. 215. The general principles for setting up the image antenna for a perfect earth are as follows:

1. The image antenna has a physical configuration that is the mirror image of the actual antenna.

2. The currents in corresponding parts of the actual and image antennas (*i.e.*, parts lying on the same vertical line and at the same distance from the earth surface) are of the same magnitude, and flow in the same direction when the corresponding parts are vertical and in the opposite direction when they are horizontal.

The wave reflected from the ground cancels the direct radiation from the antenna in certain directions and adds to the direct radiation in other

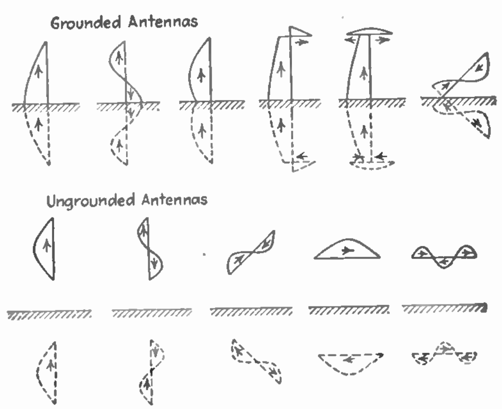


FIG. 215.—Images for common types of antennas.

directions. This is apparent from Fig. 214, where it is seen that the direct and indirect paths to a distant observer differ in length by an amount that depends upon the height of the antenna above ground and the angle of elevation θ involved. Hence the phase with which the direct and reflected waves combine will depend upon the height above ground and the angle of the radiation above the horizontal, as well as upon whether the image currents and antenna currents have the same or opposite phase.

Analysis of Ground Action in the Case of a Perfect Earth.—The resultant field produced by the radiation from an elementary antenna and its corresponding image can be analyzed with the aid of Fig. 214. Here it will be observed that the difference in distance to a distant observer from the antenna and its image is $2H \sin \theta$. This is likewise the difference in path length of the direct and reflected waves and corresponds to a phase difference of $2\pi(2H/\lambda) \sin \theta$ radians. Hence for a perfect earth, and an image of the same polarity as in the radiating antenna (*positive image*), one has

$$\left.\begin{matrix}\text{Actual radiation in}\\ \text{presence of ground}\end{matrix}\right\} = 2 \cos\left(2\pi\frac{H}{\lambda}\sin\theta\right)\left\{\begin{matrix}\text{Radiation from}\\ \text{antenna when}\\ \text{in free space}\end{matrix}\right. \qquad (125a)$$

Similarly, when the image currents are of the opposite polarity from the currents in the radiating antenna (*negative image*), then

$$\left.\begin{matrix}\text{Actual radiation in}\\ \text{presence of ground}\end{matrix}\right\} = 2 \sin\left(2\pi\frac{H}{\lambda}\sin\theta\right)\left\{\begin{matrix}\text{Radiation from}\\ \text{antenna when}\\ \text{in free space}\end{matrix}\right. \qquad (125b)$$

In these equations H/λ is the height of the center of the antenna above ground, measured in wave lengths, and θ is the angle of elevation above the horizontal.

Examination of Fig. 215 shows that the factors of Eqs. (125*a*) and (125*b*), although derived for an elementary portion of an antenna, will also give the effect of the ground in the case of ungrounded vertical or horizontal antennas of finite length. Thus for a vertical antenna that is an odd number of half wave lengths long, the effect of the ground can be calculated by means of Eq. (125*a*) on the basis of a positive image. Similarly, the effect of the ground in the case of a horizontal antenna, or a vertical antenna that is an even number of half wave lengths long, can be calculated by Eq. (125*b*) assuming a negative image. In either case the antenna height H used in the equations is the height of the center of the antenna above ground.

The nature of the factor $\left[2 \sin\left(2\pi\frac{H}{\lambda}\sin\theta\right)\right]$, which takes into account the effect of a negative image, is indicated in Figs. 216 and 217. It is seen that the ground reflection causes cancellation along the ground and at certain vertical angles. The angle of elevation of the first lobe above the horizontal also decreases as the height of the antenna above earth is increased. Consequently, in order to obtain strong radiation in directions

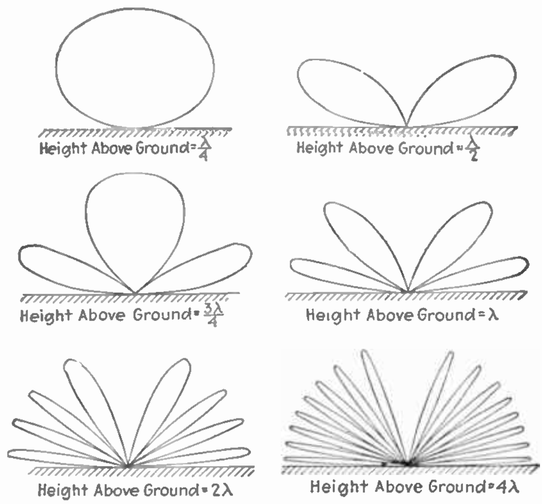


FIG. 216.—Polar diagram of the negative image factor $\left[2 \sin \left(2\pi \frac{H}{\lambda} \sin \theta\right)\right]$ for various values of H/λ, showing how the height above earth affects the directional characteristic of the antenna.

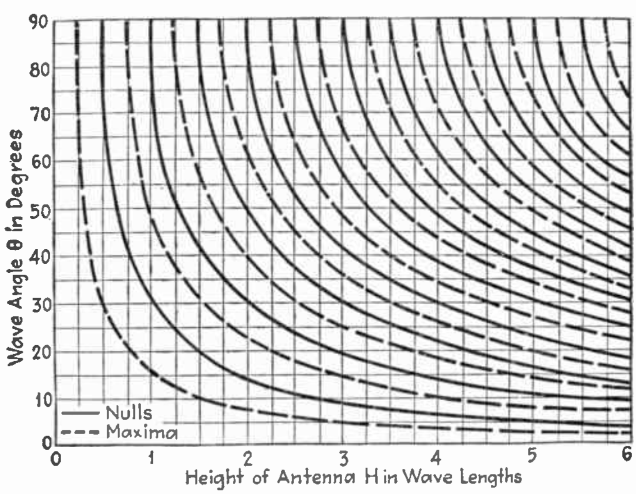


FIG. 217.—Chart showing the vertical angles at which the negative image factor $\left[2 \sin \left(2\pi \frac{H}{\lambda} \sin \theta\right)\right]$ is maximum and zero. These represent the vertical angles at which the reflections from the ground cause complete reinforcement and complete cancellation, respectively, of the radiation. With a positive image the positions of the nulls and maxima are interchanged.

approaching the horizontal in the presence of a negative image, it is necessary that the height above earth approach one wave length.

With a positive image the positions of the lobes and nulls are interchanged from the conditions shown in Figs. 216 and 217, and there is a maximum of radiation along the ground.

Effect of Imperfect Earth.—In the practical case of an imperfect earth, losses in the ground cause the reflected wave to be smaller in magnitude and to differ slightly in phase from the reflected wave obtained with a perfectly conducting earth. The modifications produced in the directional pattern by an imperfect earth are small, however, unless the earth is an unusually poor conductor and the vertical angle is large. Inasmuch as under ordinary conditions only the radiation at low or moderate vertical angles reaches the receiver, it is hence customary to assume a perfect earth in making calculations of the field radiated by an antenna.[1]

Effect of Ground on Receiving Antennas.—The reciprocity theorem shows that the effect of the ground is the same for reception as for radiation. Thus if an antenna radiates little or no energy in a particular direction as a result of the effect of the ground, it will likewise have little or no voltage induced in it by waves arriving from the same direction. The effect of the earth in the case of a receiving antenna can accordingly be calculated by Eq. (125*a*) or (125*b*), as the case may be.

117. Directional Characteristics of Simple Antennas.—The directional characteristics of transmitting antennas are important because only those waves radiated in certain directions from the transmitter will reach a particular receiving point, and all energy radiated in other directions is wasted as far as transmission to this receiver is concerned. Directional characteristics of receiving antennas are likewise important because an antenna that abstracts more energy from waves coming from the direction of the transmitter than from waves of equal strength arriving from other directions will not pick up as much interfering signals and static arriving from other directions.

The fundamental factors determining the directive characteristics of antennas can be understood by considering a series of simple cases that incorporate the various principles involved.

Resonant Wire Remote from Ground.—The directional characteristics of a single wire remote from ground and having a sinusoidal current distribution depends upon wire length as shown in Fig. 218. The directional characteristic is seen to consist of a number of lobes, the largest of which is the one making the smallest angle with respect to the wire axis. Increase

[1] The effect of an imperfect earth can be taken into account by assuming that the currents in the image antenna are less than the currents in the radiating antenna by the factor A and likewise differ in phase from the perfect ground case by an angle Φ, where A and Φ are the magnitude and phase of the reflection coefficients as calculated from Eq. (116).

in the wire length reduces the angle of this major lobe and also increases the number of minor lobes, as shown.

The directional characteristic of a long wire differs from that of an elementary antenna because the current in different parts of the long wire may not be in the same phase, and because the distance from a remote

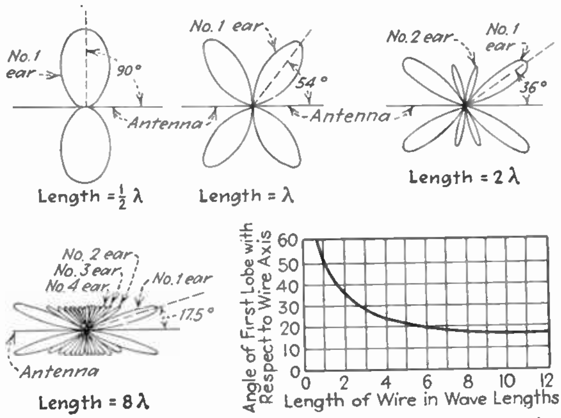


FIG. 218.—Polar diagrams showing strength of field radiated in various directions from an antenna consisting of a wire remote from the ground, together with a curve giving the angle the first lobe makes with the wire as a function of wire length.

point to various parts of the long wire will in general differ. The result is that the field radiated from different elementary sections of a long antenna add vectorially to give a sum that depends upon the direction. Thus in Fig. 219 an observer in the direction P receives no radiated field because he is equidistant from the two halves of the antenna, and these halves carry currents of opposite phase. An observer P_2 likewise receives no field, because an elementary length radiates zero field in the direction of the axis. On the other hand, the field in the direction P_1 is relatively strong because, although the radiations from the two halves of the antenna start out in this direction in opposite phase, the distance from P_1 to the two halves of the antenna is different, and cancellation does not take place.

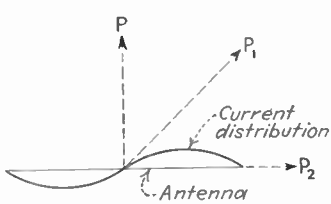


FIG. 219.—Diagram illustrating factors controlling the directional characteristic of an antenna. Note that the radiation does not cancel in the direction P_1 because of the different distances to the two ends of the antenna.

Grounded Vertical Wire.—The directional characteristic of a vertical wire grounded at the lower end depends upon the height of the antenna measured in wave lengths, as shown in Fig. 220. When this height is equal to or less than a quarter wave length, the strength of the radiated field is almost exactly proportional to the cosine of the angle of elevation.

With heights equal to or slightly greater than a half wave length there is marked concentration of radiation along the horizontal, and with heights of $3\lambda/4$ and greater most of the radiation is at a rather high vertical angle.

Grounded vertical antennas are sometimes provided with a flat-top structure, as in the fourth and fifth antennas of Fig. 215. When a flat top is present, the total radiated field consists of the vertically polarized radiation produced by the current in the vertical portion and its image, plus the horizontally polarized field radiated from the flat top and the image of the flat top. The fraction of the total field radiated from the

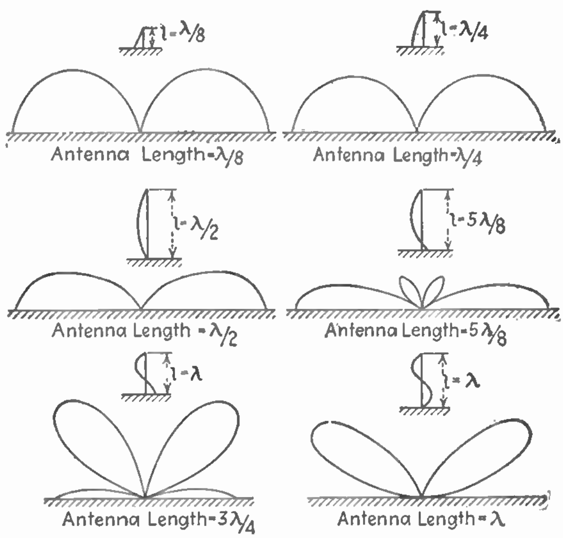
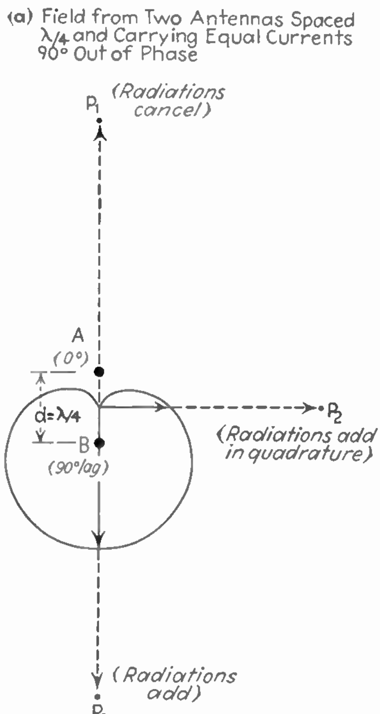


FIG. 220.—Directional characteristics in a vertical plane produced by grounded vertical antennas of varying lengths. These polar diagrams can be thought of as cross sections of a figure of revolution about the axis of the antenna, and they assume a perfectly conducting earth. A sinusoidal current distribution is also assumed.

flat top is so small in most flat-top antenna designs, however, that it is customarily neglected.

The directional characteristic of an antenna with flat top is substantially the same as the directional characteristic of a simple vertical antenna having a height somewhat greater than the height of the vertical portion of the flat-top antenna. This is because the presence of the flat top increases the current in the vertical portion of the antenna as shown in Figs. 211 and 215.

Spaced Antennas.—An important means of obtaining directivity is by combining the radiation from two or more spaced antennas. A simple example of such directivity is illustrated in Fig. 221*a*, which shows two vertical antennas spaced a quarter wave length apart, carrying equal currents, and excited so that they are 90° out of phase. To an observer at the point P_3, the radiation from the two antennas adds because, although the radiation leaves antenna A a quarter of a cycle ahead of the radiation from antenna B, it takes this first radiation a quarter cycle to travel the quarter wave length to antenna B, and the two radiations add in the direction toward P_3. However, to an observer at P_1 the radiations from the two antennas cancel because the radiation from antenna B starts a quarter cycle late and loses an additional quarter cycle in traveling to A, thereby arriving exactly in the correct phase to cancel the radiation that is starting out from A towards P_1. A distant observer at P_2 is equidistant from the two antennas, and, since the radiations start out with a phase difference of 90°, they will add up in quadrature at P_2 to give a result that is $\sqrt{2}$ times the field radiated from a single antenna. The resulting directional pattern is a cardioid as shown in Fig. 221a. By changing either the spacing, the relative phase, or both, other directional patterns are obtained as illustrated in Fig. 221b.



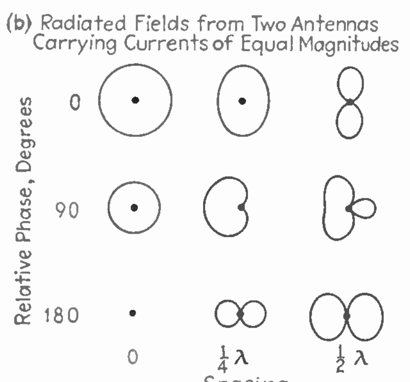


FIG. 221.—Diagram illustrating how directivity is obtained by the use of two spaced antennas, together with directional patterns obtained from a pair of spaced antennas with different spacing and phasing but carrying currents of equal magnitude. In part (b) the antennas are on a horizontal line.

More complicated systems of spaced antennas are considered in the section below on antenna arrays.

Non-resonant Long-wire Antennas.—When the end of an antenna is properly terminated, the current distribution is as illustrated in Figs. 28c and 222a. The current dies away exponentially with distance from the sending end, and there is a phase shift that increases uniformly with distance and amounts to 360° per wave length.

A long wire with such a current distribution has a directional characteristic determined by the same factors that have been considered above. However, because the law of amplitude and phase variation of the current is different, the directional characteristics will as a result be modified. The results for several typical cases are shown in Fig. 222. Here it is seen that the maximum radiation takes place at an angle, with respect to the axis of the wire, which decreases as the wire becomes longer and is affected only slightly by the attenuation. It will be noted that the radiation pattern is substantially unidirectional.

Gain of Directive Antennas.—The merit of a directional antenna is most conveniently measured in terms of the antenna *gain*, which is defined

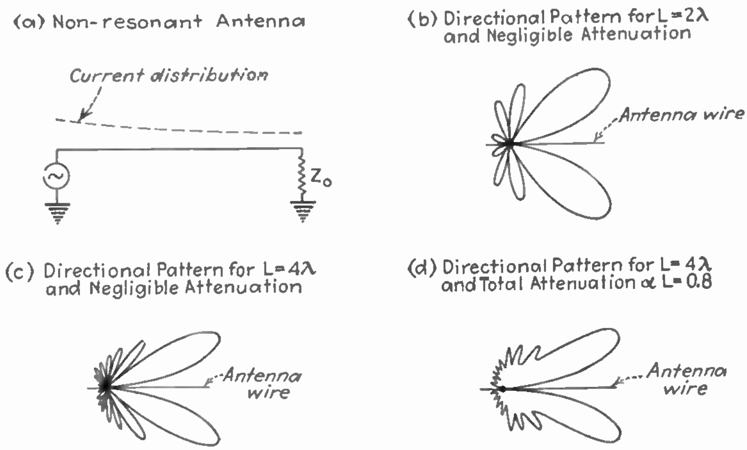


FIG. 222.—Typical directional characteristics for a non-resonant wire in space. These patterns may be considered as cross sections of a figure of revolution in which the wire is the axis.

as the power that must be supplied to a standard comparison antenna to radiate a given field strength in the desired direction, divided by the power that must be supplied to the directive antenna to accomplish the same result. Thus a gain of 100 (= 20 db) means that the directive antenna requires only $1/100$ as much power to produce a given field strength in the desired direction as does the comparison antenna. Conversely, when both antennas are supplied with the same power, the directive system will make the received energy 100 times as great (corresponding to a tenfold increase in field strength).

The comparison antenna is usually taken either as a wire one-half wave length long and at arbitrary orientation and height above earth, or as a very short vertical wire.

118. Antenna Arrays.—Directional arrangements involving a combination of two or more spaced antennas are termed antenna arrays.

The array represents one of the most widely used methods of obtaining directivity.

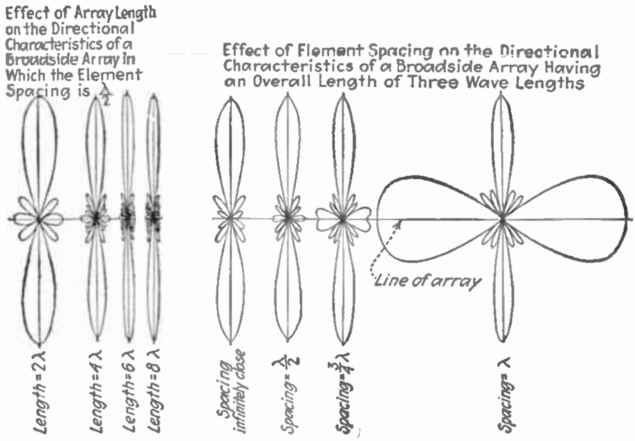


FIG. 223.—Effect of array length and element spacing on the directional characteristics of field radiated from a broadside array in a horizontal plane.

Broadside Array.—A broadside array consists of a number of antennas carrying equal currents, excited in the same phase, and spaced at uniform intervals along a straight line. The broadside array is characterized by a concentration of radiation in a plane at right angles to the line of the array, with a tendency for radiation in other directions to be canceled.

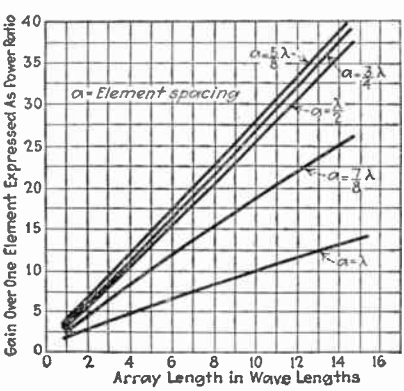


FIG. 224.—Gain of broadside array as a function of array length for various element spacings. These curves are for the case where radiation from the individual antenna is proportional to the cosine of the angle of elevation, and they give the gain over a single element of the array.

The properties of typical broadside arrays are illustrated in Fig. 223. The sharpness of the directional pattern in a plane containing the line of the array is proportional to the array length and is substantially independent of the spacing of the individual radiators, provided this spacing does not exceed a certain critical value. The gain of the array is likewise proportional to the length, as shown in Fig. 224.

The individual antennas making up the array can be of any desired type. Thus it is possible to use either horizontal or vertical radiators, or each elementary radiator of the array can itself be an array, thus giving an *array of arrays.* In any case the broadside arrangement increases the sharpness in a plane containing the line of the array. Thus the radiation can be concentrated at low vertical angles without concentrating in a horizontal plane by making the line of the array vertical.[1]

The maximum permissible spacing of the individual radiators depends upon the extent to which the individual radiators concentrate their radiation in the desired direction. If the elementary radiators have a nondirectional characteristic in the plane being considered, such as is the case in Fig. 223, then the maximum allowable spacing is $3\lambda/4$. However, if the individual radiators already possess some directivity, the spacing

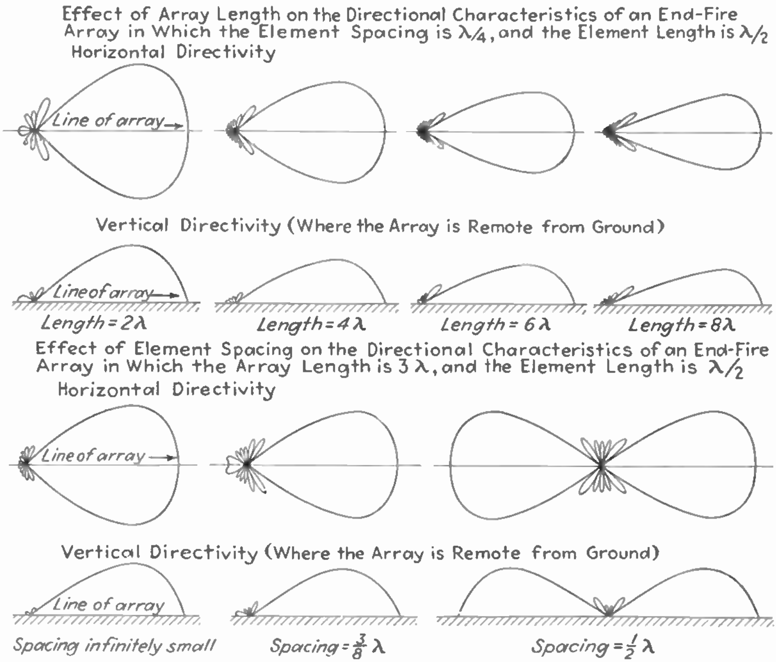


FIG. 225.—Effect of array length and element spacing on the directional characteristic of field radiated from an end-fire array.

can be increased. Thus, in a broadside array composed of horizontal radiators with their axes in the line of the array, the antennas may be spaced as much as one wave length without spoiling the directional characteristic in the horizontal plane because a horizontal wire possesses appreciable directivity in the horizontal plane. When the elementary radiators are themselves arrays with considerable horizontal directivity, spacings of two to three wave lengths are often permissible.

The directional characteristics of a broadside array can be calculated by summing up the fields produced by each element of each component of the array. Formulas for doing this are given in Sec. 119.

[1] This special form of broadside arrangement is termed a *stacked array.*

End-fire Array.—The end-fire array consists of a series of antennas arranged in a line, carrying equal currents and excited so that there is a progressive phase difference between adjacent antennas equal in cycles to the spacing between antennas in wave lengths. Thus, if the adjacent antennas are a quarter wave length apart, a phase difference of 90° between adjacent antennas is called for. The result of such a phasing system is to concentrate the radiation in the direction toward the end of the array having the most lagging phase, as illustrated in Fig. 225. The unidirectional nature of the resulting characteristic makes the end-fire array of considerable practical importance.

The principal properties of an end-fire array are illustrated in Figs. 225 and 226. The degree of concentration in both horizontal and vertical planes is proportional to the length of the array and is also independent of the spacing of the elementary radiators, provided this spacing does not exceed a critical value. The gain of the array is likewise proportional to length and nearly independent of spacing up to the critical spacing. This critical spacing depends upon the directivity of the individual radiators, and is approximately $3\lambda/8$ with antennas whose length does not exceed $\lambda/2$, but it can be greater if the individual radiators have greater directivity than a half-wave radiator.

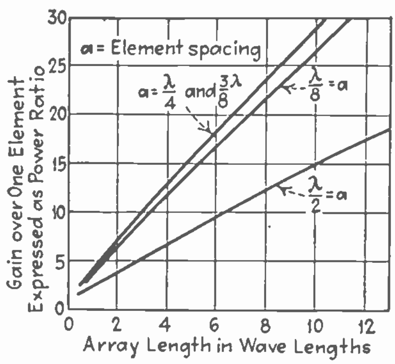


FIG. 226.—Gain of end-fire array as a function of array length for various element spacings. These curves are calculated on the assumption that the radiation from the individual antenna is proportional to the cosine of the angle of elevation.

The directive characteristics of end-fire arrays can be calculated in the same manner as the broadside array by adding up the radiations produced in the desired direction by each element of each antenna. The necessary formulas for doing this are given in Sec. 119.

Fishbone Antenna.—The fishbone antenna is illustrated in Fig. 227 and is a special form of end-fire array in which the necessary phase relations are maintained by a non-resonant line. The outstanding characteristic of the fishbone antenna is the ability to perform satisfactorily over a considerable band of frequencies without any readjustments whatsoever. This type of antenna is widely used in the reception of short-wave signals.

The fishbone antenna consists of a series of collectors arranged in colinear pairs and loosely coupled to the transmission line by small capacitances (usually the insulator capacitance) as indicated in Fig. 227. The collectors are usually but not necessarily horizontal. The phasing is obtained by taking advantage of the fact that, when a transmission line is terminated in its characteristic impedance, there is a uniform phase shift along the line of 360° per wave length, which automatically phases the collector antennas properly with respect to each other.

In the usual design the collectors are each about 0.3 wave length long at the optimum frequency, and are spaced a distance not to exceed $\lambda/12$ at this frequency. The length of the array is commonly three to five wave lengths and is limited by the reactive loading that the coupled antennas place upon the transmission line. With these proportions the antenna is effective for frequencies ranging from about 1.2 to 0.5 of the optimum frequency, or over a frequency range exceeding 2:1.[1]

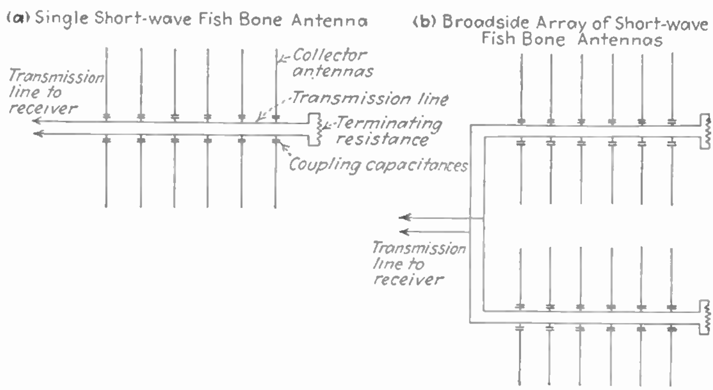


Fig. 227.—Plan view of single fishbone antenna and of an array consisting of two such antennas in broadside.

When the directivity desired in a horizontal plane is greater than that given by a single fishbone, two such arrays can be placed broadside as illustrated in Fig. 227*b*, thus giving a two-element broadside array, each component of which is a fishbone antenna.

Array of Arrays.—The antenna arrays used in practice are nearly always combinations that can be considered as an array, each element of which is itself an antenna array. A typical example of an array of arrays is illustrated in Fig. 228*a*, which shows two broadside arrays spaced an odd number of quarter wave lengths apart and excited 90° out of phase with each other. Such an arrangement gives a unidirectional pattern as shown in Fig. 228*c*, instead of the bidirectional pattern of Fig. 228*b* that is obtained with a single array.

[1] The frequency range is limited at low frequencies by the fact that very little energy is transferred between collector antennas and the line, because the collectors are so far out of resonance, and at high frequencies because, when the collectors approach resonance, the resulting reactive loading on the line alters the phase shift per unit distance along the transmission line and hence makes it impossible to maintain the proper phasing.

Many short-wave antenna arrays consist of several arrays such as shown in Fig. 228*a* stacked one above the other to give directivity in both vertical and horizontal planes. Such an antenna is illustrated in Fig. 229. Here the broadside arrangement in the horizontal direction gives

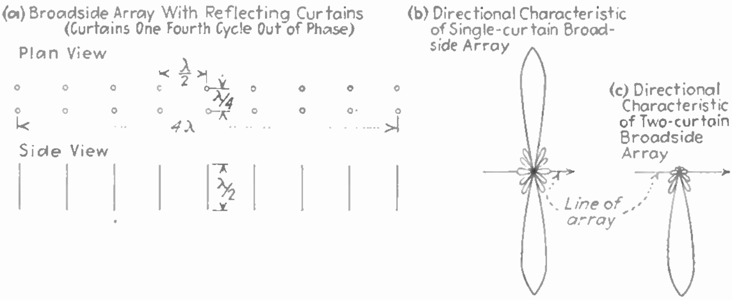


FIG. 228.—Directional characteristic of single- and double-curtain broadside array. The spacing of the two curtains can if desired be any odd multiple of a quarter wave length.

horizontal directivity, and the broadside arrangement in the vertical plane (*i.e.*, the stacking), gives directivity in a vertical plane, and the use of two curtains spaced an odd number of quarter wave lengths apart

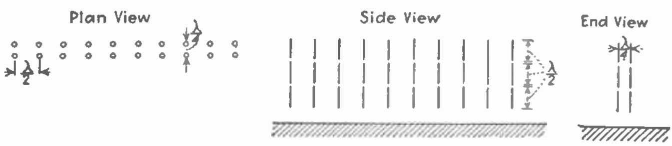


FIG. 229.—Typical antenna array consisting of a stacked broadside array having two curtains to give a unidirectional beam. The spacing between the curtains may be any odd multiple of a quarter wave length.

and with a phase difference of 90° gives a unidirectional characteristic. The result is a very intense beam of radiation confined to a small cone directed along the horizontal and aimed broadside to the antenna array.

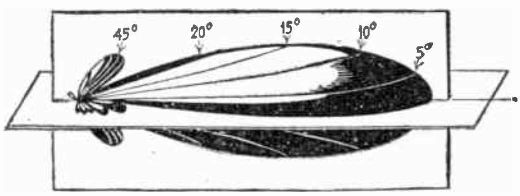


FIG. 230.—Directional characteristic in three-dimensional space of field radiated from a two-curtain broadside array. (*From Southworth.*)

The directional characteristic of such an array in three-dimensional space is illustrated in Fig. 230.

The V (or Folded-wire) Antenna.—This type of antenna is illustrated in Fig. 231. It consists of two long resonant wires arranged to form a V and excited so as to carry equal currents that are in phase opposition. The radiation from the individual wires of the V antenna has the character illustrated in Fig. 218, so that by making the apex angle twice the angle that the first lobe makes with the wire (see Fig. 218) there is a concentration of radiation in the direction of the bisector of the apex angle. The resulting directional pattern is as illustrated in Fig. 231 and

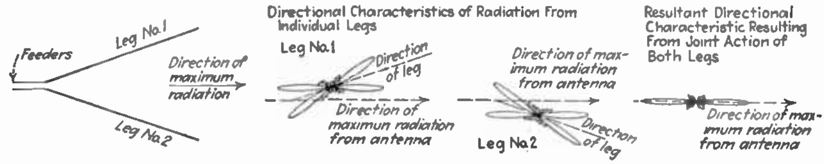


FIG. 231.—Resonant V antenna, showing how the radiation from the two legs combines to give a well-defined beam.

possesses very marked directivity in the plane of the V, particularly when the legs are long. Some directivity is also obtained in the vertical plane although this is less pronounced.

Increased directivity can be obtained by means of an array, each element of which is a V antenna. Thus the backward radiation can be eliminated by the use of two V's spaced an odd number of quarter wave lengths apart and excited 90° out of phase to give an end-fire action

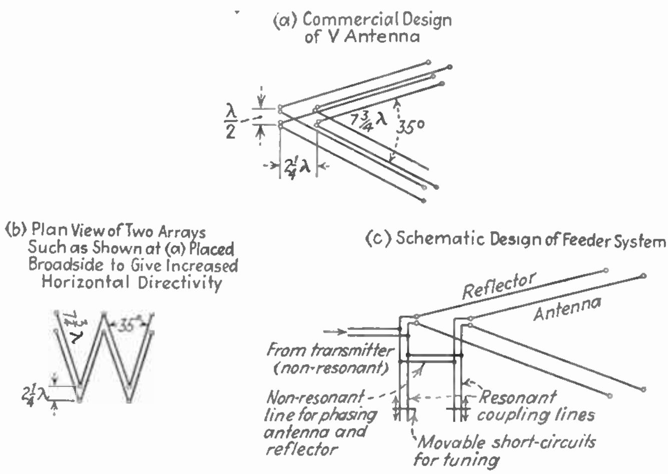


FIG. 232.—Commercial designs of resonant V antenna array together with feeder circuits.

(Fig. 232). The directivity in a vertical plane can be improved by stacking two or more V's above each other as illustrated in Fig. 232*a*. The vertical directivity is also determined to a considerable extent by the height above earth, since the radiation from a horizontal V is horizontally polarized. Greater directivity in a horizontal plane can likewise be realized by arranging V's in broadside as illustrated in Fig. 232*b*.

The gain of an array such as illustrated in Fig. 232*a* located at an average height of one wave length is approximately 39 times as compared with a half-wave antenna.

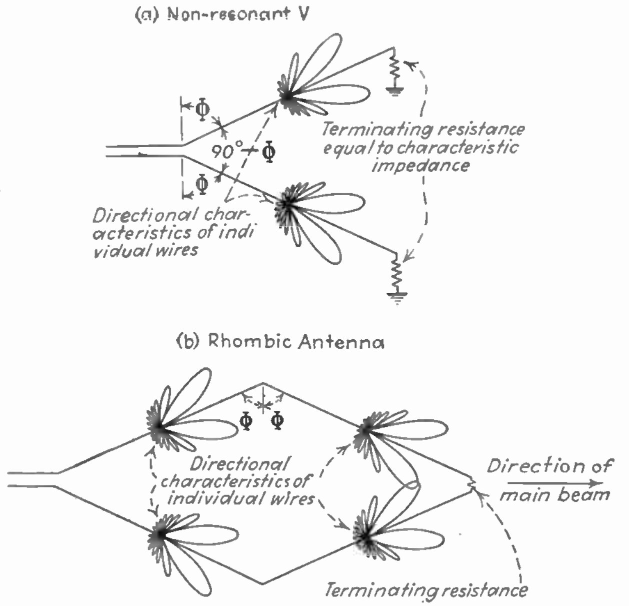


FIG. 233.—Various forms of non-resonant antenna systems.

Rhombic and V Antennas Employing Non-resonant Wires.—Directional antenna systems can also be formed by combining non-resonant wires to form an array. Thus the V antenna of Fig. 231 can be converted into a non-resonant antenna array by terminating the open ends of the V with resistances to ground equal to the characteristic impedance, as shown in Fig. 233*a*. This causes the current distribution in the wires to vary exponentially with distance as shown in Fig. 222*a*. The optimum apex angle for such a non-resonant V antenna should be such that the tilt angle Φ has the value given by the solid line in Fig. 234. When the length of the legs exceeds about two wave lengths, the apex angle is seen to vary only very slowly with the length measured in wave lengths. As a result, a non-resonant V with long legs will maintain its directive pattern reasonably well over a considerable frequency range.

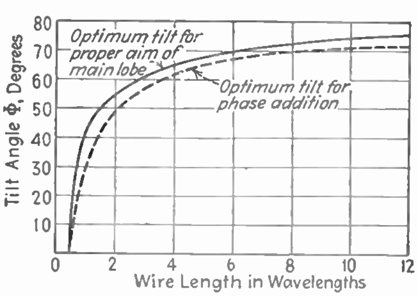


FIG. 234.—Optimum tilt angle for non-resonant antennas in space (assuming zero attenuation).

Compared with the corresponding resonant V antenna, the non-resonant arrangement has the advantage of being unidirectional and of operating over a range of frequencies without any change in adjustment. The disadvantage of the non-resonant arrangement is that considerable power is lost in the terminating resistance, and that practical difficulty is met in obtaining a ground resistance that will stay constant with varying weather conditions.

A more satisfactory form of non-resonant array is obtained by arranging four wires in the form of a diamond, or rhombus, as shown in Fig. 233*b*. By making the tilt angle Φ lie in the range between the dotted and solid lines of Fig. 234, the main lobes of radiation from each leg of the diamond are aimed in substantially the same direction and add in phase

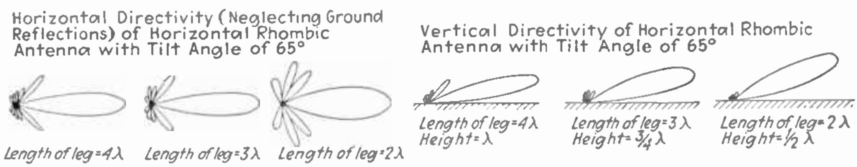


FIG. 235.—Polar diagrams showing directional characteristics of the same horizontal non-resonant rhombic antenna for three different frequencies.

in this direction. As a consequence the radiation from all legs adds in the desired direction, while tending to cancel in other directions.

The rhombic antenna is superior to the non-resonant V in that it not only has greater directivity, but also maintains this directivity better with changes in frequency. Furthermore, the terminating resistance of the rhombic antenna does not involve the earth, and therefore is not troubled with variation of ground resistance.

Typical directional characteristics obtained with the same rhombic antenna operated at different frequencies are shown in Fig. 235. The greatest directivity is obtained at the highest frequency, where the length of the legs measured in wave lengths is greatest, but the general character of the directional pattern is maintained without change over the 2:1 variation in frequency.

In the usual case where the plane of the rhombic antenna is parallel to the earth, the radiation is horizontally polarized, and the most important factor controlling the directivity in a vertical plane is the height above ground. The tilt angle and the length of the legs also influence the vertical directivity, however, and can be so chosen as to compensate to a considerable extent for lack of proper height.

119. Methods and Formulas for Calculating the Directional Characteristics of Antenna Systems.—The general method of calculating the directional characteristics of antenna systems has already been outlined in Sec. 114. Briefly it consists in determining the magnitude and phase of the field produced at a distant point by each elementary length of the antenna, and then adding the resulting fields vectorially for all the elementary lengths of the entire antenna system. This method can be applied irrespective of the geometry involved, the number of antennas, the type of antennas, the current distribution, etc.

Formulas for calculating the radiated field and directional characteristics of common antenna systems are given below.

1. Radiation from an elementary length of wire in the absence of ground [from Eq. (120)]:

$$\mathcal{E} = \frac{60\pi}{d\lambda} I(\delta l) \cos \theta \tag{126}$$

where

$\mathcal{E}$ = field strength in volts per meter
d = distance to antenna in meters
θ = angle of elevation with respect to plane perpendicular to antenna wire
λ = wave length of radiated wave in meters
δl = length of elementary antenna in meters
I = current in antenna.

2. Radiation from a resonant wire remote from ground:

a. When the wire is an odd number of half wave lengths long,

$$\mathcal{E} = \frac{60I}{d} \frac{\cos\left(\pi \dfrac{L}{\lambda} \cos \theta\right)}{\sin \theta} \tag{127a}$$

b. When the wire is an even number of half wave lengths long,

$$\mathcal{E} = \frac{60I}{d} \frac{\sin\left(\pi \dfrac{L}{\lambda} \cos \theta\right)}{\sin \theta} \tag{127b}$$

where

$\mathcal{E}$ = field strength in volts per meter
d = distance to antenna in meters
I = current in amperes at a current loop
L = length of antenna in meters
λ = wave length in meters
θ = angle measured with respect to wire axis.

3. Radiation from vertical grounded wire, assuming a perfect earth:

$$\mathcal{E} = \frac{60I}{d}\left[\frac{\cos 2\pi \dfrac{L}{\lambda} - \cos\left(2\pi \dfrac{L}{\lambda} \sin \theta\right)}{\cos \theta}\right] \tag{128}$$

where θ is the angle of elevation with respect to ground, and the remainder of the notation is as in Eq. (127).

4. Radiation from a non-resonant wire remote from ground:

$$\mathcal{E} = \frac{30I_0}{d\sqrt{1 + (\alpha/p)^2}} \frac{\sin\theta}{(1 - \cos\theta)}(1 + \epsilon^{-2\alpha L} - 2\epsilon^{-\alpha L}\cos pL)^{1/2} \qquad (129a)$$

where

$\mathcal{E}$ = field strength in volts per meter
I_0 = current entering the line
d = distance to antenna in meters
α = attenuation constant of line
$p = (2\pi/\lambda)(1 - \cos\theta)$
θ = angle with respect to wire axis
L = length of wire in meters
λ = wave length in meters (assumed same on wire as in space).

In the special case where the attenuation α can be assumed zero (*i.e.*, when the wires are so short that the current can be considered to be constant),

$$\mathcal{E} = \frac{60I_0}{d}\frac{\sin\theta}{1 - \cos\theta}\sin\left[\pi\frac{L}{\lambda}(1 - \cos\theta)\right] \qquad (129b)$$

5. Radiation from a rhombic (diamond) antenna with characteristic impedance termination, neglecting the attenuation along the wires, and with antenna remote from ground:

$$\left.\begin{matrix}\text{Relative field}\\ \text{strength}\end{matrix}\right\} = \left(\frac{\cos(\Phi - \beta)}{1 - \cos\theta\sin(\Phi - \beta)} + \frac{\cos(\Phi + \beta)}{1 - \cos\theta\sin(\Phi + \beta)}\right)$$
$$\times \sin\left\{\frac{\pi l}{\lambda}[1 - \cos\theta\sin(\Phi + \beta)]\right\}\sin\left\{\frac{\pi l}{\lambda}[1 - \cos\theta\sin(\Phi - \beta)]\right\} \qquad (130)$$

where

Φ = tilt angle of antenna (see Fig. 233)
β = bearing angle with respect to line passing through the apex having the terminating resistance, and the diagonally opposite apex
θ = angle of elevation with respect to the plane of the antenna
l = length of each side of the rhombus
λ = wave length.

6. Radiation from a resonant V antenna (see Fig. 231) when the length of each leg is an even multiple of a half wave length and the antenna is remote from ground:

$$\left.\begin{matrix}\text{Field strength}\\ \text{in plane of V}\end{matrix}\right\} = \sqrt{\mathcal{E}_a^2 + \mathcal{E}_b^2 - 2\mathcal{E}_a\mathcal{E}_b\cos\left(2\pi\frac{l}{\lambda}\sin\alpha\sin\Phi\right)} \qquad (131a)$$

where

$\mathcal{E}_a$ and $\mathcal{E}_b$ = radiation in desired direction from individual legs of antenna as given by Eq. (127)

l = length of leg

λ = length corresponding to one wave length

α = half of angle at apex

Φ = bearing angle with respect to bisector of apex.

$$\left.\begin{array}{l}\text{Radiation in vertical plane passing}\\ \text{through bisector of apex angle}\end{array}\right\} = \frac{120I}{d}\left[\frac{\sin\left(\frac{n\pi}{2}\cos\alpha\cos\theta\right)\sin\alpha}{1-\cos^2\theta\cos^2\alpha}\right] \tag{131b}$$

where

n = number of half wave lengths in each leg of antenna

α = half of angle of apex

θ = angle of elevation with respect to plane of antenna

I = current at current loop

d = distance to antenna in meters.

7. Radiation from an antenna array remote from ground and consisting of $\mathfrak{N}$ parallel planes each made up of N parallel columns where each column is made up of n individual radiating elements carrying equal currents and radiating uniformly in all directions:

$$\text{Relative field strength} = \left|\frac{\sin n\pi(a\cos\Phi\cos\theta + b)}{n\sin\pi(a\cos\Phi\cos\theta + b)}\right.$$
$$\left.\frac{\sin N\pi(A\sin\Phi\cos\theta + B)}{N\sin\pi(A\sin\Phi\cos\theta + B)}\,\frac{\sin\mathfrak{N}\pi(\mathfrak{A}\sin\theta + \mathfrak{B})}{\mathfrak{N}\sin\pi(\mathfrak{A}\sin\theta + \mathfrak{B})}\right| \tag{132}$$

where

n, N, and $\mathfrak{N}$ = number of radiators along x-, y-, and z-axes, respectively

a, A, and $\mathfrak{A}$ = spacing of adjacent radiators along x-, y-, and z-axes, respectively, measured in fractions of a wave length

b, B, and $\mathfrak{B}$ = phase displacement between adjacent radiators along x-, y-, and z-axes, respectively, measured in fractions of a cycle

θ = angle with respect to xy-plane (angle of elevation)

Φ = angle with respect to xz-plane (bearing angle).

8. Radiation from an antenna array in which the individual antennas do not radiate uniformly in all directions is obtained by calculating the directional characteristic using Eq. (132) and then multiplying the result

by the actual directional characteristic of the individual antenna involved.

9. Radiation from an array of arrays is determined by first calculating the individual directional characteristic of the elementary array. This is then multiplied by the directional characteristic of the array consisting of non-directional radiators located at the centers of the individual arrays, using Eq. (132).

10. Loop antennas are considered in Sec. 124.

11. The effect of the ground in the case of ungrounded antennas can be taken into account by the methods discussed above in connection with Eqs. (125*a*) and (125*b*). The ground effect is already included in Eq. (128) for the grounded antenna.

Agreement between Calculated and Observed Antenna Characteristics.—Experience has shown that the gain and the directional characteristics of antenna arrays calculated on the basis of a perfectly conducting earth agree satisfactorily with observed results under nearly all conditions. A comparison of theoretical and observed directional characteristics for a typical horizontally polarized antenna array is given in Fig. 236 and shows excellent agreement.

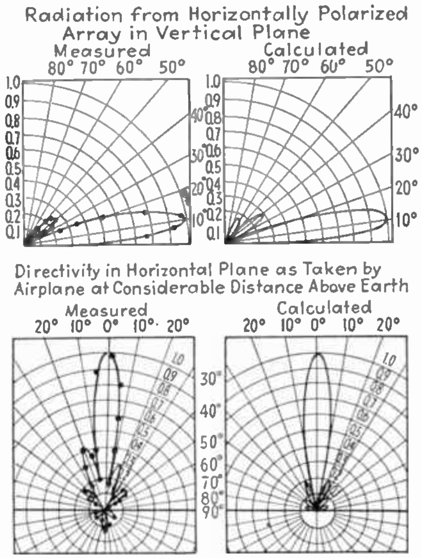


FIG. 236.—Comparison of measured and calculated distribution of radiated field about array composed of horizontal elements. The calculated results assume a perfect earth.

120. Miscellaneous. *Energy Relations in Transmitting Antennas.* The total amount of energy radiated from a transmitting system can be conveniently measured in terms of a "radiation" resistance. This is the resistance that, when inserted in series with the antenna, will consume the same amount of power as is actually radiated. The magnitude of the radiation resistance depends upon the antenna configuration and upon the point in the antenna system at which the resistance is considered as being inserted. Unless specifically stated to the contrary, it is customary to refer the radiation resistance to a current maximum.

The radiation resistance can be determined by assuming the antenna to be at the center of a large sphere, and calculating the field produced at the spherical surface by any convenient antenna current. The energy passing through each square centimeter of the surface is $0.00265\mathcal{E}^2$ watts, where $\mathcal{E}$ is the field strength in r.m.s. volts per centimeter. The total energy radiated through the entire spherical surface is then found by mathematical or graphical integration.[1]

The effect of the ground upon the fields is taken into account by assuming an image antenna. The integration to obtain the total radiated energy is then carried out only over the hemisphere above the earth's surface, instead of over the complete spherical surface as would be the case with an isolated antenna.

Energy Relations in Receiving Antennas.—The resistance component of the antenna impedance in the equivalent circuit of Fig. 213 is $R_r + R_l$, where R_r is the radiation resistance, and R_l takes into account the losses in conductors and dielectrics. The energy delivered to the load impedance connected in series with the antenna is maximum when the reactance of this load is equal in magnitude and opposite in sign to the reactance of the antenna impedance, and when the resistance component R_L of the load equals the total antenna resistance $R_r + R_l$. Under these conditions the fraction $R_L/(R_L + R_r + R_l)$ of the energy abstracted from the wave is delivered to the load to be usefully employed. At the same time the fraction $R_l/(R_L + R_r + R_l)$ is wasted in wire, ground, and other loss resistances, while the fraction $R_r/(R_L + R_r + R_l)$ is reradiated. This reradiation of energy results from the fact that when current flows in an antenna, radiation takes place irrespective of the source of voltage producing the current.

The maximum energy that an antenna can possibly abstract from a passing wave is fixed by the fact that the antenna resistance can never be less than the radiation resistance. Calculations indicate that a section of the wave front approximately a quarter of a wave length square is capable of supplying all the energy that can be abstracted by a small antenna.

Induction Fields.—An antenna carrying radio-frequency currents is surrounded by electrostatic and magnetic fields that give rise to the inductance and capacitance of the antenna wire. These fields are termed *induction* fields and are present in addition to the radiated fields given by Eqs. (120) and (121). The strength of the induction field is inversely proportional to the square of the distance, whereas the strength of the radiation field is inversely proportional to the distance. As a result, the induction fields are negligible at considerable distances from the antenna, although they are normally much stronger than the radiated fields in the immediate vicinity of the radiator. Because of this situation, measurements of radiated field must be made at distances at least 3 to 5 wave lengths away from the antenna in order to avoid errors from induction fields.

[1] A more detailed discussion of radiation resistance calculations by this as well as other methods, together with references to the literature on the subject, is given in Sec. 134 of "Radio Engineering," 2d ed.

Parasitic Antennas.—Directivity is sometimes obtained by the use of auxiliary antennas placed in the vicinity of the main antenna but not actually connected with it. When the main antenna radiates energy, voltages are induced in these parasitic antennas that cause currents to flow in them. This induced current then produces radiated fields that combine with the radiation from the main antenna. The directional characteristic of the resulting combination depends upon the tuning of the parasitic antennas and upon their spacing with respect to the radiating antenna.

In the case of reception, the parasitic antennas abstract energy from the passing wave. This energy is then reradiated to the main receiving antenna and modifies the direction characteristics.

The most common use of parasitic antennas is to produce a unidirectional effect in antenna arrays. Thus in arrays such as illustrated in Figs. 228 and 229 the rear curtain is commonly excited parasitically. The optimum spacing for unidirectional action under these circumstances is very nearly a quarter wave length, and the parasitic antenna must be tuned so that its radiation is in the proper phase to produce a substantially unidirectional characteristic.

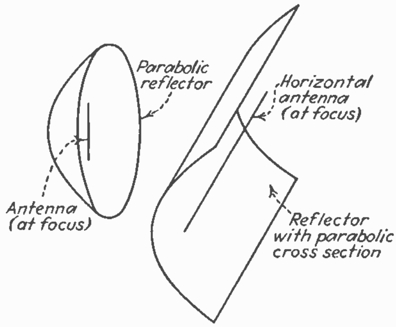


Fig. 237.—Examples of conducting reflectors for producing directional effects.

Directivity by the Use of Reflectors.—Another method of obtaining directivity is by the use of reflectors. Copper acts as a practically perfect reflector of radio waves because of its high conductivity, and so can be used as the basis of reflectors that will concentrate radiation in much the same way that reflecting surfaces are used to concentrate light radiation. Thus a parabolic reflector formed of copper, with the antenna at the focus as shown in Fig. 237, will concentrate the radiated energy in a beam in exactly the same way that a searchlight produces a beam of light. Similarly, in the case of reception, such a reflector gathers in energy arriving from the desired direction and concentrates it upon the antenna.

The usefulness of reflectors in concentrating radiation is limited by the fact that the dimensions of the reflector must be proportional to the wave length in order to form a sharp beam. The method is consequently practicable only at ultra-high frequencies.

121. Radio-frequency Transmission Lines and Impedance-matching Systems.—The output of a radio transmitter is often delivered to the antenna by means of a transmission line. Transmission lines are also used to connect receivers to receiving antennas, particularly directional antennas, special all-wave antennas, and apartment-house antenna systems. By proper design the energy can be transmitted to distances of the order ½ to 1 mile with only moderate loss.

In transmission lines used with transmitting antennas, the principal considerations are the energy lost, the cost, and the reliability. Open wire lines, similar to telephone lines except for more adequate insulation, are generally employed because of their low cost. Concentric conductor lines consisting of a central conductor surrounded by a copper tube are more expensive but can be made less subject to weather conditions and so are used to a limited extent.

The principal requirement for a transmission line used with a directional receiving antenna is that negligible energy shall be abstracted from passing radio waves. Otherwise the line will pick up energy from signals and noise that produce no effect upon the highly directional antenna, and the benefits of the directional antenna system will then be largely lost. The transmission-line arrangements usually employed for receiving antennas are the balanced four-wire transmission line, the concentric conductor line, and the twisted pair. The four-wire balanced line consists of four lines arranged at the corners of a 1- to 2-in. square, with the diagonally opposite wires connected in parallel. The wires are held in position by insulator blocks spaced 5 to 15 ft., and arranged so that each block introduces a transposition. This arrangement is nearly as good as a concentric conductor line and is much less expensive. The twisted-pair transmission line has rather high losses and so is satisfactory only when the transmission distances do not exceed a few hundred feet.

Resonant and Non-resonant Lines.—Transmission lines may be operated in such a manner as to be resonant or non-resonant, according to the load impedance at the receiving end of the line. If the receiving-end load is a resistance equal to the characteristic impedance (*i.e.*, a resistance equal to $\sqrt{L/C}$, where L and C are the inductance and capacitance of the line per unit length), then voltage and current die away uniformly as one recedes from the sending end of the line, and have a distribution that follows an exponential law as shown at Fig. 28*c*. The voltage and current at every point along the line are then in phase with each other, the ratio of voltage to current at any point is the characteristic impedance $Z_0 = \sqrt{L/C}$, and the phase drops back uniformly at a rate of 360° per wave length as the receiving end is approached.[1]

[1] In the case of lines with air insulation, the distance corresponding to a wave length on the line is almost exactly equal to the wave length of the corresponding radio wave. However, in the case of lines having appreciable dielectric insulation, as, for example, a twisted pair, the distance representing a wave length on the line is correspondingly less.

When the load impedance does not equal the characteristic impedance of the transmission line, resonance effects are produced, as illustrated in Fig. 28. With resonant lines the voltage and current follow periodic variations that repeat every half wave length, and there is a large phase difference between voltage and current except near points where the voltage or current passes through a minimum. In traveling from generator toward load there is a progressive phase lag of 180° per half wave length, but, instead of taking place at a uniform rate, as in the case of non-resonant lines, most of the shift takes place in the vicinity of the voltage and current minima.

Non-resonant lines transmit energy at unity power factor and so are more efficient and radiate less than resonant lines, and also subject the line insulation to less voltage in proportion to power transmitted. Hence non-resonant lines are always used in association with high-power transmitters and in all cases where the transmission distance is appreciable. Non-resonant lines are suitable with low-power transmitters only where the transmission distance is quite short.

Characteristic Impedance of Transmission Lines.—The characteristic impedance Z_0 of a transmission line is determined by the line construction, and at radio frequencies can for all practical purposes be considered to be a resistance defined by the equation[1]

$$Z_0 = \sqrt{\frac{L}{C}} \text{ ohms} \tag{133a}$$

where L and C are the inductance and capacitance, respectively, per unit length of line. For the two-wire and concentric-conductor transmission lines, substitution of the usual formulas for inductance and capacitance gives

$$\left.\begin{array}{l}\text{Characteristic imped-}\\ \text{ance of two-wire line}\end{array}\right\} = 276 \log_{10} \frac{b}{a} \text{ ohms} \tag{134a}$$

$$\left.\begin{array}{l}\text{Characteristic imped-}\\ \text{ance of concentric cable}\end{array}\right\} = 138 \log_{10} \frac{b}{a} \text{ ohms} \tag{134b}$$

In these equations b is the spacing between conductor centers of the two-wire line, and the inner radius of the outer conductor of the concentric cable. Similarly a represents the radius of the conductor in the two-wire

[1] The exact formula for the characteristic impedance is

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{133b}$$

where R and G are the resistance and conductance per unit length of line. At high frequencies ω is so large that R and G can be neglected, thereby resulting in Eq. (133a).

line, and the outer radius of the inner conductor in the case of the concentric line. With typical construction, the characteristic impedance is usually about 600 ohms with a two-conductor open-air line, and in the neighborhood of 75 ohms for the usual concentric conductor with air dielectric. With a twisted-pair transmission line the characteristic impedance is commonly of the order of 125 ohms, but will depend greatly upon the construction.

Impedance-matching Systems for Non-resonant Lines.—When resonances are to be avoided in transmission lines associated with a trans-

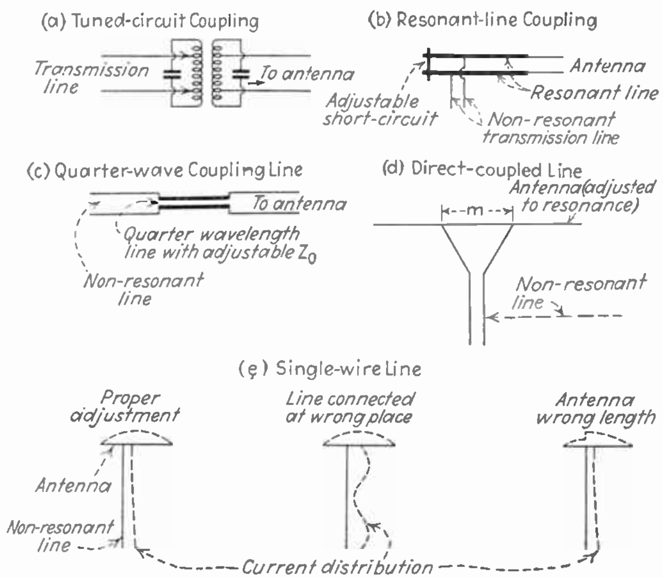


FIG. 238.—Commonly used impedance-matching systems.

mitter, it is necessary that the antenna be coupled to the line in such a way that the effective load impedance that is offered to the line is a resistance equal to the characteristic impedance of the line. Typical methods for doing this are illustrated in Fig. 238. At Fig. 238*a* a tuned circuit is used to provide the impedance matching. A resistance load is obtained by proper tuning, and the magnitude of the load can be adjusted to the correct value either by varying the mutual inductance or by varying the fraction of the coil across which the line is connected.

In the resonant coupling line at Fig. 237*b*, a resistance impedance is obtained by adjusting the short circuit on the coupling line to give resonance in conjunction with the antenna. The resulting load offered the non-resonant line is then resistive and has a magnitude determined by the point of connection. The length of resonant line should be approximately a quarter wave length when the antenna resistance is higher than the characteristic impedance of the non-resonant line, and a half wave length when the antenna resistance is lower. Arrangements employing resonant coupling lines are shown in Fig. 232.

The arrangement of Fig. 238*c* makes use of the fact that in a quarter-wave-length line $Z_s = Z_0^2/Z_a$, where Z_s is the impedance that the non-resonant transmission line sees when looking toward the matching line, Z_0 is the characteristic impedance of the matching line (a resistance), and Z_a is the antenna impedance. By adjusting the antenna so that it is in resonance, Z_a becomes a resistance. The impedance Z_s, which serves as the load for the non-resonant line, is then likewise a resistance that can be made the desired value by varying the spacing between wires of the matching line to vary Z_0. At low frequencies, where a quarter wave length is inconveniently long, artificial lines can be employed.

In Fig. 238*d* the transmission line is directly coupled to the antenna. The matching is then obtained by making the antenna length the exact value required for resonance and connecting the two wires of the line symmetrically with a spacing *m* such as to give the required impedance match. The single-wire transmission line shown at *e* is a modification of Fig. 238*d* in which the ground supplies a return circuit that is completed through the capacitance of the antenna to ground. The proper adjustment of a single-wire transmission line can be determined by observing the relative current distribution in the line and antenna as indicated in the figure.

When it is desired to avoid resonances in a transmission line associated with a receiving antenna, it is necessary that the input impedance of the receiver be a resistance equal to the characteristic impedance of the line, since in this case the antenna is the source of energy and the receiver acts as the load. Failure to match the antenna to the transmission line prevents the receiving antenna from delivering the maximum possible amount of energy to the transmission line, but does not cause resonances.

Phasing Systems.—In antenna arrays it is necessary to adjust carefully the relative phases of the various antennas. There are two basic methods of accomplishing this. The first makes use of the fact that a non-resonant transmission line in open air has a uniform phase shift of 360° per wave length. Hence any desired phase difference can be obtained by the use of a non-resonant transmission line of suitable length. Thus in the antenna system of Fig. 232 the 2¼ wave-length non-resonant line connecting antenna and reflector makes the relative phase difference 2¼ cycles, *i.e.*, 90°. Phasing of end-fire antenna arrays is always accomplished by the aid of a non-resonant line.

The second method of phasing makes use of the fact that in a resonant line the phase shifts 180° every half wave length (even though the shift does not take place uniformly). This method is commonly used in phasing broadside antenna arrays, and is illustrated in Fig. 239. Here the radiators are connected to a resonant line at voltage maxima and are spaced a half wave length apart. By connecting successive radiators to alternate sides of the line all the radiators are then excited in the same phase.

Experimental Determination of Current Distribution along Transmission Lines and Antennas.—The way in which a transmission line or antenna is operating can ordinarily be deduced by observing the current distribution. In the case of transmitters the current distribution can be measured by coupling a sensitive thermocouple instrument to the line, using one of the arrangements in Fig. 240. In the case of receiving antennas the thermocouple can be replaced by a portable receiver, and the antenna excited by means of a portable oscillator adjusted to produce radio waves of the appropriate frequency.

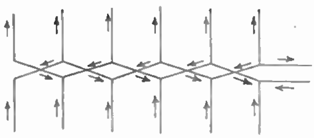


FIG. 239.—Phasing of broadside array by resonant line.

122. Practical Transmitting Antennas.—The principal considerations involved in the design of transmitting antennas are the directivity, efficiency, and cost. These matters become of increasing importance as the power of the transmitter is increased, since expenditures for improving the antenna system are then easier to justify.

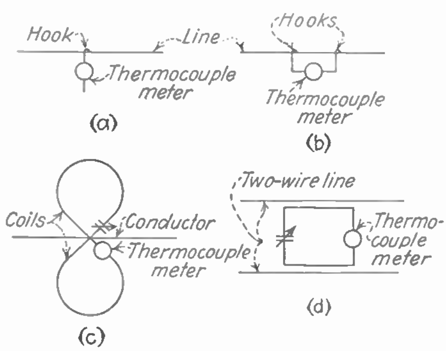


FIG. 240.—Devices for measuring the relative current distribution along a transmission line or antenna.

Short-wave Antennas.—Short-wave directive antennas find their chief use in long-distance high-power point-to-point communication. For this purpose it is desirable to concentrate the energy in a well-defined beam that is directed toward the receiving point at a vertical angle that is of the order of 10 to 25 deg. above the horizontal. The rhombic and resonant V antennas are usually preferred because they give good directivity, are easily adjusted, and can be supported upon inexpensive wooden poles. The rhombic antenna has the additional advantage that it will operate satisfactorily over a frequency range approximately 2½:1 with no readjustment, and so can be used for both the day and night frequencies. Before the resonant V and rhombic antennas were developed, broadside arrays such as illustrated in Fig. 229 were generally employed, but these have lost favor because they are difficult to tune, require expensive supporting structures, and give no better directivity.

Non-directional radiators consisting of a vertical or horizontal half-wave antenna are commonly used for amateur, marine, and short-wave broadcast work, where it is desired to transmit signals in all directions. Typical examples of such antenna systems together with typical transmission-line arrangements are shown in Fig. 241.

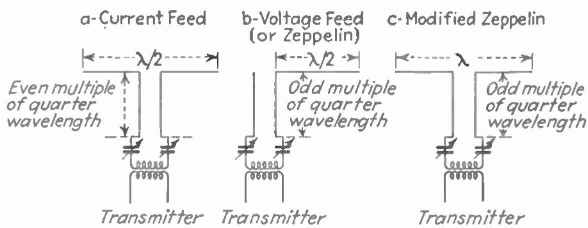


FIG. 241.—Representative types of resonant transmission lines for delivering energy to antennas.

The polarization of short-wave transmitting antennas is not important since the ionosphere causes the received wave to contain both vertical and horizontal components, irrespective of the polarization at the transmitter. Horizontal polarization is usually preferred, however, because the supporting structures required are generally not so high as for vertical radiators with the same vertical directivity.

The proper procedure for tuning an antenna depends upon the arrangement. In the case of arrays involving many radiators, such as illustrated in Fig. 229, the length of the individual parts are adjusted until the current distribution as experimentally observed has the desired character. In other cases where the antenna is associated with a resonant line such as in Figs. 232 and 241, it is possible to tune the antenna and the resonant line as a unit. At Fig. 232 this is done by adjusting the position of the short-circuiting bar, whereas in arrangements of the type shown in Fig. 241 the tuning can be accomplished by means of the variable condensers in series with the line, provided the antenna and transmission-line lengths are approximately correct to begin with. In cases such as Figs. 238*d* and 238*e*, where the antenna must offer a resistance impedance, it is necessary to adjust the antenna length very accurately to the proper value.

In general it will be found that the antenna length required for resonance is usually less than the corresponding length in wave lengths measured in free space. This is because of end effects and mutual coupling in antenna elements. Thus in the case of a half-wave antenna it is necessary to make the wire about 5 per cent shorter than a true half wave length.

The efficiency of short-wave antennas is very high because of the high radiation resistance of the antenna in proportion to size, and because the antenna is ungrounded. Even including the effect of ground losses, the over-all efficiency is commonly well in excess of 80 per cent, and most of the losses that do occur are the result of ground imperfections.

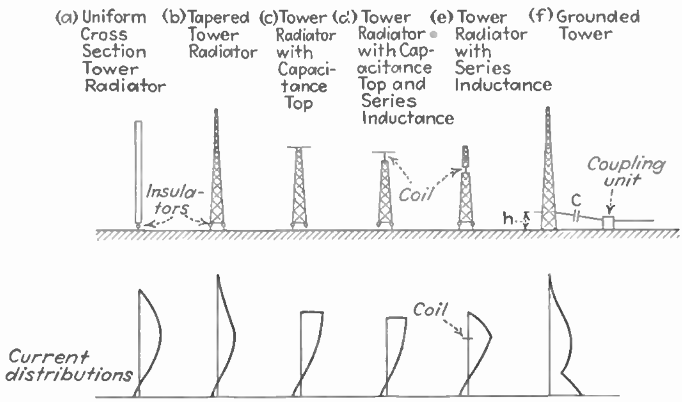


FIG. 242.—Different types of vertical radiators, with current distribution for each.

Ultra-high-frequency Antennas.—Antennas for use with ultra-high-frequency transmitters are of the same general character as those used at short waves. However, since the length corresponding to a wave length is less, it is possible to obtain greater directivity with antennas of reasonable proportions.

Transmitting Antennas for Broadcast Frequencies.—The preferred broadcast antenna consists of a structural steel tower that functions as a vertical radiator. This arrangement combines simplicity, low cost, and good electrical efficiency. Actual details of the tower antenna vary with the installation, with a number of typical arrangements shown in Fig. 242.

Broadcast antennas can be conveniently divided into two types. First are those antennas which do not attempt to obtain directivity in a vertical plane. These employ a tower having a height of the order of $\frac{1}{8}$ to $\frac{1}{4}$ wave length and give a radiated field that is almost exactly proportional to the cosine of the angle of elevation. The second class of antennas is designed to concentrate the radiation along the horizontal and to radiate very little energy at high vertical angles. This result can be accomplished by employing a vertical tower radiator having a height between 0.5 and 0.625 wave length as in Fig. 220.

The height of a vertical radiator giving the most favorable vertical directional characteristic for broadcast purposes depends somewhat upon the extent to which the tower is tapered. With a tower of uniform cross-sectional area, or with a vertical wire, the optimum height is 0.53 wave length, which corresponds to minimum high angle radiation and good concentration of energy along the horizontal. With a tower wider at the base than at the top, the optimum height is slightly greater. This is because the taper causes the current distribution to depart from a sinusoidal law, as shown in Fig. 242. The result is less current in the topmost parts of the antenna, which makes these parts less effective. However, if the effect of the taper is compensated for by a slight increase in height, the directional characteristics will then be the same as with a tower of uniform cross section. The optimum height depends upon the taper and is of the order of 0.56 to 0.62 wave length.

A comparison of the relative merits of towers of tapered and of uniform cross section therefore shows that, contrary to a rather widely held belief, identical performances can be obtained in the two cases. The only difference is that with uniform-cross-section towers it is possible to calculate the optimum height exactly, whereas with tapered towers the best height cannot be predetermined with the same exactness.

When building restrictions, nearby airports, or other considerations limit the height of a tower antenna to less than the optimum, it is possible to use some of the expedients illustrated in Fig. 242 to obtain a directional characteristic corresponding to a height greater than actually used. These expedients all operate to increase the current carried by the top part of the radiator, as indicated in the figure. In this way it is possible to obtain a vertical directivity corresponding to a simple vertical radiator as much as 50 per cent higher. These arrangements also have the advantage that by adjusting the series inductance, etc., it is possible to determine the exact optimum height experimentally with relatively little trouble. This makes it possible to take into account the effect of ground conductivity, and also eliminates the slight uncertainty as to optimum height always present with self-supporting tapered towers.

Tower radiators are normally excited by insulating the tower from ground and applying the exciting voltage across this insulator. It is possible, however, to ground the tower and then use the arrangement in Fig. 242*f*. Here the coupling loop formed by the portion h of the tower, the ground, and the connecting wire containing the capacitance C is tuned to resonance with the transmitted frequency by adjusting the capacitance C. The voltage developed across the section h of the tower is then applied to the upper portion of the tower, thereby exciting the antenna.

The efficiency of a broadcast antenna depends largely upon the ground losses occasioned by currents flowing in the earth near the antenna. These losses can be made low by providing a ground system of buried wires as shown in Fig. 243, consisting of 90 to 120 radials having a length that is preferably about a half-wave length. These radials provide a low-resistance path for the currents that flow out through the ground to charge the capacitance between tower and ground. Ground losses in the immediate vicinity of the base of an ungrounded tower can be made negligible by placing the base insulators 4 to 10 ft. above earth, connecting

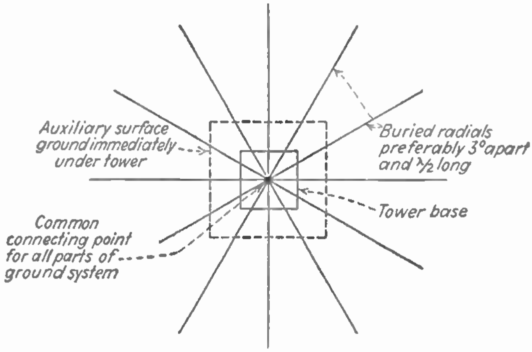


FIG. 243.—Typical ground system for vertical radiator.

the part of the legs below these insulators to the buried ground system, and placing a copper screen on the surface of the ground under the tower and connecting this screen to the ground system.

When the ground system is properly designed, the efficiency of a broadcast antenna is very high. Values in excess of 80 per cent are not unusual.

Simple antenna arrays are used to some extent in broadcast work to give horizontal directivity. Thus it is sometimes desired to concentrate most of the radiated energy toward the regions of maximum population, or to avoid interfering with another station. These objectives can ordinarily be achieved by means of a simple array consisting of two or three radiators properly phased.

Transmitting Antennas for Long Waves.—At frequencies below the broadcast band the distance corresponding to a wave length is so great that it is impracticable to obtain directivity with reasonable cost. The usual antenna system for the lower frequencies consists of a T or inverted L flat-top arrangement as illustrated in Figs. 211*e* and 211*f*, supported by two towers.

The efficiency of low-frequency antennas tends to be very small because of the low radiation resistance obtained with practical heights, and because the power losses in the ground and in the antenna tuning coil tend to be large.

Miscellaneous.—Steel towers, guy wires, and other conductors in the vicinity of an antenna tend to alter the directional characteristics and so must be located with considerable care. Steel supporting towers should not have a length that makes them resonant at the frequency being transmitted, and they should be located where the coupling to the antenna system is a minimum. Guy wires should not be too numerous, and must be broken up into lengths that are a small fraction of a half wave length by means of insulators.

Transmission lines used with transmitting antennas should approach the antenna at right angles so that there will be a minimum of coupling between the antenna and line.

Antennas for use with airplane transmitters introduce special problems. In some cases the antenna is strung from wing tip to tail, or from a short vertical pole to wing tip, etc., but such antennas have a low effective height and tend to be rather inefficient even with short waves. A more satisfactory arrangement from the point of view of efficiency is a trailing wire reeled out through a hole in the rear end of the fuselage.

123. Practical Receiving Antennas.—The main considerations involved in receiving antennas are the amount of energy that the antenna can deliver to the receiver, the directivity, the cost, and the freedom from extraneous disturbances.

A receiving antenna should abstract sufficient energy from passing waves so that even the energy abstracted from the weak radio waves representing static and other noises will under normal conditions be at least comparable with the thermal-agitation energy existing in the input of the receiver. The signal-to-noise ratio cannot then be improved by further increase in received energy, and no further improvement is possible in the antenna system as far as energy pick-up is concerned.

The amount of energy that can be abstracted by an antenna depends upon its physical size, the frequency, and the loss resistance. The abstracted energy tends to decrease with increased frequency, particularly when the antenna dimensions do not exceed a quarter of a wave length. As a result, the problem of obtaining adequate energy pick-up is most important at high and ultra-high frequencies.

When it is desired to avoid marked directional effects at broadcast and lower frequencies, it is customary to employ a single wire running to a height of 15 to 50 feet. With such an antenna the horizontal portions are relatively unimportant because waves of broadcast and lower frequencies are vertically polarized when near the earth.

The usual non-directional antenna for the reception of short-wave signals is a single wire perhaps 25 to 50 ft. long. This may be used as a vertically or horizontally polarized antenna, since short-wave signals ordinarily contain both components. However, most man-made disturbances are vertically polarized near the earth, and so can be minimized by a horizontal antenna.

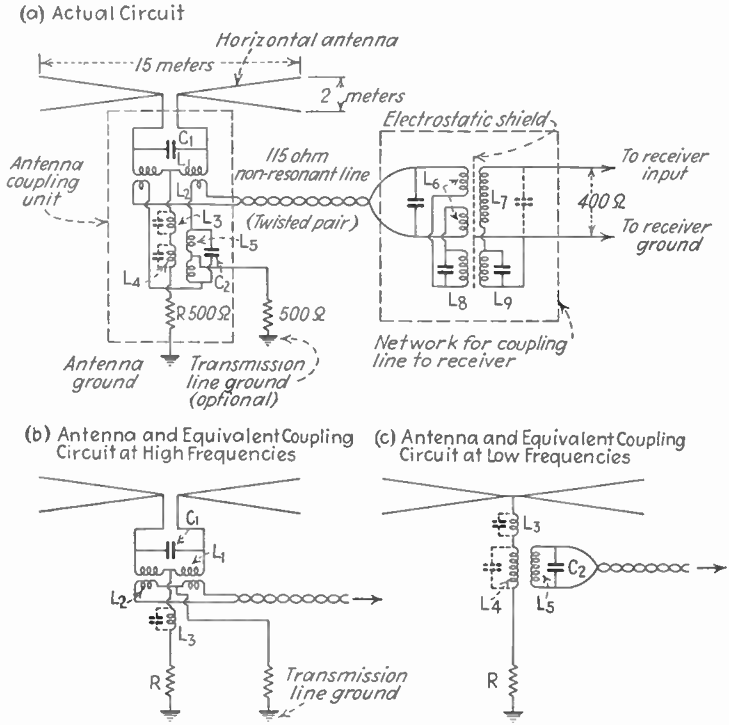


FIG. 244.—All-wave antenna system and equivalent circuits at high and low frequencies.

All-wave Antenna Systems for Broadcast Receivers.—The introduction of the all-wave receiver has led to the development of special antenna systems that give adequate energy pick-up over the entire frequency range with a minimum of directivity, and that tend to be insensitive to electrical disturbances produced by man. An arrangement of this sort is shown in Fig. 244. This consists of a horizontal antenna coupled to a transmission line by a complex network. This antenna-coupling network is so designed that at high frequencies the condenser C_2 is effectively a short circuit, causing the network to take the form shown in Fig. 244*b*. The antenna then consists of a horizontal doublet, and so tends to avoid man-made disturbances, which normally produce vertically polarized waves. At lower frequencies the coupling inductances L_1 and L_2 become so inefficient that the antenna system takes the form shown at Fig. 244*c*, which is a vertical antenna with a flat top. This transition between horizontally and vertically polarized action takes place at about 5 mc. By arranging the antenna system to change its mode of action in this way, it is possible to take advantage of the lower noise level obtained with horizontally polarized antennas at high frequencies, while retaining the vertically polarized antenna necessary at the lower frequencies where the waves are vertically polarized near the earth.

Directional Receiving Antennas.—In point-to-point radio communication it is of considerable advantage to employ directional receiving antennas, since in this way interference from static and radio signals coming from other than the desired direction can be reduced, with consequent improvement in the signal-to-noise ratio.

The most satisfactory directional antenna systems for short-wave work are the fishbone antenna of Fig. 227 and the rhombic antenna of Fig. 233*b*. Both of these are non-resonant, and so can operate over a range of frequencies without readjustment. The fishbone antenna is more complicated than the rhombic antenna, but has the advantage of a directional pattern that is substantially free of minor lobes. Other types of antennas, such as the resonant V, the broadside array, etc., have been used to some extent in reception, but are generally less desirable.

The maximum directivity usable in a receiving antenna is set by the angular deviations that can be expected from time to time in the received wave. There is also a limit to the permissible physical size of a short-wave directional antenna, since signals received at points separated by distances of the order of 10 wave lengths ordinarily have more or less random phase difference and so will not add up properly in the antenna. This causes a loss in directional characteristics, since the maintenance of the pattern requires proper phase relations for the different parts of the antenna. It is hence undesirable for the receiving antenna to be more than six to eight wave lengths in size.

In the construction of directional short-wave receiving antennas it is necessary to take special precautions to prevent the directional characteristics from being partially destroyed by the action of near-by conductors, such as towers, guy wires, and power lines. Such objects will abstract appreciable energy from waves that arrive in unwanted directions and will reradiate this energy to the receiving antenna, thereby largely neutralizing the advantages of the directional antenna. Hence it is customary to locate short-wave directional receiving antennas in spaces that are clear of trees, houses, power and telephone wires, etc., and considerable attention is also paid to minimizing the effect of towers and other metal structures that must necessarily be near the antenna.

Wooden telephone poles are particularly satisfactory when they can be used.

Problems

1. A vertical wire 10 meters long carries a current of 5 amp. of frequency 200 kc. Assuming the wire is in free space, calculate the strength of the radiated field produced at distances of 1, 10, 50, and 200 km in a direction at right angles to the axis of the wire.

2. Explain why a receiving antenna consisting of a horizontal wire has no voltage induced in it by a vertically polarized wave.

3. *a.* A particular antenna consists of a horizontal wire $2\frac{1}{2}$ wave lengths long. Sketch the nature of the current distribution in the antenna and also sketch the image antenna with its current distribution.

b. Repeat (*a*) for the case where the wire is vertical with its lower end $\lambda/4$ above ground.

4. Derive Eqs. (125*a*) and (125*b*).

5. A half-wave vertical antenna is placed so that its center is a half-wave above the ground. Assuming a perfect earth, calculate and plot the resulting directional pattern in a vertical plane.

6. An antenna consists of a horizontal wire a half wave length long and three-fourths wave length high. Calculate and plot the directional characteristics in vertical planes that: (*a*) are at right angles to the axis of the wire and (*b*) contain the wire. Assume a perfect earth.

7. Calculate and plot the directional characteristics of a resonant-wire antenna $2\frac{1}{2}$ wave lengths long and located in free space.

8. Calculate and plot the directional characteristics in a vertical plane of a vertical wire 0.53 wave length long and grounded at the lower end.

9. Three quarter-wave grounded vertical antennas spaced along a line with a spacing between adjacent antennas of a half wave length, are excited in the same phase. Calculate and plot the resulting directional characteristic for the horizontal plane.

10. An antenna system consists of two spaced grounded vertical wires carrying currents of the same magnitude and phase. Deduce a formula for the directional characteristic in a vertical plane containing the line joining the wires, assuming that the field from the individual radiator is proportional to the cosine of the angle of elevation, and that the spacing of the antennas is d/λ wave lengths.

11. In a non-resonant antenna it is possible, by the use of series condensers, to reduce the phase shift of current along the line to the point where the current everywhere is of the same phase. Discuss, qualitatively, the effect that this change can be expected to produce in the directional characteristic, with particular reference to the directions of maximum radiation.

12. If the tilt angle is not the optimum value in a non-resonant V antenna of the type shown at Fig. 233*a*, demonstrate qualitatively that, although the major lobes of radiation from the two sides of the V do not point in the desired direction, the sum of the two lobes does, provided that the tilt angle is not too far from optimum.

13. Design a horizontal rhombic antenna that will give optimum directivity at a wave length of 20 meters, when the length of each leg is limited to 50 meters and the radiation is to be concentrated at a vertical angle of 15 deg. above the horizontal. The design includes determination of antenna dimensions (in meters or feet), angles, and height above earth.

14. Calculate the directional characteristic in a vertical plane passing through the apex with terminating resistance of a rhombic antenna having a tilt angle of 60 deg., legs 3 wave lengths long, and a height above earth of one wave length.

15. Design a horizontal resonant V antenna for service at a wave length of 40 meters, when the length of each leg is 3¾ wave lengths, when the optimum vertical angle is 20 deg., and when a unidirectional characteristic is desired. The design includes specification of antenna dimensions (in meters or feet), apex angle, height above earth, and relative spacing and phasing of the two V's.

16. Derive a formula for the directional characteristic in a horizontal plane of the antenna system of Fig. 221*a*.

17. Derive a formula for the directional characteristic in the horizontal plane of an antenna system comprising two vertical wires spaced a distance d apart, carrying currents with magnitudes in the ratio A, and having a relative phase ψ.

18. *a.* Design a broadside array giving a gain of 20.

b. Design a reflecting curtain for the array of *a* that will give a unidirectional characteristic.

19. Design a fishbone antenna to receive long-distance signals in the frequency range 15 to 7.5 mc. The design includes specification of number, length, and spacing of the collector antennas and the height of the array.

20. A horizontal short-wave antenna is to be used to communicate with a receiver located 300 km distant. Assuming that the waves are reflected from the E layer, how high should the antenna be above earth?

21. When a parabolic reflector is used, explain why the focus must be an odd number of quarter wave lengths distant from the nearest part of the reflector.

22. A particular two-wire transmission line is to be constructed from No. 12 wires (diameter 0.08 in.) and must have a characteristic impedance of 600 ohms. Specify the spacing required.

23. Explain how unbalances to ground or dissymmetries in a transmission line will cause currents to flow out along the two transmission-line wires in parallel and return via the ground.

24. In (*b*) of Fig. 241, explain: (*a*) the means by which the coupling between line and antenna is obtained, and (*b*) the reason the line should be an odd number of quarter wave lengths long.

25. A half-wave transmitting antenna is to be coupled to a 600-ohm line so that the line will be non-resonant. If the connection to the antenna is made by opening the center of the antenna so that the antenna offers a resistance impedance of 73.4 ohms, design a quarter-wave coupling line of the type shown at Fig. 238*c*.

26. Calculate and plot the directional characteristic of the antenna system of Fig. 241*c*, assuming that the height above ground is one wave length, making calculations for vertical planes containing (*a*) the antenna wires and (*b*) at right angles to the antenna wires.

27. *a.* Calculate and plot the directivity in a vertical plane of grounded vertical antennas of heights $\lambda/8$, $3\lambda/8$, $\lambda/2$, 0.53λ and 0.60λ. Plot on the same polar coordinates and reduce all cases to the same radiation along the horizontal.

b. From the results discuss the effect of antenna height from the point of view of broadcast coverage.

28. Explain why a large receiving antenna is of advantage when the receiver is not particularly sensitive, but is of little or no advantage with a sensitive receiver.

29. In the all-wave receiving antenna of Fig. 244 what would be the practical effect of: (*a*) a poorer impedance match between the antenna and the line, (*b*) a poorer impedance match between the line and the receiver input, and (*c*) increasing the length of the transmission line?

CHAPTER XV

RADIO AIDS TO NAVIGATION

124. Radio Direction Finding.—Radio waves provide an important aid to navigation. Thus a ship or airplane can obtain its location by determining the direction of arrival of radio waves sent out by two or more transmitters at known locations. Similarly it is possible to determine the location of a radio transmitter by taking bearings on the radio waves at two receiving locations.

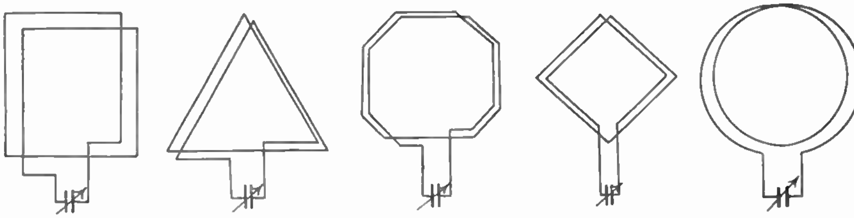


FIG. 245.—Examples of typical loop antennas.

Loop Antennas.—Most direction-finding systems make use of a loop antenna, which is essentially a coil of large cross section, as shown in Fig. 245. Passing waves induce voltages in the wires of the loop and produce a loop current that depends upon the orientation of the loop with respect to the wave front. The detailed action involved can be understood by considering the effect that vertically polarized waves produce on a vertical rectangular loop. When the plane of such a loop

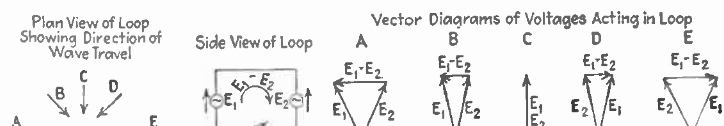


FIG. 246.—Vector diagrams showing how the voltages induced in the two sides of a loop by a passing radio wave combine to give a resultant voltage acting around the loop.

is perpendicular to the direction of travel of the waves, as shown by C in Fig. 246, the voltages induced in the two vertical sides are of equal magnitude and have the same phase. Being directed around the loop in opposite direction, however, these voltages cancel each other and result in zero loop current. On the other hand, when the plane of the loop is parallel to the direction of wave travel, as in cases of A and E of Fig.

246, the wave front reaches the two sides at slightly different times, causing a phase difference that gives rise to a resultant voltage acting around the loop. The result is then a directional response as shown in Fig. 247a. In the usual case when the loop is small compared with a

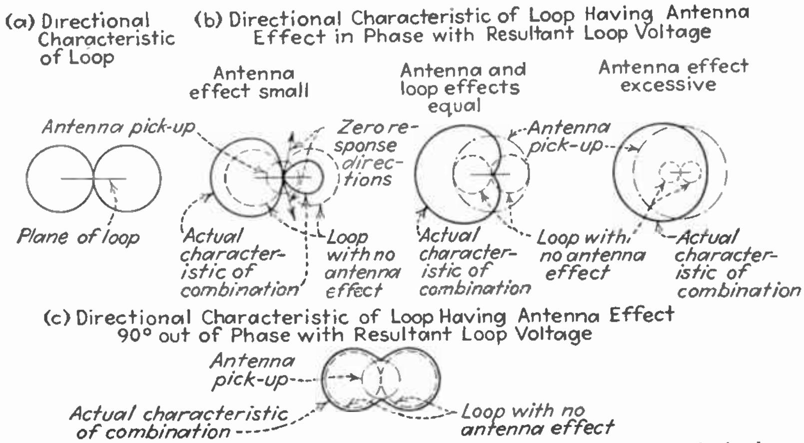


FIG. 247.—Directional characteristics of vertical loop for vertically polarized waves, with and without antenna effect. Note how a small antenna effect in phase with the loop voltage changes the relative amplitudes of the lobes and prevents the two null directions from being 180° with respect to each other, whereas antenna effect in quadrature with the loop voltage obscures the nulls.

wave length, the resultant voltage acting around the loop for any shape loop is[1]

$$\left.\begin{array}{l}\text{Resultant voltage}\\ \text{acting around loop}\end{array}\right\} = 2\pi \mathcal{E} N \frac{A}{\lambda} \cos \theta \tag{135}$$

where

A = loop area in square meters
$\mathcal{E}$ = strength of radio wave in volts per meter
N = number of turns in loop
λ = wave length of radio wave in meters
θ = direction of travel of wave with respect to plane of loop

Direction Finding with Loop Antenna.—The direction in which a radio wave travels can be determined by rotating a loop antenna until an associated radio receiver indicates zero response. The radio wave is then seen from Eq. (135) and Fig. 247a to be traveling in a direction that is perpendicular to the plane of the loop. It is customary to adjust

[1] This formula can be readily derived as follows, taking l and s as the height and width of the loop, respectively: The voltage induced in each vertical leg is $\mathcal{E}Nl$. The phase difference between the voltages is $(2\pi s \cos \theta/\lambda)$ radians, since the wave front must travel a distance $s \cos \theta$ in passing from one leg to the other. Subtracting the voltages in the two legs while taking into account this phase difference, gives Eq. (135).

the loop for minimum rather than maximum response, because the minimum is sharper than the maximum.

A simple loop antenna employed as described gives the bearing angle of the radio wave, but leaves a 180° uncertainty in the actual direction of the transmitting station. The sense of the bearing can be determined by inducing in series with the loop a small voltage derived from a vertical antenna but 90° out of phase with the voltage that the passing wave induces in the vertical antenna. The resulting direction pattern of the combined loop and vertical antenna for various amounts of antenna effect is shown in Fig. 247*b*. It is seen that when the antenna effect is not too great, one lobe of the loop pattern is enlarged, and the other is reduced. This action is caused by the fact, brought out by cases *A* and *E* of Fig. 246, that the equivalent voltage acting around the loop and representing the loop pick-up, is approximately 90° out of phase with the voltage induced in the wires of the loop and has a polarity that depends upon the direction of arrival of the radio waves. Hence when the non-directional energy pick-up of the vertical antenna is used to induce a voltage into the loop circuit with a 90° phase shift, this voltage will add to or subtract from the loop pick-up according to the direction from which the waves arrive.

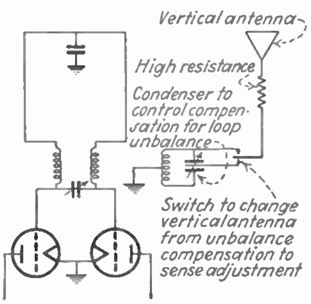


FIG. 248.—Circuit diagram of practical radio compass.

The sense arrangement in a typical direction-finding system is shown in Fig. 248. Here a vertical antenna located somewhere near the loop is provided with a high series resistance to make the antenna current substantially in phase with the voltage induced by the passing wave. This antenna circuit is arranged so that it can be inductively coupled to the loop. The voltage induced in the loop from the antenna is then 90° out of phase with the current (and hence voltage) produced in the vertical antenna by the passing wave. The series resistance in the antenna and the coupling between loop and antenna are adjusted so that the antenna effect is less than the loop pick-up, but at the same time is sufficient to produce a noticeable effect upon the directional pattern of the loop.

The practical procedure for determining the direction of arrival of radio waves with an arrangement such as shown in Fig. 248 is as follows: The bearing is first obtained by adjusting the loop for minimum response with the vertical antenna disconnected. The loop is then rotated 90° and the vertical antenna connected to one end of its coupling coil by means of a push button. The addition of the vertical antenna will then cause the receiver output to decrease or increase according to the sense of the bearing. By checking the system against a wave of known sense at the time of installation, and thereafter always making the 90° rotation of the loop in the same direction, all uncertainty as to the sense of the bearing is removed.

125. Errors in Loop Bearings.—A loop antenna will give spurious bearings unless it is electrically balanced with respect to ground. The reason for this can be understood by considering the action of the unbalanced loop of Fig. 249. When the loop is in the zero signal position the voltages induced in the vertical legs are of the same magnitude and phase, and, being directed around the loop in opposite directions, produce no resultant voltage around the loop circuit. However, because of the dissymmetry that results from grounding one side of the tuning condenser, capacitance currents flowing to ground through C_g' pass through the

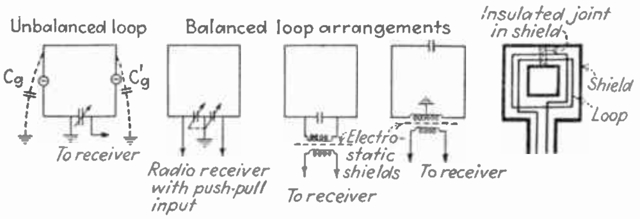


FIG. 249.—Unbalanced, balanced, and shielded loop arrangements.

tuning condenser, whereas the corresponding currents through C_g do not. This unbalance causes a signal to be delivered to the receiver even though the loop is in the position for zero response. The resulting effect is equivalent to coupling a small amount of vertical antenna pick-up into the loop circuit with a 90 degree phase shift. Zero response then corresponds to a loop position other than perpendicular to the direction of travel of the radio waves, as apparent from Fig. 247*b*. The presence of residual antenna effect can be detected by rotating the loop 180° after the zero response setting has been obtained. If there is no unbalance, the output will again be zero. However, if some antenna effect is present, a signal will appear upon reversal of the loop, because, as seen from Fig. 247*b*, the two zero directions are not exactly 180° with respect to each other in the presence of antenna effect.

Errors from unbalance can be minimized by using circuit arrangements that are symmetrical with respect to ground, such as shown in Fig. 249. It is also helpful to enclose the loop in an electrostatic shield, such as a metal housing broken by an insulated bushing as shown schematically in Fig. 249. Such a shield insures that all parts of the loop will always have the same capacitance to ground irrespective of the loop orientation, or of neighboring objects. Finally, where maximum accuracy is required, it is essential that residual unbalances be compensated for by a small balancing condenser such as provided in Fig. 248, which can be used to introduce a controllable amount of compensating antenna effect. The proper adjustment of this balancing condenser is obtained by connecting the antenna to the center plate and determining experimentally the condenser setting for which a 180° rotation of the loop does not effect the null.

The bearings obtained with a loop are influenced by the presence of wires and other conductors in the vicinity of the loop. This is because these metal objects abstract energy from the passing wave and then produce radiation and induction fields that induce spurious voltages in the near-by loop. It is accordingly always necessary to make an experimental correction curve for loop bearings unless the space near the loop is entirely free of metal objects, including those that are buried. When the influence of near-by conductors is appreciable, it is commonly found that in addition to the errors in bearing, there will be no position for which the loop output drops to zero. This arises when the spurious voltage induced in the loop contains a component 90° out of phase with the voltage acting around the loop. There is then no loop position for which the total voltage acting around the loop will be zero, and the minimum will be obscured, as in Fig. 247*c*.

Indistinct loop minima can be eliminated either by removing the offending objects, or by compensating for their induced voltages by stringing wires that will induce more or less equal, but opposite, voltages. Thus with a loop located between the foremast and smokestack of a steamship, the necessary compensation can be obtained by running one or more cables from mast to stack directly over the loop. By experimentally varying the position of such a cable, and in some cases by the use of several cables, obscure minima can ordinarily be eliminated.

A properly balanced, compensated, and calibrated loop antenna will give bearings accurate to better than one degree on radio signals from near-by transmitters operating at frequencies below 500 kc. The error becomes greater the higher the frequency and the greater the transmission distance, as the result of "night effect." Greater errors also occur when the bearings closely follow a coast line, or when the waves travel over mountainous land.

Night Effect.—A loop antenna will give correct bearings only when no horizontally polarized downcoming waves are present. The reason for this is that such waves induce voltages in the horizontal members of the loop that do not cancel out even when the plane of the loop is perpendicular to the bearing angle of the radio wave. The result is then either a false bearing, or an indistinct minimum, or both. This action is commonly termed *night effect* because at the frequencies ordinarily used in direction-finding work, *i.e.*, below 500 kc, the strength of the sky wave reflected to earth by the ionosphere, and hence the strength of the downcoming waves, is much greater at night than during the day. Night effect becomes more pronounced the higher the frequency and the greater the distance. This is because the ratio of sky-wave strength to ground-wave strength becomes higher as the frequency and distance are increased. As a consequence of night effect, it is found that at 375 kc, which is the frequency most commonly used in direction-finding work, the accuracy that can be obtained at night over sea is approximately 2° at distances up to 40 miles, and 5° up to 80 miles. During the day, *i.e.*, from about one hour after sunrise to one hour before sunset, bearings taken over sea at 375 kc are usually accurate to within 2 degrees, irrespective of distance.

126. Adcock Antenna.—The errors in bearing caused by downcoming horizontally polarized sky waves can be eliminated by replacing the loop antenna with an Adcock antenna, which in its simplest form consists of two spaced vertical antennas connected as shown in Fig. 250. The action of such an antenna, as far as vertically polarized waves are concerned, is identical with the loop. This is the case because the resultant current in the output of the Adcock antenna is proportional to the vector difference of the voltages induced in the two vertical members, exactly as is the case with the loop. Horizontally polarized downcoming waves do not affect the Adcock antenna, however, since the voltages induced in the two horizontal members are the same in phase as well as magnitude and so cancel out as a result of the circuit arrangements.

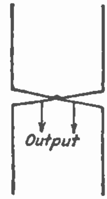


FIG. 250. Simple form of Adcock antenna.

By maintaining symmetry with respect to ground and by enclosure in an electrostatic shield, the Adcock antenna makes it possible to obtain accurate bearings to over a hundred miles at frequencies in excess of 5000 kc under conditions where a loop antenna is utterly useless. The practical value of the Adcock antenna is, however, limited by the fact that the energy pick-up is the same as that of a one-turn loop and so tends to be quite small.

127. Homing Devices.—A homing device is a form of direction finder that gives a visual indication of the orientation of the direction-finding equipment with respect to the direction of travel of a radio wave. A typical circuit arrangement is shown in Fig. 251. Here equal loop and vertical antenna outputs are combined to give a cardioid directional pattern, and the output of the receiver is rectified and passed through a zero-center direct-current galvanometer. The polarity of the loop output is continually reversed by means of a commutator so that the directional pattern is shifted back and forth between the solid and dotted cardioids of Fig. 251. A second commutator synchronous with the first reverses the terminals of the galvanometer in synchronism with reversals of the loop polarity, so that the net galvanometer current represents the difference between the receiver outputs for the two loop polarities. If the signals travel in a direction that is perpendicular to the plane of the loop, the receiver output is the same for both loop polarities. Equal d-c currents will then pass through the galvanometer alternately in opposite directions, and there will be no deflection. However, if the bearing of the signals is not perpendicular to the plane of the loop, the output will depend upon the loop polarity as seen in Fig. 251, and d-c current then flows through the galvanometer, causing a deflection to the right or left according to which side of the zero direction the signals arrive from.

Homing devices are particularly useful in guiding planes, ships, etc., to a base. Thus an airplane sent out from an aircraft carrier can be

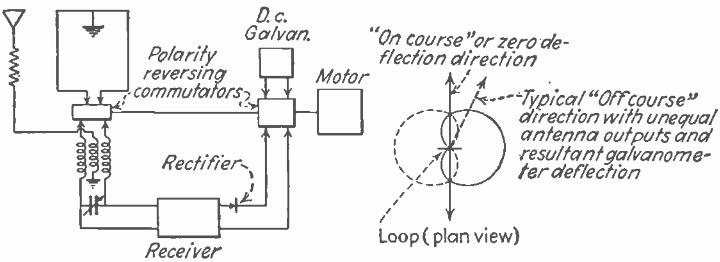


FIG. 251.—Circuit diagram of simple homing device, together with diagrams of the antenna directional patterns for the two loop polarities.

guided back to the ship by means of a homing device set to give "on-course" indication (zero galvanometer deflection) when the plane is headed in the direction of travel of the radio signals sent out from the carrier.

128. The Radio-range Beacon.—The radio range is a type of radio beacon which lays down a course in a predetermined direction. A typical arrangement consists of two crossed loop or Adcock antennas that are alternately excited from a common source of radio-frequency power. The directional characteristics of various arrangements of crossed antennas are shown in Fig. 252, where it is seen that in certain directions from the transmitter the signals from the two antennas are of equal strength. By interlocking the two antennas so that one of them is always radiating energy, and then sending out complementary signals such as N (— ·) and A (· —), the signal heard along the equisignal line is a continuous dash. In contrast with this, at points to the side of this equisignal course one or the other of the code characters dominates. The principal air routes of the United States are marked by this type of radio beacon.

The various courses of the radio range can be aligned to the actual routes followed by air travel. Methods of doing this include varying the angle between the crossed antennas, supplying unequal power to the two antenna systems, and adding a small amount of additional radiation from a vertical antenna, as illustrated in Fig. 252, and various combinations of these expedients.

The distance over which the beacon courses can be depended upon is limited by night effect, which causes the apparent course to shift. Errors in course indications can be minimized by using Adcock transmitting antennas instead of loops, since in this way the radiation of horizontally polarized waves is avoided. It is also desirable to employ a receiving antenna in which the horizontal portions are symmetrically arranged,

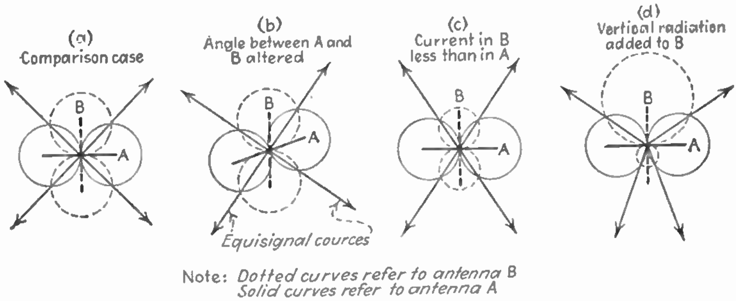


FIG. 252.—Equisignal courses from a radio range under various adjustments.

so that there will be no response to horizontally polarized waves. When these precautions are taken, radio-range course indications for frequencies around 250 kc are usually dependable up to about 100 miles both day and night.

Problems

1. *a.* Derive a formula for the effective height of a small loop antenna.

b. Calculate the effective height of a loop antenna 3 by 2 ft. with 20 turns, when the wave length is 1000 meters and $\theta = 0°$.

2. Calculate the resultant voltage acting around the loop in Prob. 1*b* when the strength of the radio signal is 10 mv per meter and $\theta = 0°$.

3. Discuss the factors that set a practical limit to the number of turns that can be used on a given-sized loop frame.

4. Describe the three-dimensional directional pattern of a loop that is in free space (remote from ground).

5. Explain why the zero-signal position of a loop is not affected by horizontally polarized waves traveling parallel to the ground, or by vertically polarized downcoming waves, but is affected by horizontally polarized downcoming waves.

6. When attempts are made to eliminate obscure minima by stringing conductors near the loop, it is sometimes found helpful to insert resistance, inductance, or capacitance in series with the cables. Explain why this is.

7. The magnitude of the "night-effect" errors obtained in direction-finding work is greater when the waves travel over land than over sea. Explain.

8. An airplane or ship, using a homing device to guide it to a base, can tell whether it is approaching toward or receding from the source of signals by intentionally deviating from the "on-course" direction and noting whether the indicator of the homing device deflects in the same or opposite direction from the ship deviation. ~~Explain why this is.~~

9. Demonstrate that the Adcock transmitting antenna is superior to the loop for radio-range work only when the ionosphere produces little change in the polarization of the sky wave, and that at high frequencies, where there is a large change in polarization produced by the ionosphere, the Adcock antenna has little if any advantage for transmitting.

10. Explain the experimentally observed fact that radio-range signals are more dependable over earth of good conductivity than over earth of poor conductivity.

CHAPTER XVI

TELEVISION

129. Elements of a System of Television.—Television is accomplished by systematically exploring the scene to be reproduced and transmitting at each instant a current that is proportional to the light intensity of the elementary area being explored at the moment. The scene is then synthesized at the receiving point by employing this transmitted current to control the light intensities of successive elementary areas corresponding to the elementary areas being explored at the transmitter. By carrying out this process with sufficient rapidity, the effect of a motion picture is obtained.

130. Scanning.—The process of exploring an image to obtain a current that varies with time in accordance with the light intensity of successive areas of the scene is called *scanning*. In all practical systems of television the portion of the picture viewed at any one time is limited by some form of aperture which has a height that is from $\frac{1}{120}$ to $\frac{1}{450}$ of the height of the image, and which is commonly square. This aperture is moved relative to the image along successive parallel slightly sloping paths that are spaced by an amount equal to the height of the aperture, as shown in Fig. 253. With such an arrangement the entire image is systematically covered by means of a series of lines, the edges of which just barely overlap.

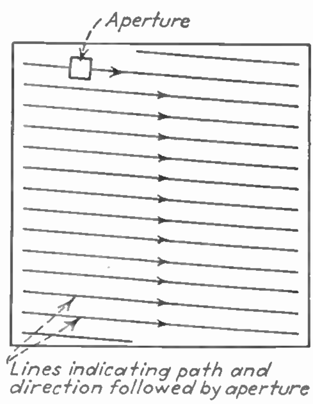


FIG. 253.—Lines indicating path followed by aperture in scanning a picture.

Many methods have been devised for carrying out the scanning operations. Of these, the two that have proven most successful for general television work are the "Image Dissector," developed by Farnsworth, and the "Iconoscope" of Zworykin.

The Image Dissector.—The principal features of this device are indicated in Fig. 254. Here C is a translucent cathode, the surface of which is coated with photosensitive material and upon which is focused the optical image of the scene to be transmitted. Each elementary area of the cathode surface accordingly gives a photoelectric emission of electrons proportional to the light intensity of that particular part of the picture. These emitted electrons are then attracted to the anode A, which is at a positive potential with respect to the cathode. By means of a suitable axial magnetic field produced by a direct current in the focusing coil, all electrons emitted from a given point on the cathode will converge together again at the anode plane, as illustrated in Fig. 254.

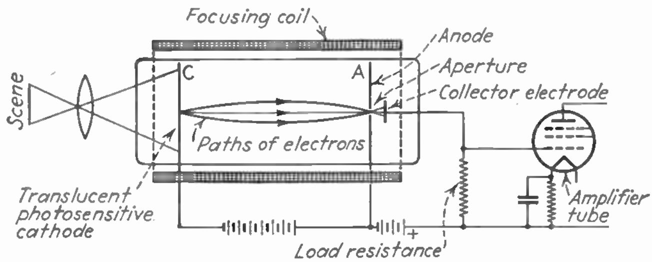


FIG. 254.—Essential features of the image dissector.

Under these conditions the distribution of electrons at the anode plane corresponds to the distribution of light intensity upon the cathode, thus giving at the anode what might be termed an *electron image* of the scene being reproduced.

The anode A is provided with a scanning aperture consisting of a small hole. The electrons passing through the aperture are collected by a collector electrode, which hence receives an electron current proportional to the light intensity of the corresponding part of the optical image. The picture is then scanned by displacing the electron image at the anode with respect to the aperture so that the part of the image that supplies electrons to the collector electrode is continuously changing in a systematic manner.

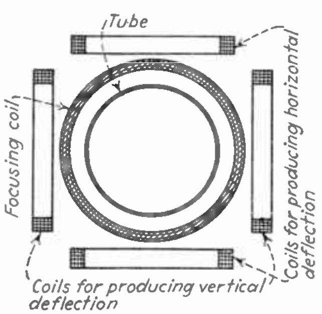


FIG. 255.—Coil system for producing magnetic deflection of electron image and for focusing. The view shown is a cross section through the middle of the tube of Fig. 254 at right angles to the axis.

This displacement of the electron image is accomplished by means of magnetic fields produced by the two pairs of coils shown in Fig. 255. Current passed through the top and bottom pair produces a magnetic field that will deflect the electron image to the right or left, and current passed through the two side coils will in the same manner deflect the electron image up or down. Scanning is accomplished by applying to the side coils a saw-toothed current wave, such as illustrated in Fig. 256, having a frequency equal to the number of times per second the scene is to be scanned, while the second pair of coils is supplied with a similar saw-toothed wave of current having a frequency equal to the number of lines per picture multiplied by the number of pictures per second. The result is to move the electron image with respect to the aperture in the manner indicated in Fig. 253.

The output of the image-dissector tube can be increased greatly by using secondary electron emission to multiply the electrons passing through the anode aperture. This is accomplished by causing the electrons that pass through the aperture to strike a surface that is especially prepared to give secondary electron emission. The secondary electrons thus produced are used to give more secondary electrons, which in turn can be made to produce still more, and so on. Since five to ten secondaries can be produced for each electron striking a sensitized surface, it is apparent that the output of the image dissector can be readily increased a thousandfold, or even more, in this way. Details of practical electron multipliers are to be found in the literature.[1]

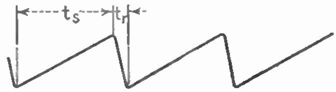


FIG. 256.—Saw-tooth wave used for scanning purposes.

The Iconoscope.—The iconoscope method of scanning can be explained by reference to Fig. 257. The optical image of the scene under consideration is focused upon the screen P consisting of a mica or other insulating plate, upon the illuminated side of which is a mosaic surface composed

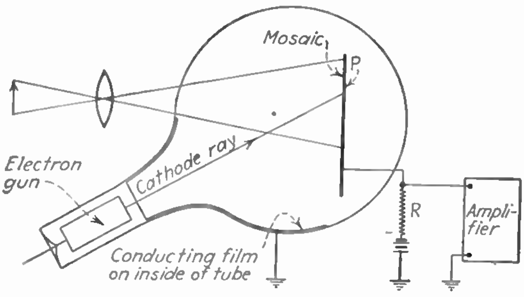


FIG. 257.—Essential features of the iconoscope.

of minute isolated globules that are photosensitive. The manufacturing process used in producing this surface is such that the individual globules are actually insulated from each other. The back side of the mica or insulating plate has a metal coating that serves as the output electrode, as shown. The optical image focused upon the mosaic photosensitive

[1] Thus see P. T. Farnsworth, Television by Electron Image Scanning, *Jour. Franklin Inst.*, vol. 218, p. 411, October, 1934; V. K. Zworykin, The Secondary Emission Multiplier—A New Electronic Device, *Proc. I.R.E.*, vol. 24, p. 351, March, 1936.

surface is scanned by means of a cathode-ray beam that is produced by an electron gun similar to that used in ordinary cathode-ray oscillograph tubes. This cathode-ray beam scans the mosaic surface by means of vertical and horizontal deflecting fields actuated by saw-toothed voltage waves similar to those used in the dissector described above.

The operation of the iconoscope depends upon the fact that each globule of the mosaic surface forms a small condenser with respect to the back plate. When light falls upon a globule, electrons are lost through photoelectric emission, and the globule becomes positively charged with respect to the back plate. This positive charge builds up at a rate proportional to the light intensity. When the cathode-ray beam strikes the globule, this positive charge is neutralized by the electrons in the cathode-ray beam, so that the cathode-ray spot can be said to discharge the globule. At the instant of discharge there is a rush of current through the resistance R proportional to the positive charge accumulated upon the globule, and hence proportional to the illumination upon the globule. It is to be noted that the current passing through the output resistance R is relatively large because the number of electrons in the discharge is equal to the total photoelectric emission since the previous scanning. This means that the total emission over one picture frame (commonly $\frac{1}{30}$ sec.) is utilized, instead of only the photoelectric emission during the instant the spot is being scanned. Although the storage action is not 100 per cent efficient, it nevertheless greatly increases the output.

131. Reproduction of Television Images at the Receiver.—All modern systems of television employ a cathode-ray tube to reproduce the picture at the receiver.[1] The intensity of the luminous spot is controlled by the strength of the incoming signals, while the position is controlled by saw-toothed deflecting waves similar to those employed for scanning at the transmitter. By synchronizing the receiver deflecting waves with the corresponding waves at the transmitter by means described below, the cathode-ray spot on the frame of the received picture will be in the same relative position as the picture element being scanned at the transmitter. The result is then the synthesis at the receiver of the original image.

132. Miscellaneous. *Scanning Waves.*—The saw-toothed scanning waves of Fig. 256 must be substantially linear during the period t_s in order for the scanning to take place properly. During the return period t_r the exact shape is not important, but the ratio of return time to scanning time t_r/t_s should be small in order to avoid lost time. The nature of the saw-toothed generator must also be such that the frequency can be readily controlled by means of pulses injected from an external source.

[1] This cathode-ray tube is often given a special name such as *oscillite*, *kinescope*, etc., when used in television, but it is still only a cathode-ray tube.

With electrostatic deflection, a saw-toothed voltage wave is required. This can be generated in a number of ways, a typical arrangement being shown in Fig. 258*a*. Here *V* is a tube that is biased beyond cut-off except during moments when a synchronizing pulse is acting on the grid. The operation can be understood by starting at the moment just after a synchronizing pulse has caused the discharge of the condenser *C*. As soon as the pulse is removed, the tube becomes non-conducting and the condenser starts to charge through the series resistance *R*. If the circuit proportions are such that the voltage across the condenser is always much

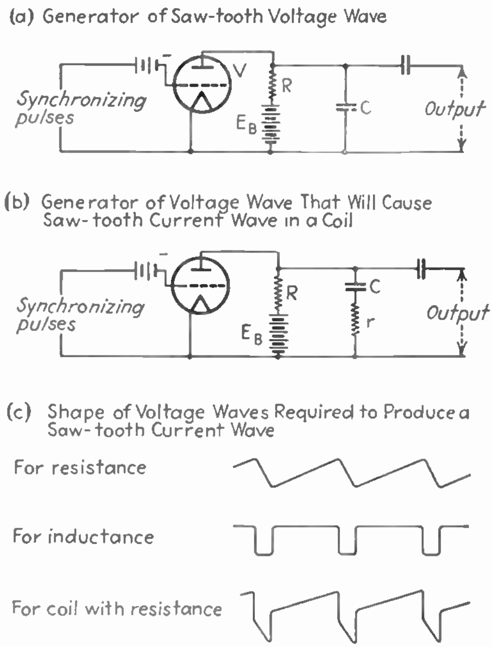


FIG. 258.—Circuit for generating a saw-tooth wave, together with voltage wave shapes required for a saw-tooth current wave through a resistance, an inductance, and a circuit with both resistance and inductance.

smaller than the supply voltage, the resulting condenser voltage will build up linearly, giving a substantially straight-line saw tooth corresponding to the portion t_s in Fig. 256. This process continues until the next synchronizing pulse arrives, when the tube momentarily passes plate current, discharging the condenser *C* rapidly and giving the return part t_r of the cycle in Fig. 256. After the passing of the synchronizing pulse, the tube again becomes non-conducting and the cycle is repeated. It will be noted that the result is a saw-toothed voltage wave having a frequency controlled by the external synchronizing pulses.

When magnetic deflection is employed for scanning, it is necessary to have a saw-toothed wave of current rather than a saw-toothed voltage wave. If the coil were a pure inductance, the required shape of the voltage wave would be a rectangular pulse having a duration equal to the return time, as shown in the second line of Fig. 258*c*. On the other hand, if the coil were predominantly resistive, the voltage wave would have the same shape as the current wave and so would be saw-toothed as in the first line of Fig. 258*c*. In the actual case, where both resistance and inductance are present, it is necessary to combine the wave shapes of resistance and inductance in the proper proportion, such as shown in the last line of Fig. 258*c*, in order to get a voltage that will produce a saw-toothed wave of current. Such a voltage wave can be obtained by rearranging the circuit of Fig. 258*a* as in Fig. 258*b*. The voltage across the condenser C in this arrangement has the same shape as in Fig. 258*a*, *i.e.*, is a saw tooth suitable for overcoming the resistance drop of the deflecting coil. The voltage across r, however, is in the form of a pulse, as shown in the second part of Fig. 258*c*, provided $r \ll R$. This is because, when the condenser discharges, there is a sudden rush of current through r, while during the charging period the current through r is small and substantially constant.

Synchronization.—In television it is necessary that perfect synchronization be maintained between positions on the image at transmitter and receiver at all times. The most satisfactory method of doing this consists in generating pulses at the transmitter which control the saw-toothed scanning waves used in the pick-up device, and which at the same time are transmitted to the receiver and there used to control the scanning of the reproducing equipment. The required pulses can be generated in a number of ways, as for example by use of a rotating disk provided with holes that allow a light to shine intermittently upon a photoelectric cell.

The usual procedure is to indicate the end of each scanning line by a short synchronizing pulse, while the end of each picture frame is marked by a much longer pulse. These pulses occupy 5 to 15 per cent of the total time, and are separated at the receiver by their difference in length.

Exact synchronization between transmitter and receiver is obtained by superimposing the synchronizing pulses upon the signal currents obtained from the pick-up tube and transmitting both to the receiver together. In order that the signal currents will not interfere with the synchronizing action, the polarity of the transmitted synchronizing pulses is made to correspond to the polarity of black, and the amplitude of the pulse is made appreciably greater than the amplitude of the signal current for a black portion of the image. In this way the synchronizing pulses are always larger than the maximum possible signal current. In order to prevent overloading of the transmitting equipment, it is common practice to pass the combined signal and synchronizing currents through a limiting amplifier which limits the peak amplitude to perhaps 50 per cent more than black. The resulting waves as transmitted are shown in Fig. 259.

By transmitting the synchronizing pulses with a polarity corresponding to that of black in the reproduced picture, and at the same time by making these pulses have a duration equal to or slightly greater than the time required for the saw-toothed scanning wave to return from the end to the beginning of its linear portion, the cathode-ray spot is blacked out during the return period.

At the receiver the synchronizing pulses are separated from the signal currents by taking advantage of the difference in amplitude. This can be conveniently done by biasing an amplifier tube sufficiently far beyond cut-off so that only the synchronizing pulses have sufficient amplitude to cause plate current to flow. The resulting pulses free of

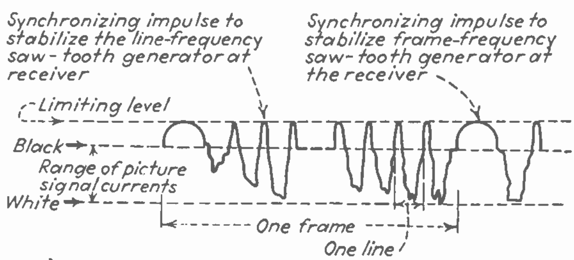


FIG. 259.—Typical television signal showing superimposed synchronizing pulses.

signal are then applied to a network, which separates the line and frame pulses by utilizing their difference in duration.

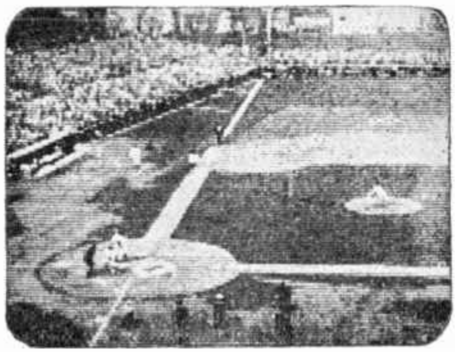
Frequency Band and Picture Detail.—The detail that can be reproduced by a television picture is determined by the number of scanning lines. Studies show that it is necessary to have at least 180 lines to reproduce a scene with any satisfaction, that 240 lines give a fair picture, and that 480 lines are practically equivalent to a home movie. At the present time it is considered that commercial television must have at least 240 lines, and there is a tendency to standardize on 350 to 450 lines. The effect of the number of lines upon the detail is shown in Fig. 260.

The frequency band required to transmit a picture having appreciable detail is very great. The worst case is where the image to be transmitted is a checkerboard pattern of alternate dark and light squares, with the side of each square equal to the width of the scanning line. In scanning such a picture the output current of the pick-up device will be alternately large and small, so that one cycle of output current corresponds to traversing two squares of the pattern. If the number of lines in the picture is a, and the ratio of width to height is R (= aspect ratio), then the total number of squares is a^2R, so that for n pictures per second the frequency band required would be $a^2nR/2$ cycles. With ordinary scenes it is found that this gives more detail in the direction of the scanning lines than in the vertical direction. Experience shows the detail is equal in the two directions when the frequency band is only about 70 per cent as great, so

$$\text{Actual frequency band} = 0.35a^2nR \tag{136}$$

The resulting frequency band required to transmit a television image having good definition is very high. Thus a 441-line picture, repeated

(*a*) 120 scanning lines



(*b*) 240 scanning lines

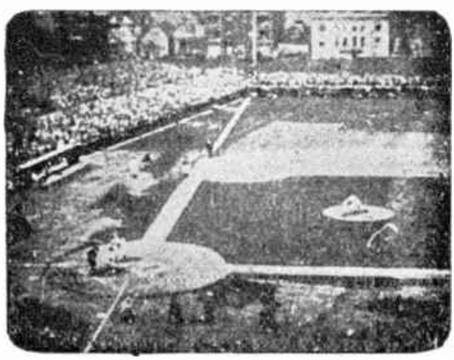


FIG. 260.—Typical television images showing effect of the number of scanning lines.

30 times per second and having an aspect ratio of $\frac{4}{3}$, requires a band at least 2,720,000 cycles wide. If 10 per cent is allowed for the synchronizing pulses, the actual frequency band that must be provided is 3,020,000 cycles. When this is modulated upon a carrier, the total width of the two side bands is thus over 6,000,000 cycles.

As a result of these side-band requirements, it is apparent that television pictures can be transmitted by wire only over special circuits. When radio transmission is to be employed, it is necessary that the carrier frequency be extremely high, such as 40 mc or more, so that the side-band width will not be too high a percentage of the carrier frequency.

Aperture Distortion.—The finite size of the aperture at the transmitter introduces distortion by making it impossible to transmit details finer than the area that the aperture represents.[1] This can be made clear by considering what happens when a simple image such as that of Fig. 261*a* is scanned. As the aperture travels from left to right, it is seen that the average light intensity of the area enclosed by the aperture varies gradually instead of suddenly from light to dark, thus reproducing a distorted pattern as shown at Fig. 261*b*. This effect is equivalent to discriminating against the higher frequency components in the signal.

Video-frequency Amplifier.—The signal currents obtained from the pick-up tube of the television transmitter are commonly termed video-frequency currents to distinguish them from audio- and radio-

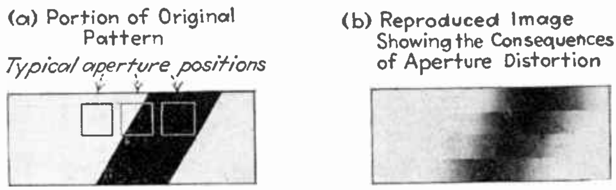


FIG. 261.—Simple image, as scanned at the transmitter, and distorted image received as a result of aperture distortion.

frequency currents. Video-frequency amplifiers for increasing the output of a pick-up tube to a level suitable for modulating a radio transmitter, and for amplifying the detector output of the receiver up to a level suitable for operating the cathode-ray tube, must handle a very wide frequency range with negligible frequency and phase distortion. The response at low frequencies must extend well below the frequency corresponding to the number of pictures per second, and the highest frequency is given by Eq. (136). Over this range the response should be substantially uniform and should be accompanied by negligible phase distortion. Phase distortion is particularly important because it is equivalent to different speeds of transmission for different frequencies, and so it obviously distorts the picture by causing certain parts of the detail to arrive either too early or too late. This is in contrast with audio-frequency work, where, because of the characteristics of the ear, the amount of phase distortion produced by ordinary amplifier circuits has negligible effect.

Flicker, Power-line Ripple, and Interlaced Scanning.—If the rate of repetition of the television image is not sufficiently high, there will be a

[1] In the case of the iconoscope, the size of the cathode-ray spot scanning the mosaic plate is the effective size of the aperture.

pronounced flicker even though the rate of repetition may be sufficient to convey movement satisfactorily. Studies indicate that the lowest repetition rate at which the flicker is not objectionable is approximately 48 times per second, which is considerably greater than the rate required to convey motion.[1]

In selecting the frame frequency it is also necessary to consider the effect of power frequency currents in the deflecting and anode circuits. If the frame frequency is not an exact submultiple of the power frequency, any stray current of power frequency in the deflecting, anode, or control circuits will cause ripples of various sorts to travel across the field with very annoying psychological effects. If the frame frequency is a submultiple of the power-line frequency, such as 15, 20, 30, or 60 for a 60-cycle power system, these disturbances, while still present, are stationary and so are much less noticeable.

Since the usual power-line frequency is 60 cycles, and since 60 has no submultiple between 30 and 60, it is therefore necessary that the repetition rate of the television picture be 60 times per second in order to avoid excessive flicker and power-line ripple. This is a relatively high frame frequency, at least three times that required to convey continuous motion, and so is undesirable because of the excessively wide frequency band that results. A fairly satisfactory solution for this situation is to scan the image 60 times per second, but to scan only alternate lines during one scanning period. Thus with a 240-line picture one would scan the odd-numbered lines during the first $\frac{1}{60}$ sec. and then scan the even-numbered lines during the next $\frac{1}{60}$ sec. As far as flicker is concerned the result is equivalent to scanning at the rate of 60 times per second, while from the point of view of detail and frequency band required, the result is substantially equivalent to scanning a 240-line picture 30 times a second. An arrangement of this sort is known as *interlaced scanning*, and is becoming recognized as the best compromise between low flicker and the narrowest possible frequency band. The chief disadvantage is the relatively complicated synchronizing system that is required to interlace the lines accurately.

133. Typical Complete System of Television.—A television transmitter consists of a pick-up tube (iconoscope or image dissector), a saw-toothed-wave generator with frequency controlled from a pulse generator, a video-frequency amplifier, and an ultra-high-frequency radio transmitter upon which the output of the video-frequency amplifier is modulated. The amplifier circuits, both video and radio frequency, must be arranged to transmit the wide frequency bands involved without appre-

[1] In motion-picture work where there are 24 frames per second, flicker is avoided by cutting off the light momentarily during the middle period of each frame so that the light flashes on the screen twice for each picture, or 48 times per second.

ciable frequency or phase distortion. In many respects this layout is similar to that of an ordinary broadcast transmitter, with the pick-up tube taking the place of the microphone and the video-frequency amplifier corresponding to the broadcast audio-frequency system. A block diagram of such a transmitter is shown in Fig. 262*a*.

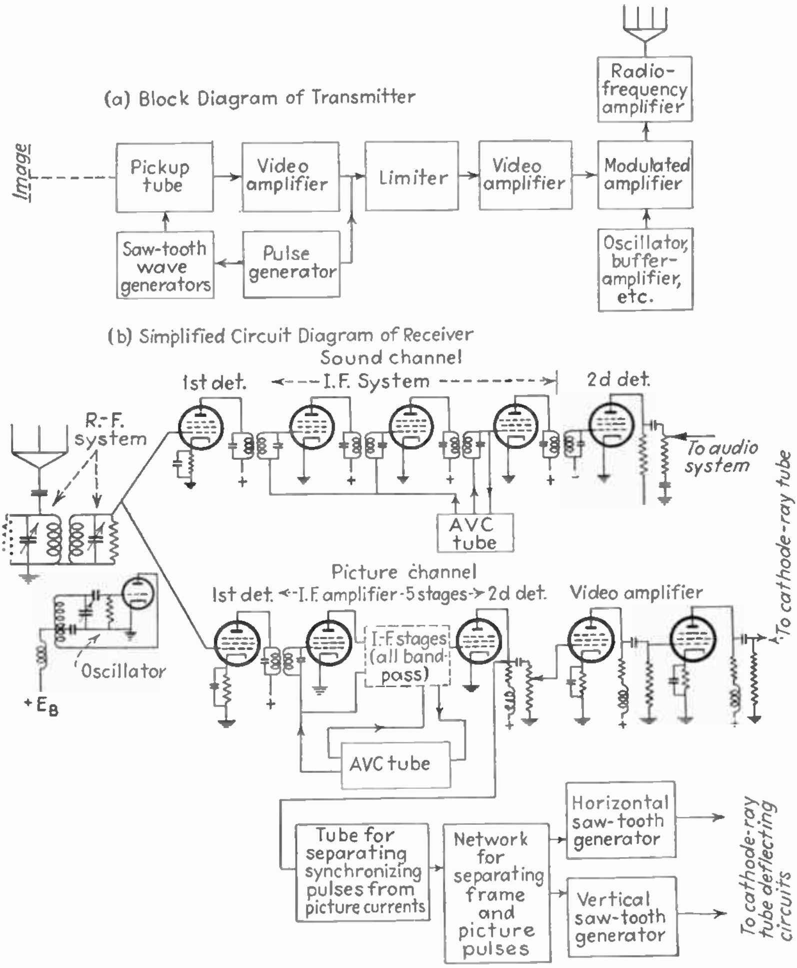


FIG. 262.—Diagram of complete television system.

A typical television receiving system is shown in Fig. 262*b*. Consider for the moment only the picture channel. This consists of a superheterodyne receiver with an intermediate frequency of perhaps 10 mc, with band-pass circuits designed to give a flat response over the necessary frequency range. The intermediate-frequency amplifier is followed by the usual second detector, and the resulting video-frequency currents, together with the accompanying synchronizing pulses, are then amplified by the video amplifier to a level suitable for controlling the spot intensity of the cathode-ray reproducing tube. Synchronizing pulses free of signal are obtained by also applying the output of the second detector to an amplifier biased so that only the pulses have enough amplitude to develop output. The pulses corresponding to the line frequency are then separated from the frame frequency by a suitable network and used to control the frequency of local saw-toothed generators that produce the currents for deflecting the spot of the cathode-ray reproducing tube.

The sound accompaniment for a television picture is most conveniently transmitted by modulating an auxiliary carrier frequency that is just outside the side-band range of the television signals. The antenna circuits and the local oscillator of the receiver can then be common to both sound and picture channels, although the sound channel requires a separate first detector and intermediate-frequency amplifier, as shown in Fig. 262*b*. The intermediate-frequency amplifier for the sound channel should have a relatively narrow response band centered about a frequency differing from the mid-frequency of the picture-channel intermediate-frequency-amplifier response band by the difference between sound and picture carriers. In this way the local oscillator automatically produces the correct beat frequency for the picture channel when the adjustment is such as to bring in the sound channel.

Problems

1. The image dissector is normally operated with an anode voltage that draws the electrons away from the cathode as fast as they are emitted. What would be the effect of reducing the anode voltage to a value so low that a space charge formed in front of the cathode?

2. In the image dissector the need for focusing arises because some of the emitted electrons have a velocity component in a direction parallel with the cathode surface. Explain why this would cause the reproduced image to be blurred if no provision for focusing were made.

3. In the iconoscope discuss the effect on the reproduced television picture of (*a*) leakage between adjacent globules and (*b*) leakage from *each* globule to the back plate but no leakage between globules.

4. If the storage process of the iconoscope were 100 per cent efficient, derive a formula giving the improvement in sensitivity over the output obtained by a simple aperture (as the image dissector without electron multiplication).

5. Explain why it would be undesirable to use an ordinary sine wave for scanning instead of the saw-toothed waves always employed.

6. *a*. What would be the effect on the reproduced image of an improper adjustment of the receiver saw-tooth oscillator for horizontal deflection such that the portion t_s in Fig. 256 was curved instead of linear?

b. What would be the effect if both sending and receiving saw-tooth oscillators for horizontal deflection had the same non-linearity mentioned in (*a*)?

7. Explain why it is not permissible to send the synchronizing pulses with a polarity corresponding to white.

8. Calculate the frequency band required to transmit pictures having 120, 240, and 480 lines, with a frame frequency of 30 and an aspect ratio of 4:3, when the video current is modulated upon a radio-frequency carrier. Allow for synchronizing pulses lasting 10 per cent of the time.

9. Sketch the shape of the output wave of the pick-up tube when one line in the pattern of Fig. 261*a* is scanned, assuming: (*a*) no aperture distortion and (*b*) normal aperture distortion.

10. Assuming that the human eye can just distinguish two objects that are separated by 2′ of arc calculate the frequency band required for "perfect reproduction" of a square television image each side of which subtends an angle of 30 deg., corresponding to the size of the screen when viewed from the average seat in a movie theatre. Assume 30 complete pictures per second.

11. When portraying rapid motion, an interlaced scanning system introduces a peculiar kind of distortion not present with a simple scanning system. Describe the effect.

CHAPTER XVII

ACOUSTICS

134. Characteristics of Audible Sounds.—The frequency range of importance in the reproduction of sound is from about 40 to approximately 16,000 cycles. Individual sounds differ greatly in the portion of this range utilized, however, as illustrated by Fig. 263. Here it is seen, for example, that a frequency range 100 to 8000 cycles is adequate

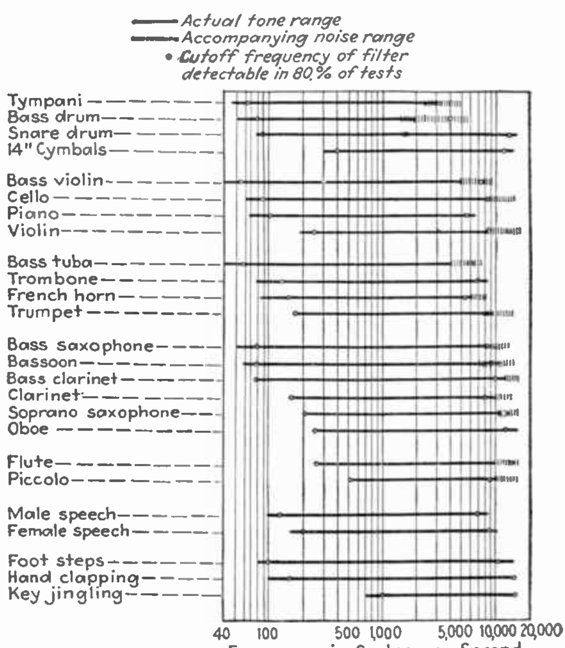


FIG. 263.—Frequency range of representative musical instruments as determined by listening tests (as given by W. B. Snow).

for the perfect reproduction of a male voice, whereas a piccolo utilizes frequencies from 500 to 16,000 cycles.

The power involved in sounds varies greatly. Thus, although speech power during ordinary conversation averages about 10 μw, peak speech powers may reach 5000 μw, and a very faint whisper may fall to 0.01 μw. This represents an intensity range of 500,000:1. Music covers an even wider power range. A large orchestra may, for example, develop a peak power of 100 watts, and power ranges of 10,000,000:1 between the loudest and softest passages are sometimes encountered.

The distribution of sound power with frequency depends upon the sound involved. With speech the most powerful sounds are distributed more or less uniformly in the frequency range 500 to 1500 cycles, although the average power is greatest in the vicinity of 500 cycles. In music the distribution of power depends upon the instrument, with most instruments constructed so as to produce more sound power at frequencies below 500 to 1000 cycles than at the higher frequencies. Noises have their energy distributed more or less uniformly over wide frequency ranges, a single noise sometimes involving all frequencies from 50 to 15,000 cycles.

135. Characteristics of the Human Ear.—The properties of the ear are of fundamental importance in sound work since it is through the medium of the ear that sound waves are observed. The frequency and amplitude range over which a normal ear receives auditory sensations is illustrated in Fig. 264. Frequencies below about 20 cycles are perceived by feeling rather than hearing, and frequencies above about 20,000 cycles are not heard by most ears. It is apparent that the sensitivity of the ear depends upon the frequency and is maximum in the range 1000 to 3000 cycles.

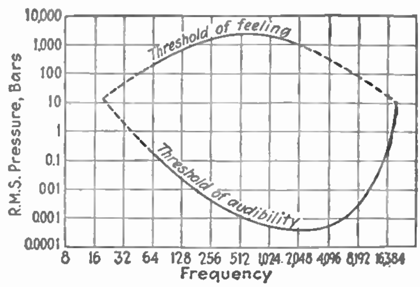


FIG. 264.—Average auditory sensation area of normal ears. The curve marked "Threshold of audibility" represents the weakest audible sound, and the curve "Threshold of feeling" represents the loudest sound that does not produce pain.

The smallest variation in sound amplitude that the ear is able to perceive is roughly a constant percentage of the original intensity. The minimum percentage of change in sound energy that is detectable is of the order of 25 per cent (1 db) in the middle range of frequencies at moderate intensities, and becomes less when the intensity is low or when the frequency is high or low.

The ear has a non-linear response to sound waves of large amplitude. The result is that with powerful sound waves the ear produces harmonics, as well as sum and difference tones, which are not present in the original sound, and yet which are actually present in the hearing organs and are perceived by the brain. Such frequencies produced in the ear are called *subjective tones* and explain a number of sound phenomena. Thus the pitch of a sound is not changed by removing the fundamental frequency since the harmonics combine in the ear to produce a difference frequency that recreates the fundamental component in the form of a subjective tone. This non-linear character of the ear also makes many radio receivers and loud-speakers at least passably acceptable by regenerating in the form of subjective tones the low frequencies that the equipment itself fails to reproduce.

Another important consequence of the non-linear character of the ear is the phenomenon known as *masking*, which appears as a deafening to high-frequency sounds caused by the presence of a lower pitched sound. Masking arises from the fact that, when the ear produces harmonics of the low frequencies, these harmonics interfere with the perception of the higher pitched sounds, which are then said to be masked. Masking is particularly important in noisy locations, since it is equivalent to a deafening. This makes it necessary to raise the voice when carrying on a conversation in a noisy location.

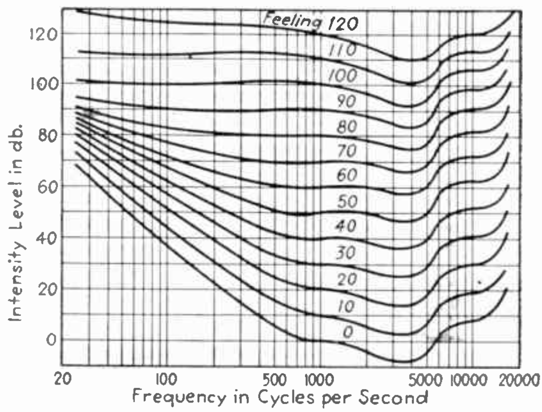


FIG. 265.—Contours giving intensity (*i.e.*, relative sound power) required for equal loudness as a function of frequency, for pure tones. The numbers on the contours give the strength of an equally loud 1000-cycle tone in decibels above the minimum audible 1000-cycle tone.

Loudness.—The magnitude of the auditory sensation produced by a sound is termed its *loudness.* Experiment shows that the relative loudness of different frequencies depends upon the absolute as well as upon the relative intensity. This is shown by Fig. 265, which presents experimental curves giving intensity of the sound (*i.e.*, sound power) required for equal loudness as a function of frequency, for the case of pure tones. It is seen from these curves that, when the intensity is lowered, as, for example, when a radio is played softly, the lower frequencies tend to disappear. Thus, if 100- and 1000-cycle tones both have intensities corresponding to the 70-db contour in Fig. 265, reducing the intensity of both by 40 db will reduce the loudness of the 1000-cycle tone to the 30-db contour, whereas the 100-cycle tone will be just at the limit of audibility and so will have disappeared.

136. Acoustics of Buildings.—The sound reaching an observer will generally differ from the sound as generated because of reflections from near-by objects. Consider, for example, the situation illustrated in Fig. 266, which shows only a few of the paths by which sound produced in a room may travel from source to observer. The direct route involving no reflections is the most important individual path, but, unless the observer is very close to the source, or unless the walls are lined with sound-absorbing material, large amounts of sound energy will reach the observer by way of the longer indirect paths involving reflections from the bounding surfaces.

The principal effects that these reflections have on the sound are as follows:

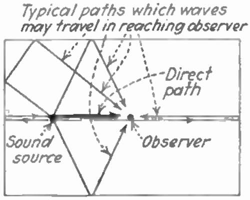


FIG. 266.—Diagram illustrating a few of the many routes that sound produced in a room may travel in reaching a listener.

1. The average intensity of the observed sound is raised because sound originally sent out in other directions is reflected back to the observer.
2. The relative amplitudes of the different frequency components of the sound may be altered as a result of selective absorption at the reflecting surfaces, which usually tend to reflect low frequencies more efficiently than high frequencies.
3. The relative amplitude of the different frequency components of the sound will always be altered as a result of interference effects resulting from the fact that the phase with which the energy traveling along the different possible paths combines depends upon the position of the observer and upon the frequency.
4. The observed sound persists for some time after the original sound has ceased as a result of the greater time it takes the sound traveling along the indirect routes to reach the observer. This effect is known as *reverberation.*

The magnitudes of the first three of these effects depend primarily upon that fraction of the total energy reaching the observing point which has traveled an indirect path, and this in turn is determined by the relative lengths of the direct and the indirect paths and the fraction of the sound-wave energy that is absorbed upon reflection. When the direct path is short, or when the bounding surfaces are of such a character as to absorb a large fraction of the energy of sound waves striking them, most of the energy reaching the observer travels along the direct path, and interference, selective absorption, etc., are not important.

Reverberation.—Reverberation in a room depends upon the ratio of the enclosed volume to the area of the bounding surfaces, and upon the average coefficient of sound absorption of these surfaces. Reverberation is measured in terms of the time after the sound source has been silenced that it takes a sound uniformly distributed throughout the room and having 1,000,000 times the minimum audible energy to die down to inaudibility. The reverberation time is commonly of the order of several seconds in large theaters and auditoriums where the sound travels long distances between reflections. In living rooms of ordinary homes the fact that the sound waves have opportunity to travel only a short distance between reflections results in rapid absorption and hence a very short reverberation time.

It might be thought that the smaller the reverberation time the better, but this is not necessarily true, because the ear normally expects a certain amount of reverberation, and because reverberation enhances certain musical and oratorical effects. Thus with music the presence of reverberation helps the different players in an orchestra to play together properly and increase the effectiveness of passages intended to convey the impression of power. Similarly, in speech the presence of reverberation increases the average intensity of the sound above the level that would be present outdoors or in a room with perfectly absorbing walls, and thereby makes it possible for a speaker's voice to reach a larger number of people.

The optimum reverberation time for a broadcast studio varies with different conditions, but is always less than for the corresponding theater or auditorium. This is because the ultimate listener receives reverberation from both the broadcast studio and the room in which the sound is reproduced.

The reverberation time can be controlled by the use of sound absorbents, and studios, auditoriums, etc., must be acoustically treated so that the optimum reverberation time is obtained. This is done by use of acoustic tiles and plasters on walls and ceiling to provide a certain amount of absorption, and by adding rugs, drapes, etc., when circumstances call for still less reverberation.

137. Effects of Distortion in the Reproduction of Sound.—In designing a system for the electrical reproduction of sound it is necessary to consider the various ways in which the original sound may be distorted and the consequences of this distortion. The most important single factor involved is the frequency range of the electrical and acoustical systems. The frequency band required for substantially perfect reproduction varies with the nature of the sound involved, as is apparent from Fig. 263, but is about 100 to 10,000 cycles for speech and 60 to 15,000 cycles for music. Experience indicates, however, that a frequency range of 80 to 8000 cycles is adequate to give excellent reproduction of all sounds except certain types of noises. The usual broadcast program has a frequency range of the order of 100 to 5000 cycles, and understandable although not natural speech can be transmitted with a much narrower range of frequencies, as illustrated by the telephone, which has a frequency range of 250 to 3000 cycles.

Experimental investigations of the effect of amplitude distortion, such as introduced by an overloaded amplifier, indicate that, when the full frequency range is reproduced, distortions of 3 to 5 per cent are detectable, and that 10 per cent is very noticeable.[1] Odd harmonics are more troublesome than even harmonics, and the amount of distortion permissible becomes less as a wider frequency range is reproduced.

The loudness with which the sound is reproduced is also of considerable importance. If the reproduced sound is at lower intensity than the original, the low frequencies appear to be weaker than they should, as explained in connection with Fig. 265, and the tendency of the high frequencies to be masked by the lower frequencies is reduced. Both these effects work in the same direction and combine to make it necessary to vary the frequency response with volume level if high-quality reproduction is to be obtained. The "tone-compensated" volume control used in many radio receivers is for the purpose of providing this type of correction.

138. Dynamic Loud-speakers Employing Paper Cones.—Nearly all loud-speakers employed in radio receivers are of the *dynamic* type, having the construction illustrated schematically in Fig. 267. Here a coil, commonly called the *voice coil,* is fastened to the apex of a paper cone. This coil is located in a magnetic field as shown and carries the audio-frequency currents that are to be transformed into sound waves. In such an arrangement the action of the magnetic field on the coil current produces a mechanical force that vibrates the paper cone and causes the radiation of sound waves. The cone must be mounted in a baffle, as shown in Fig. 267, or provided with other means for preventing the radiation from the front and back sides from canceling at low frequencies.

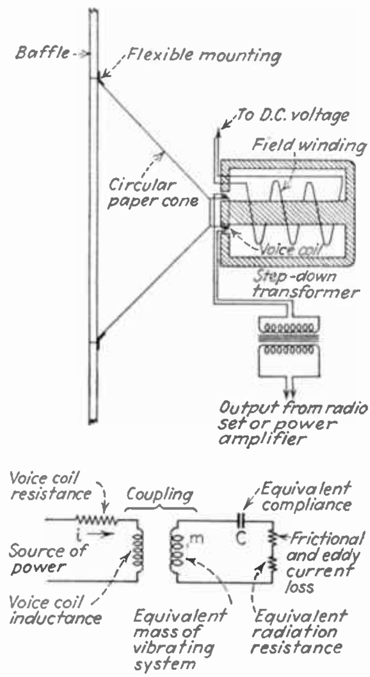


FIG. 267.—Cross section of typical dynamic type of loud-speaker and equivalent electrical circuit of electromechanical system.

The characteristics of a typical high-grade dynamic speaker are shown in Fig. 268. The frequency range is adequate for ordinary broadcast receivers, and the response is sufficiently uniform over this frequency range to sound satisfactory upon listening tests.

[1] The percentages given are the measured percentages of harmonic that result when a sine wave having an amplitude equal to the peak amplitude of the complex audio wave is applied.

Analysis of Dynamic-speaker Action.—The force that is exerted upon the voice coil as a result of the action of the magnetic field upon the voice-coil currents acts against a mechanical impedance consisting of a mass, an elastance, and a resistance. The mass consists of the effective mass that the coil-and-cone assembly offers to the frequency involved plus the fluid mass caused by the air in contact with the cone. The equivalent elastance is determined primarily by the spider supporting the voice coil and the elastic properties of the cone but is also influenced by closed air spaces, etc. The effective resistance to motion includes sound energy that is radiated, eddy-current losses, etc. The velocity of the resulting voice-coil vibration is proportional to the force divided by the mechanical

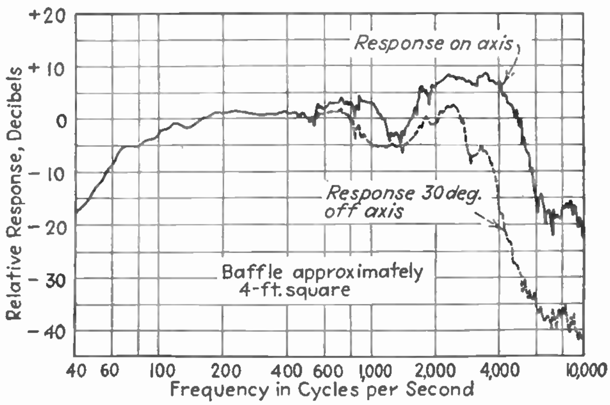


FIG. 268.—Response curve of typical cone dynamic speaker.

impedance, and the amount of sound energy radiated is proportional to the square of the velocity times the radiation component of the resistance.

The vibrating voice coil cuts across the direct-current magnetic field of the loud-speaker and so has a back voltage induced in it. This causes a voltage drop that is equivalent to adding an impedance to the voice-coil circuit, and, since this additional impedance is caused by vibration, it is termed *motional impedance.* The real and reactive energies represented by the voice-coil current flowing through the voice-coil motional impedance are the real and reactive energies, respectively, which the electrical circuits deliver to the vibrating system in order to sustain the vibrations.

A study of the electromechanical relations existing in a dynamic speaker shows that as far as the electrical circuits are concerned, the mechanically vibrating system is equivalent to a tuned secondary. This leads to the equivalent electrical circuit for the dynamic loud-speaker shown in Fig. 267, in which the mechanical force acting on the voice coil is represented by the voltage induced in the secondary, and the motional impedance is the coupled impedance. The velocity of vibration is equivalent to the secondary current in the electrical analogue.

At low frequencies (commonly up to 500 to 1000 cycles) the paper cone acts approximately as a piston diaphragm having a diameter equal to the diameter of the cone. Under these conditions the presence of the air in contact with the vibrating diaphragm produces a mechanical radiation resistance R_a per unit area, which varies with frequency as shown in Fig. 269. For frequencies low enough so that the diameter of the cone is appreciably less than a half wave length it is seen that the radiation resistance is inversely proportional to the square of the frequency. In order for the total sound power to be independent of frequency it is

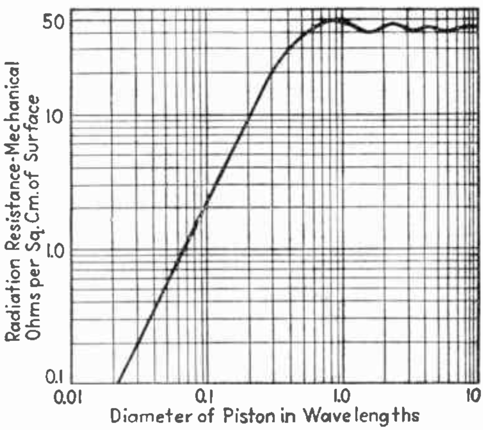


FIG. 269.—Radiation resistance of piston radiator per square centimeter of surface. (When both sides of piston radiate, the total surface is twice the area of one side.)

then necessary that the velocity be inversely proportional to frequency. This result can be readily realized by making the resonant frequency of the cone-and-coil assembly less than the lowest frequency to be reproduced. The principal impedance to motion is then supplied by the inertia of the coil mass and is proportional to frequency.

At frequencies high enough for the diameter of the cone to exceed a half wave length, the radiation resistance becomes substantially independent of frequency. With piston action and a velocity inversely proportional to frequency, the radiated power is then inversely proportional to the square of the frequency. This tendency for the response to fall off at high frequencies can be counteracted by so designing the paper cone that it ceases to operate as a piston at these higher frequencies. The cone vibrations are then in the form of waves traveling outward from the apex, with the result that the center part of the cone vibrates much more intensely than the outer edges. As the frequency is increased, the action is therefore very much as though the size of the cone were progressively decreased. This reduces the effective mass, thereby increasing the velocity of vibration and increasing the total radiated sound energy to a value greater than if piston action were maintained. The total sound power still tends to drop somewhat at the higher frequencies. However as the radiated energy tends to be concentrated more nearly along the axis of the cone the higher the frequency, it is found that the response directly in front of the cone can be made substantially constant up to rather high frequencies, although there is a falling off of the higher frequencies to the side (see Fig. 268).

The efficiency of a loud-speaker in transforming electrical energy into acoustical radiation is relatively low because the coupling between the air and the vibrating cone is so poor that very little of the driving force is actually used in doing work against the air. Under practical conditions the average efficiency of a cone speaker is less than 5 per cent.

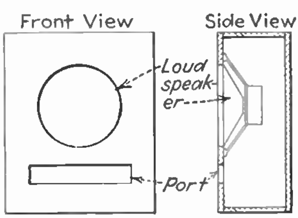


FIG. 270.—Bass-reflex system for obtaining reproduction of low frequencies.

Baffles and Bass-reflex Arrangements.—In order to reproduce low frequencies by means of a dynamic speaker, it is necessary to provide a baffle or other means to prevent the energy radiated from the two sides of the cone from canceling. At low frequencies the radiation is non-directional, so that if the distance from front to back of the cone is small compared with a wave length of sound at the frequency under consideration, there will be a tendency for the two radiations to cancel each other, since they are produced with a phase difference of 180°.

The simplest method of handling this situation is to mount the cone in some form of baffle, as illustrated in Fig. 267. In order to be effective, the baffle diameter should be of the order of a half wave length at the lowest frequency for which little or no loss in sound output is desired. For best results the baffle should have an irregular outline in order to eliminate the possibility of destructive interference between front and back radiations at certain critical frequencies.

The practical difficulty of obtaining sufficient baffle area to reproduce frequencies below 100 cycles in a receiver cabinet of reasonable size has led to the development of various devices for eliminating the baffle. A very widely used arrangement of this type is the bass-reflex arrangement illustrated in Fig. 270. Here the back of the cone is inclosed in a boxlike compartment, which is vented to the front by means of one or more holes. By properly proportioning these holes, the sound waves issuing from them at low frequencies will be 180° out of phase with the radiation from the back of the cone and so will reinforce the sound produced by the front of the cone. In this way the response can be extended to somewhat lower frequencies than would be possible if the same cabinet were used to function as an ordinary baffle.

139. Horns.—A horn is essentially an acoustic coupling device that transforms acoustic energy at a high pressure and low velocity to energy at a low pressure and high velocity. The way in which the horn accomplishes this in a practical loud-speaker can be understood by reference to Fig. 271. This shows a diaphragm driven by a moving coil and coupled to the throat of the horn by a small air chamber. Vibration of the diaphragm varies the volume of the air chamber, and hence the air pressure, which in turn causes sound waves to start out from the throat. These waves then travel along the horn, expanding in an orderly manner until large enough to transfer their energy to space without undue disturbance.

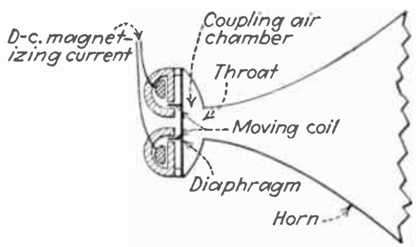


FIG. 271.—Schematic view of loud-speaker with horn.

The factors determining the behavior of a horn are the mouth area, the throat area, and the character of the taper. For proper operation the taper should be such that the cross-sectional area is proportional to an exponential function of the distance along the horn. That is,

$$\text{Area at distance } x \text{ from throat} = A_0e^{Bx} \tag{137}$$

where A_0 is the throat area, and B is a constant that determines the rate at which the horn opens out. A horn with such an exponential taper will freely transmit all sound waves that are appreciably above a critical or "cut-off" frequency, and will allow little or no energy of lower frequencies to pass through it. The transmission begins to fall off rapidly when, $f < 4000B$, with cut-off occurring at $f = 2730B$ (dimensions in centimeters). It is therefore apparent that the rate of taper is determined by the lowest frequency that is to be handled by the horn, with the necessary taper being more gradual as the frequency limit is lowered.

The mouth area of a horn determines the lowest frequency sound wave that can be transferred from the horn to free space without setting up standing waves in the air column. When the mouth diameter is $2\lambda/3$, such resonances are negligible and do not become excessive until the diameter is less than $\lambda/4$, where λ is the wave length of the sound. The mouth of a horn is hence determined by the lowest frequency that is to be handled.

The throat area of a horn determines the loading that is placed upon the diaphragm that is driving the horn. Thus when the throat is small the pressures built up in the air chamber by the diaphragm vibrations are magnet. Since the pull on the diaphragm is proportional to the square of the flux density in the air gap, then

$$\text{Pull on diaphragm} \propto (B_0 + B_s \sin \omega t)^2 = B_0^2 + 2B_0B_s \sin \omega t + \frac{B_s^2}{2}(1 - \cos 2\omega t) \quad (138)$$

The first term in the right-hand side of Eq. (138) represents the constant pull produced by the permanent magnet. The second term is a force that varies in accordance with the current passing through the receiver winding, and is proportional to the strength of the permanent magnet. This is the force that produces the desired diaphragm vibrations and is proportional to the strength of the field produced by the permanent magnet. The final term includes a double frequency distortion force that is relatively small compared with the desired force when the flux from the permanent magnet is large, but which represents the only alternating force when the permanent magnet is absent. A strong permanent magnet is hence required in order to obtain good sensitivity and low distortion.

142. Microphones.—A microphone is a device that converts sound waves into current or voltage waves. Although many types of microphones have been devised, the ones most commonly used are the carbon, velocity (or ribbon), condenser, crystal, and moving-coil microphones.

Carbon Microphones.—The carbon microphone makes use of the fact that the resistance which a mass of carbon granules offers to the flow of current depends upon the pressure applied to the mass. In a carbon microphone this mass of granules, termed a "button," is held against a diaphragm that is acted upon by the sound waves. The vibration of this diaphragm varies the pressure upon the carbon granules. This causes corresponding variations in the resistance that the granules offer to a direct-current voltage passing current from the diaphragm through the button. The construction of a high-grade microphone having buttons on each side of the diaphragm is illustrated in Fig. 274. Less expensive carbon microphones, such as those used in telephone systems and for many radio applications, have only a single button.

The response of a carbon microphone is independent of frequency up to the resonant frequency of the diaphragm. At diaphragm resonance the response depends upon the damping, and with proper adjustment can be made the same at resonance as at lower frequencies. Above resonance the response falls off rapidly. It is accordingly apparent that in order to obtain uniform response up to high frequencies with a carbon microphone, the diaphragm must be stretched to a relatively high resonant frequency.

The carbon microphone has the advantage over other microphones in that it is an amplifier. This is because the amount of electrical energy controlled by the diaphragm vibrations is more than the energy required to produce the vibrations. At the same time, the carbon microphone develops a hiss because of random changes in the resistance of the buttons to the passage of direct current. As a result the carbon microphone is used in the telephone system where the high sensitivity is a great asset, and also in amateur, police, and other radio work where the transmission of information rather than entertainment is the primary object, but is seldom used in public-address or high-quality radio-telephone work.

Moving-coil Microphones.—The moving-coil microphone makes use of a diaphragm that is vibrated by the pressure of the sound waves. Attached to the back of this diaphragm is a coil located in the field of a permanent magnet. As the diaphragm vibrates, the turns of the coil cut across the flux from the permanent magnet and accordingly have a

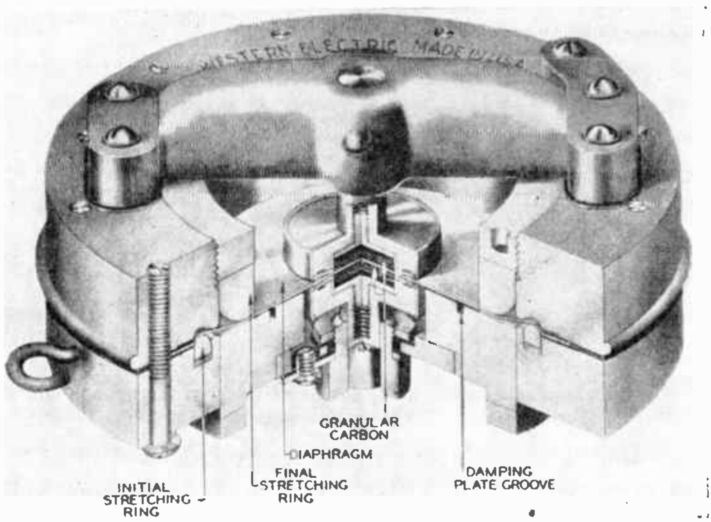


Fig. 274.—Constructional features of high-grade double-button carbon microphone.

voltage induced in them. By proper design this induced voltage can be made almost exactly proportional to the pressure of the sound waves acting against the diaphragm, and substantially independent of frequency over the range 40 to 10,000 cycles.

The moving-coil microphone is comparatively sensitive as compared with other non-amplifying microphones, has no hiss, is very rugged, and can be used outdoors even when gusty wind is present. Another important feature of the moving-coil microphone is that its output has a low impedance. This makes the output circuit relatively immune to electrostatic pick-up, with the result that very long leads can be used between the microphone and first amplifier tube. The moving-coil microphone finds wide use under practical conditions requiring high-quality reproduction.

The Velocity (or Ribbon) Microphone.—The velocity microphone is a special form of moving-coil microphone in which the moving coil consists of a ribbon of aluminum-alloy foil which is acted upon directly by the sound waves, and which has a resonant frequency below the audible range. The construction of such a microphone is shown in Fig. 275.

When a plane wave passes by a velocity microphone, the resulting force acting on the ribbon is proportional to the difference in sound pressure on the front and back of the ribbon. When the frequency is low enough so that the difference in distance to the two sides is appreciably less than a quarter wave length, the resulting force exerted against the ribbon is proportional to the frequency and to the pressure gradient, or particle velocity, of the sound wave. By making the resonant frequency of the ribbon lower than the lowest frequency to be reproduced, the ribbon offers an inertia reactance to motion that is very nearly proportional to frequency. The velocity of vibration, and hence the resulting voltage

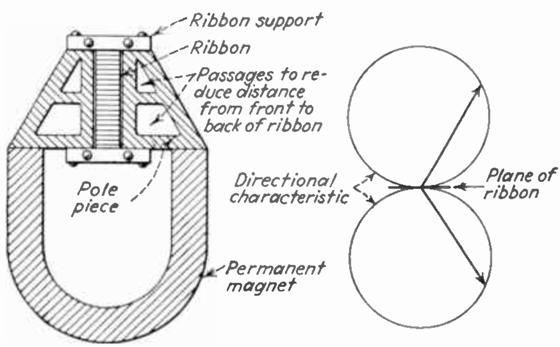


FIG. 275.—Sketch showing constructional details of velocity microphone and directional characteristic.

induced in the ribbon, are then substantially independent of frequency until the frequency is so great that the difference in distance to the front and back of the ribbon approaches a quarter wave length, when the response tends to fall off. By cutting away the pole pieces as is done in Fig. 275, this difference in distance can be made small enough to give a substantially uniform response to above 10,000 cycles.

When a velocity microphone is placed very close to the sound source, the low frequencies are overemphasized. This comes about because the pressure gradient in a spherical wave is greater in proportion to pressure than the pressure gradient of a plane wave by the factor $\sqrt{1 + (c/\omega r)^2}$, where r is the distance to the source, c is the velocity of sound, and ω is 2π times frequency.

The velocity microphone also has a pronounced directional characteristic as shown in Fig. 275. This is because sound waves arriving from the side strike both the front and back of the ribbon at the same instant and so produce no resultant force. This directivity can often be taken advantage of to minimize undesired reverberation and noise, and is helpful in controlling acoustic feedback in public-address systems.

The velocity microphone has the advantage of simplicity, an extremely good frequency response, and a low output impedance. Its principal disadvantage is that it cannot be used close to the source of sound, or outdoors in the presence of wind.

Condenser Microphone.—The condenser microphone is a condenser in which one plate is fixed and the other is a diaphragm against which the sound waves act. A direct-current potential of several hundred volts is applied between the plates of the condenser, and, as the capacitance is varied by the vibrations that the sound waves produce in the flexible plate, a corresponding voltage drop is produced in the high resistance R that is placed between the direct-current voltage and the microphone, as shown in Fig. 276.

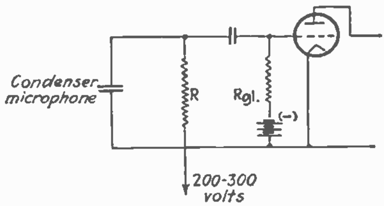


FIG. 276.—Circuit of condenser microphone.

The leads connecting a condenser microphone to its amplifier must be short in order to minimize electrostatic pickup. This is because the condenser microphone is a high-impedance device and is accordingly particularly susceptible to electrostatic fields. Short leads also keep the capacitance in shunt with the microphone a minimum and thereby avoid loss of sensitivity.

The high-frequency response of a condenser microphone is limited by the resonant frequency of the diaphragm, exactly as in the case of the carbon microphone. The response can be extended as far as desired at low frequencies by making the resistances R and R_{gl} in Fig. 276 large enough so that the equivalent resistance of these two in parallel is at least as great as the reactance of the microphone capacitance at the lowest frequency to be reproduced.

The principal use of the condenser microphone is in making sound measurements. At one time the condenser microphone was widely used in broadcast and public-address work, but it has been largely displaced in such applications by moving-coil and velocity microphones.

Pressure and Field Response Characteristics of Microphones.—Condenser, moving-coil, and similar microphones give a response that depends upon the pressure against the diaphragm, and are normally designed so that over the essential frequency range the output is proportional to pressure and independent of frequency. The calibration curve of such a microphone depends, however, upon the way in which the calibration is made, because the pressure of a sound wave traveling in free space is not necessarily the pressure that this same wave produces acting against the diaphragm of a microphone. At low frequencies, where the microphone dimensions are small compared with a wave length, the pressure against the diaphragm is a true measure of the pressure of the wave. At higher frequencies, however, reflections occur at the microphone and may cause the pressure exerted on the diaphragm to reach twice the pressure of the wave. In addition, the microphone front often forms a shallow air pocket which introduces a resonance that can still further increase the pressure against the diaphragm at high frequencies.

The relation between the output voltage of a condenser or moving-coil microphone and the pressure exerted against the diaphragm is called the *pressure calibration.* Similarly, the relation between the pressure of a sound wave in space and the output voltage that this wave develops when

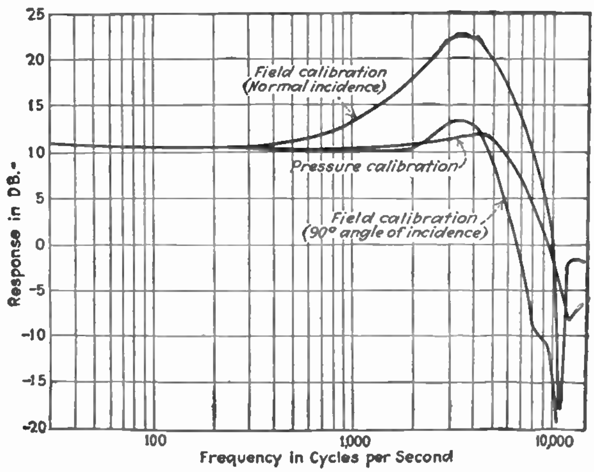


FIG. 277.—Typical pressure and field calibration of a condenser microphone.

striking the microphone is called the *field calibration.* The two are the same at low frequencies, whereas at high frequencies the field calibration depends upon the angle of incidence with which the wave strikes the microphone, as well as upon the frequency. Typical pressure- and field-calibration curves of a condenser microphone are shown in Fig. 277.

Reference Level for Expressing Microphone Sensitivity.—The sensitivity of a microphone is ordinarily expressed in terms of a reference level consisting of an output of 1 volt on an open circuit for a sound pressure of 1 bar (1 dyne per square centimeter). Thus a microphone having a sensitivity of -60 db will develop 0.001 volt on open circuit when acted upon by a sound wave having a pressure of 1 bar.

The pressure of 1 bar used in evaluating the sensitivity of a microphone is approximately the average pressure produced in ordinary speech at a distance of a few inches from the mouth.

Problems

1. Explain why subjective tones are much more pronounced when dealing with loud sounds than with weak sounds.

2. Explain why high-pitched sounds mask low frequencies only slightly if at all.

3. A particular piece of music when listened to in a theater is assumed to have a loudness level of 80 db, according to Fig. 265. If reproduced with a loudness level of 50 db in the home by a radio, calculate and plot with the aid of Fig. 265 the relative response, as a function of frequency, which the compensated tone control must provide to make up for the difference in reproduction level.

4. In view of the situation illustrated in Fig. 265, explain why the reverberation time should be different for different frequencies.

5. When a sound is reproduced in an empty auditorium, and then in the same auditorium filled with people, it is found that the average intensity of sound is much different, and there is also some difference in the quality of the sound. Discuss the differences and the factors that cause them.

6. The motional impedance of the voice coil of a dynamic speaker is found to change as the flux density of the magnetic field of the speaker is changed. Explain.

7. The motional impedance of the voice coil is reduced if mass is added to the voice coil. Explain.

8. In the frequency range where a cone functions as a piston, explain why the amplitude of vibration must be inversely proportional to frequency if the radiated power is to be independent of frequency.

9. Calculate and plot the required size of baffle as a function of the lowest frequency to be reproduced, over the range 50 to 250 cycles, and from the results discuss: (*a*) the necessity of inclosing the back side of the loud-speaker, or of using some form of baffle substitute, if frequencies below 125 cycles are to be satisfactorily reproduced by a loud-speaker mounted in an ordinary radio-receiver cabinet, and (*b*) the low-frequency response that can be expected of midget radio receivers. Assume the velocity of sound in air is 1130 ft. per sec.

10. *a*. Determine the rate of taper, the mouth area, and length of a horn that will reproduce frequencies down to 100 cycles and will have a throat area of 30 sq. cm.

b. Repeat (*a*) when the lowest frequency is 50 cycles.

c. From the results of (*a*) and (*b*) discuss the problem of reproducing low frequencies when practical considerations limit the physical dimensions of the horn.

11. In a horn speaker, the difference in distance from the throat to the nearest part of the diaphragm, and from the throat to the most distant part of the diaphragm, should be less than a half wave length at the highest frequency to be reproduced. Explain the reason for this.

12. What would be the effect upon the sound output of the horn in Fig. 272 if the sound-absorbing material were omitted from the box inclosing the back side of the cone?

13. Explain why, in order to have a flat frequency response, the diaphragm of a moving-coil microphone must vibrate with a velocity independent of frequency (amplitude of vibration inversely proportional to frequency) instead of with an amplitude independent of frequency, as in the case of the carbon and condenser microphones.

14. In the velocity microphone, increasing the distance from front to back of the ribbon increases the sensitivity in the low- and middle-frequency ranges in direct proportion to the reduction in high-frequency limit. Explain the reason for this.

15. Calculate and plot as a function of frequency, the relative response of a velocity microphone when the distance from microphone to sound source is 1, 2, and 5 ft. Cover the frequency range 50 to 500.

16. Explain why the difference between the field and pressure calibrations of a moving coil, condenser, or similar microphone at a particular frequency becomes less as the size of the microphone is reduced.

APPENDIX

FREQUENCY MEASUREMENTS

143. Frequency Measurements.—In order to avoid interference with other transmitters operating on adjacent channels, a radio transmitter must maintain its assigned carrier frequency with a high degree of accuracy. This requires that the frequency be checked either continuously or periodically and makes frequency measurements of fundamental importance in radio work.

Frequency Standards.—All accurate methods of determining frequency involve comparison of the unknown frequency with a known frequency.

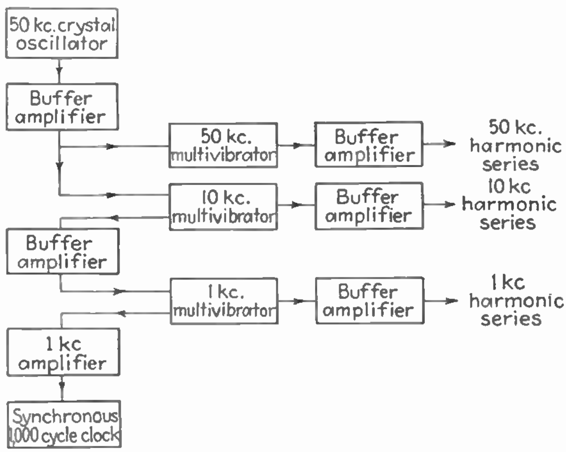


FIG. 278.—Schematic diagram of a primary frequency standard.

Practical frequency standards for use in making the comparison are conveniently classified as primary and secondary standards. A primary standard includes an oscillator that generates a frequency that is very stable and is periodically checked against the period of rotation of the earth as determined by astronomical means. Secondary standards are stable oscillators that have been calibrated by comparison with a primary standard of frequency.

The usual primary frequency standard consists of a crystal oscillator operating at a frequency of the order of 50 to 100 kc and having every possible refinement for improving the frequency stability. Associated with the stable oscillator is a train of multivibrators that operates a clock in synchronism with the oscillator frequency and also produces other frequencies related to the standard in an exact manner. The schematic diagram of a typical primary frequency standard is shown in Fig. 278. Here the crystal oscillator operates at 50 kc and controls the frequencies of multivibrators having fundamentals of 50 and 10 kc. The 10-kc multivibrator in turn controls a 1-kc multivibrator that is used to drive a synchronous clock that keeps exact time when supplied with 1000 cycles. Since the frequency supplied to the clock is always exactly one-fiftieth of the crystal frequency, the stability of the 50-kc oscillations can be measured by checking the time as kept by the clock against radio time signals. Frequencies related to the crystal frequency are obtained from the 50-, 10-, and 1-kc multivibrators, which develop harmonic series having fundamentals of 50, 10, and 1 kc, respectively, that have the same accuracy as the crystal oscillator and are available for making measurements. A properly designed primary frequency standard will maintain an accuracy approaching 1 part in 10,000,000 over long periods of time.

Secondary frequency standards may be of several types, depending upon their purpose. One very common arrangement makes use of a crystal oscillator designed for good frequency stability but without many of the refinements essential to a primary frequency standard. When such a secondary standard is used to monitor the frequency of a particular radio transmitter, as for example a broadcast transmitter, the crystal of the secondary standard is ground accordingly. The transmitter frequency is then compared directly with the secondary standard frequency by allowing the two to heterodyne together and observing the difference frequency that results. In cases where the secondary standard is to be used for general measuring purposes, the crystal is ground to some convenient frequency, such as 100 kc, and used to control a multivibrator that gives a harmonic sequence. The accuracy of a typical secondary standard employing a crystal oscillator is of the order of 10 parts per million over long periods of time.

Another type of secondary standard consists of a tuned-circuit type of oscillator, usually of the electron-coupled type. Such an oscillator is called a *heterodyne frequency meter* and is often, although not always, provided with a heterodyne detector, amplifier, and phones for observing audible beat notes between an unknown frequency and the oscillations of the frequency meter. The heterodyne frequency meter has the advantage of being continuously adjustable, whereas frequency standards based upon crystal oscillators give only a fixed frequency or a series of frequencies related to this one frequency. However, a well-designed heterodyne frequency meter has an accuracy of only about 1 part in

1000 and so is suitable only where high accuracy is not required or where the instrument is to be used for interpolation purposes.

Rough determinations of frequency are often made by the use of a wavemeter or by the use of Lecher wires. A wavemeter is a calibrated tuned circuit that is adjusted to resonance (with the frequency to be measured) by means of a variable condenser. The accuracy of such an arrangement is of the order of 1 per cent. A Lecher wire system is essentially a two-wire transmission line short-circuited at the receiving end. In such an arrangement the current and voltage vary with distance as shown in Fig. 28*b*, and the distance between adjacent minima is almost exactly a half wave length. This fact can be used to determine the frequency when the wave length is sufficiently short to make the transmission line of reasonable dimensions. The accuracy obtainable is approximately 1 per cent.

Measurement of Transmitter Frequency Using a Heterodyne Frequency Meter.—Consider first the case where the frequency of the transmitter is within the tuning range of the heterodyne frequency meter. The transmitter is loosely coupled to the frequency meter, which is then adjusted until the heterodyne detector indicates zero beat with the transmitter oscillations.[1] The frequency read on the calibration curve of the heterodyne frequency meter for this setting then gives the desired answer. In order to avoid error, the coupling between the transmitter and the frequency meter should be very loose in order to prevent automatic synchronization. It is also necessary to make sure that the loudest beat note observable is the one used in the measurements, since weak beat notes resulting from combinations of harmonics may also be present. These weaker beat notes may be used, however, to check the final results. Thus, when the heterodyne frequency meter is adjusted to one-half the transmitter frequency, there should be a weak beat note between the second harmonic of the meter and the fundamental frequency of the transmitter.

Next consider the problem of making measurements when the transmitter frequency is higher than the range covered by the oscillations of the heterodyne frequency meter. Under these circumstances the heterodyne frequency meter is adjusted so that a harmonic of its oscillations gives zero beat with the transmitter frequency. The unknown

[1] The exact position for zero beat can be determined by using an oscillating heterodyne detector to observe the beats. The oscillating detector is first adjusted to give a beat note of approximately 1000 cycles with the transmitter, after which the heterodyne frequency meter is adjusted. When the heterodyne frequency meter is within a few cycles of the transmitter frequency, the 1000 output of the detector will then wax and wane in intensity at a rate equal to the difference in frequency between the heterodyne frequency meter and the transmitter frequency, and variation of the oscillating-detector adjustment will not affect the rate of waxing and waning.

frequency is then nf, where f is the frequency as read from the frequency meter, and n is the harmonic involved. Results obtained in this way can be checked by determining the frequency-meter setting corresponding to zero beat when different harmonics are used.

Finally, there is the problem of measuring a transmitter frequency less than the range covered by the frequency meter. Here the frequency meter is adjusted to give zero beat with a harmonic of the transmitter, which then has a frequency f/n, where f corresponds to the setting of the frequency meter, and n is the harmonic. The results can be checked as above by determining the frequency meter setting corresponding to different harmonics.

Measurements of Transmitter Frequency by Means of a Series of Harmonics.—When the accuracy required is greater than obtainable with a heterodyne frequency meter, it is necessary to compare the unknown frequency with a harmonic series produced by a multivibrator that is controlled by a primary or secondary frequency standard. This comparison is ordinarily made by using a heterodyne frequency meter as an interpolation device. The heterodyne frequency meter is first adjusted just as though it were being used to determine the frequency. The harmonic series from the frequency standard is then combined with oscillations of the frequency meter. Detection of the resultant wave gives a beat note that represents the difference between the oscillations of the frequency meter and the nearest harmonic of the harmonic series. The exact frequency of the frequency meter is then this nearest harmonic plus or minus the beat note.[1] The actual frequency of the beat note can be measured directly by comparison with an audio oscillator, or can be determined approximately by interpolation between the setting of the frequency meter corresponding to zero beat with: (1) the transmitter frequency, (2) the harmonic adjacent to and just higher than the transmitted frequency, and (3) the harmonic adjacent to and just below the transmitter frequency.

Measurement of Frequency of Received Signals.—When the frequency of a distant transmitter is to be determined, the signals are tuned in on a sensitive radio receiver. The heterodyne frequency meter is then turned on, coupled into the receiver, and adjusted so that zero beat is obtained between the oscillations of the heterodyne frequency meter and the received signal, or between the fundamental of one oscillation and a harmonic of the other. This procedure effectively transfers the received signal to the frequency meter, after which the measurement is completed by following the same procedure as in the transmitter case.

[1] The plus sign is used when an increase in the frequency of the meter increases the pitch of the beat note.

INDEX

L

M

N

O

P

Q

R

U

V

W