ELECTRONIC AND RADIO ENGINEERING

FOURTH EDITION

FREDERICK EMMONS TERMAN

Professor of Electrical Engineering
Dean of the School of Engineering
Stanford University

Assisted by

ROBERT ARTHUR HELLIWELL
Associate Professor of Electrical Engineering
Stanford University

JOSEPH MAYO PETTIT
Professor of Electrical Engineering
Stanford University

DEAN ALLEN WATKINS
Associate Professor of Electrical Engineering
Stanford University

WILLIAM RALPH RAMBO
Associate Director, Applied Electronics Laboratory
Stanford University

ASIAN STUDENTS' EDITION

McGRAW-HILL BOOK COMPANY, INC.
New York Toronto London

KŌGAKUSHA COMPANY LTD.
TOKYO

1922, 1932, 1937, 1947

ELECTRONIC AND RADIO ENGINEERING

ASIAN STUDENTS' EDITION

TOSHO PRINTING CO., LTD., TOKYO, JAPAN

PREFACE

This fourth edition has the same objective as the three prior editions, namely, to provide a text and reference book that summarizes in easily understandable terms those principles and techniques which are the basic tools of the electronic and radio engineer. In keeping with current trends, increased emphasis is placed on the general techniques of electronics, without regard to the extent of their use in radio systems. This change is reflected in the new title, "Electronic and Radio Engineering," which is more descriptive of the subject matter actually covered in the present volume than is the previous title, "Radio Engineering."

The keynote continues to be thorough coverage combined with a presentation that allows the reader to study a particular topic without having to read the entire book. The level of presentation, particularly the mathematical level, remains unchanged. Thus the present volume is designed to serve as a text and reference for the same clientele that found the previous editions so useful.

To keep pace with a rapidly advancing technology, new material has been added in practically every chapter. More than half the illustrations are new, and all have been redrawn to conform to new graphic standards. A new chapter dealing with microwave tubes makes available for the first time an explanation in simple language of the basic mechanism of operation of traveling-wave tubes and backward-wave oscillators (carcinotrons). In the treatment of wideband video and tuned amplifiers, primary emphasis is placed on the rise time, overshoot, and sag, since these characteristics are more indicative of the performance under actual conditions than is the older approach in terms of amplitude and phase behavior as a function of frequency. The material on nonlinear waveforms and pulse techniques has been greatly expanded to provide more complete coverage of this important aspect of electronics. The chapter on television has been thoroughly revised, and a compact and simple explanation is given of the system of color television now standard in the United States. Increased attention is also placed on propagation phenomena involving the troposphere.

Of particular importance is the chapter on Transistors and Related Semiconductor Devices, one of the longest in the book. Here is presented a simple, straightforward explanation of the basic phenomena occurring inside the transistor, and of how these phenomena lead to the terminal characteristics. This treatment is such that it can be understood by undergraduate students; at the same time, it is sufficiently complete and fundamental to provide a firm foundation for further study of this new and very important subject.

Special attention has been given to the needs of the teacher. Because of the growth of electronics, it is no longer possible to cover every important topic adequately in a one-year course. "Electronic and Radio Engineering" provides the instructor with an opportunity to select those topics which he himself wishes to emphasize, and at the same time provides the student with a reference book of comprehensive coverage and continuing value. It will be observed that the book breaks down into three distinct parts, namely, a group of chapters dealing with circuits (components, resonant circuits, transmission lines, waveguides, and cavity resonators); a group of chapters concerned with the fundamentals of electronic engineering (vacuum tubes, transistors, amplifiers, oscillators, modulators, detectors, nonlinear waveforms, etc.), which are the heart of the book; and a concluding group of chapters concerned with radio systems and radio engineering (antennas, propagation, transmitters, receivers, television, radar, and radio aids to navigation). Thus an instructor can, if he desires, concentrate on the material concerned with fundamental electronics and regard the remaining subject matter as available to the student, should he need to extend his knowledge at a future date. Alternatively, the instructor can choose to cover a series of selected topics, for example, waveguides, wideband systems, pulse circuits, television, etc. Another possibility is to concentrate on the material concerned primarily with radio systems. Many other combinations, are, of course, possible.

An important feature for the teacher is the more than 1250 Problems and Exercises. Many of these involve numerical calculations, but more than half of them are thought questions that will require the student to give further consideration to topics covered in the text. Such Exercises can be used to extend and solidify the student's knowledge; they are also suggestive of questions suitable for use on examinations. The number of Problems and Exercises is so large that the same problem need not be assigned to a class more often than once every two or three years.

The collaborators listed on the title page have made important contributions to the preparation of this volume. Dr. Helliwell worked on the sections dealing with ionospheric propagation, and Dr. Pettit is in large measure responsible for the general character of the chapter dealing with transistors and semiconductors. The treatment of traveling-wave tubes and backward-wave oscillators is due to Dr. Watkins. William Rambo prepared the background material used in revising the presentation on radar. In addition, acknowledgment is made to Dr. B. H. Wadia, Bruno Ludovici, and Arthur Vassilaides, graduate students at Stanford, for assistance in preparing illustrations.

FREDERICK EMMONS TERMAN

CONTENTS

CHAPTER 1

THE ELEMENTS OF A SYSTEM OF RADIO COMMUNICATION

1-1. Radio Waves. Electrical energy that has escaped into free space exists in the form of electromagnetic waves. These waves, which are commonly called radio waves, travel with the velocity of light and consist of magnetic and electric fields that are at right angles to each other and also at right angles to the direction of travel. If these electric and magnetic fluxes could actually be seen, the wave would have the appearance indicated in Fig. 1-1. One-half of the electrical energy contained

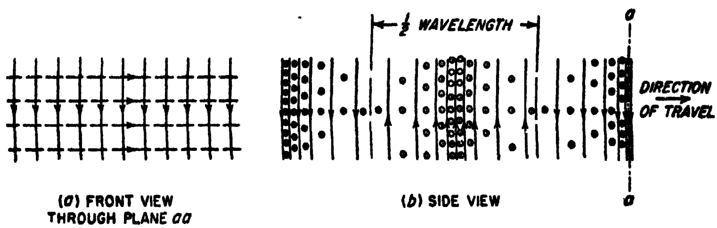


FIG. 1-1. Front and side views of a vertically polarized wave. The solid lines represent electric flux; the dotted lines and the circles indicate magnetic flux.

in the wave exists in the form of electrostatic energy, while the remaining half is in the form of magnetic energy.

The essential properties of a radio wave are the frequency, intensity, direction of travel, and plane of polarization. The radio waves produced by an alternating current will vary in intensity with the frequency of the current and will therefore be alternately positive and negative as shown in Fig. 1-1*b*. The distance occupied by one complete cycle of such an alternating wave is equal to the velocity of the wave divided by the number of cycles that are sent out each second and is called the wavelength. The relation between wavelength λ in meters and frequency f in cycles per second is therefore

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

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.

(1-1) to have a long wavelength, while a high frequency corresponds to a short wavelength.

The strength of a radio wave is measured in terms of the voltage stress produced in space by the electric field of the wave, and it is usually expressed in microvolts stress per meter. Since the actual stress produced at any point by an alternating wave varies sinusoidally from instant to instant, it is customary to consider the intensity of such a wave to be the effective value of the stress, which is 0.707 times the maximum stress in the atmosphere during the cycle. The strength of the wave measured in terms of microvolts per meter of stress in space is also exactly the same voltage that the magnetic flux of the wave induces in a conductor 1 m long when sweeping across this conductor with the velocity of light.

The minimum field strength required to give satisfactory reception of a wave depends upon a number of factors, such as frequency, type of signal involved, and amount of interference present. Under some conditions radio waves having signal strengths as low as 0.1 μv per m are usable. Occasionally signal strengths exceeding 1000 μv per m are required to ensure entirely satisfactory reception at all times. In most cases the weakest useful signal strength lies somewhere between these extremes.

A plane parallel to the mutually perpendicular lines of the electric and electromagnetic flux is called the wavefront. The wave always travels in a direction at right angles to the wavefront, but whether it goes forward or backward depends upon the relative direction of the lines of magnetic and electric flux. If the direction of either the magnetic or electric flux is reversed, the direction of travel is reversed; but reversing both sets of flux has no effect.

The direction of the electric lines of flux is called the direction of polarization of the wave. If the electric flux lines are vertical, as shown in Fig. 1-1, the wave is vertically polarized; when the electric flux lines are horizontal and the electromagnetic flux lines are vertical, the wave is horizontally polarized.

Propagation of Radio Waves of Different Frequencies. As radio waves travel away from their point of origin, they become attenuated or weakened. This is due in part to the fact that the waves spread out.

In addition, however, energy may be absorbed from the waves by the ground or by the ionized regions in the upper atmosphere termed the ionosphere, and the waves may also be reflected or refracted by the ionosphere, or by conditions within the lower atmosphere, or by the ground. The resulting situation is quite complex and differs greatly for radio waves of different frequencies, as shown in Table 1-1, which summarizes the behavior of different classes of radio waves.

1-2. Radiation of Electrical Energy. Every electrical circuit carrying alternating current radiates a certain amount of electrical energy in the form of electromagnetic waves, but the amount of energy thus radiated is extremely small unless all the dimensions of the circuit approach the order of magnitude of a wavelength. Thus, a power line carrying 60-cycle current with a 20-ft spacing between conductors will radiate practically no energy because a wavelength 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-m

TABLE 1-1
CLASSIFICATION OF RADIO WAVES

Class	Frequency range	Wavelength range	Propagation characteristics	Typical uses
Very low frequency (VLF)	10–30 kc	30,000–10,000 m	Low attenuation at all times of day and of year; characteristics very reliable	Long-distance point-to-point communication
Low frequency (LF)	30–300 kc	10,000–1000 m	Propagation at night similar to VLF but slightly less reliable; daytime absorption greater than VLF	Long-distance point-to-point service, marine, navigational aids
Medium frequency (MF)	300–3000 kc	1000–100 m	Attenuation low at night and high in daytime	Broadcasting, marine communication, navigation, harbor telephone, etc.
High frequency (HF)	3–30 Mc	100–10 m	Transmission over considerable distance depends solely on the ionosphere, and so varies greatly with time of day, season, and frequency	Moderate and long-distance communication of all types
Very high frequency (VHF)	30–300 Mc	10–1 m	Substantially straight-line propagation analogous to that of light waves; unaffected by ionosphere	Short-distance communication, television, frequency modulation, radar, airplane navigation
Ultra-high frequency (UHF)*	300–3000 Mc	100–10 cm	Same	Short-distance communication, radar, relay systems, television, etc.
Super-high frequency (SHF)*	3000–30,000 Mc	10–1 cm	Same	Radar, radio relay, navigation

* Frequencies higher than about 2000 Mc are frequently referred to as microwave frequencies.

wavelength of this radio wave. From these considerations it is apparent that the size of radiator required is inversely proportional to the frequency. High-frequency waves can therefore be produced by a small radiator, while low-frequency waves require a large antenna system for effective radiation.

Every radiator has directional characteristics as a result of which it sends out stronger waves in certain directions than in others. Directional characteristics of antennas are used to concentrate the radiation toward the point to which it is desired to transmit, or to favor reception of energy arriving from a particular direction.

1-3. Generation and Control of Radio-frequency Power. The radio-frequency power required by a radio transmitter is practically always obtained from a vacuum-tube oscillator or amplifier. Vacuum tubes can convert d-c power into a-c energy for all frequencies from the very lowest up to 30,000 Mc, or even higher. Under most conditions the efficiency with which this transformation takes place is in the neighborhood of 50 per cent or higher. At frequencies up to well over 1000 Mc, the amount of

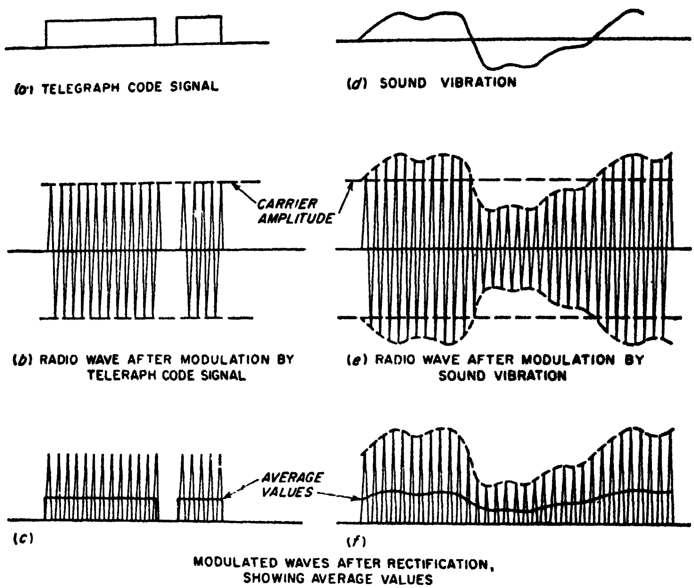


FIG. 1-2. Diagram showing how a signal may be transmitted by modulating the amplitude of a radio wave, and how the original signal may be recovered from the modulated wave by rectification. For the sake of clarity the radio frequency is shown as being much lower than would usually be the case.

power that can be generated continuously by vacuum tubes is of the order of kilowatts.

Modulation. If a radio wave is to convey a message, some feature of the wave must be varied in accordance with the information to be transmitted. One way to do this, termed *amplitude modulation,* consists in varying the amplitude of the radiated wave. In radio telegraphy, this involves turning the radio transmitter on and off in accordance with the dots and dashes of the telegraph code, as illustrated in Fig. 1-2*b*. In radio-telephone transmission by amplitude modulation the radio-frequency wave is varied in accordance with the pressure of the sound wave being transmitted, as shown in Fig. 1-2*e*. Similarly in picture transmission, the amplitude of the wave radiated at any one time is made proportional to the light intensity of the part of the picture that is being transmitted at that instant.

Intelligence may be transmitted by other means than by varying the amplitude. For example, one may maintain the amplitude constant and vary the frequency that is radiated in accordance with the intelligence, thus obtaining *frequency modulation.* This results in a wave such as shown in Fig. 1-3*b*, which is to be compared with the corresponding amplitude-modulated wave of Fig. 1-3*a*. Frequency modulation is widely used in very high-frequency communication systems.

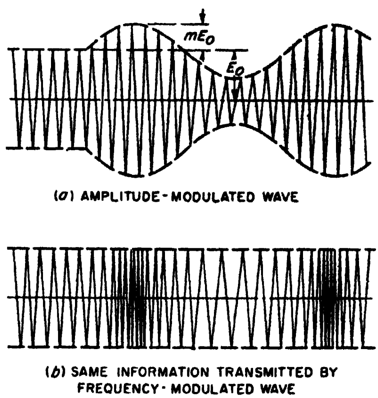


FIG. 1-3. Character of waves produced by amplitude modulation and by frequency modulation, where the modulation is sinusoidal in both cases. For the sake of clarity the radio frequency is shown much lower than would usually be the case.

1-4. Reception of Radio Signals. In the reception of radio signals it is first necessary to abstract energy from the radio wave passing the receiving point. Any antenna capable of radiating electrical energy is also able to absorb energy from a passing radio wave. This occurs because the electromagnetic flux of the wave, in cutting across the antenna conductor, induces in the antenna a voltage that varies with time in exactly the same way as does the current flowing in the antenna radiating the wave. This induced voltage, in association with the current that it produces, represents energy that is absorbed from the passing wave.

Since every wave passing the receiving antenna induces its own voltage in the antenna conductor, 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 transmitting stations and is carried out by the use of resonant circuits which can be made to discriminate very strongly in favor of a particular 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.*

Although intelligible radio signals have been received from radio transmitters thousands of miles distant, using only the energy abstracted from the radio wave by the receiving antenna, much more satisfactory reception can be obtained if the received energy is amplified. This amplification may be applied to the radio-frequency currents before detection, in which case it is called radio-frequency amplification; or it may be applied to the rectified currents after detection, in which case it is called audio-frequency amplification. The use of amplification makes possible the satisfactory reception of signals from waves that would otherwise be too weak to give an audible response. The only satisfactory method of amplifying radio signals that has been discovered is by the use of vacuum tubes or transistors. Before vacuum tubes were discovered, radio reception had available only the energy abstracted from the radio wave by the receiving antenna.

Detection. The process by which the message being transmitted is reproduced from the modulated radio-frequency current present in the receiver is called *detection*, or sometimes *demodulation*. With amplitude-modulated waves, detection is accomplished by rectifying the radio-frequency currents to produce a current that varies in accordance with the modulation of the received wave. Thus, when the modulated wave shown at *e* of Fig. 1-2 is rectified, the resulting current, shown at *f*, is seen to have an average value that varies in accordance with the amplitude of the original signal. In the transmission of code signals by radio, the rectified current reproduces the dots and dashes of the telegraph code, as shown at Fig. 1-2*c*, 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, since otherwise the telephone receiver would give forth a succession of unintelligible clicks.

The detection of a frequency-modulated wave involves two steps. First, the wave is transmitted through a circuit in which the relative response depends upon the frequency. The wave that then emerges from the circuit is amplitude-modulated, since as the frequency of the constant-amplitude input wave changes, the output amplitude will follow the variation of circuit transmission with frequency. The resulting amplitude-modulated wave is then rectified.

1-5. Nature of a Modulated Wave. A sine wave conveys very little information since it repeats over and over again. When a wave is modulated, either in amplitude or frequency, it is no longer a simple sine wave, but is instead a mixture of several waves of slightly 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 amplitude modulation shown in Fig. 1-3*a*, the amplitude of the radio-frequency oscillation is given by $E = E_0 + mE_0 \sin 2\pi f_s t$, in which E_0 represents the average amplitude, f_s the frequency at which the amplitude is varied, and m the ratio of amplitude variation from the average to the average amplitude, which is called the ***degree of modulation.*** The equation of the amplitude-modulated wave can be hence written as

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

in which f is the frequency of the radio oscillation. Multiplying out the right-hand side of Eq. (1-2) gives

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

By expanding the last term into functions of the sum and difference angles by the usual trigonometric formula, the equation of a wave with simple sine-wave amplitude modulation can be written in the form

$$e = E_0 \sin 2\pi f t + \frac{mE_0}{2} \cos 2\pi(f - f_s)t - \frac{mE_0}{2} \cos 2\pi(f + f_s)t \tag{1-3}$$

Equation (1-3) shows that the wave with sine-wave modulation consists of three separate waves. The first of these, represented by the term $E_0 \sin 2\pi f t$, is called the *carrier*. Its amplitude is independent of the presence or absence 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 that of the carrier frequency by an amount equal to the modulation frequency, while the frequency of the other is more than that of the carrier by the same amount. These two components, called *sideband frequencies*, carry the intelligence that is being transmitted by the modulated wave. The frequency of the sideband components relative to the carrier frequency is determined by the modulation frequency. The relative amplitude of the sideband components is determined by the extent of the amplitude variations that are impressed upon the wave, i.e., by the degree of modulation.

When the modulation is more complex than the simple sine-wave amplitude variation of Fig. 1-3*a*, the effect is to introduce additional sideband components. Thus, if the wave of a radio-telephone transmitter is amplitude-modulated by a complex sound wave containing pitches of 1000 and 1500 cycles, the modulated wave will contain one pair of 1000-cycle sideband components and one pair of 1500-cycle sideband components.

The analysis of a frequency-modulated wave is somewhat more complex but leads to an analogous result. The principal difference is that the frequency-modulated wave not only contains the same sideband frequencies as does the corresponding amplitude-modulated wave, but in addition contains higher-order side bands. Thus, if a wave has its frequency varied at a rate of 1000 times per second, the resulting modulated wave will contain not only a pair of 1000-cycle sideband components, but in addition a pair of 2000-cycle sideband components, possibly a pair of 3000-cycle sideband components, etc. The amplitude of these various sideband pairs will depend upon the extent and upon the rate of frequency variation.

Significance of the Sidebands. The carrier and sideband frequencies are not a mathematical fiction, but have a real existence, as 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 sideband frequencies can be considered as being generated as a result of varying the wave. They are present only when the wave is being varied, and their magnitude and frequency are determined by the character of the modulation.

It is apparent that the transmission of intelligence requires the use of a band of frequencies rather than a single frequency. Speech and music of the quality reproduced in standard broadcasting involve frequency components from about 100 cycles up to 5000 cycles; when modulated upon a carrier wave, the total bandwidth involved is therefore 10,000 cycles. If this entire band is not transmitted equally well through space, and by the circuits in both transmitter and receiver through which the modulated wave must pass, then the sideband frequency components that are discriminated against will not be reproduced in the receiving equipment with proper amplitude, and a loss in quality will result. With telegraph signals, the required sideband is relatively narrow because the amplitude of the signals is varied only a few times a second, but a definite frequency band is still required. If some of the sideband components of the code signal are not transmitted, the received dots and dashes tend to be rounded off and run together, and may become indistinguishable.

1-6. The Decibel. The decibel (abbreviated db) is a logarithmic unit used in communication work to express power ratios. If the powers being compared are P_1 and P_2, then

$$\text{Decibels} = 10 \log_{10} \frac{P_2}{P_1} \qquad (1\text{-}4)$$

The sign associated with the number of decibels indicates which power is greater; thus a negative sign means P_2 is less than P_1.

The decibel has no other significance than that given in Eq. (1-4). Thus, if decibels are used to express amplification, this simply means that the presence of the amplification increases the power output by the number of decibels attributed to the amplification. Again, under many conditions relative power is proportional to the square of the voltage E (or current I, or field B, etc.). Under these conditions

$$\text{Decibels} = 20 \log_{10} \frac{E_2}{E_1} = 20 \log_{10} \frac{I_2}{I_1} = 20 \log_{10} \frac{B_2}{B_1}, \text{ etc.} \qquad (1\text{-}5)$$

These relations must be used with caution, however, as they hold only when the resistance associated with E_2 (or I_2 or B_2) is the same as associated with E_1 (or I_1 or B_1).

TABLE 1-2
(a) POWER, VOLTAGE, AND CURRENT RATIOS FOR ASSIGNED DECIBEL VALUES

Db	Current and voltage ratio		Power ratio		Db	Current and voltage ratio		Power ratio	
	Gain	Loss	Gain	Loss		Gain	Loss	Gain	Loss
0.0	1.00	1.000	1.00	1.000	10	3.16	0.316	10.00	0.100
0.2	1.02	0.977	1.05	0.955	12	3.98	0.251	15.8	0.063
0.4	1.05	0.955	1.10	0.912	14	5.01	0.200	25 1	0.040
0.6	1.07	0.933	1.15	0.871	16	6.31	0.158	39.8	0.025
0.8	1.10	0.912	1.20	0.832	18	7.94	0.126	63.1	0.016
1.0	1.12	0.891	1.26	0.794	20	10.00	0.100	100.0	0.010
1.5	1.19	0.841	1.41	0.708	25	17.8	0.056	3.16×10^{2}	3.16×10^{-3}
2.0	1.26	0.794	1.58	0.631	30	31.6	0.032	10^{3}	10^{-3}
2.5	1.33	0.750	1.78	0.562	35	56.2	0.018	3.16×10^{3}	3.16×10^{-4}
3.0	1.41	0.708	2.00	0.501	40	100.0	0.010	10^{4}	10^{-4}
3.5	1.50	0.668	2.24	0.447	45	177.8	0.006	3.16×10^{4}	3.16×10^{-5}
4.0	1.58	0.631	2.51	0.398	50	316	0.003	10^{5}	10^{-5}
4.5	1.68	0.596	2.82	0.355	60	1,000	0.001	10^{6}	10^{-6}
5	1.78	0.562	3.16	0.316	70	3,160	0.0003	10^{7}	10^{-7}
6	2.00	0.501	3.98	0.251	80	10,000	0.0001	10^{8}	10^{-8}
7	2.24	0.447	5.01	0.200	90	31,600	0.00003	10^{9}	10^{-9}
8	2.51	0.398	6.31	0.158	100	100,000	0.00001	10^{10}	10^{-10}
9	2.82	0.355	7.94	0.126	120	1,000,000	0.000001	10^{12}	10^{-12}

(b) DECIBEL EQUIVALENT OF POWER, VOLTAGE, AND CURRENT RATIOS

Ratio	Db equivalent		Ratio	Db equivalent		Ratio	Db equivalent	
	Power	Voltage or current		Power	Voltage or current		Power	Voltage or current
10^{-6}	−60.00	−120.00	1.2	0.79	1.58	10	10.00	20.00
10^{-5}	−50.00	−100.00	1.4	1.46	2.92	12	10.79	21.58
10^{-4}	−40.00	−80.00	1.6	2.04	4.08	14	11.46	22.92
0.001	−30.00	−60.00	1.8	2.55	5.10	16	12.04	24.08
0.003	−25.23	−50.46	2.0	3.01	6.02	18	12.55	25.10
0.005	−23.01	−46.02	2.5	3.98	7.96	20	13.01	26.02
0.01	−20.00	−40.00	3.0	4.77	9.54	25	13.98	27.96
0.03	−15.23	−30.46	3.5	5.44	10.88	30	14.77	29.54
0.05	−13.01	−26.02	4.0	6.02	12.04	40	16.02	32.04
0.10	−10.00	−20.00	4.5	6.53	13.06	50	16.99	33.98
0.15	−8.24	−16.48	5.0	6.99	13.98	60	17.78	35.56
0.20	−6.99	−13.98	5.5	7.40	14.81	80	19.03	38.06
0.30	−5.23	−10.46	6.0	7.78	15.56	100	20.00	40.00
0.40	−3.98	−7.96	6.5	8.13	16.26	10^{3}	30.00	60.00
0.50	−3.01	−6.02	7.0	8.45	16.90	10^{4}	40.00	80.00
0.60	−2.22	−4.44	7.5	8.75	17.50	10^{5}	50.00	100.00
0.80	−0.97	−1.94	8.0	9.03	18.06	10^{6}	60.00	120.00
1.00	0.00	0.00	9.0	9.54	19.08	10^{7}	70.00	140.00

The practical value of the decibel arises from its logarithmic nature. This permits the enormous ranges of power involved in communication work to be expressed in terms of decibels without running into inconveniently large numbers, while at the same time permitting small ratios to be conveniently expressed. Thus, 1 db represents a power ratio of approximately 5:4, while 60 db represents a ratio of 1,000,000:1. The logarithmic character of the decibel also makes it possible to express the ratio of input to output powers of a complicated circuit as the sum of the decibel equivalent of the ratios of the input to output powers of the different parts of the circuit that are in cascade.

Table 1-2 gives a convenient summary of decibel values.

CHAPTER 2

CIRCUIT ELEMENTS

2-1. Inductance. A current flowing in an electrical circuit produces magnetic flux that links with (i.e., encircles) the current. The effect of this flux is expressed in terms of a property of the circuit called the *inductance*.

Inductance can be defined as the flux linkages per ampere of current producing the flux; i.e.,

$$\begin{matrix}\text{Inductance } L \\ \text{in henrys}\end{matrix} = \frac{\text{flux linkages}}{\text{current (amperes) producing flux}} \times 10^{-8} \quad (2\text{-}1)$$

A flux linkage represents one flux line encircling the circuit current once. Thus in Fig. 2-1 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.

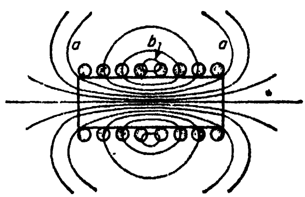


FIG. 2-1. Flux and current distribution in typical single-layer air-cored inductance coil. The current density is indicated by the density of shading.

Calculation of Inductance. The inductance of an electrical circuit is computed by assuming a convenient current flowing in the circuit. The magnetic flux produced by this current is then calculated, and the total number of flux linkages that results is counted. The inductance in henrys is this total number of flux linkages multiplied by 10^{-8} and divided by the circuit current.

Formulas have been derived by this procedure that give the inductance for all commonly used types of air-cored coils.[1] It is thus neither necessary nor desirable to guess at the number of turns and coil dimensions required to obtain a desired inductance. For example, the inductance of a single-layer solenoid, such as shown in Fig. 2-1, is given by the relation

$$\text{Inductance in microhenrys} = Fn^2d \quad (2\text{-}2)$$

[1] A comprehensive collection of such formulas is given by F. E. Terman, "Radio Engineers' Handbook," pp. 48–64, McGraw-Hill Book Company, Inc., New York, 1943.

where n = number of turns

d = diameter of coil measured to center of wire

F = constant that depends only upon the ratio of length to diameter, given in Fig. 2-2

The quantity F depends in a complicated way upon the ratio of coil length to diameter, since the geometrical distribution of the flux produced by the current in the coil does not follow a simple mathematical law. However, once the relationship represented by F has been determined,

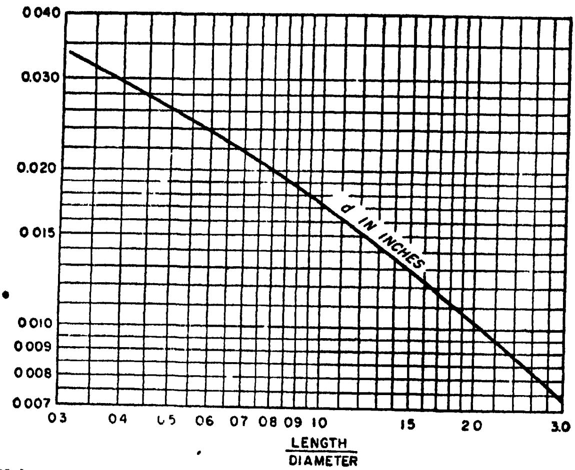


FIG. 2-2. Values of constant F for use in Eq. (2-2), to obtain the inductance of single-layer solenoids.

the value of F can be computed once for all, and presented by a curve, such as Fig. 2-2, or by a table.

The inductance of all coils with air cores 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.

The inductance of all air-cored 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, and depth of winding), then the larger coil will have twice the inductance of the smaller one. This rule results from the fact that the cross section of the flux paths is proportional to the square of the linear dimension of the coil, while the length of these paths varies directly as the linear dimension.

In calculating the inductance of coils with magnetic cores, the flux is determined in accordance with the usual methods of making magnetic circuit calculations, taking into account air gaps, leakage and fringing flux, etc. It is also necessary to assume the proper value of permeability, as discussed below.[1] To the extent that the permeability of the core material can be considered as constant, the inductance of a coil with a magnetic core is proportional to the square of the number of turns and to the first power of the size, just as in the air-cored case.

Inductance of a Connecting Wire. The inductance associated with a connecting wire depends on the wire diameter, and can be minimized by making the diameter large. This results from the fact that when the wire diameter is small the length of the flux paths immediately outside of the wire is less than if the diameter is large. As a result the small wire is circled by more flux and hence has higher inductance.

An alternative means of achieving a low-inductance connection consists in employing a conductor comprising two or more spaced wires connected in parallel. If three wires are employed, they should be placed at the corners of an equilateral triangle; in a four-wire system the individual wires would be at the corners of a square, etc. Such arrangements give the first approximation to a solid conductor of large diameter, and will have less inductance the greater the diameter of the individual wires and the greater the spacing between the wires connected in parallel.

Initial and Incremental Permeability; Incremental Inductance. The permeability of a magnetic material is defined as the ratio B/H of the flux density to the magnetizing force, and depends upon the flux and the material. The permeability at very low flux densities, termed the *initial permeability*, is of particular importance in communication systems, where the current is commonly very weak. The initial permeability of magnetic materials is nearly always much less than the permeability at somewhat higher flux densities.

Coils having magnetic cores are frequently used in communication work under conditions where there is a large d-c magnetization upon which is superimposed a small a-c magnetization. Under these conditions, one is interested in the inductance that is offered to the superimposed alternating current. This is called the *incremental inductance*, and the corresponding permeability of the magnetic material is termed the incremental permeability.

Incremental permeability, and hence incremental inductance, depend upon the magnitude of both d-c and a-c magnetizations, and upon the previous magnetic history of the core. When a core that has been

[1] Such calculations are discussed in "Components Handbook" (vol. 17, Radiation Laboratory Series), chap. 4, McGraw-Hill Book Company, Inc., New York, 1949.

quencies. The addition of cobalt to nickel-iron alloys introduces the possibility of obtaining substantially constant permeability up to moderate flux densities, combined with extremely low hysteresis loss at low flux densities and almost zero residual induction and coercive force (see page 17). Such alloys are termed perminvars, and also possess in large degree the high-permeability features of nickel-iron alloys such as permalloy.

Iron-cobalt alloys containing from 36 to 50 per cent cobalt are characterized by a saturation flux density appreciably higher than that of silicon steel. Such alloys also have higher incremental permeability at high d-c magnetizing forces than do other magnetic materials.

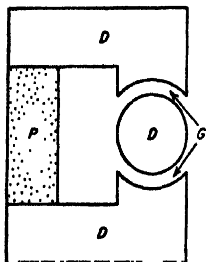


FIG. 2-5. Typical magnetic circuit involving a permanent magnet.

Magnetic cores that are nonconducting have been developed for use in radio-frequency coils.[1] They are composed of mixtures of ferrites, and have a resistivity so high that eddy-current losses are negligible in solid cores even at frequencies higher than 1 Mc. At the same time, such core material has a relatively high initial permeability, a value of 500 being typical. These nonconducting magnetic cores are not suitable for use in power transformers, however, as they saturate at low flux densities.

2-2. Permanent Magnets.[2] Permanent magnets now find many uses as a result of the development in recent years of improved permanent magnet materials. A typical system involving a permanent magnet is illustrated in Fig. 2-5. Here P is the permanent magnet, G is an air gap in which it is desired that the permanent magnet produce magnetic flux, and D denotes soft-iron pole pieces of low magnetic reluctance. In such an arrangement, the permanent magnet can be thought of as being a generator of magnetomotive force that acts on an external circuit (load) consisting of the magnetic circuit $DGDGD$ that is external to the permanent magnet.

Assume that the permanent magnet in the system of Fig. 2-5 is magnetized to saturation and that the magnetizing force is then removed. The resulting situation that exists in the magnet corresponds to a point somewhere on the part of the hysteresis loop lying in the upper left-hand quadrant of Fig. 2-3. This section of the hysteresis curve, shown enlarged in Fig. 2-6, is termed the demagnetization curve, and gives the principal characteristics of the permanent magnet. The flux density B_r

in the magnet for zero magnetizing force is termed the *residual induction*, while the demagnetizing force H_c which makes the flux density in the magnet zero is termed the *coercive force*.

For an operating condition of the magnet in Fig. 2-6 corresponding to point C', the flux density in the magnet is B', and the total flux generated by the permanent magnet is $B'A$ where A is the cross-sectional area of the permanent magnet. Also for the same operating point C', each unit length of the permanent magnet produces a magnetomotive force H'; hence the total magnetomotive force that is applied to the external circuit (*DGDGD* in Fig. 2-5) by the permanent magnet is $H'l$, where l is the length of the permanent magnet. The operating point C' accordingly assumes a position on the demagnetization curve such that $H'l/B'A$ equals the reluctance of the external magnetic circuit.

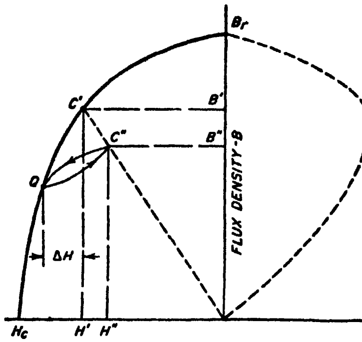


FIG. 2-6. Demagnetization and energy-product curves of a permanent magnet, showing minor hysteresis loop ciated with stabilization.

Design Principles. The magnetic energy developed by the permanent magnet in the external system *DGDGD* in Fig. 2-5 is proportional to the product $(B'A)$ $(H'l)$ of magnetic flux and magnetomotive force associated with the external circuit. Thus the magnetic energy available in the external circuit per unit volume of the permanent magnet is proportional to the product BH of the demagnetization curve, as plotted in Fig. 2-6.

It is now possible to state the principal design considerations of systems involving permanent magnets. First, the permanent magnet should be operated at a point on the demagnetization curve where the energy product BH is at or near its maximum; this operating point is a characteristic of the magnetic material involved, and defines a magnetomotive force H' per unit length and a flux density B' for the permanent magnet. Next, the cross section A of the magnet is given a value such that, when the flux density in the magnet has the value of B', the total flux $B'A$ will have the value desired for the external magnetic circuit. Finally, the length l of the permanent magnet is made such that $H'l$ will equal the magnetomotive force required to develop the required flux $B'A$ in the external magnetic circuit.

A permanent magnet operating at a point such as C' in Fig. 2-6 will have the flux density permanently changed when subjected to a transient action that momentarily reduces the flux density below B'. Thus assume that a transient demagnetizing current (or a momentary increase in reluctance) shifts the operating point from C' to Q. If this added effect is now removed, the operating point does not return to C'. Rather, it moves to a new point C'', corresponding to some new value of flux density B'' less than B', but such that the ratio $H''l/B''A$ still equals the reluctance of the external magnetic circuit; when this reluctance is linear, as when it arises from an air gap, then C'' lies on a straight line joining C' and the origin, as shown in Fig. 2-6. If the transient added force is applied a second time, the operating point will now return to Q, and upon removal of the added force will go back to C'', following the paths shown.

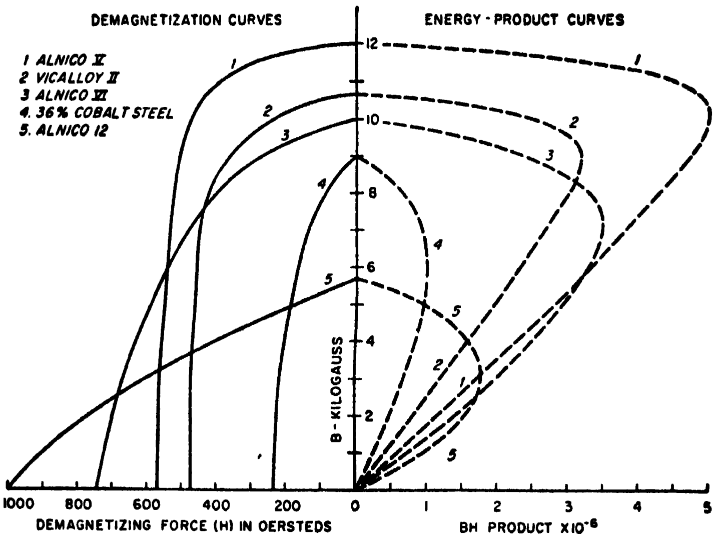


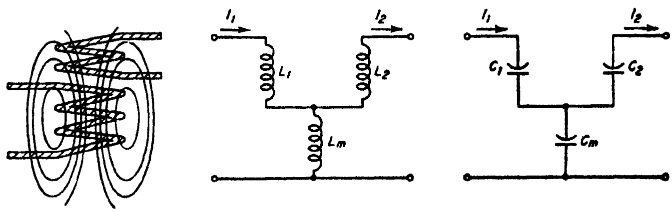
FIG. 2-7. Demagnetization and energy-product curves of typical permanent magnet materials.

Thus a permanent magnet system can be stabilized against added effects by initially subjecting the system to an added demagnetizing force ΔH; this will reduce the energy that the magnet supports in the system external to the magnet, but it prevents subsequent transient demagnetizing effects from producing a permanent change in the system provided their amplitudes do not exceed the demagnetizing force ΔH used in stabilization.[1] It will be noted that QC'' is a minor hysteresis loop analogous to the right-hand half of loop 1 in Fig. 2-3.

Permanent-magnet Materials.[2] Many different types of permanent-magnet materials have been developed; the characteristics of representative examples are illustrated in Fig. 2-7. The magnetic properties depend upon the composition and require proper cold working and heat-treatment to be fully developed. Heat-treatment is sometimes carried out in the presence of a strong magnetizing field; in this case the material when used should be magnetized in the same direction as when heat-treated.

The choice between different materials for a particular application is determined, not only by the energy product, but also by cost, by ease of fabricating, by whether the magnet is to be used in an external system of high or low reluctance, etc. In general, the better permanent-magnet materials are very difficult to work. Thus the alnicos (aluminum-nickel-iron alloys) are hard, weak, and brittle, and are commonly cast to approximate shape and then finished by grinding to exact size; they cannot be machined, drilled, or tapped.

2-3. Mutual Inductance and Coefficient of Coupling. *Mutual Inductance.* When two inductance coils are so placed in relation to each other



(*a*) INDUCTIVELY COUPLED (*b*) COUPLED BY COMMON INDUCTANCE (DIRECT COUPLING) (*c*) COUPLED BY COMMON CAPACITANCE

FIG. 2-8. Several simple methods of coupling two circuits.

that flux lines produced by current in one of the coils link with the turns of the other coil as shown in Fig. 2-8*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*, which is defined by the relation

$$\left.\begin{array}{l}\text{Mutual}\\ \text{inductance}\\ M \text{ in henrys}\end{array}\right\} = \frac{\left\{\begin{array}{l}\text{flux linkages in second coil}\\ \text{produced by current in first coil}\end{array}\right\}}{\text{current in first coil}} \times 10^{-8} \qquad (2\text{-}3)$$

$$= \frac{\left\{\begin{array}{l}\text{flux linkages in first coil}\\ \text{produced by current in second coil}\end{array}\right\}}{\text{current in second coil}} \times 10^{-8} \qquad (2\text{-}4)$$

Formulas (2-3) and (2-4) are equivalent and give the same value of mutual inductance. The flux linkages produced in the coil that has no current in it are counted just as though there were a current in this coil, so that the number of times a flux line would encircle an imaginary coil current is the number of linkages contributed by this particular line. In adding up the flux linkages it is important to note that different flux lines may conceivably link with the same coil in opposite directions, in which case

C 14

the total number of linkages is the difference between the sums of positive and negative linkages. The mutual inductance may therefore be positive or negative depending upon the direction of the linkages.

The problem of calculating mutual inductance is similar in all respects to the problem of computing inductance, and formulas have been worked out by which the mutual inductance can be calculated with good accuracy in all the ordinary types of configurations.

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. Thus, when all linkages are in the same direction, the total inductance of the series combination exceeds by $2M$ the sum of the individual inductances of the two coils.

Coefficient of Coupling. 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 the mutual inductance M that is actually present to this maximum possible value of mutual inductance is called the *coefficient of coupling*, which can therefore be expressed by the relation

$$\text{Coefficient of coupling} = k = \frac{M}{\sqrt{L_1L_2}} \tag{2-5}$$

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-cored coils 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.

General Case of Coupled Circuits. 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 capacitor, 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. 2-8. *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 impedance that are present in the two circuits.* That is,

$$k = \frac{Z_m}{\sqrt{Z_1Z_2}} \tag{2-6}$$

where Z_m is the impedance common to the two circuits, and Z_1 and Z_2 are the total impedances of the *same kind* in the two circuits. When applied to case b in Fig. 2-8, where the coupling is furnished by the common inductance L_m, the total inductances of the two circuits are $L_1 + L_m$ and $L_2 + L_m$, respectively, and Eq. (2-6) reduces to

$$\text{Coefficient of coupling } k \text{ for Fig. 2-8}b \quad \frac{L_m}{\sqrt{(L_1 + L_m)(L_2 + L_m)}} \tag{2-7}$$

In Fig. 2-8c the coupling element is a common capacitance C_m, and the coefficient of coupling is[1]

$$\text{Coefficient of coupling for Fig. 2-8}c = \frac{\sqrt{C_1C_2}}{\sqrt{(C_m + C_1)(C_m + C_2)}} \tag{2-8}$$

2-4. Skin Effect in Coils and Conductors at Radio Frequencies. The effective resistance offered by conductors to radio frequencies is considerably more than the ohmic resistance measured with direct currents. This is because of an action known as *skin effect,* which causes the current to be concentrated in certain parts of the conductor and leaves the remainder of the cross section to contribute little or nothing toward carrying the current.

A simple example of skin effect, and one that makes its nature clear, 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. 2-9. 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; this is because of the greater number of flux

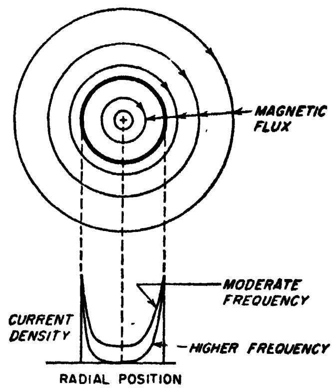


FIG. 2-9. Isolated round conductor, showing magnetic flux paths, and also typical current distributions.

[1] Equation (2-7) is derived as follows: In Fig. 2-8c, the primary circuit has C_1 and C_m in series and so has an equivalent capacitance of $C_1C_m/(C_1 + C_m)$ while the equivalent capacitance of the secondary is similarly $C_2C_m/(C_2 + C_m)$. The coupling reactance is $1/\omega C_m$, while the primary and secondary reactances are $(C_1 + C_m)/\omega C_1C_m$ and $(C_2 + C_m)/\omega C_2C_m$, respectively. The coefficient of coupling is then

$$k = \frac{1/\omega C_m}{\sqrt{\dfrac{C_1 + C_m}{\omega C_1C_m}\,\dfrac{C_2 + C_m}{\omega C_2C_m}}}$$

which reduces to Eq. (2-7).

linkages existing in the central region. At radio frequencies, the reactance of this extra inductance is sufficiently great to affect seriously the flow of current, most of which 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 true or effective resistance is increased, since in effect the useful cross section of the wire is very greatly reduced. The types of current distribution obtained in typical cases of skin effect in a round wire are shown in Fig. 2-9.

When skin effect is present, the current is always redistributed over the conductor cross section in such a way as to make most of the current flow where it is encircled by the smallest number of flux lines. This general principle controls the distribution of current, irrespective of the shape of the conductor involved. Thus, with a conductor consisting of a thin flat strip, such as shown in Fig. 2-10, the current flows primarily along the edges, where it is surrounded by the smallest amount of flux, and the true or effective resistance will be high because most of the strip carries very little current. This illustration makes clear that it is not the amount of conductor surface that determines the resistance to alternating current, but rather the way in which the conductor material is arranged.

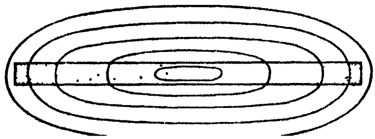


FIG. 2-10. Flux lines and current distribution in a thin strip at radio frequency, showing how skin effect causes the current to concentrate at the edges of the strip. The current density is indicated by the density of the shading.

The ratio that the effective a-c resistance bears to the d-c resistance of a conductor is commonly called the *resistance ratio.* It increases with frequency, with conductivity of the conductor material, and with the size of conductor. This results from the fact that a higher frequency causes the extra inductance at the center of the conductor to have a higher reactance. Similarly, a greater conductivity makes the reactance of the extra inductance of more importance in determining the distribution of current, while a greater cross section provides a larger central region. It is to be noted, however, that a larger conductor always has less radio-frequency resistance than a smaller one because, although the ratio of a-c to d-c resistance is less favorable, this is more than made up by the greater amount of conductor cross section present.

Skin Effect at High Frequencies.[1] When the frequency is sufficiently high, substantially all of the current in a conductor is confined to a region very close to the surface. The current density then falls off with depth from the surface in accordance with the relation

[1] An excellent discussion of skin effect at very high frequencies is given by Harold A. Wheeler, Formulas for the Skin Effect, *Proc. IRE*, vol. 30, p. 412, September, 1942.

$$\frac{\text{Current at depth } z}{\text{Current at the surface}} = \qquad (2\text{-}9)$$

Here z and δ are in the same units, and δ is a quantity called the ***skin depth*** that is given by the equation

$$\delta = 5033 \sqrt{\frac{\rho}{\mu f}} \qquad (2\text{-}10)$$

where δ = skin depth, cm
ρ = resistivity of conductor, ohms per centimeter cube
f = frequency, cycles
μ = magnetic permeability of core material (permeability of air equals unity), for low flux densities (that is, μ is the initial permeability)

For copper at 20°C this reduces to

$$\text{Skin depth of copper in cm} = \frac{6.62}{\sqrt{f}} \qquad (2\text{-}11)$$

At 1 Mc the skin depth in copper is thus 0.0066 cm, or 0.0026 in. The phase of the current at depth z lags the current at the surface by z/δ radians. At a depth from the surface corresponding to one skin depth, the current density has dropped to 36.8 per cent of the value at the surface, and the phase of the current lags the current at the surface by 1 radian.

Equation (2-9) is valid whenever the radius of curvature of the conductor surface is at least several times the skin depth, provided the effective thickness of the conductor is at the same time at least three or four skin depths.

The power loss associated with the current flowing under any particular portion of the conductor surface is the same as though this current were uniformly distributed down to a depth δ. Thus, in an isolated round wire, where the current is uniformly distributed over the surface, the effective resistance at high frequencies is the d-c resistance of a hollow cylindrical shell having the same outer diameter as the wire and possessing a thickness δ. The d-c resistance of a strip of surface one skin depth thick, one centimeter long, and one centimeter wide is sometimes called the ***surface resistivity;*** it is the resistivity that is offered to the flow of current at very high frequencies.

Proximity Effect—Skin Effect in Coils. When two or more adjacent conductors are carrying current as in a coil, the current distribution in any one conductor is affected by the magnetic flux produced by the adjacent conductor as well as by the magnetic flux produced by the current in the conductor itself. This effect, termed ***proximity effect,*** ordinarily causes the true or effective resistance to be greater than in the case of simple skin effect and is particularly important in radio-frequency inductance coils.

The current distribution under conditions where proximity effect is present follows the same law as for simple skin effect; i.e., the current density is greatest in those parts of the conductor encircled by the smallest number of flux lines. This is illustrated in Fig. 2-1, where the approximate current density is illustrated by relative shading.

Litz Wire. The effective a-c resistance of a conductor can be made to approach the d-c resistance at low and moderate radio frequencies by forming the conductor from a number of strands of small enameled wire connected in parallel at their ends, but insulated throughout the rest of their length and thoroughly interwoven. If the stranding is properly done, each wire will, on the average, link with the same number of flux lines as every other wire, and the current will divide evenly among the strands. If at the same time each strand is of small diameter, it will have relatively little skin effect over its cross section, so all of the material is equally effective in carrying the current. Such a stranded cable is called a *litz* conductor.

Practical litz conductors are very effective at frequencies below about 1000 kc, but as the frequency becomes higher the benefits disappear. This is because irregularities of stranding, and capacitance between the strands, cause a failure to realize the ideal condition at very high frequencies.

2-5. Capacitors and Dielectrics. A capacitor is formed wherever an insulator (i.e., dielectric) separates two conductors between which a difference of potential can exist.

Capacitor Losses and Their Representation. A perfect capacitor when discharged gives up all the electrical energy that was supplied to it in charging. Actual capacitors never realize this ideal perfectly but, rather, dissipate some of the energy delivered to them. Most of the loss in ordinary capacitors occurs in the dielectric, although at very high frequencies skin effect also causes an appreciable loss to occur in the capacitor leads and electrodes. At very high voltages corona may occur and contribute to the loss.

The merit of a capacitor from the point of view of freedom from losses is usually expressed in terms of the power factor of the capacitor.[1] The power factor represents the fraction of the input volt-amperes that is dis-

[1] The merit of a capacitor or of a dielectric can also be expressed in terms of the angle by which the current flowing into the capacitor fails to be 90° out of phase with the applied voltage. This angle is termed the *phase angle* of the capacitor. The power factor is the sine of the phase angle. The tangent of the phase angle is termed the *dissipation factor*. The reciprocal of the dissipation factor is termed the capacitor Q and is the ratio of the capacitor reactance to the equivalent series resistance. With ordinary dielectrics, the phase angle is so small that the power factor, the dissipation factor, and the reciprocal of capacitor Q are for all practical purposes equal to each other and to the phase angle expressed in radians. Thus a power factor of 0.01 represents a phase angle of 0.573° and a capacitor Q of 100.

sipated in the capacitor. To the extent that the losses in the capacitor are a result of dielectric losses, the power factor of the capacitor is also the power factor of the dielectric and is practically independent of the capacitor capacitance, the applied voltage, the voltage rating, or the frequency (unless polar effects are involved). Values of power factor of some typical dielectrics are given in Table 2-1.

TABLE 2-1
CHARACTERISTICS OF TYPICAL DIELECTRICS AT RADIO FREQUENCIES WITH NORMAL ROOM TEMPERATURE

Material	Dielectric constant	Power factor
Air	1.00	0.000
Mica (electrical)	5–9	0 0001–0.0007
Glass (electrical)	4.5 7.00	0.002–0.016
Bakelite derivatives	4.5 7.5	0.02–0.09
Wood (without special preparation)	3-5	0.03–0 07
Mycalex	8	0.002
Steatite materials	6.1	0.002–0.004
Polystyrene	2.4–2.9	0.0002
Polyethelene	2.3	0.00015–0.0003
Rutile (titanium dioxide)	90–170	0.0006

Although the power factor of a capacitor is determined largely by the type of dielectric used in the capacitor, it is also affected by the conditions under which the dielectric operates. In particular, the power factor tends to become higher as the temperature is raised, and is likewise adversely affected by high humidity and by the absorption of moisture.

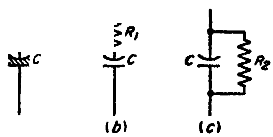


FIG. 2-11. Representation of imperfect capacitor by a perfect capacitor of same capacitance with series resistance, and by a perfect capacitor ith shunt resistance.

Equivalent Series and Shunt Resistance. The action of a capacitor in an electrical circuit is taken into account by replacing the actual capacitor with a perfect capacitor associated with a resistance. This resistance may be connected in series, as in Fig. 2-11*b*, or in parallel, as in Fig. 2-11*c*. The value of the series or shunt resistance is so selected that the power factor of the perfect capacitor associated with the resistance is the same as the power factor of the actual capacitor. The value of the series resistance R_1 can be computed in terms of the power factor, capacitor capacitance C, and frequency f in the usual way, and when the power factor is low (i.e., when $R_1 << 1/\omega C$), then R_1 is given to a high degree of accuracy by the equation

$$\text{Series resistance} = R_1 = \frac{\text{power factor}}{2\pi f C} \qquad (2\text{-}12)$$

In the same way, the shunt resistance that can be used to represent the actual losses of the capacitor is related to the power factor, capacitance, and frequency to a high degree of accuracy by the equation

$$\text{Shunt resistance} = R_2 = \frac{1}{(2\pi fC)\ (\text{power factor})} \tag{2-13}$$

Polar and Nonpolar Dielectrics. Molecules of some dielectrics are polar, while other dielectrics consist of molecules that are not polar. In the case of polar molecules, the dielectric constant under a-c conditions is increased as a result of the rotation of the polar molecules under the influence of the applied voltage. The extent to which this polar action is effective depends, however, upon the frequency and the temperature. Thus, if the frequency is made sufficiently high, the polar molecules are not able to follow the alternations of the applied field, and the dielectric constant drops. Moreover, the frequency at which this transition occurs is less the lower the temperature. As a result, temperature and frequency affect the capacitance of a capacitor possessing a polar dielectric in the manner shown in Fig. 2-12.

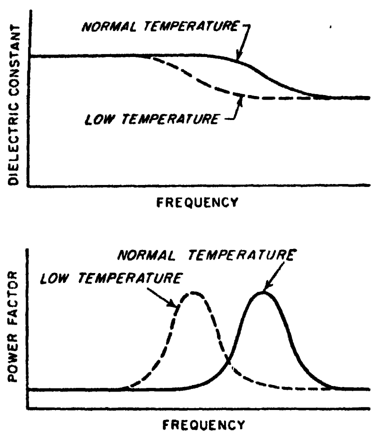


FIG. 2-12. Variation of dielectric constant and power factor of a polar dielectric as a function of frequency, for two temperatures.

The power factor of a polar dielectric shows a pronounced peak when under conditions where the dielectric constant corresponds to partial polar action, as shown in Fig. 2-12. The power factor of a polar dielectric hence becomes quite large at certain combinations of temperature and frequency.

Nonpolar molecules do not exhibit these changes in dielectric constant under temperature and frequency changes. The power factor of nonpolar dielectrics likewise does not exhibit peaks of loss such as shown in Fig. 2-12.

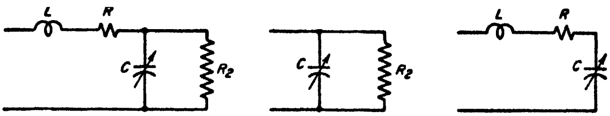
2-6. Capacitors for Electronics. In electronics the principal uses made of capacitors are for tuning resonant circuits, for blocking d-c voltages from parts of an electrical circuit while permitting alternating voltages to pass through, for obtaining transients with specified time constants, and for by-passing or short-circuiting alternating voltages. By-pass capacitors are frequently but not always subjected to a d-c potential.

A wide variety of dielectrics are used in capacitors designed for radio work, and new types are continually finding important applications. Among the types of importance are air; solid dielectrics such as mica, plastic films, certain ceramics, and paper; and electrolytic films.

Capacitors with Air Dielectric. Air dielectric finds its principal use in variable capacitors for tuning resonant circuits.

Although air is a perfect dielectric with zero power factor, air capacitors have losses because of the insulating material used to mount the two sets of plates, and also because of the skin-effect resistance of the leads, plates, rods, and washers, through which the capacitor current flows.

An air-dielectric capacitor can be represented by the equivalent electrical network in Fig. 2-13*a*. Here C is the capacitance of the capacitor



(*a*) ACTUAL EQUIVALENT CIRCUIT (*b*) EQUIVALENT CIRCUIT AT LOW FREQUENCIES (*c*) EQUIVALENT CIRCUIT AT HIGH FREQUENCIES

FIG. 2-13. Equivalent electrical circuits of a variable air condenser.

while R_2 is the equivalent shunt resistance introduced by the presence of the solid dielectric. . The capacitor inductance L in Fig. 2-13*a* takes into account the magnetic flux associated with the current flowing in the capacitor; it is proportional to the physical dimensions of the capacitor. The resistance R represents the series resistance of the leads, washers, connecting rods, etc. It increases with frequency as a result of skin effect and is proportional to the square root of the frequency at high frequencies.

At low and moderate frequencies the effects of the inductance L and the series resistance R are negligible, and the capacitor equivalent circuit reduces to Fig. 2-13*b*. At very high frequencies, on the other hand, the power loss in R becomes very much larger than in R_2, and the equivalent circuit has the form shown in Fig. 2-13*c*.

At very high frequencies the reactance of the series inductance L is not negligible compared with the reactance of the capacitor capacitance. This causes the apparent capacitance of the capacitor as observed at the terminals to be greater than the actual capacitance according to the relation[1]

$$\text{Apparent capacitance} \quad \frac{C}{1 - \omega^2 LC} \tag{2-14}$$

where $\omega = 2\pi$ times frequency, and L and C are as shown in Fig. 2-13.

[1] This results from the fact that, neglecting losses,

$$\text{Reactance at terminals} = \frac{1}{\omega C_{app}} = \left(\frac{1}{\omega C} - \omega L\right)$$

Solving for the apparent capacitance C_{app} gives Eq. (2-14).

Capacitors with Solid Dielectrics. Solid dielectrics are used in most fixed and in some adjustable capacitors. The dielectrics most commonly employed include mica, paper, plastic films, and ceramics.

Mica is characterized by low electrical losses, stability, high leakage resistance to d-c voltages, and high voltage strength. It is, however, relatively expensive. Mica capacitors find their chief use in small fixed capacitors for by-passing radio-frequency currents or blocking off d-c voltages, and in resonant circuits or in filters where a stable low-power-factor capacitor is required.

In capacitors employing paper as the dielectric the electrodes are either aluminum foil or are metal films evaporated directly on to the paper. In either case, the assembly is rolled into a bundle which is then vacuum-treated, impregnated with oil or wax, and sealed against moisture. Paper capacitors are inexpensive in proportion to capacitance, and are relatively compact in proportion to voltage rating. Such capacitors are used primarily for by-pass and blocking purposes. The power factor of paper capacitors is of the order of 0.5 per cent, and although the leakage current when subjected to direct voltages is somewhat greater than that of mica capacitors, it is not large.

Thin plastic films have been developed that are suitable for use as a capacitor dielectric in place of paper. Capacitors of this type using polystyrene dielectric have electrical qualities such as power factor, dielectric absorption, and insulation resistance superior even to mica capacitors.

Ceramics based on titanium dioxide mixtures find extensive use as dielectrics of small capacitors.[1] Dielectrics of this type are characterized by a high dielectric constant, a low to very low power factor, and a very high voltage rating. The temperature coefficient of such capacitors depends upon the actual ceramic mixture used and can be made either negative or positive as desired. Ceramic capacitors are used extensively for blocking and by-pass purposes where small mica capacitors have heretofore been employed, and have advantages of compactness and high voltage ratings. Ceramic dielectric capacitors have also found a wide field of usefulness in resonant circuits and other similar applications, where a negative temperature coefficient provided by a ceramic capacitor can be used to compensate for the positive temperature coefficient of associated coils and of capacitors of other types.

Capacitors with solid dielectric can be represented by the same equivalent electrical circuit shown in Fig. 2-13 for air-dielectric capacitors. The only difference is that all the capacitance C in this equivalent circuit is now associated with solid dielectric. As a result, at low and moderate frequencies the capacitor power factor almost exactly equals the power

[1] A survey of such ceramics is given by B. H. Marks, Ceramic Dielectric Materials, *Electronics*, vol. 21, p. 116, August, 1948.

factor of the dielectric and is independent of the capacitance of the capacitor and also of the frequency except in so far as polar molecules affect the behavior of the dielectric. At very high frequencies, the power factor increases with increasing frequency as a result of the skin-effect losses in leads and conductors. Also, the apparent capacitance at very high frequencies drops off because of the series inductance, in accordance with Eq. (2-14).

The voltage rating of capacitors with solid dielectrics is subject to two basic limitations: (1) If the applied voltage exceeds the insulation strength of the dielectric, the dielectric will spark through or at least deteriorate rapidly. (2) The temperature of the capacitor must not be permitted to rise excessively as a result of dielectric losses. This second limitation is the ruling one for all except d-c voltages and for very low frequencies. Inasmuch as the relationship between losses and temperature rise depends upon the design of the capacitor with respect to such matters as heat removal, it is not possible to give any general rules regarding voltage ratings. It is to be noted, however, that the voltage rating will drop rapidly as the frequency increases because of the increase in loss with frequency. Thus a particular low-loss air-cooled mica capacitor capable of standing 10,000 volts at low-frequencies was found by test to have a rating of 180 volts at a frequency of 10 Mc. Special cooling methods, such as the use of an air blast, will increase greatly the rating on a capacitor, and water cooling is still more effective.

Electrolytic Capacitors. The electrolytic capacitor makes use of the fact that certain metals, notably aluminum and tantalum, when placed in a suitable solution and made the positive electrodes, form a thin insulating surface film. This film is capable of withstanding considerable voltage and has a high electrostatic capacitance per unit area of film. It is the result of electrochemical action, and is formed by applying positive voltage to the electrode. The thickness of the film, and hence also the capacitance obtained per unit area of surface, depend largely upon the voltage used in this forming process. Typical voltage ratings of electrolytic capacitors range from 25 up to about 500 volts. Constructional details vary but, typically, the electrodes are of etched aluminum foil, thus giving maximum surface area. They are separated by paper or gauze, saturated with an electrolyte that is commonly a fudgelike solid, and the entire assembly is wound into a roll and mounted in a waxed cardboard tube or box.

Electrolytic capacitors are widely used for filter and by-pass purposes in situations where a superimposed d-c voltage is present. Compared with capacitors of the solid dielectric type, electrolytic capacitors have a very high power factor and appreciable leakage conductance to the superimposed d-c potential; they also vary in capacitance and loss with time, frequency, and temperature. However, for many purposes these features are unimportant and, in proportion to capacitance and voltage rating, electrolytic capacitors are the least expensive and most compact available. They are, however, subject to progressive deterioration with time and so have limited life, and their dependability is appreciably less than that of paper capacitors designed for the corresponding applications.[1]

2-7. Coils for Resonant Circuits. Coils intended for use in resonant circuits must have very low losses and small distributed capacitance. Both air-cored and magnetically cored coils are used for resonant circuits, with the choice depending upon circumstances.

Methods of Expressing Coil Losses—Coil Q. The principal causes of energy loss in air-cored coils are skin effect in the conductor, proximity effect resulting from the interaction between nearby turns, dielectric losses associated with the distributed capacitance of the coil, and eddy-current losses in shields and other neighboring metallic objects present within range of the magnetic field of the coil. In the case of coils with magnetic cores, the principal cause of energy loss is usually core loss, although factors such as skin-effect resistance of the wire and also dielectric loss are sometimes likewise of significance.

For purposes of circuit analysis the coil losses are commonly expressed in terms of an equivalent resistance, which when placed in series with the coil inductance will account for all the power losses actually observed. The most convenient way to express the merit of the coil is, however, in terms of the ratio of the reactance ωL of the coil to this equivalent series resistance R. This ratio approximates the reciprocal of the coil power factor, and is usually referred to by the symbol Q; that is,

$$Q = \frac{\text{coil reactance}}{\text{equivalent series resistance}} = \frac{\omega L}{R} \tag{2-15}$$

It is convenient to express the characteristics of a coil in terms of Q because the Q in the operating range of the coil usually varies only moderately with frequency; moreover, the value of Q corresponding to a good coil is substantially the same irrespective of the frequency for which the coil was designed. The tendency for the coil Q to remain constant with frequency arises from the fact that, as the frequency increases, all the losses also increase, so that the *ratio* of coil reactance to resistance tends to be much more nearly constant with frequency than is either the reactance or the resistance of the coil.

Distributed Capacitance of Coils. In a coil there are small capacitances between adjacent turns, between turns that are not adjacent, between terminal leads, between turns and ground, etc. Some of the different capacitances that may exist in a typical air-cored coil are shown in Fig.

[1] By substituting tantalum for the less expensive aluminum foil electrodes, it is possible to increase greatly the reliability; see M. Whitehead, Tantalum Electrolytic Capacitors, *Bell Lab. Record*, vol. 28, p. 448, October, 1950.

2-14. Each of the various capacitances associated with the coil stores a quantity of electrostatic energy that is determined by the capacitance involved and the fraction of the total coil voltage that appears across it. The total effect that the numerous small coil capacitances have can be represented to a high degree of accuracy by assuming that they can be replaced by a single capacitor of appropriate size shunted across the coil terminals. This equivalent capacitance is called either the distributed capacitance or the self-capacitance of the coil; it causes the coil to show parallel resonance effects under some conditions (see Sec. 3-2).

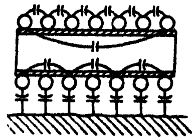


FIG. 2-14. Some of the coil capacitances that contribute to the distributed capacitance of a single-layer coil.

In multilayer coils the distributed capacitance will be high unless arrangements are used that prevent turns from different parts of the winding from being located close to each other. Thus, in the two-layer winding shown in Fig. 2-15*a*, in which the turns are numbered in order, the first and last turns are adjacent; the capacitance between the turns at opposite ends of the winding then stores an undesirably large amount of electrostatic energy. This can be avoided by the use of the bank winding, shown at *b*. Here the adjacent turns represent parts of the coil that are close together electrically, while the ends of the winding, which are far apart electrically, are also far apart physically. Alternative approaches consist in using many layers with few turns per layer, as in Fig. 2-13*c*, or in spacing the layers, as in Fig. 2-15*d*. The common "universal" multilayer coil represents a convenient mechanical method of utilizing these principles to achieve low distributed capacitance.

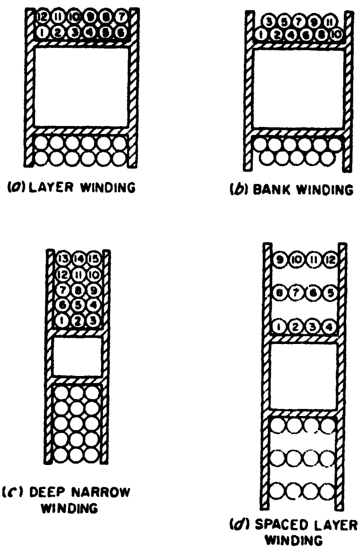


FIG. 2-15. Several types of multilayer windings.

The distributed capacitance of a coil that is to be used in a resonant circuit must be small. This is because the distributed capacitance limits the highest frequency to which the coil can be tuned, and also introduces losses that become serious at the higher frequencies. These losses are dielectric losses occurring in the coil form, in the wire insulation, and in any other dielectric that may be in the electrostatic fields associated with the coil.

Air-cored Coils for Resonant Circuits. Air-cored coils are widely used in radio receivers and almost universally used for the resonant circuits of radio transmitters. Single-layer coils are generally employed for frequencies above 500 to 1500 kc, while at lower frequencies multilayer coils are typical, as they give the desired inductance compactly. Multilayer coils, generally of the bank-wound type, also find some use at broadcast frequencies (535 to 1600 kc).

In designing a single-layer coil, the highest Q in proportion to size is obtained when the length of the winding is somewhat less than the diameter of the coil.[1] The number of turns required is then determined by

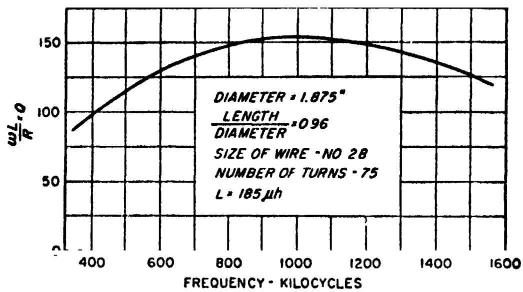


FIG. 2-16. Variation of Q with frequency for a typical air-cored coil.

the exact ratio of length to diameter that is selected, by the diameter, and by the inductance desired; when these factors are all settled, the optimum wire size corresponds to a conductor diameter that is between 0.5 and 0.75 times the distance between the centers of adjacent turns. If one compares the Q of two coils having the same inductance and the same ratio of length to diameter but different physical size, then the coil that is larger will have the highest Q provided it is wound with wire of optimum size.

The design of multilayer coils is more involved than that of single-layer coils because of the increased number of variables. In general, best results are obtained if the coil is relatively "loose," i.e., if the copper occupies only a small fraction of the actual winding cross section. Again, as in the case of single-layer coils, larger physical size will result in a higher Q associated with a given inductance value, and also requires a larger wire.

The Q of a typical air-cored coil varies with frequency in the manner illustrated in Fig. 2-16. With increasing frequency, the Q first rises slowly with frequency, then goes through a broad maximum, and finally drops at very high frequencies. The rise is due to the fact that the

[1] A discussion of coil losses under idealized conditions in which dielectric effects are neglected is given by G. W. O. Howe, The Q Factor of Single-layer Coils, *Wireless Eng.*, vol 26, p. 179, June, 1949. An excellent discussion of coils for high frequencies is given by D. Pollack, The Design of Inductances for Frequencies between 4 and 25 Megacycles, *Trans. AIEE*, vol. 56, p. 1169, September, 1937.

inductive reactance of a coil is proportional to frequency, whereas the resistance due to skin effect cannot increase faster than the square root of the frequency; hence the ratio $Q = \omega L/R$ tends to rise with increasing frequency. If skin effect accounted for all the losses, the Q at very high frequencies would be proportional to the square root of frequency. However, dielectric losses arising from the coil form, the cotton or enamel insulation on the wire, etc., give rise to a resistance in series with the coil that is proportional to the cube of the frequency. At very high frequencies these dielectric losses become comparable to the skin-effect losses, and cause the coil Q to drop off.

The best conductor to use in an air-cored coil depends upon the frequency and coil design. In general, solid wire is used for frequencies above 1500 kc. Litz wire will give lower losses than the corresponding solid wire at frequencies below about 500 kc and will give some advantage for small multilayer coils in the frequency range 500 to 1500 kc.

A value of Q in the range 50 to 200 is typical of a good fairly small air-cored coil such as would be used for resonant circuits in a radio receiver. A Q of 10 or 20 is considered to be quite low, while Q values in excess of 300 are high and can ordinarily be achieved only by the use of coils that are physically large, such as are used in radio transmitters. These numbers are applicable for coils of all frequency ranges and inductance values.

Magnetic-cored Coils for Resonant Circuits. Coils with magnetic cores find extensive use at radio frequencies. The principal problem involved in using magnetic cores at radio frequencies is that of preventing eddy-current losses in the core material from becoming excessive. The permeability of magnetic materials does not drop off with frequency until the frequency is of the order of 10^{11} cycles. The hysteresis loss is proportional to the frequency, but since the coil reactance is likewise proportional to the frequency, the Q is not adversely affected by hysteresis loss at radio frequencies. In contrast, the eddy-current loss for a given core is proportional to the square of the frequency, whereas the reactance is proportional only to the frequency. Thus when the frequency becomes sufficiently high, eddy-current losses dominate the situation, and the coil Q drops.

The eddy-current losses can be kept low at high frequencies by arranging the magnetic material in the form of very fine particles or dust, produced either by chemical or mechanical means. These particles are coated with an insulating film, mixed with a suitable proportion of binder, pressed to the desired shape, and baked. In this way one obtains a core in which the individual magnetic particles are very finely subdivided, with resulting low eddy-current losses; such an arrangement is often called a "dust" or powder core. It is possible to make "dust" cores which have low losses at frequencies as high as 150 Mc.

The details of the magnetic core depend upon the application for which the coil is intended, and the frequency range over which the core is to operate. At audio and the lower radio frequencies, the core material is commonly made in the form of rings (toroidal core), so that a closed magnetic path may be obtained. Cores designed for use at these frequencies usually have an effective initial permeability that is quite high, such as 75 to 125, corresponding to a core involving relatively coarse particles combined with a minimum of insulating material and binder. At higher frequencies, the usual practice is to employ a single-layer winding of fine wire on a form that snugly fits an open core made in the form of a cylindrical slug with a large length/diameter ratio, as in Fig. 2-17. Also, as the frequency is increased, the effective permeability that it is practical to employ in a core becomes less because the size of the particles of magnetic material in the core must be reduced, and the proportion of core material to binder and insulation becomes proportionally less. Thus cores that are suitable for use in coils operating at 100 Mc have, typically, an effective permeability of only 2 to 4, while cores for use at frequencies around 1 Mc have permeabilities from 10 to 30.

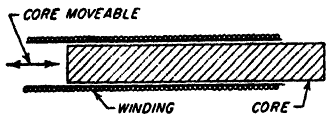


FIG. 2-17. Coil with slug-type magnetic core arranged so that the inductance can be varied by moving the core.

The particular magnetic material used likewise depends on frequency. For audio and the lower radio frequencies molybdenum permalloy is common; at the higher radio frequencies it is customary to use iron or magnetite, a natural iron oxide.

An alternative means of obtaining a magnetic core with low eddy-current losses is to employ a nonconducting magnetic material such as mentioned on page 16. Such material is suitable for use up to frequencies above 20 Mc; however, above some limiting frequency the dielectric and residual loss effects in the nonconducting magnetic material may adversely affect the behavior. In frequency ranges for which they are suitable, nonconducting magnetic cores result in coils having Q's as high as, or higher than, values typical of dust cores; at the same time a nonconducting core possesses considerably greater permeability than can be used in a dust core at the same frequency, and so has the advantage of compactness.[1]

Magnetic cores are particularly desirable when it is necessary to obtain a reasonable Q such as 25 to 100 in a very compact coil. They find extensive use in radio receivers. When a magnetic-cored coil is used in a resonant circuit, it is customary to employ a fixed tuning capacitance; the resonant frequency is then adjusted by varying the position of the slug

[1] A discussion of the properties obtainable in coils employing nonconducting ferrite cores is given by Strutt, *loc. cit.*

core, as indicated in Fig. 2-17. Such *permeability tuning*, as it is called, represents a means of tuning a resonant circuit that is often preferred to the alternative arrangement consisting of a fixed air-cored coil and a variable capacitance.

Radio-frequency Choke Coils. A radio-frequency choke coil is an inductance designed to offer a high impedance to alternating currents over the frequency range for which the coil is to be used. This result is obtained by making the inductance of the coil high and the distributed capacitance low, and by so proportioning that the inductance is in parallel

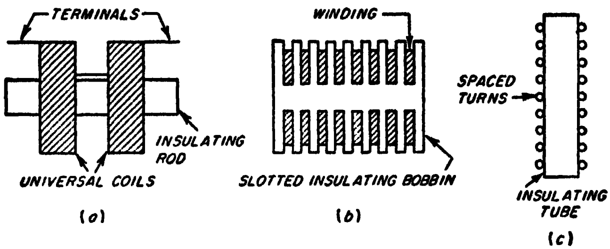


FIG. 2-18. Typical examples of radio-frequency choke coils.

resonance with the distributed capacitance somewhere in the desired operating range of frequencies.

A typical radio-frequency choke coil consists of one or more universal-wound coils mounted on an insulating rod, or of a series of "pies" wound in deep narrow slots in a slotted bobbin. A long single-layer solenoid is likewise sometimes used. Examples of radio-frequency choke coils are shown in Fig. 2-18.

The performance obtainable from a radio-frequency choke can generally be improved by the proper use of slug-type magnetic cores, which increase the inductance, and hence the impedance of the coil, without materially affecting the distributed capacitance.

2-8. Shielding of Magnetic and Electrostatic Fields. Under many conditions it is necessary to confine magnetic and electrostatic fields to a restricted space. This result is accomplished by using a shield composed of suitable material to enclose completely the space to be shielded.

The Shielding of Magnetic Flux at Radio Frequencies; Conducting Shields. The most practical shield for magnetic flux at radio frequencies is made of material having low electrical resistivity, such as copper or aluminum. Magnetic flux in attempting to pass through such a shield induces voltages in the shield which give rise to eddy currents. These eddy currents oppose the action of the flux, and in large measure prevent its penetration through the shield. In this way the flux is restricted to the interior of the shield, as illustrated in Fig. 2-19*c*.

To be effective a conducting shield should have a thickness *a* that is at least several times the skin depth δ as defined by Eq. (2-10). Under these conditions, and assuming that the radius of curvature of the shield is large compared with the skin depth, the ratio of the tangential components of the magnetic field intensities existing on the two sides of the shield is[1]

$$\text{Ratio of magnetic fields} = \epsilon^{a/\delta} \qquad (2\text{-}16)$$

Since the energy associated with a magnetic field is proportional to the square of the field intensity, the attenuation in decibels of the tangential component that is introduced by the conducting magnetic shield is

$$\text{Shield attenuation} = 8.69\,\frac{a}{\delta} \quad \text{db} \qquad (2\text{-}17)$$

Joints which interfere with the eddy currents by adding resistance to the eddy-current paths greatly reduce the effectiveness of a conducting shield. However, a joint parallel to the lines of current flow does not adversely affect the shielding unless it results in an open hole. This is true even if there is failure to make contact, so that the shield lacks continuity. These effects of joints are explained by the fact that the shielding is produced by the eddy currents; if the eddy currents are not disturbed, then the shielding resulting from action is not affected.

Power is dissipated in a conducting shield because the eddy currents must flow through the resistance of the shield material. When the thickness of the shield is considerably greater than the skin depth, the power loss in the shield can be determined by making use of the fact that the total magnitude I of the eddy currents in a strip of shield 1 cm wide is related to the density B in lines per square centimeter of the tangential component of flux that is adjacent to the surface of that part of the shield according to the relation

$$I = \frac{10B}{4\pi} \qquad (2\text{-}18)$$

The current I flows along the surface of the shield in a direction that is at right angles to the flux lines adjacent to the shield. For purposes of calculating power dissipation, this current can be considered as uniformly distributed to a thickness of one skin depth; it therefore encounters a resistance that is the surface resistivity of the material as calculated by skin-effect considerations (see page 23). Thus the total power loss in a shield can be obtained by first determining by some means the distribution of the tangential component of the magnetic flux adjacent to the surface of the shield. The distribution of current over the surface of the shield is next obtained with the aid of Eq. (2-18). The energy loss in each square

[1] The effect of a conducting shield on the component of magnetic field that is normal to the shield follows a different law, see B. Boston, Screening at V.H.F., *Wireless Eng.*, vol. 25, p. 221, July, 1948.

centimeter of shield surface is then determined by assuming that this current flows through a d-c resistance corresponding to a conductor that is one skin depth thick. The power consumed by a conducting shield is derived from the source of energy producing the magnetic field.

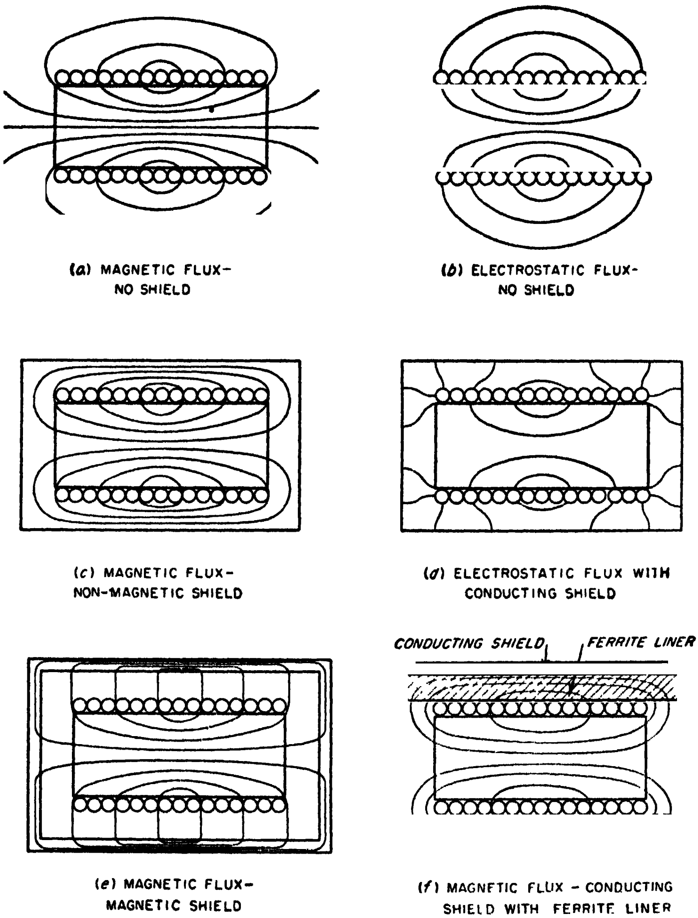


FIG. 2-19. Paths of electrostatic and magnetic-flux lines about the same coil with and without magnetic and nonmagnetic shields.

Shielding of D-C and Low-frequency Magnetic Fields; Magnetic Shields. When shielding against unidirectional magnetic fields is required, a shield composed of magnetic material is employed. Such a shield tends to short-circuit the flux lines which attempt to extend through the shield, as shown in Fig. 2-19*e*. The effectiveness of a magnetic shield is directly proportional to the thickness of the shield, since the reluctance that the shield offers to magnetic flux is inversely proportional to thickness. Joints or air gaps which add reluctance to the flux paths must be avoided. The degree of shielding achieved by a given total thickness of material can be increased by dividing the given thickness of magnetic material into two or more concentric shields separated by air spaces.

Magnetic shields must have high initial permeability to be effective. They are accordingly composed of high-permeability alloys such as permalloy; steel or iron is not a satisfactory material because of its low initial permeability. Since the desirable magnetic properties of permalloy and similar materials are adversely affected by mechanical strains, such as are introduced by drilling, punching, bending, etc., magnetic shields must be properly heat-treated *after* fabrication, to relieve these strains and develop the desirable magnetic properties.

Magnetic shields can be used for shielding alternating fields as well as d-c fields. In particular, they find extensive use at audio and power frequencies, particularly 60 cycles, where conducting shields would have to be excessively thick to be effective. The shielding action of a magnetic shield at these lower frequencies is achieved in part because of the short-circuiting action of the magnetic material on magnetic flux and in part because of eddy currents which cause the shield to act simultaneously as a conducting shield.

Magnetic shields of high-permeability material are also more effective at radio frequencies than are copper or aluminum shields. At these higher frequencies they act as conducting shields, but because of their high permeability have less skin depth. Thus the shielding obtained with a given thickness of material is greater. However, conducting material such as copper is less expensive per pound than magnetic material such as permalloy, is easier to fabricate, and requires no heat-treatment. Hence nonmagnetic conducting shields are generally used in preference to magnetic shields for alternating fields when the frequency is high enough so that the required degree of shielding can be obtained with a reasonable thickness of conducting shield; the only practical exception is when a conducting shield employs a liner of nonconducting magnetic material, as discussed below.

Electrostatic Shielding. Electrostatic shielding is obtained by enclosing the space to be shielded by a conducting surface. Accordingly, the magnetic and conducting shields for magnetic flux lines discussed above also serve as electrostatic shields. However, fairly effective electrostatic shielding can be obtained by a metal mesh made of any good to fair electrical conductor, which would be a rather poor shield for magnetic flux.

It is possible to shield electrostatic flux without simultaneously affecting the magnetic field by surrounding the space to be shielded with a conducting cage that is made in such a way as to provide no low-resistance path for the flow of eddy currents, while at the same time offering a metallic terminal upon which electrostatic flux lines can terminate. Thus, the secondary winding of a transformer may be shielded electrostatically from the primary by a shield having an insulated gap located in such a manner as to prevent the shield from becoming a short-circuited turn. This is illustrated in Fig. 2-20. Another type of electrostatic shield that does not affect the magnetic flux is illustrated below in connection with Prob. 2-45.

Energy loss is associated with electrostatic shielding as a result of the fact that the charging current induced in the shield produces currents that must flow through the surface resistance of the shield. However,

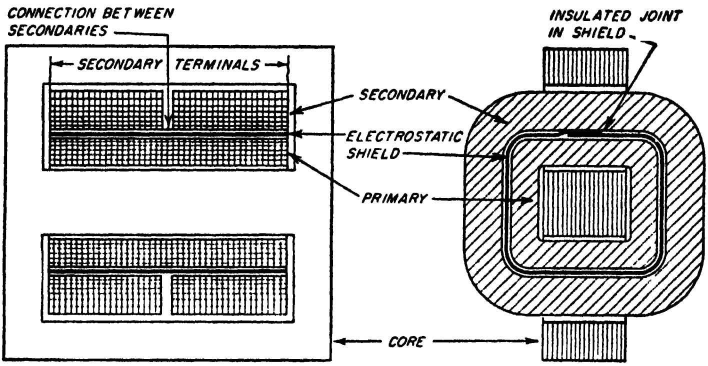


FIG. 2-20. Transformer with electrostatic shield between primary and secondary.

under most conditions these currents are quite small, so that the associated loss is generally insignificant. This is in contrast with shields for alternating magnetic flux, where the shield losses are very often substantial.

Effect of Shielding on Coils.[1] The magnetic and electric fields associated with a coil are frequently confined by placing the coil in a shield can composed of aluminum or copper. Such a shield increases the coil's distributed capacitance and effective resistance, and reduces its inductance. The distributed capacitance is increased as a result of the capacitance between the shield and various parts of the coil (see Fig. 2-19*d*). The inductance is decreased because the conducting shield restricts the magnetic flux lines to the space within the shield; this decreases the cross section of the magnetic circuit and thus reduces the flux linkages. The effective resistance of the coil is increased because the energy consumed by the eddy currents flowing in the shield is supplied by the coil.

The degree of shielding obtained at radio frequencies by enclosing a coil in a copper or aluminum container is very high, and if a reasonable clear-

[1] The quantitative relations involved are discussed by Howe, *op. cit.*; A. G. Bogle, The Effective Inductance and Resistance of Screened Coils, *J. IEE* (*Wireless Soc.*), vol. 15, p. 221, September, 1940.

ance is provided between the shield and the coil, the properties of the coil are not seriously impaired. In general, the clearance between the shield and the coil should everywhere be not less than the coil radius. Under such conditions, the presence of the shield will not reduce the coil Q by more than 20 per cent, while the reduction in inductance will be still less.

A shielded inductance that is very compact can be achieved by making the conducting shield only slightly larger in diameter than the coil, and then using a liner of nonconducting magnetic material to fill the space between coil and shield, as shown in Fig. 2-19*f*. The high permeability of the magnetic material provides a low-reluctance return path for the magnetic flux outside the coil, with the result that the inductance obtained is greater than when the coil is unshielded as in Fig. 2-19*a*. This makes it practical to reduce the physical size required to obtain a given inductance at a desired value of Q. In addition, the shielding is more effective than when only a conducting shield is used (Fig. 2-19*c*), because both magnetic and conducting shielding is simultaneously obtained.

PROBLEMS AND EXERCISES

2-1. If the flux shown in Fig. 2-1 is produced by a current of 0.01 amp, estimate the coil inductance. (Assume that Fig. 2-1 gives a two-dimensional representation of the actual flux lines present in the three-dimensional coil.)

2-2. A single-layer coil is to have an inductance of 220 μh and is to be wound on a form having a diameter of 2 in. If the ratio of length to diameter is 1.5, determine the distance between centers of adjacent turns of the winding.

2-3. A single-layer solenoidal coil having 60 turns on a winding 3 in. long and 3 in. in diameter possesses an inductance of 187 μh. Without using Fig. 2-2, determine:

a. How many turns would be required to obtain the same inductance if the core were 2 in. in diameter and 2 in. long.

b. How many turns would be required to obtain an inductance of 400 μh with a winding 4 in. long and 4 in. in diameter.

2-4. *a.* On a hysteresis loop similar to that of Fig. 2-3, show a minor hysteresis loop originating at point 1 on the main loop, but corresponding to a substantially larger value of alternating magnetization.

b. Repeat for the same alternating magnetization as in (*a*), but with the d-c magnetization corresponding to point 3, instead of point 1.

2-5. The incremental inductance at low alternating magnetization of a particular iron-cored coil having 1000 turns is 10 henrys with no d-c saturation, and 4 henrys when carrying a d-c magnetizing current of 0.1 amp. When the number of turns is reduced to 500, it is found that the inductance without d-c saturation is reduced to 2.5 henrys, or exactly one-fourth of the previous value, whereas with a d-c magnetizing current of 0.1 amp the incremental inductance is somewhat greater than one-fourth of 4 henrys. Explain.

2-6. A coil uses a silicon-steel core composed of material having the characteristics given in Fig. 2-4, and the core is assembled with negligible air gap. If the incremental inductance is 5.4 henrys with no d-c magnetization and low alternating flux density, what will be the incremental inductances with d-c magnetizations sufficient to produce 1, 2, and 3 ampere turns per cm?

2-7. *a.* A permanent magnet that is required to produce a large amount of flux in a low-reluctance magnetic circuit will be short and thick. Explain.

b. A permanent magnet that is required to produce a small amount of flux in a high-reluctance circuit will be long and thin. Explain.

2-8. A permanent magnet of Alnico V is required to establish a flux density of 2000 lines per sq cm in an air gap 1.2 cm long and having an effective cross section of 20 sq cm. Determine the length and cross section of the magnet required, assuming stabilization is not necessary.

2-9. Explain why a permanent magnet stabilized as in Fig. 2-6 will have to be both larger in cross section and longer than an unstabilized magnet in order to produce the same flux in a given external circuit.

2-10. A particular permanent-magnet system employs a cylindrical magnet of Alnico V having a diameter of 1.0 in. and a length of 0.6 in. If Alnico XII is used instead, calculate the diameter and length required to give the same result. Assume that the permanent-magnet material is used under optimum conditions in both cases and that stabilization is not required.

2-11. A primary coil having an inductance of 100 μh is connected in series with a secondary coil of 240 μh, and the total inductance of the combination is measured as 146 μh. Determine (*a*) the mutual inductance, (*b*) the coefficient of coupling, and (*c*) the inductance that would be observed if the terminals of one of the coils were reversed.

2-12. Two circuits are to be coupled by a common capacitor using the circuit of Fig. 2-8*c*. If the total capacitance required in the primary circuit is 150 $\mu\mu$f, while the total capacitance required in the secondary circuit is 100 $\mu\mu$f, determine the value of the common capacitance C_m in Fig. 2-8*c* to give a coefficient of coupling of 0.02. [Note: In solving this problem do not attempt to use Eq. (2-8).]

2-13. In two circuits coupled as in Fig. 2-8*b*, $L_1 = 0.05$ henry, $L_2 = 0.08$ henry, and $k = 0.4$. Determine (*a*) the required value of L_m, and (*b*) the total primary and total secondary inductances.

2-14. Explain why two coils that have their axes, respectively, parallel to, and at right angles to, the line joining the coil centers will have zero mutual inductance.

2-15. Two single-layer air-cored coils are located coaxially end to end, as illustrated in Fig. 2-8*a*. It is found that, if a long cylindrical magnetic core is slipped inside of these coils so that it is common to both coils, the mutual inductance is increased more than is the self-inductance of the individual coils. Explain.

2-16. What effect does the redistribution of current associated with skin effect have on the inductance? Explain.

2-17. *a.* Calculate the skin depth in coppor for 1 kc, 1 Mc, and 1000 Mc, and tabulate the results.

b. Repeat for aluminum.

2-18. Parts formed of brass, steel, etc., are sometimes silver- or copper-plated to reduce the effective resistance to radio frequencies. If copper plating is employed, and the part is to be used in the frequency range 5 to 20 Mc, recommend a minimum thickness for this plating, and give the reasoning upon which this recommendation is based.

2-19. Inductances (and also shields) are sometimes plated to reduce corrosion and improve appearance. The resistivity of the plating material suitable for this purpose is usually much higher than the resistivity of the material that is plated. What criterion must the thickness of the plating satisfy if the effective resistance of the plated conductor is to approach closely the resistance obtained without plating?

2-20. A No. 14 copper wire (diameter 0.0641 in.) has a d-c resistance of 0.2525 ohm per 100 ft. Calculate its resistance at 10 Mc, and at 3000 Mc, and tabulate these three values of resistance alongside of one another.

2-21. What diameter must a copper wire have if its resistance is not to exceed 3.0 ohms per 100 ft at 10 Mc?

2-22. A conductor consisting of a thin-walled tube will have much less resistance at very high frequencies than a solid wire of the same d-c resistance. Explain.

2-23. In a conductor consisting of a tube of specified outside diameter, the resistance at very high frequencies will be almost independent of wall thickness if this thickness exceeds several skin depths, but will be roughly inversely proportional to wall thickness when the thickness is small compared with the skin depth. Explain these observations.

2-24. Determine what mathematical approximation is involved in each of the following statements:

a. The phase angle in radians is equal to the power factor.

b. The reciprocal of capacitor Q is equal to the power factor of the capacitor.

c. The dissipation factor is equal to the phase angle.

2-25. On the basis of the information given in Fig. 2-12 and the associated discussion, sketch curves analogous to Fig. 2-12 but showing qualitatively how the dielectric constant and power factor would vary as a function of temperature for (*a*) a low frequency, and (*b*) a high frequency.

2-26. *a*. A mica capacitor with power factor 0.0005 has a capacitance of 0.001 μf. Assuming skin-effect resistance to be negligible, what is the equivalent series resistance of the capacitor at frequencies of 1000, 100,000, and 10,000,000 cycles?

b. What is the equivalent shunt resistance for the same conditions?

2-27. A certain air capacitor employing mycalex insulation has a power factor of 0.0003 at 1000 cycles. What will its power factor be at this same frequency and same capacitance if the mycalex insulation is replaced by polystyrene insulation of the same geometrical configuration?

2-28. The power factor of a capacitor at very high frequencies is roughly proportional to f^n where f is the frequency. What is the value of n?

2-29. Show that the power factor of a given variable air capacitor at low frequencies is independent of frequency but increases inversely with capacitance setting.

2-30. At very high frequencies, does an increase in frequency cause the power factor of a variable air capacitance for a given capacitance setting to become greater, less, or unchanged? Give an adequate justification for the answer chosen.

2-31. In a variable air capacitor the ratio of the power factor at a given high frequency to the power factor at a given low frequency becomes greater as the capacitance setting increases. Explain.

2-32. In a capacitor having a capacitance of 0.001 μf, the equivalent series inductance of the leads, etc., is 0.1 μh. At what frequency does the apparent capacitance differ from the true capacitance by 10 per cent?

2-33. A certain capacitor 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. What is the equivalent series resistance and power factor at 10,000 kc when the total capacitance is 100 $\mu\mu$f (90 $\mu\mu$f from air and 10 $\mu\mu$f from bakelite)? Neglect skin-effect losses.

2-34. A particular mica capacitor having a capacitance of 0.001 μf has a power factor of 0.0005 at a frequency of 1000 cycles, while at 10 Mc the power factor has risen to 0.001. From this information deduce the values of R and R_2 applicable in Fig. 2-13 at 10 Mc.

2-35. The capacitor of Prob. 2-26 is able to stand a d-c potential of 5000 volts and is capable of dissipating safely 3 watts of heat.

a. At what frequency will heating begin to limit the voltage rating?

b. What is the voltage rating at frequencies of 1, 1000, and 10,000 kc?

2-36. 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 $Q = 50$.

2-37. Explain why the distributed capacitance of a coil is always increased by the wax or other coating used for protection against moisture.

2-38. On the basis of proximity and skin effects, explain why it is reasonable to expect that the maximum coil Q would be obtained with a wire not so large as to leave very little clearance between adjacent turns, and not so small as to make this clearance become a large fraction of the spacing between centers of adjacent turns.

2-39. In a coil with a magnetic slug core as in Fig. 2-17, removing the core will reduce the inductance less in a system using a core designed for 100 Mc than in a system using a core designed for 1 Mc. Explain.

2-40. A copper shield is required to reduce the magnetic flux density by 60 db. What shield thickness is required at (*a*) 1 kc, (*b*) 1 Mc, and (*c*) 1000 Mc?

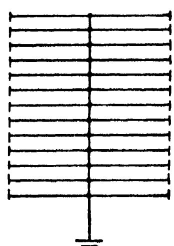
2-41. Derive Eq. (2-17) from Eq. (2-16).

2-42. A particular magnetic shield attenuates d-c magnetic fields by 20 db. What will the attenuation be if the shield thickness is doubled?

2-43. A conducting magnetic shield is composed of permalloy having an initial permeability of 15,000 and a resistivity of 17 μohms per cm cube. Calculate (*a*) the thickness which this material must have to be 5 skin depths thick at 60 cycles, and (*b*) the thickness which copper must have to achieve the same degree of shielding.

2-44. Explain why magnetic material in powdered form, such as used in magnetic cores for radio frequencies, is not suitable for use as a shield of alternating magnetic fields.

2-45. A grid of wires such as shown in the accompanying figure will provide electrostatic shielding without magnetic shielding provided the structure (shown dotted in



PROB. 2-45

the illustration) supporting the sides of the shield is an insulator. However, if the material of the supporting structure is a conductor, then magnetic as well as electrostatic fields are shielded at least to some extent. Explain.

2-46. When a nonmagnetic shield can surrounds a solenoidal coil, it is observed that the shielding of the magnetic field 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 the effectiveness of the shield is very seriously reduced if the joint is in a plane that contains the axis of the coil. Explain.

2-47. If it is necessary that a magnetic shield for d-c fields have a joint, how should this joint be oriented with respect to the direction of the magnetic flux that is being shielded?

CHAPTER 3

PROPERTIES OF CIRCUITS WITH LUMPED CONSTANTS

3-1. Series Resonance. A circuit consisting of an inductance, capacitance, and resistance all in series, as in Fig. 3-1, is called a series resonant or series tuned circuit. When a constant voltage of varying frequency is applied to such a circuit, the current that flows depends upon frequency in the manner shown in Fig. 3-1. At low frequencies, the capacitive reactance of the circuit is large and the inductive reactance is small. Most of the voltage drop is then across the capacitor, while the current is small and leads the applied voltage by nearly 90°. At high frequencies, the inductive reactance is large and the capacitive reactance low, resulting in a small current that lags nearly 90° behind the applied voltage, and most of the voltage drop is across the inductance. In between these two extremes there is a frequency, *called the resonant frequency, at which the capacitive and inductive reactances are exactly equal* and, consequently, neutralize each other; there is then only the resistance of the circuit to oppose the flow of current. The current at the resonant frequency is accordingly equal to the applied voltage divided by the circuit resistance, and is very large if the resistance is low.

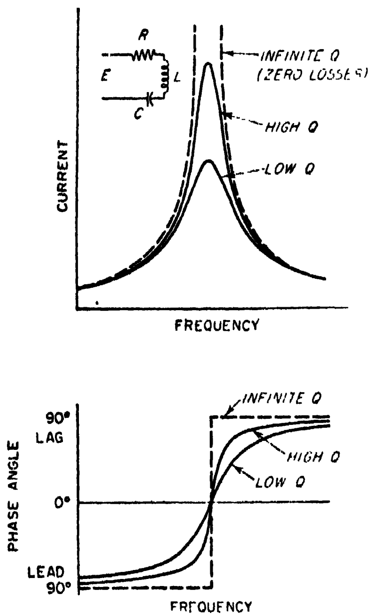


FIG. 3-1. Magnitude and phase angle of current in a series resonant circuit as a function of frequency for constant applied voltage and different circuit Q's.

A resonance curve such as illustrated in Fig. 3-1 finds extensive use in selective systems for separating a desired a-c signal from signals of other frequencies. For frequencies in the vicinity of resonance corresponding to a carrier wave and its sideband frequencies, the response is nearly uniform and is quite large. However, at frequencies differing greatly from resonance the response is relatively small, with the result that signals of such frequencies, i.e., the unwanted signals, are severely discriminated against.

The characteristics of a series resonant circuit depend primarily upon the ratio of inductive reactance ωL to *circuit* resistance R, i.e., upon $\omega L/R$. This ratio is frequently denoted by the symbol Q and is called the circuit Q.[1] Most of the loss in the usual resonant circuit is due to coil resistance because the losses in a properly constructed capacitor are small in comparison with those of a coil. The result is that the circuit Q ordinarily approximates the Q of the coil alone, which was discussed in Sec. 2-7.

The general effect of different circuit resistances, i.e., different values of Q, is shown in Fig. 3-1. It is seen that, when the frequency differs appreciably from the resonant frequency, *the actual current is practically independent of the circuit resistance and is very nearly the current that would be obtained with no losses.* On the other hand, the current at the resonant frequency is determined solely by the resistance. The effect of increasing the resistance of a series circuit is, accordingly, to flatten the resonance curve by reducing the current at resonance without significantly affecting the behavior at frequencies differing appreciably from resonance. This broadens the top of the curve, giving a more nearly uniform current over a band of frequencies near the resonant point, but does so by reducing the ability of the circuit to discriminate between voltages of different frequencies.

Analysis of Series Resonant Circuit. The elementary voltage, current, and impedance relations of series resonant circuits are discussed in every book on alternating currents. The basic quantitative relations are listed below for convenient reference.

$$\text{Resonant frequency} = f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{3-2}$$

$$Z_s = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{3-3a}$$

$$|Z_s| = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \tag{3-3b}$$

[1] The circuit Q can also be defined as

$$Q = 2\pi \frac{\text{energy stored in circuit}}{\left\{\begin{matrix}\text{energy dissipated in circuit}\\ \text{during one cycle}\end{matrix}\right.} \tag{3-1}$$

This relation follows from the fact that the energy stored in the inductance L when the current I is maximum (i.e., when all the stored energy is in the inductance) is $I^2L/2$, where I is the peak current. At the same time, the energy lost per cycle in the circuit resistance R is $I^2R/2f$, where f is the frequency.

$$\tan\theta = \frac{\omega L - (1/\omega C)}{R} \tag{3-4}$$

$$I = \frac{E}{Z_s} = \frac{E}{R + j[\omega L - (1/\omega C)]} \tag{3-5}$$

$$\text{Current at resonance} = I_0 = \frac{E}{R_0} \tag{3-6}$$

$$\text{Voltage across inductance} = j\omega LI \tag{3-7a}$$

$$\text{Voltage across capacitor} = \frac{-j}{\omega C} I \tag{3-7b}$$

where E = voltage applied to circuit
I = current flowing in circuit, amp
f = frequency, cycles
$\omega = 2\pi f$
$Q = \omega L/R$
R = total series resistance of tuned circuit
L = inductance, henrys
C = capacitance, farads
Z_s = impedance of series circuit
θ = phase angle of impedance
Subscript $_0$ denotes values at resonant frequency

At frequencies near resonance the voltages across the capacitor and the inductance will both be very much greater than the applied voltage. This is possible because at frequencies near resonance the voltages across the capacitor 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.

At resonance, where the circuit current is E/R, Eqs. (3-7) show that the voltage across the inductance (or capacitor) is then Q times the applied voltage; i.e., *there is a resonant rise of voltage in the circuit amounting to Q times.* Since a typical value of Q is of the order of 100, a series resonant circuit will thus develop a high voltage even with small applied potentials. At frequencies differing from resonance the voltage developed across the inductance (or capacitor) falls off. In the vicinity of resonance the resulting curve of voltage as a function of frequency has a shape that for all practical purposes can be considered to be the same as the corresponding curve of current as a function of frequency (see Fig. 3-1). The reason for this is that most of the resonance effects exist in a very narrow frequency band, typically representing a frequency variation of only a few per cent. Over this frequency range the term ωL (or $1/\omega C$) in Eqs. (3-7) is so nearly constant that to a first approximation the voltage developed across the circuit can be considered to be proportional to the circuit current.

Universal Resonance Curve. Equations (3-5) and (3-6) can also be rearranged to express the ratio of current actually flowing to the current at resonance, in terms of the circuit Q and the fractional deviation of the frequency from resonance. This leads to the universal resonance curve of Fig. 3-2.[1]

In the universal resonance curve, the frequency is expressed in terms of a parameter a that represents f_0/Q cycles, as defined in Fig. 3-2. Thus $a = 1.0$ when the cycles off resonance equal f_0/Q cycles, $a = 2$ when the number of cycles off resonance is $2f_0/Q$, etc.

The use of Fig. 3-2 in practical calculations can be illustrated by two 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. 3-2 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°.

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.

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

$$a = {}^{10}\!/\!_{1000} \times 125 = 1.25$$

Reference to Fig. 3-2 shows that for $a = 1.25$ the response is reduced by a factor 0.37 and that the phase of the current is 68° leading.

The only assumption involved in the universal resonance curve is that Q is assumed to be the same at the frequency being considered as at the resonant frequency. *When this is true, the universal resonance curve*

[1] The equation of the universal resonance curve is obtained as follows: The ratio of Eq. (3-5) to Eq. (3-6) gives

$$\frac{\text{Actual current}}{\text{Current at resonance}} = \frac{R_0}{R + j[\omega L - (1/\omega C)]} = \frac{R_0}{R + j[(\omega^2 LC - 1)/\omega C]}$$

Now define the fractional deviation δ of the frequency from resonance, according to the relation

$$\omega = \omega_0(1 + \delta)$$

Substituting this expression for ω and remembering that $\omega_0 L = 1/\omega_0 C$, one obtains

$$\frac{\text{Actual current}}{\text{Current at resonance}} = \frac{R_0}{R + j\left[\frac{(1+\delta)^2 - 1}{1+\delta}\right]\omega_0 L} = \frac{1}{\frac{R}{R_0} + jQ\delta\left(\frac{2+\delta}{1+\delta}\right)},$$

When Q is constant, the radio-frequency resistance is proportional to frequency so that $R/R_0 = \omega/\omega_0 = (1 + \delta)$, which when substituted yields

$$\frac{\text{Actual current}}{\text{Current at resonance}} = \frac{1}{1 + \delta + jQ\delta\left(\frac{2+\delta}{1+\delta}\right)} \qquad (3\text{-}8)$$

Figure 3-2 is then obtained by substituting $a = Q\delta$.

involves no approximations whatsoever. Over the limited range of frequencies near resonance represented in Fig. 3-2, the variation in Q in practical cases is so small as to introduce negligible (i.e., less than 1 per cent) error from the use of the curve, when the value of Q existing at resonance is used in determining the parameter a.

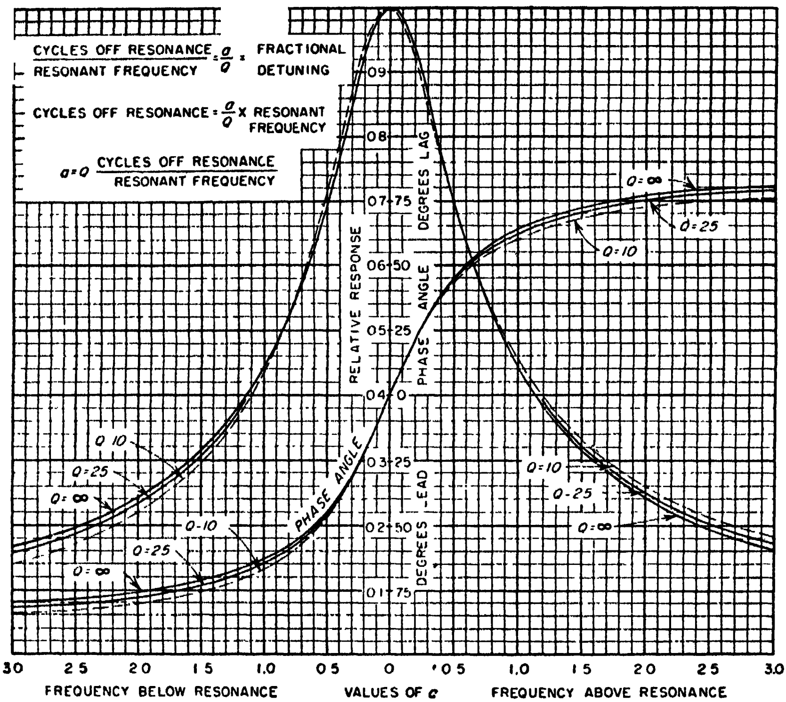


FIG. 3-2. Universal resonance curve for series resonant circuit. This curve can also be applied to the parallel resonant circuit by considering the vertical scale to represent the ratio of actual parallel impedance to the parallel impedance at resonance. When applied to parallel circuits, the angles shown in the figure as leading are lagging, and vice versa.

The universal resonance curve is useful because it is independent of the resonant frequency of the circuit and of the ratio of inductance to capacitance, and because it is substantially independent of circuit Q. It thus follows that *all resonance curves have the same relative shape irrespective of resonant frequency, Q, or ratio of inductance to capacitance of the circuit.*

Working Rules for Estimating Sharpness of Resonance. Since the curves for different values of Q are almost identical in Fig. 3-2, particularly in the neighborhood of the resonant frequency, it is possible to state several easily remembered working rules that will enable one to estimate the sharpness of any resonance curve with an error of less than 1 per cent when only the Q of the circuit is known.[1] These rules follow:

Rule 1. *When the frequency of the applied voltage deviates from the resonant frequency by an amount that is* $1/2Q$ *of the resonant frequency, the current that flows is reduced to* 70.7 *per cent of the resonant current, and the current is* 45° *out of phase with the applied voltage. Thus the frequency band* B *over which the response is at least* 70.7 *per cent of that at resonance (i.e., within* 3 *db of resonance) is* $B = f_0/Q$, *where* f_0 *is the resonant frequency.*

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 $1/250$ of 1000 kc, or 4000 cycles off resonance, and to 44.7 per cent of the resonant current for a frequency deviation of $1/125$ of 1000 kc, or 8000 cycles. Since the resonant rise of voltage in this circuit is 125 ($=Q$) times, the rise of voltage is very nearly $0.7 \times 125 = 87.5$ times when the frequency is 4000 cycles off resonance, and is very close to $0.45 \times 125 = 56.25$ times at a frequency 8000 cycles from resonance.

Practical Calculation of Resonance Curves. The proper procedure for calculating a resonance curve is to start by determining the current at resonance, using Eq. (3-6). The working rules can then be applied to obtain the response at frequencies $1/2Q$ and $1/Q$ on either side of resonance. This gives a picture of the sharpness of resonance and is sufficient for many purposes. However, if additional points in the vicinity of resonance are needed, they can be calculated with the aid of Fig. 3-2.

At frequencies too far off resonance to come within the range of the universal resonance curve, the *magnitude* of the current can be determined with an accuracy sufficient for nearly all practical purposes by neglecting the resistance R in Eq. (3-5). The phase angle of the current under these conditions is obtained from Eq. (3-4).

The above procedure for calculating resonance curves is much superior to making calculations based directly upon Eq. (3-5). The use of the universal resonance curve in the vicinity of the resonant frequency not only reduces the amount of labor involved but also greatly improves the accuracy under ordinary conditions. This is because resonant circuit formulas such as Eq. (3-5) contain a term $\left(\omega L - \frac{1}{\omega C}\right)$ which involves the difference of two quantities which near resonance are nearly equal in magnitude. In order to obtain this difference without more than 1 per

[1] An error of 1 per cent is nearly always permissible in calculations of radio-frequency circuits. This is because the effective circuit constants at radio frequencies are very seldom known to an accuracy that involves an error of less than 1 per cent.

cent error, five-place logarithms must ordinarily be employed. Slide-rule calculations are never permissible. Neglecting the resistance at frequencies too far off resonance to come within the range covered by the universal resonance curve enormously reduces the labor involved in calculating magnitudes, and introduces an error of less than 1 per cent of the magnitude at resonance. This accuracy is ample for all ordinary purposes, and the error is undetectable when resonance curves are plotted.

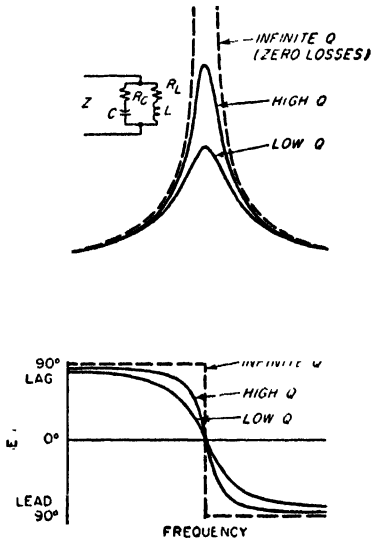


FIG. 3-3. Magnitude and phase angle of impedance of a parallel circuit as a function of frequency for different circuit Q's.

3-2. Parallel Resonance. A parallel circuit consisting of an inductance branch in parallel with a capacitance branch offers an impedance of the character shown in Fig. 3-3. Such a circuit is termed a parallel resonant or parallel tuned circuit.

When a voltage is applied to such a system, then 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 lagging line or circuit current and a low lagging circuit impedance. At high frequencies, the inductance has a high reactance compared with the capacitance, resulting in a large leading line current and a correspondingly low circuit impedance that is leading in phase. 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 capacitance branch are equal; being 180° out of phase, they then neutralize and leave only a small resultant inphase current flowing in the line. The impedance of the parallel circuit will then be a very high resistance, as is brought out in Fig. 3-3.[1]

A comparison of Figs. 3-1 and 3-3 shows that the impedance curve of a parallel circuit is similar in character to the current curve of a series circuit. In particular, increasing the resistance of a parallel resonant

[1] In obtaining a parallel resonance curve experimentally by measurements of applied voltage and line current, extreme care must be taken to ensure that the applied voltage contains no harmonics. This is necessary because at resonance the circuit impedance is extremely high to the fundamental component of the applied voltage and very low to the harmonic components, with the result that even a small harmonic-voltage component will cause line currents that mask the small fundamental component.

circuit lowers and flattens the peak of the resonance curve, just as in the analogous series resonance case. This similarity is considered below in greater detail.

The relationship between the line and branch currents in a parallel circuit is illustrated in Fig. 3-4. It will be noted that, unlike the line or circuit current, which shows a resonance effect, the currents in the individual branches of a parallel circuit vary only slightly in the vicinity of resonance and are relatively large. At resonance the two branch currents have similar magnitudes, and being almost (but not quite) out of phase they add up to a very small resultant current, thus giving a high circuit impedance. As the frequency departs from resonance the two branch currents become slightly unequal in magnitude; this causes the line current to increase as shown, which means lowered circuit impedance. It is characteristic of parallel resonant circuits that for frequencies near resonance the current flowing in the branches, commonly referred to as the circulating current, is much larger than the line current, i.e., than the current supplied to the circuit.

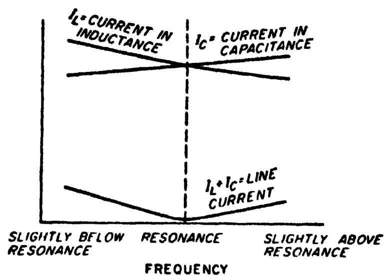


FIG. 3-4. Relationship of line and branch currents in a parallel resonant circuit in the vicinity of resonance.

The fundamental relations of a parallel resonant circuit are derived in every introductory book on a-c circuit theory, and are listed below for convenient reference.

$$\text{Parallel impedance} = Z = \frac{Z_c Z_L}{Z_c + Z_L} = \frac{Z_c Z_L}{Z_s} \tag{3-9}$$

$$\text{Line current} = \frac{E}{Z} \tag{3-10}$$

$$\text{Inductive branch current} = \frac{E}{Z_L} = \frac{E}{R_L + j\omega L} \tag{3-11}$$

$$\text{Capacitive branch current} = \frac{E}{Z_C} = \frac{E}{R_C - (j/\omega C)} \tag{3-12}$$

where E = voltage applied to circuit

$Z_c = R_c - (j/\omega C)$ = impedance of capacitive branch

$Z_L = R_L + j\omega L$ = impedance of inductive branch

$Z_s = Z_c + Z_L$ = series impedance of circuit

Z = impedance of circuit when connected in parallel

$R_s = R_c + R_L$ = total series resistance of circuit

$\omega = 2\pi$ times frequency

$Q = \omega L/R_s$ = circuit Q

These equations are fundamental to every parallel circuit, irrespective of the circuit Q, the frequency, or the division of resistance between the branches.

Quantitative Relations in Parallel Resonant Circuits with Moderate or High Q, and the Use of Universal Resonance Curve. When the Q of a parallel resonant circuit is not too low (e.g., of the order of 10 or more), the quantitative relations become quite simple. To begin with, it is then permissible to assume that the circuit has maximum impedance and unity power factor at the same frequency, which is also the frequency at which the same circuit is in series resonance as given by Eq. (3-2). In contrast, when the circuit Q is low, this is not necessarily the case, as discussed below.

When the circuit Q is not too low, the exact expressions of Eqs. (3-9) and (3-10) can be simplified, without introducing appreciable error, by neglecting the resistance components of the impedances Z_L and Z_C in the numerator of Eq. (3-9). When this is done[1]

$$\text{Parallel impedance} = Z = \frac{(\omega_0 L)^2}{Z_s} \qquad (3\text{-}13)$$

At resonance $Z_s = R_s$, and this becomes

$$\text{Parallel impedance at resonance} = \frac{(\omega_0 L)^2}{R_s} = (\omega_0 L)Q \qquad (3\text{-}14)$$

In these equations ω_0 is the value of ω at resonance. It will be noted from Eq. (3-14) that *at resonance the impedance of a parallel circuit is a resistance that is Q times the reactance of one of the branches.*[2] It can, therefore, be said that the parallel arrangement of inductive and capacitive branches causes a resonant rise of impedance of Q times the impedance that would be obtained from either branch alone. It is thus apparent that very high impedances can be developed by parallel resonance. This is one of the most important properties of parallel resonance.

Under conditions where the circuit Q is not too low, the resonance curve of the parallel impedance of a circuit can be considered to have the *same shape as the resonance curve of the series current* in a circuit consisting

[1] This transformation is carried out as follows: If the resistance components in the numerator of Eq. (3-12) 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,

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

[2] It also follows from Eq. (3-14) and Eqs. (3-10) to (3-12) that at resonance the branch currents are Q times as large as the line current, provided the resistance components in Eqs. (3-11) and (3-12) are small compared with the associated reactive components.

of the same inductance, capacitance, and resistances connected in a series instead of a parallel arrangement. This follows from the fact that a comparison of Eqs. (3-5) and (3-13) shows that both the parallel impedance and the series current are equal to a constant divided by Z_s. *Consequently, the universal resonance curve and the working rules that were applied for estimating the sharpness of resonance of the series circuit also apply to the case of parallel resonance when the circuit Q is moderate or high.* The only difference is that the signs of the phase angles are now reversed, the phase of the parallel impedance being leading at frequencies higher than resonance and lagging at frequencies below resonance.

The proper procedure for calculating the impedance of a parallel resonant circuit of moderate or high Q is therefore similar to that used

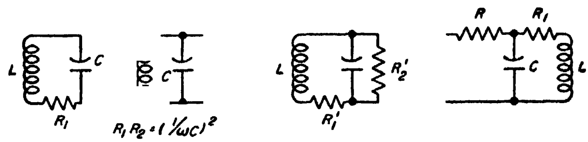


FIG. 3-5. Forms of parallel resonant circuits involving a shunt resistance and shunt feed.

with a series resonant circuit. The first step is to determine the resonant frequency and the impedance at resonance, using Eqs. (3-2) and (3-14). Next, the working rules are applied to obtain the 70.7 and the 44.7 per cent points on either side of resonance. This gives the general picture of the sharpness of resonance and is sufficient for many purposes. If a more complete curve is desired in the vicinity of resonance, one may make use of the universal resonance curve of Fig. 3-2. Finally, at frequencies so far off resonance as to be outside the range of the universal resonance curve, one may determine the magnitude of the impedance by using Eq. (3-13), but neglecting the circuit resistance R when making the calculation. The power-factor angle of the impedance thus obtained is the negative of the corresponding angle for series resonance, as given by Eq. (3-4).

Parallel Resonant Circuits with Shunt Resistances and with Parallel Feed. The two types of parallel resonant circuits shown in Fig. 3-5*a* are equivalent to each other provided the resistances R_1 and R_2 are properly, related, and provided also that the circuit Q is not too low. To determine the relationship that must exist between R_1 and R_2, one notes that R_1 can be thought of as being associated with capacitance C as its equivalent series resistance, while R_2 can be regarded as an equivalent shunt resistance of the same capacitance C. Assuming that the circuit Q is not too low [i.e., that $(1/\omega C)/R_1 >> 1$, and that $R_2/(1/\omega C) >> 1$], then from Eqs. (2-12) and (2-13) one has

$$R_1R_2 = \left(\frac{1}{\omega C}\right)^2 \tag{3-15}$$

Although the relationship between R_1 and R_2 is seen by Eq. (3-15) to depend on frequency, it is common practice to determine the relation between R_1 and R_2 at resonance, and then to assume that the values at resonance also hold for all frequencies in the vicinity of resonance. This approximation is equivalent to assuming that the right-hand term of Eq. (3-15) is constant at the value it has at resonance.[1] Since ω changes by only a small percentage in the limited frequency range around resonance, this assumption is not far from the truth, and the error it introduces is quite small.

The parallel resonant circuit of Fig. 3-5*b* can be transformed to the circuits of Fig. 3-5*a* by converting R_2' to an equivalent series resistance R_1 or transforming R_1' to an equivalent shunt resistance R_2. By use of Eqs. (2-12) and (2-13), respectively, this leads to the following relations between the circuits of *a* and *b* in Fig. 3-5 for the resonant frequency:

$$\left.\begin{array}{l}\text{Total effective}\\ \text{series resistance}\end{array}\right\} = R_1 = R_1' + \frac{(\omega_0 L)^2}{R_2'} \tag{3-18a}$$

$$\left.\begin{array}{l}\text{Shunt resistance}\\ \text{equivalent to } R_1'\end{array}\right\} = R_{eq} = \frac{(\omega_0 L)^2}{R_1'} \tag{3-18b}$$

$$\left.\begin{array}{l}\text{Total effective}\\ \text{shunt resistance}\\ \text{including } R' \text{ and } R_{eq}\end{array}\right\} = R_2 = \frac{R_{eq}R_2}{R_{eq} + R_2'} \tag{3-18c}$$

The above analysis is of practical importance for two reasons. In the first place, it shows that to a high approximation, the effect produced by shunting a resistance across a parallel resonant circuit is merely to lower the effective Q of the circuit. The resonant frequency is unchanged, however, and the impedance curve still has the shape of a resonance curve as given by the universal resonance curve. In the second place, the analysis provides a simple means of determining the quantitative effect that a shunt resistance produces on the properties of a resonant circuit.

Still another form of parallel resonant circuit that is frequently encountered is shown in Fig. 3-5*c*, where a resistance R is connected in series with

[1] At resonance, one can write $\omega_0 L = 1/\omega_0 C$, where ω_0 is the value of ω at the resonant frequency. Under these conditions the following useful relations apply to Fig. 3-5*a*:

$$R_1R_2 = (\omega_0 L)^2 \tag{3-16}$$

$$\left.\begin{array}{l}Q \text{ of circuit}\\ \text{at resonance}\end{array}\right\} = \frac{\omega_0 L}{R_1} = \frac{R_2}{\omega_0 L} \tag{3-17a}$$

$$\left.\begin{array}{l}\text{Parallel impedance}\\ \text{at resonance}\end{array}\right\} = \omega_0 LQ = R_2 \tag{3-17b}$$

the parallel circuit. The behavior of arrangements of this type is analyzed in Sec. 3-7.

Parallel Circuits with Low Q. The entire discussion of parallel resonance given above except for Eqs. (3-9) to (3-12) assumes that the Q of the parallel circuit is at least reasonably high (i.e., of the order of 10 or more). In the general case when the circuit Q is low, the curve of circuit impedance as a function of frequency still has a shape that resembles a resonance curve unless the circuit Q approaches or is less than unity. However, the maximum impedance no longer necessarily occurs at the frequency of series resonance, and the condition of unity power factor does not necessarily occur either at the frequency of series resonance or when the impedance is a maximum. The actual behavior for any given Q depends upon the division of resistance between the inductive and capacitive branches, as illustrated in Fig. 3-6 for typical cases.

An important consideration in the use of low-Q resonant circuits occurs when such a circuit is tuned to resonance with a given frequency by varying either the inductance or capacitance of the circuit. If, for example, the tuning is accomplished by varying the capacitance, then, if all the circuit losses are in the inductive branch, the capacitance setting that makes the circuit impedance maximum also corresponds to unity power factor. If, however, part or all of the circuit resistance is in the capacitive branch, then the capacitance setting that makes the circuit impedance maximum at an assigned frequency does not correspond to the capacitance setting for which the circuit power factor is unity. This is illustrated in Fig. 3-6. Similarly, if the tuning is accomplished by varying the inductance, then the situation is reversed, and maximum impedance and unity-power-factor conditions coincide only if all the circuit losses are concentrated in the capacitive branch. These properties of parallel resonant circuits with low Q are often of considerable importance in connection with the resonant circuits of Class C amplifiers such as used in radio transmitters.

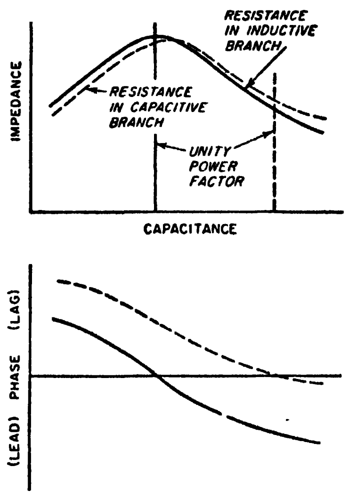
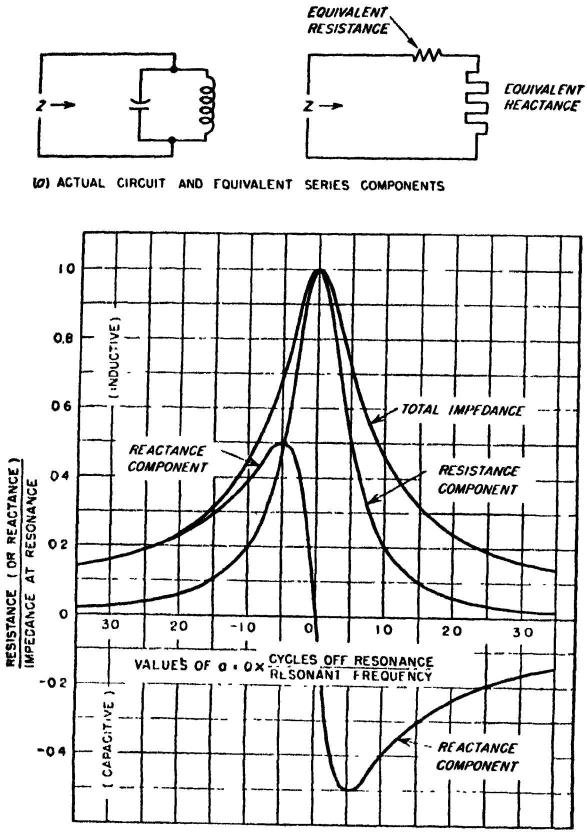


FIG. 3-6. Typical characteristics of parallel resonance circuits having low Q.

Components of Parallel Impedance. The parallel impedance as calculated by Eq. (3-9) or (3-13) can be thought of as equivalent to a resistance in series with a reactance, as shown in Fig. 3-7*a*. When the circuit Q is sufficiently high for Eq. (3-13) to apply, then these resistance and reactance components will be found to vary with frequency in the manner shown in Fig. 3-7*b*, which is a universal curve derivable directly from the universal resonance curve of the parallel circuit. It will be noted that the resistance component has a shape superficially similar to that of a resonance curve, but differs in that it has steeper sides. In particular,



(b) UNIVERSAL CURVE OF IMPEDANCE COMPONENTS

FIG. 3-7. Representation of parallel impedance in terms of equivalent series resistance and reactance components, together with universal curve giving these components as a function of frequency in a parallel resonant circuit having a relatively high *Q*.

the resistance drops to 50 per cent of the resonant impedance at frequencies corresponding to the 70.7 per cent points of the impedance curve (i.e., when the number of cycles off resonance equals the resonant frequency divided by 2*Q*). It will also be noted that the reactance curves are characterized by maxima and minima which occur at the 70.7 per cent points of the resonance curve and which have peak amplitudes that are exactly 50 per cent of the impedance at resonance as given by Eq. (3-14).

An application of these concepts is supplied by the case of a coil having distributed capacitance. With respect to its terminals, such a coil is represented by the left-hand circuit of Fig. 3-7*a*, and accordingly behaves as shown in Fig. 3-7*b*. Below the frequency at which the distributed capacitance is resonant with the inductance, the system is equivalent to a resistance in series with an inductive reactance. The apparent inductance represented by this equivalent reactance depends on frequency, however, rising with frequency until just before resonance is reached, and then dropping rapidly. The apparent inductance becomes zero at the parallel resonant frequency, *while for higher frequencies the coil has a capacitive reactance and is therefore equivalent to a small capacitor.* The

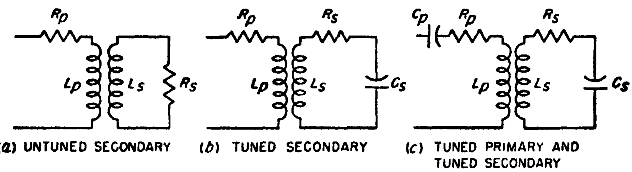


FIG. 3-8. Various types of inductively coupled circuits commonly encountered in electronics.

apparent resistance of the coil increases rapidly with the frequency until a maximum is reached at the resonant frequency, beyond which the resistance rapidly diminishes. These effects are all direct consequences of the properties of parallel resonant circuits, and can be readily deduced by an examination of Fig. 3-7 or of Eqs. (3-13) and (3-14). The behavior of an inductance coil with self-capacitance can accordingly be calculated just as one would determine the characteristics of any other parallel circuit.

3-3. 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. This results in induced currents and a transfer of energy from the first or primary circuit to the coupled or secondary circuit. Several examples of inductively coupled circuits commonly encountered in electronics are shown in Fig. 3-8.

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

Rule 1. *As far as the primary circuit is concerned, the effect that the presence of the coupled secondary circuit has is exactly as though an imped-*

ance $(\omega M)^2/Z_s$ *had been added in series with the primary,*[1] where M = mutual inductance, $\omega = 2\pi f$, and Z_s = series impedance of secondary circuit when considered by itself. The equivalent impedance $(\omega M)^2/Z_s$ which the presence of the secondary adds to the primary circuit is called the *coupled* (or reflected) *impedance* and, since Z_s is a vector quantity having both magnitude and phase, the coupled impedance is also a vector quantity, having resistance and reactance components.

Rule 2. *The voltage induced in the secondary circuit by a primary current of* I_p *has a magnitude of* ωMI_p *and lags behind the current that produces it by* 90°. *In complex quantity notation the induced voltage is* $-j\omega MI_p$.

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 if the primary were absent.*[2] The secondary current therefore has a magnitude $\omega MI_p/Z_s$, and in complex quantity representation is given by $-j\omega MI_p/Z_s$.

These three rules hold for all frequencies and all types of primary and secondary circuits, both tuned and untuned. The procedure to follow in computing the behavior of a coupled circuit is (1) to determine the primary current with the aid of Rule 1; (2) to compute the voltage induced in the secondary, knowing the primary current and using Rule 2; and (3) to 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 by writing down the circuit equations for the primary and secondary. These equations are

$$E = I_pZ_p + j\omega MI_s$$
$$\text{Induced voltage} = -j\omega MI_p = I_sZ_s$$

where Z_p is the series impedance of the primary and E is the voltage applied to the primary. Solving this pair of equations to eliminate I_s gives

$$E = I_p\left[Z_p + \frac{(\omega M)^2}{Z_s}\right] \tag{3-19}$$

This relation shows that the effective primary impedance with secondary present is $Z_p + (\omega M)^2/Z_s$, of which the second term represents the coupled impedance arising from the presence of the secondary.

[2] Some readers may wonder why it is that, although the secondary circuit couples an impedance into the primary, the primary is not considered as coupling an impedance into the secondary. The explanation for this is as follows: The effect that the secondary really has upon the primary circuit is to induce a back voltage in the primary proportional to the secondary current. This back voltage represents a voltage drop occurring in the primary circuit and is the same voltage drop that results when the primary current is assumed to flow through the hypothetical coupled impedance. The impedance that the secondary couples into the primary is hence a means of taking into account the voltage that the secondary current induces into the primary. The voltage that is induced in the secondary circuit by the primary current is taken into account by Rule 3, so that no coupled impedance need be postulated as present in the secondary to take into account the effect of the primary.

$$\left.\begin{array}{l}\text{Impedance coupled into primary}\\ \text{circuit by secondary}\end{array}\right\} = \frac{(\omega M)^2}{Z_s} \tag{3-20}$$

$$\text{Equivalent primary impedance} = Z_p + \frac{(\omega M)^2}{Z_s} \tag{3-21}$$

$$\text{Primary current} = I_p = \frac{E}{Z_p + (\omega M)^2/Z_s} \tag{3-22}$$

$$\text{Voltage induced in secondary} = -j\omega M I_p \tag{3-23}$$

$$\text{Secondary current} = \frac{-j\omega M I_p}{Z_s} = \frac{-j\omega M E}{Z_p Z_s + (\omega M)^2} \tag{3-24}$$

In these equations Z_p is the series impedance of the primary considered as though the secondary were removed, E is the applied voltage, and the remaining notation is as previously used. The primary and secondary impedances Z_p and Z_s, respectively, are vector quantities, so that Eqs. (3-16) to (3-20) are all vector equations.

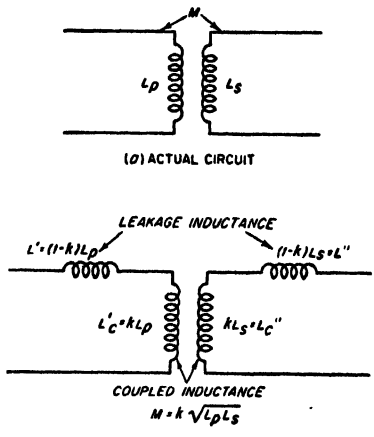


FIG. 3-9. Inductively coupled circuit represented as a transformer with coupled and leakage inductances.

Inductively Coupled Circuit as a Transformer. The inductively coupled circuit is a transformer, and the theory of the inductively coupled circuit that is given above is the general theory of transformers. The method commonly used to analyze the behavior of 60-cycle power transformers, which involves the use of leakage inductance, magnetizing current, and turn ratio, is a special form of the general theory that is convenient when the coupling coefficient k between the primary and secondary windings approaches unity. However, when the coupling coefficient k is small, then the use of Eqs. (3-20) to (3-24) is preferable.

The equivalent transformer circuit represented by two coils coupled together with mutual inductance M is shown in Fig. 3-9. Here the total primary inductance L_p is broken up into a leakage inductance L' and a coupled inductance L'_c, while the secondary is likewise broken up into leakage inductance L'' and a coupled inductance L''_c. Each leakage inductance is considered as having no coupling whatsoever to the other winding, while the coupled inductances L'_c and L''_c are taken as having a coefficient of coupling equal to unity. The values of these inductance components in terms of the coefficient of coupling and the primary, secondary, and mutual inductances are given in the figure. In the representation of Fig. 3-9b, turn ratio has practical significance only when the coefficient of coupling k approaches unity; when the coefficient of coupling is small, as for example, 0.01, then the primary and secondary inductances are practically entirely leakage inductances. Under these conditions the voltage induced in the secondary may be much smaller than the voltage applied to the primary terminals, even when the secondary winding has many more turns than does the primary.

3-4. Analysis of Some Simple Inductively Coupled Circuits. In this and the next section, the types of coupled circuits most commonly encountered in electronics work will be analyzed by the principles given above.

In studying the behavior of a coupled circuit the first step is always to examine the nature of the coupled impedance $(\omega M)^2/Z_s$. When the coupled impedance is small, then the primary current is very nearly the same as though no secondary were present, and the effects produced in the secondary circuit by the primary current will likewise be small. The coupled impedance will be low if the mutual inductance M is very small (i.e., if there is small coupling), or if the secondary impedance is very high, for example, if the secondary is open-circuited. In contrast, consider the case when the coupled impedance $(\omega M)^2/Z_s$ is large, either because of large M or small Z_s, or both. The voltage and current relations that exist in the primary circuit are then affected to a considerable extent by the presence of the coupled secondary, and a very considerable transfer of energy to the secondary occurs.

When determining the effect produced by the coupled impedance, it is important to note that this impedance has the same phase angle as does the secondary impedance Z_s, but with the exception that the sign of the phase angle is reversed. Thus, if the secondary impedance is inductive and has an angle of 30°, the impedance coupled in series with the primary circuit by the action of the secondary has a capacitive phase angle of 30°. The physical significance of this change in sign of the phase angle becomes apparent from the examples considered below. A particularly important case occurs when the secondary impedance Z_s is a pure resistance; under these conditions the coupled impedance will also be a resistance.

The energy consumed by the secondary circuit is the energy represented by the primary current flowing through the resistance component of the coupled impedance.

Coupled Circuit with an Untuned Secondary Consisting of a Resistance and Inductance. This arrangement is illustrated in Fig. 3-8a, and is the type of coupled circuit that results when a resistance is connected across the terminals of the secondary inductance, or, alternatively, is the case where the secondary load is a resistance and an inductance in series. Such a secondary consists of an inductance L_s in series with a resistance

R_s. The coupled impedance is accordingly

$$\text{Coupled impedance} \quad \frac{(\omega M)^2}{Z_s} \quad \frac{(\omega M)^2}{R_s + j\omega L_s} \tag{3-25}$$

Multiplying both numerator and denominator by $R_s - j\omega L_s$ gives

$$\text{Coupled impedance} = \frac{R_s}{R_s^2 + (\omega L_s)^2}(\omega M)^2 - j\frac{\omega L_s}{R_s^2 + (\omega L_s)^2}(\omega M)^2 \tag{3-26}$$

Examination of Eq. (3-26) shows that the coupled impedance introduced into the primary circuit by a resistance-inductance secondary consists of a resistance in series with a capacitive reactance. The effect of the coupled resistance is to increase the effective resistance that appears between the primary terminals. The effect of the coupled capacitive reactance is to neutralize a portion of the primary inductance, thereby reducing the equivalent inductance that is observed between the terminals of the primary coil. The physical explanation of the fact that an inductive secondary produces a capacitive coupled reactance is that such a secondary causes some of the inductive reactance already possessed by the primary to be neutralized. This is done electrically by postulating a capacitive reactance of suitable magnitude in series with the primary.[1]

A special case of considerable importance is that for which the resistance R_s of the secondary circuit in Fig. 3-8*a* is negligible compared with the inductive reactance of the secondary. This situation will arise when the secondary coil is short-circuited, or when the secondary load is a low-loss inductance. To the extent that the resistance of the secondary circuit can be neglected, the coupled resistance introduced into the primary by the presence of such a secondary is zero; the only effect produced by the presence of the secondary is then to reduce the effective inductance that exists between the primary terminals. The percentage reduction in the equivalent primary inductance in such a situation depends only upon the coefficient of coupling between the primary and secondary circuits. If $k = 1.0$, the primary inductance is completely neutralized.[2]

A shield surrounding a coil, or a piece of metal such as a panel located in the magnetic field of a coil, represents a coupled secondary circuit that consists of an inductance in series with a resistance. Such an arrange-

[1] Although the coupled impedance is capacitive and so neutralizes part of the primary inductance, it is impossible to obtain a resultant capacitive reactance in the primary circuit by very large coupling since, with the maximum coupling that can possibly exist ($k = 1$), it will be found that the coupled capacitive reactance can never be greater than the value that will just neutralize all the inductive reactance of the primary.

[2] For other values of k, it can be shown by manipulating Eqs. (3-21) and (3-26) that the equivalent primary inductance is $L_p(1 - k^2)$.

ment can, accordingly, be analyzed as above. Thus the effect of a shield or metal panel on a coil is to reduce the equivalent inductance and to increase the apparent resistance observed at the coil terminals; these effects, moreover, become greater the larger the coupling, i.e., the smaller the spacing between the primary coil and the metal secondary. It is also to be noted that if the secondary resistance is low, as will be the case if the shield or metal panel is made of a good conductor such as copper or aluminum, then the principal effect produced by the presence of the metal near the coil is to reduce the equivalent inductance of the coil; under these circumstances the increase in equivalent coil resistance is only nominal. It will be noted that these conclusions derived from the viewpoint of coupled circuits are all consistent with the qualitative conclusions stated in Sec. 2-8, relative to the effect that shielding has on the properties of a coil.

Coupled Circuits with Untuned Primary and Tuned Secondary. A circuit of this type is shown in Fig. 3-8*b*. Here one has

$$\text{Coupled impedance} = \frac{(\omega M)^2}{Z_s} = \frac{(\omega M)^2}{R_s + j[\omega L_s - (1/\omega C_s)]} \qquad (3\text{-}27)$$

An examination of this expression shows that, in the limited frequency range in which the principal resonance effects take place when the secondary Q is not too low, the numerator is substantially constant, whereas the denominator represents the series impedance of the secondary circuit. This is, therefore, an equation of the same general type as Eq. (3-13) for parallel resonance. *The coupled impedance produced by a tuned secondary circuit consequently varies with frequency according to the same general law as does the parallel impedance of the secondary circuit* (see Fig. 3-3). The absolute magnitude of the curve, however, depends upon the mutual inductance. This arrangement thus provides a means whereby the impedance of a parallel resonant circuit can be transformed in magnitude. Comparison of Eqs. (3-13) and (3-27) shows that the transformed impedance appearing in the primary circuit is $(M/L_s)^2$ times the actual parallel impedance of the resonant secondary circuit.

A special case of the circuit of Fig. 3-8*b* that is of particular importance occurs when the primary resistance R_p is the plate resistance of a vacuum tube. One then has the equivalent circuit of the transformer-coupled tuned radio-frequency amplifier. In this instance one is interested in the curve showing the variation of the secondary current (or of the voltage developed across the secondary capacitor C_s)[1] as the frequency is varied

[1] The voltage across the secondary capacitor C_s is equal to the product of the secondary current and the reactance $1/\omega C_s$ of this capacitor. In the limited frequency range represented by the vicinity around resonance ω changes very little in comparison with the variation of the secondary current. Hence, to a first approximation the voltage developed across the capacitor can be considered as being equal to

about resonances. When $R_p >> \omega L_p$, the curve of secondary current (or of voltage across the secondary capacitor) varies with frequency according to a resonance curve having the same resonant frequency as the secondary circuit, but possessing a slightly lower Q. When the reactance ωL_p of the primary inductance is not negligible compared with the primary resistance R_p, the curve of secondary current as a function of frequency still has the shape of a resonant curve. However, the frequency at which the secondary current (or voltage across the secondary capacitor) is maximum is now slightly higher than the resonant frequency of the secondary. A typical example of this is shown by the dotted curve of Fig. 3-10. The analysis that leads to these conclusions is presented in Sec. 3-7.

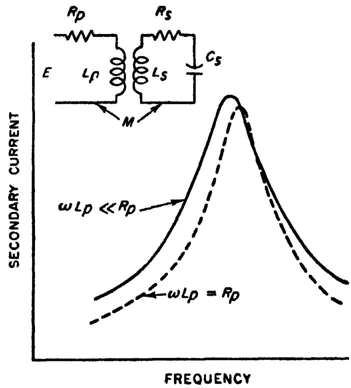


FIG. 3-10. Variation of secondary current as a function of frequency in a coupled system in which the secondary is a resonant circuit and the primary is untuned, showing that the secondary circuit follows a resonance curve, which, however, has a lower Q than that of the secondary circuit taken alone.

3-5. Behavior of Systems Involving Resonant Primary and Resonant Secondary Circuits. *Primary and Secondary Circuits Resonant at the Same Frequency and Having Q's That Are Equal and Not Too Low.* When two resonant circuits having equal Q's that are not too low are tuned to the same frequency and coupled together, the resulting behavior depends very largely upon the degree of coupling, as seen from Fig. 3-11.[1] When the coefficient of coupling is small, the curve of primary current as a function of frequency is substantially the series resonance curve of the primary circuit considered alone. The secondary current is small and varies with frequency in such a way as to be much more peaked than the resonance curve of the secondary circuit considered as an isolated circuit. As the coefficient of coupling is increased somewhat, the curve of primary current becomes broader, as a result of a reduction in the primary current at resonance and an increase in the primary current at frequencies

the product of the secondary current and a constant. In the immediate vicinity of resonance the curve of voltage across this secondary capacitor therefore has very nearly the same shape as does the curve of secondary current.

[1] The phase shift is not shown in Fig. 3-11, but varies ±180° about the phase at the resonant frequency. Thus the total shift in phase between input voltage and output current as the frequency varies through resonance is 360° This is in contrast with systems having only one tuned circuit; the total phase shift then varies over the range ±90°, or a total of 180°.

slightly off resonance. At the same time the secondary-current peak becomes higher and the curve of secondary current somewhat broader.

These trends continue as the coefficient of coupling is increased until the coupling is such that the resistance which the secondary circuit couples into the primary at resonance is equal to the primary resistance. This is called the *critical coupling* and causes the secondary current to

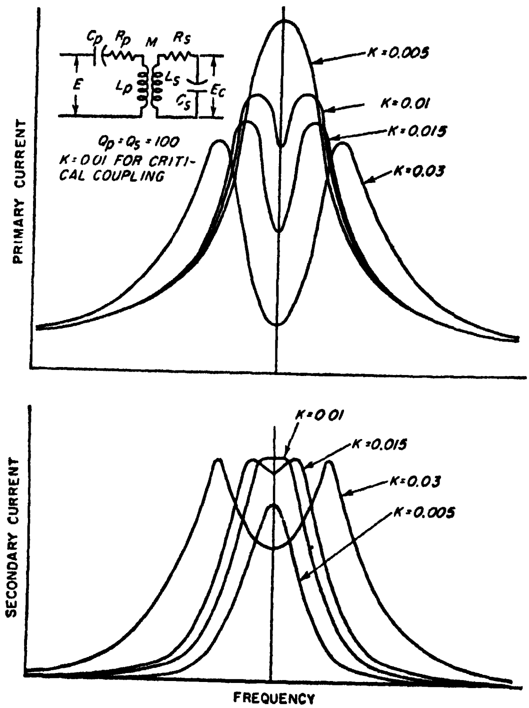


FIG. 3-11. Curves showing variation of primary and secondary currents with frequency for different coefficients of coupling when the primary and the secondary are separately tuned to the same frequency.

have the maximum value it can attain. The curve of secondary current is then somewhat broader than is the resonance curve of the secondary circuit considered alone, and has a relatively flat top. The primary current now has two peaks, being greater at frequencies just off resonance than at the resonant frequency.

As the coefficient of coupling is increased beyond the critical value, the double humps in the primary current become more prominent and the peaks spread farther apart. The curve of secondary current now also begins to display double humps, with the peaks becoming more pronounced and spreading farther apart as the coupling increases. The value of the primary current at the peaks becomes smaller the greater the coupling, but in the secondary circuit not only do the two peaks have substantially the same height, but this height is also independent of the coefficient of coupling provided only that the coupling is not less than the critical value. The reason for the above behavior centers around the way in which the coupled impedance $(\omega M)^2/Z_s$ varies with frequency. Consider first the total primary-circuit impedance. This consists of the actual self-impedance of the primary plus whatever impedance the secondary circuit couples into the primary. The type of coupled impedance produced by a tuned secondary has already been discussed; it is substantially a parallel resonance curve having a shape corresponding to the Q of the secondary circuit and an amplitude determined by the mutual inductance. The coupled impedance is hence maximum at resonance and is then a resistance. At frequencies below resonance the coupled impedance is inductive and at frequencies above resonance it is capacitive, as shown in Fig. 3-7.

When this coupled impedance is added to the self-impedance of the primary circuit, the effect at resonance is to increase the effective primary resistance above the value that would exist in the absence of the secondary. This causes the primary current at resonance to be reduced in all cases by the presence of the secondary. At frequencies somewhat below resonance the coupled impedance is largely inductive whereas the primary self-impedance is largely capacitive. The coupled inductive reactance then neutralizes some of the primary capacitive reactance, lowering the primary circuit impedance and increasing the primary current. The situation is somewhat similar for frequencies above resonance except that now the coupled reactance is capacitive and neutralizes some of the inductive reactance which the primary circuit otherwise has at frequencies above resonance. Consequently, the net effect of the coupled impedance is to lower the primary current at the resonant frequency and to raise the current at frequencies somewhat off resonance. The magnitude of this effect depends upon the coefficient of coupling, being small when the coupling is small. However, when the coupling is of the order of magnitude of the critical value or greater, the coupled impedance becomes sufficient to be the major factor in determining the impedance of the primary circuit. In particular, at resonance the primary current tends to be relatively small because of the very large coupled resistance, while there is a frequency on each side of resonance at which the coupled reactance exactly neutralizes the primary reactance, giving zero reactance for the total primary circuit impedance and causing the flow of a large primary current. This is the cause of the double-humped curves of primary current for high couplings, such as shown in Fig. 3-11.

The curve of secondary current is determined by the secondary impedance, and by the voltage induced in the secondary by the primary current.

The induced voltage varies with frequency in almost exactly the same way as does the primary current, since the magnitude of the induced voltage is $\omega M I_p$; and in the limited frequency range in which the resonance effects take place, ω changes very little. As a result of this, the curve of secondary current has a shape that is almost exactly the product of the shape of the curve of primary current and the shape of the resonance curve of the secondary circuit. Since the latter curve is sharply peaked, the secondary current is much more peaked than the primary current, as is clearly evident in Fig. 3-11.

At low coefficients of coupling, the curve of secondary current is particularly sharp, being substantially the product of the resonance curves of the primary and secondary circuits. As the coupling increases, the primary-current curve becomes broader, thereby making the secondary curve less sharp. At the same time, the amplitude of the secondary-current peak increases because of the increased coupling. When the coefficient of coupling reaches the critical value, the secondary current has the maximum value it can attain. Under these conditions the dip in primary current in the vicinity of resonance has a curvature that is exactly opposite from the curvature of the resonance curve of the secondary circuit. The result is that the curve of secondary current now has a very flat top in the immediate vicinity of resonance. As the coupling is increased beyond the critical value, the secondary-current peak splits into two peaks, both of which have amplitudes substantially the same as the secondary-current peak at critical coupling. The separation between these peaks increases with coupling and is substantially the same as the separation of the peaks of primary current when the peaks are pronounced.

The voltage developed across the secondary capacitor is equal to the reactance of this capacitor times the secondary current; thus it can readily be calculated once the current curve is known. For most purposes, it is sufficient to assume that the curve of voltage developed across the capacitor has the same shape as the curve of secondary current. One is interested primarily in the behavior about resonance, and the capacitor reactance changes very little in the limited frequency range consequently involved when the circuit Q's are not too low.

The exact shapes of curves such as those of Fig. 3-11 can be calculated with the aid of Eqs. (3-20) to (3-24). Such computations are, however, complicated and tedious. The usual practical procedure is accordingly to determine (1) the response at resonance, (2) the frequencies at which the peaks of secondary response occur when this response curve has double humps, (3) the heights of these two peaks, and (4) the response at one or two other frequencies so chosen as to simplify the calculations. In this way, it is possible, with a minimum of work, to obtain a good semiquantitative picture of the behavior. The following nomenclature in addition to that of Fig. 3-11 will be used in the discussion of the quantitative relations:

E_c = voltage across secondary capacitor
E = voltage applied in series with primary
k = actual coefficient of coupling
k_c = critical coefficient of coupling
Q_p = Q of primary circuit
Q_s = Q of secondary circuit

At resonance, the series impedances of the primary and secondary circuits are resistances, and the response in the secondary is given by the relation[1,2]

$$\frac{\left.\begin{matrix}\text{Voltage across secondary}\\ \text{capacitor at resonance}\end{matrix}\right\}}{\left.\begin{matrix}\text{Voltage applied in series}\\ \text{with primary}\end{matrix}\right\}} = \frac{E_c}{E} = \sqrt{\frac{L_s}{L_p}}\,\frac{k}{k^2 + (1/Q_pQ_s)} \tag{3-28}$$

The secondary response has its maximum value when the coefficient of coupling has a value k_c such that

$$k_c = \frac{1}{\sqrt{Q_pQ_s}} \tag{3-29}$$

This value of coupling is called the *critical coefficient of coupling* and is the condition where the resistance that the secondary circuit couples into the primary circuit at resonance is equal to the resistance of the primary circuit, i.e., when $(\omega M)^2/R_s = R_p$.

When the coefficient of coupling equals the critical value and if $Q_p = Q_s$, then the curve of secondary current (or voltage) as a function of frequency has the maximum flatness that is possible in the vicinity of resonance. The shape of this curve is shown in Fig. 3-12, together with the resonance

[1] This follows from Eq. (3-24) by substituting

$$Z_s = R_s,\ Z_p = R_p,\ E_c = I_s/j\omega C_s = -j\omega L_s I_s$$

to give

$$\frac{E_c}{E} = \frac{-j\omega M}{R_pR_s + (\omega M)^2}(-j\omega L_s)$$

Dividing both numerator and denominator by $\omega^2 L_p L_s$ gives

$$\frac{E_c}{E} = \frac{-(M/\sqrt{L_pL_s})\,\sqrt{L_s/L_p}}{\dfrac{R_p}{\omega L_p}\dfrac{R_s}{\omega L_s} + \dfrac{M^2}{L_pL_s}}$$

Equation (3-28) is then obtained by substituting $M^2/L_pL_s = k^2$ and dropping the minus sign.

[2] It is to be noted that Eqs. (3-28) and (3-29) are not limited to the case where $Q_p = Q_s$, although the rest of the discussion in this section does assume $Q_p = Q_s$.

curve of a simple tuned circuit having the same Q as the primary or secondary. It will be noted that the coupled circuit case has a bandwidth between the 70.7 per cent response points that is $\sqrt{2}$ times as great as for the single tuned circuit. The shapes of the two curves differ greatly, the coupled system being flatter in the center and much deeper on the sides.

When the coefficient of coupling exceeds the critical value, then for $Q_p = Q_s$, double humps will always occur in the secondary response

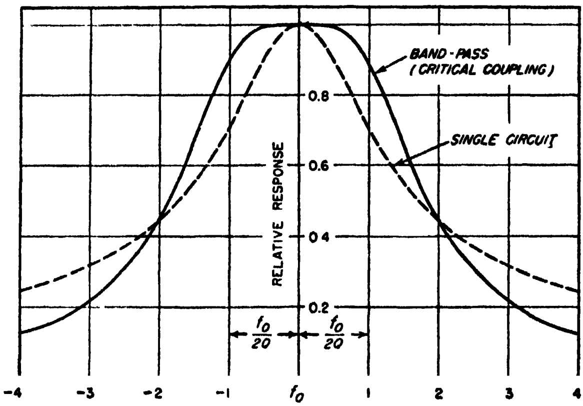


FIG. 3-12. Relative response of a bandpass system provided by two identical circuits critically coupled, together with the resonance curve of a single circuit such as used in the bandpass system.

curve. If the coefficient of coupling is at least several times the critical value these humps are quite pronounced and occur at frequencies that differ from the resonant frequency f_0 by approximately $\pm kf_0/2$ cycles.[1] When the peaks of secondary response are not pronounced, i.e., when the actual coefficient of coupling does not greatly exceed the critical value, then these peaks are somewhat closer together than indicated by this simple relation [see Eq. (3-30) and Fig. 3-16].

When the circuit Q's are equal and not too low, the peaks of the secondary current for $k > k_c$ will have almost exactly the same height as the resonant peak of secondary current at critical coupling. This relation holds irrespective of the exact location on these peaks provided only that

[1] An analysis that does not contain these restrictions leads to the more precise relation

$$\frac{\text{Frequency at peak of secondary voltage}}{\text{Resonant frequency of tuned circuits}} = \sqrt{1 \pm k\left[1 - \frac{k_c^2}{2k^2}\left(\frac{Q_p}{Q_s} + \frac{Q_s}{Q_p}\right)\right]^{1/2}} \quad (3\text{-}30)$$

the coefficient of coupling involved is small compared with unity[1] and that the Q's are not too low.

When double peaks occur in the secondary response curve, additional information on the shape of the response curve can be easily obtained by taking advantage of the fact, illustrated in Fig. 3-13, that the response equals or exceeds the response at resonance over a frequency band that is $\sqrt{2}$ times the width of the frequency band between coupling peaks, as calculated from Eq. (3-30).

At frequencies that are sufficiently high or low relative to the resonant frequency to lie well on the sides of the response curve, one can neglect the resistances of the primary and secondary circuits when calculating the magnitude of the secondary response. This greatly simplifies calculations while introducing relatively little error in magnitudes.

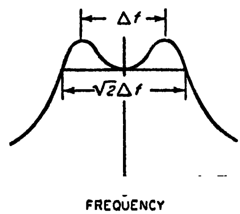


FIG. 3-13. Relationship between bandwidth and width between secondary peaks, existing when two circuits resonant at the same frequency are coupled together.

The Effects Produced by Unequal Q's. The behavior of two coupled circuits resonant at the same frequency is modified in several respects when $Q_p \neq Q_s$. The secondary response at resonance is still given by Eq. (3-28), and is maximum when the coefficient of coupling has the critical value as defined by Eq. (3-29). However, double peaks do not now appear until the coupling is somewhat greater than the critical value, and the magnitude of the response at the secondary peaks when they do appear is less than the response with critical coupling.

Coupled Resonant Circuits Tuned to Slightly Different Frequencies. Consider the case of two circuits resonant at slightly different frequencies and coupled together. When $Q_p = Q_s$, the response curve of secondary current (or voltage) has almost exactly the same shape as would be obtained if the circuits were both tuned to the same frequency and the coefficient of coupling were increased to a value k_{eq} such that

$$k_{eq} = \sqrt{k^2 + \left(\frac{\Delta}{f_0}\right)^2} \tag{3-31}$$

where k is the actual coefficient of coupling, Δ is the difference between the resonant frequencies of the primary and secondary circuits, and f_0 is the frequency midway between the primary and secondary resonant frequencies. *Hence detuning primary and secondary circuits slightly has*

[1] If the coefficient of coupling is not small compared with unity, then the relative heights of the individual peaks of voltage developed across the secondary capacitor will be very nearly inversely proportional to the square of the ratio of the frequencies at which the respective peaks occur. Under these conditions the low-frequency peak will be slightly higher than the high-frequency peak, although the average height of the two peaks will still approximate the response with critical coupling.

approximately the same effect on the shape of the secondary-current curve as increasing the coefficient of coupling when there is no detuning.

In the more general case of detuning where the circuit Q's are not the same, the secondary-response curve is no longer symmetrical about the mean resonance frequency.

Shunt-fed and Shunt-loaded Coupled Circuits. In all the examples of coupled resonant circuits considered so far, the input voltage has been applied in series with the primary circuit. In many practical circumstances, however, the excitation is applied to the system as illustrated in

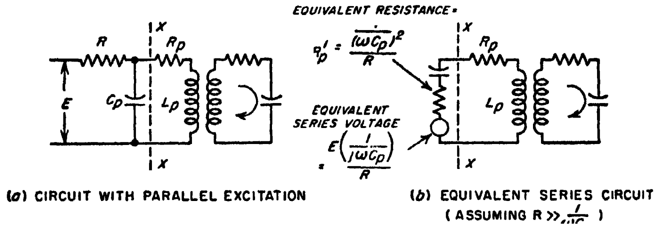


FIG. 3-14. Two coupled resonant circuits in which the primary circuit is excited by shunt feed.

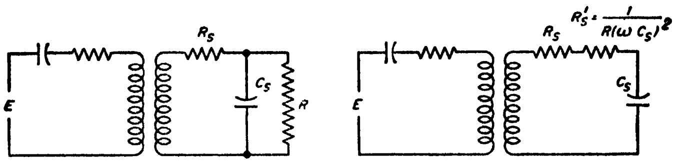
Fig. 3-14a. This arrangement is analogous to the shunt-fed parallel resonant circuit discussed in connection with Fig. 3-5c.

The shunt-fed arrangement of Fig. 3-14a can be reduced to the equivalent series-fed arrangement of Fig. 3-14b by means of Thévenin's theorem, as explained in Sec. 3-7. Examination of the circuit of Fig. 3-14b shows that in a limited frequency range such as represented by the region about resonance, the equivalent voltage acting in series with the circuit is substantially constant. However, there is now an added resistance R_p' in the primary circuit that is equal to the equivalent series resistance that would be obtained by assuming that the resistance R is a shunt resistance for the primary capacitance C_p. The rest of the system is unchanged.

The principal effect of exciting a system of coupled circuits by parallel instead of series feed is accordingly to introduce some added resistance in the primary that lowers the effective value of Q_p. This effect will be slight in the usual case where the resistance R is very large compared with the reactance $1/\omega C_p$ of the capacitor C_p. Under these conditions, shunt feed and series feed accordingly give essentially the same shaped curves of secondary response as a function of frequency.

In systems involving two coupled resonant circuits, resistances are often placed in shunt with the primary and secondary resonant circuits for the purpose of adjusting the effective Q's of the primary and secondary circuits to desired values. Such resistances are sometimes placed across both primary and secondary circuits, while in other cases they are used only across the primary, or only across the secondary. An example where a resistance R is shunted across the secondary capacitor is shown in Fig. 3-15. In each case, a resistance in shunt with a particular resonant circuit of a coupled system has the same effect on that resonant circuit as it does when this resonant circuit is isolated, instead of being part of a coupled system. Hence a shunt resistance can be replaced by an equivalent series resistance, such as R'_s in Fig. 3-15. The effect of a shunting resistance is accordingly to lower the effective Q of the resonant circuit with which it is associated, as discussed in connection with Fig. 3-5.

Bandpass Action in Two Coupled Resonant Circuits. When two resonant circuits having $Q_p = Q_s$ are tuned to the same frequency and



(a) CIRCUIT WITH SHUNT RESISTANCE LOADING

(b) EQUIVALENT CIRCUIT

FIG. 3-15. Two coupled resonant circuits in which a shunt resistance loading is used to control secondary Q.

coupled together with critical coupling, the response characteristic of the secondary circuit is as shown in Fig. 3-12. As compared with the response of a simple resonant circuit with the same 70.7 per cent points, the response of the coupled system is found to be much flatter on top, and much steeper on the sides. Such an arrangement is often termed a *bandpass filter* because to a first approximation it responds equally well to a band of frequencies centered on the common resonant frequency, and rather sharply discriminates against frequencies outside of this band. Such bandpass characteristics are particularly desirable when handling modulated waves, because by proper adjustment of the bandwidth of the filter, the response can be made practically the same to the carrier and to all of the important sideband frequencies contained in the wave. In contrast with this, an ordinary resonant circuit has a response that is rounded on top, as shown dotted in Fig. 3-12, and so discriminates against the higher sideband frequencies in favor of the lower sideband frequencies and the carrier.

The bandpass characteristic that is best for most purposes corresponds to a coefficient of coupling equal to the critical value. For this case, still, assuming $Q_p = Q_s$, the design equations giving the required values of k and Q to realize a given bandwidth B are[1]

$$k = k_c = \frac{B}{\sqrt{2}\, f_0} \qquad Q_p = Q_s = \frac{1}{k_c} \tag{3-32}$$

[1] With unequal circuit Q's the formulas will be slightly different for equivalent results, since the curve with flattest top now corresponds to a coefficient of coupling greater than the critical value.

where B = the bandwidth between the 70.7 per cent response points, cycles

f_0 = center frequency of passband (i.e., resonant frequency of tuned circuits)

k_c = critical coefficient of coupling

Effect of Varying Q in Coupled Systems Tuned to the Same Frequency. (Coefficient of Coupling Constant). Additional insight into the characteristics of coupled circuits can be gained by considering what happens to the secondary response curve as the Q's of the primary and secondary circuits are changed, while keeping the coefficient of coupling constant. The effects observed are illustrated in Fig. 3-16 for a particular case. This example brings out clearly the fact that as the peaks of the response become less pronounced, they tend to move toward each other, and that at frequencies appreciably off resonance, the response differs only negligibly from the response calculated on the assumption of infinite Q (zero circuit loss).

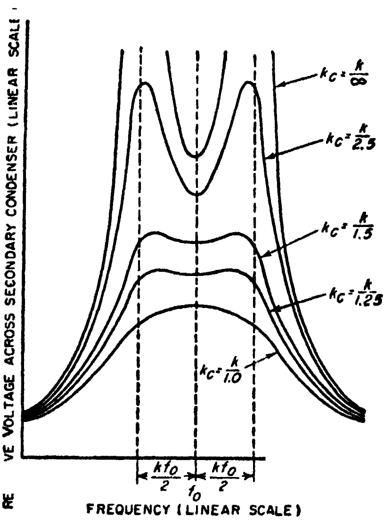


FIG. 3-16. Curves of secondary response when two circuits resonant at the same frequency are coupled together, showing the effect of varying the circuit Q's while maintaining the coefficient of coupling unchanged.

3-6. Generalized Coupled Circuits. Energy can be transferred from one circuit to another by a variety of coupling methods, in addition to the inductive coupling just considered. Thus, in Fig. 3-17*a* the coupling consists of an inductance L_m common to the two circuits; in Fig. 3-17*b* the coupling is provided by a capacitance C_m common to the two circuits, and in Fig. 3-17*c* by a capacitance C'_m that connects the two circuits involved. Also, an infinite variety of more complicated coupling systems can be built up from the basic elements of mutual inductance, common inductance, common capacitance, and connecting capacitance. Simple examples of such combined couplings are shown in Fig. 3-17*d* and *e*.

The behavior of all these coupled circuits follows the same general character as that discussed for inductive coupling. Thus, the secondary circuit can be considered as producing an equivalent coupled impedance in the primary circuit while the primary circuit can be considered as inducing in the secondary a voltage that gives rise to the secondary current.

The simplest method of analyzing these various forms of coupled circuits is to take advantage of the fact that all of them can be reduced to the simple coupled circuit of Fig. 3-17*f*, provided suitable values are assigned to Z_p, Z_s, and M. The rules that determine the values of these quantities in the simple equivalent circuit are as follows:

1. The equivalent primary impedance Z_p of the equivalent circuit is the impedance that is measured across the primary terminals of the actual circuit when the secondary circuit has been opened.

2. The secondary impedance Z_s of the equivalent circuit is the impedance that is measured by opening the secondary of the actual circuit and

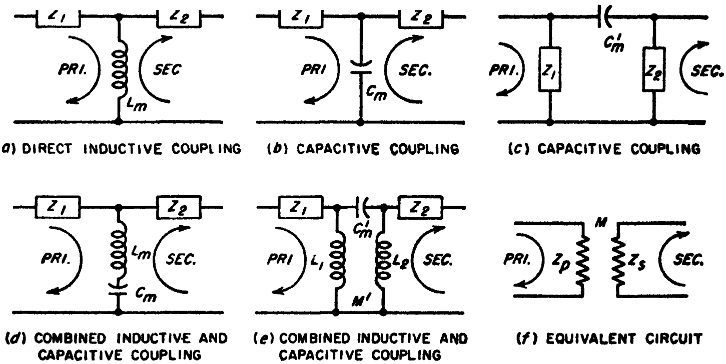


FIG. 3-17. Examples of methods whereby circuits may be coupled.

determining the impedance between these open points when the primary is open-circuited.

3. The equivalent mutual inductance M is determined by assuming a current I_p flowing into the primary circuit. The voltage which then appears across an open circuit in the secondary is equal to $-j\omega MI_p$.

In making use of the equivalent circuit of Fig. 3-17*f*, it is to be remembered that the values of Z_p, Z_s, and M may all vary with frequency, so that it is generally necessary to determine a new equivalent circuit for each frequency at which calculations are to be made.

After the actual coupled circuit has been reduced by the above procedure to its equivalent form shown in Fig. 3-14*f*, one can then apply the formulas that have already been derived for inductively coupled circuits, using the appropriate values M, Z_s, Z_p as determined for the equivalent circuit. This procedure has the advantage of using the same fundamental formulas to handle all types of coupling and makes it possible to carry on the analysis in the same manner for all cases. The method is particularly convenient in the handling of complex coupling networks such as illustrated in Fig. 3-17*d* and *e*.

The quantity M that appears in the equivalent circuit represents the effective coupling that is present between the primary and secondary circuits. It is not necessarily a real mutual inductance of the inductive type, but rather a sort of mathematical fiction that gives the equivalent effect of whatever coupling is really present. If the actual coupling is capacitive, the numerical value of M will be found to be negative; if the coupling is of a complex type representing both resistive and reactive coupling, the numerical value of M will be found to have both real and imaginary parts. This need introduce no uncertainty, however, since the proper procedure is to take the value of M as it comes and substitute it with its appropriate sign and phase angle whenever M appears in the expressions previously derived for inductively coupled circuits.

When this analysis is applied to capacitively coupled circuits, such as those illustrated in Fig. 3-18, the results are essentially the same as for

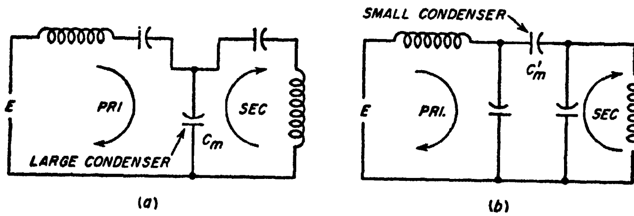


FIG. 3-18. Two methods of capacitively coupling two resonant circuits.

inductive coupling. Thus, when primary and secondary are both tuned to the same frequency, the secondary-current characteristic has two humps if the coupling is large, i.e., if capacitor C_m is small or C'_m large, while there is only one peak of secondary current when the coupling is small, i.e., when capacitor C_m is large.

Circuits having combined electromagnetic and electrostatic coupling, such as those at *d* and *e* of Fig. 3-17, behave as ordinary coupled circuits except that the coefficient of coupling varies with frequency. Thus, in the case of circuit *d*, the circuit is capacitively coupled at low frequencies and inductively coupled at high frequencies because the coupling combination of C_m in series with L_m has capacitive and inductive reactance under these respective conditions. In between, at the resonant frequency of L_m and C_m, there is no coupling and $k = 0$. The arrangement shown at *e* acts similarly as a circuit with a coefficient of coupling that varies with frequency. Circuits having combined electrostatic and electromagnetic coupling find application where it is desired to obtain a coefficient of coupling that varies with frequency, as is commonly the case in tuned amplifiers and antenna-coupling circuits of radio receivers.

3-7. Thévenin's Theorem. *According to Thévenin's theorem, any linear network containing one or more sources of voltage and having two terminals behaves, in so far as a load impedance connected across these terminals is concerned, as though the network and its generators were equivalent to a simple generator having an internal impedance Z and a generated voltage E, where E* *is the voltage that appears across the terminals when no load impedance is connected and Z is the impedance that is measured between the terminals when all sources of voltage in the network are short-circuited.*[1,2]

This theorem means that any network and its generators, represented schematically by the block in Fig. 3-19*a*, can be replaced by the equivalent circuit shown in Fig. 3-19*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.

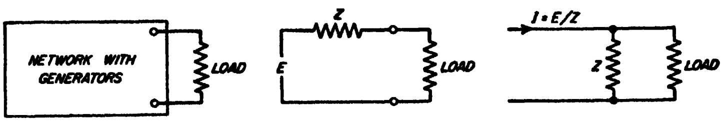


FIG. 3-19. Diagrams illustrating how Thévenin's and Norton's theorems can be used to simplify a complicated network containing generators.

Thévenin's theorem offers a very powerful means of simplifying networks, particularly when a load impedance is connected across the output terminals of a complicated network. Two examples will be used to illustrate this. First, consider the circuit of Fig. 3-10, which is redrawn in Fig. 3-20*a*. If one takes the secondary capacitor C_2 as the load impedance and applies Thévenin's theorem to the network to the left of C_2, the result is Fig. 3-20*b*, in which the equivalent generator voltage is the voltage induced in the secondary inductance L_2 when the secondary is open-circuited, and the equivalent generator impedance consists of the inductance L_2 and the resistance R_2 in series with the impedance which is coupled into L_2 by a secondary circuit consisting of L_1 shunted by the resistance R_1. The coupled impedance produced by such a secondary circuit has been previously considered; it is equivalent to adding capacitive reactance and resistance in series. The resistance causes the effective Q of the secondary-response curve to be reduced, while the series capaci-

[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 observed 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.

[2] An alternative circuit that is also equivalent to Fig. 3-19*a* is given in Fig. 3-19*c*. Here the network with its generators is replaced by a constant current I that is delivered to a system consisting of the source impedance Z in shunt with which is the load impedance, where I is the output current of the network when the output terminals are short-circuited, and is $I = E/Z$. The equivalence of the arrangements at *a* and *c* in Fig. 3-19 is sometimes referred to as *Norton's theorem.*

tive reactance tends to raise the apparent resonant frequency by an amount that becomes greater the higher the ratio $\omega L_1/R_1$. This accounts for the behavior of the curves of Fig. 3-10.

The second example is furnished by Fig. 3-14*a*. This circuit may be simplified by considering that the load is represented by the circuit to the right of the line xx, the generator being the voltage E acting in series with the resistance R and the capacitance C_p. Such a generator can be reduced immediately by Thévenin's theorem to the form shown to the left of the line xx in Fig. 3-14*b*. Here it is to be noted that the equivalent generator resistance R'_p is the series resistance equivalent to a shunt resistance R associated with the capacitance C_p, as given by Eqs. (2-12) and (2-13).

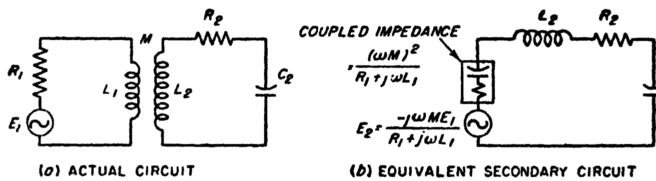


FIG. 3-20. Application of Thévenin's theorem to simplify and explain the behavior of the system of Fig. 3-10, consisting of a tuned secondary and untuned primary circuit coupled together.

3-8. Impedance Matching. A load connected across the output terminals of a network, such as represented schematically by Fig. 3-19*a*, can be matched to the source of power in either of two ways. When the load impedance has the same magnitude and phase angle as the equivalent generator impedance Z defined by Thévenin's theorem (see Fig. 3-19*b*), the load is said to be matched to the generator or source of power on an *image-impedance* basis. The term "image" arises from the fact that the impedances on the two sides of the output terminals are images of each other. When the load impedance is not identical with the generator impedance and it is desired to obtain impedance matching on an image basis, it is then necessary to transform the load to the correct impedance to match the generator. This transformation can be accomplished with the aid of an appropriate network of reactances or, in simple cases, by means of a transformer.

Alternatively, a load impedance may be matched to a source of power in such a way as to make the power delivered to the load a maximum.[1] This is accomplished by making the load impedance the conjugate of the generator impedance as defined by Thévenin's theorem. That is, the load impedance must have the same magnitude as the generator impedance, but the phase angle of the load is the negative of the phase angle of the generator impedance. This method of matching is shown schemat-

[1] The power delivered to the load under these conditions is termed the *available power* of the power source.

ically in Fig. 3-21. It will be noted that the reactive component of the load is then in series resonance with the reactive component of the generator impedance; i.e., the load reactance is the correct value to "tune out" the generator reactance. The resistance components of the load and generator impedances are then matched on an image-impedance basis. Such impedance matching to obtain maximum power delivered to the load is a common operation in communication circuits. It is carried out by transforming the equivalent series resistance of the load to a value equal to the resistance component of the generator impedance by the use of suitable networks and transformers, and then adding reactance to the load as required to resonate with the generator reactance.

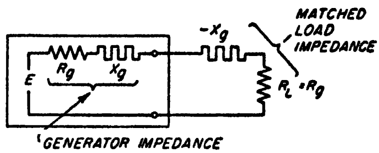


FIG. 3-21. Load impedance matched to generator in such a way as to give maximum power in the load.

It will be noted that, when the generator impedance is resistive, the conditions corresponding to matching on an image-impedance basis are identical with those corresponding to matching for maximum power output delivered to the load. Otherwise, the two conditions are not the same, and matching on an image-impedance basis then does not result in maximum possible power being delivered to the load, although it is often still used to maintain appropriate impedance relations in a system of networks.

PROBLEMS AND EXERCISES

3-1. The coil of Fig. 2-16 is tuned to resonance at 1000 kc by a capacitor having a power factor of 0.001. What is the circuit Q?

3-2. In Prob. 3-1, what tuning capacitance is required?

3-3. A variable capacitor 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. The coil and associated wiring have a distributed capacitance of 20 $\mu\mu$f.

a. What size inductance coil is required to make the lowest frequency 530 kc?

b. Calculate the exact tuning range with the coil selected.

3-4. A series circuit is resonant at 800 kc and has an inductance of 160 μh and a circuit Q of 75. Calculate and plot the magnitude of the current that flows when 1 volt is applied to the circuit, carrying the curves out to 40 kc on each side of resonance. In making these calculations use the working rules and 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.

3-5. In Prob. 3-4 calculate the exact response at 40 kc above resonance, taking into account the circuit resistance, and compare the results with those obtained when the circuit resistance is neglected.

3-6. Assume that a series resonant circuit employs the coil of Fig. 2-16, and that the tuning capacitor has negligible losses.

a. Calculate and plot from 500 to 1500 kc 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.

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

3-7. 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 the same applied voltage. From this information determine the Q of the circuit.

3-8. 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.

3-9. In variable capacitors 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 the angle of rotation than if semicircular plates were employed.

3-10. What is the highest effective Q that a tuned circuit may have when it must respond to a band of frequencies 10,000 cycles wide (5000-cycle sideband frequencies) with a response always at least 70.7 per cent of the response at resonance, assuming carrier frequencies of 50, 500, 5000, and 50,000 kc?

3-11. *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 out to 60 kc on each side of resonance. Use the working rules and 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, and plot the results on the same axes as the results of (*a*).

3-12. Calculate and plot as a function of frequency the parallel impedance at resonance when the coil of Fig. 2-16 is tuned with a capacitor of negligible losses and when the resonant frequency is varied from 500 to 1500 kc.

3-13. A tuned circuit is required to have a parallel impedance of 6000 ohms and a Q of 12. If the resonant frequency is 300 kc determine the inductance, capacitance, and resistance that the circuit must have.

3-14. Using the same tuned circuit as in Prob. 3-4, but connected for parallel resonance, calculate and plot curves as a function of frequency from 760 to 840 kc for (*a*) magnitude and phase angle of parallel impedance; (*b*) line current, and current in each branch, when the applied potential is 10 volts (assume that all the circuit resistance is in the inductive branch); and (*c*) reactance and resistance components of the impedance of (*a*).

3-15. The coil of Fig. 2-16 is tuned to resonance at 1000 kc with a capacitor having negligible losses. Transform this circuit to the form shown in the right-hand part of Fig. 3-5*a* by determining R_2.

3-16. The circuit of Fig. 2-16 is tuned to resonance at 1000 kc with a capacitor having negligible losses, and is then shunted by a resistance R_2' of 100,000 ohms.

a. Determine the equivalent shunt resistance R_2 for such an arrangement (see Fig. 3-5).

b. Calculate the Q of this system.

3-17. If a parallel resonant circuit is shunted by a resistance R_2' and if the parallel resonant impedance of the unshunted circuit is R_0, prove that the shunt resistance R_2' reduces the equivalent Q of the circuit by the factor $R_2'/(R_2' + R_0)$.

3-18. In a low-Q parallel circuit in which the losses are all in the inductive branch, prove that, when the capacitance is varied, the capacitance that makes the parallel circuit impedance have unity power factor for a given frequency also makes this impedance have maximum magnitude at this same frequency.

3-19. A particular coil has an inductance of 180 μh at 1 kc and an apparent inductance of 200 μh at 1400 kc. Determine the distributed capacitance of the coil.

3-20. Primary and secondary coils have inductances of 75 and 300 μh, respectively, and 1 volt is applied to the primary circuit. Assuming the resistances of the coils are negligible, calculate the voltage induced in the secondary as a function of coefficient of coupling from $k = 0$ to $k = 1.0$.

3-21. Draw an equivalent transformer circuit for the coils of Prob. 3-20, for the case where the mutual inductance is 50 μh.

3-22. *a.* Explain the effect of a short-circuited turn upon the inductance and Q observed at the terminals of a coil, using coupled-circuit theory.

b. Indicate qualitatively the differences that would be expected if the short-circuited turn were the end turn of a single-layer solenoid, as in Fig. 2-1, as against being a turn near the center.

3-23. Two identical coils each having $Q = 100$ and an inductance of 200 μh are coupled together with a mutual inductance of 50 μh. If the secondary coil is short-circuited, calculate (*a*) the coupled resistance and coupled reactance at a frequency of 600 kc, (*b*) the total resistance and reactance of the primary circuit, and (*c*) the effective Q of the primary circuit including effect of the coupled impedance.

3-24. Describe a procedure for experimentally determining the coefficient of coupling between a coil and its shield can, assuming that the shield has negligible resistance.

3-25. Derive the formula in the second footnote on page 61 for the equivalent primary-circuit inductance in the presence of an inductive secondary with zero losses.

3-26. An air-cored coil is placed near a brass panel. Describe in a qualitative way the effect that copper plating this panel will have on the inductance and Q observed at the coil terminals.

3-27. The coil of Fig. 2-16 is coupled to a primary coil with a mutual inductance of 50 μh. If the secondary coil is tuned to resonance by means of a capacitor having negligible loss, calculate and plot the coupled impedance at the resonant frequency of the secondary as this resonant frequency is varied from 500 to 1500 kc.

3-28. The coil of Fig. 2-16 is coupled to a primary circuit having an inductance of 75 μh, and is tuned to resonance at 1000 kc with a capacitor having negligible losses. Calculate the impedance coupled into the primary circuit at 1000 kc as a function of coefficient of coupling from $k = 0$ to $k = 1.0$.

3-29. In the circuit of Fig. 3-8*b*, what general effect is produced on the phase and magnitude of the coupled impedance at the resonant frequency of the secondary by shunting the secondary capacitor C_s by a resistance R_2?

3-30. Explain why in Fig. 3-11 a flat-topped secondary-circuit curve (like $k = 0.01$) can be obtained only if the primary-current curve has pronounced double peaks.

3-31. Derive Eq. (3-29) from Eq. (3-28).

3-32. 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.

3-33. The coupling between the circuits of Prob. 3-32 is adjusted to make the coefficient of coupling have a value 0.03, 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? Assume the two peaks have equal heights.

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

d. Over what frequency range will the secondary response equal or exceed the secondary response at resonance?

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

3-34. The circuits of Prob. 3-32 are coupled with a coefficient of coupling of 0.1. Determine the frequencies at which the secondary-current peaks occur, and give the approximate ratio of voltages across the secondary at frequencies corresponding to the low- and high-frequency peaks.

3-35. The two circuits of Prob. 3-32 are coupled with a mutual inductance of 2.8 μh(k = 0.02).

a. Calculate and plot the resistance and reactance components of the coupled impedance out 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 total primary impedance in the presence of the secondary.

3-36. If, in Prob. 3-35, the mutual inductance had a value of 1 μh, then to what frequencies would it be necessary to tune the primary and secondary circuits in order to obtain the same shape of secondary-response curve as is actually obtained for the conditions given in Prob. 3-35?

3-37. In a shunt-feed circuit such as illustrated in Fig. 3-14, the tuned circuits are the same as in Prob. 3-32, and the shunt-feed resistance R is 100,000 ohms. What is the equivalent primary Q under these conditions?

3-38. The two resonant circuits in Fig. 3-15 are the same as in Prob. 3-32. What value must R have to make the effective Q of the secondary equal to 40?

3-39. A particular bandpass filter is to be used to handle a wave in which the highest modulation frequency is 4000 cycles. The carrier frequency of the wave is 456 kc. If the primary and secondary inductances are both 2 mh and if it is desired just barely to avoid double humps in the response curve, specify the proper coefficient of coupling and the proper circuit Q's, assuming equal primary and secondary Q's.

3-40. Two identical tuned circuits are used in a shunt-feed bandpass arrangement. The circuits are resonant at 450 kc, have inductances of 2.0 mh, and Q's of 80. The shunt-feed resistance has a value of 300,000 ohms. A bandwidth between 70.7 per cent response points of 30 kc is desired.

a. Calculate required values of circuit Q's, assuming $Q_p = Q_s$.

b. Determine the resistance that must be shunted across the secondary capacitor to make the effective Q of the secondary circuit have the required value.

c. Determine the resistance that must be shunted across the primary capacitor to make the effective Q of the primary have the required value when the effect of both the shunt-feed resistance and the primary-circuit resistance are taken into account.

3-41. According to Fig. 3-16 the response at resonance will increase as the Q is increased while leaving the coefficient of coupling unchanged.

a. Demonstrate that this result is predicted by Eq. (3-28).

b. Determine the ratio of response at resonance for zero circuit losses to the response for k = 0.01 when the circuit losses make k = 0.01 correspond to critical coupling.

3-42. Calculate the coefficient of coupling in the circuit of Fig. 3-18*b* when $C_1 = C_2 = 100\ \mu\mu f$, and $C'_m = 1.5\ \mu\mu f$.

3-43. Signals in the frequency range of 550 to 1500 kc are to be handled by means of a bandpass filter. If the circuits are assumed to have $Q = 100$ over this frequency range, and if the adjustment is such that $k = 0.01$ at 1000 kc, discuss how the width and shape of the passband will vary with resonant frequency when the tuning is obtained by varying the primary and secondary capacitors simultaneously and when the coupling is (*a*) inductive as shown at Fig. 2-8*b*, and (*b*) capacitive as shown at Fig. 2-8*c*. Assume that the circuit elements that provide the coupling do not change as the capacitors are varied. Illustrate the discussion with the aid of sketches showing in a general way the relative character of the response curves to be expected at 550, 1000, and 1500 kc for each type of coupling.

3-44. Explain how the magnitudes of the Thévenin-theorem equivalent voltage and impedance for a complex network can be determined experimentally from an open- and short-circuit test at the output terminals of the network, using only a voltmeter and an ammeter.

3-45. In Fig. 3-20 (also Fig. 3-10) the secondary circuit has an inductance of 150 μh, and is resonant at 1000 kc. If $R_1 = 10{,}000$, $L_1 = 150\ \mu h$, and $M = 100\ \mu h$, calculate the frequency at which the peak of secondary response occurs.

3-46. A primary circuit has an inductance of 1 mh and a resistance of 150 ohms connected in series. A secondary coil is coupled to the primary coil and delivers power to a load consisting of the secondary coil, a resistance of 50 ohms, and a tuning capacitance, all in series. If the impedance that the secondary circuit couples into the primary circuit is considered to be the load impedance of the primary circuit, determine the mutual inductance required between the two circuits and the reactance that the secondary circuit must have if the load is to match the generator on a maximum-power basis.

3-47. *a.* In order to demonstrate impedance matching for maximum-power transfer, write the equation of power P delivered to a rheostat as a function of its resistance R when connected to a d-c generator of internal resistance R_s and open-circuit voltage E_s. Show that this equation has a maximum for $R = R_s$.

b. Plot a graph of the equation of (*a*), showing P/P_{max} versus R/R_s, where P is the actual power when the load resistance is R, and P_{max} is the power when $R = R_s$. By how many decibels is the power reduced for the following cases of mismatch: (1) $R = 0.5R_s$ and (2) $R = 2R_s$?

CHAPTER 4

TRANSMISSION LINES

4-1. Voltage and Current Relations on Radio-frequency Transmission Lines in Terms of Traveling Waves.[1] Transmission lines find many uses in radio work. They are employed, not only to transmit energy, but also as resonant circuits at very high frequencies, as measuring devices at high frequencies, as aids to obtain impedance matching, etc.

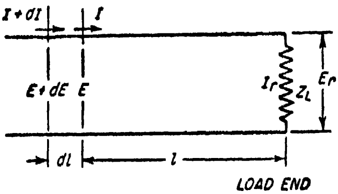


FIG. 4-1. Transmission line, showing elementary length dl.

Basic Transmission-line Equations. Consider the voltage and current relations that exist in a very short length dl of the transmission line shown in Fig. 4-1. In this short distance the voltage between the wires changes an amount dE as a result of the voltage drop produced by the line current I flowing through the resistance $R\,dl$ and reactance $j\omega L\,dl$ of the length dl. Likewise, the current changes a small amount dI in the length as a result of the flow of current between the wires through the capacitance $C\,dl$ and conductance $G\,dl$ caused by the voltage that exists between these wires. Referring to Fig. 4-1, one can accordingly write the equations

$$
\begin{aligned}
dE &= I \times (\text{impedance of length } dl) \\
&= I(R + j\omega L)\,dl \\
dI &= E \times (\text{admittance of length } dl) \\
&= E(G + j\omega C)\,dl
\end{aligned}
$$

[1] This material on transmission lines is a review and summary of those concepts and relations that are most widely used in radio work. It presupposes at least a little previous familiarity with the subject, and therefore should not be regarded as a self-supporting presentation of transmission-line theory. The reader desiring to gain a comprehensive understanding of transmission lines, or desiring the derivation of the equations made use of here, should consult one of the several excellent textbooks that are available on the subject, for example, H. H. Skilling, "Electric Transmission Lines," McGraw-Hill Book Company, Inc., New York, 1951; Walter C. Johnson, "Transmission Lines and Networks," McGraw-Hill Book Company, Inc., New York, 1950. More limited treatments of transmission lines, typically of chapter length, are to be found in most textbooks on communication engineering; these are adequate as an introduction to the material presented here.

Rearranging,

$$\frac{dE}{dl} = (R + j\omega L)I = ZI \tag{4-1a}$$

$$\frac{dI}{dl} = (G + j\omega C)E = YE \tag{4-1b}$$

where E = voltage across line at distance l from receiving end
I = current in line at distance l from receiving end
l = distance measured from load end of line
R = resistance per unit length, ohms
L = inductance per unit length, henrys
C = capacitance per unit length, farads
G = conductance per unit length, mhos
$Z = (R + j\omega L)$ = line series impedance per unit length, ohms
$Y = (G + j\omega C)$ = line shunt admittance per unit length, mhos
$\omega/2\pi$ = frequency, cycles

Simultaneous solution of Eq. (4-1) gives[1]

$$\frac{d^2E}{dl^2} = ZYE \tag{4-2a}$$

$$\frac{d^2I}{dl^2} = ZYI \tag{4-2b}$$

Equations (4-2*a*) and (4-2*b*) are not independent of each other, since they are related through Eqs. (4-1*a*) or (4-1*b*).

Equations (4-2*a*) and (4-2*b*) are the standard differential equations of wave propagation and have solutions of the form

$$E = E_1\epsilon^{\sqrt{ZY}l} + E_2\epsilon^{-\sqrt{ZY}l} \tag{4-3a}$$

$$I = I_1\epsilon^{\sqrt{ZY}l} + I_2\epsilon^{-\sqrt{ZY}l} \tag{4-3b}$$

where E_1, E_2, I_1, and I_2 are constants of integration whose values are determined by the boundary conditions, i.e., by the load impedance and the magnitude of the voltage applied to the system. Although four constants appear in Eqs. (4-3), actually only two of them are independent since it can be readily shown that[2]

$$I_1 = \frac{E_1}{\sqrt{Z/Y}} = \frac{E_1}{Z_0} \tag{4-4a}$$

$$I_2 = \frac{-E_2}{\sqrt{Z/Y}} = \frac{-E_2}{Z_0} \tag{4-4b}$$

[1] These results are obtained by differentiating Eq. (4-1*a*), and then substituting Eq. (4-1*b*) to eliminate the resulting dI/dl. This gives

$$\frac{d^2E}{dl^2} = Z\frac{dI}{dl} = ZYE$$

Equation (4-2*b*) is obtained in an analogous manner.

[2] These relations are obtained by substituting Eq. (4-3*a*) in Eq. (4-1*a*), and then comparing the result with Eq. (4-3*b*).

Here

(4-5)

The final solution of the differential Eqs. (4-1*a*) and (4-1*b*) of the transmission line can accordingly be written as

$$E \quad E_1\epsilon^{\sqrt{ZY}l} + E_2\epsilon^{-\sqrt{ZY}l} = E' + E'' \tag{4-6a}$$

$$\frac{E_1}{\sqrt{Z/Y}}\epsilon^{\sqrt{ZY}l} \quad \frac{E_2}{\sqrt{Z/Y}}\cdot^{-\sqrt{ZY}l} = I' + I'' \tag{4-6b}$$

In these equations $Z_0 = \sqrt{Z/Y}$ is termed the *characteristic impedance* of the line. In the case of radio-frequency lines, Z_0 can nearly always be assumed to be a pure resistance, as discussed on page 88.

The quantity $\sqrt{ZY}$ is called the *propagation constant* of the line. It is a complex quantity, having a real part α called the *attenuation constant* and an imaginary part β termed the *phase constant.* That is

$$\sqrt{ZY} = \alpha + j\beta \tag{4-7}$$

4-2. Interpretation of Transmission-line Equations in Terms of Traveling Waves. The voltage and current existing on a transmission line as given by Eqs. (4-6) can be conveniently expressed as the sum of the voltages and currents of two waves. One of these waves can be regarded as traveling toward the receiving or load end of the line, and is called the *incident wave* because it is incident upon the load. The second wave can be thought of as traveling from the load toward the generator end of the line; it is termed the *reflected wave,* and is generated at the load by reflection of the incident wave.′ These two waves are identical in nature except for consequences arising from their different directions of travel.

The Incident Wave. The incident wave consists of the voltage component E' of Eq. (4-6*a*) associated with the current component I' of Eq. (4-6*b*). For such a wave it follows that everywhere on the line

$$\frac{E'}{I'} = Z_0 \tag{4-8}$$

The magnitude $|E'|$ of the incident wave becomes larger as the distance l from the load increases, according to the relation

$$|E'| = |E_1\epsilon^{(\alpha+j\beta)l}| = |E_1|\epsilon^{\alpha l} \tag{4-9}$$

In this equation E_1 is the vector value of the voltage of the incident wave at the load end of the line, and α is the attenuation constant,[1] as defined

[1] The unit of α in Eq. (4-9) is the neper. In discussing attenuation of lines, values of α (or of αl) are, however, frequently described in decibels. The relation between nepers and decibels is

$$\text{Attenuation in decibels} = 8.686\alpha \tag{4-9a}$$

by Eq. (4-7). The quantity αl, the total attenuation of the line, is commonly called simply the line attenuation.

The phase of the incident wave advances β radians per unit distance from the load, where β is the phase constant as defined by Eq. (4-7). Hence the phase position of the incident wave at a distance l from the load leads the phase position at the load by βl radians.

The incident wave on the transmission line can therefore be described as a voltage accompanied by a current that is everywhere in phase with,

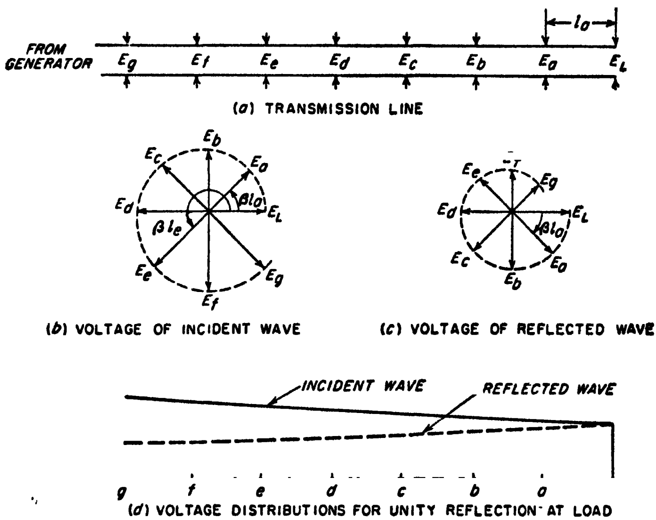


FIG. 4-2. Diagrams illustrating behavior of the voltage of the incident and reflected waves on a transmission line. The case shown assumes that the reflection coefficient at the load is unity, and that the line attenuation is only moderate. The clock diagrams show magnitude and phase of the voltage at increasing values of distance from the load.

and proportional to, the voltage, with the voltage and hence current decreasing exponentially in magnitude and dropping back uniformly in phase as the load is approached. Such a distribution is illustrated in Fig. 4-2, and can be represented by the equations

$$E' = E_1 \epsilon^{(\alpha + j\beta)l} \qquad (4\text{-}10a)$$

$$I' = \frac{E'}{Z_0} = \frac{E_1}{Z_0} \epsilon^{(\alpha + j\beta)l} \qquad (4\text{-}10b)$$

The incident wave is said to travel toward the load because it gets smaller as the load is approached and because its phase at a distance from the generator corresponds to the phase that existed at the generator at an earlier time proportional to distance. These are properties of a wave propagating away from a source. The velocity of propagation, called the *phase velocity*, is discussed below in connection with Eq. (4-19b).

The Reflected Wave. The reflected wave is identical with the incident wave except that it is traveling toward the generator. The reflected wave thus consists of the component voltage E'' of Eq. (4-6*a*) associated with a current component I'' such that everywhere on the line

$$\frac{E''}{I''} = -Z_0 \tag{4-11}$$

This differs from Eq. (4-8) only by the negative sign, which arises from the fact that the current in the reflected wave travels toward the generator, whereas the current in the incident wave travels toward the load.

The magnitude $|E_2|$ of the reflected wave becomes smaller as the wave travels away from the receiver (i.e., as l increases) according to the relation

$$|E''| = |E_2\epsilon^{-(\alpha+j\beta)l}| = |E_2|\epsilon^{-\alpha l} \tag{4-12}$$

Here E_2 is the vector value of the reflected wave at the load. Equation (4-12) is similar to Eq. (4-9) except for the negative sign in the exponent; this denotes a decrease in magnitude with increasing distance l from the receiver.

The phase of the reflected wave drops back β radians for each unit of distance that the wave travels toward the generator. Thus the reflected wave at a distance l from the load lags the phase position at the load by βl radians.

As a result of these properties, the reflected wave on the transmission line can be described as a voltage accompanied by a current proportional to the voltage and flowing *away* from the load, with the voltage, and hence current, decreasing exponentially in magnitude and dropping back uniformly in phase as the distance from the load *increases*. Such a distribution is illustrated in Fig. 4-2, and can be represented by the equations

$$E'' = E_2\epsilon^{-(\alpha+j\beta)l} \tag{4-13a}$$

$$I'' = -\frac{E''}{Z_0} = -\frac{E_2}{Z_0}\epsilon^{-(\alpha+j\beta)l} \tag{4-13b}$$

Relation of Incident and Reflected Waves—Reflection Coefficient. The reflected wave is generated at the load as a result of reflection of the incident wave by the load impedance. This reflection is of such a character as simultaneously to meet the following conditions: (1) The voltage and current of the incident wave at the load must satisfy Eq. (4-8); (2) the voltage and current of the reflected wave at the load must satisfy Eq. (4-11); (3) the load voltage E_L is the sum of the voltages of the incident and reflected waves at the load, that is, $E_L = E_1 + E_2$; (4) the load current I_L is the sum of the currents of the incident and reflected waves at the load, that is, $I_L = I_1 + I_2$; and (5) the vector ratio E_L/I_L must equal the load impedance Z_L.

The vector ratio E_2/E_1 of the voltage of the reflected wave to the voltage of the incident wave at the load is termed the *reflection coefficient* of the load. Simultaneous solution of the above five relations leads to the result

$$\text{Reflection coefficient} = \rho = \frac{E_2}{E_1} = \frac{(Z_L/Z_0) - 1}{(Z_L/Z_0) + 1} \tag{4-14}$$

The reflection coefficient has both magnitude and phase, and so is a vector quantity. Although Eq. (4-14) is expressed in terms of the situation at the load, the ratio E''/E' of the voltages of the reflected and incident waves at a distance l from the load can be termed the reflection coefficient at the point l. It will be noted that when $\alpha = 0$ (i.e., zero losses on the line), the reflection coefficient everywhere has the same magnitude, and equals the reflection coefficient of the load. However, when $\alpha \neq 0$, then the reflected wave becomes smaller and the incident wave larger with increasing distance from the load, causing $|\rho|$ to decrease correspondingly. The quantitative relation is

$$|\rho_b| = |\rho_a|\epsilon^{-2\alpha(l_b - l_a)} \tag{4-15}$$

where $|\rho_a|$ and $|\rho_b|$ are the magnitudes of the reflection coefficients at distances l_a and l_b, respectively, from the load.

The relation between the load voltage and current and the voltages of the incident and reflected waves at the load can be deduced from the above five required conditions. It is

$$E_1 = \frac{E_L}{1 + \rho} = \left(\frac{E_L + I_L Z_0}{2}\right) \tag{4-16a}$$

$$E_2 = \rho E_1 = \frac{\rho}{1 + \rho} E_L = \left(\frac{E_L - I_L Z_0}{2}\right) \tag{4-16b}$$

The corresponding currents are given by Eqs. (4-8) and (4-11).

Line Voltage and Current. The actual voltage and current existing on a transmission line are the sum of the voltages and currents, respectively, of the incident and reflected waves, as given by Eqs. (4-6), with the values for E_1 and E_2 defined as in Eqs. (4-16).[1] Although the equations of the transmission line appear complicated, the character of the voltage and current distributions that they lead to under different conditions can be readily understood with the aid of the typical examples considered in Sec. 4-4.

4-3. Transmission-line Constants. The electrical properties of a transmission line are determined by the inductance L, capacitance C,

[1] This result can also be written in an equivalent form in terms of hyperbolic functions:

$$E = E_L \cosh(\alpha + j\beta)l + I_L Z_0 \sinh(\alpha + j\beta)l \tag{4-17a}$$

$$I = I_L \cosh(\alpha + j\beta)l + \frac{E_L}{Z_0} \sinh(\alpha + j\beta)l \tag{4-17b}$$

series resistance R, and shunt conductance G, per unit length of line. The inductance and capacitance can be calculated by the usual formulas for transmission lines, except that at radio frequencies there are negligible magnetic-flux linkages inside the conductor as a result of skin effect; this means that one should omit the small term in the low-frequency inductance formulas that does not involve the dimensions. The series resistance of radio-frequency lines is controlled by skin effect, and so is proportional to the square root of the frequency. The shunt conductance is determined by the dielectric loss. With air insulation the shunt conductance is therefore negligible, but with solid dielectric such as used in twisted-pair and coaxial cables, the shunt conductance will be proportional to the product of frequency, power factor, and dielectric constant.

The electrical properties of the transmission line enter into the equations of the line through the characteristic impedance Z_0 and the propagation constant $\sqrt{ZY}$ as defined by Eqs. (4-5) and (4-7). At radio frequencies it is nearly always permissible to assume that $\omega L >> R$, and $\omega C >> G$. To the extent that this is true, one can rewrite Eqs. (4-5) and (4-7) as follows.

$$\sqrt{\frac{L}{C}} \tag{4-18a}$$

$$\alpha = \frac{R}{2Z_0} + \frac{GZ_0}{2} \tag{4-18b}$$

$$\beta = \omega\sqrt{LC} \tag{4-18c}$$

The characteristic impedance Z_0 is the ratio of voltage to current in an individual wave [see Eqs. (4-8) and (4-11)]; it is also the impedance of a line that is infinitely long or the impedance of a finite length of line when $Z_L = Z_0$. It will be noted that at radio frequencies the characteristic impedance is a resistance that is independent of frequency. Typical values for the characteristic impedance are of the order of 200 to 800 ohms for two-wire lines with air insulation, and 20 to 100 ohms for coaxial cables.

The attenuation constant of radio-frequency lines as given by Eq. (4-7) increases with frequency; this follows from Eq. (4-18b), and the fact that at high frequencies the series resistance and shunt conductance are proportional to the square root and the first power of frequency, respectively. With air insulation the conductance G is negligible, and the attenuation is due almost entirely to the skin-effect resistance of the conductors. However, in lines possessing solid dielectric, such as twisted-pair and many coaxial cables, the situation is more involved. Conductor resistance loss is then responsible for most of the attenuation at low frequencies, while the dielectric loss is the cause of most of the attenuation when the frequency is sufficiently high.

The phase constant β of a radio-frequency line is seen from Eq. (4-18c) to be proportional to frequency, and to the square root of the product LC of the line inductance and capacitance, but is independent of line resistance or conductance. The use of dielectric insulation, as is common in coaxial cables, increases the capacitance of the line, and thereby makes β larger in proportion to $\sqrt{k}$ where k is the dielectric constant of the insulation.

Wavelength and Phase Velocity. The distance λ that a wave must travel along the line in order for the total phase shift to be 2π radians is defined as the *wavelength* λ of the line. Thus, since $\beta\lambda = 2\pi$,

$$\lambda = \frac{2\pi}{\beta} \tag{4-19a}$$

In the case of radio-frequency lines with air dielectric, λ approximates the free-space wavelength of a radio wave of the same frequency. In the case of cables with solid dielectric having a dielectric constant k, the wavelength is very closely the free-space wavelength divided by $\sqrt{k}$.

A wavelength λ at a frequency f corresponds to a velocity $v_p = f\lambda$. This is termed the *phase velocity* of the line, i.e.,

$$\text{Phase velocity} = f\lambda = \frac{2\pi f}{\beta} \tag{4-19b}$$

In radio-frequency lines having air dielectric, the phase velocity approximates very closely the velocity of light. In lines with solid-dielectric insulation, the phase velocity is the velocity of light divided by the square root of the dielectric constant of the insulation.

4-4. Examples of Voltage and Current Distributions on Transmission Lines. The various ways in which the voltage and current may be distributed along a transmission line can be understood by considering in detail a number of special cases. In the discussion of these examples to follow, it is assumed that the attenuation constant α is small; this is done in order to simplify the phenomena involved. The modifications introduced when the attenuation constant is not small are discussed in Sec. 4-5.

Transmission Line with Open-circuited Load. When the load impedance is infinite, Eq. (4-14) shows that the coefficient of reflection will be $\underline{1/0}$. Under these conditions the incident and reflected waves will have equal magnitudes at the load, and the reflection will be such that the voltages of the incident and reflected waves have the same phase. As a result, the voltages of the two waves add arithmetically so that at the load $E_1 = E_2 = E_L/2$. Under these conditions it follows from Eqs. (4-8) and (4-11) that the currents of the two waves are equal in magnitude but opposite in phase; they thus add up to zero load current, as must be the case if the load is open-circuited.

Consider now how these two waves behave as the distance l from the load increases. The incident wave advances in phase β radians per unit length, while the reflected wave lags correspondingly; at the same time magnitudes do not change greatly when the attenuation constant α is small. The vector sum of the voltages of the two waves is then less than the arithmetic sum, as illustrated in Fig. 4-3*a*, for $l = \lambda/8$. This tendency continues until the distance to the load becomes exactly a quarter wavelength, i.e., until $\beta l = \pi/2$. The incident wave has then advanced 90° from its phase position at the load, while the reflected wave has dropped back a similar amount. The line voltage at this point is thus the arithmetic *difference* of the voltages of the two waves, as shown in Fig. 4-3*a*, for $l = \lambda/4$, and it will be quite small if the attenuation is small. The resultant voltage will not be zero, however, because some attenuation will always be present, and this causes the incident wave to be larger and the reflected wave smaller at the quarter-wave length point than at the load, where the amplitudes are exactly the same.

As the distance to the load increases to a value greater than a quarter wavelength, the phase of the incident wave continues to advance, while that of the reflected wave continues to lag. As a consequence, the voltages of the two waves depart increasingly from the condition of phase opposition existing at the quarter-wavelength point, and give a resultant value that becomes larger with increasing distance. This tendency continues until the distance from the load is a half wavelength (that is, $\beta l = \pi$); at this point the phases of the two waves have respectively advanced, and retarded, by 180°. The result is that the voltages now have the same relative phase relation with respect to each other as existed at the load, and so add arithmetically as at the load to give a large resultant line voltage. At greater distances than a half wavelength the cycle starts to repeat, as illustrated in Figs. 4-3*a* and 4-4*a*.

The voltage distribution on the open-circuited transmission line that results from this process is shown in Figs. 4-3*a* and 4-4*a*. It is characterized by voltage maxima at points that are even multiples of a quarter wavelength distant from the load, and by deep voltage minima at points that are odd multiples of a quarter wavelength from the load.

The current distribution associated with this voltage is also illustrated in Fig. 4-4*a*. The current distribution has minima where the voltage has maxima, and vice versa. This arises from the fact that the current of the reflected wave has the opposite phase from the reflected voltage [see Eq. (4-11)]. As a result, the currents in the two waves add where the voltages subtract, and subtract to give a minimum where the voltages add.

It will be noted that the variations in both the voltage and current distributions repeat their general character each half wavelength. This is characteristic of all distributions on transmission lines.

Transmission Line with Short-circuited Load. When the load end of the line is short-circuited, that is, $Z_L = 0$, reference to Eq. (4-14) shows that the reflection coefficient has the value $-1.0/\underline{0°} = 1.0/\underline{180°}$. As in

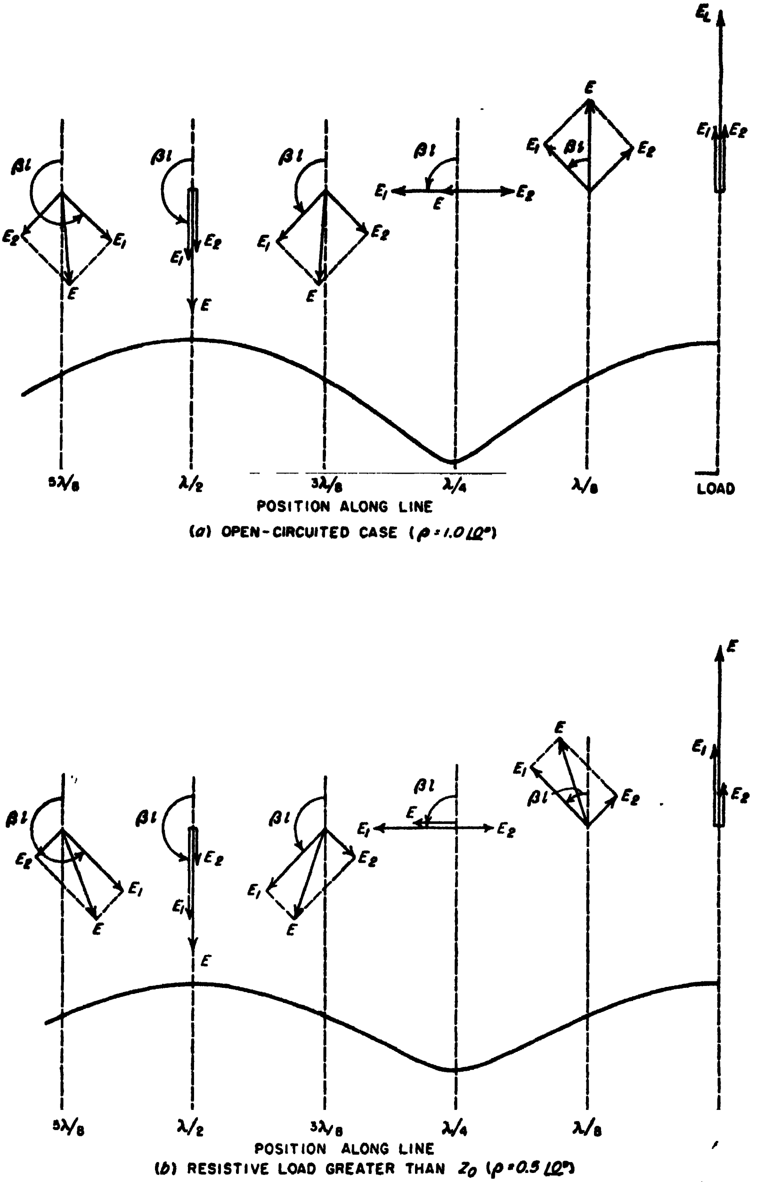


FIG. 4-3. Vector diagrams showing manner in which the incident and reflected waves combined to produce a voltage distribution on the transmission line. The cases shown correspond to a reflection in which the phase of the voltage is unchanged by reflection; it is also assumed that the attenuation of the line is quite small.

the open-circuited case, the reflected wave has an amplitude equal to the amplitude of the incident wave. However, the reflection now takes place with reversal in phase of the voltage, and without change in phase of the current. The result is that the current in each individual wave at the load is half of the load current, while the voltages in the two waves add up at the load to a resultant of zero voltage, as obviously is required across a short circuit.

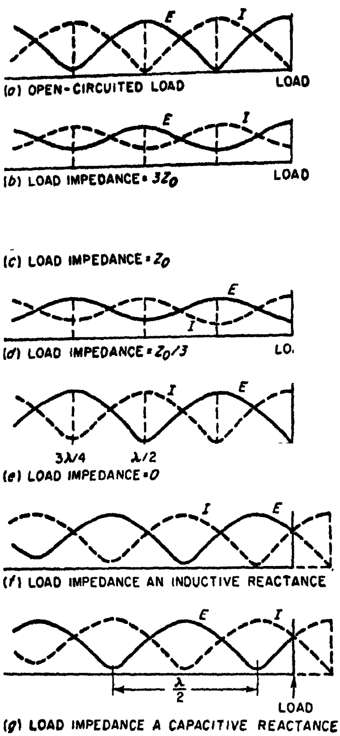


FIG. 4-4. Types of voltage and current distributions produced on a transmission line by different load impedances. It is assumed that the transmission line has low attenuation, and a characteristic impedance that is resistive.

If one now examines the situation as the distance from the load increases, the incident wave advances in phase while the reflected wave lags correspondingly, exactly as in the case of the open-circuited load. However, since it is now the currents that add at the load end of the line and the voltages that subtract, one obtains the distribution of voltage and current illustrated in Fig. 4-4e. This differs from the corresponding distributions of the open-circuited load case only in that voltage and current are interchanged. That is, with the short-circuited load the voltage on the line goes through minima at distances from the load that are even multiples of a quarter wavelength, and through maxima at distances that are odd multiples of a quarter wavelength. As before, the positions of the current maxima correspond to the voltage minima, and vice versa.

Characteristic Impedance Load. When the load impedance is equal to the characteristic impedance, the reflection coefficient is zero; i.e., there is no reflected wave. Under these conditions the voltage and current both increase exponentially with increasing distance from the load, as illustrated in Fig. 4-4c.

The physical significance of the situation where the reflection coefficient is zero (i.e., when $Z_L = Z_0$) is that the vector ratio of the voltage to current required by the load is exactly the same as that present in the incident wave. The load is therefore able to absorb completely the incident wave. With any other value of load impedance this is not possible, and a reflected wave is then produced.

Intermediate Values of Load Impedance. When the load impedance is a resistance greater than the characteristic impedance, the reflected wave produced at the load is smaller than the incident wave, but has the same phase angle as in the open-circuited case.[1] As a result, the voltage and current distributions go through successive maxima and minima at exactly the same places as for the open-circuited load. However, since the reflected wave is smaller than the incident wave, the minima are not as deep in proportion to the load voltage; this is illustrated in Fig. 4-4*b*. Vector diagrams showing how the voltages of the incident and reflected waves add to give the line voltage in this case are shown in Fig. 4-3*b*; a comparison with the corresponding diagrams of Fig. 4-3*a* shows in detail why and how the situation is modified when the reflected wave is smaller than the incident wave.

When the load impedance is a resistance that is smaller in magnitude than the characteristic impedance of the line, then the reflected wave is smaller than the incident wave, and has the same phase relation with respect to the incident wave as in the short-circuited load case. Under these conditions, the voltage and current distributions possess maxima and minima at exactly the same points as for the short-circuited load, but the maxima are not as large and the minima are less deep. This is illustrated in Fig. 4-4*d*.

Reactive Loads. Next consider the case where the load impedance is a pure reactance. Study of Eq. (4-14) shows that if the characteristic impedance can be assumed to be a resistance, the reflection coefficient for Z_L reactive is unity irrespective of the magnitude of the load reactance; however the phase angle of the reflection coefficient will depend upon the ratio of the load reactance to characteristic impedance. The consequences of this situation are illustrated in Fig. 4-4*f* and *g*. With a reactive load impedance, the voltage and current distributions vary in the same way, and to the same extent, as with the open-circuited (or short-circuited) load case. However, a reactive load impedance causes the minima of these curves to be displaced with respect to the position of the minima for an open-circuited line.

If one takes the open-circuit distribution as a reference, then a capacitive load causes the first minimum in the voltage distribution to occur closer to the receiver than a quarter wavelength, as illustrated in Fig. 4-4*g*. This comes about because for capacitive loads the phase angle of the

[1] For the reflection coefficient to have a phase angle of exactly 0 or 180°, it is necessary that the load impedance have the same phase angle as the characteristic impedance. In the case of radio-frequency transmission lines, this means a load that for all practical purposes is resistive.

reflection coefficient is negative, i.e., the reflected wave at the load lags behind the incident wave. Thus with a capacitive load the distance from the load at which the reflected wave lags 180° behind the incident wave is less than a quarter wavelength. In contrast, an inductive load causes the first voltage minimum to occur at a distance from the load that is greater than a quarter wavelength, as illustrated in Fig. 4-4f. This results from the fact that the phase angle of the reflection coefficient is positive in this case. With both inductive and capacitive loads, the displacement of the minima from their open-circuited position is greater the lower the load reactance. It is also to be noted that the effect of a reactive load is merely to displace the position of the minima; the distance between the adjacent minima still remains a half wavelength, just as in the open- and short-circuited cases.

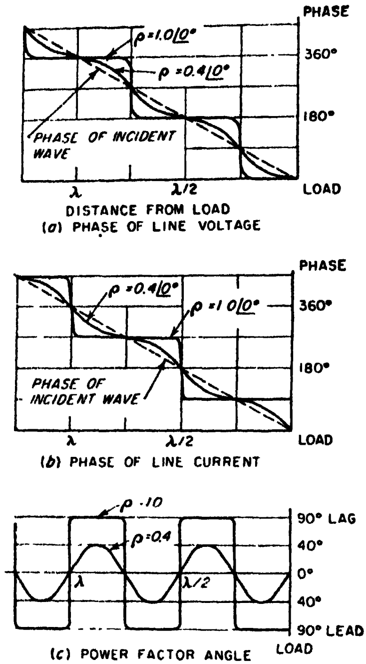


FIG. 4-5. Phase relations on a transmission line for two typical conditions. In these curves, the voltage of the incident wave at the load is used as the reference phase, and the line attenuation is assumed to be small.

Load impedances that have both resistive and reactive components will result in voltage and current distributions in which the variation in amplitude along the line is less than in the open- and short-circuited cases because the reflection coefficient is less than unity, as in Fig. 4-4b and d. However, at the same time the maxima and minima are shifted along the line in the same direction as when the load is purely reactive.

Phase Relations in Voltage and Current Distributions. The phase of the voltage and current in an individual wave drops back β radians per unit length in the direction in which the wave travels. Thus, when the load impedance equals the characteristic impedance so that only the incident wave is present, the line voltage and current advance in phase at the uniform rate of β radians per unit length as one goes from the load to the generator. The total phase shift is 2π radians per wavelength under these conditions.

When the load impedance does not equal the characteristic impedance, the phase relations are complicated by the presence of the reflected wave. The phase of the resulting line voltage (or current) then oscillates about the phase of the voltage (or current) of the incident wave, as illustrated in Fig. 4-5. The phase shift under these conditions tends to be concentrated in regions where the voltage (or current) goes through a minimum; this is increasingly the case as the reflected wave approaches equality with the incident wave. However, irrespective of the relative amplitudes of the incident and reflected waves, the phase of both voltage and current will advance exactly π radians (180°) when the distance toward the generator decreases by a half wavelength. Although in the absence of a reflected wave the variation in phase is at a uniform rate within this distance, this is not the case when a reflected wave is present.

4-5. The Effect of Attenuation on Voltage and Current Distribution—Lossless Lines. The voltage and current distributions illustrated in

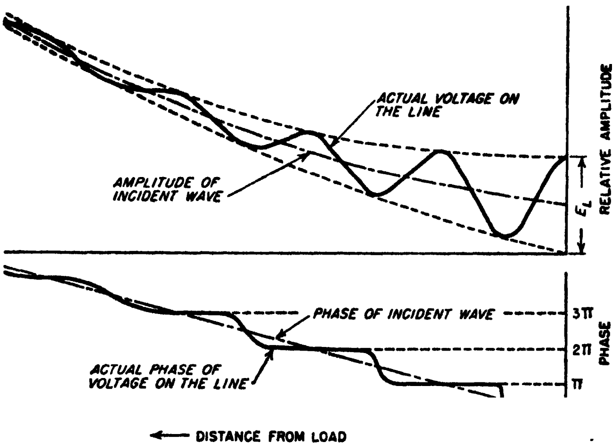


FIG. 4-6. Variation of voltage amplitude and phase with distance along a transmission line having such high attenuation that toward the generator end the reflected wave is attenuated at a very small size.

Fig. 4-4 assume that the total attenuation αl of the line is small compared with unity. Under these conditions the amplitude of the incident, and also of the reflected wave, changes only slightly in traveling the entire length of the line.

When the attenuation of the line is relatively large the incident wave then increases rapidly in amplitude as one goes toward the generator. Similarly the reflected wave decreases rapidly in size as it recedes from the load. The resulting behavior is as illustrated in Fig. 4-6; at a considerable distance from the load the reflected wave becomes so small that the voltage and current begin to approximate the values that would exist for the case $Z_L = Z_0$, irrespective of the actual value of the load impedance. The progressive change in the ratio of reflected to incident waves that is caused by attenuation produces corresponding effects on the phase behavior. These are also illustrated in Fig. 4-6, which shows that the actual phase departs less and less from the phase of the incident wave as the reflected wave becomes smaller.

Transmission Lines with Zero Losses. The behavior of an idealized transmission line with zero losses is important because under many circumstances, and for many purposes, it is permissible to neglect the losses associated with practical radio-frequency transmission lines.

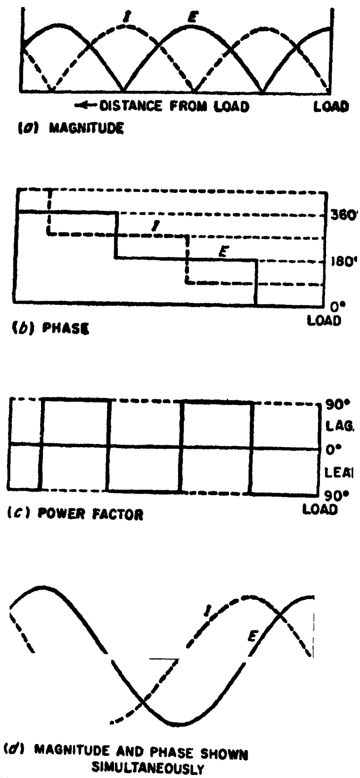


FIG. 4-7. Voltage, current, and phase relations on an open-circuited transmission line having zero losses.

When the resistance and conductance of a transmission line are zero, the attenuation constant α is likewise zero, and the incident and reflected waves on the transmission line suffer no change in amplitude as they travel from one end of the line to the other. The voltage and current distributions that result are then similar to those of Fig. 4-4, except that all the maxima (and minima) are of the same height.

When the reflection coefficient of the load is unity, corresponding to an open- or short-circuited or reactive load, the curves giving the distribution of voltage and current on the loss-free line are sections of half sine waves that go to zero at the minima, as shown in Fig. 4-7*a*. In this case the phase of the voltage (or current) jumps 180° at each minimum, as indicated in Fig. 4-7*b*. The distribution curves of the lossless line are hence commonly drawn as shown in Fig. 4-7*d*, which simultaneously indicates both magnitude and phase by using negative amplitudes to indicate the polarity reversal associated with a 180° phase shift.

Since the waves on a lossless line do not change in amplitude as they travel along the line, the reflection coefficient in such a system is everywhere constant and equal to the reflection coefficient at the load, as given by Eq. (4-14). Similarly, the standing-wave ratio (see below) is everywhere the same on a lossless line.

4-6. Standing-wave Ratio. The character of the voltage (or current) distribution on a transmission line can be conveniently described in terms of the ratio of the maximum amplitude to minimum amplitude possessed by the distribution. This quantity is termed the ***standing-wave ratio*** (often abbreviated SWR); thus in Fig. 4-8,[1]

$$\text{Standing-wave ratio} = S = \frac{E_{max}}{E_{min}} \tag{4-20}$$

Alternatively, the standing-wave ratio may be defined in terms of maximum and minimum current; for any particular line the standing-wave ratio at a given region on the line will be the same whether defined in terms of the voltage or current distribution.

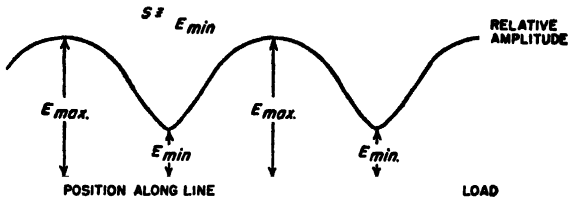


FIG. 4-8. Diagram illustrating nomenclature used in defining the standing-wave ratio.

In terms of the amplitudes $|E_1|$ and $|E_2|$ of the incident and reflected waves respectively, the standing-wave ratio can be written

$$S \quad \frac{|E_1| + |E_2|}{|E_1| - |E_2|} \tag{4-21}$$

The standing-wave ratio is seen from Eq. (4-21) to be a measure of the amplitude ratio of the reflected to incident waves. Thus a standing-wave ratio of unity denotes the absence of a reflected wave, while a very high standing-wave ratio indicates that the reflected wave is almost as large as the incident wave. Theoretically, for the case of zero attenuation, the standing-wave ratio will be infinite when the load is either open- or short-circuited, or is a lossless reactance.

The standing-wave ratio S is one means of expressing the magnitude of the reflection coefficient; the exact relation between the two is

$$S \tag{4-22a}$$

or

$$|\rho| \quad \frac{S-1}{S+1} \tag{4-22b}$$

This relationship is illustrated graphically in Fig. 4-9.

The importance of the standing-wave ratio arises from the fact that it can be very easily measured experimentally. Moreover, the standing-

[1] This definition of standing-wave ratio is sometimes called voltage standing-wave ratio (VSWR) to distinguish it from the standing-wave ratio expressed as a power ratio, which is $(E_{max}/E_{min})^2$.

wave ratio indicates directly the extent to which reflected waves exist on a system. In addition, standing-wave measurements provide an important means of measuring impedance, as discussed in Sec. 4-9.

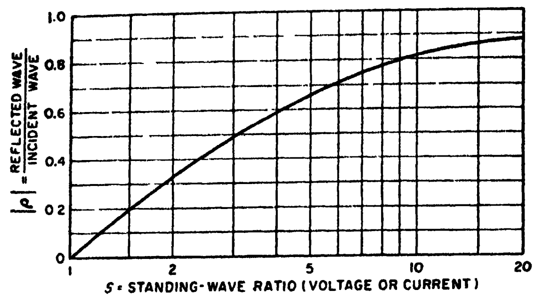


FIG. 4-9. The relationship between standing-wave ratio and magnitude $|\rho|$ of reflection coefficient.

4-7. Impedance and Power-factor Relations in Transmission Lines. The expression "transmission-line impedance" applied to a point on a transmission line signifies the vector ratio of line voltage to line current at that particular point. This is the impedance that would be obtained if the transmission line were cut at the point in question, and the impedance looking toward the load were measured on a bridge.

When the load impedance equals the characteristic impedance, only the incident wave is present, and the line impedance is everywhere equal to the characteristic impedance. The line impedance is also equal to the characteristic impedance under conditions where the total attenuation αl to the load is so great that the reflected wave is of negligible amplitude compared with the incident wave. Under these conditions the impedance of the transmission line is independent of conditions at the load.

When a reflected wave is present, the impedance will be alternately greater and lower than the characteristic impedance, as illustrated in Fig. 4-10. Since the line current is always a minimum when the voltage is maximum, and vice versa, the impedance maxima and minima coincide with the voltage maxima and minima, respectively. The magnitude of the line impedance therefore varies cyclically with a periodicity of a half wavelength. If the line losses are low and the reflection coefficient of the load is not too close to unity, the line impedance repeats almost exactly in successive half-wave intervals, as illustrated in Fig. 4-10*a*. However, when the reflection coefficient at the load approaches unity (large standing-wave ratio), then the line attenuation, even if small, will cause the peaks of impedance to diminish in amplitude at progressively larger distances to the load, as in Fig. 4-10*b*.

The power factor of the line impedance varies according to the standing-wave situation. When the load impedance equals the characteristic impedance, there is no reflected wave and the power-factor angle of the line is zero, corresponding to a resistive impedance. However, when a reflected wave is present, the power-factor angle is zero only at the points on the line where the voltage goes through a maximum or a minimum. At other points the power-factor angle will alternate between leading and lagging at intervals of a quarter wavelength, as shown in Figs. 4-10 and 4-5*c*. When the line is short-circuited at the receiver (Fig. 4-10*b*), or if

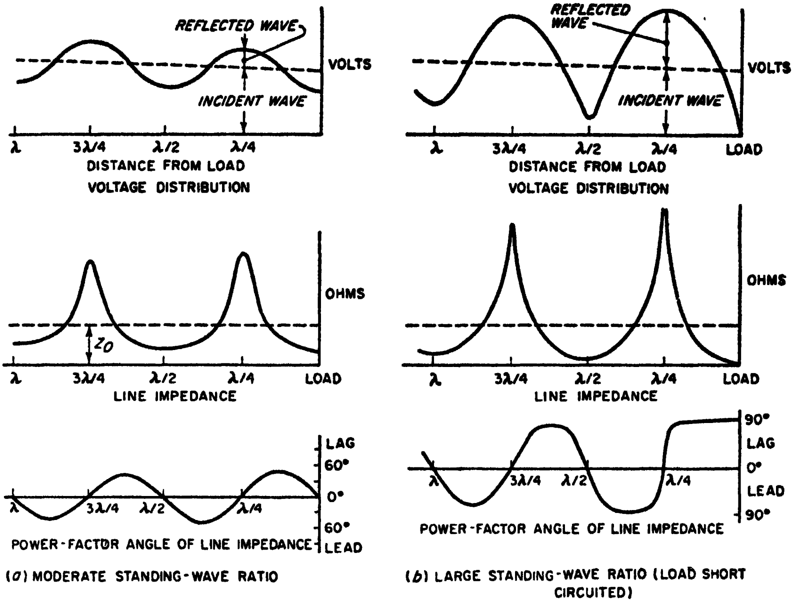


FIG. 4-10. Magnitude and power factor of line impedance with increasing distance from the load, for load impedances that are, respectively, a resistance less than the characteristic impedance, and a short circuit. These diagrams assume that the attenuation of the line is small.

the load is a resistance less than the characteristic impedance so that the voltage distribution is of the short-circuit type (Fig. 4-10*a*), the power factor is inductive (lagging) for lengths corresponding to less than the distance to the first voltage maximum, and thereafter alternates between capacitive and inductive at intervals of a quarter wavelength. Similarly, with an open-circuited receiver, or with a resistance load greater than the characteristic impedance so that the voltage distribution is of the open-circuit type (Fig. 4-5), the power factor is capacitive for lengths less than the distance to the first voltage minimum. Thereafter, the power factor alternates between capacitive and inductive at intervals of a quarter wavelength, exactly as in the short-circuited case.

If one considers the impedance at the generator end of a transmission line of fixed length under conditions where the frequency of measurement is progressively increased, the impedance will vary in magnitude with frequency in much the same manner as with increasing length. Thus, with a short-circuited load, the line impedance will go through successive maxima at frequencies that make the line length correspond to one-quarter, three-quarters, five-quarters, etc., of a wavelength, and will go through minima at frequencies that correspond to line lengths measured in wavelengths that are an even number of quarter wavelengths. This is illustrated in Fig. 4-11.

The extent to which the power factor of the line impedance varies with changes in length, or changes in frequency, depends upon the standing-wave ratio at the point on the line where the power factor is observed.

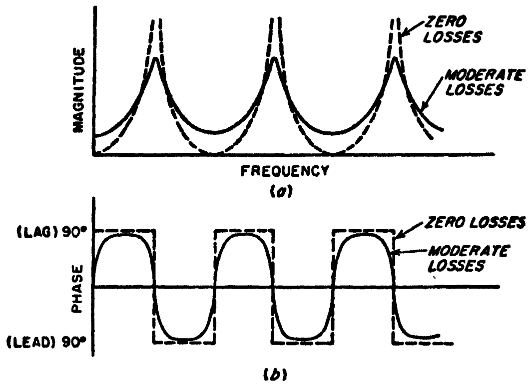


FIG. 4-11. Effect of variation in frequency on the magnitude and phase of the sending-end impedance of a short-circuited transmission line of fixed length.

If the standing-wave ratio is large, and the line losses low (solid curve in Fig. 4-11), the power-factor angle will approach 90° except in the immediate vicinity of the voltage maxima and minima. Then the power-factor angle suddenly shifts between nearly +90° and nearly −90°, as illustrated in Fig. 4-11*b* and Fig. 4-5*c*. In fact, in the case of a short-circuited or open-circuited ideal line of zero losses, the power-factor angle is exactly 90° everywhere except at the voltage maxima and minima, as illustrated in Fig. 4-7*c*, and by the dotted line in Fig. 4-11*a*. On the other hand, if the standing-wave ratio is small or moderate, the maximum range over which the power-factor angle varies about unity power factor will be correspondingly less than 90° (see Figs. 4-5*c* and 4-10*a*).

4-8. Transmission-line Charts—the Smith Chart. The various properties of a transmission line can be presented graphically in an almost endless variety of charts. The most useful graphical representations, however, are those which give the impedance relations that exist along a lossless line for different load conditions.

The Smith chart shown in Fig. 4-12 is the most widely used transmission-line chart of this class.[1] This diagram is based on two sets of orthogonal circles. One set represents the ratio R/Z_0, where R is the resistance component of the line impedance, and Z_0 is the characteristic impedance (which for a lossless line is a resistance). The second set of circles represents the ratio jX/Z_0, where X is the reactive component of the line impedance. These coordinates are so chosen by means of a conformal transformation that conditions on the lossless line corresponding to a given standing-wave ratio (or what is the same thing, a given magnitude of the load reflection coefficient) lie on a circle having its origin at the center of the chart.

The standing-wave ratio S corresponding to any particular circle is equal to the value of R/Z_0 at which the circle crosses the horizontal axis on the right-hand side of the chart center (see Prob. 4-25). This same circle intersects the horizontal axis to the left of the center at a value of R/Z_0 such that $1/S = R/Z_0$. Intersections with the horizontal axis that are on the left of the chart center represent voltage minima; intersections with the horizontal axis on the right of the center correspond to voltage maxima.

Moving around a given standing-wave circle is equivalent to traveling along a lossless transmission line on which the standing-wave ratio corresponds to the circle involved; thus the successive values of impedance indicated by a given circle correspond to the line impedances at successive points along the lossless line. Distance on the actual transmission line is directly proportional to the angle of rotation around the standing-wave circle, with one complete revolution corresponding to exactly a half wavelength on the transmission line. Thus in Fig. 4-12 the distance between points on the line where the impedance conditions are represented by P and Q on the chart is 0.05 wavelength, because P and Q lie on the same circle, and radial lines OPA and OQB drawn from the center of the chart are displaced by 0.05λ on the outer scale; this corresponds to 36° angular displacement, or $36/720 = 0.05$ wavelength.[2] Travel around the circle in a clockwise direction is toward the generator, whereas travel in a counterclockwise direction is toward the load; this fact is marked on the periphery of the chart.

The impedance at any point on a transmission line for a given load

[1] P. H. Smith, Transmission Line Calculator, *Electronics*, vol. 12, p. 29, January, 1939; P. H. Smith, An Improved Transmission Line Calculator, *Electronics*, vol. 17, p. 130, January, 1944. Graph paper and a plastic calculator are commercially available. A paper covering the theoretical foundations of the Smith chart, and its relation to the so-called rectangular chart, is H. L. Krauss, Transmission Line Charts, *Elec. Eng.*, vol. 68, p. 767, September, 1949.

[2] Distances greater than a half wavelength are handled by going around the standing-wave circle as many times as required. Thus the distance OA to OB actually represents $0.05\lambda + n\lambda/2$, where n can be any interger, including zero.

condition, including the load impedance, is represented by a point properly located on the Smith chart. Thus P in Fig. 4-12 corresponds to the impedance $Z_0(0.98 + j0.7)$, and lies on the circle centered at O that corresponds to a standing-wave ratio of 2 (because the circle through P intersects the R/Z_0 axis on the right of the chart at $R/Z_0 = 2$). If the line were terminated with a load having an impedance corresponding to

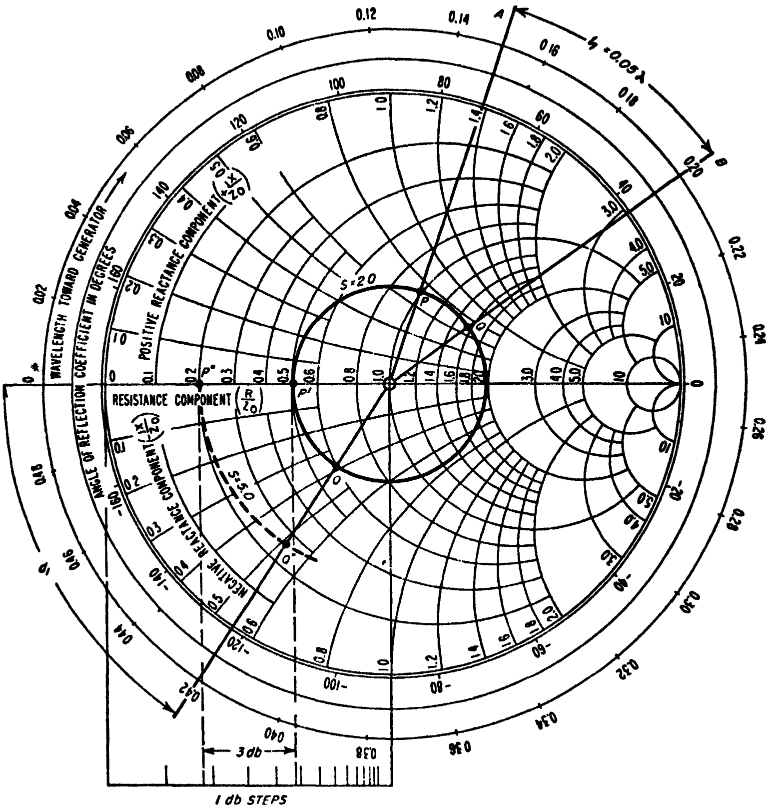


FIG. 4-12. The Smith chart.

P, then the standing-wave ratio that would exist on the line would be 2.0; the impedance at other points along the line could be obtained by traveling clockwise around the circle passing through P by an amount indicated by the calibration on the periphery of the chart. For example, at a distance 0.05λ toward the generator from P, the line impedance is $Z_0(1.56 + j0.7)$, corresponding to point Q, while 0.27λ distant from P, the impedance corresponds to Q' and is $Z_0(0.6 - j0.38)$. Again, if the load impedance corresponded to the value Q, the standing-wave ratio would still be 2, but the impedance at Q' would now be the line impedance at a distance 0.22λ from the receiver, since Q' is 0.22λ around the circle in a clockwise direction from Q.

The Smith chart thus shows very simply and directly the standing-wave ratio corresponding to a given load impedance. It also shows the line impedance at any desired point, given the standing-wave ratio and the impedance at any other point on the line, for example, the load impedance. From the standing-wave ratio, one can obtain the magnitude of

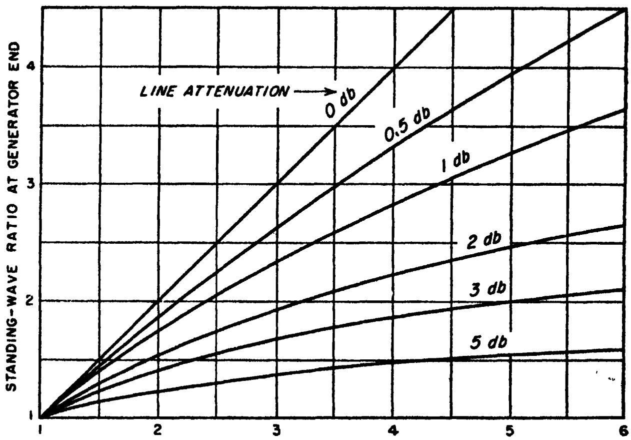


FIG. 4-13. Relationship between standing-wave ratios at two points on a transmission line, for different values of line attenuation between these points.

the reflection coefficient from Eq. (4-22b) or Fig. 4-9. The phase angle of the reflection coefficient is given on the chart periphery. Thus for point P one has $\rho = 0.33\underline{/72°}$. The Smith chart can also be used to determine impedance from data obtained from standing-wave measurements; this is discussed in Sec. 4-9.

Effect of Line Attenuation. The Smith chart assumes that the line attenuation is zero. Under these conditions the standing-wave ratio is everywhere constant, and the chart implies that this is the case. When attenuation is present it is, however, still possible to use the Smith chart by using Fig. 4-13 to correct for the change in standing-wave ratio with position.[1] The method of doing this is made clear by the following example.

Example. Assume that the conditions existing at some point on the line correspond to P in Fig. 4-12; this may be the generator end of the line although it is not so limited. It is then desired to know the line impedance at a point 0.23λ closer to the load when

[1] The curves in Fig. 4-13 are obtained by combining Eqs. (4-15), (4-22a), (4-22b), and (4-9a).

the total attenuation for the line length of 0.23λ is known to be 3.0 db rather than zero. The first step is to ignore the line attenuation and travel *counterclockwise* around the circle passing through P for a distance corresponding to 0.23λ. This brings one to point Q', which corresponds to the line impedance that would exist at the desired point if the line had no attenuation. However, Fig. 4-13 shows that a line attenuation of 3 db causes a standing-wave ratio of 2.0 at the generator end of the section of line to correspond to a standing-wave ratio of 5.0 at the load end. A circle, shown dotted in Fig. 4-12, is then drawn corresponding to this standing-wave ratio. The intersection of this circle with the radial line OQ' then defines a point Q'' on the chart that corresponds to the desired impedance, taking into account the line attenuation;[1] this impedance is $Z_0\,(0.26 - j0.52)$.

4-9. Impedance Measurements Using Standing-wave Ratios.[2] The impedance at very high frequencies is commonly determined with the aid of standing waves. This is done by using the unknown impedance as the load impedance of a line having low losses. The resulting standing-wave ratio is then observed experimentally and, in addition, the distance from the receiver to the first voltage minimum is observed. From this information one can, with the aid of a Smith chart, readily determine the unknown impedance.

Example 1. Suppose that a standing-wave ratio of 2.0 is observed and that the first voltage minimum is 0.08λ from the load. One would then enter the Smith chart at the point P', which corresponds to a voltage minimum for a standing-wave ratio that is 2.0, and would then travel along this circle of constant standing-wave ratio toward the load a distance 0.08λ thus arriving at point Q'. The coordinates of this point are $0.6 - j0.38$, and multiplying these numbers by the value of Z_0 for the transmission line gives the impedance of the terminating load, which is the impedance to be determined.

Example 2. Assume that once again the standing-wave ratio is observed to be 2.0, but that it is now inconvenient to measure the actual distance from the load to the first voltage minimum. The procedure then consists in first connecting the unknown impedance across the end of the line and observing the position of some convenient voltage minimum. Next, the unknown impedance is replaced by a short circuit, and the position of the first voltage minimum *on the load side* of the original minimum is observed. Assume that this minimum is 0.35λ toward the load from the original minimum. It is then permissible to regard this new minimum as the equivalent position of the load. This follows from the fact that on a lossless line impedances repeat exactly each half wavelength. Therefore one enters the Smith chart at point P', which corresponds to the voltage minimum with the load connected, and travels 0.35λ *toward the load* along the circle for $S \mp 2.0$. This leads to point P, which has the coordinates $0.98 + j0.7$; these numbers multiplied by Z_0 then give the unknown

[1] Smith charts are sometimes provided with an auxiliary decibel scale that can be used to determine the effect of the attenuation on the radius of the standing-wave circle. Such a scale is shown in Fig. 4-12, and is calibrated so that each unit on the auxiliary scale represents the change in circle radius associated with 1 db attenuation. Thus starting with a standing-wave circle of radius OQ' in Fig. 4-12, the circle passing through Q'' is drawn with a radius that is 3.0 units different on the decibel scale than OQ' as shown, because the line attenuation is 3.0 db.

[2] An extensive summarizing discussion of this subject is given by F. E. Terman and J. M. Pettit, "Electronic Measurements," pp. 135–152, McGraw-Hill Book Company, Inc., New York, 1952.

impedance. An alternative procedure would be to note that if the reference point is taken as the minimum with the unknown connected at the load, then when the line is short circuited, the first minimum on the generator side of this reference point is 0.5 − 0.35 = 0.15λ *toward the generator.* Entering the chart at P' as before, one could therefore proceed 0.15λ toward the generator (i.e., a distance −0.15λ toward the load). This also brings one to point P.

Equipment for Experimental Determination of Standing-wave Ratio for Impedance Measurements. The standing-wave ratio on a transmission

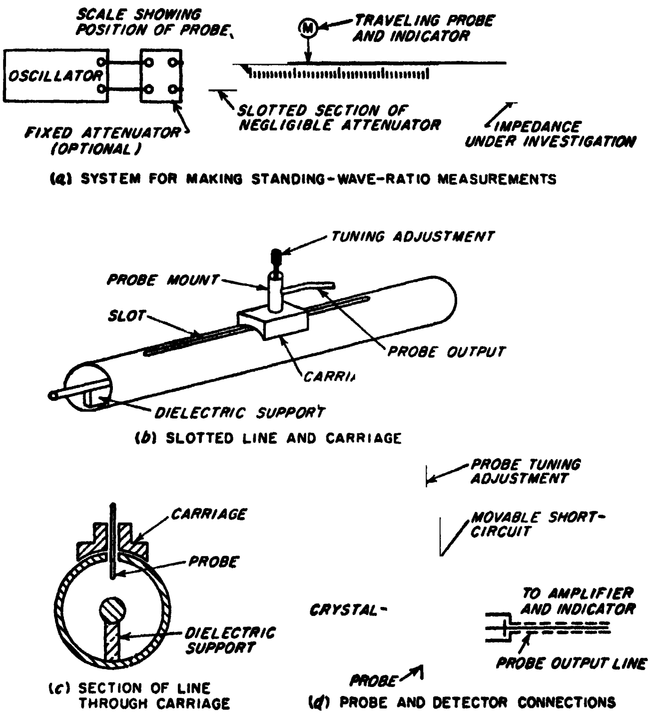


FIG. 4-14. Details of a slotted-line type of standing-wave detector for a coaxial line.

line can be observed by exploring along the length of the line with a pickup arrangement that will indicate the strength of either the electric field (line voltage) or the magnetic field (line current), in the vicinity of the line. A typical example of such a *standing-wave detector* that is suitable for coaxial systems is illustrated in Fig. 4-14. This arrangement consists of a section of coaxial line having air insulation and a longitudinal slot in the outer conductor, as shown. Mounted on this slotted section is a traveling carriage carrying a probe that projects through the slot toward the center conductor, as shown. To this probe there is connected some form of power- or voltage-indicating device, often a simple detector. An oscillator is connected to one end of the slotted line, while the other end is connected to the unknown impedance or, alternatively, to the input of the transmission line that is to have its standing-wave ratio observed. The standing-wave pattern is then obtained by moving the carriage (and hence the probe position) and observing the resulting variations in the probe output.

4-10. Transmission Lines as Resonant Circuits and as Circuit Elements.[1] A transmission line can be used to perform the functions of a resonant circuit. Thus, if the line is short-circuited at the load, then at frequencies in the vicinity of a frequency for which the line length is an odd number of quarter wavelengths long, the impedance will be high and will vary with frequency in the vicinity of resonance (i.e., frequency corresponding to quarter wavelength) in exactly the same manner as does the impedance of an ordinary parallel resonant circuit. It is therefore possible to describe resonance on a transmission line in terms of the impedance at resonance and the equivalent Q of the resonance curve.[2]

At very high frequencies, the parallel impedance at resonance and the obtainable circuit Q are far higher than can be realized with lumped circuits. In high-frequency lines having air insulation the losses all arise from skin effect in the conductors, and one has with copper conductors

For concentric lines:[3]

$$Q = 0.0839 \sqrt{f}\, bH \tag{4-23a}$$

$$Z_s = 11.11 \sqrt{f}\, bF \tag{4-23b}$$

For two-wire lines (neglecting radiation losses):

$$Q = 0.0887 \sqrt{f}\, bJ \tag{4-24a}$$

$$Z_s = \frac{23.95 \sqrt{f}\, bG}{n} \tag{4-24b}$$

[1] For further information, including particularly a derivation of the basic relations, see F. E. Terman, Resonant Lines in Radio Circuits, *Elec. Eng.*, vol. 53, p. 1046, July, 1934. In this paper it was demonstrated for the first time that the resonance curve of a transmission line has the same shape as the resonance curve of a circuit with coil and capacitor, and so can be described by specifying a Q.

[2] The Q in such a situation can be defined in terms of the detuning required to reduce the response to 70.7 per cent of the response at resonance, in accordance with Rule 1 on p. 49; alternatively, one may employ Eq. (3-1).

[3] Examination of Fig. 4-15 shows that in an air-insulated coaxial line of given outer radius b, Q will be maximum when the inner conductor has a size such that $b/a = 3.6$, corresponding to $Z_0 = 77$ ohms. These are also the proportions for minimum power loss in a transmission line operated with $Z_L = Z_0$. However, the maximum power that can be transmitted without exceeding a given voltage gradient occurs when $b/a = 1.65$, giving $Z_0 = 30$ ohms.

where Q = circuit Q defined from the resonance curve so that $Q = f_0/2\Delta f$, where f_0 is the resonant frequency and Δf is the number of cycles off resonance at which the response is 70.7 per cent of the response at resonance

Z_s = sending end or input impedance

f = frequency, cycles

b = inner radius of outer conductor of a concentric line, or spacing of wire centers in two-wire line, cm

a = outer radius of inner conductor in concentric line, or wire radius in two-wire line, cm

n = number of quarter wavelengths in the line

F, G, H, J = constants determined by b/a and given in Fig. 4-15

Substitution of reasonable values in these equations leads to surprising results. Thus, at a wavelength of 150 cm (200 Mc), a concentric line with copper conductors and air insulation in which $b/a = 3.6$ and with a diameter of outer conductor of 5 cm (2 in.) possesses a Q of approximately 3000; when the line length is a quarter wavelength long (approximately 15 in.), the resonant impedance is over 250,000 ohms. Because of favorable properties such as these, together with the fact that the physical size of a resonant line is relatively large in proportion to wavelength as compared with a coil-and-capacitor combination, resonant transmission lines find extensive use as resonant circuits at the higher radio frequencies, particularly at frequencies of the order of 100 Mc and greater.

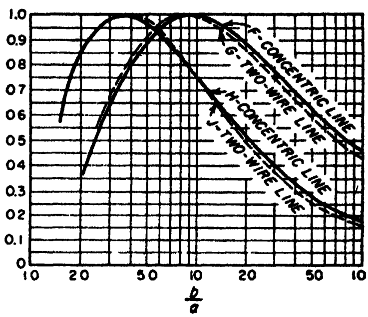


FIG. 4-15. Factors F, G, H, and J for use in Eqs. (4-23) and (4-24).

A behavior corresponding to that of a series resonant circuit can be obtained from a transmission line that is an odd number of quarter wavelengths long and open-circuited at the receiver. Under these conditions, the voltage at the load is much higher than the applied voltage, as is apparent from Fig. 4-4. Furthermore, at frequencies near resonance the voltage step-up varies with frequency in exactly the same manner as does a resonance curve, and has an equivalent Q given by Eq. (4-23a) or (4-24a) as the case may be. The voltage step-up ratio is, however, $Q \times 4/\pi n$, instead of Q as in the case of the ordinary series resonant circuit.

Transmission lines can be used to provide low-loss inductances or capacitances by employing the proper combination of length, frequency, and termination. Thus a line short-circuited at the load will offer an inductive reactance when less than a quarter wavelength long, and a capacitive reactance when between a quarter and a half wavelength long. With an open-circuited load, conditions for inductive and capacitive reactances are interchanged.

4-11. Impedance Matching in Transmission Lines.[1] Energy is transmitted most efficiently by a transmission line when no reflected wave is present.[2] However, only under exceptional cases will the load impedance be a resistance that is exactly equal to the characteristic impedance of the line. Thus, to obtain transmission of energy with maximum efficiency, it is necessary to provide means for matching the actual load impedance to the characteristic impedance of the line. Again, it is often desired that the line impedance be independent of the distance to the load. Likewise in making many types of measurements in systems involving transmission lines, it is frequently desirable, and in some cases very necessary, that there be no reflected wave present.

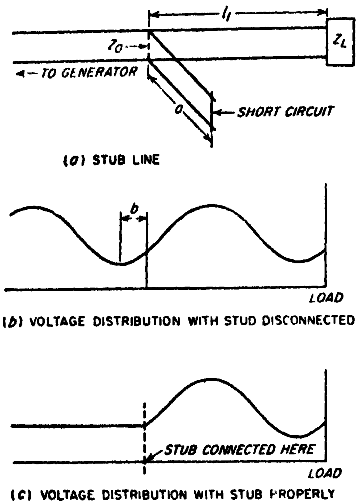


FIG. 4-16. Impedance matching by means of a short-circuited stub line. Although the arrangement shown is a two-wire system, coaxial lines may be employed.

At the lower radio frequencies a load can be matched to the characteristic impedance of a line by associating with the load a network of reactances that tunes out the load reactance and simultaneously transforms the resulting resistance to a value equal to the characteristic impedance of the line. This is discussed further in Sec. 4-12.

At very high and microwave frequencies, impedance matching is normally achieved with the aid of transmission-line techniques. The *stub* line arrangement of Fig. 4-16 is a common example. Here a short

[1] For additional information of a design character see T. E. Moreno, "Microwave Transmission Data," pp. 103–110, McGraw-Hill Book Company, Inc.. New York, 1948.

[2] When the characteristic impedance is a resistance, as is always the case at high frequencies, one can consider that the incident wave delivers energy to the load and that the reflected wave carries energy from the load back toward the generator. If the load impedance does not equal the characteristic impedance, i.e., if the load is not matched to the line, then some of the incident energy is reflected by the load and travels a round trip over the line, dissipating power in the line without delivering energy to the load. Thus the ratio of energy lost in the line to power dissipated in the load is increased by reflection.

section of short-circuited transmission line is connected in shunt with the transmission line. The distance l_1 from the load, and the length a of the stub, are so chosen that the reflected wave produced by the shunting impedance of the shunt line is equal in magnitude and opposite in phase to the reflected wave existing on the line at this point as a result of the reflection from the load impedance Z_L. Thus, although a reflected wave is present in the length l_1 because of reflection from Z_L, there is no reflected wave on the generator side of the stub line as a result of the cancellation of the two reflected waves.[1]

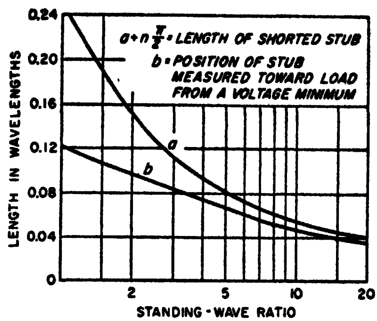


FIG. 4-17. Design information giving the length and position required for a short-circuited stub in order to obtain impedance matching. If desired, the stub line may be made any convenient multiple of a half wavelength greater than a.

The practical design of a stub-line system of this type can be readily carried out with the aid of Fig. 4-17, which gives the length a of the stub[2] and its position b with respect to a voltage minimum of the standing-wave pattern existing in the absence of the stub. A stub line used in this way will enable any load impedance to be matched to the characteristic impedance of a transmission line provided only that the load is not an open-circuit, short-circuit, or pure reactance.

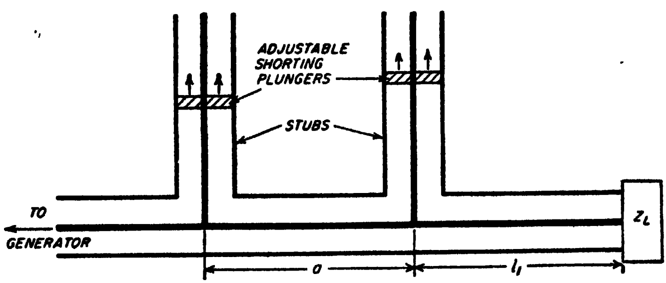


FIG. 4-18. Double-stub impedance-matching system.

Another arrangement often used to match a load to a transmission line is the two-stub system of Fig. 4-18. Here two spaced stubs whose lengths are individually controllable are shunted across the line near the load as

[1] Another way of expressing this situation is to say that the stub position and length are so selected that the input impedance of line l_1 shunted by the input impedance of stub line a will equal the characteristic impedance Z_0.

[2] The stub length can actually be made any convenient number of half wavelengths plus the value a given by Fig. 4-17.

shown. This arrangement has the advantage that trial-and-error adjustment of the impedance-matching system can be made without the necessity of providing a connection that can be slid along the transmission line. The arrangement is thus particularly suitable for coaxial transmission lines, as it avoids the mechanical problems involved in moving the position of a shunting stub along a coaxial line. The disadvantage of the two-stub system is that the range of load impedances that can be matched to the transmission line is limited. Thus, in the typical case where the spacing between stubs is made an eighth wavelength, an impedance

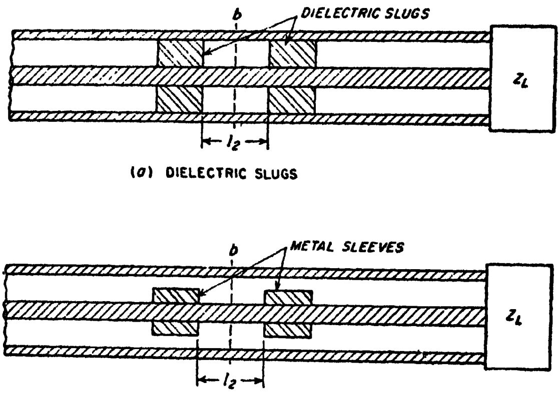


FIG. 4-19. Double-slug system for impedance matching in a concentric-line system.

match can be obtained only if the conductance component of the impedance at the stub nearest the load and looking toward the load is less than $2/Z_0$.*

The impedance-matching systems of Fig. 4-19 are termed two-slug tuners, and make use of two spaced elements that can be either dielectrics, as at *a*, or metal sleeves that reduce the clearance between inner and outer conductors, as at *b*. These arrangements operate by introducing a reflected wave that is adjusted to produce a reflection equal in magnitude and opposite in phase to the reflected wave produced by the load impedance. The phase of the reflection introduced in this way is controlled by moving the slugs along the line while maintaining the spacing l_2 between them constant. The magnitude of the reflected wave can be controlled with little effect on the phase by moving the two slugs equal amounts in opposite directions (i.e., by changing l_2 while keeping the slugs symmetrical with respect to reference line *b*). Like the two-stub arrange-

* When this requirement is not satisfied, an impedance match can still be obtained by increasing the distance l_1 from the double-stub tuner to the load by a quarter wavelength. This is because of the impedance-transforming action of a quarter-wave line, as discussed in connection with Eq. (4-31).

ment, slug tuners are limited in the range of load impedances that can be matched to a given line.

When the load impedance is resistive, or when it can easily be made resistive by tuning, the impedance-matching problem is considerably simplified. It is then merely necessary to transform the resistance actually present to a resistance that is equal to the characteristic impedance of the line. Transmission-line techniques that can be used to achieve this result, in addition to those discussed above, include the use of a quarter-wave transformer and a tapered section, as discussed in connection with Figs. 4-27 and 4-28, respectively.

Nonreflecting Terminations for Ultra-high-frequency and Microwave Transmission Lines.[1,2] In some circumstances, particularly in measurement work, it is necessary to terminate a transmission line so that the reflected wave is as small as possible. In many cases this condition must be realized for a substantial band of frequencies. The problem of achieving a nonreflecting load contrasts with the case where one starts with an assigned load impedance that is to absorb the power and desires to match this load as well as possible to the transmission line.

A simple and effective means of obtaining a nonreflecting load impedance is to connect the end of the transmission line involved to a length of transmission line having high loss but the same characteristic impedance as the line being terminated. This arrangement is illustrated in Fig. 4-20*a*. An incident wave reaching such a termination will proceed into the lossy line and will be completely absorbed if the attenuation of the lossy line is sufficient. For example, if the attenuation of this line is 20 db, then even if the reflection coefficient at the end of the lossy line is unity, the reflected wave emerging from the junction of the two lines will be 40 db weaker than the incident wave, corresponding to a reflection coefficient of 0.01, or a standing-wave ratio of 1.02.

Lossy lines must be especially designed so that the total attenuation required can be achieved in a reasonable length. Flexible cable is commercially available for these applications in which the attenuation has been intentionally made very high by the use of insulation having high radio-frequency losses, and by employing resistance wire for the center conductor of the cable. In lines having air insulation, high attenuation can be obtained in coaxial systems by plating a high-resistivity coating on the center conductor of the coaxial line to give high skin-effect losses; in the case of two-wire open-air lines it is customary to obtain a high attenuation by using resistance wire or iron wire for the conductors.

[1] For further information on this subject see F. E. Terman and J. M. Pettit, "Electronic Measurements," sec. 14-7, McGraw-Hill Book Company, Inc., New York, 1952.

[2] Emphasis is placed here on the higher frequencies. At short-wave and lower frequencies lumped resistance terminations are entirely satisfactory.

An alternative type of nonreflecting termination that is particularly suitable for coaxial systems with air insulation makes use of a tapered section of lossy dielectric arranged as illustrated in Fig. 4-20*b*. The taper provides a gradual transition between the nonattenuating and the attenuating regions, so that no reflection is produced in spite of the fact that the dielectric changes the characteristics of the line. The lossy dielectric can be some type of plastic loaded with conducting material. Nonreflecting terminations of this type have the advantage that the total length of the termination is relatively small compared with the length of

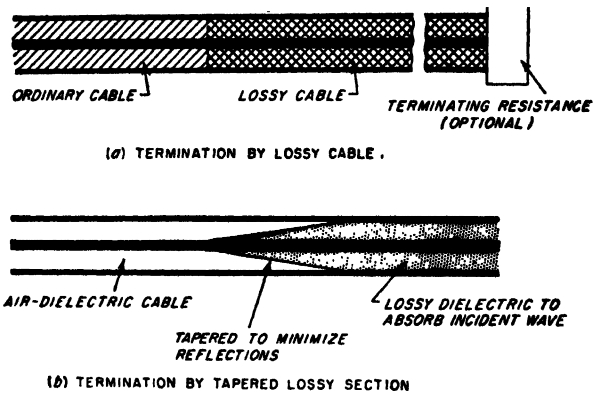


FIG. 4-20. Nonreflecting terminations for coaxial transmission lines, suitable for use at very high frequencies.

a lossy cable required to achieve a similar result. This difference arises from the fact that the taper makes it possible to work up to a very much higher attenuation per unit length without reflection than can be obtained in a uniform structure such as a lossy cable.

4-12. Artificial Lines. An artificial line is a four-terminal network composed of resistance, inductance, and capacitance elements. In so far as the terminals are concerned, such a network can be considered as being the equivalent of some transmission line when symmetrical about the mid-point, or a combination of a transmission line and a transformer when unsymmetrical.[1]

It can be demonstrated that any four-terminal network can have its properties at any one frequency represented, in so far as the terminals are concerned, by three independent constants. From this it follows that the most general artificial lines possible can be represented at any one frequency by three independent impedances. These can be arranged either in the form of a T or a π, as in Figs. 4-21*a* and 4-21*b*.[2] The L network

[1] The unsymmetrical case is also equivalent to a tapered transmission line.

[2] It will be noted that the T and π can be drawn as Y and Δ arrangements of impedances, respectively.

shown in Fig. 4-22 is a special case of the more general three-element network in which one of the impedance arms has become either zero or infinity.

The characteristics of a four-terminal artificial line can be expressed, from the transmission-line point of view, in terms of a propagation

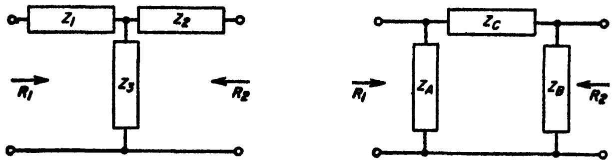


FIG. 4-21. General T and π networks.

constant $\alpha + j\beta$ that has exactly the same significance as in ordinary transmission-line theory (see Secs. 4-1 and 4-2), together with two characteristic impedances (or resistances), one associated with one pair of terminals and the other with the other set of terminals. When the network is symmetrical about its mid-point, i.e., when $Z_A = Z_B$ for the π network, or $Z_1 = Z_2$ for the T network, these two characteristic impedances are identical.[1] However, when the network is unsymmetrical, the two characteristic impedances differ, and the transmission line, in addition to introducing a certain attenuation and phase shift, also introduces a transformation of the characteristic impedance. The artificial line is then equivalent to a line plus a transformer or, what is the same thing, to a tapered line. as discussed below in connection with Fig. 4-27.

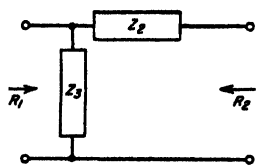


FIG. 4-22. General L network.

Artificial lines find extensive use in radio work for impedance matching

[1] The relations between the impedances of such a symmetrical artificial line and the constants Z_0 and $\alpha + j\beta$ of the equivalent transmission line are:

For symmetrical T section of Fig. 4-21*a* ($Z_1 = Z_2$):

$$Z_0 = \sqrt{Z_1^2 + 2Z_1Z_3} \qquad (4\text{-}25a)$$

$$\cosh(\alpha + j\beta) = 1 + \frac{Z_1}{Z_3} \qquad (4\text{-}25b)$$

For symmetrical π section of Fig. 4-21*b* ($Z_A = Z_B$):

$$Z_0 = \quad Z_AZ_C \qquad (4\text{-}26a)$$

$$\cosh(\alpha + j\beta) = 1 + \frac{Z_C}{Z_A} \qquad (4\text{-}26b)$$

Corresponding formulas for unsymmetrical artificial lines (that is, $Z_1 \neq Z_2$ and $Z_A \neq Z_B$) are given by F. E. Terman, "Radio Engineers' Handbook," p. 208, McGraw-Hill Book Company, Inc., New York, 1943.

and for introducing phase shifts. Networks for these purposes are composed of reactive elements having the smallest possible resistance and conductance in order that the artificial line will consume little or no energy. In this way it is possible to realize an artificial line which has almost zero attenuation and which simultaneously has resistive values for the characteristic impedances. The only effect that the presence of such an artificial line has on a traveling wave, other than the transformation in impedance level that may be present, is the introduction of a phase shift of β radians in the wave involved.[1]

Design of T and π Reactive Networks.[2] The design of an ideal network composed of reactive impedances with zero losses to give assigned values R_1 and R_2 of characteristic impedance and to introduce a desired phase shift β can be carried out with the aid of the following relations:

For T section:

$$\begin{aligned} Z_1 &= -j\frac{R_1 \cos\beta - \sqrt{R_1R_2}}{\sin\beta} \\ Z_2 &= -j\frac{R_2 \cos\beta - \sqrt{R_1R_2}}{\sin\beta} \\ &-j\frac{\sqrt{R_1R_2}}{\sin\beta} \end{aligned} \tag{4-27}$$

For π section:

$$\begin{aligned} Z_A &= j\frac{R_1R_2 \sin\beta}{R_2 \cos\beta - \sqrt{R_1R_2}} \\ Z_B &= j\frac{R_1R_2 \sin\beta}{R_1 \cos\beta - \sqrt{R_1R_2}} \\ Z_C &= j\sqrt{R_1R_2}\sin\beta \end{aligned} \tag{4-28}$$

The reactances obtained from these equations are inductive or capacitive according to whether their sign is + or −, respectively. R_1 and R_2 are the two values of characteristic impedance associated with the network. The angle β in Eqs. (4-27) and (4-28) is the angle by which the phase of the wave reaching the output terminals of the network lags behind the phase that the corresponding wave had at the input terminals; a negative value of β is possible and indicates that passage of the wave through the network advances the phase. It is to be noted that this phase shift is the same irrespective of the direction in which the wave travels through the network. A single reactive T or π section is capable of transforming the impedance level from any assigned resistance R_1 to any other resistance

[1] It is customary to discuss the behavior of an artificial line in terms of the incident and reflected waves that would exist on the equivalent transmission line. Although these wave trains cannot exist physically on the artificial line, the behavior, in so far as the terminals are concerned, is exactly as though they were present.

[2] For further information on design details see F. E. Terman, "Radio Engineers' Handbook," pp. 210-215, McGraw-Hill Book Company, Inc., New York, 1943.

R_2, without restriction on the values of these resistances, and is capable of introducing phase shifts of any desired value between 0 and $\pm 180°$.

In case the load (or for that matter the generator) has a reactive impedance component, this reactance can be used to supply part of the reactance required by the network. For example, if a load impedance $R_L + jX_L$ is connected to the right-hand terminals of the T network of Fig. 4-21*a*, then one would consider X_L to be part of the impedance Z_2 of the impedance-matching network. In the same way, if the load is regarded as a resistance shunted by a reactance, then the shunting reactance can be used to supply part of the shunt impedance Z_B of the π section of Fig. 4-21*b*.

L Reactive Networks. An L network composed of reactive impedances is able to transform from one arbitrarily assigned characteristic impedance to a second arbitrarily assigned characteristic impedance. However, since the L network contains only two circuit elements, the phase shift β introduced by the L section is determined by the ratio of these two impedances. The design equations of a reactive L network in terms of the characteristic impedances R_1 and R_2 at the two pairs of terminals are, assuming the configuration of Fig. 4-22, and that $R_1 > R_2$,

$$\begin{aligned} Z_2 &= \pm j\sqrt{R_2(R_1 - R_2)} \\ Z_3 &= \mp jR_1\sqrt{\frac{R_2}{R_1 - R_2}} \end{aligned} \tag{4-29}$$

One may employ either the two top signs, or the two bottom signs. The phase shift β corresponding to the characteristic impedances R_1 and R_2 is

(4-30)

4-13. Directional Couplers.[1] A directional coupler is a device that couples a secondary system only to a wave traveling in a particular direction on a primary line, and ignores entirely the wave traveling in the opposite direction.

Loop-type Directional Coupler. A number of types of directional couplers have been devised. One example is illustrated in Fig. 4-23*a*. This is a coaxial arrangement in which the secondary system consists of lines A and B interconnected by coupling loop D that projects into the primary line in such a manner as to be subjected to the simultaneous influence of the electric and magnetic fields produced by the waves traveling on the primary line.

The operation of this arrangement will now be explained. Assume that a wave is traveling on the primary line toward the right. The

[1] For a further discussion, together with an extensive list of references on the subject, see Terman and Pettit, *op. cit.*, p. 57.

electric field of this wave induces a charge on the loop D that produces a wave in part A of the secondary system, and also a wave in part B. The equivalent circuit that describes this action is illustrated in Fig. 4-23b; it consists of a voltage E_1 that is applied to coaxial systems A and B in parallel through series capacitance C_1, producing currents as indicated by the arrows. At the same time loop D links with the magnetic flux from the wave in the primary line, and therefore has a voltage E_2 induced in series with it, as illustrated by the equivalent circuit of Fig. 4-23b. This series voltage gives rise to an additional wave in A and likewise a second wave in part B. These magnetically induced waves are characterized by currents that flow in the directions indicated by the arrows.

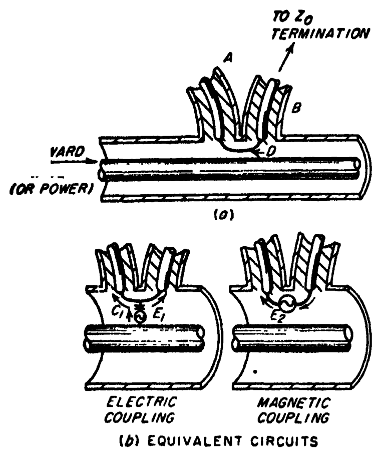


FIG. 4-23. Loop type of directional coupler for coaxial line, together with equivalent circuits that take into account the effects produced by the electric and magnetic fields on the primary line.

The two waves in section A produced by magnetic and electrostatic coupling, respectively, are of the same polarity and so add, while the two waves produced in section B are of opposite polarity and so tend to cancel each other. It is accordingly apparent that if the electric and magnetic couplings are so proportioned that the waves induced by the magnetic effect have the same amplitudes as the waves induced by the electric coupling, then complete cancellation takes place in section B. When this is the case, then a wave traveling to the right in the primary line will induce only one resultant wave in the secondary system, namely, a wave that travels in the direction of A. No wave is induced that travels in direction B.

The relative magnitude of electric and magnetic couplings in Fig. 2-23a can be readily controlled by the design of the coupling loop D. The electric coupling depends on the amount of electric field that terminates on the loop, and so is determined by the length of the loop and by the width (or diameter) of its conductor. Similarly, the magnetic coupling is determined by the amount of magnetic flux that links with the loop, and so is determined by the area enclosed between the loop and the outer conductor and by the orientation of the plane of the loop with respect to the axis of the line.

Assuming that the coupling arrangement in Fig. 4-23 has been designed so that a wave traveling to the right on the primary system produces no induced wave in part B, then consider the effect of a wave traveling to the left in the primary system. The component waves induced in A and B by the electric and magnetic fields in the primary coaxial line will again be equal to each other, since their relative magnitudes are not affected by the direction of travel of the primary wave. However, the polarity of the waves produced by magnetic coupling will now be reversed with respect to the polarity of the induced waves resulting from electric coupling. Accordingly, the two waves induced in A now cancel each other, while the two waves induced in B add. Consequently, a wave traveling to the left in the primary line produces no effect in section A, but does produce an induced wave traveling to the right in section B. By terminating B of the secondary system in its characteristic impedance, this induced wave is absorbed. The final result is that any wave traveling to the left in section A is determined only by the wave traveling to the right in the primary system, and is independent of the presence or absence of a wave traveling to the left in the primary system. Thus one has achieved a directional coupling system.

It is to be noted that to obtain the directional action it is absolutely necessary that B be terminated in its characteristic impedance. If the impedance terminating B produces a reflection, the resulting reflected wave will return along line B, pass through the coupling loop, and enter A. The actual wave traveling to the left in A will then be the resultant of the desired effect produced by the wave traveling to the right in the primary line and an undesired effect proportional to the product of the amplitude of the wave traveling to the left in the primary system and the reflection coefficient at the termination of B.

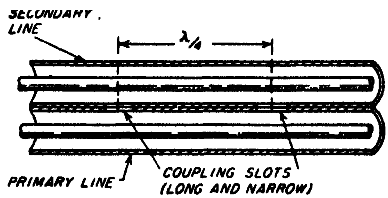


FIG. 4-24. Example of two-hole directional coupler for coaxial line.

Two-hole Coupler. A quite different type of directional coupling system is shown in Fig. 4-24. This is known as a two-hole coupler, and consists of primary and secondary systems which are coupled *either* electrically or magnetically at two points separated by an odd multiple of a quarter wavelength. It is essential that the coupling at each of these two points be either primarily electric or primarily magnetic. This result can be achieved by using probes (for electric coupling), loops (for magnetic coupling), or suitably shaped and oriented slots that favor either one or the other type of coupling.[1] In the two-hole coupler a wave traveling to the right in the primary system gives rise to a wave that also

[1] Details relating to the use of slots as a means of coupling are discussed on p. 133. The narrow axially oriented slots in Fig. 4-24 provide coupling that is predominately electrostatic.

travels to the right in the secondary system, but not to a wave traveling to the left; similarly, a wave traveling to the left in the primary system gives rise to a wave traveling to the left in the secondary system, but not to a wave traveling to the right. This result comes about because although each hole induces waves that travel in the secondary system in both directions away from the coupling point, the induced waves traveling in the favored direction away from the two holes add in phase, while those in the reverse direction cancel exactly if they are of equal amplitude, provided the holes are an odd multiple of a quarter wavelength apart.

Directivity and Coupling in Directional Couplers. In an ideal directional coupler, the secondary system will respond only to a wave traveling in the favored direction on the primary line. In actual directional couplers, mechanical imperfections, frequency differing from the design value, second-order effects, etc., will ordinarily result in a small output being produced by a wave traveling in the backward direction. The ratio of the responses to waves traveling in the two directions on the primary is called the *directivity* of the coupling system, and is commonly expressed in decibels. Thus a directivity of 30 db means that the undesired induced wave is 30 db weaker (representing only one-thousandth as much power) than the desired induced wave when equal waves travel in opposite directions on the primary line.

The ratio of power induced in the secondary system by a wave traveling in the desired direction on the primary line to the power of this wave on the primary line is called the *coupling* of the directional coupler. The coupling is ordinarily expressed in decibels, and represents the attenuation introduced by the coupling system.

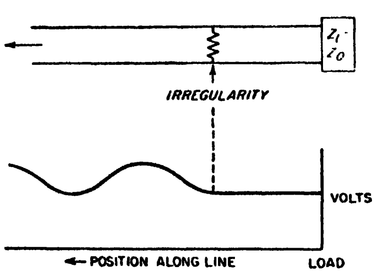


FIG. 4-25. Diagram illustrating standing waves produced on the generator side of an irregularity in the case of a transmission line terminated with a load impedance equal to the characteristic of the line.

4-14. Miscellaneous Aspects of Transmission Lines. *Transmission-line Irregularities—Discontinuity Capacitance.* When a wave traveling along a transmission line encounters an isolated discontinuity, it is partially reflected; i.e., while a portion of the wave continues to travel down the line, another portion of the wave is reflected backwards. Thus, in a transmission line terminated with a load equal to the characteristic impedance, an irregularity at some point on the line as shown in Fig. 4-25 will cause standing waves to exist on the generator side of the irregularity, as indicated.

Irregularities may be introduced in many ways. Typical causes are sharp bends, insulating supports, joints possessing resistance, coupled circuits, and extraneous objects that affect the electric or magnetic field, such as probes, dielectric or metal bodies, etc.

A type of irregularity that is particularly important at very high frequencies results from the distortion of the electric and magnetic fields associated with a change in line geometry. Consider, for example, a coaxial line in which the characteristic impedance changes abruptly as a result of a change in diameter of the outer conductor, as illustrated in Fig. 4-26*a*. It can be shown[1] that the distortion of the electric and magnetic fields in the vicinity of the junction is equivalent to shunting a capacitance across the junction, as shown in Fig. 4-26*b*, in addition to whatever effects are caused by the change in characteristic impedance. This "discontinuity capacitance" is ordinarily only a few tenths of a micromicrofarad; however, at ultra-high frequencies and higher frequencies its reactance becomes low enough to affect the behavior significantly.

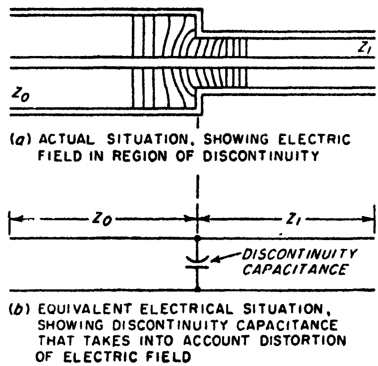


FIG. 4-26. Transmission line with discontinuity in the characteristic impedance, showing distortion of the electric field that results, and how this is taken into account by postulating a discontinuity capacitance at the point of irregularity in addition to the discontinuity in characteristic impedance.

A discontinuity capacitance is ordinarily present whenever a geometrical change occurs. Thus, in Figs. 4-16 and 4-18, the change in geometry at the points where the stubs are connected to the lines has an effect equivalent to a small capacitance connected in shunt across the coaxial line at the junction point. This shunting capacitance is in addition to the shunting action of the stub.

Tapered Transmission Lines.[2] A length of transmission line in which the characteristic impedance varies gradually and continuously from one value to another is said to be tapered. A traveling wave passing through such a section will have its ratio of voltage to current transformed in accordance with the ratio of the characteristic impedances involved.

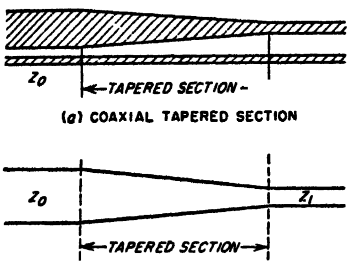
The requirement for a satisfactory taper is that the change in characteristic impedance per wavelength must not be too large; otherwise the

[1] See J. R. Whinnery, H. W. Jamieson, and T. E. Robbins, Coaxial Line Discontinuities, *Proc. IRE*, vol. 32, p. 695, November, 1944.

[2] For further information see Wilbur N. Christensen, The Exponential Transmission Line Employing Straight Conductors, *Proc. IRE*, vol. 35, p. 576, June, 1947; Charles E. Burrows, Exponential Transmission Line, *Bell System Tech. J.*, vol. 17, p. 555, October, 1938; Harold A. Wheeler, Transmission Line with Exponential Taper, *Proc. IRE*, vol. 27, p. 65, January, 1939; Moreno, *op. cit.*, pp. 53–55.

tapered section will introduce a reflection. That is, if the change in characteristic impedance per wavelength is excessive, then the tapered section acts as a lumped irregularity rather than producing merely a gradual transformation.

From these considerations it follows that a tapered section of transmission line acts as a perfect impedance transformer at the higher frequencies. However, as the frequency is lowered, such a section finally fails to be satisfactory as an impedance transformer, because the distance represented by a wavelength, and hence the change in characteristic impedance per wavelength, becomes greater. Thus as the frequency is reduced the taper introduces an increasingly large reflection. The practical lower-frequency limit of usefulness of a tapered section that thereby results corresponds to the frequency for which the characteristic impedance changes by a factor between about 1.3 and 4.0 per wavelength, with the exact value depending upon the standing-wave ratio that can be tolerated.



(b) TWO-WIRE TAPERED SECTION

FIG. 4-27. Tapered line.

A line can be tapered by varying the spacing of the conductors in the case of a two-wire line, or by varying the diameter of the inner (or outer) conductor in the case of a concentric line. The ideal type of taper is one in which the characteristic impedance changes uniformly with length, so that the higher derivatives of the rate of change of characteristic impedance with length are minimized. However, nearly as satisfactory results are obtained by the much more practical arrangement shown in Fig. 4-27, in which the spacing varies linearly with distance. Such straight-line tapers are accordingly used in ordinary practice.

Quarter- and Half-wave Transformers. Sections of transmission lines that are exactly a quarter wavelength or a half wavelength long have unique impedance-transforming properties that are frequently made use of in radio work. Thus consider the situation illustrated in Fig. 4-28. When the length l of the line is exactly an odd number of quarter wavelengths, then to the extent that the losses in l can be neglected, the impedance looking into the system is

$$Z_s = \frac{Z_0^2}{Z_L} \tag{4-31}$$

where Z_0 is the characteristic impedance of the line l. When the load impedance Z_L is a resistance, the effect of the line is thus to transform this resistance into another resistance Z_s that is inversely proportional to the resistance Z_L. Again, when Z_L is a capacitive reactance, then the impedance-transforming action of the line causes Z_s to be an inductive reactance having a magnitude inversely proportional to the capacitive reactance of Z_L.

In an arrangement such as illustrated in Fig. 4-28, the ratio of impedance transformation obtained can be varied by adjusting the characteristic impedance Z_0 of the connecting transmission line l. In the case of a two-wire line, this is readily accomplished by varying the spacing between the conductors that form the line. With coaxial lines, one can change the diameter of the inner conductor, or can move the inner conductor so that it is eccentric with respect to the outer conductor.

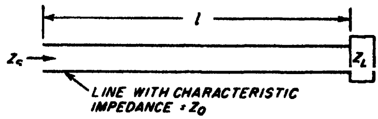


FIG. 4-28. Transmission line as an impedance transformer.

When the transmission line of Fig. 4-28 is exactly a whole number of half wavelengths long, then

$$Z_s = Z_L \tag{4-32}$$

This relation holds irrespective of the characteristic impedance of the line provided only that the line losses can be neglected. The half-wave line is thus a one-to-one impedance transformer. A typical practical application of such an arrangement is to provide a short circuit across an inaccessible pair of terminals. This can be achieved by connecting a transmission line to these terminals and then placing a short circuit across the line at an accessible point that is exactly a whole number of half wavelengths away from the terminals across which it is desired that a short circuit exist.

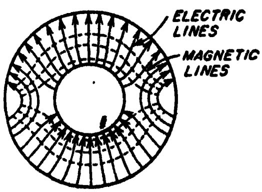


FIG. 4-29. First higher mode in a coaxial transmission line.

Higher-order Modes. When the spacing between the two wires of a transmission line exceeds a half wavelength, or when the circumference of a coaxial line exceeds a wavelength, it is possible for energy to propagate down the transmission line by using configurations of electric and magnetic fields that differ from the field arrangements ordinarily associated with transmission lines. These special configurations are termed *higher-order modes.* The first such higher-order mode that can exist on a coaxial transmission line is illustrated in Fig. 4-29. Fields of this particular type will propagate freely provided that the arithmetic mean circumference exceeds the wavelength λ' in the cable, i.e., when

$$\lambda' < 2\pi \frac{b + a}{2} \tag{4-33}$$

where a and b are the radii of the inner and outer conductors, respectively. Modes of still higher order are also possible on two-wire lines.

The amplitude of the higher mode (or modes), compared with the amplitude of the normal mode, is determined by the extent to which the method of applying voltage to the cable produces a field configuration corresponding to the higher mode (or modes). However, even if a higher mode is produced at the terminals of a transmission line, the mode will not propagate along the line unless the wavelength is less than the cutoff value given by relations such as Eq. (4-33). This can happen in ordinary cables and lines only at the higher microwave frequencies.

Loaded Lines. A loaded line is an ordinary transmission line to which lumped elements, usually capacitances or inductances, are added at

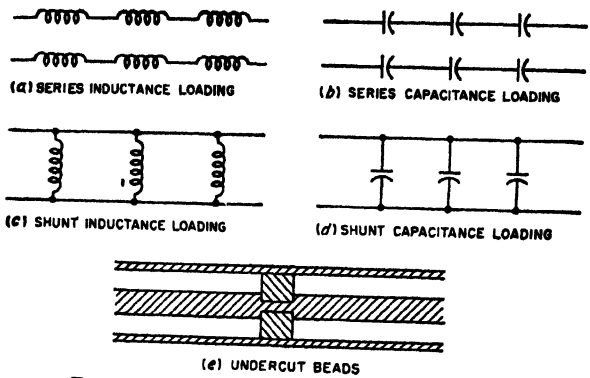


FIG. 4-30. Examples of loaded transmission lines.

regular intervals, as illustrated in Fig. 4-30. When these lumped loading impedances are spaced uniformly at distances that do not appreciably exceed a quarter wavelength, they act almost exactly as though their impedances were uniformly distributed. However, if the frequency is so high that the spacing appreciably exceeds a quarter wavelength, then the loading impedances act as irregularities that tend to prevent transmission.

The most common use of loading is in connection with telephone cables where inductance coils are commonly added at regular intervals, as in Fig. 4-30*a*. Such inductance loading makes the equivalent inductance per unit length of the transmission line greater than the actual inductance of the unloaded line, thereby increasing the characteristic impedance and lowering the velocity of propagation. Also, if most of the line losses result from the series resistance of the line, the attenuation constant is reduced by inductive loading. Loading by means of series capacitances, as in Fig. 4-30*b*, reduces the characteristic impedance and increases the phase velocity to a value greater than that of light, while shunt inductances as in Fig. 4-30*c* increase both characteristic impedance and the phase velocity. These results follow from Eqs. (4-18*a*) and (4-18*c*), by noting that series inductive loading increases the equivalent line inductance, and series capacitive loading reduces it, while shunt loading similarly increases the equivalent line capacitance if capacitive, and reduces it if inductive.

An important case of loading is provided by the beads sometimes used to support the center conductor of a concentric cable having air insulation. These beads introduce localized additions to the line capacitance and so represent shunt capacitive loading, as illustrated in Fig. 4-30*d*. This has the effect of increasing the effective line capacitance and thereby lowering the characteristic impedance and the velocity of phase propagation, as well as fixing an upper frequency beyond which the line does not behave properly. In order to overcome these effects, the beads are sometimes undercut, as in Fig. 4-30*e*. Here the reduction in diameter of the center conductor is so chosen as to make up for the increased dielectric constant of the space occupied by the bead, as well as the discontinuity capacitance introduced by the bead. In this way the characteristic impedance of the section containing the bead can be made the same as that of the portion of the line having only air insulation.

PROBLEMS AND EXERCISES

4-1. Assume that Fig. 4-1 is modified so that the length l is measured from the generator or sending end of the line instead of from the load end as in Fig. 4-1. Set up the differential equations of the line in terms of this notation, and obtain a solution to the transmission line analogous to Eqs. (4-6), but in terms of the amplitudes E_1', E_2', I_1', and I_2' of the individual waves at the generator end of the line.

4-2. Redraw Fig. 4-2*b*, *c*, and *d* for (*a*) an attenuation that is considerably greater than in Fig. 4-2, and (*b*) zero attenuation.

4-3. In a transmission line 100 ft long terminated so that only the incident wave is present, the power at the load end of the line is 1.2 db less than at the generator end. What is the value of α per foot?

4-4. Derive Eq. (4-14).

4-5. In a transmission line in which $Z_0 = 50$ ohms, calculate and plot the reflection coefficient as a function of load resistance for load resistances ranging from 0 to 250 ohms.

4-6. In a transmission line in which $Z_0 = 50$ ohms and which has a reactive load, calculate and plot the magnitude and phase angle of the reflection coefficient as a function of load reactance in the range from $-j100$ ohms to $+j100$ ohms.

4-7. Derive Eq. (4-15). In doing this, start by assuming that the incident wave at point a is E_a, and then express the magnitudes of the various waves at a and b in terms of E_a and $|\rho_a|$.

4-8. Derive Eq. (4-16*a*).

4-9. *a.* The line conductance will be negligible in a transmission line with air dielectric. Under these conditions the attenuation constant α of a radio-frequency line is proportional to the square root of the frequency. Explain.

b. In a coaxial transmission line with solid dielectric, the dielectric losses at extremely high frequencies will be very much greater than the losses resulting from the line resistance. Under these conditions, how does the attenuation constant α vary with frequency?

4-10. A transmission line with air dielectric is 20 m long. What is the line length, measured in wavelengths, and what is the value of β at frequencies of 10 and 100 Mc?

4-11. Show vector diagrams and curves for the current distributions that go with the voltage distributions in Fig. 4-3*a* and *b*. On the resulting curves show the voltage distributions of Fig. 4-3 by dotted lines.

4-12. Draw curves and vector diagrams similar to those of Fig. 4-3*b*, except for the case $\rho = 0.5\underline{/180^\circ}$.

4-13. Calculate the exact distance from the load in wavelengths at which the first voltage maximum occurs in Fig. 4-4*f* when $Z_L/Z_0 = j0.5$.

4-14. Sketch the voltage distribution on a low-loss transmission line in the manner shown in Fig. 4-4, but for the case where $Z_L/Z_0 = 2.0\underline{/45^\circ}$.

4-15. Sketch curves of voltage and current distribution on a low-loss transmission line analogous to the curves of Fig. 4-4, but for the following load impedance conditions. (Note: In each case calculate location and relative amplitude of minima and maxima accurately, and show these minima and maxima in correct positions and in correct magnitudes in the sketch, carrying the curves for a distance of slightly more than one wavelength from the load end of the line.)

a. Reflection coefficient at load $= 0.2\underline{/0^\circ}$.
b. Reflection coefficient at load $= 0.8\underline{/0^\circ}$.
c. Reflection coefficient at load $= 0.8\underline{/45^\circ}$.
d. Reflection coefficient at load $= 1.0\underline{/-45^\circ}$.

4-16. Derive a formula giving the distance from the load to the first minimum of voltage in terms of the phase shift β per unit length of line and the phase angle δ of the coefficient of reflection of the load.

4-17. Sketch curves analogous to those of Fig. 4-5, except applying to a short-circuited load.

4-18. Sketch curves analogous to those of Fig. 4-6, except assume that the attenuation of the line is approximately twice as great as in Fig. 4-6. For purposes of comparison, sketch the solid curves from Fig. 4-6 on the same axes.

4-19. Derive Eq. (4-22*a*) starting with Eq. (4-21).

4-20. Calculate and plot the standing-wave ratio as a function of Z_L/Z_0 for resistive loads, for values of this ratio ranging from 0.1 to 10.0.

4-21. Prove that resistive loads of R_1 and R_2 will produce the same standing-wave ratio provided $R_1R_2 = Z_0^2$.

4-22. Sketch curves similar to those of Fig. 4-10*a* and *b*, except applying to cases where the load is (*a*) a resistance greater than the characteristic impedance, and (*b*) an open circuit, respectively.

4-23. In a transmission line having negligible losses, derive formulas giving, respectively, the maximum impedance and the minimum impedance that can occur anywhere on the transmission line in terms of magnitude of the reflection coefficient of the load and the characteristic impedance of the line.

4-24. Sketch curves similar to those of Fig. 4-10*b* except applying to a transmission line having considerably greater attenuation.

4-25. When the load impedance of a transmission line is a resistance R_L, prove that $S = R_L/Z_0$ when $R_L > Z_0$, and likewise that $S = Z_0/R_L$ when $R_L < Z_0$.

Note that this proof shows that the standing-wave ratio corresponding to any particular circle on the Smith chart is given by the intersection of this circle with the horizontal axis, as stated in the second paragraph on page 101.

4-26. In a particular transmission line the load impedance is such that

$$Z_L = (0.8 - j0.6)Z_0$$

With the aid of the Smith chart, determine the standing-wave ratio on the line, and the magnitude and phase angle of the reflection coefficient.

4-27. Assuming that the line of Prob. 4-26 has negligible losses, plot the magnitude and phase of the line impedance as a function of distance from the load up to a distance slightly greater than one wavelength. Make use of the Smith chart to determine the resistive and reactive components of the impedance.

4-28. An impedance of $35 + j75$ ohms is connected across the load end of a transmission line having a characteristic impedance of 60 ohms.

a. With the aid of the Smith chart, and assuming that the line has negligible losses, determine the standing-wave ratio produced on the line, and also the input impedance of the line when the line length is 1.8λ.

b. If the total attenuation of the line is 1.4 db, determine the standing-wave ratio and the line impedance at the generator end of the line.

c. Tabulate the results from (*a*) and (*b*) side by side, and explain in physical terms how attenuation accounts for the differences observed.

4-29. With the aid of the Smith chart determine the magnitude and the phase angle of an impedance which, when placed at the receiving end of a transmission line having the characteristic impedance $R_0 = 75$ ohms, would account for an observed standing-wave ratio of 1.65 with a voltage distribution such that the voltage minima with a short-circuited load are 0.2λ closer to the load than the voltage minima produced by the impedance to be measured.

4-30. Same as Prob. 4-29, except that $S = 2.10$, and the minima with a short-circuited load are 0.10λ closer to the *generator* than the minima produced by the impedance to be determined.

4-31. A concentric transmission line having copper conductors and air insulation is short-circuited at the receiving end and is to be in quarter-wavelength resonance at a frequency of 100 Mc. Determine (*a*) the smallest diameter of the outer concentric line for which a Q of 5000 can be obtained, and (*b*) the sending-end impedance of the line in *a*.

4-32. A resonant quarter-wave coaxial transmission line 25 cm long has $b = 1$ cm and $b/a = 3.6$. Determine the resonant frequency, Q, and standing-end impedance.

4-33. A load impedance is connected to a transmission line and is found to produce a standing-wave ratio of 2.0. The first voltage minimum occurs at a distance of 0.4 wavelength from the load. Design a stub-line impedance-matching system for this situation.

4-34. A load impedance of $70\underline{/30°}$ is connected to a concentric transmission line having a characteristic impedance of 50 ohms. Calculate the resulting standing-wave ratio and the location of the voltage minima. From this information specify the length and position of a stub line that will match the load to the transmission line.

4-35. Assume that the double-slug tuner of Fig. 4-19*b* is adjusted to give an impedance match. Will this impedance match be destroyed if the right-hand slug is then displaced a half wavelength to the right, while leaving the position of the other slug unchanged? Explain.

4-36. A short-circuited lossy line is used to terminate a transmission line. How much total attenuation must the lossy line have if the standing-wave ratio on the terminated line is not to exceed $S = 1.10$? Assume the lossy line is open-circuited.

4-37. Design a reactive T network that at 1000 kc will match a load impedance of 100 ohms to a line having a characteristic impedance of 50 ohms, and introduce a phase shift of 30° leading in the load current.

4-38. Design a reactive T network that will match a load impedance of $100 + j50$ ohms to a 50-ohm line, and introduce a phase shift of 30° leading in the load current.

4-39. Explain how the directional coupler of Fig. 4-23 can be arranged so that the wave in the secondary section B is proportional to the wave traveling to the left on the primary line and is not affected by the primary wave traveling to the right.

4-40. Explain how one could measure the magnitude of the reflection coefficient on a line by apparatus including (*a*) a directional coupler of the type illustrated in Fig. 4-23, and (*b*) two instruments suitable for measuring voltage on transmission lines.

4-41. In the directional coupler system of Fig. 4-23, assume that the left-hand side of the secondary system (i.e., line A) is terminated by a load equal to the characteristic impedance but that the right-hand is not. Under these conditions prove that the intensity of the secondary wave traveling to the right in B is proportional to the strength of the primary wave traveling to the left, irrespective of the presence or absence of a primary wave traveling toward the right, but that the secondary wave traveling to the left in A is dependent on *both* the primary wave traveling to the left and the primary wave traveling to the right.

4-42. Give a detailed explanation of why the two-hole directional coupler of Fig. 4-24 theoretically can give ideal directional coupler action only when the hole spacing is exactly $n\lambda/4$, where n is odd. Include a justification for the fact that increasing the spacing by a half wavelength makes no difference.

4-43. A two-hole coupler such as illustrated in Fig. 4-24 is operated at a frequency 5 per cent higher than the value that makes the hole spacing exactly $\lambda/4$. What is the directivity in decibels caused by this incorrect operating frequency?

4-44. Sketch a curve similar to that of Fig. 4-25, except for an irregularity that is a series resistance equal in magnitude to the characteristic impedance. Be careful to show the correct standing-wave ratio on the generator side of the irregularity, as well as the correct location of the minima with respect to the irregularity, and also show the voltage drop in the series resistance.

4-45. In Fig. 4-25, the irregularity consists of a shunt discontinuity capacitance of 0.2 $\mu\mu$f. Determine the standing-wave ratio on the generator side of the irregularity at 100 and 10,000 Mc, assuming that the characteristic impedance of the line is 50 ohms.

4-46. Two coaxial lines having characteristic impedances of 50 and 100 ohms, respectively, are to be joined by a tapered section. If it is desired that the reflections introduced by the tapered section be kept very small in the frequency range 2000 to 11,000 Mc, determine the minimum length of tapered section that can be used.

4-47. A load resistance of 300 ohms is to be matched to a two-wire transmission line having a characteristic impedance of 600 ohms by means of a quarter-wave matching line. What characteristic impedance must the matching line have?

4-48. From the behavior of incident and reflected waves, demonstrate the correctness of Eq. (4-32).

4-49. In a particular coaxial transmission line, $b/a = 3.6$ and $b = 1$ cm. What is the shortest wavelength that can be transmitted on the line without danger of a higher-order mode being generated?

4-50. Explain with the aid of Eq. (4-18*c*) why the different types of loading illustrated in Fig. 4-30*a* to *d* have the effects on phase velocity summarized on page 122.

CHAPTER 5

WAVEGUIDES AND CAVITY RESONATORS

5-1. Waveguides—General Considerations.[1,2] A hollow conducting tube used to transmit electromagnetic waves is termed a waveguide. At ultra-high and microwave frequencies, waveguides provide a practical alternative to transmission lines for the transmission of electrical energy.

Any configuration of electric and magnetic fields that exists inside a waveguide must be a solution of Maxwell's equations. In addition, these fields must satisfy the boundary conditions imposed by the walls of the guide. To the extent that the walls are perfect conductors there can therefore be no tangential component of electric field at the walls. Many different field configurations can be found that meet these requirements. Each such configuration is termed a *mode*.

A critical examination of the various possible field configurations or modes that can exist in a waveguide reveals that they all belong to one or the other of two fundamental types. In one type, the electric field is everywhere transverse to the axis of the guide, and has no component

[1] The practical possibilities of waveguides as transmission systems for very high-frequency waves was discovered independently and almost simultaneously by W. L. Barrow and G. C. Southworth. Fundamental papers on the subject include: W. L. Barrow, Transmission of Electromagnetic Waves in Hollow Tubes of Metal, *Proc. IRE*, vol. 24, p. 1298, October, 1936; G. C. Southworth, Hyper-frequency Wave Guides—General Considerations and Experimental Results, *Bell System Tech. J.*, vol. 15, p. 284, April, 1936; L. J. Chu and W. L. Barrow, Electromagnetic Waves of Hollow Metal Tubes of Rectangular Cross Section, *Proc. IRE*, vol. 26, p. 1520, December, 1938.

[2] The discussion given here of waveguides is intended to provide a description of their more important characteristics, together with formulas for calculating quantitatively their principal characteristics. The rigorous derivation of the quantitative relations existing in waveguides is a specialized subject that would take more space than is available in a book of this type. The reader who wishes to study the techniques by which waveguide equations are derived is referred to Ramo and Whinnery, "Field and Waves of Modern Radio," John Wiley & Sons, Inc., New York, 1944; H. H. Skilling, "Fundamentals of Electric Waves," John Wiley & Sons, Inc., New York, 1948. An excellent discussion of the physical phenomena involved in waveguides is given by H. G. Booker, The Elements of Wave Propagation Using the Impedance Concept, *J. IEE*, vol. 94, part III, p. 171, May, 1947. Useful summary information on waveguide techniques is given by M. H. L. Prece, Waveguides, *J. IEE*, vol. 93, part IIIA, no. 1, p. 33, 1946; T. E. Moreno, "Microwave Transmission Data," McGraw-Hill Book Company, Inc., New York, 1948.

anywhere in the direction of the guide axis; the associated magnetic field does, however, have a component in the direction of the axis. Modes of this type are termed *transverse electric* or *TE modes* (also sometimes called *H* modes). In the other type of distribution, the situation with respect to the fields is reversed, the magnetic field being everywhere transverse to the guide axis while at some places the electric field has components in the axial direction. Modes of this type are termed *transverse magnetic* or *TM modes* (also sometimes called *E* modes).[1] The different modes of each class are designated by double subscripts, such as TE_{10}, as explained below.

The behavior of a waveguide is similar in many respects to the behavior of a transmission line. Thus waves traveling along a guide have a phase velocity, and are attenuated. When a wave reaches the end of a guide it is reflected unless the load impedance is carefully adjusted to absorb the wave; also an irregularity in a waveguide produces reflection just as does an irregularity in a transmission line. Again, reflected waves can be eliminated by the use of an impedance-matching system, exactly as with a transmission line. Finally, when both incident and reflected waves are simultaneously present in a waveguide, the result is a standing-wave pattern, such as illustrated in Fig. 4-4, that can be characterized by defining a standing-wave ratio.

In some other respects waveguides and transmission lines are unlike in their behavior. The most striking difference is that a particular mode will propagate down a waveguide with low attenuation only if the wavelength of the waves is less than some critical value determined by the dimensions and the geometry of the guide. If the wavelength is greater than this critical *cutoff* value, the waves in the waveguide die out very rapidly in amplitude even when the walls of the guide are of material having infinite conductivity. Different modes have different values of cutoff wavelength; the particular mode for which the cutoff wavelength is greatest is termed the *dominant* mode.

5-2. Rectangular Waveguides. The most frequently used type of waveguide has a rectangular cross section, as illustrated in Fig. 5-1. In such a guide, the preferred mode of operation is the dominant mode.

Field Configuration of the Dominant Mode in a Rectangular Waveguide. At wavelengths less than the cutoff value, the electric and magnetic fields representing the dominant mode in a rectangular waveguide have the character illustrated in Fig. 5-2. Here the electric field is transverse to the guide axis, and extends between the two walls that are closest together, i.e., between the top and bottom of Fig. 5-1. The intensity of this elec-

[1] Following this system of designation, the field configuration normally associated with a coaxial line is sometimes called the TEM mode, because both the electric and magnetic fields are transverse to the axis of the line. The higher-order coaxial mode illustrated in Fig. 4-29 is a TE mode, since the electric field is everywhere transverse to the line.

tric field is maximum at the center of the guide, and drops off sinusoidally to zero intensity at the edges, as shown. The magnetic field is in the form of loops which lie in planes that are at right angles to the electric field, i.e., planes parallel to the top and bottom of the guide in Fig. 5-1. The magnetic field is the same in all of these planes, irrespective of the position of the plane along the y axis.

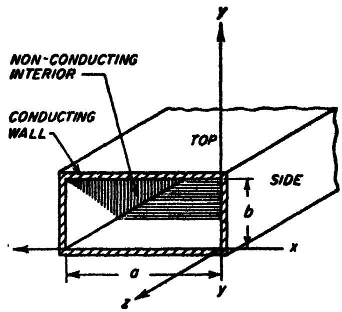


FIG. 5-1. Rectangular waveguide, illustrating notation.

This field configuration travels along the waveguide axis (in the z direction in Fig. 5-1).[1] As it travels a distance l down the guide, the amplitude will be reduced by the factor $\epsilon^{-\alpha l}$ as a result of energy losses in the walls of the guide, and the wave will drop back in phase βl radians, just as in the analogous transmission-line case, where α and β are termed the *attenuation constant* and *phase constant* respectively.

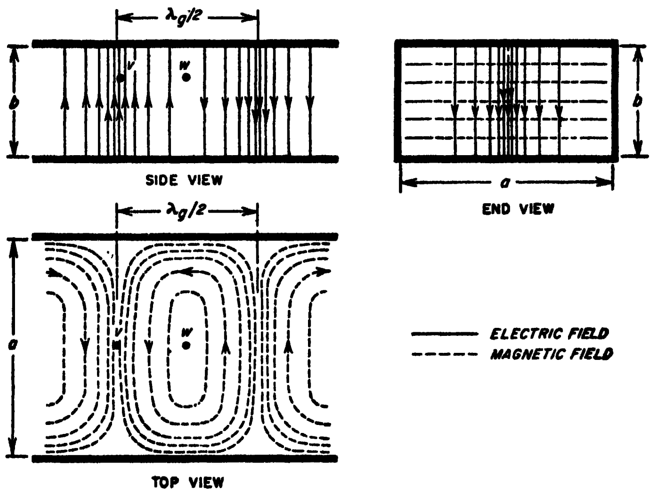


FIG. 5-2. Field configuration of the dominant or TE_{10} mode in a rectangular waveguide.

The field configuration representing the dominant mode illustrated in Fig. 5-2 is seen to be a transverse electric mode and is designated as the

[1] Thus Fig. 5-2 can be regarded as representing a snapshot of the fields as they exist at some particular moment. As this field configuration travels down the guide, the fields at any given point vary sinusoidally in amplitude. Thus, although the fields at position w in Fig. 5-2 have zero intensity, a quarter cycle later the fields will have moved a distance $\lambda_g/4$, and the amplitude at w will then be the same as the amplitude shown for position v in Fig. 5-2.

TE_{10} mode. The subscript 1 means that the field distribution in the direction of the long side of the waveguide (x direction in Fig. 5-1) contains one-half cycle of variation. The subscript 0 indicates that there is no variation in either the electric or magnetic field strength in the direction of the short side (y axis) of the guide.

The equations giving the fields at frequencies above cutoff for the dominant mode in a rectangular waveguide filled with air are as follows:

$$\begin{aligned} E_z &= E_x = B_y = 0 \\ E_y &= A\frac{\omega a}{\pi}\sin\frac{\pi x}{a}\sin(\omega t - \beta z) \\ B_z &= -A\cos\frac{\pi x}{a}\cos(\omega t - \beta z) \\ B_x &= \frac{\beta}{\omega}E_y \end{aligned} \tag{5-1}$$

where E = electric field intensity, abvolts por cm
B = magnetic field intensity, gauss
$\omega/2\pi$ = frequency
t = time
A = an arbitrary constant of amplitude

The quantities a, x, y, and z have meanings indicated in Fig. 5-1. Subscripts x, y, and z indicate components in these respective directions. Finally β, the phase constant, has the value given by Eq. (5-4) below.

Cutoff Wavelength in a Rectangular Waveguide. Field configurations such as those illustrated in Fig. 5-2 can exist and propagate down a guide only when the frequency is such that the free-space wavelength is greater than a certain critical value termed the *cutoff wavelength,* commonly denoted as λ_c. For the dominant mode in rectangular waveguide, the cutoff wavelength is exactly twice the width a of the guide. That is

$$\left.\begin{array}{l}\text{Cutoff wavelength based}\\ \text{on free-space conditions}\end{array}\right\} = \lambda_c = 2a \tag{5-2}$$

If the frequency is less than the cutoff value, so that the free-space wavelength is greater than λ_c, then the waves attenuate rapidly with distance down the guide, as discussed in Sec. 5-8, instead of propagating freely.

The fact that a waveguide must have a dimension approaching a wavelength in order for the fields to propagate limits the practical use of waveguides to extremely high frequencies. For example, to transmit 300 Mc the guide width must exceed 20 in.

Each mode that can exist in a waveguide has its own cutoff wavelength The dominant mode is by definition the particular mode having the largest possible cutoff wavelength (lowest cutoff frequency). Accordingly, there is usually a frequency range between the dominant and the next higher mode in which only the dominant mode will propagate freely. By so proportioning a waveguide that the frequency to be transmitted lies in this range, all higher modes are suppressed after traveling a short distance down the guide; thereafter the only fields present in the guide will be those of a single pure mode, the dominant mode. In the case of a rectangular guide so proportioned that $a = 2b$, such single-mode operation occurs for free-space wavelengths lying between $2a$ and a. This matter is discussed further on page 138.

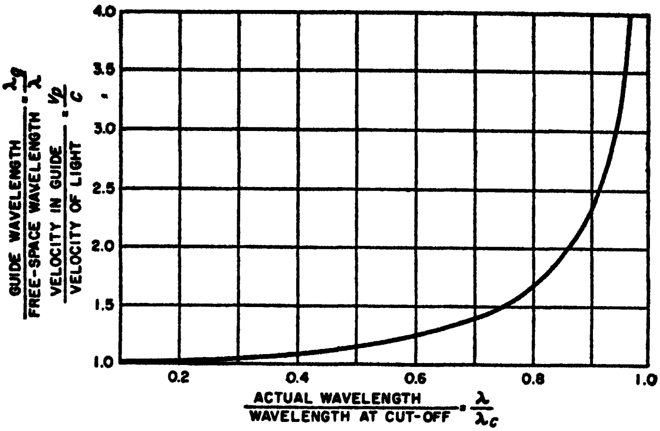


FIG. 5-3. Variation of phase velocity and wavelength in waveguides as a function of ratio of actual wavelength to the cutoff wavelength.

Guide Wavelength, Phase Constant, Group and Phase Velocity. The axial length λ_g corresponding to one cycle of variation of the field configuration in the axial direction (see Fig. 5-2) is termed the *guide wavelength.* It is related to the free-space wavelength λ and the cutoff wavelength λ_c according to the equation

$$\text{Guide wavelength} = \lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_c)^2}} \qquad (5\text{-}3)$$

Results calculated from Eq. (5-3) are plotted in Fig. 5-3. It will be noted that the guide wavelength exceeds the wavelength in free space, with the ratio of the two becoming increasingly large as the cutoff wavelength is approached.

The guide wavelength λ_g also represents the distance that a wave travels down the guide when undergoing a phase shift of 2π radians. Accordingly, the phase constant β, representing the phase shift per unit distance traveled by the wave, has the value

$$\beta = \frac{2\pi}{\lambda_g} = \frac{2\pi}{\lambda}\sqrt{1 - \left(\frac{\lambda}{\lambda_c}\right)^2} = \sqrt{\frac{\omega^2}{c^2} - \frac{\pi^2}{a^2}} \qquad (5\text{-}4)$$

where c is the velocity of light. It will be noted that the phase constant β has the same significance in waveguides as in transmission lines.

The quantity $v_p = f\lambda_g$ is the distance the wave travels in f cycles (i.e., one second) and so has the dimension of a velocity. Termed the ***phase velocity***,[1] it is related to the velocity of light c by the equation

$$\frac{\text{Phase velocity}}{\text{Velocity of light}} = \frac{v_p}{c} = \frac{\lambda_g}{\lambda} = \frac{1}{\sqrt{1 - (\lambda/\lambda_c)^2}} \tag{5-5}$$

This relation is plotted in Fig. 5-3. It is seen from Eq. (5-5) that the velocity of phase propagation always exceeds the velocity of light. In

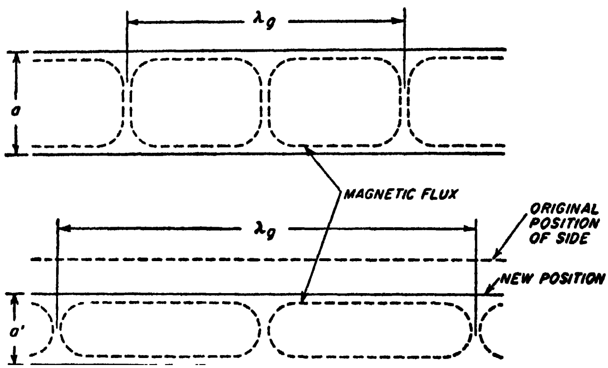


FIG. 5-4. Top view of TE_{10} field in a rectangular waveguide, showing the effect upon the guide wavelength λ_g of reducing the guide dimension a.

particular, as the frequency is lowered so that it approaches the cutoff value, the phase velocity increases and becomes indefinite at cutoff. Similarly, as the guide width is reduced so that the cutoff wavelength is made to approach the free-space wavelength, the phase velocity and λ_g increase and β decreases. This behavior arises from the fact that, as the width of the guide is reduced while keeping the frequency constant, the field configuration required to satisfy Maxwell's equations is affected in the manner shown in Fig. 5-4; specifically, compressing the flux sidewise by narrowing the guide is compensated for by an axial expansion that increases the guide wavelength and hence the phase velocity.

Currents in Waveguide Walls. The fields inside a waveguide induce currents that flow on the inner surface of the walls and that can be considered to be associated with the magnetic flux adjacent to the wall. The

[1] The phase velocity is an apparent velocity deduced from the rate of phase change with position along the axis. The actual velocity with which a pulse of energy travels is termed the group velocity v_{gr}, and is related to v_p and c by the equation $v_p v_{gr} = c^2$. Thus the group velocity is less than the velocity of light to the extent that the phase velocity is greater. This matter is discussed further on p. 142.

relationship between flux density at the surface of the wall and the current flowing in the wall is given by Eq. (2-18). The direction in which the current flows at any point in the wall is at right angles to the direction of the adjacent magnetic flux. The resulting lines of instantaneous current flow in the walls of a rectangular guide for the dominant mode are illustrated in Fig. 5-5. In the sides of the guide the current everywhere flows vertically, since the magnetic flux in contact with the side walls lies in planes parallel to the top and bottom sides of the guide. In the top and bottom of the guide there are a transverse component of current proportional to the axial component B_z of magnetic field, and an axially flowing current component proportional at any point to the transverse magnetic field B_x.

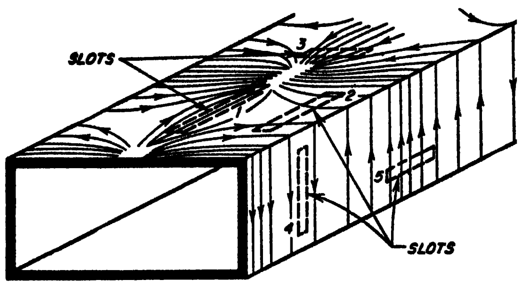


FIG. 5-5. Paths of current flow in the walls of a rectangular waveguide when propagating the dominant mode, showing slots transverse and parallel to the lines of current flow.

The current in the guide walls penetrates in accordance with the laws of skin effect, as given in Eq. (2-9). The depth of penetration is accordingly inversely proportional to the square root of the frequency. At the very high frequencies at which waveguides are used, this penetration is very small, and the walls provide practically perfect shielding.

Coupling and Leakage through Slots and Holes in Waveguide Walls. A hole or joint or slot in the waveguide wall introduces the possibility that energy will leak from the guide to outer space. When this happens, the fields inside the guide are affected, thereby introducing an irregularity with resulting reflection. The coupling thus introduced by a hole in the guide wall may be either to the electric or magnetic fields inside the guide. Electric coupling occurs when electrostatic flux lines that would normally terminate on the guide wall are able to pass through the hole into outside space. Magnetic coupling results when the hole or slot interferes with the current flowing in the guide wall. With either type of coupling, both electric and magnetic fields will be present outside the guide. Thus electric flux leaking through the hole will induce currents on the outer surface of the guide that produce a magnetic field. Again, when magnetic flux leaks through a hole, the associated interference with the flow of currents in the wall produces a voltage across the hole that gives rise to an electric field that will extend outside of the guide.

The nature and magnitude of the coupling in any particular case depend upon the size, shape, and orientation of the coupling hole, and upon the thickness of the guide wall. The factors involved can be understood by considering the effects produced by long narrow slots oriented in various ways, as illustrated in Fig. 5-5. Thus slot 1, which is transverse to the magnetic field inside the guide and so produces a minimum of interference with currents in the guide wall, introduces little or no magnetic coupling. It will, however, permit electric coupling if the slot width is great enough in proportion to the wall thickness to permit a reasonable number of electric flux lines to pass through the slot. However, if the slot is in the nature of a joint representing two surfaces fitted together, or is very narrow, then the electric coupling will be negligible. Similarly, long, narrow slot 4 produces little magnetic coupling because it is transverse to the magnetic flux and therefore interferes only negligibly with the flow of current in the guide wall; neither does it produce electric coupling because there is no electric field terminating on the side wall. Such a slot will therefore have negligible effect even if it is quite long. In contrast, slot 5, while causing no electric coupling, introduces a substantial amount of magnetic coupling to outside space through the fact that its long dimension is parallel to the magnetic field in the guide; this slot is hence oriented in such a manner as to permit easy escape of magnetic flux lines and to interfere to a maximum extent with the wall currents. This coupling is fully effective even if the slot is quite narrow, since it is necessary only that the slot interrupt the flow of current in the wall. Slots 2 and 3 in Fig. 5-5 also give rise to magnetic coupling, because they interfere with the flow of current in the guide wall. In the case of slot 2, the amount of magnetic coupling will be greater the farther the slot is to the side of the center line of the guide. Slots 3 and 2 will also simultaneously introduce electric coupling to the extent that the slot is wide enough in relation to the wall thickness to permit the passage of electric flux. In the case of slot 2, the electric coupling becomes less the farther the slot is from the center line, because the intensity of the electric field terminating on the top and bottom sides of the guide becomes less as the side walls are approached.

Attenuation. The propagation of energy down a waveguide is accompanied by a certain amount of attenuation as a result of the energy dissipated by the current induced in the walls of the guide. The magnitude of this current at any point is determined by the intensity of the magnetic field adjacent to the wall at that point, as explained above. The resistivity that the induced currents encounter is determined by the skin effect of the wall as discussed in Sec. 2-4, and is proportional to the square root of the frequency and the square root of the resistivity of the material of which the wall is composed.

The total energy loss in a waveguide can be calculated by summing up the I^2R loss in the top, bottom, and two sides of the guide for each unit area over a length corresponding to a half wavelength. This is done for the magnetic field distribution actually present, as calculated by Eqs. (5-1), assuming the field at any one point varies sinusoidally with time; under these circumstances the rms value of the field (and current) determines the time average of the power loss occurring at the point.[1]

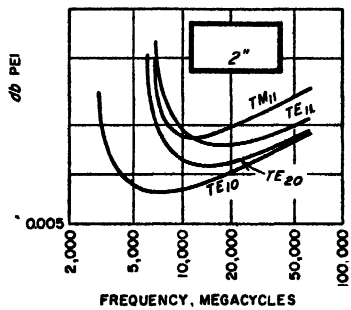


FIG. 5-6. Attenuation of different modes in a particular rectangular copper waveguide as a function of frequency.

The energy loss is conveniently expressed in decibels attenuation per unit length. With rectangular guides the loss has the general behavior illustrated in Fig. 5-6. It will be noted that for each mode there is a particular frequency for which the attenuation is a minimum. This is a result of two opposing tendencies. Thus as the frequency is lowered the skin depth becomes greater, causing the effective resistivity of the walls to decrease. At the same time, as the frequency approaches the cutoff value for the mode in question, the group velocity decreases. This causes the magnetic fields adjacent to the walls to become rapidly stronger for a given rate of energy flow down the guide.

5-3. Higher Modes in Rectangular Waveguides. The dominant mode is only one of an infinite series of field configurations that can exist in a waveguide. Fields for several of the higher-order modes that are possible in a rectangular waveguide are illustrated in Fig. 5-7.[2] In addition to TE modes, these include TM types, in which the magnetic flux lines lie in planes that are at right angles to the axis of the guide.

These various modes are designated by double subscripts, such as TE_{10}, TE_{20}, TE_{11}, TE_{mn}, TM_{11}, TM_{21}, and TM_{mn}. In this system of nomenclature the first subscript denotes the number of half-period variations of the electric (or magnetic) field in the transverse plane in the direction of the long side of the rectangle (along the x axis in Fig. 5-1); the second subscript denotes the number of half-period variations of the same field in the direction of the short side of the rectangle (along the y axis in Fig. 5-1).

[1] Formulas for the attenuation of different modes in rectangular waveguides are given by Moreno, *op. cit.*, chap. 8; they are also to be found in most handbooks.

[2] Equations for the fields of the various higher modes are to be found in many reference books; for example, see Moreno, *op. cit.*, p. 115.

Each mode has its own cutoff wavelength, guide wavelength, phase constant, and phase and group velocities. Equations (5-3) to (5-5) giving relations between these quantities apply to the higher-order modes as well as to the dominant mode [except for the right-hand form of Eq. (5-4)].

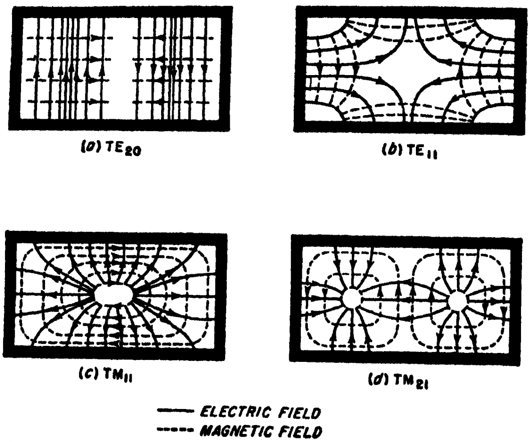


Fig. 5-7. Field configurations in the transverse plane for the first four higher modes in a rectangular waveguide.

The cutoff wavelength in the general case is given by the relation

$$\lambda_c = \frac{2a}{\sqrt{(m^2) + (na/b)^2}} \qquad (5\text{-}6)$$

Here a and b have the significance shown in Fig. 5-1, and m and n are, respectively, the first and second subscripts describing the mode. Equation (5-1), giving the cutoff wavelength of the dominant mode, is a special case of Eq. (5-5), in which $m = 1$ and $n = 0$. Results from Eq. (5-6) are tabulated in Table 5-1 for a few of the lowest-order modes for rectangular

TABLE 5-1
CUTOFF WAVELENGTHS IN WAVEGUIDES

Rectangular guide $a = 2b$		Square guide $a = b$		Circular guide radius $= r$	
Mode	Cutoff wavelength	Mode	Cutoff wavelength	Mode	Cutoff wavelength
TE_{10}	$2a$	TE_{10}	$2a$	TE_{11}	$3.42r$
TE_{01}	a	TE_{01}	$2a$	TM_{01}	$2.61r$
TE_{20}	a	TE_{11}	$1.4a$	TE_{21}	$2.06r$
TE_{11}	$0.89a$	TM_{11}	$1.4a$	TE_{01}	$1.64r$
TM_{11}	$0.89a$	TE_{20}	a	TM_{11}	$1.64r$

guides that are square ($a/b = 1$), and for the shape $a/b = 2$ that is customarily used.

Generation of Different Waveguide Modes. Any actual configuration of electric and magnetic fields existing in a waveguide can be regarded as being the result of a series of modes that are superimposed upon one another. If the magnitude, phase, and position along the axis of each individual mode is properly chosen, then the sum of the fields of the individual modes can be made to equal any actual electric and magnetic fields that can be present. Modes in waveguides are thus analogous to the harmonics of a periodic wave, since a periodic wave of arbitrary shape

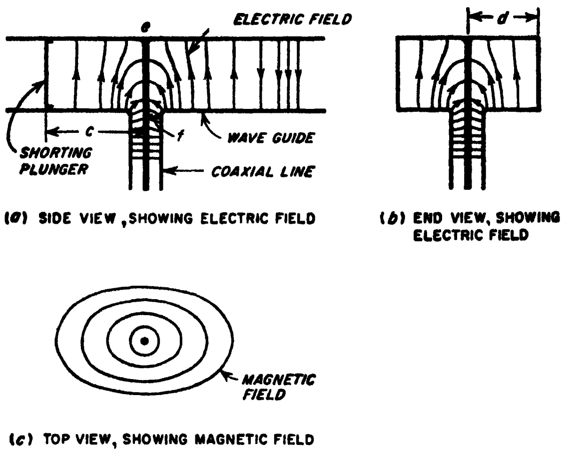


FIG. 5-8. Launching of TE_{10} wave in a waveguide excited by a coaxial line.

can always be considered as being represented by the sum of a series of properly chosen harmonic components.

The magnitude of each component mode associated with a given configuration of fields is determined by the character of the field distribution involved. For example, consider the arrangement illustrated in Fig. 5-8, where a concentric line delivers energy to a waveguide as a result of the electric and magnetic fields produced in the waveguide by the extension *ef* of the center coaxial conductor that extends from the bottom to the top of the guide. Current in *ef* generates a magnetic field in the guide, which lies in planes parallel to the top and bottom sides of the guide. At the same time, the voltage drop along *ef*, and the consequent difference in voltage thereby produced between the top and bottom of the guide, result in electric fields being produced as shown. This configuration suggests the TE_{10} mode, in that the magnetic field lies in planes parallel to the top and bottom of the guide, while the electric field is vertical and is maximum midway between the sides of the guide. Thus the TE_{10} is the largest single component in the field configuration of Fig. 5-8. The difference between the field configuration of this mode and the actual field present is then accounted for by the presence of a succession of higher-order modes, each of which is of smaller amplitude than the TE_{10} component. These higher-order modes will be primarily TE types, since examination of Fig. 5-8 indicates that, except to a very minor extent, the electric field is everywhere almost exactly transverse to the guide axis. Again, since the coupling element *ef* is located midway between the sides, the system is symmetrical with respect to the center of the guide; this means that, for this particular situation, no mode can be present that is unsymmetrical about the guide center; i.e., modes such as the TE_{20} or TE_{40} cannot exist.

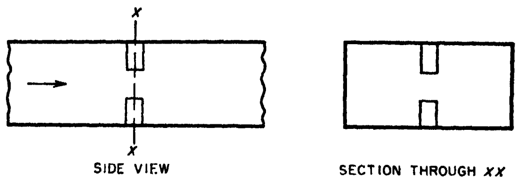


FIG. 5-9. Rectangular waveguide with vertical stub posts extending into the guide from the top and bottom sides.

While the modes initially present in a waveguide are determined by the field configuration used to excite the guide, new modes are generated whenever the field is distorted. For example, consider the situation in Fig. 5-9, where an obstacle in the form of a pair of metal posts is present in the guide, and assume that a TE_{10} mode is traveling down the guide. The posts distort both the electric and magnetic fields, which, therefore, in the vicinity of the posts can no longer have the configuration corresponding to a pure TE_{10} mode. The resulting distorted configuration can, however, be represented by a TE_{10} mode of different amplitude from that which would be present in the absence of the posts, plus superimposed higher-order modes.

It is thus seen that an irregularity transforms a portion of an original mode into new modes. This is true irrespective of the exact nature of the irregularity, which, for example, can be a bend, a twist, a constructional irregularity, etc., instead of a post. Also any arrangement for absorbing energy from the waveguide (i.e., a load termination) can in general be expected to distort the field and generate new modes unless special care is taken to avoid this result.

Suppression of Unwanted Modes. An attempt is usually made to operate waveguides so that only a single pure mode is present. In this way coupling systems and terminations can be designed on the basis of a definitely known type of field pattern. In most cases, the dominant mode is preferred because the guide then has the smallest possible dimensions, and the undesired modes can be very simply eliminated.

A dominant mode, free of higher-order modes, can be obtained by taking advantage of the fact that the dominant mode has the lowest cutoff frequency of all possible modes. Thus, by proportioning the guide so that it is large enough to transmit the dominant mode while too small to permit propagation of any other mode, the higher-order modes do not travel down the guide, but rather are confined to the region where they are generated.

In rectangular guides, mode suppression of this character is most effective when the guide is so proportioned that $a/b = 2$ in Fig. 5-1. With these proportions, there is a two-to-one frequency range over which only the dominant mode propagates (see Table 5-1, page 136). In contrast, if the guide were made square, the TE_{01} mode would have the same cutoff wavelength as the TE_{10} mode, and there would be no frequency range over which only a single mode could propagate. Because of considerations such as this, rectangular guides are practically always proportioned so that $a/b = 2$, as this ratio gives the best mode separation of all possible proportions.

Modes which are beyond cutoff, and so cannot propagate, are sometimes termed *evanescent* modes. They represent localized field distributions, i.e., induction fields, that introduce reactive effects but do not carry energy away from the point of origin as does the dominant mode. For example, if the waveguide in Fig. 5-9 is so proportioned that only the dominant mode can propagate, the end result of the field distortion introduced by the post will be equivalent to introducing an irregularity in the waveguide that causes a portion of the dominant wave to be reflected as though from a reactive load. In addition, there will be induction fields in the immediate vicinity of the irregularity that represent reactive energy obtained from the incident dominant mode. However, if in Fig. 5-9 the waveguide were made sufficiently large to permit some of the higher-order modes produced by the post to propagate in the guide, these modes would then travel away from the post, carrying energy with them that was derived from the incident dominant mode. The remaining modes, of such high order as to be unable to propagate, would still be evanescent modes, and would give rise to reactive effects.

Another method of suppressing undesired modes consists in modifying the guide in such a manner that fields of undesired modes are interfered with, while fields of the desired mode are not affected. An example of such a *mode filter* is illustrated in Fig. 5-10. Here the metal vanes do not affect the fields of the TE_{m0} modes, but do interfere with both the electric and magnetic fields of any TM or TE_{0n} mode that might be present. Thus such an arrangement is an effective means of suppressing the transverse magnetic mode in a rectangular waveguide.

An obvious means of mode suppression is to arrange matters so that as far as possible the undesired modes are never generated. This means exciting the waveguide in such a manner that the initial field configuration resembles the desired mode as much as possible, and then avoiding irregularities, including terminations, that introduce distortions in the field pattern. For example, a means of launching the waves in the guide that produces only transverse vertical electric fields that do not vary in strength in the vertical direction will not generate any TM mode, or any of the TE_{0n} series of modes. Further, if the launching system is also symmetrical with respect to the center of the long side of the rectangular guide, the only modes present will be of the TE_{m0} type, where m is odd.

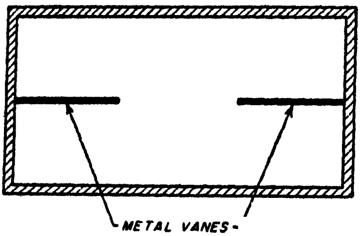


FIG. 5-10. Simple mode filter.

5-4. Physical Picture of Propagation in Rectangular Waveguides. It is possible to explain many of the properties of waveguide propagation by means of a simple physical picture. To do so, start by considering two parallel conducting planes; these planes will later define the top and bottom walls of a rectangular waveguide. A plane radio wave such as illustrated in Fig. 1-1 will propagate freely in the space between these surfaces provided the electric field is vertical. Such a wave travels with the velocity of light, and its electric and magnetic fields are everywhere in time phase. Some of the details involved are portrayed in Fig. 5-11. This wave can

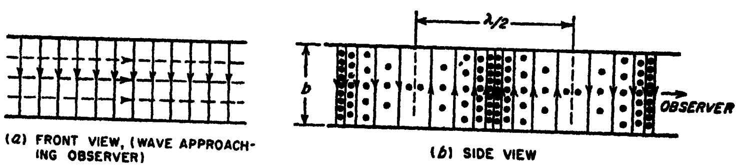


FIG. 5-11. Electric and magnetic fields of a plane radio wave that is propagating between two parallel conducting planes.

also be represented by successive crests spaced a wavelength apart, as illustrated in Fig. 5-12a, where θ is the direction of travel of the wave with respect to some reference axis. A second similar wave, differing in that the direction of travel with respect to the same reference axis is $-\theta$, is illustrated in Fig. 5-12b.

If now both waves are simultaneously present in the space between the conducting planes, one obtains the situation pictured in Fig. 5-12c. A close examination of Fig. 5-12c shows that if the two waves have equal amplitudes, then in vertical planes indicated by the heavy dotted lines, cc and dd, the electric fields of the two waves are equal and opposite and so cancel. The transverse components of the magnetic fields likewise cancel at cc and dd, causing the resultant magnetic field at these planes to be parallel to lines *cc* and *dd*. Vertical conducting sheets can accordingly be placed along *bb* and *cc* without affecting either magnetic or electric fields in any respect. These vertical conducting surfaces, together with the conducting horizontal planes, define a rectangular waveguide with conducting walls. The fields inside this guide satisfy the boundary conditions imposed by the walls, and also satisfy Maxwell's equations in the space inside the guide. The resultant field configuration obtained by adding the fields of these two plane waves that travel at angles θ and $-\theta$, respectively, is the TE_{10} mode; this is illustrated by the dotted lines in Fig. 5-12*c*, which show the resultant magnetic flux and are seen to correspond to the magnetic-flux distribution given in Fig. 5-2.

Study of Fig. 5-12*c* shows that it is now possible to consider that the fields inside the waveguide are the result of a pair of electromagnetic waves that travel back and forth between the sides of the guide, following a zigzag path as illustrated in Fig. 5-12*d*. Each time such a wave strikes the conducting side wall, it is reflected with reversal of the electric field, with an angle of reflection equal to the angle of incidence, as illustrated.

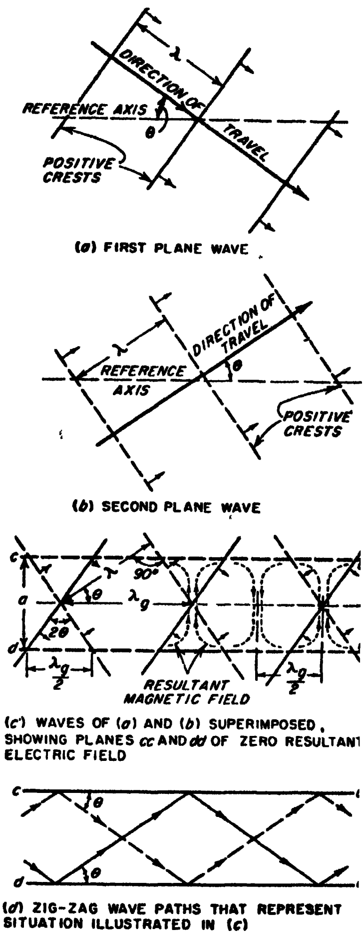


FIG. 5-12. Steps involved in building up a physical picture of propagation in a rectangular waveguide.

The guide wavelength λ_g for the situation in Fig. 5-12*c* is the distance along the axis between points in the guide where the positive crests coincide. It will be noted that the guide wavelength λ_g exceeds the free-space wavelength λ of the plane wave by an amount that will increase as θ becomes larger. Various relations follow from the geometry of Fig. 4-42*c*; thus

$$\cos \theta = \frac{\lambda}{\lambda_g} \tag{5-7a}$$

$$\tan \theta = \frac{\lambda_g/4}{a/2} = \frac{\lambda_g}{2a} \tag{5-7b}$$

Combining to eliminate θ gives

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda/2a)^2}} \tag{5-8}$$

This is equivalent to Eq. (5-3) when it is noted that $\lambda_c = 2a$.

Examination of the geometry of Fig. 5-12*c* reveals that, if the distance between successive crests is increased (i.e., free-space wavelength λ increased), then if the electric fields are to cancel along planes *cc* and *dd*, it is necessary that θ be increased. As the free-space wavelength approaches closer and closer to the cutoff wavelength, θ thus becomes increasingly large, and the zigzag path of the waves becomes increasingly transverse, as illustrated in Fig. 5-13.

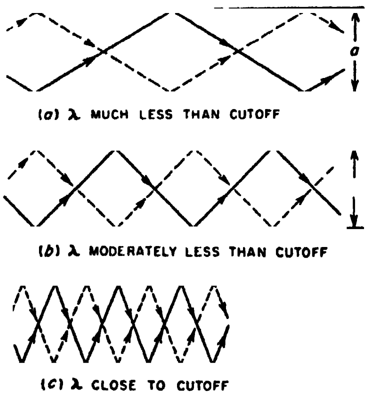


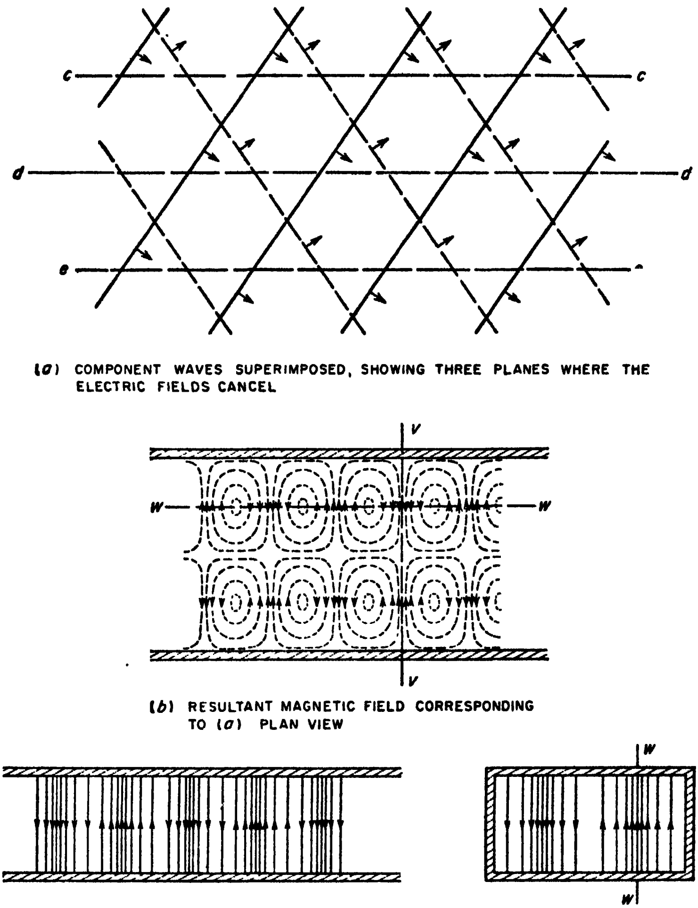
FIG. 5-13. Paths followed by a wave traveling back and forth between the sides of a waveguide for values of free-space wavelengths differing from the cutoff wavelength by various amounts.

The fact that the guide wavelength λ_g in Fig. 5-12*c* exceeds the free-space wavelength causes the phase velocity in the rectangular guide to exceed the velocity of light. At the same time, the individual waves themselves advance more slowly down the guide than the velocity of light, since the individual waves travel by a zigzag path. This rate at which the waves progress down the guide is the *group* velocity and corresponds to the rate at which a pulse of energy would travel. As the free-space wavelength approaches more closely the cutoff wavelength of the guide (i.e., as $\lambda \rightarrow 2a$), the phase velocity becomes progressively larger and the group velocity progressively less. In the limit, at the cutoff wavelength, the waves travel back and forth between the sides of the guide at right angles to the axis ($\theta = 90°$). Under these conditions nothing at all travels down the guide, so the group velocity is zero, while the phase velocity is infinite.[1]

This picture that has been developed of wave propagation in a rectangular guide can be readily extended to take into account the higher-order modes. For example, in Fig. 5-12*c*, it is apparent that there are also other vertical planes in which the electric fields of the two component waves cancel exactly; one such plane is indicated by *ee* in Fig. 5-14*a*. If

[1] An excellent discussion of the significance of group and phase velocities is given by J. A. Stratton, "Electromagnetic Theory," pp. 330–340, McGraw-Hill Book Company, Inc., New York, 1941; also see H. H. Skilling, "Electric Transmission Lines," pp. 369–373, McGraw-Hill Book Company, Inc., New York, 1951.

now vertical conducting sheets are placed at *cc* and *ee* instead of *cc* and *dd*, one again has formed a rectangular waveguide inside of which are fields (illustrated in Fig. 5-14*b* and *c*) that satisfy all of the required conditions; this particular configuration is the TE_{20} mode.



(*a*) COMPONENT WAVES SUPERIMPOSED, SHOWING THREE PLANES WHERE THE ELECTRIC FIELDS CANCEL

(*b*) RESULTANT MAGNETIC FIELD CORRESPONDING TO (*a*) PLAN VIEW

SIDE VIEW THROUGH *WW* END VIEW THROUGH *VV*

(*c*) RESULTANT ELECTRIC FIELD CORRESPONDING TO (*a*), IN VERTICAL PLANES

Fig. 5-14. Physical picture showing how the TE_{20} mode arises in a rectangular waveguide.

5-5. Circular Waveguides. It might be thought that waveguides with circular cross sections would be preferred to guides with rectangular cross sections, just as circular pipes are commonly used for carrying water and fluids in preference to rectangular pipes. However, circular waveguides have the disadvantage that there is only a very narrow range

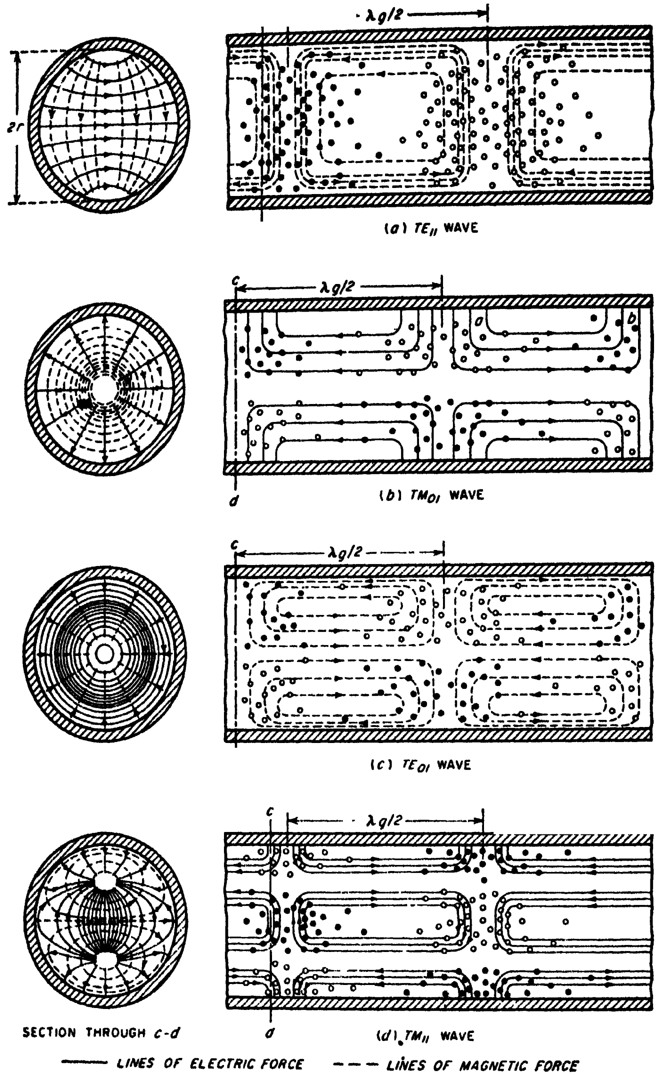


Fig. 5-15. Field configuration of the dominant TE_{11} mode, and of the first few higher-order modes in a circular waveguide.

between the cutoff wavelength of the dominant mode and the cutoff wavelength of the next higher mode. Thus the frequency range over which pure mode operation is assured is relatively limited. Also, because of its circular symmetry, the circular guide possesses no characteristic that positively prevents the plane of polarization of the wave from rotating about the guide axis as the wave travels. As a result, circular waveguides are used only under special circumstances, for example, where it is necessary to introduce a rotating joint into a waveguide system.

Field configurations for the more important circular modes are illustrated in Fig. 5-15. As with rectangular guides, these modes may be classified as transverse electric (TE) or transverse magnetic (TM), according to whether it is the electric or magnetic lines of force that lie in planes perpendicular to the axis of the guide. The different modes are designated by a double subscript system analogous to that for rectangular guides.[1]

The wavelength corresponding to cutoff for a particular mode in a circular guide is proportional to the diameter of the waveguide, with the exact relationship being given by the equation

$$\text{Cutoff wavelength} = \lambda_c = \frac{2\pi r}{\mu} \tag{5-9}$$

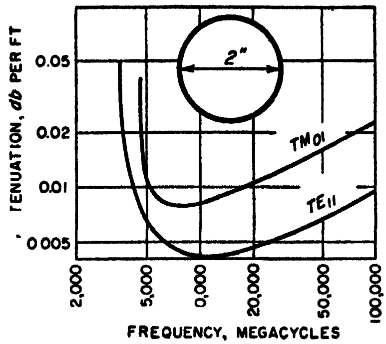


FIG. 5-16. Attenuation as a function of frequency of the dominant and first higher-order mode in a particular circular waveguide with copper walls.

where r is the guide radius and μ is a constant that depends on the order of the mode.[2] Results of Eq. (5-9) for the first few modes are tabulated in Table 5-1.

The TE_{11} circular mode (see Fig. 5-15) has the longest cutoff wavelength, and is accordingly the dominant circular mode. The next higher circular mode is the TM_{01} mode, for which the cutoff wavelength is 0.76 times that of the dominant mode. The corresponding ratio is 0.5 for the first two modes in a rectangular guide with $a/b = 2$. Thus the ratio of frequencies over which only the dominant mode will propagate is over 50 per cent greater for the rectangular guide than for the circular guide.

The guide wavelength λ_g in a circular guide is greater than the wavelength λ in free space, just as in the rectangular guide. In fact, Eq. (5-3) applies to circular as well as to rectangular guides. The velocity of phase propagation is λ_g/λ times the velocity of light in all cases.

A wave traveling down a circular guide is attenuated as a result of

[1] For example, in the TM_{om} mode, the magnetic field is circular, and m is the number of cylinders, including the boundary of the guide, to which the electric vector is normal. Rules for determining the subscripts for the various possible cases are given in "Standards on Radio Wave Propagation—Definitions of Terms Related to Guided Waves," Institute of Radio Engineers, New York, 1945.

[2] For TE_{nm} waves, μ is the mth root of the equation $J'_n(x) = 0$, and for TM_{nm} waves, it is the mth root of the relation $J_n(x) = 0$.

power dissipated in the walls by the induced wall currents, exactly as in the case of a rectangular guide. Curves of attenuation as a function of frequency are given in Fig. 5-16 for the first two modes in a particular guide. These are similar in character to the corresponding curves of Fig. 5-6 for the rectangular guide, in that the attenuation passes through a minimum at a frequency that is moderately greater than the cutoff frequency.[1]

5-6. Reflected and Incident Waves, Field Distributions, and Standing-wave Ratio in Waveguides. As indicated previously, the field configuration in the waveguide behaves in the same way as a wave on a transmission line. That is, the electric and magnetic fields associated with a particular mode, such as the TE_{10} mode, travel down the guide at the phase velocity. At the end of the guide, or at an irregularity, a reflection is produced that creates a similar field configuration traveling in the opposite direction. As in the analogous transmission line case, the reflection coefficient at a point can be defined as the ratio of the reflected to incident wave at that point in the guide.

The superposition of incident and reflected waves in a waveguide gives rise to amplitude distributions along the guide that are of exactly the same character as the voltage and current distributions encountered in transmission lines (illustrated in Fig. 4-4) provided one interprets the electric and magnetic fields of the guide as being equivalent, respectively, to the voltage and current of the transmission line. Thus a short-circuited receiver (zero voltage or zero electric field at the receiving end of the system) gives a distribution in which the resultant electric field is maximum at distances from the load corresponding to an odd number of quarter wavelengths based on λ_g, the guide wavelength. At the same time, the resultant magnetic field is maximum at the load, and at distances from the receiver corresponding to an even number of quarter wavelengths. Resistive loads of the incorrect value to absorb the incident wave completely will give partial reflections, but with the maxima and minima in the distribution occurring at the same places as in the corresponding open- and short-circuited cases. On the other hand, load impedances that have a reactive component will have the minima displaced, exactly as in the case of the transmission line.

The extent to which a reflected wave is present in a waveguide can be conveniently expressed in terms of a standing-wave ratio. As applied to a waveguide, the standing-wave ratio has the same significance and

[1] An exception to this otherwise general behavior is the TE_{01} mode, sometimes called the "smoke ring" mode, in which the attenuation decreases steadily with increasing frequency and becomes zero at infinite frequency. This result comes about through the fact that in this mode the magnetic field adjacent to the walls of the guide becomes progressively weaker as the ratio of free-space to cutoff wavelengths becomes less. In the limit, at infinite frequency, this magnetic field becomes zero, resulting in zero current induced in the walls.

usefulness as in the analogous transmission-line situation, provided that one remembers that the magnetic and electric fields in the guide correspond respectively to current and voltage in the transmission line.

Any irregularity in a waveguide will give rise to reflections and hence will establish standing waves, just as does a load impedance that is not matched to the waveguide. Thus bends, twists, joints, probes, mechanical imperfections, pieces of dielectric, etc., all give rise to reflections, the magnitude of which can be expressed in terms of the resulting standing-wave ratio.[1]

Transmission-line Equivalent of a Waveguide System. In dealing with a waveguide system possessing an irregularity, it is commonly convenient to regard the arrangement as though it were a transmission line possessing a corresponding irregularity. The characteristic impedance of this equivalent transmission line can be taken as the waveguide impedance defined in whatever manner is most convenient (see below). The impedance of the transmission-line irregularity[2] (and also the load impedance) is then assigned the value such that in relation to the characteristic impedance the resulting reflection coefficient associated with the transmission-line irregularity will be the same as the reflection coefficient actually produced in the waveguide by the irregularity. The standing-wave situation existing on the equivalent transmission line is then the same in every respect as is actually present on the waveguide; an example is given on page 149.

5-7. Impedance Relations in Waveguides. *Waveguide Impedance.* In a transmission line, one can define a characteristic impedance that is determined by the geometry of the line and which holds for all frequencies. In contrast, there are several different ways in which a "characteristic impedance" can be defined for a waveguide, and each of these definitions gives a different numerical result. In addition, the waveguide impedance for a given guide will be a function of frequency irrespective of how defined.

One commonly used approach is to define the impedance associated with a waveguide as the ratio of the transverse components of the electric to magnetic field strength. This is termed the *wave impedance;* for a guide with air dielectric it is given by the formulas

For TE waves:

$$\text{Wave impedance} = 377 \frac{\lambda_g}{\lambda} \qquad \text{ohms} \tag{5-10}$$

[1] The quantitative effects produced by bends, twists, etc., are summarized by N. Elson, Rectangular Waveguide Systems, *Wireless Eng.*, vol. 24, p. 44, February, 1947; also see Moreno, *op. cit.*, pp. 162–169.

[2] In many cases an irregularity is more satisfactorily represented by a simple T or π network, or a simple resonant circuit, than by a single circuit element. Examples of such cases are given in Figs. 5-19*c* and 5-28.

For TM waves:

$$\text{Wave impedance} = 377 \frac{\lambda}{\lambda_g} \quad \text{ohms} \tag{5-11}$$

Here λ and λ_g are the free-space and guide wavelengths, respectively. The wave impedance has the desirable feature that it is independent of the physical proportions or shape of the guide, or of the transmission mode, except in so far as these affect the guide wavelength λ_g. The concept of wave impedance is particularly useful in the study of waveguide discontinuities and loads.

Another approach is to define the impedance of a waveguide as the ratio of the *maximum* value of the transverse voltage developed across the guide to the total longitudinal current flowing in the guide walls for a traveling wave when no reflected wave is present. On this basis, the waveguide impedance Z_0 for the TE_{10} mode in an air-filled rectangular guide is

$$Z_0 = 377 \frac{\lambda_g}{\lambda} \frac{\pi}{2} \frac{b}{a} \tag{5-12}$$

This definition of waveguide impedance is useful in the design of systems for coupling waveguides to coaxial lines, such as illustrated in Fig. 5-8. It must be used with some caution, however, because in contrast with transmission lines, the fields of a guide are not uniformly distributed over the cross section.

Impedance Matching in Waveguides. Reflected waves are generally to be avoided in waveguides for exactly the same reasons that they are avoided in transmission lines. One method of achieving this result in a waveguide is to arrange matters so that the load impedance that is used will completely absorb the incident fields exactly as they arrive, so that there is nothing left over to be reflected; this corresponds to characteristic impedance termination in a transmission line. A second approach to the problem is to create a reflected wave near the load that is equal in magnitude but opposite in phase from the wave reflected by the load; in this way the two reflected waves cancel each other. Most commonly both methods of impedance matching are used simultaneously. That is, the system is initially so arranged that the load provides as good an impedance match as is possible to obtain with reasonable effort, and then what reflected wave still remains is eliminated by the use of an impedance-matching system that introduces a neutralizing reflection.

Numerous waveguide arrangements have been devised for introducing a controllable reflection. Some of these are analogous to the impedance-matching arrangements employed in transmission lines (described in Sec. 4-11), while others are unique to waveguides.

The waveguide analogue of the stub line of Fig. 4-16 is the stub guide or T section illustrated in Fig. 5-17. Two possibilities are to be distinguished.[1] At *a* the reactance at the input of the stub guide is effectively in series with the equivalent transmission line of the guide, while with the stub as in *b*, the reactance introduced by the stub is in shunt in the equivalent transmission line circuit of the guide. This is shown schematically at *c* and *d*, respectively. The magnitude of the reflection introduced by such a stub guide is controlled by the position of the short-circuiting plunger in the stub guide. The phase of the reflected wave produced by the stub is determined by the position of the stub in relation

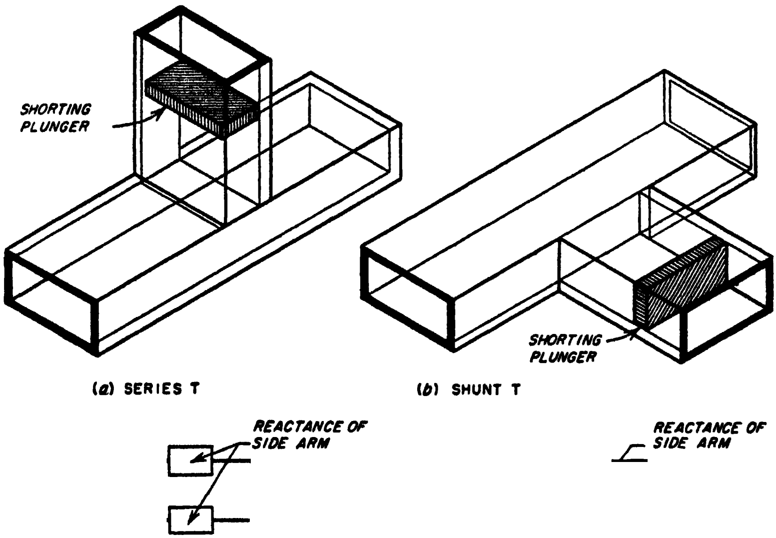


FIG. 5-17. Waveguides provided with tuning stubs in arrangements analogous to that of Fig. 4-18, together with equivalent transmission-line circuits.

to a minimum of the standing-wave pattern existing in the absence of the stub. Thus, to eliminate a reflected wave using a single stub, it is necessary to be able to vary not only the effective length of the stub, but also its distance to the load. This latter requirement makes a single stub arrangement unsatisfactory in systems that must be adjusted by trial and error, since there is no simple way that the position of the stub can be continuously varied. When trial-and-error adjustment is required, one can, however, employ two waveguide stubs spaced approximately $n\lambda_g/8$, where n is odd, to give the waveguide equivalent of the two-stub tuner of Fig. 4-18.

An alternative to the waveguide stub is an adjustable screw or probe

[1] The arrangements at *a* and *b* are often referred to as *E* and *H* stubs, respectively, because the axis of the stub is parallel to the *E* lines and *H* plane, respectively, in the main guide.

that projects into the waveguide in a direction parallel to the electric field, as illustrated in Fig. 5-18. Such an arrangement has the same effect as shunting a capacitive load across the equivalent transmission line of the waveguide, with the susceptance of this capacitive load increasing with penetration into the guide up to the point where the equivalent penetration is a quarter of a wavelength.[1] Thus the extent to which such a probe (or screw) projects into the waveguide determines the magnitude of the compensating reflection, while the position of the probe with respect to the standing-wave pattern that is to be eliminated determines the phasing of the reflected wave. When it is necessary that the axial position of the probe or screw be adjustable experimentally, this can be achieved by providing the guide with a longitudinal slot located in the middle of the broad side, as shown dotted in Fig. 5-18. As pointed out in connection with Fig. 5-5, such a slot (labeled 1 in this figure) produces a minimum of interference with the fields inside the guide, and has little tendency to radiate energy. Where it is desirable to avoid the use of a slot, one can instead employ two spaced probes in an arrangement analogous to that of Fig. 4-18.

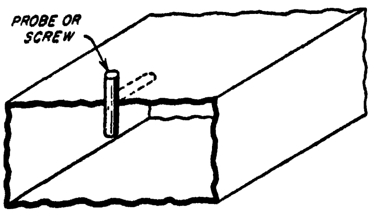


FIG. 5-18. Adjustable screw (or probe) for producing an adjustable reflection for impedance-matching purposes.

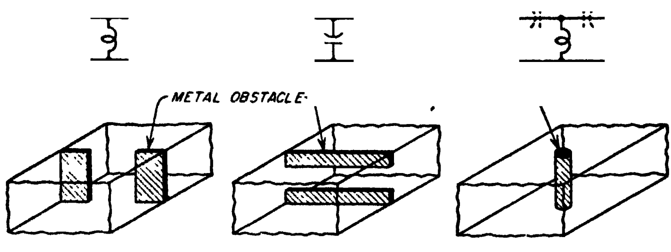


FIG. 5-19. Examples of obstacles used in waveguides to introduce reflection, together with equivalent transmission-line systems.

Another impedance-matching system consists of a thin metallic barrier, or "window," placed at right angles to the axis of the guide, as illustrated in Fig. 5-19. The arrangements at *a* and *b* introduce, respectively, inductive and capacitive shunts in the equivalent transmission-line circuit of the waveguide as shown, the magnitudes of which depend upon the size

[1] When the equivalent penetration is exactly a quarter wavelength, the probe becomes resonant. The system then acts as though a series resonant circuit of low resistance was connected in shunt with the waveguide; thus at exact resonance the probe acts as a shunt of very low resistance.

of the opening. A conducting cylindrical post going from top to bottom of a rectangular waveguide, as at *c*, produces an inductive shunt susceptance[1] having a magnitude determined by the size of the waveguide, the diameter of the post, and the post position in the transverse plane. Still another type of obstacle is illustrated in Fig. 5-9. Reflections introduced by obstacles such as illustrated in Fig. 5-19 cannot be conveniently adjusted experimentally. These arrangements are of practical use, however, in systems where a reflected wave of known and unvarying character is to be neutralized.[2]

Impedance Matching with Resistive Loads. There is the theoretical possibility of matching a resistance load directly to the waveguide in such a manner as to avoid a reflected wave; this eliminates the need of introducing a compensating reflection. Thus consider the situation illustrated in Fig. 5-20, *case a*, where the load consists of a resistance R_L connected between the top and bottom planes of the guide midway between the sides, and an odd multiple of a quarter of a guide wavelength away from a short circuit.[3] If the load resistance R_L is now equal to the waveguide impedance Z_0 as defined by Eq. (5-12), then the incident wave will be absorbed without reflection. If the resistance R_L that is to be used differs from the value called for by Eq. (5-12), one can change the guide impedance as required

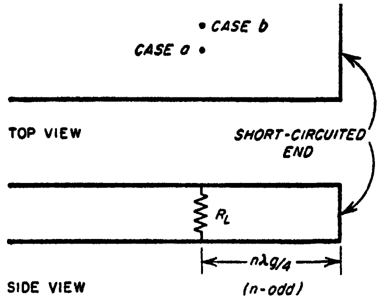


FIG. 5-20. Resistance load connected between top and bottom sides of waveguide.

[1] Actually the equivalent circuit will be a simple shunt inductance only when the diameter of the post is not more than a few per cent of the guide width. With thicker posts, the equivalent circuit includes series capacitances in addition to the shunt inductance, as shown dotted in Fig. 5-19*c*. These capacitances become larger (i.e., have lower reactance) the smaller the post diameter and have negligible reactance in the case of very thin posts The T network shown in Fig. 5-19*c* will accurately represent the behavior of even a very thick post over a wide range of frequencies when the inductance and capacitances of the equivalent section are properly chosen. Similarly, if the strips forming the windows at *a* and *b* are not thin, the obstacle is represented more accurately by a T network than by a single shunting reactance.

[2] Quantitative analysis of the structures shown in Fig. 5-19, and also of other forms of obstacles, is given in "Waveguide Handbook" (vol. 10, Radiation Laboratory Series, chap. 5, McGraw-Hill Book Company, Inc., New York, 1951; also see Moreno, *op. cit.*, chap. 9.

[3] The short circuit placed an odd multiple of a quarter of a guide wavelength distant from the load is necessary because if the guide is continued indefinitely beyond the resistance, then R_L would act merely as a shunt irregularity in the guide. Alternatively, if the guide simply ended at the point where the resistance was connected, then part of the energy of the incident wave would be radiated from the open end of the guide rather than being dissipated in the resistance.

by varying the height b of the guide, using a gradual taper as illustrated in Fig. 5-21a to avoid introducing a reflection. A variation consists in tapering only the center portion of the guide to form a ridge, as in Fig. 5-21b.

An alternative arrangement, suitable for use when the load resistance is less than the guide impedance, consists in placing R_L off center as indicated by *case b* of Fig. 5-20. This subjects R_L to less voltage than case a, and so gives an impedance-transforming action.[1] A similar effect is also obtained by making the distance from R_L to the short circuit differ from an odd multiple of a quarter wavelength.

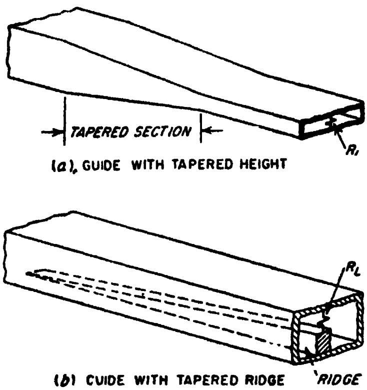


FIG. 5-21. Waveguides terminated with resistance loads.

In actual practice, arrangements of the type illustrated in Fig. 5-20 usually introduce a discontinuity capacitance. When this is the case, no adjustment of the resistance match will eliminate completely the reflected wave; to achieve such a result some additional impedance-matching adjustment, such as obtainable with a probe or a stub guide, must also be used.

Nonreflecting Loads. In systems involving waveguides it is often necessary, particularly in measurement work, to provide a termination that

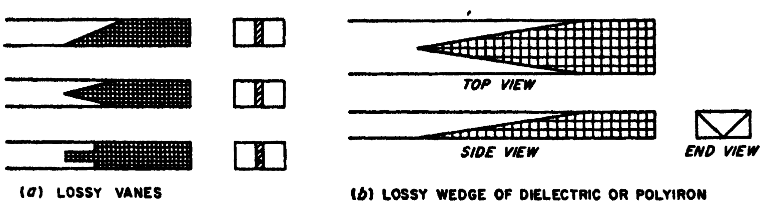


FIG. 5-22. Examples of nonreflecting terminations for waveguides.

will completely absorb any wave going down the guide, irrespective of the exact frequency of this wave, and without any adjustment being required.

This result is most conveniently achieved by absorbing the wave in a lossy section tapered so gradually as to introduce no reflection. Examples of such sections are illustrated in Fig. 5-22; these involve lossy vanes, or wedges of lossy dielectric or iron dust core material, tapered on the enter-

[1] The off-center connection causes the resistive impedance that the guide presents to the coaxial line to be less than Eq. (5-12) by the factor $\cos^2 (\pi x/a)$, where x is the distance off center and a is the guide width.

ing edge, and having a sufficient length to absorb an entering wave almost completely.[1]

5-8. Waveguide Behavior at Wavelengths Greater than Cutoff.[2] When a waveguide is excited at a wavelength greater than cutoff, the behavior is entirely different from the behavior at wavelengths less than cutoff. In particular, the electric and magnetic fields now decay exponentially with distance at a very much more rapid rate than is accounted for by energy losses in the walls. The rate of this attenuation, moreover, depends only on the ratio λ/λ_c of the free-space wavelength to the cutoff wavelength; unlike waves shorter than the cutoff wavelength the attenuation is independent of the material of the guide walls. The exact law of attenuation can be derived by application of the fundamental field equations, and is

$$\left.\begin{array}{l}\text{Attenuation in}\\ \text{db per unit length}\end{array}\right\} = \alpha = \frac{54.6}{\lambda_c}\sqrt{1 - \left(\frac{\lambda_c}{\lambda}\right)^2} \qquad (5\text{-}13a)$$

When the actual wavelength is much greater than cutoff ($\lambda >> \lambda_c$), then

$$\alpha \approx \frac{54.6}{\lambda_c} \qquad (5\text{-}13b)$$

TABLE 5-2
ATTENUATION FORMULAS FOR CUTOFF ATTENUATORS

Mode	Attenuation, db per unit length	Value of λ_c
Circular waveguides of radius r		
TE_{11}	$\frac{16.0}{r}$	$3.42r$
TM_{01}	$\frac{20.9}{r}$	$2.61r$
TE_{01}	$\frac{33.3}{r}$	$1.64r$
Rectangular guide of width a and height b		
TE_{10}	$\frac{27.3}{a}$	$2a$
TE_{11} and TM_{11}	$\frac{27.3}{a}\sqrt{1 + \left(\frac{a}{b}\right)^2}$	$\frac{2a}{\sqrt{1 + \left(\frac{a}{b}\right)^2}}$

[1] For further discussion of nonreflecting terminations for waveguides see F. E. Terman and J. M. Pettit, "Electronic Measurements," p. 639, McGraw-Hill Book Company, Inc., New York, 1952.

[2] The original paper on this subject was by Daniel E. Harnett and Nelson P. Case The Design and Testing of Multirange Receivers, *Proc. IRE*, vol. 23, p. 578, June, 1935.

Here λ is the free-space wavelength and λ_c is the cutoff wavelength, measured in the same units of length used in expressing the attenuation. Equations (5-13) apply to all modes of propagation in all types of waveguides. The resulting relation between the rate of attenuation and the guide dimensions is given in Table 5-2 for cases of particular interest.

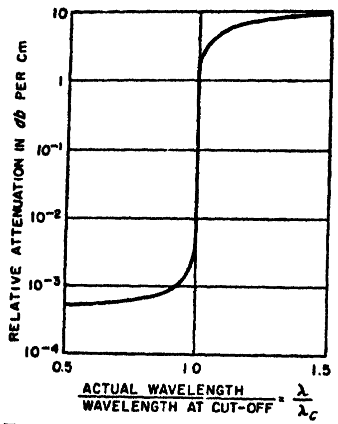


FIG. 5-23. Typical variation of attenuation in a waveguide with frequency in the vicinity of cutoff.

Frequency enters into the expression for attenuation only through the term λ_c/λ in Eq. (5-13a). When this is small, the attenuation is substantially independent of frequency.[1] As the wavelength approaches the cutoff value, the rate of attenuation will diminish in accordance with Eq. (5-13a). This is illustrated in Fig. 5-23. As cutoff is very closely approached, a rapid transition takes place, as shown, and when the wavelength is less than cutoff, the attenuation assumes the comparatively low value associated with wall losses.

Waveguides operated at wavelengths greater than cutoff, termed *waveguide attenuators*, are often used as attenuators in signal generators. The usual arrangement for this purpose, illustrated in Fig. 5-24, involves exciting the guide, which may be either circular or rectangular, with a coil, the axis of which is at right angles to the axis of the guide. The pickup system then consists of a similar coil with its axis parallel to the

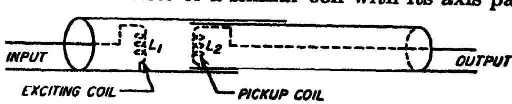


FIG. 5-24. Schematic diagram of typical waveguide attenuator.

axis of the exciting coil. Such an arrangement uses the TE_{10} mode in the rectangular case, and the TE_{11} mode when the guide is circular.[2] The

[1] This neglects the variation with frequency of the depth of current penetration into the wall. To obtain maximum accuracy when using Eqs. (5-13), the effective internal dimensions of the waveguide should be taken as extending into the walls a distance equal to one-fourth the skin depth as given by Eq. (2-10). Since this skin depth varies with frequency, the value to be assigned λ_c will likewise vary with frequency, increasing slightly as the frequency is reduced, and hence introducing a small additional cause of variation of attenuation per unit length.

[2] These are the dominant modes and are employed because they attenuate more slowly with distance than do higher-order modes. Hence when the dominant mode is initially mixed to some extent with higher-order modes, then the mode becomes increasingly pure as one goes down the guide away from the exciting coil.

output of such an attenuator is varied by adjusting the distance between the pickup coil and the exciting coil. The change in output produced by a known displacement of the pickup coil can be calculated from the waveguide dimensions, using Table 5-2 or Eqs. (5-13). The waveguide operated at wavelengths greater than cutoff hence provides a simple and reliable way of introducing known changes in the output.[1]

5-9. Miscellaneous Aspects and Properties of Waveguides. *Coupling between Coaxial Lines and Waveguides.* Numerous arrangements have been devised for coupling a coaxial transmission line to a waveguide so that power may flow from one transmission system into the other. A typical example[2] is illustrated in Fig. 5-8. As viewed by the coaxial transmission line, the waveguide in this arrangement behaves like a resistance equal to the waveguide impedance as defined by Eq. (5-12). In addition, there is a reactive effect associated with the coupling as a result of the inductance of the length of conductor extending across the waveguide, and also as a result of evanescent modes present at the junction. In order to obtain an impedance match between a coaxial line and waveguide such that power will pass from one system to the other without producing a reflected wave, it is therefore necessary not only to match the characteristic impedance of the coaxial line properly to the waveguide impedance, but in addition a compensating reactance must be introduced at the coupling point. A simple method of producing the required neutralizing effect consists in adjusting the distance c in Fig. 5-8a so that the shunt reactance observed by the coaxial line, when looking toward the short-circuited end of the waveguide, is equal and opposite to the shunt reactance associated with the coupling system.[3]

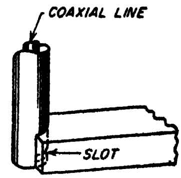


FIG. 5-25. Waveguide-to-coaxial-line coupling system based upon a slot in the outer conductor of the coaxial line.

A very different approach to the problem of coupling a waveguide to a coaxial line is illustrated in Fig. 5-25. Here a transverse slot in the outer conductor of the coaxial line allows magnetic flux to leak from the line into the waveguide. At the same time, the slot interrupts the flow of current in the outer conductor of the coaxial line, thereby creating a voltage difference across the slot that produces an electric field between the top and bottom sides of the waveguide. In this way a wave on the

[1] A more extensive discussion of waveguide attenuators is given by Terman and Pettit, *op. cit.*, p. 656.

[2] The detailed design of systems of this type is given by Seymour B. Cohn, The Design of Simple Broad-band Waveguide-to-coaxial-line Junctions, *Proc. IRE*, vol. 35. p. 920, September, 1947.

[3] The resistance that the waveguide offers the coaxial line can, when desired, be reduced by placing the coupling point off center, i.e., by making distance d in Fig. 5-8b less than half the guide width.

coaxial line introduces electric and magnetic fields into the waveguide that correspond roughly to the fields of the dominant mode. Conversely, a dominant mode traveling down the waveguide will excite a wave on the coaxial system.

Waveguide Directional Couplers. It is possible to devise directional-coupling systems involving waveguides that are analogous to the transmission-line directional-coupling arrangements discussed on page 115. The waveguide equivalent of the two-hole directional coupler of Fig. 4-24 is illustrated in Fig. 5-26. Directional coupling between a waveguide and a coaxial system is also possible. Thus, if the primary line in Fig. 4-23 is replaced by a waveguide, one obtains directional coupling between a waveguide primary and a coaxial secondary system.

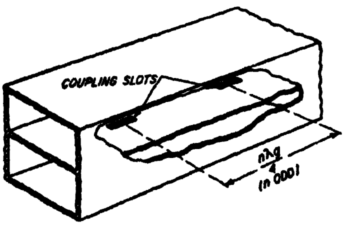


FIG. 5-26. Two-hole directional coupler for waveguide analogous to the two-hole coaxial line coupler of Fig. 4-24. The narrow slot parallel to the guide axis provides coupling that is predominantly electrostatic.

Magic T. The waveguide arrangement illustrated in Fig. 5-27*a*, termed a *magic T*, possesses many of the qualities of a bridge. Thus if the two side outlets *C* and *D* have the same length and are terminated identically, then power delivered to the system at *A* divides at the junction and flows equally to *C* and *D*, with no output whatsoever being obtained at *B*; similarly, power supplied at *B* divides between *C* and *D* and none of it appears at *A*. On the other hand, if power is delivered to the system at *A* and the terminations at *C* and *D* are not identical, then there will be an output at *B* proportional to the difference between the waves reflected at *C* and *D*.

This behavior can be explained as follows: A wave of the dominant mode traveling down *A* cannot turn the corner into *B*, because the orientation of the electric field in *A* is such that in turning into *B* the electric field would necessarily have to be parallel to the long dimension of the guide. For this field configuration, guide *B* will have a cutoff wavelength less than the wavelength of the wave arriving from *A*, provided the proportions and absolute dimensions of the system are properly chosen. The waves arriving from *A* can, however, divide and travel in directions *C* and *D*, it merely being necessary for the electric flux to turn corners into similar guides. Equal reflections from *C* and *D*, upon reaching the junction of the magic T, will divide between *A* and *B*. The portions entering *A* from *C* and *D* are in phase and so combine to give standing waves in *A*. However, the portions of these reflected waves that attempt to enter *B* do so as a result of the electric vector turning a corner as illustrated in Fig. 5-27*b*, and it will be noted that the reflections from *C* and D when entering B are of opposite polarity and so tend to cancel. This cancellation is complete if the reflected waves from C and D are identical upon arrival at the junction, in which case there is no transmission to B. However, if the reflected waves produced at C and D are not identical in magnitude and phase as they arrive at the junction, then there will be a resulting component entering B that is proportional to the vector difference of the two waves.[1,2]

The Resonant Obstacles in Waveguides.[3] When certain types of obstructions are placed in waveguides, a resonant effect is introduced that is equivalent to shunting the equivalent transmission line of the guide with either a series or shunt resonant circuit, as the case may be. The quarter-

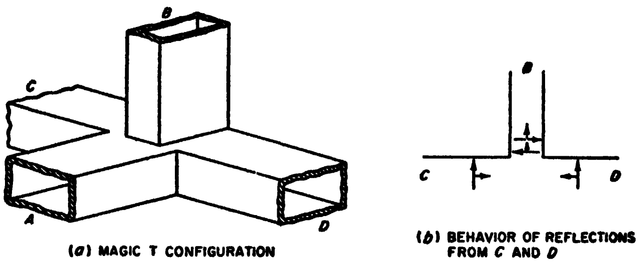


FIG. 5-27. Magic-T arrangement.

wave resonant post discussed in connection with Fig. 5-18 is an example, and is equivalent to a series resonant system connected across the guide. Another example is provided by a rectangular window in a rectangular waveguide, illustrated in Fig. 5-28*a*, which acts as a parallel resonant shunt. In contrast, an obstacle having the configuration shown in Fig. 5-28*b* acts as a series resonant shunt, the resonance occurring at a frequency determined largely by the peripheral length of the obstructing rectangular ring. Many other forms of resonant obstacles are also possible.

Obstructions that behave as shunting series-resonant systems will transmit energy rather freely at all frequencies except those in the immediate vicinity of the series resonant frequency, where the shunting impedance is so low as to reflect nearly all of the energy. In contrast,

[1] This explanation assumes that the discontinuity capacitances existing at the common junction of the magic-T configuration have been neutralized by the introduction of appropriate inductive irregularities, such as a window of the type illustrated in Fig. 5-19*a*.

[2] Another waveguide arrangement, known as the *hybrid ring*, has properties similar to those of the magic T, and can be regarded as an alternative arrangement. Various forms of the hybrid ring are described by W. A. Tyrell, Hybrid Circuits for Microwaves, *Proc. IRE*, vol. 35, p. 1294, November, 1947.

[3] Further material on this subject, particularly design information, is given by Moreno, *op. cit.*, pp. 150–157; also see "Waveguide Handbook," *op. cit.*, chap. 5.

an obstacle that acts as a parallel resonant shunt will have no effect on the transmission at the resonant frequency of the obstacle, but at all frequencies differing appreciably from the resonant frequency it will introduce a low shunting reactance that permits very little energy to be transmitted past the obstacle.

Ridged Waveguides.[1] Under some circumstances there is an advantage in providing a rectangular waveguide with a ridge analogous to the ridge shown in Fig. 5-21*b* except not tapered. This increases the cutoff wavelength, and widens the frequency range over which only the dominant mode will propagate. Thus a ridged structure has advantages when physical compactness is important, and when the guide is to be used over

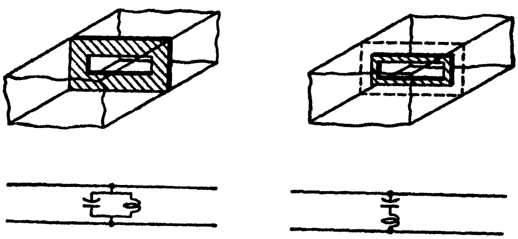


FIG. 5-28. Typical resonant obstacles, together with their equivalent transmission-line circuits.

an unusually wide frequency range. At the same time, the attenuation of the ridged structure per unit length is greater than for the corresponding rectangular guide. The impedance of the ridged structure analogous to the impedance defined in Eq. (5-12) is less than in a rectangular guide; this is sometimes an advantage when matching impedances (see Fig. 5-21*b*), or when coupling a coaxial line to a waveguide.

Comparison of Waveguides and Coaxial Transmission Lines. Waveguides find their principal use in the transmission of power at wavelengths of the order of 10 cm or less, under conditions where low attenuation or high power-carrying capacity is important. The power losses in a waveguide will be of the order of one-third as great as in a comparable coaxial line having air dielectric with supporting insulators, and the superiority is many times greater as compared with the best flexible cable. The power-carrying capacity of a waveguide as limited by flashover is likewise from three to ten times as great as that of a standard coaxial line having air dielectric with supporting insulators, and may be of the order of thousands of times as great as that of a flexible cable with solid dielectric.

[1] For further details see Seymour B. Cohn, Properties of Ridge Wave Guide, *Proc. IRE*, vol. 35, p. 783, August, 1947.

A waveguide must have a size that is a reasonable fraction of a wavelength. This is an advantage at very short wavelengths, such as 1 cm, where coaxial lines with proportions that avoid higher modes are prohibitively small. However, at wavelengths much greater than 10 cm, the waveguide becomes undesirably large and so then finds use only in special applications. Other things being equal, waveguides also have the advantage in mechanical simplicity over coaxial lines with air insulation and dielectric support.

5-10. Cavity Resonators.[1] Any space enclosed by conducting walls possesses a resonant frequency for each particular type of field configuration that can exist in the space. Resonators of this type, commonly called *cavity resonators*, find extensive use as resonant circuits at extremely high frequencies. Their behavior is analogous to that of coil-and-capacitor combinations, but for microwave frequencies cavity resonators have the advantages of reasonable dimensions, simplicity, remarkably high Q, and very high shunt impedance.

Cavity resonators can take many forms, since any enclosed surface, irrespective of how irregular its outline, forms a cavity resonator. The simplest cavity resonator is a length of circular or rectangular waveguide short-circuited at each end to form a cylinder or rectangular prism, respectively. A spherical cavity is also of interest from a theoretical point of view, although not very useful in a practical way. Cavities such as illustrated in the lower half of Fig. 5-29, in which the opposite sides are brought close together to form a reentrant structure, are of importance when an electron beam is passed through the cavity, as in klystron tubes.[2] In such arrangements, the electric field is very strong in the gap formed by the reentrant sections, thus permitting effective interaction with electrons passing across this gap.

Cavity resonators can also be derived from coaxial lines. For example, a line short-circuited at each end, as in Fig. 5-30*a*, is resonant whenever the length is a multiple of a half wavelength. Alternatively, it is possible to arrange a coaxial transmission line, as illustrated in Fig. 5-30*b*; this can be regarded as a line short-circuited at one end and open at the other end except for the localized capacitance between the center conductor and

[1] Resonant cavities were introduced to radio by W. W. Hansen, A Type of Electrical Resonator, *J. Appl. Phys*, vol. 9, p. 654, October, 1938. Useful information on properties of cavities is given by Moreno, *op. cit.*, pp. 210–241; Terman and Pettit, *op. cit.*, pp. 204–210; I. G. Wilson, C. W. Schramm, and J. P. Kinzer, High Q Resonant Cavities for Microwave Testing, *Bell System Tech. J.*, vol. 25, p. 408, July, 1946; J. P. Kinzer and I. G. Wilson, Some Results on Cylindrical Cavity Resonators, *Bell System Tech. J.*, vol. 26, p. 410, July, 1947; End Plate and Side Wall Currents in Circular Cylinder Cavity Resonator, *ibid.*, vol. 26, p. 31, January, 1947.

[2] Properties of such resonators are given by T. E. Moreno, *op. cit.*; also see W. W. Hansen and R. D. Richtmyer, On Resonators Suitable for Klystron Oscillators, *J. Appl Phys.* vol. 10, p. 189, March, 1939.

the conducting surface that closes the end of the line. It is also possible to regard the cavity of Fig. 5-30*b* as a reentrant cavity analogous to that of Fig. 5-29*f*.[1]

Modes in Cavities. As in waveguides, it is possible for many different types of field configurations, or modes, to exist in a cavity. Associated with each such mode is a resonant frequency that is determined by the particular field configuration involved and by the cavity dimensions.

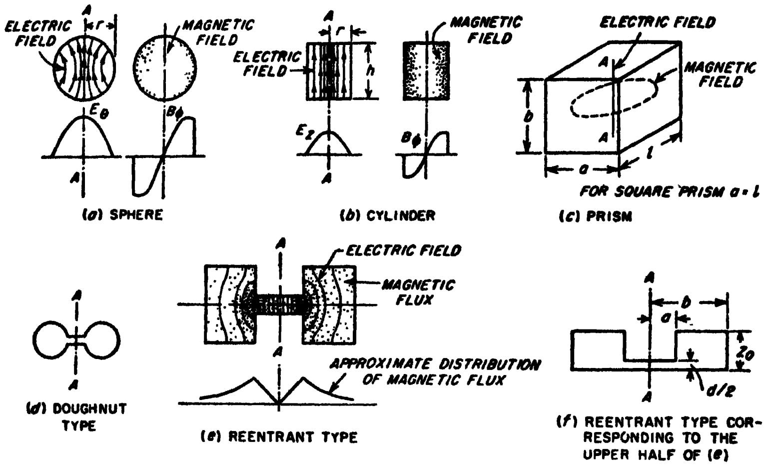


FIG. 5-29. Examples of cavity resonators. All these resonators except the prism are shown as cross sections of figures of revolution. The field distributions shown for certain of the resonators correspond to the distributions with the dominant mode of operation.

Thus each cavity resonator possesses an infinite number of resonant frequencies. As in the case of the waveguide, the lowest resonant frequency associated with a particular cavity is termed the *dominant mode*, while the remaining resonant frequencies are referred to as higher-order modes.

The cavity modes can in many cases be associated with waveguide modes. Thus in the case of the rectangular prism of Fig. 5-29*c*, a TE or TM wave traveling in the l direction will be in resonance whenever the frequency is such as to make the cavity length l a multiple of half of a guide wavelength for the mode in question. An analogous situation also

[1] Resonant lines of this type are sometimes termed hybrid lines, since as the center conductor is shortened in length compared with the length of the outer conductor (see *c* and *d* of Fig. 5-30), the behavior, including field configurations, is intermediate between that of a resonant line and that of a cylindrical cavity. The properties of coaxial cavity resonators are discussed by W. L. Barrow and W. W. Mieher, Natural Oscillations of Electrical Cavity Resonators, *Proc. IRE*, vol. 28, p. 184, April, 1940; also see W. W. Hansen, On the Resonant Frequency of Closed Concentric Lines, *J. Appl. Phys.*, vol. 10, p. 38, January, 1939.

exists with cylindrical cavities. However, in the case of the cylindrical cavity, it happens that the dominant mode corresponds to the field configuration illustrated in Fig. 5-29*b*, for which there is no waveguide counterpart. In contrast, the dominant mode for the rectangular prism corresponds to the TE_{10} waveguide mode traveling along the axis that is longest when measured in guide wavelengths. In reentrant cavities, the dominant mode corresponds to a field configuration of the type illustrated in Fig. 5-29*e*; here the electric field is most intense in the gap.

Modes in a cavity are classified as transverse electric (TE) or transverse magnetic (TM) modes, corresponding as far as possible to the analogous waveguide modes. The particular mode of any such class is then commonly designated by three subscripts. Thus the field configuration

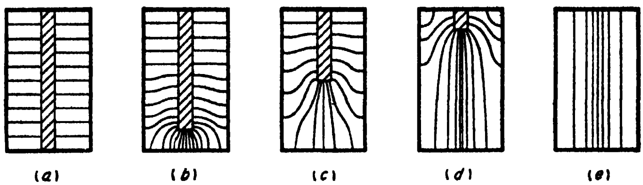


FIG. 5-30. Transition from concentric line to cylindrical cavity, showing electric fields for various intermediate or hybrid cases.

shown in Fig. 5-29*b* is the TM_{010} mode. Here TM denotes that the magnetic field lies in planes transverse to the axis of the cylinder, while the first and third subscripts denote, respectively, that the variation of the magnetic field is zero with radial direction and with position along the axis, and the second subscript indicates that there is one-half cycle of variation in the field along a radial line passing from one edge of the cylinder to the other edge. Again, the field configuration indicated in Fig. 5-29*c* is the TE_{101} mode, meaning that the electric field is transverse to an axis in the l direction and that the variation of the electric field is one-half cycle, zero, and one-half cycle in the a, b, and l directions, respectively.

A cavity resonator possesses many more modes than does the corresponding waveguide. For example, in the rectangular prism of Fig. 5-29*c*, there are an infinite number of TE_{10n} modes for each of the three axes of the prism. Thus a triple infinity of modes exists in the rectangular prism corresponding to the single infinity of TE_{n0} waveguide modes. As a result, at frequencies appreciably greater than that corresponding to resonance at the dominant mode, it is found that the resonant frequencies of cavities will be extremely closely spaced. This results in an impossible situation if one wishes to obtain pure mode operation; at the same time, it is an advantage if one desires to make it as easy as possible for the cavity to resonate with an arbitrary exciting frequency.

Resonant Frequency of Cavity Resonators. A resonant frequency of a cavity resonator corresponds to a possible solution of Maxwell's equations for the electric and magnetic fields within the resonator. The resonant frequencies (or wavelengths) can be calculated mathematically for geometrical shapes such as spheres, cylinders, and rectangular prisms and some idealized forms of reentrant sections. Formulas for the resonant wavelength of the dominant mode are given in Table 5-3 for spheres, cylinders, and square prisms. In the case of prisms it will be noted that the length l corresponds to $\lambda_g/2$ for the corresponding TE_{10} waveguide mode.[1]

The resonant wavelength is proportional in all cases to the size of the resonator; i.e., if all dimensions are doubled, the wavelength corresponding to resonance will likewise be doubled. This fact simplifies the construction of resonators of shapes that cannot be calculated. To obtain a resonator operating exactly at a desired frequency, one first constructs a resonator of convenient size and of the desired proportions and measures the resulting resonant wavelength. The ratio of the desired resonant wavelength to this wavelength gives a scale factor that is applied to every dimension of the test model to obtain the dimensions of the desired resonator.

The resonant frequency of a cavity resonator can be changed by altering the mechanical dimensions, by coupling reactance into the resonator, or by means of a copper paddle. Small changes in mechanical dimensions can be achieved by flexing walls, while large changes require some type of sliding member. Reactance can be coupled into the resonator through a coupling loop in the manner discussed below, thus affecting the resonant frequency. A copper paddle placed inside the resonator will affect the normal distribution of flux and tend to alter the resonant frequency by an amount that can be controlled by the orientation of the paddle.

Q of Cavity Resonators. The Q of a cavity resonator has the same significance as for an ordinary resonant circuit. It can be defined on the basis that when the response has dropped to 70.7 per cent of the response at resonance, the cycles off resonance are the resonant frequency divided by $2Q$ (see Rule 1, page 49). In the case of cavity resonators, it is also sometimes convenient to base the definition of Q upon Eq. (3-1), namely,

$$Q = 2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}} \tag{5-14}$$

The energy stored is proportional to the square of the magnetic flux density integrated throughout the volume of the resonator, while the energy lost per cycle in the walls is proportional to the skin depth and to the square of the magnetic flux density integrated over the surface of the

[1] Design data for reentrant cavities of the type illustrated in Fig. 5-29*d* and *e* are given by Moreno, *op. cit.*, pp. 230–238.

cavity. Thus, to obtain high Q, the resonator should have a large ratio of volume to surface area, since it is the volume that stores energy and it is the surface area that dissipates energy. As a consequence, resonators such as spheres, cylinders, and prisms can in general be expected to have higher Q's than corresponding resonators with pronounced reentrant sections.

TABLE 5-3
PROPERTIES OF CAVITY RESONATORS FOR DOMINANT MODE

Type of cavity	Sphere	Cylinder	Square prism
Figure illustrating notation	5-29*a*	5-29*b*	5-29*c*
Wavelength λ_0 at resonance	$2.28r$	$2.61r$	$1.414a$
Q	$0.318\frac{\lambda_0}{\delta}$	$0.383\frac{1}{1+(r/h)}\frac{\lambda_0}{\delta}$	$0.353\frac{1}{1+(a/2b)}\frac{\lambda_0}{\delta}$
Shunt impedance across AA at resonance	104.4	$72\frac{h}{r}\frac{1}{1+(r/h)}\frac{\lambda_0}{\delta}$	$120\frac{b}{a}\frac{\ }{1+(a/2b)}\frac{\ }{\delta}$

All dimensions are in centimeters.
δ = skin depth as defined by Eq. (2-10)
$= 6.62/\sqrt{f}$ cm for copper, where f is in cycles

Quantitative analysis leads to the formulas given in Table 5-3 for the Q of the dominant mode of spheres, cylinders, and square prisms. Some typical values of Q obtainable in practical cavity resonators are given in Table 5-4. It will be noted that the values are extremely high compared with those encountered in ordinary resonant circuits (e.g., Fig. 2-16). This is true even in the case of the reentrant cavity.

TABLE 5-4
PROPERTIES OF TYPICAL CAVITY RESONATORS WHEN OPERATING IN THE DOMINANT MODE

Resonator	Dimensions, cm	Resonant wave-length λ_0, cm	Q (copper walls)	Shunt resistance, ohms (copper walls)
Sphere	$r = 5$	11.4	28,000	9.7×10^6
Cylinder	$r = h/2 = 5$	13.0	24,000	9.1×10^6
Square prism (cube)	$a = b = l = 10$	14.1	23,000	7.8×10^6
Reentrant (Fig. 5-29*f*)	$a = 0.81$ $b = 1.69$ $z_0 = 1.82$ $d/2 = 0.20$	12.8 (approx.)	4,000 (approx.)	0.17×10^6 (approx.)

The Q of resonators of the same proportions but of different size will be proportional to the square root of the resonant wavelength. This arises from the fact that, whereas the ratio of volume to wall surface is proportional to a resonant wavelength, the skin depth (and hence the energy dissipation per unit of surface) is proportional to the square root of the wavelength.

Shunt Impedance of Cavity Resonators. The shunt impedance of a cavity resonator between two surfaces, such as those intersected by the axis AA in Fig. 5-29, can be defined as the square of the line integral of voltage along a path such as AA divided by the power loss in the resonator when excited to give the voltage used in the line integration. This impedance corresponds to the parallel resonant impedance of a tuned circuit, and at resonance becomes a resistance termed the shunt resistance of the resonator.

The shunt resistance obtained with spheres, cylinders, and square prisms operating in the dominant mode can be calculated from the formulas given in Table 5-3. Values of shunt resistance for the dominant mode in several typical cases are given in Table 5-4, and are seen to be very large compared with the shunt resistances obtainable with ordinary resonant circuits. It is further to be noted that although the shunt resistance of the reentrant cavity is much less than that of the other cavities, this impedance is developed across such a short distance that the impedance per unit length is of the same order of magnitude as the maximum value obtainable with other geometries.

5-11. Coupling to Cavity Resonators. To make use of a cavity resonator it must be coupled in some manner to a transmission line or

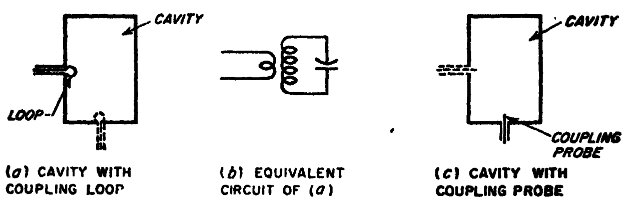


Fig. 5-31. Loop and probe coupling to cavity resonator.

waveguide. One means of accomplishing this is to employ a small loop so oriented as to link with magnetic flux lines existing in the desired mode of operation, as illustrated in Fig. 5-31a. A current passed through such a loop will then excite oscillations of this mode; conversely, oscillations existing in the resonator will induce a voltage in the coupling loop. The combination of the coupling loop and cavity resonator is equivalent to the inductively coupled system of Fig. 5-31b. In such a system, the ratio of the impedance that the cavity couples into the loop to the shunt resistance of the cavity resonator is equal to the square of the ratio of the coupled flux to the total magnetic flux lying to one side of the cylinder axis.[1] The

[1] When the plane of the coupling loop is at right angles to the direction of the flux lines, the loop area is in the most favorable position for enclosing magnetic flux; then if the loop is located at a position where the magnetic flux density approximates the

magnitude of the magnetic coupling can be readily controlled by the orientation of the loop, and its location with respect to the magnetic field. Thus the coupling is reduced to zero when the plane of the loop is rotated so that it is parallel to the magnetic flux. Also the coupling will be low if the loop is placed at a point of low magnetic flux density; thus a loop near the vertical axis, as shown dotted in Fig. 5-31*a*, will have little coupling to the dominant mode.

Coupling to a cavity can also be achieved by means of a probe as illustrated in Fig. 5-31*c*. Here the electric flux of the desired mode terminates on the probe, inducing a current in it; conversely a voltage applied to the probe produces electric fields inside the cavity that excite oscillations. This is thus a form of capacitive coupling, the magnitude of which is determined (1) by the surface that the probe exposes to the electric field of oscillations of the desired mode and (2) by the intensity of the electric field at the position of the probe. Thus maximum coupling is obtained in a cylindrical cavity operating in the TM_{010} mode when the probe is located on the axis as shown; the coupling to this mode will be zero if the probe projects into the cavity from the side wall instead of the end (dotted probe in Fig. 5-31*c*).

Still another method of coupling to a cavity is by means of a hole or slot. The principles involved in this situation are the same as in the corresponding waveguide case, and are discussed in detail on page 133.

PROBLEMS AND EXERCISES

5-1. Sketch fields corresponding to the side view in Fig. 5-2, for three successive values of time each differing by one-quarter of a cycle. Show the three cases one above the other.

5-2. Sketch field distributions similar to those of Fig. 5-2, for $\lambda = 1.5a$, being careful to show λ_g and a to scale.

5-3. A particular rectangular waveguide has a width of 2 in. and a height of 1 in. What is the lowest frequency wave that will be transmitted by this waveguide?

5-4. Calculate and plot the ratio of phase shift per unit length in a rectangular waveguide (dominant mode) to the phase shift per unit length in a coaxial transmission line having air dielectric, as the dimension a of the waveguide is varied from 0.55λ to λ.

5-5. A wave having a frequency of 10,000 Mc travels down a rectangular guide for which dimension $a = 2$ cm. Calculate the value of β per cm, and compare the result with the value of β that would be obtained at the same frequency on an air-filled coaxial line.

5-6. What is the ratio v_p/c at a frequency such that the guide width a is exactly $\lambda_g/2$?

---

average flux density in the cavity, one has to a rough approximation:

$$\left.\begin{matrix}\text{Impedance coupled}\\ \text{into loop}\end{matrix}\right\} = \left(\frac{\text{area of loop}}{\text{half of cross-sectional area of cavity}}\right)\left(\begin{matrix}\text{shunt impedance}\\ \text{of cavity}\end{matrix}\right) \quad (5\text{-}15)$$

5-7. Draw curves similar to those of Fig. 5-4, but for $a = 0.55\lambda$ and $a = 0.7\lambda$. Be careful to show a and λ_g to scale.

5-8. In a waveguide operating at a wavelength of 3 cm, calculate the depth in the copper walls at which the current density is reduced to 0.0001 of the density at the inner surface of the walls.

5-9. In the waveguide of Fig. 5-5, discuss how the distribution of the current flowing in the walls will change at times differing by (*a*) one-half cycle, and (*b*) one-quarter cycle.

5-10. In Fig. 5-5, what would be the consequences of making holes 1 and 4 round instead of rectangular, assuming that the area of the hole is the same in each case?

5-11. In Fig. 5-5, what effects are produced on the electric coupling by making slot 1 half as long and twice as wide, thus keeping the area of the opening unchanged?

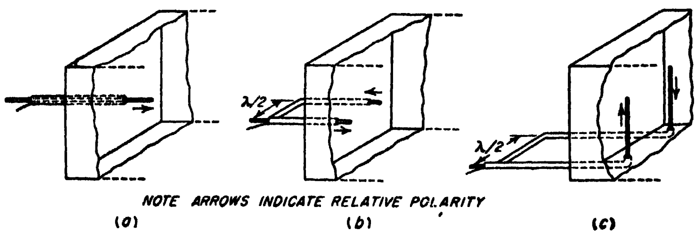
5-12. An incident TE_{10} wave of 5000 Mc travels down the guide of Fig. 5-6. How far must it go before the amplitude is reduced to 70.7 per cent of the initial amplitude?

5-13. A rectangular waveguide has dimensions 2.5 by 5 cm. Determine λ_g, β, and phase velocity at a wavelength of 4.5 cm for the dominant mode and the first higher-order mode, and tabulate results side by side.

5-14. What are the lowest frequencies for which the waveguide of Prob. 5-13 will transmit (*a*) the dominant mode, and (*b*) the first higher-order mode?

5-15. A rectangular waveguide is 2 by 3 cm. What are the cutoff wavelengths for the dominant and the first two higher-order modes?

5-16. What higher-order modes will tend to be excited in a waveguide by the coaxial line exciting systems illustrated in the attached figure?



PROB. 5-16

5-17. Suggest an arrangement involving a coaxial line terminating in a loop for exciting the TM_{11} mode in a rectangular waveguide.

5-18. In a rectangular guide in which $a = 4$ cm, calculate and plot cutoff wavelength as a function of b/a for $b/a = 0$ to $b/a = 1$, for TE_{10}, TE_{01}, TE_{20}, TE_{11}, and TM_{11} modes.

5-19. Which of the following modes will be unaffected by the posts of Fig. 5-9 (assuming the posts are located midway between the vertical sides): TE_{10}, TE_{01}, TE_{20}, TM_{11}, TM_{21}?

5-20. Draw a diagram similar to Fig. 5-12, but assume $\theta = 70°$. Be sure to mark the distances representing λ and λ_g. Compare the ratio λ_g/λ_c with the ratio for Fig. 5-12, and show that the result is consistent with Fig. 5-3.

5-21. Explain how Fig. 5-14 is consistent with the fact, deduced from Eq. (5-6), that with a given guide, the cutoff wavelength for the TE_{20} mode will always be exactly twice the cutoff wavelength for the TE_{10} mode.

5-22. Sketch curves for the TE_{11} mode, analogous to those given in the right-hand part of Fig. 5-15*a*, for a frequency of 3000 Mc in a circular waveguide when the *diameter* of the guide is (*a*) 6.3 cm, and (*b*) 8 cm. Draw the curves to full-scale size, and be careful to show λ_g correctly.

5-23. In a circular waveguide in which the radius is 1.5 in. ($r = 3.8$ cm) calculate the value of β and the phase velocity for the dominant mode at a wavelength of 10 cm.

5-24. One has available in a wall a circular hole of 2 in. *diameter* through which a waveguide is to be passed. If it is desired to obtain the longest possible cutoff wavelength, what are the relative merits of the following guides: (*a*) circular; (*b*) rectangular, $a = b$; (*c*) rectangular, $b/a = 0.5$; and (*d*) rectangular, $b/a \rightarrow 0$?

5-25. A long, narrow longitudinal slot is to be cut in the wall of a circular waveguide. Assuming the fields in the guide are as shown in Fig. 5-15*a*, where should this slot be located on the circumference of the guide if it is desired that the slot provide (*a*) coupling to the electric field, but not to the magnetic field inside the guide, and (*b*) magnetic coupling but no electric coupling?

5-26. Plot the voltage (i.e., electric field) distribution of the standing-wave pattern as a function of position from 0 to 20 cm from the receiver for a rectangular waveguide carrying a wave having a free-space wavelength of 10 cm, when the receiving end of the waveguide is short-circuited and the waveguide dimensions are 6 by 3 cm. Show the position of the minima and maxima accurately. Neglect attenuation.

5-27. The rectangular waveguide illustrated in Fig. 5-6 carries a TE_{10} wave of 5000 Mc. If the standing-wave ratio produced at the load end of the waveguide is 2, what will be the standing-wave ratio 100 ft from the load end of the line?

5-28. In a 6 by 3 cm rectangular guide, calculate and plot the waveguide impedance Z_0, as a function of frequency, from cutoff to twice the cutoff frequency.

5-29. How should a longitudinal vane projecting radially inward from the side of a circular waveguide carrying the TE_{11} mode be arranged to serve as (*a*) a mode filter, and (*b*) as an impedance-matching device?

5-30. Explain why in Fig. 5-18 a probe projecting into the guide from the side, with its axis horizontal, will be of no assistance in impedance matching for the TE_{10} mode, but would be useful in the case of TE_{01}, TE_{11}, and TM_{11} modes.

5-31. The power transmitted down a rectangular waveguide is to be delivered to a 50 ohm load resistance that is connected between the top and bottom sides of the guide, and matched by a tapered section, as in Fig. 5-21*a*. If the guide on the input side of the taper is 2.5 by 1.25 in. and the frequency is 3000 Mc, then what is the required height on the load side of the taper?

5-32. The 50-ohm load in Prob. 5-31 is matched to the guide by being placed off center, as in case *b* in Fig. 5-20, instead of by tapering the guide. How far to the side of the center line should the load resistance be placed?

5-33. In a particular rectangular waveguide attenuator based upon the TE_{10} mode, it is desired that the attenuation be exactly 10 db per in. Determine the width that the waveguide must have, assuming that the wavelength is many times the waveguide width.

5-34. In a circular waveguide attenuator, it is found that at a particular distance from the source of excitation there is an undesired TE_{01} mode present which is 30 db weaker than the desired TE_{11} mode. If the fields are now examined at a position where the attenuation to the TE_{11} mode is increased by 48 db, how strong is the undesired TE_{01} mode output compared with the TE_{11} output?

5-35. *a.* Repeat Prob. 5-34, but assume that the modes are interchanged; i.e., assume that initially the TE_{01} mode (which is now the desired mode) is 30 db stronger than the TE_{11} mode.

b. Explain why in this case the attentuation in decibels per inch will be different for large values of attenuation as compared with small values of attenuation.

5-36. In a circular waveguide attenuator using the TE_{11} mode, what will be the effect of rotating the pickup coil 90° about the axis of the guide?

5-37. A coaxial line is to be coupled to a waveguide in the manner illustrated in Fig. 5-8. If a good impedance match is desired, show that the guide cannot have the proportion $a/b = 2.0$ if the characteristic impedance of the coaxial line is in the range 50 to 100 ohms.

5-38. If the slot in Fig. 5-25 were replaced by a round hole aligned with the center axis of the guide, what mode would be excited in the waveguide by energy in the coaxial line?

5-39. Suggest a means by which a coaxial line could be coupled to a circular waveguide in such a manner as to excite the TE_{11} mode.

5-40. Describe an experimental means which could be used to measure the terminating impedance actually existing on a waveguide without removing this unknown terminating impedance from its guide, and involving a magic-T junction and a calibrated adjustable terminating impedance.

5-41. A waveguide system possesses an obstacle, the exact nature of which is not known, although the position of the obstacle is. Explain how, by the aid of standing-wave measurements, one can determine whether the obstacle is inductive, capacitive, series resonant, or shunt resonant.

5-42. Sketch curves showing qualitatively how the electric and magnetic fields are distributed in a cylindrical cavity resonator operating in the TE_{11} waveguide mode, under conditions where the cavity is a half guide-wavelength long.

5-43. Derive the formula given in Table 5-3 for the wavelength of a square prism type of cavity from the properties of the TE_{10} mode in a rectangular waveguide.

5-44. In a cavity that is a rectangular prism operating in the TE_{10} waveguide mode, it is found that the resonant frequency is independent of the dimension b in Fig. 5-29c. Explain how this is consistent with waveguide theory.

5-45. Derive a formula for the resonant frequency of a cylindrical cavity formed by a section of length h of the guide shown in Fig. 5-15a, short-circuited at both ends, and operating in the TE_{11} waveguide mode in such a manner that one-half cycle of field variation occurs in the h direction.

5-46. Show that, when a cylindrical cavity is operated in the waveguide TE_{11} mode, the resonant frequency depends on both the radius and length of the cavity.

5-47. A particular cavity with copper walls is found to have a Q of 10,000. The walls are then plated with a material having a resistivity seven times that of copper. What value will the Q then have, assuming that the plating is relatively thick?

5-48. A cylindrical cavity has a radius of 2 in. and is 6 in. long. Calculate the resonant frequency for the mode illustrated in Fig. 5-29b (the dominant mode), the circuit Q, and the shunt impedance, assuming copper walls. Tabulate the results.

5-49. A sphere, cylinder, and square prism (cube) are all so proportioned as to have the same resonant wavelength of 12.8 cm. Calculate Q, shunt impedance, and shunt impedance per unit length of shunt path, for each case. Tabulate the results, and include in the table the corresponding results from Table 5-4 for the reentrant cavity of Fig. 5-29f. Also give in the tabulation the largest linear dimension (i.e., diameter, or length of side) for each resonator.

5-50. A coupling loop 1 cm in diameter is inserted in the cylindrical cavity resonator of Table 5-4, as in Fig. 5-31a. Calculate the approximate value of the impedance that the resonator will couple into this loop at resonance, considering the resonator as a secondary, and the loop as a primary.

5-51. *a.* Describe how to locate a probe so as to couple to the TE_{101} cavity mode illustrated in Fig. 5-29c.

b. How could a loop be arranged to couple to the same mode? Give both the location of the loop and the required orientation of its plane.