RADIO COMMUNICATION SERIES
BEVERLY DUDLEY, CONSULTING EDITOR

ELECTRICAL TRANSMISSION IN STEADY STATE

RADIO COMMUNICATION SERIES

BEVERLY DUDLEY · *Consulting Editor*

Bond—
RADIO DIRECTION FINDERS

Frank—
PULSED LINEAR NETWORKS

Hund—
FREQUENCY MODULATION

King—
ELECTROMAGNETIC ENGINEERING
I. FUNDAMENTALS

Preisman—
GRAPHICAL CONSTRUCTIONS FOR VACUUM TUBE CIRCUITS

Selgin—
ELECTRICAL TRANSMISSION IN STEADY STATE

ELECTRICAL TRANSMISSION IN STEADY STATE

BY

PAUL J. SELGIN

Research Engineer, Farnsworth Television and Radio Corporation; Formerly Instructor, Polytechnic Institute of Brooklyn

First Edition

New York *London*

McGRAW-HILL BOOK COMPANY, INC.

1946

ELECTRICAL TRANSMISSION IN STEADY STATE

PRINTED IN THE UNITED STATES OF AMERICA

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

This book grew out of a set of mimeographed notes prepared by the author for a series of lecture courses, offered at the Polytechnic Institute of Brooklyn in 1942 and 1943 under the sponsorship of the War Training Program.

The notes, and the lectures, were intended chiefly as background material supplementing the practical information that the students (mostly engaged in industrial work) already had in communications and related subjects. There was no attempt to cover these subjects comprehensively.

The same may be said of this book, although it bears but little resemblance to the original notes and has an altogether different character. The book aims primarily at broadening and strengthening the foundations for a superstructure of technological knowledge which the reader may be about to acquire or may have acquired through the exercise of his profession.

It is fairly well agreed among teachers of Communications and Radio Engineering that students of these subjects generally approach them without adequate theoretical preparation. Yet it may be unwise to offer more background theory to the student not yet acquainted with the practical aspects of the art, because of his natural desire to get away from abstract ideas and absorb factual information of a descriptive character.

Later, however—perhaps in the senior year, or in graduate school, or even after some years of industrial experience—the same student or engineer may well profit from a thorough overhaul of his knowledge of fundamental theory, especially if, at the same time, he is shown how to use modern timesaving methods of analysis and computation, which may help him in his work. This going-over need not "start from scratch," as the concepts of elementary alternating-current theory, for example, will be familiar to such a reader: these are the concepts that sink in and become second nature through constant reference, in contrast to others that are apt to fade into a dim background.

What is needed perhaps as much as anything else is a fresh slant, showing how many different ideas, seemingly unrelated because acquired in connection with totally different problems, are in fact closely associated.

If it is to meet these requirements, a book such as this must be built upon a solid mathematical framework and cannot be "light reading." Yet it must remain accessible to readers of average mathematical preparation. When unfamiliar branches of mathematics must be called upon, these should be adequately introduced, taking advantage of the fact that the physical problem often throws as much light on the method used in its analysis, as it derives from the analysis itself. It is hoped, for example, that the reader may become familiar with the fundamentals of complex-function theory through its frequent applications to network and impedance transformation theory as carried out in the book.

Many recent books, essentially different from this in character and purpose, devote much attention to subjects requiring a working knowledge of vector analysis, such as the general solution of dynamic field problems in three dimensions, usually as a preliminary to the treatment of radiation and wave guides.

In compiling this book it was felt that its main objective would not have been furthered by the inclusion of these subjects, in view of the existence of excellent recent treatments, and because radiation and wave guides are not often among the subjects that require strengthening or integrating, but are more likely to be approached as totally new subjects. On the other hand, the prerequisite of vector analysis may be expected to constitute a serious obstacle. Maxwell's equations are applied in this book to circuit and line analysis without recourse to the vector-analytical form: yet the transition from the quasi-static to the dynamic state is emphasized, thus laying the groundwork for the next step, namely, the expression of Maxwell's equations in differential form and their application to radiation theory.

It may be objected that other subjects of somewhat limited interest, and likely to be new to the reader, are treated here—for example, the multisection transformer, generalized selectivity, and the exponential line. These discussions are offered as examples of the more advanced type of problem which may be attacked with the help of the methods presented. It was also felt that some of this material, not yet covered in the literature, may be of use to practicing engineers.

One of the guiding principles of this book is that essential ideas require repeated presentations of increasing thoroughness, to be fully assimilated. For this reason, the book will appear to go over the same ground more than once on occasion; sometimes, a subject will be briefly mentioned, under the assumption that the reader does not require formal introduction; later, the same point may undergo a more thorough treatment.

The author acknowledges his indebtedness to Professors Robin Beach, Frank Canovaciol, George Hoadley, and Ernst Weber, and to Dr. William Maclean of the Brooklyn Polytechnic Institute, for their helpful advice and suggestions. To Mr. Lewis Winner, editor of *Communications*, the author is particularly grateful for his valuable assistance in the preparation of the manuscript and illustrations.

PAUL J. SELGIN.

FORT WAYNE, IND.,
March, 1946.

CONTENTS

ELECTRICAL TRANSMISSION IN STEADY STATE

CHAPTER I

INTRODUCING THE FOUR-TERMINAL NETWORK AS A CIRCUIT ELEMENT: SOME GENERAL PRINCIPLES

1.1. Circuit, Field, and Network Problems. Engineering problems are usually solved by considering, in place of the true physical situation, an ideal one about which relatively simple statements, capable of mathematical expression, can be made.

Electrical engineering problems lend themselves to this process of idealization to a varying degree and may be distinguished accordingly.

Electrical systems are frequently idealized in that they are considered as a combination of *circuit branches*, or *two-poles*. The branches are interconnected by their terminals (two for each). Each branch has a *voltage* across it and a *current* through it, both uniquely defined (Sec. 12.7). The branches are frequently assumed to be *linear* (Sec. 1.7), in which case branch voltage and branch current are related by a linear differential equation—the *branch equation* (Sec. 12.7). Certain fundamental properties of voltage and current, following directly from their definition, impose conditions, known as *Kirchhoff's* laws (Sec. 12.1), which, added to the branch equations, assign all voltages and currents in terms of any one.

The theory built upon these premises is commonly known as *circuit theory*, and we may class under the heading of *circuit problems* those which may be adequately handled by this theory. Generally speaking, circuit theory ceases to be valid when the electrical quantities change very rapidly with time; hence it cannot accurately handle very high frequency transmission problems. Often circuit theory, although valid, is not the most convenient method of approach, in the sense that the branch is not the most convenient unit into which the system may be resolved.

At the other end of the scale, we find situations which may only be handled, if at all, by direct application of the laws of electromagnetism to each volume element of the physical system. In their treatment, voltage and current may not be used as variables because they are no longer uniquely defined (Chap. XII). The field vectors **E** and **H, D** and **B,** used only implicitly in circuit problems, now appear as the variables. The fundamental laws of electromagnetism, *Maxwell's* equations (Sec. 12.2), expressed in differential form, mutually relate these variables, and their integration solves the problem.

This approach idealizes only with regard to the intimate structure of matter and occasionally simplifies the geometry of the problem. Its limitations arise not from restricting assumptions but from analytical difficulties.

Problems requiring this type of approach are known as *field problems;* many of them, requiring the field approach because of the high value of frequency, go under the heading of *microwave problems.*

In addition to circuit and field problems, a third group of problems is subject to yet another method of approach. Problems of this group generally arise in the study of complex or extensive systems. They are solved by dividing the system into units or parts, which in turn are individually and separately subject to either the circuit or the field approach but are often regarded as independent functional *wholes* characterized by numerical *constants* (Sec. 2.2). In the absence of any other widely accepted, precise designation, the term *network problems* will be used for this group.

In the present treatment, the term *network* will refer to *any part of an electrical system having more than two points of contact with the rest.* In a narrower sense, *network* will often be used to indicate a linear, passive four-pole (Sec. 1.7).

A network may or may not include, as the word suggests, a number of interconnected branches; some very important networks (for example, transformers and lines) are not so constituted. On the other hand, a combination of branches, however complex, having less than three points of contact with the outside is *not* a network according to the present definition;

if it has two points of contact, it is a branch or two-pole; if one or none at all, a closed system.

An alternative, perhaps more suitable, designation for *network* is *n-pole* (four-pole, six-pole, etc.). The term *two-pole* is the equivalent of *branch.*

Network problems might well be called *transmission problems* because all transmission systems (whether for power or communication) are studied most conveniently by subdivision into networks.

It may be said more specifically that the unit of transmission systems is the *four-terminal network* or *four-pole.* This will be substantiated in the next section by generic considerations valid for both electrical and mechanical systems. Later on in the chapter, the reader will find outlined the functions that individual transmission networks are required to perform and the assumptions upon which is based the solution of network problems.

1.2. Two-point Transmission. Networks and Mechanisms. It is possible to state quite generally that if there is energy interchange between two parts of a closed system, electrical or mechanical, and if this interchange depends on contact, *there must be no less than two distinct points of contact between the two parts.* The two points must be *distinct* in the sense that if they are located within a single area of contact no energy will flow. Situations in which energy is transmitted entirely by radiation, or by sound, are not included in the statement.

The statement is self-evident with regard to most electrical systems, with some possible exceptions.

Consider, for example, an isolated metal sphere that is being charged through a single wire. In appearance, the sphere receives energy only through the charging wire, or at any rate by virtue of it. (The location of the flow of energy need not be discussed at this point.) Actually, this is not correct. As the sphere receives its charge an equal and opposite charge must go to ground; energy must be associated not with a single charge, but with both (as in a condenser). The receiver is therefore the part of the system which includes the sphere and ground and connects to the source (the electric generator) at two points: the charging wire and the grounding wire.

There is a mechanical parallel to the above. A crane lifts a bale of cotton, apparently supplying it with potential energy through a single point of contact (the hook). But here again, the energy of the lifted weight does not change, except by virtue of its position in the earth's gravitational field. Therefore, a combination of the earth and the bale of cotton should be considered as receiving energy. Such a system acquires energy because a part of it is separated from the rest against the force of gravitation. So did the condenser of the preceding example acquire energy, when its two conductors acquired equal and opposite charges. We shall return to the analogy between charge and separation (Sec. 1.5); for the present we observe that two points of contact are necessary for energy interchange in the example, the contact with the lifted weight and the contact with ground.

A bow releases an arrow. The arrow appears to have received kinetic energy by contact at a single point. But how much exactly is this energy? We say $\frac{1}{2}mv^2$, where v is referred to the earth. But if the arrow were to reach some planet other than the earth it would release on collision some value of energy not equal to $\frac{1}{2}mv^2$. We cannot assign a value to the arrow's kinetic energy except in relation to the earth and under the tacit assumption that its motion will end on the earth, where it started. Two masses having a mutual velocity (the earth and the arrow) constitute a system endowed with a definite amount of kinetic energy; not so a single moving mass.

In the examples chosen, two-point transmission is not evident because there is no question of a closed system unless we include ground. Mechanically, *ground* is any fixed rigid framework, electrically, *ground* is any conductor at ground potential. We may thus differentiate between closed systems that include ground and those that do not. Electrical systems of the first type are called *unbalanced;* of the second, *balanced.* It should be added that some systems are considered balanced when the contacts with ground carry only direct current. Likewise, in mechanics, a rotating mass is balanced when it does not subject its supports to vibrational stresses.

Having called to mind the need of two points of contact for energy interchange, we are made aware of the fundamental part played by *four-terminal networks* in electrical transmission and of their analogy with *mechanisms.*

If one part of an electrical system is the *source* of energy, it must have no less than two points of contact with the rest. The same may be said of the part where energy is ultimately utilized, or the *receiver*. What is left of the system is, therefore, in the simplest case a four-terminal network; or it may be resolved into a series of such networks.

It is clear that the task of handling the energy as it passes from generator to receiver must fall to networks; hence the advantage in considering the four-terminal network rather than the two-pole as the functional unit of transmission systems; hence, also, the analogy of networks and mechanisms. We shall make use of this analogy in briefly reviewing the ways in which networks handle electrical energy.

1.3. Functions of Networks. The connection between source and receiver may simply have the purpose of bridging the intervening distance. Thus we have, to continue the electromechanical analogy, such devices as belt drives and shafts and their electrical counterpart, the transmission line.

While we need not think of a line as a net of current paths, we cannot, in general, ignore the capacitive current flowing across the wire spacing. As a circuit, the line would have to be subdivided into an infinite number of branches; as a network, it may be considered as a whole or subdivided into sections of arbitrary length.

Because of their uniformity of structure, transmission lines are sometimes set apart from other networks and treated in much the same way as uniform elastic media are treated. They can, however, be brought within the framework of general network theory. As a matter of fact, the reasoning involved is substantially the same, whatever the method of approach. The approach through network theory has the advantage that it does not depend on the solution of a differential equation, which can be *shown* to be correct and general but is not obtained by an intuitive deduction. Moreover, the network approach has the advantage of being equally applicable to *lumped* structures.

Situations in which transmitter and receiver are joined by a simple uniform line are the exception rather than the rule. They are no more common than mechanical systems in which the transmission ratio is 1 to 1. The reduction gear and various lever actions are obviously paralleled by the transformer; but, many and varied as are mechanical transmission problems, the number and variety of transformer problems (and the word is used here in its broader sense) are still greater. The reader will be introduced later on (Sec. 8.1) to this broader meaning of transformation.

The branch of mechanics that provides the closest parallels to transforming networks, as included in the more general definition, is that of acoustics. Unfortunately, however, acoustical mechanisms are no more intuitive in their operation than their electrical counterparts. The situation is often reversed.

Many of the functions performed by networks are corrective in character and come about as the result of the strict requirements imposed upon communication systems. In this respect a sharp distinction must be made between the transmission of energy as an end in itself and the transmission of energy for the purpose of communication. In the second case the efficiency of transmission is relatively unimportant; on the other hand, the *distribution* of the energy over the frequency spectrum, or over a part of it, must remain virtually unaltered. Or, if altered for the purpose of more convenient handling, it must eventually, before reaching the receiver, be brought back to its original form.[1]

It is to be expected that the networks which, as in the case of lines, span the distance between transmitter and receiver, or as in the case of transformers, make possible the transfer of appreciable energy, will have efficiency (the word is used here without precision) variable with frequency; or, in other words, that they will to some extent *discriminate* against the energy in certain frequency intervals. In limiting cases, they may be *selective* and discriminate against all frequencies except for those in a very narrow interval. Sometimes discrimination is intentional; without it, it would not be possible to design systems in which many transmitters and many receivers are interconnected and utilize a single line. Often, however, it must be corrected by networks designed expressly for this purpose.

Further correction may be required to eliminate distortion of another kind: that arising when the *time lag*, or time interval

[1] The reader is referred to the literature[(1)] on the fundamentals of *Fourier* analysis, which underlies the concept of distribution of energy over the frequency spectrum.

elapsing between transmission and reception, varies with the frequency.

Systems in which several transmitters are linked to as many receivers have been mentioned. Such systems must include networks with more than four terminals; four-terminal network theory must therefore be extended to cover networks with more than four terminals, although in practice, owing to the fact that it is possible to think of all the transmitters and all the receivers as coming together at two junctions, the extension is a relatively simple matter.

The majority of communication systems include components which perform the functions of transmission, generation, and reception simultaneously. We do not often think of these functions being performed by an amplifier, because the word *amplifier* alone describes all three functions. However, an amplifier does receive signal energy (quite apart from battery energy), although in small quantities; it transmits energy, although sometimes not by deliberate design, and generates energy. A common approach to the study of amplifiers is the subdivision into sections, or stages; this is generally possible because the direct interchange or transmission of energy between stages is so small that its "second-order effects" may be neglected.

Sometimes this is not possible, and then we must again resort to the network concept and extend this to include networks in which energy is generated, as distinct from *passive* networks. This extension (like the other extension previously considered) is not fundamental, as long as it does not require rejection of the two fundamental hypotheses upon which elementary network theory is based: the hypothesis of *steady state* and the hypothesis of *linearity*. These hypotheses will be considered later (Sec. 1.6).

1.4. Chains of Networks. The subdivision of amplifiers into stages suggests the possibility of breaking down any complex transmission network into sections, for the purpose of analyzing it or designing its components on the assumption that their operation is not mutually affected.

Clearly, the four-terminal network that connects transmitter and receiver is generally far too complex to be considered as a single unit. Furthermore, specific functions cannot be assigned to separate components, or units, of this complex network unless the operation of each unit, when connected or inserted in the system, can be made to depend *exclusively* upon the design of the unit itself.

In order to make this possible, the individual unit networks must be connected together according to certain rules; and the more closely these rules are followed, the more accurately can the operation of the system be predicted.

Networks joined together in this way are said to constitute a *chain.* Some special condition or other must be met at each junction of the chain. Three types of chains will be considered: *iterative, image,* and *maximum power transfer* chains (Secs. 2.1 and 2.10).

It is not necessary for a chain to extend all the way from transmitter to receiver. The same conditions that prevail at each junction must, however, be fulfilled at the junctions of the chain with the remainder of the system.

1.5. Electrical and Mechanical Quantities Referred to a Junction. Duality. The phrase "conditions that prevail at a junction" is rather noncommittal and will have to be given more precise significance.

Before specifying these conditions it will be worth while, perhaps, to amplify the concept of junction and review the significance of electrical quantities as referred to a junction. Once again, the mechanical analogy will provide an intuitive basis for discussion.

We define a junction as *a pair of terminals;* more precisely, it is the combination of two areas on the boundary between parts of a system, through which there is flow of current.

Not any cut through an electrical system is necessarily a junction, but a network is bounded by two junctions. A network may or may not possess intermediate junctions.

In dealing with networks, electrical quantities are referred to some particular junction. As a matter of fact, this is true of circuits as well (Sec. 1.1), although perhaps not explicitly.

Hence, there will be reference to the power through a junction, voltage at a junction, current through a junction, although occasionally other equivalent expressions, such as *input* or *output terminals*, will be used.

It was shown that not only electrical, but also mechanical energy is transmitted through at least two points, or a junction.

Hence, it is with reference to a junction that a parallel can be traced between the electrical quantities mentioned above and their mechanical counterparts.

In order to transmit mechanical power, we must have a *force* and a *velocity*, whose scalar product is not zero. More specifically, if power is transmitted through only two points of contact, the transmitting system must exert two equal and opposite forces along the line through both points, and the points must have a relative velocity along the same line. The power is the product of this velocity by the force on either point. We can express this as follows: the power transmitted through a junction is the product of *force* and *velocity* at the junction.

Note that only *one* force and *one* velocity are mentioned, although there are two points. This is because the velocity is a relative matter, and the force likewise. As an example, if power is transmitted to a spring by compressing it, we think of the force compressing the spring, not of two forces.

Now consider the electrical case. The power transmitted through a junction is the product of the *instantaneous voltage* and the *instantaneous current* at the junction. Both these quantities, not the voltage alone, should be understood to have reference to a pair of points. A voltmeter must be connected to two points to show a reading, but so must an ammeter. An ammeter does not measure current at a point; it measures current at the junction between the meter itself and the rest of the system, which happens to be the same as the current at the junction of the load whose input current we are measuring with the rest of the system.

The identical reasoning can be applied to force with regard to dynamometers. It is not always necessary, of course, and not always convenient to think in terms of junctions rather than points. In dealing with networks, however, it is a useful habit.

The circumstance that mechanical power is the product of force and velocity (as defined), while electrical power is the product of voltage and current, would seem to indicate that each of the two electrical quantities corresponds to one of the mechanical quantities. We can, in fact, build (or imagine) mechanical models of given electrical systems in which, junction for junction, the *current* is equal to the *velocity*, the *voltage* is equal to the *force*. Or we can make the current equal to the force, the velocity equal to the voltage. In both cases the analogy can be carried out for all situations, at least in theory.

Since these analogies cannot be considered in detail here, the literature[2] should be consulted for further information. A tabulation of the corresponding quantities is given below, however, as it brings out the existence of a *dual* quantity for every electrical (or mechanical) quantity. For example, the current is the dual of the voltage (both can be considered analogous to the velocity), or the velocity is the dual of the force (both can be considered analogous to current).

Electrical	$Q = Ce = \int i\,dt$	i	e/L	e	L	C	$\Phi^* = Li = \int e\,dt$	G	R	i/C
Mechanical	s	v	a	f	m	$\frac{ds}{df}$	$mv = \int f\,dt$	$\frac{dv}{df}$	$\frac{df}{dv}$	$\frac{df}{dt}$
Electrical	$\Phi = Li = \int e\,dt$	e	i/C	i	C	L	$Q = Ce = \int i\,dt$	R	G	e/L
Mechanical	$mv = \int f\,dt$	f	$\frac{df}{dt}$	v	$\frac{ds}{df}$	m	s	$\frac{df}{dv}$	$\frac{dv}{df}$	a

* Total linked flux.

In the tabulation above, which uses standard notation, dual quantities (both electrical or both mechanical) are in the same column. The *principle of duality* may be stated as follows: *If an equation is written with regard to a system, another equation may be written, where every quantity is replaced by its dual; the new equation is valid for a new system, dual of the first.* Examples of dual systems will be pointed out as they present themselves.

1.6. The Hypothesis of Steady State. The study of networks can be greatly simplified, while remaining adequate to handle many practical situations, if it is based upon two limiting assumptions: the hypothesis of *steady state* and the hypothesis of *linearity.*

The *hypothesis of steady state* can be expressed as follows: *It is assumed that the amounts of energy transmitted through any junction over two successive equal time intervals, at any particular frequency, are not appreciably different, provided that the period of this frequency is small compared to the intervals considered.*

The hypothesis could be given quantitatively by specifying in numbers what is meant by *appreciably different* and *small compared to.* Such precision would be quite arbitrary, however.

We can only say that, in a relative sense, situations in which the difference is smaller, for larger intervals, are more adequately studied by methods based on the steady-state assumption.

To illustrate the hypothesis, suppose an oscillogram of the voltage at the junction has been taken over a period of time. *Any portion* of the oscillogram, corresponding to a particular time interval, can be analyzed into Fourier components,[1] by graphical methods or otherwise. In general, different components will be obtained for different time intervals. The sum of the components will not reproduce the original pattern except over the specified time interval.

Suppose we divide the oscillogram into intervals of $\frac{1}{10}$ sec. and carry out the analysis for each. In each interval we will find a number of sine wave components, large or small, depending on the shape of the pattern and on the accuracy of the analysis. Let us suppose that some particular component, say 1,000 cycles per second, is present in all intervals and varies in amplitude no more than 10 per cent between successive intervals. For ordinary purposes, we are then satisfied that the assumption of steady state is valid, at least for the 1,000 c.p.s. component. If we can verify that the same condition prevails for all components, using in each case intervals large compared with the period, then we are dealing with a steady-state signal.

This example does not reveal the importance of the steady-state hypothesis unless we make it the object of further reasoning.

We have succeeded in resolving the oscillogram into component voltages, each of which is sinusoidal within successive equal intervals extending over a large number of cycles. Each component has the same frequency in all the intervals and varies in amplitude slightly from one interval to the next.

Had we chosen our intervals infinitely small, each component would appear to build up or decay gradually, rather than in steps. This conclusion, although plausible, can only be guessed at, because the Fourier analysis can only be carried out over finite intervals.

We can now define the steady-state assumption in a more intuitive manner. A steady-state signal can be considered as the sum of individual steady-state components, *gradually varying* in amplitude.

The value of the assumption is this: Each *steady-state compo-*

nent, although not sinusoidal, *can be regarded as such for all practical purposes.* In fact, we can imagine a second Fourier analysis, whereby each component is in turn resolved into second-order components included, like the side bands of a modulated signal, within a narrow frequency interval. This is true independently of the time interval chosen for the second Fourier analysis. For each steady-state component, energy is transmitted only within a narrow range of frequencies, equivalent to a single frequency for practical purposes, and more exactly so when the amplitude variations of the component are more gradual.

It is clear that if we defined a *steady-state signal* as a periodic signal *i.e.*, one whose Fourier components do not depend on the interval of time selected for analysis, we would exclude from this classification all practical circumstances. In terms of sound, this definition would apply only to an even, uniform note, indefinitely sustained.

If a steady-state signal (as defined originally) is impressed upon a linear network (Sec. 1.7), each steady-state component may be considered as acting independently. Although not sinusoidal, each component occupies a narrow frequency band.

If we know how the network transmits a *sinusoidal* signal of frequency anywhere within this narrow band, then we know with sufficient accuracy how the steady-state component is transmitted.

In consequence, a plot of the frequency against the quantities (whatever they are), which define transmission, will enable us to predict how *any* steady-state signal is transmitted, thus (and here lies the simplification) obviating the necessity of considering each new wave form as a fresh problem.

The same information can be applied in certain cases to the transmission of signals not included in the steady-state category, or *transients*, although other methods are often preferable. A transient pulse can be regarded as the summation of an infinite number of Fourier components, known as the *Fourier integral*, if we suppose the analysis to be carried out over a period of infinite length.(1) The Fourier integral can, however, be evaluated only in particular cases.

If the rigid definition of steady-state signal is used (a periodic signal of indefinite duration), then all communication signals must be regarded as transients, approaching steady state to a variable extent. The loose definition which has been adopted affords a criterion for judging how closely steady state is approached.

For a given type of signal, however far from periodic, the bulk of the energy is transmitted within a specific frequency band, whose extent can be determined experimentally by suitable discriminating networks. It can be further ascertained that certain frequencies contribute more than others to the transmission of intelligence. This is true in particular with regard to speech.[(3)]

Upon such considerations is based the engineering specification of transmission networks. This generally calls for a *distortion* (Sec. 4.2) not exceeding some given value over a given frequency range.

1.7. The Hypothesis of Linearity. As a consequence of the steady-state assumption, we are now in a position to give exclusive attention to the transmission of *a single frequency*—or, which is the same thing, to the transmission of sine wave signals.

Let us suppose, then, that a sine wave voltage, or signal, appears at a junction through which electrical energy is transmitted. This voltage is fully identified by the corresponding *voltage vector*, $\mathbf{V}_0$.[1] We can say briefly, if without precision, that the voltage at the junction is $\mathbf{V}_0$. The vector notation will often be omitted in the context, but it is understood, unless otherwise specified, that equations apply to the *vector*, and not *effective*, voltages and currents.

If the junction transmits energy into a *linear* system, which is passive (*i.e.*, generates no energy) and receives energy only through the junction in question, then voltages and currents at all junctions of the system will likewise be sinusoidal, of the same frequency as $\mathbf{V}_0$. Moreover, the ratios between $\mathbf{V}_0$ and any other voltage, and the ratios between $\mathbf{V}_0$ and any current, will be complex numbers independent of the amplitude of $\mathbf{V}_0$. In other words, in such a system all voltages and currents are *linearly related.*

[1] The following notation is used:

i, v, e	instantaneous values of current, voltage, e.m.f.
I, V, E	effective values of current, voltage, e.m.f. often used for current, voltage, e.m.f. vectors in the context
$\hat{I}$, $\hat{V}$, $\hat{E}$	amplitudes of current, voltage, e.m.f.
$\mathbf{I}$, $\mathbf{V}$, $\mathbf{E}$	current, voltage, e.m.f. vectors

Physical systems are linear only as long as the voltage or current at each junction remains below some maximum value; the hypothesis of linearity can be considered satisfied by the system if this maximum is never attained in ordinary operating conditions.

All solutions of linear network problems make use of this hypothesis in some form or other. The most general solution for the terminal voltages and currents of a linear four-terminal network follows directly from the mathematical expression of the hypothesis, namely, a system of two linear equations in four variables, the voltages and currents at the two ends (Sec. 2.7). But important particular cases can be studied without writing these equations explicitly.

Instead of writing the equations, we can draw conclusions from a universal principle, *the principle of superposition*, whose validity is based, in turn, on the assumption of linearity.

This principle is best understood with reference to a linear system, upon which several electromotive forces are simultaneously impressed, for example, $\mathbf{E}_1$ and $\mathbf{E}_2$. An electromotive force (e.m.f.) is considered impressed at a junction when the voltage there is independent of all other voltages or currents in the system, being due to an external cause (Sec. 2.5).

If $\mathbf{E}_2$ were reduced to zero, the voltage at a junction J would be proportional to $\mathbf{E}_1$, because the system is linear. Hence

$$\mathbf{V}_J(O, \mathbf{E}_1) = \mathbf{A}_1\mathbf{E}_1$$

where $\mathbf{A}_1$ is a complex constant. If $\mathbf{E}_1$ is zero, we have

$$\mathbf{V}_J(\mathbf{E}_2, 0) = \mathbf{A}_2\mathbf{E}_2$$

By virtue of the superposition principle, we may write

$$\mathbf{V}_J(\mathbf{E}_1, \mathbf{E}_2) = A_1\mathbf{E}_1 + A_2\mathbf{E}_2 \tag{1}$$

The extension to the case when many e.m.fs. are impressed is immediate. Instead of the voltage $\mathbf{V}_J$, the current $\mathbf{I}_J$ can be considered.

It can be shown rather simply that the superposition principle is not valid for nonlinear systems. Suppose $\mathbf{E}_1$ and $\mathbf{E}_2$ to be in series. Then, for a nonlinear system

$$\mathbf{V}_J(O, \mathbf{E}_1) = f(\mathbf{E}_1)$$
$$\mathbf{V}_J(\mathbf{E}_2, O) = f(\mathbf{E}_2)$$

where $f(\mathbf{E})$ stands for a nonlinear function of $\mathbf{E}$. By the superposition principle, we would have

$$\mathbf{V}_J(\mathbf{E}_1, \mathbf{E}_2) = f(\mathbf{E}_1) + f(\mathbf{E}_2)$$

whereas actually the two e.m.fs. in series are equivalent to a single e.m.f. $E_1 + E_2$. Hence

$$\mathbf{V}_J(\mathbf{E}_1, \mathbf{E}_2) = f(\mathbf{E}_1, \mathbf{E}_2)$$

The last two lines are not equivalent unless the system is linear.

The concept of electromotive force, upon which the principle of superposition is predicated, deserves some attention. It is impossible to impress an e.m.f. across two points; no generator has terminal voltage independent of the current. However, any *linear* generator is *equivalent*, as far as the external circuit is concerned, to an e.m.f. independent of the load in series with an *internal impedance*. In fact this may be considered as the definition of a linear generator.

If the generator e.m.f. vanishes, then the generator becomes a linear passive system. The internal impedance of the generator has the value that would be measured under these circumstances. An e.m.f. of zero volts can be thought of as a shorting connection.

Any linear generator, no matter how complex, can be represented in this way (Thévenin's theorem, Sec. 2.6). The equivalence is only valid, however, as far as conditions *outside* the generator terminals are concerned.

Linear generators can also be represented by a *current*, whose value does not depend on conditions outside the terminals, in parallel with an internal admittance (Norton's theorem, Sec. 2.6). The *equivalent generator current* (which we shall abbreviate *e.g.c.*) becomes equivalent to an open circuit (absence of connection) when its value is zero, and the internal admittance has the value which would be measured in this case.

In most generators, rotary or otherwise, (the term includes such devices as microphones, etc.) the e.m.f. is generated by changes of magnetic flux and bears a direct relation to mechanical quantities, such as speed.

Generators which depend on changes of electric flux are not common. In such cases, however, the e.g.c. would bear a direct relation to the speed, hence the Norton equivalence would be more suitable.

Both equivalences can be used conveniently to represent the generator action of vacuum tubes. For the same generator, the two equivalent circuits are dual systems (Sec. 1.5).

A problem relating to generator equivalences is given in Sec. 1.9.

1.8. Impedances at the Junctions and Power Relations. In the preceding sections, we have been dealing with concepts of such a general nature that it would have been difficult to represent them symbolically. The two fundamental hypotheses, however,

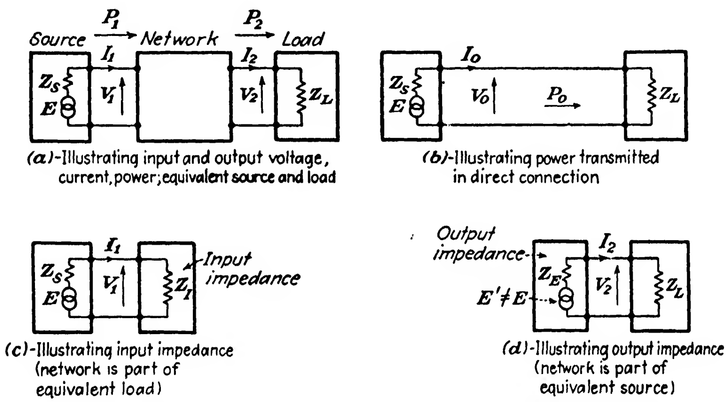


FIG. 1.—*Basic concepts relating to a four-terminal network.*

have narrowed down the object of discussion to such an extent that it can now be shown in simple symbols, as in Fig. 1*a*.

Generator, four-terminal network, and receiver (or, more briefly, source, network, and load) are shown as three boxes, which means that they are closed and energy can only be exchanged through the terminals. This was our first assumption (Sec. 1.2).

Furthermore, within the source and load boxes, equivalent circuits have been drawn. This is possible because of the linearity hypothesis. The load is equivalent (for anything *outside* its terminals) to an impedance, whose value is the complex ratio between voltage and current at the load terminals, therefore (Sec. 1.7), a constant for a given frequency.

The source is equivalent to an impedance in series with a sine-wave e.m.f. This particular form of e.m.f. can be assumed, thanks to the steady-state hypothesis.

The e.m.f. frequency f and its amplitude or effective value

E are generally considered as the *known data*, independently assigned; and so are the source and load impedances Z_s and Z_l.

Therefore, when we use equivalent representations for source and load, we are merely doing what is done in all problems, putting the data of the problem in a usable simple form.

We could use an equivalent circuit representation for the network, too; in fact, we would have a choice of several, the equivalent Π, the equivalent T, etc. All these would involve, in general, *three* distinct impedances. We would choose one of these forms, if we were interested for some reason or other in replacing the network by that particular equivalent circuit. Because this is not our present purpose, we shall leave the network box empty.

Putting the network into an equivalent form certainly does not give us the ultimate answer to the problem, although it may be a step toward that end. We are ultimately interested in finding, and describing by an adequate set of numbers, the effect that the network has on the transmission of power. As was pointed out when discussing the functions of networks, in many cases efficient power transmission is the only object; in other cases, efficiency is desirable at some frequencies, inefficiency at others. But always, efficiency of transmission is of paramount importance; hence we must somehow find an expression for it.

Considerable progress will have been made in this direction once we have determined the *power input* P_1 and the *power output* P_2. The definition of these quantities is implicit in Fig. 1*a*. But a third value of power must be introduced, if we want a true reference, the *power in direct connection* P_0. This is the power that would be transmitted from source to load if they were directly connected at a common junction (Fig. 1*b*).

Once these three powers are known, we are in a position to compare the power output which is actually available at the load with the power input which the source must deliver; or, alternatively, with the power which *would* be available at the load if the network were not *inserted*, everything else remaining the same. These comparisons can be expressed numerically, either by the ratios of the powers involved or by their logarithms (Sec. 1.9).

Power comparison is not the whole answer. The phase of both voltage and current is generally different at the two ends of the network. These phase differences, or *phase shifts*, are generally not equal. When they are large, however, they are nearly the same in a relative sense, so that we can speak of a single value of *phase shift.* The phase shift, in radians, divided by the angular frequency ω, in radians per second, gives the *time lag,* in seconds. This represents the time taken by a variation in signal amplitude in traveling through the network (Sec. 4.2).

In most cases, additional information is necessary. We must remember that both source and load, which we have so simply represented, are in reality complex systems. Just as we have reduced whatever lies *to the right of junction* 2 to a simple equivalent impedance, we might want to do the same thing with respect to junction 1; in other words, extend the boundaries of the load still further. To do this, we must find the impedance "looking toward load" at the input terminals, with source disconnected. This is called the *input impedance* (Fig. 1*c*).

Similarly, we can define the *output impedance* as the impedance measured toward the source at the output, with load disconnected and with *zero e.m.f.* at the source, in accordance with the definition of internal impedance in a generator (Sec. 1.7).

In conclusion, we are looking for four things: a power ratio,[1] a phase shift, and two impedances. The impedances, being complex, are really two numbers each, hence we are looking for six answers in all. Furthermore, we must remember that this information must, in general, be available over a range of frequencies (Sec. 1.6), so that the solution calls for six curves.

This number checks with other considerations. If we knew the ratios of three of the voltage and current vectors (Fig. 1) to the fourth, we would have all the answers. But there again, we would need three complex ratios, or six numerical data.

The six data which we have taken as descriptive of transmission through the network when this is connected in a system obviously depend on the system itself, *i.e.*, on the load and source impedances, or *terminations*, as well as on the internal structure of the network. We shall call them *transmission data.*

On the other hand, the network itself is completely identified by any one of several sets of six quantities, or *parameters.* For example, the three impedances of the equivalent T (six scalar

[1] There are two possible power ratios, as was explained, but each depends on the other when the impedances are known.

parameters) can be used for this purpose. So can the Π impedances, or the two short-circuit impedances together with the transfer impedance (Sec. 2.7), and so on.

One of these possible sets of network parameters is especially convenient because, if the terminations are correctly chosen, *it becomes identical* to the set of six transmission data (transmission loss, phase shift, input and output impedances). In other words, the network may be identified by parameters which, under special conditions often approached, also describe its performance. These parameters are called the *network constants.*[1]

For the time being, we shall remain on general ground and give some attention to the methods used for expressing power and voltage ratios.

1.9. Transmission Loss. Insertion Loss. It was pointed out that two distinct power ratios may serve to describe the efficiency of transmission. They are

$$(a) \qquad \frac{P_2}{P_1} = \frac{\text{power output}}{\text{power input}}$$

and

$$(b) \qquad \frac{P_2}{P_0} = \frac{\text{power output}}{\text{power in direct connection}}$$

One or the other of these two ratios may be significant. The choice of a criterion for power comparison will be taken up in Sec. 6.1. Sometimes the two ratios coincide; as can be readily verified, this is true when the load and input impedances (Fig. 1) are equal.

In a linear system, both ratios are constant for a given frequency. This follows from the fact that any given impedance or resistance depends only on frequency. The transmitted power at any junction is I^2R, where I is the effective current and R the resistive component of impedance looking toward load.

Suppose the effective value of current doubles at the junction. It must then double right through the system. Hence, all values of power increase fourfold.

If all values of power increase by a factor of four, the ratios between them remain the same.

[1] This designation is convenient and commonly used in spite of the fact that the parameters in question vary with the frequency.

Power ratios are seldom given explicitly. Numbers equal to their common or natural logarithms, multiplied by constant numerical coefficients, are given instead, as a measure of comparison. These numbers, or the symbols for them, appear in all computations and analytical work. They are called *losses.*

The *neper loss* $L(N)$ is half the natural logarithm of the power ratio, where the power output appears in the denominator. Thus

(*c*) $L_T(N) = \frac{1}{2} \ln P_1/P_2$ is the *neper transmission loss.*

(*d*) $L_I(N) = \frac{1}{2} \ln P_0/P_2$ is the *neper insertion loss.*

The *decibel loss* L (db) is the common logarithm of the power ratio, multiplied by 10. Thus

(*e*) L_T (db) $= 10 \log P_1/P_2$ is the *db transmission loss.*

(*f*) L_I (db) $= 10 \log P_0/P_2$ is the *db insertion loss.*

If figures are given for these losses, they are usually followed by the abbreviation N or db, as if such data were "in nepers" or "in decibels," considering the neper and the decibel as units. The advantage of thus extending the meaning of the word *unit* is open to question. The phrase "a loss of 3 db" means nothing more than "a decibel loss of 3."

The neper loss multiplied by the factor $20/\ln 10 = 8.68$ *is equal to the decibel loss.*

When the insertion loss and the transmission loss are equal, we used instead the term *attenuation,* the symbol for which is α (Sec. 2.2).

The existence of two equivalent methods for the logarithmic evaluation of ratios has historical reasons. The neper loss was introduced for basic reasons of mathematical convenience, as will appear later (Sec. 2.2). Other advantages incidental to the logarithmic evaluation came to be more widely appreciated than the early reason for its adoption. These advantages do not depend on the use of natural logarithms, so that this was largely discarded in favor of the common, or decimal logarithm, which is easier to manipulate in a superficial sense: ratios of 1 to 10, 1 to 100, and so on, are round numbers when evaluated in decibels.

It should be noted that the insertion loss (both neper and db) can be obtained from the ratio of effective *currents* or *voltages* at the load terminals directly. For instance,

$$L_I(N) = \frac{1}{2} \ln \frac{P_0}{P_2} = \frac{1}{2} \ln \frac{I_0^2 R_L}{I_2^2 R_L} = \ln \frac{I_0}{I_2}$$

where I_0 is the root-mean-square current in direct connection, R_L the load resistance.

By extension, logarithmic evaluation, particularly in the decibel form, came to be used for voltage and current ratios even when

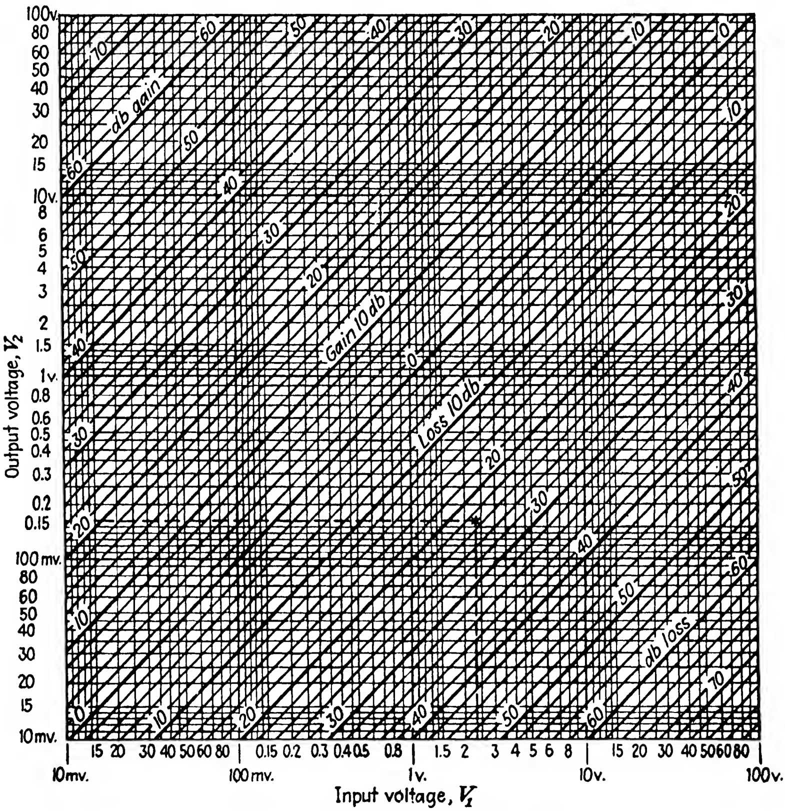


Fig. 2.—*Decibel conversion chart.*

these ratios are *not* the square roots of the corresponding power ratios.

In such cases, the terms *decibel drop* and *neper drop* should be used in place of *decibel loss* and *neper loss*. For example, the decibel drop from a high value of voltage V_1 to a low value V_2 is

$$D \text{ (db)} = 10 \log \frac{V_1^2}{V_2^2} = 20 \log \frac{V_1}{V_2}$$

If the two voltages were applied to identical impedances, then the decibel drop in voltage would be equal to the decibel loss in power.

The negative of the *drop* is the *gain*. Thus

$$G \text{ (db)} = -D \text{ (db)} = 20 \log \frac{V_2}{V_1}$$

The term *gain* does not imply a gain in power. V_2 might be, and generally is, the output voltage, V_1 the input voltage of a stage of amplification. A chart for quickly obtaining the decibel loss (or drop) and the gain when the two voltages are known (or vice versa) is given in Fig. 2.

The decibel gain is more convenient to handle than the corresponding voltage ratio, because the gains of successive stages simply add to give the total gain. Furthermore, plots of gain against frequency have the advantage of greater accuracy, for a given size, than plots of the voltage ratio. Finally, decibel figures bear a direct relation to the response of the ear.(3)

These and other considerations have spread the use of logarithmic evaluations, particularly in the decibel form, to many fields, notably acoustics.

1.10. Illustrative examples.

Equivalent generators. The following measurements were taken across the terminals of a generator of unknown characteristics, except that it is known to be inductive:

1. In short circuit:

$$I_{sc} = 1.5 \text{ amp. r.m.s.}$$

2. In open circuit:

$$V_{oc} = 31.2 \text{ volts r.m.s.}$$

3. With unknown resistive load:

$$I_r = 0.7 \text{ amp. r.m.s.}$$
$$V_r = 22.1 \text{ volts r.m.s.}$$

The frequency was constant for all measurements. Equivalent series (Thévenin) and parallel (Norton) circuits are required for the generator at the same frequency.

a. Thévenin generator. In open circuit, the e.m.f. must equal the terminal voltage. Hence

$$E = 31.2 \text{ volts r.m.s.}$$

In short circuit, the current is

$$I_{sc} = \frac{E}{|Z_s|}$$

where $|Z_s|$ is the magnitude of the internal impedance, which is then equal to

$$|Z_s| = \frac{E}{I_{sc}}$$

Measurements (3) were taken with the following resistive load:

$$R = \frac{V_r}{I_r} = 31.6 \text{ ohms}$$

The total loop impedance $|Z_s + R|$ under conditions (3) is given by

$$|Z_s + R| = \frac{E}{I_r}$$

Expanding Z_s in the rectangular form, we have

$$R_s^2 + X_s^2 = \frac{E^2}{I_{sc}^2}$$

and expanding $(Z_s + R)$,

$$(R_s + R)^2 + X_s^2 = \frac{E}{I_r^2}$$

Subtracting the last two lines and expressing R_s,

$$R_s = \frac{E^2\left(\dfrac{1}{I_r^2} - \dfrac{1}{I_{sc}^2}\right) - R^2}{2R}$$

Expressing X_s,

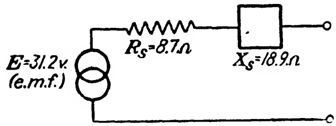
$$X_s = \pm\sqrt{\frac{E^2}{I_{sc}^2} - R_s^2}$$

Substituting numerical values,

$$R_s = \frac{(31.2)^2\left[\left(\dfrac{1}{0.7}\right)^2 - \left(\dfrac{1}{1.5}\right)^2\right] - (31.6)^2}{2 \times 31.6} = 8.7 \text{ ohms}$$

$$X_s = \sqrt{\left(\frac{31.2}{1.5}\right)^2 - (8.7)^2} = 18.9 \text{ ohms}$$

(The reactance must be positive because the generator is inductive.) Hence, the equivalent circuit below:



b. Norton generator. Any equation involving electrical quantities is equally valid for the dual system when dual quantities (Sec. 1.5) are substituted. Hence we have for G_s, dual of R_s

$$G_s = \frac{U^2\left(\frac{1}{V_r^2} - \frac{1}{V_{oc}^2}\right) - G^2}{2G}$$

and for B_s

$$B_s = \pm\sqrt{\frac{U^2}{V_{oc}^2} - G_s^2}$$

In these equations, U is the equivalent generator current, dual of the e.m.f. This is equal to the short-circuit current, because in short circuit no current flows in the internal-shunt admittance. Hence

$$U = I_{sc} = 1.5 \text{ amp.}$$

G is the reciprocal (and dual) of R; hence

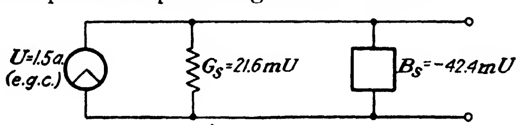
$$G = \frac{1}{R} = 0.0316 \text{ mho}$$

Substituting these and the measured values in the expression for G_s and B_s, we have

$$G_s = \frac{(1.5)^2[(1/22.1)^2 - (1/31.2)^2] - (0.0316)^2}{2 \times 0.0316} \times 10^3$$
$$= 21.6 \text{ millimhos}$$

$$B_s = -\sqrt{\left(\frac{1.5}{31.2}\right)^2 - (0.0216)^2} \times 10^3 = -42.4 \text{ millimhos}$$

whence the equivalent parallel generator below:



c. Check. The equivalent parallel generator should produce the measured open-circuit voltage (31.2 volts). Hence, this voltage must cause the currents in the shunt branches G and B to have values which, added vectorially, equal the equivalent generator (or short-circuit) current. Letting I' and I'' stand for these currents (I' current in conductance, I'' current in susceptance),

$$I' = 21.6 \times 31.2/10^3 = 0.674 \text{ amp.}$$
$$I'' = -42.4 \times 31.2/10^3 = -1.322 \text{ amp.}$$

Adding vectorially,

$$\sqrt{I'^2 + I''^2} = \sqrt{(0.674)^2 + (1.322)^2} = 1.485 \text{ amp.}$$

Previously we had obtained

$$U = 1.5 \text{ amp.}$$

The error is within slide-rule accuracy. A similar check on the series generator would call for vector addition of the voltages through internal resistance and reactance under short-circuit conditions.

Neper losses. Decibel voltage drop. The following measurements have been taken on a four-terminal network transmitting steady-state, a-c power:

1. Power input:

$$P_1 = 3.5 \text{ watts}$$

2. Power output:

$$P_2 = 1.2 \text{ watts}$$

3. Input voltage:

$$V_1 = 8.5 \text{ volts r.m.s.}$$

4. Output voltage:

$$V_2 = 5.8 \text{ volts r.m.s.}$$

With the network replaced by direct connections, the power transmitted to the load was

5. Power in direct connection

$$P_0 = 4.1 \text{ watts}$$

The following data are to be computed:

a. The load and input conductances
b. The neper insertion loss
c. The neper transmission loss
d. The decibel and neper input to output voltage drops

a. Load conductance:

$$G_L = \frac{P_2}{V_2^2} = \frac{1.2}{33.6} = 0.0357 \text{ mho}$$
$$\therefore G_L = 35.7 \text{ millimhos}$$

Input conductance:

$$G_I = \frac{P_1}{V_1^2} = \frac{3.5}{72.1} = 0.0485 \text{ mho,}$$
$$\therefore G_I = 48.5 \text{ millimhos}$$

b. Neper insertion loss:

$$L_I(N) = \tfrac{1}{2} \ln \frac{P_0}{P_2} = \tfrac{1}{2} \ln 3.42 = 0.615$$

c. Neper transmission loss:

$$L_T(N) = \tfrac{1}{2} \ln \frac{P_1}{P_2} = \tfrac{1}{2} \ln 2.42 = 0.535$$

d. Input to output voltage drop, decibels:

$$D(\text{db}) = 20 \log \frac{V_1}{V_2} = 20 \log 1.465 = 3.316$$

Conversion to neper:

$$D(N) = \frac{3.316}{8.68} = 0.382$$

CHAPTER II

THE CONSTANTS OF NETWORKS AND THEIR SIGNIFICANCE

2.1. Iterative Chains. The network problem has been stated in a very general way. At this point, two courses are open: we can either seek a general solution, or we can make further assumptions leading to a particular solution.

We will consider the second alternative and give prior attention to the study of *chain-connected* networks (Sec. 1.4).

Networks forming part of a chain are so related to the rest that their performance, as part of the chain, can be inferred from data, or *parameters*, which depend only on the network's *internal* configuration.

We would expect a chain to be uniform in some respects. The simplest type of chain consists of an indefinite number of identical networks connected end to end, or in "cascade." It must surely be possible to predict the performance of a network forming part of such a chain, when only the network parameters are known.

Consider (Fig. 3*a*) such an indefinite homogeneous chain. Note that there are many ways of connecting identical networks end to end, just as there are, for example, many ways of coupling cars in a train. Assuming that the cars have fixed seats, the order of connection is immaterial, but each car has a front and a rear which must not be exchanged. We will therefore stipulate, to begin with, that all the networks face the same way, *i.e.*, are similarly oriented.

It is clearly evident that the impedance toward load is the same at all junctions of the chain we are considering, and likewise the impedance toward source.

If there is any doubt regarding this, let us suppose the chain to be cut at some junction and consider the half chain extending indefinitely toward the load. If we remove an arbitrary number of networks from the near end of this half chain, an infinite number still remain; therefore, the new half chain is identical to the original one. It follows that the impedance toward load is the same at all junctions. The same argument applies to the impedance toward source.

These two recurrent impedances Z_1 and Z_2 are the *iterative* impedances of the chain. Instead of the infinite chain, we may consider a section of it, or even a single network, provided it is *terminated* in these impedances (Fig. 3*b*). It clearly does not matter, as far as the operation of the network is concerned, whether it is part of a homogeneous chain or not, provided the

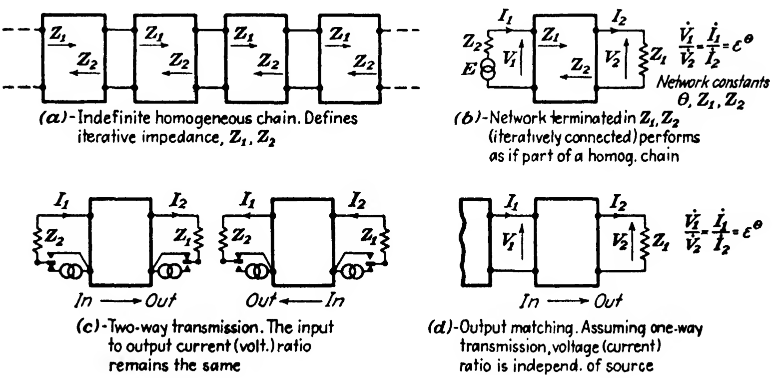


FIG. 3.—*Iterative connection.*

source and the load are correctly represented by the equivalent circuits of Fig. 3*b*.

Nor is it necessary to consider an indefinite chain in order to define the iterative impedances. In the system of Fig. 3*b*, the load and input impedances are equal, and likewise the source and output impedances. Hence, the iterative impedances are simply defined as those values which, when used as source and load impedances of a given network, produce identical values of output and input impedance, respectively. In other words, they reproduce or repeat themselves from one end of the network to the other (hence the word *iterative*).

Evidently, the iterative impedances depend on the internal structure of the network exclusively and may therefore be regarded as network parameters.

It was anticipated in Sec. 1.8 that it might be possible to select the network parameters so that, under special circumstances, they are equal to the six quantities which describe network performance.

We see now that this is at least partly possible for the special circumstance of iterative connection. In fact, for the iteratively connected network, the input and output impedances (used to describe network performance, Sec. 1.8) are equal to the iterative impedances (network parameters).

We have agreed to call these specially selected parameters *network constants*, as is customary, in spite of the fact that they vary with frequency. The iterative impedances are network constants. Being complex, they take care of four out of the six real numbers which, in one way or another, are necessary to define the network.

Two more real parameters remain open to choice. This means that two networks having the same iterative impedances are not necessarily equal. If they were, iterative chains would all be homogeneous, and different functions could not be assigned to the constituent networks.

The two remaining parameters, in order to be classed as "constants," would have to be equal to two of the remaining descriptive quantities. One of them could be the *transmission loss* (or alternatively the *insertion loss*) of the iteratively connected network; the other would have to be the *voltage phase shift* (or alternatively the *current phase shift*) of the iteratively connected network.

As it happens, there is only one choice, not two, because the insertion and transmission loss of iteratively connected networks are equal, and likewise the voltage and current phase shifts. Both propositions follow immediately, as the reader will readily verify, from the equality of input and load impedances, and, respectively, output and source impedances.

2.2. Network Constants. To summarize our conclusions: Among the many possible sets of parameters, we have selected as means of identifying the network those that are equal to the *transmission data* describing network performance, for the case of iterative connection. They are

1. *The iterative impedance* Z_1 (equal to both load and input impedances when the network is iteratively connected).

2. *The iterative impedance* Z_2 (equal to both source and output impedances when the network is iteratively connected).

3. *The attenuation* α (*neper or decibel*) (equal to the insertion loss, transmission loss, voltage and current drop, for the iterative connection).

4. *The phase shift* β (*in degrees or radians*) (equal to the phase difference of either current or voltage vectors for iterative connection).

Constants (3) and (4) are advantageously combined into a single complex number. If the attenuation is in nepers and the phase shift in radians, *this complex number is identically equal to the natural logarithm of the ratio of the voltage* (*current*) *vectors* for the iterative connection. It is then written

$$\theta = \alpha + j\beta$$

where α is the neper attenuation, β the phase shift in radians. θ is called *propagation* or *transfer constant.*

The significance of the transfer constant in terms of voltage or current ratio is verified immediately. We may write

$$\mathbf{V}_1 = \hat{V}_1 e^{j(\omega t+\beta_1)} \qquad \mathbf{V}_2 = \hat{V}_2 e^{j(\omega t+\beta_2)}$$

Hence

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\hat{V}_1}{\hat{V}_2} e^{j(\beta_1-\beta_2)} \tag{2}$$

Since α, the neper attenuation, is also equal to the neper voltage drop, we have

$$\alpha = \ln \frac{\hat{V}_1}{\hat{V}_2}$$

whence

$$\frac{\hat{V}_1}{\hat{V}_2} = e^{\alpha} \tag{3}$$

The phase shift in radians is obviously given by

$$\beta = \beta_1 - \beta_2 \tag{4}$$

Hence, substituting (3) and (4) into (2),

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = e^{\alpha+j\beta} = e^{\theta} \tag{5}$$

or

$$\theta = \ln \frac{\mathbf{V}_1}{\mathbf{V}_2} \tag{6}$$

To some readers, the logarithm of a complex quantity may not have a clear meaning. The literature on functions of complex variables[4] should be consulted. But the definition of the transfer constant itself may help in clearing up this point.

The transfer constant is equal to the natural logarithm of a complex quantity; at the same time, we have shown it to be equal to the natural logarithm of the *magnitude* ($\alpha = \ln V_1/V_2$) plus j times the *angle* in radians [$j(\beta_1 - \beta_2)$]. This, in general, is the significance of the logarithm of a complex quantity.

The reason for selecting earlier (Sec. 1.9) the neper rather than any other logarithmic evaluation of ratios is now apparent. In order to evaluate the logarithm of a complex ratio, we must take the neper expression of the magnitude of the ratio and add j times the angle of the ratio, in radians.

If we took the decibel expression and added j times the angle in radians, we would have a meaningless number.

Thanks to the introduction of the transfer constant, the network constants may now be given as *three complex numbers*, namely

1. and 2. The iterative impedances Z_1 and Z_2

3. The transfer constant θ

2.3. The Principle of Reciprocity. The transfer constant of a network has been defined as the logarithm of the input to output voltage vector ratio (Eq. 6) when the network is iteratively connected. This ratio and the corresponding current ratio are, of course, equal.

The definition appears ambiguous in one respect. Consider a system where both terminations are capable of transmitting energy, although not at the same time. This is not just an academic hypothesis; all systems which provide "two-way" transmission are in this category.

Such a system may be shown as in Fig. 2*c*, with the understanding that both e.m.fs. are not in operation at the same time (one or the other is always zero).

The network of Fig. 2*c* is iteratively connected, hence the ratio of input to output voltage vectors must be e^{θ}. But, which is the input and which the output? If θ is defined with one direction of transmission in mind, will the definition hold for the other direction as well?

The answer is in the affirmative, because the ratio of input to output voltage (or current) vectors is the same for iteratively connected networks, no matter which end is the input and which the output.

This proposition is contained, as a corollary, within a much broader principle known as the *principle of reciprocity*. Like the principle of superposition, this is universally valid for linear systems, whether mechanical or electrical; and, like the principle of superposition, it is best understood with reference to a linear system on which two e.m.fs. $\mathbf{E}_1$ and $\mathbf{E}_2$ are impressed at the same time (see Sec. 1.7).

Let $\mathbf{I}_1$ be the current at the junction where $\mathbf{E}_1$ is impressed, or the current through $\mathbf{E}_1$, and $\mathbf{I}_2$ the current through $\mathbf{E}_2$. Then, by the superposition principle, we have (Sec. 1.7, Eq. 1)

$$\begin{aligned} \mathbf{I}_1(\mathbf{E}_1\mathbf{E}_2) &= y_{11}\mathbf{E}_1 + y_{12}\mathbf{E}_2 \\ \mathbf{I}_2(\mathbf{E}_1\mathbf{E}_2) &= y_{21}\mathbf{E}_1 + y_{22}\mathbf{E}_2 \end{aligned} \tag{7}$$

where all the coefficients are complex constants for the particular frequency under consideration.

The principle of reciprocity is contained in the equality

$$y_{12} = y_{21} \tag{8}$$

Now suppose that first $\mathbf{E}_1$, then $\mathbf{E}_2$, are given a common value $\mathbf{E}_0$ while in each case the remaining e.m.f. is zero. The currents in the two cases will be

$$\begin{aligned} \mathbf{I}_1(\mathbf{E}_0\, 0) &= y_{12}\mathbf{E}_0 \\ \mathbf{I}_2(0\, \mathbf{E}_0) &= y_{21}\mathbf{E}_0 \end{aligned}$$

Hence

$$\mathbf{I}_1(\mathbf{E}_0\, 0) = \mathbf{I}_2(0\, \mathbf{E}_0) \tag{9}$$

which is a widely quoted statement of the principle.

In the above, voltages could have been considered in place of currents, and independent currents (e.g.cs.) in place of e.m.fs., by virtue of the principle of duality (Sec. 1.5).

Various proofs of the principle of reciprocity have been given. It must be remembered, however, that neither the principle of superposition nor that of reciprocity is proved by showing that it applies in particular cases. Neither is it proved with the help of equivalent circuits. These can be set up in the first place only as a consequence of the principles themselves. For example, the T structure equivalent to a network for transmission in one direction can readily be shown to have a transfer constant independent of direction. However, if we reverse the direction, how valid is the equivalence? Only as valid as the principle of reciprocity.

Returning to the iteratively connected network of Fig. 2*c*, if an e.m.f. E at the near end (for example, at the left) sends an output current I_2 into the far end, then the same E at the far end will send I_2 into the near end (9). Thus, if the e.m.f. simply changes ends without changing in value, the output current remains equal.

Now consider the input current. When transmission is from the near end, the input current (at the near end) is

$$\mathbf{I}_1 = \frac{\mathbf{E}}{Z_1 + Z_2}$$

When transmission is from the far end, we have at the far end the input current

$$\mathbf{I}_1 = \frac{\mathbf{E}}{Z_2 + Z_1}$$

In conclusion, input and output currents simply change ends. We have therefore shown that, if the e.m.f. goes from one end to the other without changing value, the input to output current vector ratio remains the same. If this is true, it must remain the same for all values of the e.m.f. Obviously, the same can be said of the voltage vector ratio and of the transfer constant. Hence this is uniquely defined.

2.4. Output Matching. We defined the network constants in terms of their significance in the particular condition of iterative connection. For example, the transfer constant is equal to the natural logarithm of the input to output voltage ratio *for the iterative connection.* It should be noted that this, while a sufficient condition, is not a *necessary* condition for the equality to hold.

The input to output voltage vector ratio is still equal to e^θ, even when the network is not iteratively connected, provided the *load impedance* has the iterative value. In other words, this ratio (as well as the current ratio) is unaffected by the source impedance or anything else pertaining to the source.

This premise is easily verified. The system which includes load and network is a passive linear system. Any two voltages in this system (Sec. 1.7) are in constant ratio for the frequency under consideration; this ratio may be regarded as a parameter of the system. We cannot affect it by changing anything outside the system.

The expression *matching terminations* is often used to designate the source and load impedances of a network in iterative connection (although the phrase may have other meanings). If a matched network is one whose terminations are both matching, we will use the term *output matched* for one whose load only is matching, while the source is not specified. Obviously, while a matched network is matched independently of the direction of transmission, an output-matched network is no longer output matched if the direction is inverted. Output matching is insufficient in two-way transmission systems.

A network has the same transmission loss, insertion loss, phase shift, and input impedance, whether it is *matched* (iteratively connected) or simply *output matched* (iteratively connected with regard to the load only). The *output impedance*, however, differs in the two cases, as can be immediately verified.

If the direction of transmission through an output-matched network is inverted, the network ceases to be output matched and all the above quantities change. In this case the source, not the load impedance, would have the iterative value; the load would have some generic value upon which all the quantities (except the output impedance) would depend.

2.5. Fictitious Replacements. The logic behind some of the methods of network analysis is rather delicate and deserves at least as much attention as the algebraic manipulation.

The superposition system (Sec. 1.7) was explained with reference to a linear system upon which several e.m.fs. were impressed. In stating the principle, the e.m.fs. were implicitly regarded as the causes to which all voltages and currents were due.

This is correct enough. The e.m.f. is distinguished from all other voltages (and the e.g.c. from all other currents) by its *independence* from the circuit, or system, upon which it is impressed by an outside agency. In other words, it is an *independent variable.* For example, the e.m.f. of an alternator is an independent variable as far as the alternator and its circuit are concerned; it depends on the driving speed and the excitation, factors that have no connection with the alternator circuit proper.

On the other hand, the voltage across two points of a circuit depends, in general, upon the load on which impressed. Only a generator of zero impedance, which cannot be realized, would have voltage independent of load.

For this reason an e.m.f. may be thought of as a generator of zero impedance.

A physical generator, of finite internal impedance, is equivalent to an e.m.f. (impedanceless generator) in series with a passive load.

In addition, it is possible to imagine *any* two-pole, generator or other, replaced by a *fictitious* generator of zero impedance, or fictitious e.m.f.

Consider, for example, the terminal voltage of a generator, given by

$$\mathbf{V} = \frac{\mathbf{E}Z_l}{Z_s + Z_l} \tag{10}$$

where E is the generator e.m.f., Z_s and Z_l the source and load impedances. Let us imagine the generator replaced by a pure e.m.f. (or impedanceless generator) which can be adjusted by an automatic control.

Suppose the load Z_l is varied. The value of V for the actual generator would vary according to (10). But we can imagine the automatic control of the fictitious impedanceless generator so set up that it responds to variations in the load and varies the e.m.f. also according to (10). Under these conditions, the fictitious impedanceless generator and the actual generator are equivalent.

Electromotive forces such as that in the preceding example, whose value is not independent of the system parameters, are called *fictitious* or *dependent.*

It is clear that any branch or two-pole forming part of an electrical system may be ideally replaced by a *fictitious e.m.f.* equal under all conditions to the voltage appearing across the two-pole. This *fictitious replacement* will not disturb the remainder of the system.

It is equally possible to replace the two-pole by a *fictitious e.g.c.* equal under all conditions to the current flowing through the two-pole. A fictitious e.g.c. is a generator of infinite internal impedance, whose terminal current is the required function of the system parameters.

In the case of a passive branch, the fictitious e.m.f. may be written IZ, where I is the branch current and Z the branch impedance; the fictitious e.g.c. may be written VY, where V is the branch voltage and Y the branch admittance. The first of these two statements is usually given prominence. It goes under the name of *compensation theorem.*

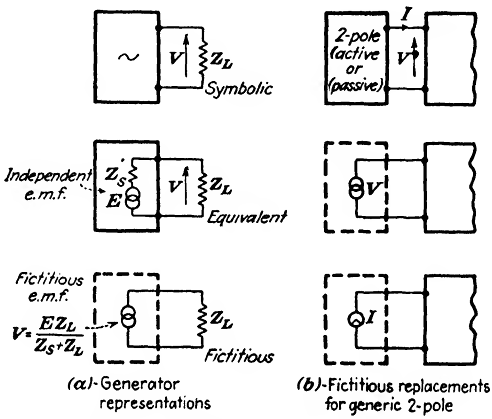


FIG. 4.—*Fictitious replacements.*

The difference between a fictitious replacement and an equivalent circuit should be kept in mind. An equivalent circuit is completely identified by parameters (such as impedances) and independent variables (independent e.m.fs.). The quantity or quantities which identify a fictitious replacement are functions of the remainder of the system.

In the limiting cases when a branch does not conduct current (open branch) or has no voltage across it (shorted branch) there is latitude in the choice of a fictitious replacement, namely

1. A branch in open circuit (open branch) may be replaced by a fictitious e.m.f. equal to the voltage across the branch terminals (open-circuit voltage), *in series with an arbitrary impedance.*

2. A branch in short circuit (shorted branch) may be replaced by a fictitious e:g.c. equal to the short-circuit current *in parallel with an arbitrary admittance.*

Evidently, absence of all connection may be regarded as an open branch and a shorting connection as a shorted branch.

Statement 1 will appear plausible if the reader thinks, for example, of placing a 110-volt battery across a 110-volt d-c line (positive terminal to positive terminal) across an arbitrary resistance. No current will flow, although an open circuit (absence of connection) has been replaced by a branch potentially able to carry current. However, if the line voltage fluctuates, current will start to flow unless by some means we adjust the battery e.m.f. to follow the line-voltage fluctuations.

2.6. Thévenin's and Norton's Theorems. With the help of fictitious replacements, the Thévenin and Norton equivalences for linear generators may be stated with greater precision. In Sec. 1.7, these equivalences were introduced simply as definitions of linear generators.

Consider Fig. 5*a*, a generator in open circuit. The open circuit, considered as a branch of zero admittance or open branch, may be replaced by a fictitious e.m.f. equal to the open-circuit voltage V_{oc}, in series with an arbitrary impedance Z_l (statement 2, Sec. 2.5).

The replacement is possible because the component currents due to the generator and to the fictitious e.m.f., adding by the superposition principle, cancel or balance out. We can write an equation to this effect

$$I_l + \frac{V_{oc}}{Z_l + Z_s} = 0 \qquad (11)$$

where I_l is the component of current through Z_l due to the generator and Z_s the generator impedance measured when the generator is inactive.

In (11) we have an expression for I_l, the current flowing through a generic load connected to the generator terminals. This is

$$I_l = \frac{V_{oc}}{Z_l + Z_s} \qquad (12)$$

This expression is true for any Z_l, as this was arbitrarily selected. Therefore, for any load, the generator is equivalent to an e.m.f. $-V_{oc}$ in series with the impedance Z_s. The minus sign of V_{oc} appears when we place this e.m.f. inside the generator terminals instead of out. It means that the internal e.m.f. of the generator and the open-circuit voltage are opposed as far the open-circuit current is concerned, thereby causing it to vanish. They must be considered to have equal signs with regard to any branch added in shunt to the open circuit. The equivalence may be expressed as follows:

Thévenin's theorem. For a given frequency, any *generator* (active linear system) is equivalent to an e.m.f. equal to the

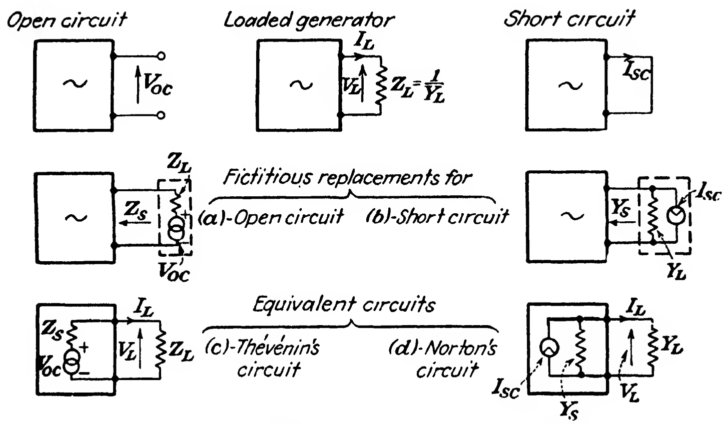


FIG. 5.—*Thevenin and Norton equivalences.*

open-circuit voltage, in series with an impedance having the value measured at the generator terminals at the same frequency, in absence of load, and with the generator inactive.

The statement of Norton's theorem (Fig. 5*b*) closely parallels the above. Norton's theorem is implicit in Thévenin's theorem and in the principle of duality. A statement of the theorem is the following:

Norton's theorem. For a given frequency, any *generator* (active linear system) is equivalent to an e.g.c. (equivalent generator current) equal to the short-circuit current, in parallel with an admittance having the value measured across the generator terminal at the same frequency, in absence of load, with the generator inactive.

It should be noted that the internal admittance of the Norton generator is the reciprocal of the internal impedance of the

Thévenin generator, since both are measured under the same conditions.

2.7. Measurable Parameters. Going back, temporarily, to the network with generic terminations, the general method of solution will now be indicated. Rather than distinguish between load and source, we will suppose both terminations to be active. Figures 6*a* and 6*b* represent the Thévenin and Norton equivalent circuits for such a system. Note that $Y_a = 1/Z_a$, $Y_b = 1/Z_b$.

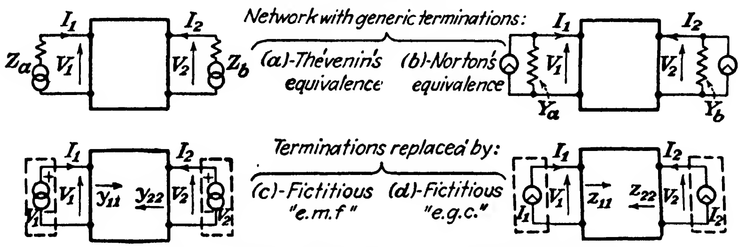


FIG. 6.—*Network between generic terminations.*

Both terminations may, for convenience, be replaced by fictitious e.m.fs. equal to V_1 and V_2, or by fictitious e.g.cs., equal to I_1, I_2. Figure 6*c* shows the first replacement, Fig. 6*d* the second. The superposition principle may be applied in order to find the terminal currents in terms of the e.m.fs. V_1 and V_2, or the terminal voltages in terms of I_1 and I_2. The following equations result:

(Fig. 6*c*)

$$\begin{aligned} \mathbf{I}_1 &= y_{11}\mathbf{V}_1 + y_{12}\mathbf{V}_2 \\ \mathbf{I}_2 &= y_{21}\mathbf{V}_1 + y_{22}\mathbf{V}_2 \end{aligned} \qquad (13)$$

(Fig. 6*d*)

$$\begin{aligned} \mathbf{V}_1 &= z_{11}\mathbf{I}_1 + z_{12}\mathbf{I}_2 \\ \mathbf{V}_2 &= z_{21}\mathbf{I}_1 + z_{22}\mathbf{I}_2 \end{aligned} \qquad (14)$$

where, in agreement with the principle of reciprocity,

$$y_{12} = y_{21} \qquad\qquad z_{12} = z_{21}$$

The above equations are similar to (7) (Sec. 2.3), but there is this difference: In Eqs. (7), the currents were expressed in terms of *independent* variables (E_1 and E_2), which must be considered as the given data; Eqs. (13) and (14), on the other hand, merely correlate four dependent variables together. They cannot lead to a solution without some additional information, since there are two equations in four variables (obviously there are only two independent equations, written in two alternative forms). The additional information must be supplied by the expressions for the fictitious e.m.fs. V_1, V_2 (or e.g.cs. I_1, I_2), namely

(Fig. 6*a*)

$$\mathbf{V}_1 = \mathbf{E}_1 - Z_a\mathbf{I}_1$$
$$\mathbf{V}_2 = \mathbf{E}_2 - Z_b\mathbf{I}_2$$

(Fig. 6*b*)

$$\mathbf{I}_1 = \mathbf{U}_1 - Y_a\mathbf{V}_1$$
$$\mathbf{I}_2 = \mathbf{U}_2 - Y_b\mathbf{V}_2$$

If, instead of a single network between known terminations, we had considered a system of many networks connected together, the problem would require simultaneous solution of many equations. In any case, three complex coefficients (y_{11}, y_2 and y_{22} or z_{11}, z_{12}, and z_{22}) would have to be known for each network. The significance of the coefficients becomes clear when we observe the following:

for $\mathbf{V}_2 = 0$

$$y_{11} = \frac{\mathbf{I}_1}{\mathbf{V}_1}$$
$$y_{12} = y_{21} = \frac{\mathbf{I}_2}{\mathbf{V}_1}$$

for $\mathbf{V}_1 = 0$

$$y_{22} = \frac{\mathbf{I}_2}{\mathbf{V}_2}$$
$$y_{12} = y_{21} = \frac{\mathbf{I}_1}{\mathbf{V}_2}$$

for $\mathbf{I}_2 = 0$

$$z_{11} = \frac{\mathbf{I}_1}{\mathbf{V}_1}$$
$$z_{12} = z_{21} = \frac{\mathbf{V}_2}{\mathbf{I}_1}$$

for $\mathbf{I}_1 = 0$

$$z_{22} = \frac{\mathbf{V}_2}{\mathbf{I}_2}$$
$$z_{12} = z_{21} = \frac{\mathbf{V}_1}{\mathbf{I}_2}$$

Hence the following definitions for the coefficients:

1. y_{11} is the *near end or driving point short-circuit admittance* (admittance measured at the near end with the far end shorted).

2. y_{22} is the *far end short-circuit admittance* (measured at the far end with the near end shorted).

3. $y_{12} = y_{21}$ is the *transfer admittance*, equal to the ratio

$$\frac{\text{Current at the shorted end}}{\text{Voltage at the other end}}$$

4. z_{11} is the *near end or driving point open-circuit impedance* (impedance measured at the near end with the far end open).

5. z_{22} is the *far end open-circuit impedance* (measured at the far end with the near end open).

6. $z_{12} = z_{21}$ is the *transfer impedance,* equal to the ratio

$$\frac{\text{Voltage at the open end}}{\text{Current at the other end}}.$$

Any three of the six coefficients define the network, hence when three are given the rest can be obtained from them. Of the identities correlating the coefficients with one another, one in particular will be found useful, namely

$$y_{11}z_{11} = y_{22}z_{22} \tag{15}$$

To verify this, let $V_2 = 0$ in Eqs. (13) and (14). Then we have

$$y_{11} = \frac{\mathbf{I}_1}{\mathbf{V}_1} \qquad y_{21} = \frac{\mathbf{I}_2}{\mathbf{V}_1} \qquad \frac{z_{11}}{z_{21}} = -\frac{\mathbf{I}_1}{\mathbf{I}_2}$$

Hence

$$\frac{y_{11}}{z_{22}} = -\frac{y_{21}}{z_{21}}$$

Likewise, if V_1 is set equal to zero, we obtain

$$\frac{y_{22}}{z_{11}} = -\frac{y_{12}}{z_{12}}$$

which, compared with the above, yields (15).

The network coefficients, unlike the network constants, are never equal to the transmission data (transmission loss, phase shift, input and output impedance). In themselves they do not tell us much about network performance. On the other hand, they can be *measured* (or computed) on the network—the short-circuit admittances and open-circuit impedances being preferable, in this respect, to the transfer coefficients.

We will therefore refer to y_{11}, y_{22}, z_{11}, and z_{22} as *measurable parameters* and consider them as known quantities. They represent a starting point from which to attack the network problem. In a generic case, the problem is attacked by writing four or more equations in which these parameters appear as coefficients and solving them simultaneously. In the particular case of iterative connection, all the information we need is contained in the network constants; hence, all we have to do is to express the network constants in terms of the measurable parameters. This will be the object of the following section.

2.8. Network Constants in Terms of Measurable Parameters. Consider a network in iterative connection: Two equivalent representations (Norton and Thévenin) are possible for the source termination, and both will be used (Figs. 7*a* and 7*b*).

The values of current and voltage in the network so connected will be called, for brevity, matched values ($\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{V}_1$, $\mathbf{V}_2$). They are mutually related as follows:

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\mathbf{I}_1}{\mathbf{I}_2} = e^{\theta}$$

$$\frac{\mathbf{V}_1}{\mathbf{I}_1} = Z_1 = \frac{1}{Y_1}; \qquad \frac{\mathbf{V}_2}{\mathbf{I}_2} = Z_2 = \frac{1}{Y_2}$$

$$\mathbf{I}_1 = \frac{\mathbf{E}}{Z_1 + Z_2}$$

where θ is the transfer constant, Z_1 and Z_2 the iterative impedances (Sec. 2.1).

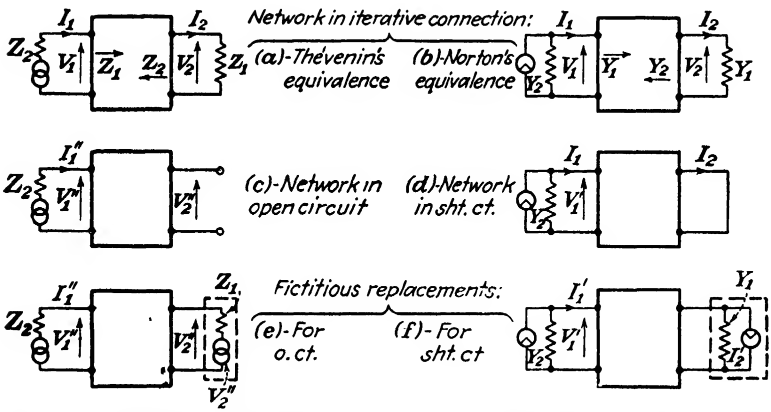


FIG. 7.—*Network in iterative connection. Derivation of the network constants.*

Now consider the same network, with the same source but with the *output shorted*, and let the currents and voltages for this case (short-circuit values) be denoted by primed symbols (I_1' I_2' V_1' V_2'). For convenience, the page will be divided into two columns and all work relating to the short-circuited network will occupy the left-hand column.

On the right, the parallel steps will be carried out on the same network (with the same source), but with *output open*, and the corresponding values will be denoted by double-primed symbols (I_1'', and so on).

SHORT-CIRCUITED NETWORK (Fig. 7*d*)

The short circuit across the output can be replaced, as in (2), Sec. 2.5, by a fictitious e.g.c. equal to the short-circuit current I_2', shunted by the admittance Y_1 (Fig. 7*f*).

I_2' can be found as follows: The output voltage due to I_2' alone and the output voltage due to the source alone (the matched value), adding by the superposition principle, must cancel. Hence

$$\mathbf{V}_2 - \frac{\mathbf{I}_2'}{Y_1 + Y_2} = 0$$

$$\therefore \mathbf{I}_2' = \mathbf{V}_2(Y_1 + Y_2) = \mathbf{V}_1(Y_1 + Y_2)e^{-\theta}$$

Knowing I_2', we can write the other short-circuit values, adding the components due to I_2' to the matched values:

$$\mathbf{I}_1' = \mathbf{I}_1 + \mathbf{I}_2' \frac{Y_2}{Y_1 + Y_2} e^{-\theta} = \mathbf{I}_1 + \mathbf{V}_1 Y_2 e^{-2\theta}$$

$$\therefore \mathbf{I}_1' = \mathbf{I}_1\left(1 + \frac{Y_2}{Y_1} e^{-2\theta}\right)$$

$$\mathbf{V}_1' = \mathbf{V}_1 - I_2' \frac{e^{-\theta}}{Y_1 + Y_2} = \mathbf{V}_1(1 - e^{-2\theta})$$

OPEN-CIRCUITED NETWORK (Fig. 7*c*)

The open circuit across the output can be replaced, as in (1), Sec. 2.5, by a fictitious e.m.f. equal to the open-circuit voltage, V_2'', in series with the impedance Z_1 (Fig. 7*e*).

V_2'' can be found as follows: The output current due to V_2'' alone and the output current due to the source alone (the matched value), adding by the superposition principle, must cancel. Hence

$$\mathbf{I}_2 - \frac{\mathbf{V}_2''}{Z_1 + Z_2} = 0$$

$$\therefore \mathbf{V}_2'' = \mathbf{I}_2(Z_1 + Z_2) = \mathbf{I}_1(Z_1 + Z_2)e^{-\theta}$$

Knowing V_2'', we can write the other open-circuit values, adding the components due to V_2'' to the matched values:

$$\mathbf{V}_1'' = \mathbf{V}_1 + \mathbf{V}_2'' \frac{Z_2}{Z_1 + Z_2} e^{-\theta} = \mathbf{V}_1 + \mathbf{I}_1 Z_2 e^{-2\theta}$$

$$\therefore \mathbf{V}_1'' = \mathbf{V}_1\left(1 + \frac{Z_2}{Z_1} e^{-2\theta}\right)$$

$$\mathbf{I}_1'' = \mathbf{I}_1 - \mathbf{V}_2'' \frac{e^{-\theta}}{Z_1 + Z_2} = \mathbf{I}_1(1 - e^{-2\theta})$$

Dividing the above lines by one another, and noting that

$$\frac{\mathbf{I}_1'}{\mathbf{V}_1'} = y_{11} \text{ (near end short circuit admittance)}$$

and that

$$\frac{\mathbf{I}_1}{\mathbf{V}_1} = Y_1$$

$$\frac{\mathbf{V}_1''}{\mathbf{I}_1''} = z_{11} \text{ (near end open circuit impedance)}$$

$$\frac{\mathbf{V}_1}{\mathbf{I}_1} = Z_1$$

SHORT-CIRCUITED NETWORK
(Fig. 7*d*)

we obtain the following:

$$y_{11} = \frac{Y_1 + Y_2e^{-2\theta}}{1 - e^{-2\theta}} = \frac{Y_1e^{\theta} + Y_2e^{-\theta}}{e^{\theta} - e^{-\theta}} \qquad (16)$$

OPEN-CIRCUITED NETWORK
(Fig. 7*c*)

$$z_{11} = \frac{Z_1 + Z_2e^{-2\theta}}{1 - e^{-2\theta}} = \frac{Z_1e^{\theta} + Z_2e^{-\theta}}{e^{\theta} - e^{-\theta}} \qquad (17)$$

The entire derivation, for both short and open circuit, may evidently be repeated with the network turned about. This merely interchanges the subscripts 1 and 2. Thus

$$y_{22} = \frac{Y_2e^{\theta} + Y_1e^{-\theta}}{e^{\theta} - e^{-\theta}} \qquad (18)$$

$$Z_{11} = \frac{Z_2e^{\theta} + Z_1e^{-\theta}}{e^{\theta} - e^{-\theta}} \qquad (19)$$

Equations (16), (17), (18), (19) are four equations in the three unknowns θ, Z_1 (or Y_1), and Z_2 (or Y_2). Clearly, the equations are not independent. In fact, the parameters y_{11}, y_{22}, z_{11}, z_{22}, are correlated by Eq. (15), namely

$$y_{11}z_{11} = y_{22}z_{22}$$

hence any one of the four equations could have been obtained from the other three.

The solution of the equations, leading to explicit expressions for the network constants θ, Z_1, and Z_2, is purely a matter of algebraic manipulation. We will continue for a time the division of the work in two columns, which permits a more orderly arrangement.

Subtracting (18) from (16),

$$y_{11} - y_{22} = Y_1 - Y_2 \qquad (20)$$

Subtracting (19) from (17),

$$z_{11} - z_{22} = Z_1 - Z_2 \qquad (21)$$

This is a remarkable result. It shows, for one thing, that if the network measures the same from the two ends, either in open or short circuit, its iterative impedances are the same.

Eliminating Y_2 in (16), (18),

$$y_{11}e^{\theta} - y_{22}e^{-\theta} = Y_1(e^{\theta} + e^{-\theta}) \qquad (22)$$

Eliminating Z_2 in (17), (19),

$$z_{11}e^{\theta} - z_{22}e^{-\theta} = Z_1(e^{\theta} + e^{-\theta}) \qquad (23)$$

Multiplying (22) by (23) and recalling (15),

$$(e^{\theta} + e^{-\theta})^2 = y_{11}z_{11}(e^{2\theta} + e^{-2\theta}) - y_{11}z_{22} - z_{11}y_{22}$$

In view of the identities

$$(e^\theta + e^{-\theta})^2 = e^{2\theta} + e^{-2\theta} + 2 = 4 \cosh^2 \theta$$
$$(y_{11} + y_{22})(z_{11} + z_{22}) = 2y_{11}z_{11} + y_{11}z_{22} + y_{22}z_{11}$$

the above can be put in the form

$$\cosh \theta = \tfrac{1}{2}\sqrt{\frac{(y_{11} + y_{22})(z_{11} + z_{22})}{y_{11}z_{11} - 1}} \tag{24}$$

[*General expression for transfer constant*]

Equation (24), like all equations involving the hyperbolic functions of a complex variable, is not in a form suitable for numerical computation. The reader will find, at the end of the chapter and elsewhere, examples of the algebraic procedure used to obtain α and β from hyperbolic functions of θ. Alternatively, maps of the hyperbolic functions can be used.

Equations (20) and (21) lead to expressions for the iterative impedances. From (20) we have

$$y_{11} - y_{22} = \frac{1}{Z_1} - \frac{1}{Z_2}$$
$$\therefore\ Z_2 - Z_1 - Z_1Z_2(y_{11} - y_{22}) = 0$$

Combining the above with (21),

$$(y_{11} - y_{22})Z_1^2 - (y_{11} - y_{22})(z_{11} - z_{22})Z_1 + (z_{11} - z_{22}) = 0$$

Similarly,

$$(y_{11} - y_{22})Z_2^2 + (y_{11} - y_{22})(z_{11} - z_{22})Z_2 + (z_{11} - z_{22}) = 0$$

Solving for Z_1 and Z_2, and noting that

$$\frac{z_{11} - z_{22}}{y_{11} - y_{22}} = \frac{z_{11}}{y_{22}}\,\frac{1 - z_{22}/z_{11}}{y_{22}/y_{11} - 1} = -\frac{z_{11}}{y_{22}}$$

we have for the iterative impedances

$$\left\{\begin{aligned} Z_1 &= \sqrt{\frac{z_{22}}{y_{11}} + \left(\frac{z_{11} - z_{22}}{2}\right)^2} + \frac{z_{11} - z_{22}}{2} \\ Z_2 &= \sqrt{\frac{z_{22}}{y_{11}} + \left(\frac{z_{11} - z_{22}}{2}\right)^2} + \frac{z_{22} - z_{11}}{2} \end{aligned}\right. \tag{25}$$

[*General expression for the iterative impedances*]

2.9. Reversible Networks. The expressions for the network constants simplify very materially in the event that

$$Z_1 = Z_2$$

This is, actually, by far the most important case. It is usual and necessary for communication systems to have interchangeable terminations, and an iterative chain linking such terminations must present the same impedance at any junction in both directions. As a result, the two iterative impedances must coincide, and it makes no difference which way each network is facing.

In this case the common value of Z_1 and Z_2 is written Z_0 and called the *characteristic impedance*, and the network is often referred to as *symmetrical*. Because this word sometimes has other connotations, we shall use the word *reversible* instead. This term implies that the network can be inverted, or turned about, without effect on the transmission, which is true. The other designation is mathematically correct in that it refers to the matrix of the network coefficients; however, it would seem to indicate a symmetry of structure which the network often does not possess.

All work on transmission lines and much of the work relating to *lumped* networks consider as a starting point reversible networks in the iterative connection. For the sake of brevity, such networks, when so connected, will be referred to as *matched reversible networks* or simply *matched networks*. However, it must be understood that matched has a broader meaning than iteratively connected.

The measurable parameters reduce to two for reversible networks, in agreement with (20) and (21). They are customarily given as two impedances: the short-circuit impedance:

$$Z_{sc} = \frac{1}{y_{11}} = \frac{1}{y_{22}}$$

and the open-circuit impedance:

$$Z_{oc} = z_{11} = z_{22}$$

Equations (16) and (18) are now identical and may be written

$$Z_{sc} = Z_0 \tanh \theta \tag{26}$$

Similarly, (17) and (19) go into the common form

$$Z_{oc} = Z_0 \coth \theta \tag{27}$$

Dividing (26) by (27), we obtain

$$\tanh \theta = \sqrt{\frac{Z_{sc}}{Z_{oc}}} \tag{28}$$

[*Expression for transfer constant of reversible networks*]

(By the use of well-known identities correlating the hyperbolic functions, the general expression (24) readily reduces to the above form for the reversible network, as the reader will verify.) Multiplying (26) by (27), we have

$$Z_0 = \sqrt{Z_{sc}Z_{oc}} \tag{29}$$

[*Expression for characteristic impedance*]

2.10. Image Impedances, Maximum Power Transfer. The iterative chain has particular significance because it is the only possible connection which makes the voltage vector ratio equal to the current vector ratio and therefore permits the transfer constant to be uniquely defined. It is not, however, the only possible chain. The only requirement for chain connection, according to the definition given, is that each network in the chain must have terminations which depend on the network parameters in a prescribed manner.

If the network is so terminated that its *input* impedance is equal to the *source* impedance, while the *output* impedance equals the *load* impedance, it is said to be *image connected.* This condition is fulfilled for all members of an *image chain.* It follows that at all junctions of the image chain the impedance is the same in both directions (Fig. 8*a*).

If the networks are reversible the image chain and the iterative chain become coincident.

The importance of the image chain is brought out when we consider a chain of reversible networks, each network having an internal junction separating two equal halves, symmetrically placed (Fig. 8*b*). With regard to the outer junctions, the chain is both image and iterative. This is also true with regard to the inner junctions, which also separate the chain into reversible networks, provided the impedance at all the inner junctions is the same. But if both outer and inner junctions are considered as boundaries between individual networks, then the chain is no longer iterative; it is, however, still an image chain.

The image chain is more flexible than the iterative chain.

It permits nonreversible networks to operate between identical terminations. It further permits the *impedance level* to change from junction to junction, to suit the requirements of networks having special functions.

Consider, for example, an iterative, reversible chain between equal terminations, of characteristic impedance Z_0. It may be necessary to insert into the chain a reversible network of characteristic impedance Z_0'. Obviously, this cannot be done without additional insertions. Between the new network and the original ones, matching networks must be added whose function

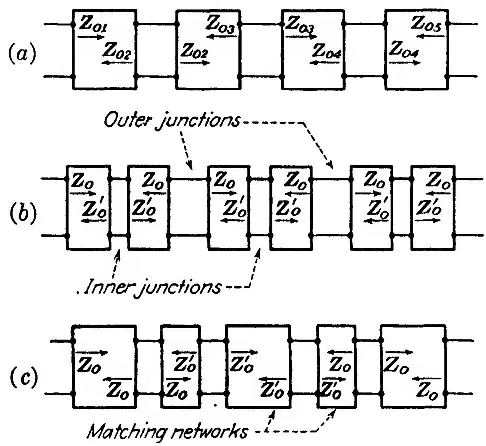


FIG. 8.—*Image chains.*

is that of providing correct terminations for both the new network and the others (Fig. 8*c*).

These matching networks obviously cannot be reversible, nor can they be iteratively connected. They must be selected so that their *image impedances* are equal to Z_0 and Z_0', respectively. and operate as parts of an image chain.

The image impedances are defined as those values which, when used as terminations for a given network at a given frequency, cause the impedance toward load to equal the impedance toward source, at both ends of the network.

For a reversible network, the image impedances coincide with the characteristic impedance Z_0. Nonreversible networks can be considered in conjunction with identical networks, symmetrically connected (Fig. 8*b*). In other words, regarding the nonreversible network as a half, the two possible whole networks (both reversible) are considered. Their characteristic impedances Z_{01} and Z_{02} are the image impedances of the half.

The transmission loss of the image-connected half network is equal to half the transmission loss of either whole. As for the insertion loss, this is seldom of importance, since insertion losses of individual networks do not add up to the over-all insertion loss, except in the case of iterative connection.

Image chains form the basis of conventional filter design.(5)

A third type of chain is obtained when the impedances in the two directions, at each junction, are not given by equal complex numbers but by *conjugate* numbers. As will be shown in connection with the coupling problem (Sec. 8.2), this is the condition for *maximum power transmission* through the junction. In other words, if an equivalent generator has impedance $R + jX$, it will transmit the highest possible power to a load $R - jX$. Chains which meet this requirement at all junctions may be called *maximum power transfer* chains.

Chains of this type may be realized only approximately, as they impose serious limitations upon the component networks (Sec. 9.2).

If a chain has resistive impedance of uniform value, looking in both directions at all junctions, it has the characteristics of all three types (iterative, image, and maximum power transfer). Transmission lines, under favorable conditions, are in this category.

In the next chapter, the basic theory of four-terminal networks, particularly that of reversible networks in matched operation, will be used as a tool for the analysis of transmission lines.

2.11. Applications and examples.

Iterative chains. A number of identical resistive L sections in cascade (Fig. 9) are to be used in a variable attenuator designed for one-way transmission. The load impedance is 600 ohms, resistive. Each section must provide 2 db attenuation.

Let R_s, R_p stand for the series and shunt arms of each section. The iterative impedance toward load, R_1, equals the input impedance of a section terminated in R_1. Hence

$$R_1 = R_s + \frac{R_1 R_p}{R_1 + R_p}$$

$$\therefore\ R_1^2 = R_s(R_1 + R_p)$$

The attenuation A must equal twenty times the log of the input to output voltage ratio. Hence

$$A = 20 \log \frac{R_1}{R_1R_p/(R_1 + R_p)}$$

$$\therefore R_1 = R_p\left(\log^{-1}\frac{A}{20} - 1\right)$$

The above may be written

$$R_p = \frac{R_1}{\log^{-1}(A/20) - 1}$$

Hence

$$R_s = R_1\left(1 - \frac{1}{\log^{-1}(A/20)}\right)$$

Numerically,

$$R_1 = 600 \qquad \log\frac{A}{20} = 1.26$$

$$R_p = 2{,}305 \text{ ohms} \qquad R_s = 123 \text{ ohms}$$

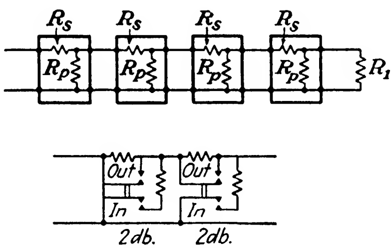


FIG. 9.—*Attenuator sections.*

Hence the arrangement of Fig. 9. Switches are designed to exclude sections; they short the series branch and open the shunt branch simultaneously.

Suggested Exercises. Find the iterative impedance toward source for the above sections (networks). Verify that

$$\frac{1}{R_2} - \frac{1}{R_1} = \frac{1}{R_p}$$

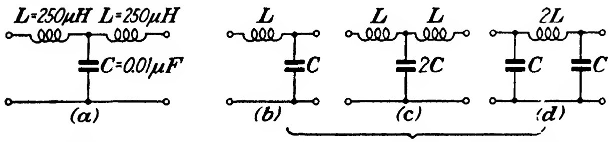
Design a chain of T sections, providing attenuation up to 100 db in 1-db steps, between 600-ohm terminations, using the least number of sections.

Network constants. Compute the constants of the (low-pass filter) T section of Fig. 10a.

The open- and short-circuit impedances are, by inspection,

$$Z_{oc} = j\left(\omega L - \frac{1}{\omega C}\right)$$

$$Z_{sc} = j\omega L + \frac{1}{j[\omega c - (1/\omega L)]}$$



w=100,000 r.p.s. (2.11.2.) L=0.1H C=0.1μF w=5000 r.p.s. (2.11.3.)

FIG. 10.—*Low-pass filter sections.*

To simplify the algebraic work, and for other reasons; it is convenient to use dimensionless parameters. Thus, the above may be written

$$Z_{oc} = jr\left(n - \frac{1}{n}\right)$$

$$Z_{sc} = jr\left(n - \frac{1}{n - (1/n)}\right)$$

having let

$$r = \sqrt{\frac{2}{C}} \qquad n = \omega\sqrt{2C}$$

Multiplying, clearing fractions, and taking the positive square root,

$$Z_0 = \sqrt{Z_{oc}Z_{sc}} = r\sqrt{2 - n^2}$$

Dividing, and proceeding likewise,

$$\tanh\theta = \sqrt{\frac{Z_{sc}}{Z_{oc}}} = \left|\frac{n}{n^2 - 1}\right|\sqrt{n^2 - 2}$$

Numerically,

$$L = 250\ \mu H \qquad C = 0.01\ \mu F \qquad \omega = 100{,}000 \text{ r.p.s.}$$

Hence

$$r = 158 \qquad n = 0.158$$
$$Z_0 = 222 \text{ ohms} \qquad \tanh\theta = j0.228$$

α and β may be computed by expanding $\tanh\theta$, as follows:

$$\tanh\theta = \frac{e^{2\theta} - 1}{e^{2\theta} + 1}$$

$$\therefore e^{2\theta} = \frac{1 + \tanh\theta}{1 - \tanh\theta}$$

Numerically,

$$e^{2\theta} = \frac{[1 + (0.228)^2]e^{j\tan^{-1} 0.228}}{[1 + (0.228)^2]e^{-j\tan^{-1} 0.228}} = e^{2j(\tan^{-1} 0.228)}$$

Hence

$$\theta = \alpha + j\beta = j \tan^{-1} 0.228$$
$$\therefore \alpha = 0$$
$$\beta = \tan^{-1} 0.228 = 12°50'$$

Actually, in this particular case this procedure is not necessary. Because of the identity

$$\tanh jx = j \tan x$$

the result could have been obtained by inspection. This is not true when α and β are both $\neq 0$.

Suggested Exercise. Assign arbitrary values to the network constants (preferably keeping the two iterative impedances equal), and solve for Z_{oc} and Z_{sc} by means of (26) and (27). Note that, in some cases, the results represent vectors in the second and third quadrants, corresponding to negative resistances. Barring such cases, investigate the field of variation for Z_0 and θ (see also group 6 in the Bibliography).

Image impedances. Find the image impedances of the L section in Fig. 10*b*.

The image impedances are found by taking the characteristic impedances of the T and Π sections formed by joining two identical L sections in the two possible symmetrical combinations.

For the T section (Fig. 10*c*) we may use the equation

$$Z_0 = r\sqrt{2 - n^2}$$

However, in place of n, we must write $n\sqrt{2}$, and in place of r, $r/\sqrt{2}$, if we want to preserve the notation

$$r = \sqrt{\frac{L}{C}} \qquad n = \omega\sqrt{LC}$$

where L and C are the values for the L section. Hence

$$Z_{01} = r\sqrt{1 - n^2}$$

The configuration of the Π section (Fig. 10*d*) is in every respect the dual to that of the T section. We can therefore use the last line for this case also; however, Z becomes Y, and L

becomes C; hence, r becomes $1/r$. Thus

$$\frac{1}{Z_{02}} = \frac{1}{r}\sqrt{1 - n^2} \qquad \therefore\ Z_{02} = \frac{r}{\sqrt{1 - n^2}}$$

Numerically

$$\begin{aligned} L &= 0.1 \text{ henry.} & C &= 0.1\ \mu & \omega &= 5{,}000 \text{ r.p.s.} \\ r &= 1{,}000 \text{ ohms} & n &= 0.5 \\ Z_{01} &= 867 \text{ ohms} \\ Z_{02} &= 1{,}150 \text{ ohms} \end{aligned}$$

Suggested Exercise. Consider an L section with series arm of impedance Z_a shunt arm of admittance Y_b. Obtain the image impedances in terms of Z_a and Y_b. Observe that the expressions for the image impedance at one end and the image admittance at the other are the same, except that Z_a takes the place of Y_b.

Note. The reader should bear in mind that, while there exists an equivalent T and an equivalent Π for any four-terminal network, in some cases these are not physically realizable. As for the equivalent L, this exists only in particular cases. L networks, although nonreversible, have only two independent constants (7).

CHAPTER III

NETWORK THEORY APPLIED TO UNIFORM NONREFLECTING TRANSMISSION LINES

3.1. The Section of a Long Line as a Four-terminal Network. The conclusions of the preceding chapters find immediate and useful application in the study of the transmission line, a finite section of which will now be considered in place of the generic four-terminal network.

In order to study the line, certain data must be taken as the starting point. These data must be connected with, and deducible from, the physical structure of the line (its length, the distance between conductors, the size and nature of the conductors and of the dielectric between them). Our first object is to define quantities suitable for this purpose. The relation between these and the line constants will be investigated later (Sec. 3.4).

3.2. Series Impedance. Shunt Admittance. It is impossible to draw a schematic representing a finite section of line because this can be likened to a network having an infinite number of meshes. A crude approximation (Fig. 11*a*) includes a finite number of meshes. Upon this approximate model may be based considerations which are valid regardless of the number of meshes and therefore apply to the line itself without loss of rigor.

Suppose, for instance, that all the *shunt* or transverse connections of Fig. 11*a* are severed. The line section degenerates into a simple series impedance connecting source and load, or adjoining sections (Fig. 11*b*). This is the *series impedance* of the line section. This impedance is obviously independent of the arbitrary number of meshes used in the model and is, also obviously, in direct proportion to the length of the section. Its value *per unit length* of line (*i.e.*, for a line section of unit length) is the *series impedance per unit length* of the line, Z_s. It may be defined as the impedance which would be measured across the input of a unit length section, if the output were shorted and if the distributed shunt admittance were removed.

Now suppose that all the series branches of Fig. 11*a* were shorted or replaced by direct connections. This reduces the section of line to a single shunt connection across both pairs of terminals (Fig. 11c). The admittance of such a connection will

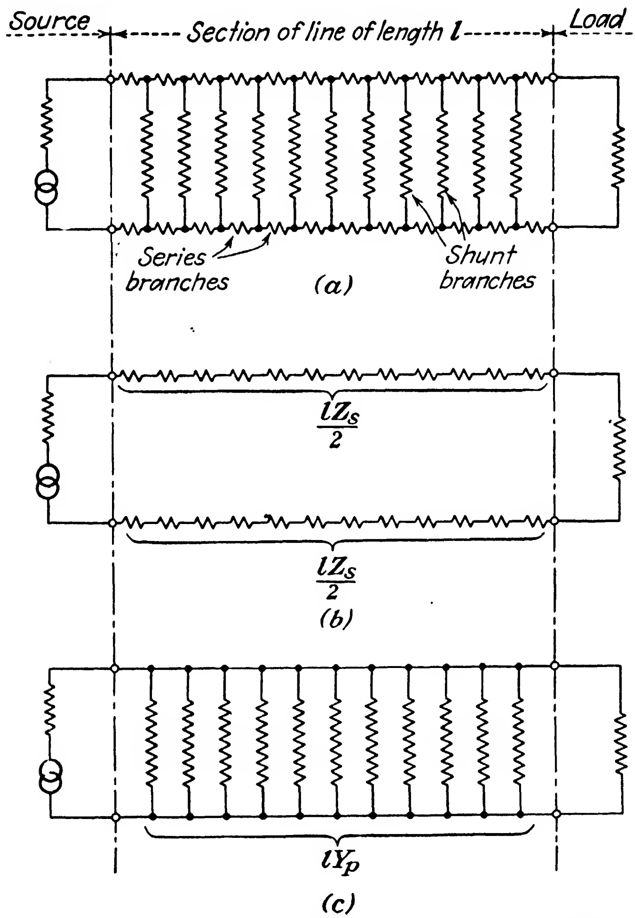


FIG. 11.—*Approximate representation of a section of line, showing significance of series impedence and shunt admittance.*

naturally be proportional to the length of the line section, and independent of the number of shunt branches in the model initially considered. It is the *shunt admittance* of the section. Its value *per unit length* is the *shunt admittance per unit length* of the line, Y_p. This may be defined as the admittance that could be measured across the input of a unit length section, if the output were open and the distributed series impedance replaced everywhere by impedanceless connections.

As has been pointed out, the quantities so defined would be valueless as a starting point if they could not be linked to the line's physical characteristics. The reader will appreciate that the above pictures in which series impedance and shunt admittance appear independently are far easier to analyze than the composite one where these two quantities appear together. Therefore, for the time being it is not necessary to go any further in the study of Z_s and Y_p, and these may be treated as known parameters.

3.3. Element of Line. The series impedance of a section of line of length Δl is clearly equal to $Z_s\,\Delta l$, and likewise $Y_p\,\Delta l$ is the shunt admittance. According to the preceding definitions, we could consider $Z_s\,\Delta l$ equal to the short-circuit impedance of the section, Z_{sc} (Sec. 2.9), if it were possible to assume that the shunt admittance of the short-circuited section had no effect and could be removed without any change in input current or voltage.

The short-circuited section is shown in Fig. 12*a*, again by an

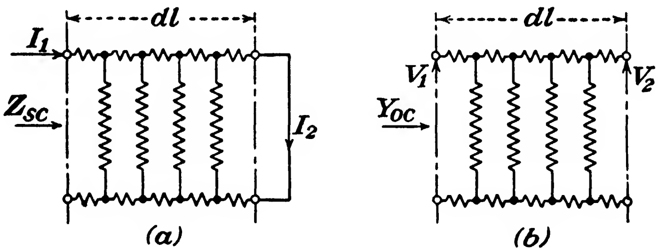


FIG. 12.—*Element of line.*

approximate representation with a limited number of transverse connections. It is evident that if the length Δl is small, these connections will absorb only a negligible fraction of the total input current, so that their severance has very little effect. Going to the limit, if the length becomes infinitesimal we are justified in saying that the shunt connections carry no current and may be neglected, hence

$$Z_{sc} = Z_s\,dl \tag{30}$$

for a section of infinitesimal length dl, or *element* of the line.

Considering Fig. 12*b*, we see that the drop in voltage from the input to the output terminals is negligible in a very short section. Hence all series branches have only a negligible voltage across them and may be short-circuited without appreciable change. In the limiting case of an infinitesimal section of length dl, or line element, we may say that the input admittance for short circuit is equal to the shunt admittance of the element, or

$$Y_{oc} = \frac{1}{Z_{oc}} = Y_p\, dl \tag{31}$$

3.4. Constants of the Line. A chain consisting of an infinite number of line elements connected in cascade will now be considered. Such a chain is equivalent to a section of line of finite length l.

If Z_0 is the characteristic impedance of a single element (Sec. 2.3), the characteristic impedance of the chain, or section of line, is also Z_0. If this value of impedance is used to terminate the section, this will operate as an output-matched chain of reversible networks having an infinite number of junctions. If any junction of such a chain is disconnected, the impedance on the load side of the junction is always Z_0. This value of impedance will be measured across any two opposite points of a line, if the line is cut at those points, and if the line is terminated in the same impedance at the far end. The impedance Z_0 is therefore called the *characteristic impedance of the line.* A line used to transmit signals to distant points should be terminated in its characteristic impedance; when this is the case the line is said to *match* the receiver or other load, and vice versa. Nonmatched operation will be taken up in Chap. V.

If the definition of transfer constant is called to mind (Sec. 2.2) it will be seen that the transfer constant of a chain of n identical networks is $n\theta$, if θ is the transfer constant of each. As there are $1/dl$ line elements of length dl in the unit length of line, if θ is the transfer constant of a unit length section, $\theta\, dl$ will be its value for the element.

On the basis of Eqs. (30) and (31), the values of image impedance Z_0 and transfer constant $\theta\, dl$ for an element of line will now be found. The first of these is also the *characteristic impedance* of the line; the second, divided by dl, yields the *transfer constant per unit length* of the line. Using Eqs. (28) and (29), we may write

$$
\begin{aligned}
Z_0 &= \sqrt{Z_{sc}Z_{oc}} = \sqrt{\frac{Z_s}{Y_p}} \\
\tanh\,(\theta\, dl) &= \sqrt{\frac{Z_{sc}}{Z_{oc}}} = dl\,\sqrt{Z_sY_p}
\end{aligned}
\tag{32}
$$

Note that in the above line, tanh $\theta\, dl$ is identical to $\theta\, dl$, this being an infinitesimal quantity, as can be quickly verified by examining the series expansion for hyperbolic tangent (Sec. 5.6). Accordingly, the line constants will now be rewritten in simpler form

$$Z_0 = R_0 + jX_0 = \sqrt{\frac{Z_s}{Y_p}} \tag{33}$$

[Characteristic impedance]

$$\theta = \alpha + j\beta = \sqrt{Z_sY_p} \tag{34}$$

[Transfer constant per unit length (mile) of line]

In the above expressions, using the mile as the unit of length, Z_s represents the series impedance and Y_p the shunt admittance per mile of line. α and β are the attenuation and phase shift per mile. Equation (33) tells how the line should be terminated and (34) gives the transmission characteristics per mile *for the correctly terminated line.* These equations are in vector form, and further analysis is necessary in order to make them directly useful for computation. The following sections will be devoted to this purpose.

3.5. Dissipation Factors. The series impedance and shunt admittance per unit length of line, Z_s and Y_p, which served as basis for the derivation of Eqs. (33) and (34), must now be reconsidered and expressed in terms of parameters having more direct engineering significance.

Letting L stand for the inductance per mile, which is therefore a constant, and R for the resistance per mile for both wires, we have

$$Z_s = R + j\omega L \tag{35}$$

For the value of L in terms of wire spacing, see Sec. 4.4.

In perfect analogy with the above, Y_p may be expressed in terms of G and C, conductance and capacitance per mile of line

$$Y_p = G + j\omega C \tag{36}$$

ω as usual stands for the angular frequency $2\pi f$. For values of C in terms of wire spacing, see Sec. 4.4.

Expressions for α, β, R_0, and X_0 (see Sec. 3.4) could be found in terms of R, L, G, and C and the results would be of value. However, these results would not lend themselves to convenient handling, except for a few cases of particular interest. For the results to have more significance, it is found, in this as in many other cases, that dimensionless parameters must be used.

There are other advantages in the use of parameters of this type; they generally have greater value in the comparison of systems whose electrical dimensions are widely different. The numerical value of such parameters is usually a good indication of the general character of the line, or any other system, to which they apply. Moreover, they eliminate confusion with regard to units and conversions.

Parameters of this type, when not in general use in the literature, will be introduced and explained as the need arises; this slight departure from current usage is justified, in the author's opinion, by the simpler expressions which result.

The well-known ratio Q, equal to X/R or B/G, is an example of a widely used dimensionless parameter. Later on we shall make a distinction between Q and Q_0, another parameter defining selectivity (Sec. 9.5). Although less popular than Q, its reciprocal, the *dissipation factor* d is also used on occasion. In the treatment of lines d is preferable, as it leads to simpler expressions.

Accordingly, let us define the following:

a. The series dissipation factor

$$d_s = \frac{R}{\omega L} \tag{37}$$

where R is the series resistance per unit length, L the series inductance; and

b. The shunt dissipation factor

$$d_p = \frac{G}{\omega C} \tag{38}$$

where G and C are the shunt conductance and capacity per unit length, respectively.

These two parameters are all that is needed to account for the *losses* present in the line.

3.6. The Lossless Line. Impedance Number. Time Lag and Frequency Number. Using the dissipation factors, Eqs. (33) and (34), giving values for the line constants, expand as follows:

a. The characteristic impedance:

$$Z_0 = \sqrt{\frac{Z_s}{Y_p}} = \sqrt{\frac{R + j\omega L}{G + j\omega C}}$$

hence

$$Z_0 = \sqrt{\frac{L}{C}} \sqrt{\frac{d_s + j}{d_p + j}} \tag{39}$$

b. The transfer constant (per unit length):

$$\theta = \sqrt{Z_s Y_p} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

hence

$$\theta = \omega \sqrt{LC} \sqrt{(d_s + j)(d_p + j)} \tag{40}$$

Let us now consider the values taken when d_s and d_p are negligibly small. Because in this case, as we shall see, the attenuation α vanishes, the line in question would be incapable of dissipating power and may be properly called a *lossless* line.

Often, a physical line may be considered lossless for practical purposes. Besides, the values obtained neglecting losses are useful for reference. By comparing the actual values to these, we obtain dimensionless parameters which can be used to some advantage.

The lossless value of Z_0 is a pure resistance. From (39)

$$Z_{0(d_s = d_p = 0)} = R_{0(d_s = d_p = 0)} = \sqrt{\frac{L}{C}} \tag{41}$$

Comparing the actual, or "lossy" value to the above, we are led to a pure number, usually complex, as follows:

Characteristic impedance number:[1]

$$z_0 = \frac{Z_0}{Z_{0(d_s = d_p = 0)}} = \sqrt{\frac{d_s + j}{d_p + j}} \tag{42}$$

Note that the above equation contains dimensionless quantities only, therefore lending itself well to graphical methods of evaluation (Sec. 3.4).

[1] Later, the term *impedance number* will be used in a somewhat different sense, to signify the ratio of a given impedance to the characteristic impedance of a line or network which is being studied.

The lossless value of the transfer constant is a pure imaginary. We have, in fact (from Eq. 40)

$$\theta_{(d_p=d_s=0)} = \omega\sqrt{LC}\sqrt{-1} = j\omega\sqrt{LC}$$

hence

$$\begin{aligned} \alpha_{(d_p=d_s=0)} &= 0 \\ \beta_{(d_p=d_s=0)} &= \omega\sqrt{LC} \end{aligned} \tag{43}$$

A lossless line, as we anticipated, does not attenuate but produces a *phase shift* of value $\omega\sqrt{LC}$ per unit length.

Since the voltage (or current) vector rotates uniformly with time at the angular velocity ω, the phase shift introduced by the line could be duplicated by "setting the clock back" a given number of seconds. We could say that the line *retards* the signal by this time interval, or that it produces a *time lag* of so many seconds. If steady state is assumed, this time lag is not particularly significant; it has importance only when there are time variations, which is true in practice, of course. One might object that the present method of analysis does not apply to changing conditions, except when they are very gradual (Sec. 1.6). There is no valid answer to this objection; however, we shall later confirm the results predicated on steady state by other methods not based on this hypothesis (Sec. 13.2).

The time lag caused by a mile of lossless line is readily expressed. We may argue that the time lag is to one second as the phase shift due to the mile of line is to the vector rotation during a second, ω. Hence

$$t = \frac{\beta}{\omega} \tag{44}$$

[*Time lag per mile of line*]

The above is general. Now for a *lossless* line we have the value of β given in (43). Hence

$$t_{(d_p=d_s=0)} = \frac{\beta_{(d_p=d_s=0)}}{\omega} = \sqrt{LC} \tag{45}$$

The reciprocal of the *time lag per unit length* is called *velocity of propagation,* although we cannot think of this quantity as a velocity in the ordinary sense unless we assume some kind of disturbance being propagated along the line, which again contradicts the hypothesis of steady state. For a lossless line, this velocity takes the value

$$\Omega = \frac{1}{\sqrt{LC}}$$

If, in addition to the absence of losses, we also assume absence of magnetic materials or of any dielectric other than gases or a vacuum and furthermore consider the currents to flow only on the *surface* of the conductors, Ω can be calculated by the methods of field theory (Chap. XIII) and turns out to be the same as the velocity of electromagnetic waves in space, or the velocity of light. For practical purposes, this is true of all physical open wire lines with reasonably small dissipation.

We shall find it convenient to express the ratio

$$n = \omega \sqrt{LC} = \frac{\omega}{\Omega} \tag{46}$$

which we shall call *frequency number*. For open wire lines, using the mile as the unit of length, n is approximately equal to

$$n = \frac{2\pi f}{186{,}000} = 33.8 \times 10^{-6} f$$

In terms of n, Eq. (40) may be written

$$\frac{\theta}{n} = \sqrt{(d_s + j)(d_p + j)} \tag{47}$$

The remainder of the chapter will be devoted to various methods for evaluating the line constants in cases when dissipation cannot be neglected.

★ 3.7. Expressions for the Line Constants in Terms of the Line Parameters d_s and d_p. To expand (42), we take the squares of both sides, rationalize, and separate into real and imaginary parts. Thus

$$\begin{cases} r_0^2 - x_0^2 = \dfrac{1 + d_p d_s}{1 + d_p^2} \\ r_0 x_0 = \frac{1}{2} \dfrac{d_p - d_s}{1 + d_p^2} \end{cases} \tag{48}$$

★ Sections marked by star are largely taken up by the mathematical steps leading to formulas useful in computation but not necessarily essential to a good understanding of the theory. The reader may omit such sections without loss of logical continuity.

Now we let, for brevity

$$A = \frac{1 + d_p d_s}{d_p - d_s}$$

and divide the second line into the first, obtaining

$$\frac{r_0}{x_0} - \frac{r_0}{x_0} = 2A$$

Now, solving for r_0/x_0,

$$\frac{r_0}{x_0} = A \pm \sqrt{1 + A^2} \tag{49}$$

Solving for x_0/r_0

$$\frac{x_0}{r_0} = -A \pm \sqrt{1 + A^2} \tag{50}$$

Multiplying (49) and (47), expanding, and taking the square root,

$$r_0 = \frac{1}{\sqrt{2}} \sqrt{\frac{\sqrt{(1 + d_p{}^2)(1 + d_s{}^2)} + 1 + d_p d_s}{1 + d_p{}^2}} \tag{51}$$

[*Characteristic resistance of the line*]

The uncertainty on signs is eliminated by making d_p and d_s vanish in the above expression and comparing the result with the expression derived from (44) by removing d_p and d_s in that equation. Now, multiplying (40) and (47),

$$x_0 = \pm \frac{1}{\sqrt{2}} \sqrt{\frac{\sqrt{(1 + d_p{}^2)(1 + d_s{}^2)} - 1 - d_p d_s}{1 + d_p{}^2}} \tag{52}$$

[*Characteristic reactance of the line*]

We cannot remove once for all the ambiguity in the sign of x_0. The minus sign applies when $d_s > d_p$, which is true in all practical cases.

The identical procedure is applied to Eq. (45). It leads to the following:

$$\frac{\alpha}{n} = \frac{1}{\sqrt{2}} \sqrt{\sqrt{(1 + d_p{}^2)(1 + d_s{}^2)} - 1 + d_p d_s} \tag{53}$$

[*Attenuation per unit length*]

$$\frac{\beta}{n} = \frac{1}{\sqrt{2}} \sqrt{\sqrt{(1 + d_p{}^2)(1 + d_s{}^2)} + 1 - d_p d_s} \tag{54}$$

[*Phase shift per unit length*]

In the above expressions (51) to (54):

α = attenuation, nepers per mile
β = phase shift, radians per mile
$n = \omega \sqrt{LC}$ = frequency number
$r_0 = R_0 \sqrt{C/L}$ = resistance number
$x_0 = X_0 \sqrt{C/L}$ = reactance number
$Z_0 = R_0 + jX_0$ = characteristic impedance
$d_s = R/\omega L$ = series dissipation factor
$d_p = G/\omega C$ = shunt dissipation factor
R, G, L, C = distributed resistance, conductance, inductance, capacity

(For simplified expressions valid in many cases, see Sec. 3.6.)

3.8. Mapping of Line Constants. Equations (51), (52), (53), and (54) provide means for obtaining the line constants, θ and Z_0, having measured or obtained in some way the parameters d_s and d_p and the frequency number n. The necessary computation, however, is laborious, since in general the constants have to be evaluated over a range of frequencies.

The line constants can be determined much more easily by the use of appropriate charts or *maps*. A visual appreciation of the quantities involved and their variation is also gained by means of such devices.

Consider a plane whose coordinates are $Q_s = 1/d_s$ and $Q_p = 1/d_p$. Each point of the plane corresponds to only one value for each of the quantities α/n, β/n. All points for which α/n has a particular value, joined together, define a constant α/n line. A family of such lines, drawn over a sector of the Q_s/Q_p plane and adequately spaced, constitutes a map of the quantity α/n. Other maps are similarly defined. These maps are obtained by manipulation of Eqs. (51) to (54).

Figure 13 shows maps of α/n and β/n. Both maps are symmetrical about the 45-deg. line through the origin; for this reason, each is drawn on one side of this line only, to avoid crowding. For the same reason, it does not matter whether Q_s is taken as the abscissa and Q_p as the ordinate, or vice versa. Each scale may be used indifferently for Q_s or Q_p, hence both the abscissa and ordinate scales are marked Q without a subscript. These scales are logarithmic, to ensure uniform accuracy and provide much wider ranges.

Once the point corresponding to a given line at a given frequency (the *point of operation*) is located on the map, the points of operation for other values of frequency become immediately available. Consider the point P of coordinates Q_s and Q_p. Both $Q_s = \omega L/R$ and $Q_p = \omega C/G$ are proportional to frequency (provided this does not exceed a certain value; roughly, this is true over the audio range). From the above expressions, we obtain

$$\frac{Q_s}{Q_p} = \frac{GL}{CR} \tag{55}$$

This is the equation of the locus described by P as the frequency is varied. It represents a straight line passing through the origin. This line is the *path of operation* for the line in question at variable frequency, on the Q_p, Q_s plane.

This construction is based upon the use of arithmetic (linear) scales for Q_p and Q_s. If logarithmic scales are used, the actual coordinates of the point of operation are $\log Q_s$ and $\log Q_p$. In this case the equation for the path of operation can be obtained by taking the logarithms of (55), as follows:

$$\log Q_s - \log Q_p = \log \frac{GL}{CR} \tag{56}$$

Equation (56) represents, on the plane $\log Q_p$ vs. $\log Q_s$, a straight line with a positive slope of 45 deg. Upon this path of operation can be traced a frequency scale, which is the same for all possible conditions, being logarithmic and having projections upon the axes of coordinates entirely similar to the scales used for Q_s and Q_p.

This frequency scale accompanies the maps of Figs. 13 and 14. By the use of a simple procedure, described in the legend accompanying the maps, the required constants can be obtained for many values of frequency in one operation.

With reference to the maps of Fig. 14, these include lines of constant $|z_0|$ and of constant $\underline{/z_0}$. These two symbols stand for the magnitude and angle of the impedance number

$$z_0 = \frac{Z_0}{\sqrt{L/C}}.$$

These lines have been drawn on separate halves of the sheet, just as in the case of α/n and β/n (Fig. 13), although they are

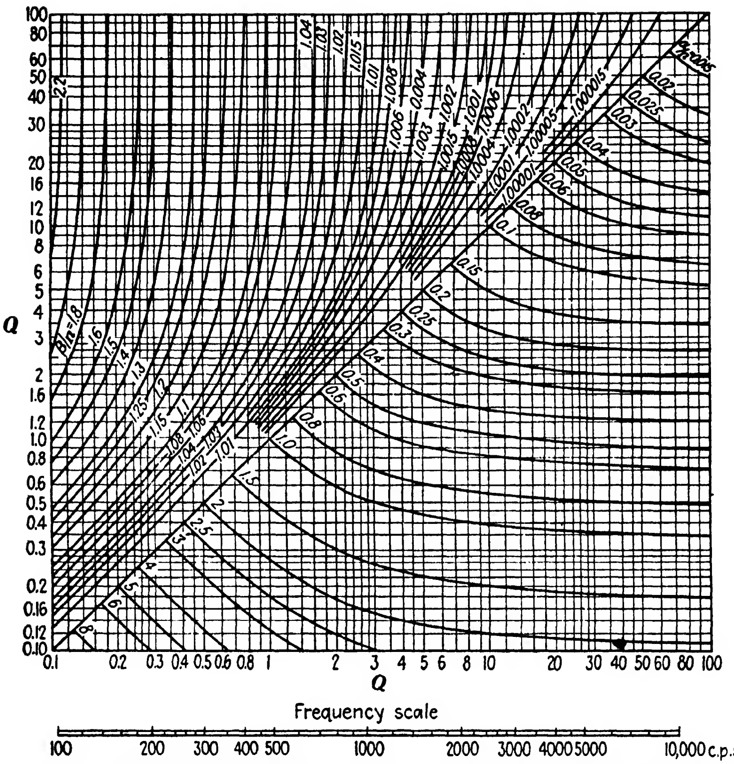
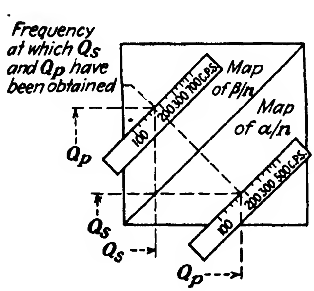


FIG. 13.—*Attenuation and phase-shift maps for matched transmission lines.*



Symbols used:
α = attenuation, nepers/mile
β = phase shift, radians/mile
$n = \omega\sqrt{LC}$ = frequency number
Q stands for either $\begin{cases} Q_s = \omega L/R \\ \text{or } Q_p = \omega C/G \end{cases}$
L, R, C, G = inductance, capacity, resistance, conductance per mile.

To read α/n and β/n over a range of frequencies, copy as much of frequency scale as needed on a straight edge. Transfer this upon maps in both positions shown on left. Frequency scale is logarithmic, valid only up to upper limit of audio range. For examples, see Sec. 4.5.

not symmetrical about the $Q_s = Q_p$ line. However, we note from Eq. (44) that substitution of d_s for d_p and vice versa changes z_0 to its reciprocal.

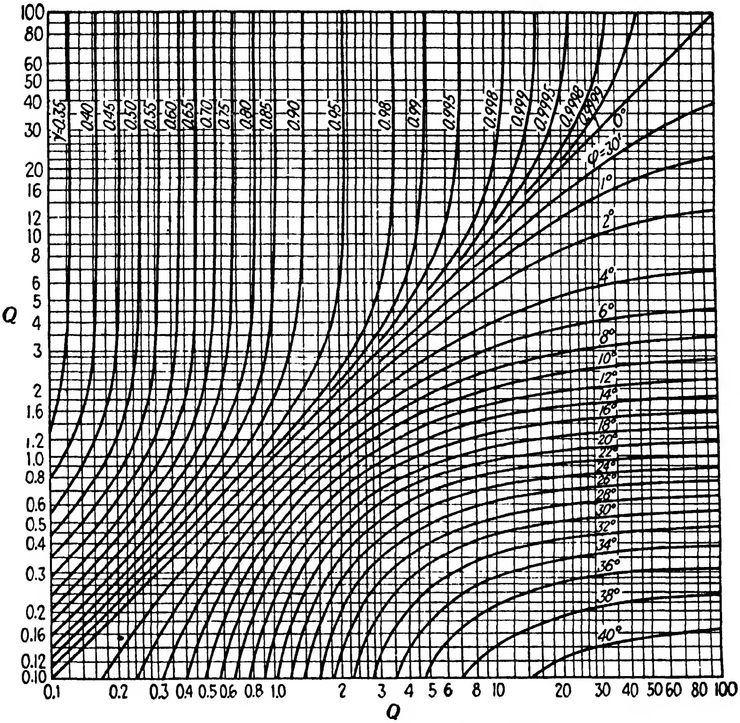


FIG. 14.—*Impedance map for matched transmission lines.*

Magnitude of impedance number $\mathfrak{z}_0 = Z_0 \sqrt{\frac{C}{L}}$ is given by $\begin{cases} \gamma \text{ when } Q_s > Q_p \\ 1/\gamma \text{ when } Q_p > Q_s \end{cases}$

Angle of characteristic impedance Z_0 is given by . . . . $\begin{cases} \varphi \text{ when } Q_s > Q_p \\ -\varphi \text{ when } Q_p > Q_s \end{cases}$

Q stands either for $Q_s = \omega L/R$, or $Q_p = \omega C/G$

For other symbols and for use of frequency scale, see Fig. 13. (The same frequency scale can be used.) For examples, see Sec. 4.5 and Fig. 13.

Hence, interchange of Q_s with Q_p will change $|z_0|$ to $\frac{1}{|z_0|}$ and $\underline{/z_0}$ to $-\underline{/z_0}$. This makes it necessary to interpret the values obtained according to the key which accompanies the maps.

3.9. Example of Line Constant Evaluation. In the following example, we will suppose that the distributed series impedance and shunt admittance have been measured on an open wire line (see also Sec. 4.4). The line parameters will be computed, and, by means of the maps described in the preceding section, the line constants will be evaluated over the lower end of the audio-frequency range. Plots of the constants against frequency appear on Fig. 15.

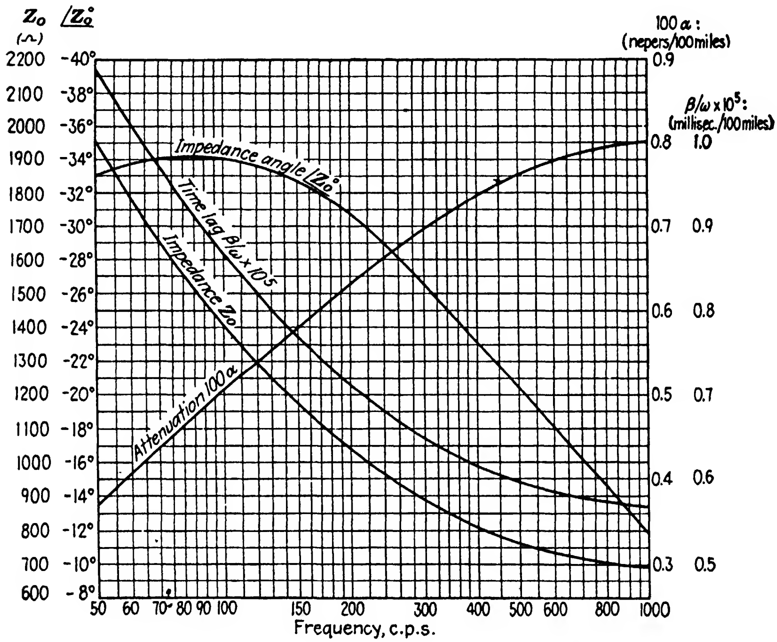


FIG. 15.—*Constants of a* 0.104-*in. open wire line.*

Open wire line of 0.104-*in. wire.* The following data are given:
Resistance per mile $R = 10.4$
Inductance per mile $L = 3.67$
Conductance per mile $G = 0.8$
Capacitance per mile $C = 0.00835$
Computation of Q_s and Q_p at 1,000 c.p.s.:

$$Q_s = \frac{\omega L}{r} = \frac{1{,}000 \times 2\pi \times 3.67 \times 10^{-6}}{10.4} = 2.22$$

$$Q_p = \frac{\omega C}{g} = \frac{1{,}000 \times 2\pi \times 0.00835 \times 10^{-6}}{0.8 \times 10^{-6}} = 65.5$$

Computation of $\sqrt{LC}$ and $\sqrt{\frac{L}{C}}$:

$$\sqrt{LC} = \sqrt{3.67 \times 10^{-3} \times 0.00835 \times 10^{-6}} = 5.53 \times 10^{-6}$$

$$\sqrt{\frac{L}{C}} = \sqrt{\frac{3.67}{0.00835} \times 10^{3}} = 663$$

TABULATION OF VALUES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Quantity	f	$100n$	α/n	β/n	γ	φ	100α	$\beta/\omega \times 10^5$	$\lvert Z_0\rvert$	$/Z_0$
Units	c.p.s.					deg.	nep. 100 miles	millisecs 100 miles	Ω	deg.
Description of quantity and method of obtaining		$100n = 100w\sqrt{LC} = 2\pi \times f \times 5.53 \times 10^{-4}$	From map, Fig. 13	From map, Fig. 13	From map, Fig. 14	From map, Fig. 14	Attenuation per 100 miles $= (\#2) \cdot (\#3)$	Time lag per 100 miles $= \beta/n \times \sqrt{LC} \times 10^5 = (\#4) \times .553$	Characteristic impedance $= \sqrt{\frac{2}{C}}\frac{1}{\gamma} = \frac{663}{(\#5)}$	Characteristic impedance angle $= -\varphi = -(\#6)$
	50	.174	2.25	1.96	.34	33	.392	1.085	1950	−33
	70	.2435	1.8	1.76	.4	34	.438	.973	1660	−34
	100	.348	1.45	1.57	.47	34	.505	.868	1410	−34
	140	.486	1.17	1.41	.56	33	.569	.78	1185	−33
	200	.695	0.9	1.285	.64	30.7	.625	.71	1035	−30.7
	300	1.043	0.67	1.17	.75	26.5	.7	.647	884	−26.5
	400	1.39	0.52	1.12	.82	22.8	.723	.62	808	−22.8
	500	1.74	0.44	1.076	.87	20	.765	.595	762	−20
	700	2.435	0.32	1.043	.915	15.5	.779	.577	724	−15.5
	1000	3.48	0.23	1.021	.955	11.8	.8	.567	694	−11.8

These results are plotted in the curves of Fig. 15. The significance of the time lag β/ω is explained in Sec. 4.2.

★ 3.10. Lines of Low Parallel Dissipation. No assumption has been made, so far, regarding the values of d_s and d_p as they occur in most practical cases. In the next chapter this matter will be taken up in some detail; however, it is convenient to note at this point that the following conditions are fulfilled in a large number of situations:

First condition:

$$d_s \gg d_p \tag{57}$$

Second condition:

$$d_p \ll 1 \tag{58}$$

In order to assign a definite meaning to the above inequalities, let it be assumed that $x \ll 1$ when $x^2 < 0.01$, or, in other words, when x^2 can be neglected with respect to unity with an error of 1 per cent or less.

Conditions (57) and (58), thus defined, hold with the following exceptions:

Condition (57) does not generally hold for loaded open wire lines if treated on the assumption of uniform loading (Sec. 4.4).

Condition (58) does not generally hold below frequencies of 150 c.p.s. (open wire) and 25 c.p.s. (cables).

Consider the form that Eqs. (51), (52), (53), and (54) take when the above conditions are met and errors of less than about 1 per cent are tolerated. Take, for example, Eq. (53),

$$\frac{\alpha}{n} = \frac{1}{\sqrt{2}} \sqrt{\sqrt{(1 + d_p^2)(1 + d_s^2)} - 1 + d_p d_s}$$

We may write

$$\sqrt{1 + d_p^2} = 1 + e \tag{59}$$

e being approximately equal to $d_p^2/2$, first term of the series expansion. Substituting in (53),

$$\frac{\alpha}{n} = \sqrt{\frac{1}{2}\left(\sqrt{1 + d_s^2} - 1 + e\sqrt{1 + d_s^2} + d_s d_p\right)}$$

It is convenient to introduce the following quantity, function of d_s only

$$S_1 = \sqrt{\frac{\sqrt{1 + d_s^2} - 1}{2}} \tag{60}$$

whence

$$\sqrt{1 + d_s^2} = 2S_1^2 + 1$$

and

$$\frac{\alpha}{n} = S_1 \sqrt{1 + e + \frac{e}{2S_1^2} + \frac{d_s d_p}{2S_1^2}} \tag{61}$$

Noting that all the terms beneath the radical of the above are positive, we may neglect the term e with very small error [see condition (58)]. Furthermore, since e and $d_p^2/2$ are approximately equal, the term $e/2S_1^2$ is also negligible if we can show the following term to be negligible:

$$\frac{d_p^2}{4S_1^2} = \frac{d_p^2}{2(\sqrt{1 + d_s^2} - 1)}$$

For any given d_p, this term decreases with increasing d_s. Therefore, if the term can be neglected when $d_s = 10d_p$, it can also be neglected when $d_s > 10d_p$; hence, from (57), if the term is negligible for $d_s = 10d_p$, it is negligible for all values of d_s consistent with the premises and with the given d_p. Letting $d_s = 10d_p$, we have

$$\frac{d_p^2}{4S_1^2} = \frac{d_p^2}{2(\sqrt{1 + 100d_p^2} - 1)}$$

As d_p takes all values consistent with (58), the above varies between 0.01 and approximately 0.012. This last figure is the highest possible value of the term $e/2S_1^2$, and, taking for the term e its highest possible value, which is approximately 0.005, we find that in the most unfavorable case $(d_s d_p/2S_1^2 = 0)$. The omission of both terms introduces an error of about 0.85 per cent. We may, therefore, without exceeding 1 per cent error, write

$$\frac{\alpha}{n} = S_1 \sqrt{1 + \frac{d_s d_p}{2S_1^2}} \tag{62}$$

Considering now the term $d_s d_p/2S_1^2$, we note that, for a given d_p, this term is largest when the expression

$$\frac{d_s}{2S_1^2} = \frac{d_s}{\sqrt{1 + d_s^2} - 1}$$

has the greatest value. This expression depends on d_s only and has ∞ for a limit as d_s approaches 0. The expression can easily be shown to increase as d_s decreases for all positive values of d_s. Hence it will have its highest value, consistent with our premises and with the given d_p, when d_s has its lowest value, which is $d_s = 10d_p$. In this case we have

$$\frac{d_s d_p}{2S_1^2} = \frac{10d_p^2}{\sqrt{1 + 100d_p^2} - 1}$$

As d_p takes all permissible values, this quantity ranges between 0.2 and 0.24. It cannot, therefore, be neglected as a rule, although in many cases its value is much smaller. However, since this value is never greater than unity, (62) can be expanded in a series as follows:

$$\frac{\alpha}{n} = S_1 \left\{ 1 + \frac{d_s d_p}{4S_1^2} - \frac{d_s^2 d_p^2}{32S_1^4} + \cdots \right\}$$

The highest possible value of the third term of this series is

$$\frac{(0.24)^2}{8} = 0.0072$$

Omission of this term introduces an error never greater than 0.72 per cent. It should be noted that this error is in the opposite sense to the error of less than 0.85 per cent previously introduced, and both errors reach their peak for the same conditions; hence we may, with a total error probably much less than 1 per cent, write the final equation

$$\frac{\alpha}{n} = S_1 + \frac{d_s d_p}{4S_1}$$

Reasoning along similar lines, Eqs. (51) to (54) can be put in the approximate forms tabulated below. Their graphical counterpart is the curve of Fig. 16, accompanied by a self-explanatory procedure for quickly finding the constants of lines or cables applicable in the vast majority of cases. The terms $d_s d_p/4S_1$ and $d_s d_p/4S_2$, generally small, are allowed for in the form of an additive correction.

The expressions given below are correct within 1 per cent provided $d_s > 10d_p$, $d_p < 0.1$.

$$\frac{\alpha}{n} = S_1 + \frac{d_s d_p}{4S_1} \tag{63}$$

[*Attenuation*]

$$x_0 = -S_1 + \frac{d_s d_p}{4S_1} \tag{64}$$

[*Characteristic reactance*]

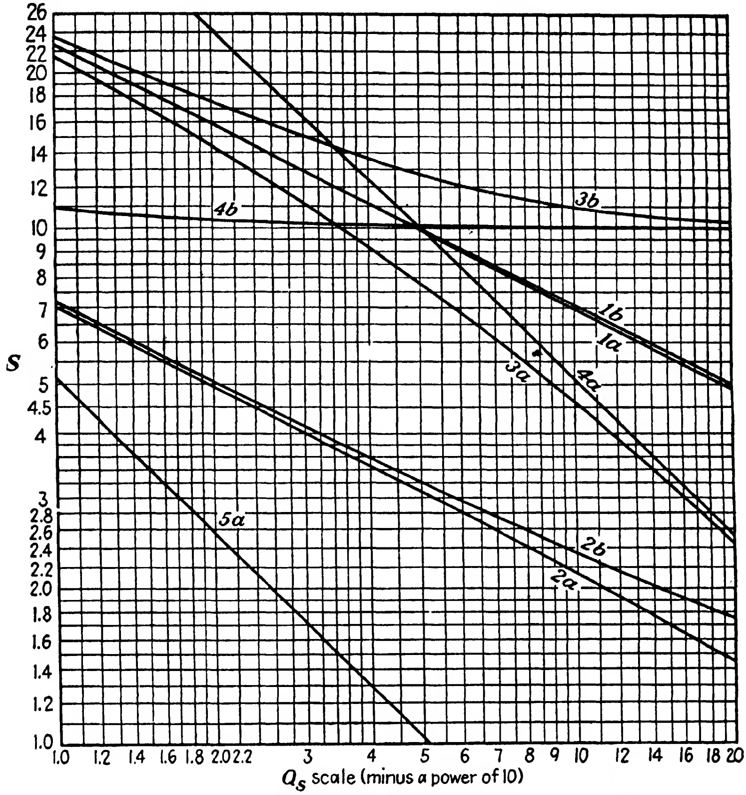


FIG. 16.—*Constants of lines with low shunt dissipation.*

Q_s range:	To find α or X_0: 1. Read S on curve No.	2. Divide S by	3. Add the correction*	4. For α (nep./mile) multiply by	5. For X_0 (ohms) multiply by	To find β/ω or R_0: 1. Read S on curve No.	2. Divide S by	3. Subtract the correction*	4. For β/ω (μ sec/mile) multiply by	5. For R (ohms) multiply by
0.001 to 0.02	1a	1	$\frac{0.25}{SQ_pQ_s}$	$n=2\pi f\sqrt{LC}$	$-\sqrt{\frac{L}{C}}$	1b	1	$\frac{0.25}{SQ_pQ_s}$	$\sqrt{LC}\times 10^6$	$\sqrt{\frac{L}{C}}$
0.01 to 0.2	2a	1	$\frac{0.25}{SQ_pQ_s}$	$n=2\pi f\sqrt{LC}$	$-\sqrt{\frac{L}{C}}$	2b	1	$\frac{0.25}{SQ_pQ_s}$	$\sqrt{LC}\times 10^6$	$\sqrt{\frac{L}{C}}$
0.1 to 2	3a	10	$\frac{2.5}{SQ_pQ_s}$	$n=2\pi f\sqrt{LC}$	$-\sqrt{\frac{L}{C}}$	3b	10	$\frac{2.5}{SQ_pQ_s}$	$\sqrt{LC}\times 10^6$	$\sqrt{\frac{L}{C}}$
2 to 20	4a	100	$\frac{25}{SQ_pQ_s}$	$n=2\pi f\sqrt{LC}$	$-\sqrt{\frac{L}{C}}$	4b	10	$\frac{2.5}{SQ_pQ_s}$	$\sqrt{LC}\times 10^6$	$\sqrt{\frac{L}{C}}$
10 to 50	5a	100	$\frac{25}{SQ_pQ_s}$	$n=2\pi f\sqrt{LC}$	$-\sqrt{\frac{L}{C}}$	.. $\beta/\omega = \sqrt{LC}\times 10^6\ \mu$ sec/mile; $R_0 = \sqrt{\frac{L}{C}}\ \Omega$				

* Do not apply correction if less than 1 per cent of total. Method is correct to 1 per cent except for frequencies below 150 c.p.s. (open wire) or 25 c.p.s. (cable) and for loaded open wire lines.
See example, Sec. 3.11.

$$\frac{\beta}{n} = S_2 - \frac{d_s d_p}{4S_2} \tag{65}$$

[*Phase shift*]

$$r_0 = S_2 + \frac{d_s d_p}{4S_2} \tag{66}$$

[*Characteristic resistance*]

In the above expressions

$$S_1 = \sqrt{\frac{\sqrt{1 + d_s^2} - 1}{2}} \qquad S_2 = \sqrt{\frac{\sqrt{1 + d_s^2} + 1}{2}}$$

All other symbols are as for Eqs. (51) to (54).

3.11. Example of Line Constant Evaluation for a Line of Low Shunt Dissipation. The following values have been obtained by direct measurement on a 22 A.W.G. cable (see Sec. 3.12):

$$Q_s = 0.0293 \qquad \omega = 5{,}000 \text{ r.p.s.}$$

$$Q_p = 209 \qquad \omega\sqrt{LC} = n = 0.0427$$

$$10^6 \times \sqrt{LC} = 8.54$$

$$\sqrt{\frac{L}{C}} = 117 \text{ ohms}$$

The constants can be found on the curves of Fig. 16 as follows:

a. To find α and X_0 (attenuation and characteristic reactance)

1. Q_s is in the range from 0.01 to 0.2; hence S_1 is read on curve 2*a*

$$S_1 = 4.08$$

2. No division is necessary

3. The correction is

$$\frac{1}{4 \times 4.08 \times 0.0293 \times 209} = 0.01$$

The correction is therefore negligible compared with S_1 and need not be added.

4. The attenuation is as follows:

$$\alpha = 4.08 \times n = 4.08 \times 0.0427 = 0.1745 \text{ nep./mile}$$

5. The characteristic reactance is

$$X_0 = -4.08 \times \sqrt{\frac{L}{C}} = -4.08 \times 117 = -477 \text{ ohms}$$

b. To find β/ω and R_0:

1. Read S_2 on curve 2*b*

$$S_2 = 4.2$$

2. No division necessary
3. Correction negligible
4. The time lag in microseconds per mile shows the following result:

$$10^6 \times \frac{\beta}{\omega} = 4.2 \times 10^6 \times \sqrt{LC} = 4.2 \times 8.54 = 35.9\ \mu\ \text{sec./mile}$$

5. The characteristic resistance R_0 is

$$R_0 = 4.2 \times \sqrt{\frac{L}{C}} = 4.2 \times 117 = 491 \text{ ohms}$$

Note. The approximate cable formulas (Sec. 4.5) could have been used in the above example with small error (about 1.4 per cent). For higher frequencies and for lines of other types the error would be large.

3.12. Measurement of Line Parameters. It will be noted that the four real constants of a line, α, β, R_0, and X_0, are fully established, through Eqs. (51) to (54) or (63) to (66), in terms of frequency and of the following four parameters:

$d_s = R/\omega L =$ series dissipation factor
$d_p = G/\omega C =$ shunt dissipation factor
$L =$ distributed inductance per mile of length
$C =$ distributed capacitance per mile of length

In general, six real quantities identify a network; but a mile of line is a reversible network (29); hence, four parameters are sufficient.

The four parameters may be obtained analytically from the physical dimensions; however, considerable mathematical difficulty may be encountered, especially in the case of d_s and d_p.

On the other hand, a knowledge of the parameters is not required if the constants are measured directly, or if the open- and short-circuit impedances of a long section are known with good accuracy, in which case the general equations for the constants of a reversible network can be used (Sec. 2.9). The difficulty in this case is experimental; long sections are needed, as well as special equipment.[6]

It is often convenient to determine the parameters by measurements which can be performed easily on short sections and then substitute the values in the equations or use the corresponding charts, such as those of Figs. 13, 14, and 16.

It was pointed out (Sec. 3.3) that the impedance of an element of line of length dl, short-circuited at the far end, is (30)

$$Z_{sc} = Z_s \, dl$$

Hence, and from (35)

$$Z_{sc} = (R + j\omega L) \, dl$$

The Q factor of the short-circuited element, Q_{sc}, is the ratio X/R for the element, hence

$$Q_{sc} = \frac{\omega L}{R} = \frac{1}{d_s}$$

It is impossible, of course, to perform measurements on infinitely short elements. However, if the element is short compared to the wavelength used (less than about 2 per cent of the wavelength) it can be shown that (30) still holds with error less than 1 per cent (Sec. 5.8). Hence, d_s can be measured conveniently up to frequencies in the neighborhood of 50 or 100 mc. At higher frequencies, several measurements can be taken for decreasing length and the curve extrapolated to zero length.

The factor $Q_{oc} = 1/d_p$ can be measured in the same way whenever necessary, except that the far end of the section under test is kept open (disconnected).

The inductance per unit length, L, can be obtained at the same time as Q_{sc}; the capacitance C is obtained together with Q_{oc}.

These measurements are conveniently performed on a commercial Q meter. The procedure is exactly the same as if the short-circuited section of cable were a coil of which the inductance and Q factor are required and the open-circuited section, a condenser whose capacity and Q factor are to be measured.

The values of L and C obtained in this measurement are, of course, those for a section of length l (whatever length is used). The values per unit length result when we divide the measured values by l.

At the lower audio frequencies the measurement of Q and L is apt to present difficulties, especially for pairs in cable, because of the large capacitors required to resonate the small inductance of the section. At these frequencies accurate results may be obtained by simply using the d-c values of R and G and computing L and C by means of Eqs. (76) and (77). At ultrahigh frequencies, on the other hand, the method discussed in Sec. 10.9 may be used.

The literature should be consulted[7] for general information regarding the measurement of inductance, capacitance, and Q.

CHAPTER IV

DISTORTION IN TRANSMISSION LINES

4.1. Voltages and Currents Along the Matched Line. The problem of establishing the over-all performance of a matched line has been reduced, in the preceding two chapters, to the determination of the line dissipation factors and characteristic frequency. We have as yet learned nothing, however, about those aspects of line transmission which can only be understood through a study of conditions all along the line rather than at the terminals exclusively.

To this end, let us obtain, for a matched line, expressions giving the instantaneous voltage and current at any point.

This purpose can best be achieved by plotting instantaneous values of voltage against distance for two consecutive values of time. Such a plot will indicate how the voltage (or the current) is distributed along the line at any instant, and how this distribution varies with time. The familiar mechanical concept of *wave propagation* will emerge from this discussion. Voltage only will be considered because the current is proportional to the voltage and in constant phase relation to the voltage everywhere, their ratio being the characteristic line impedance Z_0.

The voltage at a point x miles from the sending end will be (Secs. 2.2 and 3.4)

$$\mathbf{V}_x = \mathbf{V}_1 e^{-\theta x} = \mathbf{V}_1 e^{-(\alpha + j\beta)x} \tag{67}$$

Time will be counted from the instant when the input voltage V_1 is at a positive maximum. Then

$$\mathbf{V}_1 = V_1 e^{j\omega t}$$
$$\therefore \mathbf{V}_x = V_1 e^{-\alpha x - j(\beta x - \omega t)} \tag{68}$$

This expression gives V_x for any distance and time. The instantaneous value v_x is the real part of (68). Hence, recalling that $e^{j\theta} = \cos\theta + j\sin\theta$,

$$v_x = V_1 e^{-\alpha x} \cos(-\beta x + \omega t) \tag{69}$$

Let t have some particular value, 0 for instance. At this time we have

$$v_x = V_1 e^{-\alpha x} \cos(-\beta x) \tag{70}$$

which is an expression of instantaneous voltage in terms of distance. Its plot is a cosine wave of amplitude decreasing exponentially towards the receiving end (Fig. 17). $-\beta x$ is the phase of the voltage, which is 0 at the transmitter: by definition,

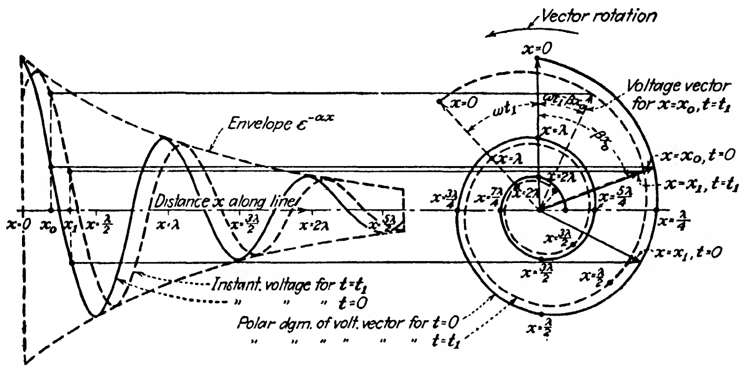


FIG. 17.—*Voltage distribution along the matched line.*

it becomes -2π at a point λ miles from the transmitter, λ being the *wavelength* in miles. Hence

$$\lambda = \frac{2\pi}{\beta} \tag{71}$$

Now consider a point at the distance x_0 from the transmitter. Its voltage has the phase $-\beta x_0$. If an interval t_1 is allowed to elapse, the whole distribution of Fig. 17 will have changed and in particular the phase of point x_0 will have changed. But there will exist another point x_1, whose voltage will now have the phase previously associated with x_0, indicating that the wave has traveled a distance $x_1 - x_0$. To express x_1 in terms of x_0, we must state that the phase at x_0 for time 0 equals the phase at x_1 for time t_1, or

$$-\beta x_0 = -\beta x_1 + \omega t_1$$

From this we have

$$\frac{x_1 - x_0}{t_1} = \frac{\omega}{\beta}$$

showing that ω/β is the ratio between the travel of the wave

$x_1 - x_0$ and the time elapsed t_1. This quantity is called the *wave velocity*, or *velocity of propagation.*

4.2. Phase and Amplitude Distortion. The wave velocity acquires particular significance when steady state (which has heretofore been assumed) suffers a change. We will suppose that the signal (a steady sine-wave voltage) suddenly changes at the transmitter by a given amount, and thereafter remains at the new value. If, at the instant of the change, the phase angle of the sending end voltage was Φ, we may conclude that the change will be felt at any particular point down the line, when the voltage *at that point* reaches the same phase angle Φ, which will be after some time, as was shown. We may say that the change is *propagated,* or travels down the line with the wave velocity ω/β, although the expression for this velocity was derived on the premise of steady state, which is inconsistent with the concept of change.

The interval $(\beta l/\omega)$, representing the time elapsed while a disturbance travels the length l, is the time lag for this length of line. $\beta/\omega \times 10^5$ is the time lag in milliseconds per 100 miles (see Fig. 13).

If the signal has complex wave form or consists of several superimposed voltages of different frequencies,[(1)] a change in signal strength or amplitude will affect all those frequencies. If the wave velocity, hence the time lag, is different for each frequency, at the receiver the signal components will change amplitude at different times, instead of simultaneously. For example, the high note of a chord may last after the low note has stopped.

This evidently constitutes a serious disadvantage in the transmission of intelligence over long lines or cable. The resulting effect is called *phase distortion.* To be free of phase distortion, a line must be so constructed that its wave velocity is constant with frequency.

Likewise, if the attenuation varies with frequency, the mutual *proportion* of the components of a sound will be affected by transmission, giving rise to *amplitude distortion.* This is sometimes called *frequency distortion.*

A completely distortionless line will, in consequence, fulfill the following two conditions:

1. ω/β constant with frequency (eliminating phase distortion)

2. α constant with frequency (eliminating amplitude distortion)

4.3. Distortionless Line. If d_p and d_s have a common value d Eqs. (53) and (54) simplify to the following:

$$\frac{\alpha}{n} = d \qquad \frac{\beta}{n} = 1 \tag{72}$$

In terms of ω and distributed line constants, if

$$\frac{R}{L} = \frac{G}{C} \tag{73}$$

then

$$\alpha = R\sqrt{\frac{C}{L}} = G\sqrt{\frac{L}{C}} \qquad \lambda = \frac{1}{f\sqrt{LC}} \qquad \frac{\omega}{\beta} = \Omega = \frac{1}{\sqrt{LC}} \tag{74}$$

It is clear that any line for which Eq. (73) holds within the operating frequency range is free of both amplitude and phase distortion, as it fulfills the conditions of the preceding section (α constant, ω/β constant). The characteristic impedance of such a line is purely resistive, as Eqs. (51) and (52) reduce to the following:

$$R_0 = \sqrt{\frac{L}{C}} \qquad x_0 = 0 \tag{75}$$

A distortionless line can be realized by properly selecting the constants, as (73) indicates, provided that it can be assumed that resistance and conductance do not change with frequency.

4.4. Open Wire Line. Summing up the conclusions of the preceding sections, we may say that to secure satisfactory transmission the two dissipation factors d_s and d_p should be as nearly equal as possible, and both should be low. As in many similar situations, the attainment of these conditions conflicts with economic necessity. We will now investigate what the dimensions of a line should be in order to secure theoretical freedom from distortion. The scope of the discussion will be limited to open wire lines at audio frequencies; as regards cables, of course, the dimensions are hardly subject to choice, and for open wire lines at high frequencies, the above criterion does not apply.

The condition for absence of distortion (Eq. 73) involves R, L, G, and C. In the audio range these are all independent of frequency. They are controlled by the wire diameter, the spacing between wires, the wire material, and the nature of the dielectric. For open wire lines, the dielectric is air, and we will assume the material to be pure copper. Given the character of this discussion, exact data are not required.

The inductance of a mile of line, defined in Sec. 3.5, is given in millihenrys as follows:[6]

$$L = 1.481 \log \frac{b}{a} + L_0 \text{ mh} \tag{76}$$

At audio frequencies L_0 has the value of 0.16. This value decreases for increasing frequencies.

The capacitance, also defined in Sec. 3.5, has the value[6]

$$C = \frac{0.01941}{\log b/a} \mu\text{f} \tag{77}$$

In both these expressions, b is the spacing between wire centers and a is the wire radius.

The resistance per mile of a pair of copper wires of radius a in millimeters, at audio frequencies, is approximately

$$R = \frac{17.5}{a^2} \text{ ohms} \tag{78}$$

The conductance G is affected by many factors and will be estimated later. Expressing (73) in terms of wire spacing, wire radius, and conductance, with the help of Eqs. (76), (77), and (78), we have

$$R = \frac{GL}{C}$$

$$\therefore \frac{17.5}{a^2} = G \times 10^3 \times \left[76.3\left(\log \frac{b}{a}\right)^2 + 8.24 \log\left(\frac{b}{a}\right)\right]$$

Practical considerations impose that b/a be a fairly large number, hence the second term in the brackets may be neglected for the present purpose. Consequently

$$\left(\log \frac{b}{a}\right)^2 = \frac{229 \times 10^{-6}}{Ga^2}$$

Expressing G in micromhos for convenience, taking the square root, and expressing b in meters, we have finally

$$b = a \log^{-1}\left(\frac{15.13}{a\sqrt{G}} - 3\right) \tag{79}$$

where b = spacing of wires, in meters, for distortionless line
a = radius of wire in mm. (copper wire)
G = conductance per mile in μmhos

Assuming a conductance of 1 μmho as representative of average conditions, we obtain

For a wire of 1 mm. diameter, $b = 0.9 \times 10^{27}$ m.
For a wire of 2 mm. diameter, $b = 1.35 \times 10^{12}$ m.
For a wire of 5 mm. diameter, $b = 2.8$ km.
For a wire of 10 mm. diameter, $b = 9.1$ m.
For a wire of 20 mm. diameter, $b = 32.6$ cm.

We may conclude that the distortionless open wire line, for audio frequencies at least, is not an economic possibility, as it would require extremely heavy conductors. If conductors of normal size are used, the series dissipation factor d_s is considerably larger than the shunt factor d_p. At frequencies exceeding 10,000 cycles and lower than about 500 kc., both factors become so small that distortion is never serious. At the audio frequencies a reduction of d_s is very desirable, as it not only reduces distortion but materially improves the efficiency of transmission.

The problem has been solved by the insertion of inductors at regular intervals along the line. This increases the average value of L, hence decreases d_s and enables the engineer to satisfy Eq. (72) without having resort to prohibitive dimensions.

The inductors used for this purpose are known as *loading coils*. Their use was first suggested by Heaviside and further developed by Pupin.

Most loaded lines are disuniform structures and cannot be studied within the framework of uniform line theory. The reader is referred to the literature on the subject.[(6)] However, at the lower audio frequencies, lines loaded at regular intervals (*lump* loaded) operate as if the loading were continuous. To study the effect of loading, the reader should repeat the example of Sec. 3.9, giving L a value higher by 30 mh, and increasing R by 1 ohm. These are representative values for the increase in average distributed inductance and resistance due to standard "light" loading.

4.5. Approximate Expressions for Conductors in Cable. When the two conductors are almost contiguous as in a cable, the limitations presented by the open wire line become much more serious. Because of the proximity of the wires, the capacity is large and the inductance practically negligible, so the d_s (except at very high frequencies) is very large compared to unity and d_p very small. Accordingly, approximate expressions may be obtained from the general equations (53) and (54) by neglecting d_s with respect to unity and unity with respect to d_s. These are

$$\frac{\alpha}{n} = \frac{\beta}{n} = r_0 = x_0 = \sqrt{\frac{d_s}{2}} \tag{80}$$

Hence

$$R_0 = X_0 = \sqrt{\frac{r}{2\omega c}} \tag{81}$$

$$\alpha = \beta = \sqrt{\frac{\omega c r}{2}} \tag{82}$$

$$\frac{\omega}{\beta} = \sqrt{\frac{2\omega}{cr}} \tag{83}$$

The approximation involved in these expressions is clearly shown on the curves of Fig. 16; if the expressions were rigorous, curves 1_a and 1_b, 2_a and 2_b, etc., would coincide throughout.

Both attenuation α and wave velocity ω/β are proportional to the square root of frequency. In a long cable, the difference in attenuation and time delay between the lowest and highest speech frequencies is very considerable.

Loading materially improves the performance of a cable. It is not practicable, however, to load a cable to the point where its performance approaches that of the distortionless line previously considered (Sec. 4.3).

CHAPTER V

REFLECTION AT THE TERMINALS OF TRANSMISSION NETWORKS AND LINES. USE OF MAPPING METHODS

5.1. Reason for the Use of Matching Terminations. Up to this point, the analysis of transmission lines has been carried out under the assumption, formulated in Sec. 3.4, of matched operation. If, as this assumption implies, the load or receiver impedance is equal to Z_0, the characteristic impedance of the line, then the entire problem of analyzing or predicting the line performance reduces to the determination of the two complex constants, Z_0 and $\theta = \alpha + j\beta$, which has been carried out in Chaps. III and IV.

As was stressed in Chap. II, the performance of a network is affected by the terminations (source and load). A section of line is no exception to this rule. We have considered a section of line a mile long, terminated in such a manner that the load impedance was reproduced at the input; this impedance, and the logarithm of the complex input to output voltage (or current) ratio for such a setup, have been selected as the *constants* Z_0 and θ for the line. Conversely, if the constants are known, then the performance of the line in question is fully determined, provided the load has the matching value Z_0.

This condition is very often satisfied in practice. It has to be met, more or less carefully, in all systems designed for the transmission of a *wide range* of frequencies, because a mismatch would bring with it fluctuations of the voltage and current ratios with frequency (Sec. 5.5).

The distinction between *matched* and *output-matched* operation should be called to mind here (Sec. 2.4). The points previously raised in this connection will be repeated here with special reference to the line, as it is important that they should be clearly understood.

A line is output matched when only the load has the matching value Z_0. This implies, of course, that transmission is always in the same direction. Under these conditions, the input impedance is Z_0 and the voltage (current) ratio is $e^{\theta l}$, just as if both terminations were matched. To put it more generally, the ratios between any two of the four quantities—input current and voltage, output current and voltage—are unaffected by the source impedance (Sec. 1.7).

For what reason, then, is it necessary to match a line at *both* ends and not at the output only? Primarily because very few lines transmit only in one direction. Terminations generally function alternatively as transmitters and receivers. In addition, although none of the ratios is affected by the source impedance, the values themselves are; the input current, for instance, is inversely proportional to the sum of source and input impedances, hence a change in source impedance will affect the input current and every other value in the same ratio.

Consequently, the power transmitted by the line depends on the source impedance. A given source transmits maximum power into an impedance equal to the conjugate of the source impedance (Sec. 2.10). In particular, if the input impedance is a resistance, the source impedance should be an equal resistance if it is to deliver all the power of which it is capable. Actually, the characteristic impedance of a line, hence the input impedance if the load is matching, is not a pure resistance but generally approaches a resistance, so that when the source matches the line the condition for maximum power is approximately satisfied.

In conclusion, whenever the transmission line is used as a means for conveying energy to a distant point for the purpose of communication, matched operation is the rule. The rule may be followed more or less rigidly and the question of predicting how the line will behave for any particular value of load is obviously of more than academic interest. Moreover, the study of lines in the extreme cases, when the far end is open- or short-circuited, provides methods of fault location (Sec. 5.8).

On the other hand, sections of line have come to be used more and more frequently for purposes other than transmission over long distances. Generally speaking, they replace coils and condensers when the wavelength is of the order of the dimensions of apparatus. Because this use is restricted to very high frequencies—currently known as vhf—it is customary to distinguish between *telephonic lines* and *vhf lines.* Although there is no fundamental difference, vhf applications stress aspects of line theory that are not important at telephonic frequencies, and the behavior of the line with generic terminations becomes particularly relevant.

This subject will now be approached by extending general network theory to cover operation with nonmatching loads.

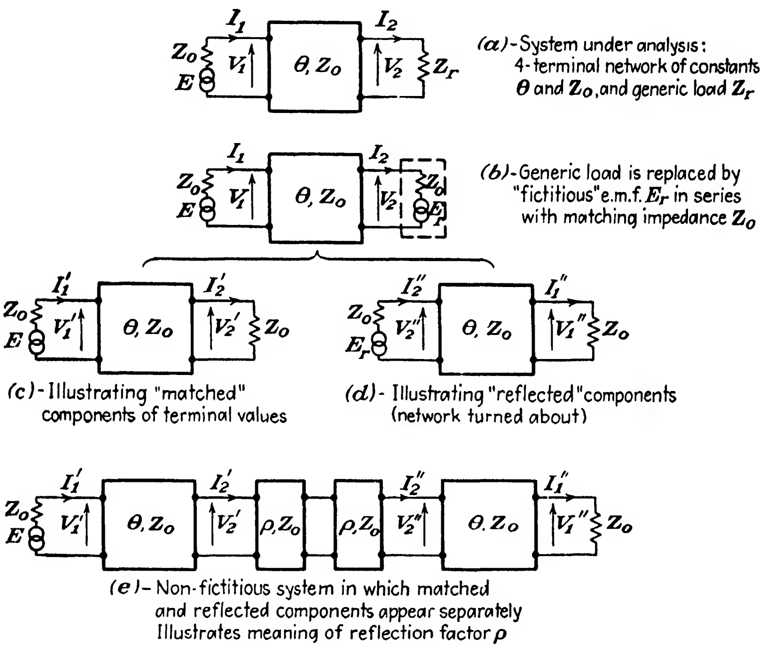


FIG. 18.—*Analysis of network operation under generic load conditions.*

We must therefore go back to the generic four-terminal network and obtain expressions for the ratios between the terminal values when the load is arbitrarily chosen. (The source impedance does not affect these ratios, as we know.) Steady state (Sec. 1.6) and linearity (Sec. 1.7) will be assumed, as before.

5.2. Analysis of the Four-terminal Network with Generic Load. Consider the system of Fig. 18*a*. Note the differences between this system and that of Fig. 6*a* (four-terminal network between generic terminations). Chiefly, it is now assumed that the network is reversible, of known constants, that the load is passive, and that the source impedance is Z_0, the characteristic value.

Such a system may safely be taken as the basis for discussion because line sections are reversible; their constants are known (or can be evaluated by the means indicated in Chap. III); transmission is never in both directions at the same time; and finally, since we are interested in the *ratios* of the terminal values, the source impedance is immaterial and may be set at any convenient value.

The conclusions drawn from this analysis will apply not only to the line but to any reversible network of known constants (such as symmetrical filter sections). The same conclusions may be reached by methods suitable to the line only, as will be shown (Chap. XIII). The present treatment, being more general and somewhat less complex from an analytical standpoint, should have precedence.

We shall rely chiefly, as in Sec. 2.8, on the principle of superposition. Using a fictitious replacement (Sec. 2.5) we will imagine the load Z_r replaced by a matching load Z_0 in series with a fictitious e.m.f., $\mathbf{E}_r$ (Fig. 18*b*). The "dual" replacement could be used equally well, but there would be no advantage in doing so.

The fictitious character of $\mathbf{E}_r$ should be fully appreciated. Its value, unlike that of E, depends on the parameters of the system and can be obtained by computing the current $\mathbf{I}_2$ in the output twice, first by Thévenin's theorem, and secondly, on the basis of Fig. 18*b*, by the superposition principle. For the two systems in Figs. 18*a* and 18*b* to be equivalent, the two values thus obtained must be identically equal. This condition yields an expression for $\mathbf{E}_r$.

When $\mathbf{E}_r$ is known, we can, again by the superposition principle, write the partial values of terminal voltage and current, due respectively to $\mathbf{E}_r$ and $\mathbf{E}$, and add them, obtaining the actual values as they appear in the original system.

As usual, let $\mathbf{V}_1$, $\mathbf{I}_1$, $\mathbf{V}_2$, $\mathbf{I}_2$, stand for the terminal values to be evaluated (Fig. 18*a*). Let us apply Thévenin's theorem (Sec. 2.6) to find $\mathbf{I}_2$. We have

$$\mathbf{I}_2 = \frac{\mathbf{V}_{oc}}{Z_0 + Z_r} \tag{84}$$

In the above, $\mathbf{V}_{oc}$ remains to be determined. This is not difficult, however. Consider Fig. 18*c*; all the terminal values for this system may be considered known; they are the matched values (Sec. 2.8). They are known in the sense that they can be written in terms of **E** and of the network constants, as follows:

$$\mathbf{V}_1' = \frac{\mathbf{E}}{2} \qquad \mathbf{V}_2' = \frac{\mathbf{E}}{2}e^{-\theta} \qquad \mathbf{I}_1' = \frac{\mathbf{E}}{2Z_0} \qquad \mathbf{I}_2' = \frac{\mathbf{E}}{2Z_0}e^{-\theta} \tag{85}$$

Thévenin's system may again be used to express $\mathbf{I}_2'$ in terms of the open-circuit voltage

$$\mathbf{I}_2' = \frac{\mathbf{V}_{oc}}{2Z_0} \tag{86}$$

Combining the last line with (84), we have

$$\mathbf{I}_2 = \mathbf{I}_2' \frac{2Z_0}{Z_0 + Z_r} \tag{87}$$

The first expression for $\mathbf{I}_2$ has now been obtained. To obtain the second expression, based on the system of Fig. 18*b*, we must evaluate the difference of the two partial currents, $\mathbf{I}_2'$ (Fig. 18*c*) and $\mathbf{I}_2''$ (Fig. 18*d*) due to **E** and $\mathbf{E}_r$, respectively. Note that the difference rather than the sum must be taken because of the convention as to the positive direction of the two currents, as shown by the arrows. Accordingly, we have

$$\mathbf{I}_2 = \mathbf{I}_2' - \mathbf{I}_2'' = \mathbf{I}_2' - \frac{\mathbf{E}_r}{2Z_0} \tag{88}$$

We may now compare (88) with (87) and derive an expression for $\mathbf{E}_r$. Thus

$$\mathbf{I}_2' \frac{2Z_0}{Z_0 + Z_r} = \mathbf{I}_2' - \frac{\mathbf{E}_r}{2Z_0}$$

Evaluating $\mathbf{I}_2'$ (85), the above becomes

$$\mathbf{E}_r = \mathbf{E}\frac{Z_r - Z_0}{Z_r + Z_0}e^{-\theta} \tag{89}$$

The system of Fig. 18*b* is now fully determined. Before proceeding with the evaluation of the terminal values and their ratios, however, the concept of *reflection* should be introduced, and with it the definitions of *reflection coefficient* and *reflection constant.*

5.3. Reflection. The double-primed components of terminal voltage and current (Fig. 18*d*), which, superimposed to the matched values (Fig. 18*c*), result in the actual values for the network with generic load (Fig. 18*a*), are called *reflected* values.

This designation may be justified as follows: The values in question would be the actual terminal values if the load, or receiver, were active; actually, the receiver is passive. Hence we are led to assume that the receiver acts as a mirror, sending back, or reflecting, a part of the signal impressed upon it.

It should be kept in mind, however, that the reflected values have no separate existence. It is possible to solve for the terminal values without having resort to the separation into components, by the general method of Sec. 2.7; the answer would then be in terms of the network parameters.

The network and load of Fig. 18*a* constitute a passive system driven at the input terminals. This is the basic mental picture that the reader should retain throughout the steady-state analysis of networks and lines.

The other picture, more artificial but also more revealing in some respects, is that of two superimposed disturbances propagated in opposite directions, and such that their mutual ratio is determined at the output by the load impedance, while their sum at the input is the driving voltage (if voltage is being considered). This concept will be stressed later (Sec. 7.5). In making use of it, the reader is cautioned against assuming that the reflected disturbance can be reflected again at the source if this does not have the matching value, and so on. If a transient pulse were applied, this multiple reflection would actually take place.[8] But under steady-state conditions, the direct and reflected waves add up precisely to the value of the driving voltage at the input terminals, and there is no further reflection.

5.4. Reflection Coefficient. Reflection Constant. The ratio between matched and reflected values at the output terminals (generally a complex number) is the *reflection coefficient* k, a quantity which depends on the load and characteristic impedances; when the latter is fixed once for all, k may be used to define a particular load impedance.

Let us express k. By definition, we have

$$k = \frac{\mathbf{I}_2''}{\mathbf{I}_2'}$$

Hence, and from Eqs. (85) and (88),

$$k = \frac{\mathbf{E}_r/2Z_0}{\mathbf{E}e^{-\theta}/2Z_0} = \frac{\mathbf{E}_r}{\mathbf{E}}\, e^{\theta} \tag{90}$$

Now, substituting the value of E_r (89)

$$k = \frac{Z_r - Z_0}{Z_r + Z_0} \tag{91}$$

[*Expression of reflection coefficient*]

Using the ratios $z_r = Z_r/Z_0$ and $y_r = Y_r/Y_0$, we have the forms

$$k = \frac{z_r - 1}{z_r + 1} = \frac{1 - y_r}{1 + y_r} \tag{92}$$

The reflection coefficient is extensively used in the literature and will be used in the present treatment whenever convenient.

However, k is essentially a ratio of currents, or voltages. (The same expression for it would be obtained if the output voltages had been used.) The logarithms of such ratios generally take the place of the ratios themselves, for a number of reasons. Accordingly, we will define the *reflection constant* as follows:

$$\rho = \sigma + j\tau = -\tfrac{1}{2}\ln k = \tfrac{1}{2}\ln\frac{Z_r + Z_0}{Z_r - Z_0} \tag{93}$$

[*Definition of reflection constant*]

Alternatively, any one of the following forms may be used to define ρ:

$$k = \frac{Z_r - Z_0}{Z_r + Z_0} = e^{-2\rho} \tag{94}$$

$$z_r = \frac{Z_r}{Z_0} = \coth \rho \tag{95}$$

$$y_r = \frac{Z_0}{Z_r} = \tanh \rho \tag{96}$$

The reflection constant is often considered as the transfer constant of an imaginary network (usually a section of line), hence no special notation for it has appeared in the literature. However ρ, like k, may be used to define an impedance. It is therefore simply another form of evaluating impedances, while θ is the constant of a four-terminal network. A separate notation for the two quantities is certainly desirable.

With the help of the reflection constant, we can now devise a system, entirely free of fictitious elements, where both the matched and the reflected values of Fig. 18*c* and 18*d* appear simultaneously at different points. Imagine (Fig. 18*e*) a chain of four matched networks, of which the first and fourth are identical to the original network, while the two middle networks have constants Z_0 and ρ. As can easily be verified, the terminal values of the first network are the matched values of Fig. 18*c*; those of the fourth network, the reflected values of Fig. 18*d*.

By this scheme, the meaning of electrical reflection can be made apparent. The type of reflection that everyone has in mind is reflection of light at oblique incidence, when the incident and reflected rays are physically separate. Such a situation is electrically reproduced in Fig. 18*e*. What actually happens in a network with generic load corresponds to reflection at normal incidence.

The reflection constant is now endowed with new significance. It is the transfer constant of a hypothetical network, which handles the signal twice (first in one direction, then in the other) before it is sent back toward the input. Likewise, light may be attenuated and retarded at the point where it is reflected, an infinitesimal film, which has to be crossed twice, accounting for these changes.

5.5. Voltage and Current Ratios. The matched and reflected components of voltage and current at the terminals will now be tabulated and added. Their values can be arrived at, by inspection of Fig. 18*e*, and tabulated (page 93).

The mutual ratios of the terminal values result from the tabulation. These are independent of the source e.m.f. and of the source impedance; they are functions only of the network constants and of the load impedance. Thus we have, for the input to output voltage ratio (see tabulation):

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{1 + e^{-2(\theta+\rho)}}{1 + e^{-2\rho}}\, e^{\theta} = \frac{\cosh(\theta + \rho)}{\cosh \rho} = \cosh \theta + y_r \sinh \theta \qquad (97)$$

[*Input to output voltage ratio*]

Similarly, for the current ratio

$$\frac{\mathbf{I}_1}{\mathbf{I}_2} = \frac{1 - e^{-2(\theta+\rho)}}{1 - e^{-2\rho}}\, e^{\theta} = \frac{\sinh(\theta + \rho)}{\sinh \rho} = \cosh \theta + z_r \sinh \theta \qquad (98)$$

[*Input to output current ratio*]

Note. $y_r = Z_0/Z_r = \tanh \rho$; $z_r = Z_r/Z_0 = \coth \rho$

When $Z_r = Z_0$ (matching load), both the above reduce to

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\mathbf{I}_1}{\mathbf{I}_2} = \cosh \theta + \sinh \theta = e^{\theta}$$

confirming that the values of terminal voltage and current which serve to define the transfer constant are those of the output matched network (Eq. 6).

The expressions we have derived for the voltage and current ratios show that these ratios vary with the load impedance and with θ in a rather complicated manner. It is quite possible for either ratio to be less than unity in magnitude, in which case the voltage (or current) has a higher value at the output than at the input. This occurrence was observed in the operation of lines when line theory had not yet been developed, and given the name of *Ferranti* effect (see Sec. 5.10). In general, we may deduce from the form of the expressions that both magnitude and angle of the ratios depend on the value of β, the phase constant. In the case of lines, in particular, this means that if the input voltage is constant, the output will vary with the frequency, since β is sensibly proportional to ω. This is one reason why a departure of the load impedance from the matching value (a mismatch) is undesirable in telephonic systems.

Matched components	$V_1 = \frac{E}{2}$	$I_1 = \frac{E}{2Z_0}$	$V_2' = \frac{E}{2}e^{-\theta}$	$I_2 = \frac{E}{2Z_0}e^{-\theta}$
Reflected components	$V_1'' = \frac{E}{2}e^{-2(\theta+\rho)}$	$I_1'' = \frac{E}{2Z_0}e^{-2(\theta+\rho)}$	$V_2'' = \frac{E}{2}e^{-(\theta+2\rho)}$	$I_2'' = \frac{E}{2Z_0}e^{-(\theta+2\rho)}$
Total values: Input terminals	$V_1 = V_1' + V_1'' = \frac{E}{2}[1 + e^{-2(\theta+\rho)}]$	$I_1 = I_1' - I_1'' = \frac{E}{2Z_0}[1 - e^{-2(\theta+\rho)}]$		
Output terminals			$V_2 = V_2' + V_2'' = \frac{E}{2}e^{-\theta}(1 + e^{-2\rho})$	$I_2 = I_2' - I_2'' = \frac{E}{2Z_0}e^{-\theta}(1 - e^{-2\rho})$

5.6. Input Impedance. Most important among the ratios of terminal values is the *input impedance*. This quantity determines the amount of power delivered by the source into the network. If the network is nondissipative, there is no difference between this power and the power received by the load.

The difference between load and input impedances may be very great. In such cases, the network has the function of transforming the load impedance.

A discussion of impedance transformation will be found in Chap. VIII. The relation between load and input impedance will be taken up here in a general way.

Taking the ratio of input voltage and current, we have

$$Z_i = \frac{\mathbf{V}_1}{\mathbf{I}_1} = Z_0 \frac{1 + e^{-2(\theta+\rho)}}{1 - e^{-2(\theta+\rho)}} = Z_0 \coth(\theta + \rho)$$

The above, combined with (95), brings out the desired relation

$$z_i = \coth(\theta + \rho) \tag{99}$$

[*Input impedance number*]

$$z_r = \coth \rho \tag{100}$$

[*Load impedance number*]

$$\left[z_i = \frac{Z_i}{Z_0} \qquad z_r = \frac{Z_r}{Z_0} \right]$$

Equations (99) and (100) are not in a form suitable for computation but lend themselves to an interpretation from which originate various graphical methods of solving transmission problems, one of which (based on the *hyperbolic tangent map*) will be taken up shortly; the others, later on (Secs. 8.7 and 10.5). The compact form of (99) makes it an easy equation to remember; other forms of the same equation, currently found in the literature, may be obtained by expanding it; for example, the following:

$$Z_i = Z_0 \frac{Z_r \cosh\theta + Z_r \sinh\theta}{Z_0 \cosh\theta + Z_r \sinh\theta} \tag{101}$$

The above must be expanded further in order to compute Z_i, with the help of ordinary tables of hyperbolic and circular functions. On the other hand, if maps of the hyperbolic functions of complex variables are available, (99) can be used directly, at a great saving of time.

For the reader who is not familiar with the use of maps of functions of a complex variable, the subject is briefly reviewed in the following section.

5.7. Map of the Hyperbolic Tangent of a Complex Variable. Consider the equation

$$z = \coth \rho$$

Analytically, for every value of the complex variable z, the equation assigns a value to the complex variable ρ, and vice versa.

Geometrically, *complex numbers* represent *points on a plane*. For example, $a + jb$ represents a point whose rectangular coordinates are a and b; $Ae^{j\varphi}$, a point whose polar coordinates are A and φ.

A *complex variable* is a complex number that may have any one of a set of values. It represents a point which may be anywhere within a given region of a plane. For example, z, depending on the particular value which the load has in each case, occupies a particular position on its plane; however, if we assume the load conductance to be positive at all times, certain regions of the plane will be barred to it.

It is convenient to assign a particular plane to each variable. Thus we have, for instance, the z plane and the ρ plane (Fig. 19). As the load impedance varies, both the z point and the ρ point will move to new positions in their respective planes.

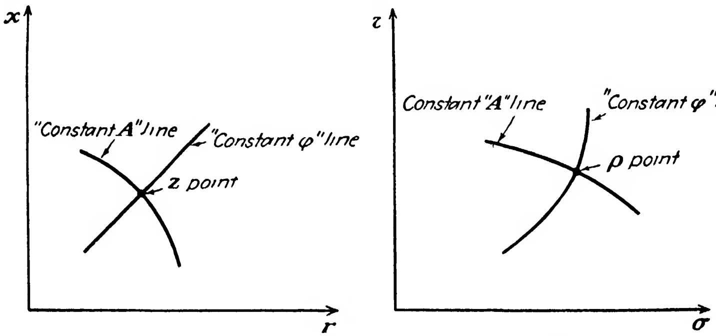
Suppose now that we know the value of z for some given case and want to find the corresponding value of ρ. Let A and φ stand for the polar coordinates of z (rectangular coordinates could be used, but not as conveniently, as examples will show).

It is easy enough to locate the z point on its plane, once the scale of A has been agreed upon. This may be done with a protractor and compasses, since we know A and φ. Polar coordinate paper would make this operation still easier. The lines on this paper are, analytically, lines of constant A (circles about the origin) and lines of constant φ (straight lines through the origin). Jointly, they form a map whereby a point of given polar coordinates may be located readily. For example, if $z = 2.5\underline{/30^\circ}$, the line $A = 2.5$ is followed to its intersection with the line $\varphi = \pi/6$ (30°).

Since for every z point there is a corresponding ρ point, while the former describes the line $A = 2.5$, the latter must describe some definite line on its own plane (Fig. 19). The second line is a transformation of the first; it is still a constant A line, but it has been transferred to a new plane, thereby acquiring a new shape (hence the word, *transformation*).

Evidently, if all the constant A and constant φ lines had been transferred to the ρ plane, it would be easy to locate the ρ point; these lines would constitute a *map* of the function $z = \coth \rho$ on the ρ plane. Such a map may be considered as a special kind of graph paper—a *transformation* of the polar coordinate paper.

Any figure may be transformed from one plane to the other with the help of maps of this type. If the figure is small, its shape will remain substantially the same, because the trans-



z plane

$$A = |z| = \sqrt{r^2 + x^2}$$
$$\varphi = \underline{/z} = \tan^{-1}\frac{x}{r}$$

ρ plane

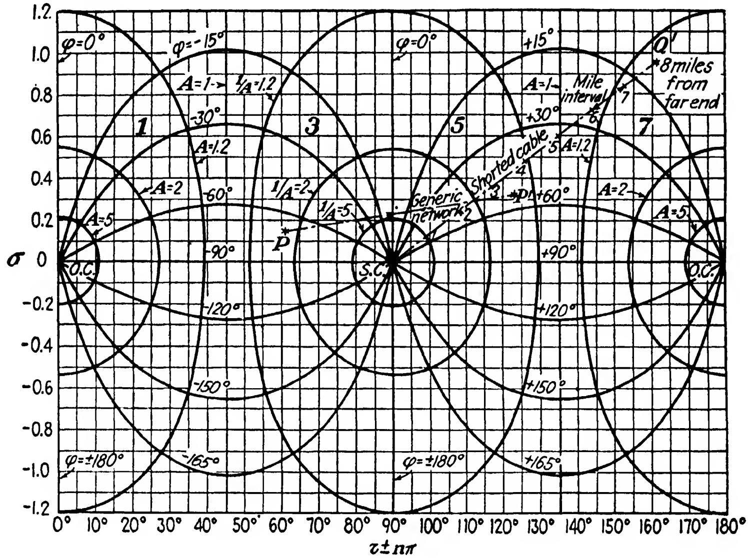
$$A = |\coth \rho| = \sqrt{\frac{\sinh^2 \sigma + \cos^2 \tau}{\sinh^2 \sigma + \sin^2 \tau}}$$
$$\varphi = \underline{/\coth \rho} = \tan^{-1}(\tanh \sigma \tan \tau) - \tan^{-1}(\coth \sigma \tan \tau)$$

FIG. 19.—*Illustrating the transformation of constant A and constant B lines from the z plane to the ρ plane if z = coth ρ.*

formation does not change the *angle* between two lines at their common point. Hence, such transformations are known as *isogonal* or *conformal*, and the branch of geometry that deals with them is *conformal mapping.*[18]

A map of the function $z = \coth \rho$ on the ρ plane is shown on Fig. 20*a* and 20*b*. It can also serve as a map of the reciprocal function $y = \tanh \rho$, and is called accordingly *map of the hyperbolic tangent of a complex variable.* The map of Fig. 20*a* contains only a few lines; it must be used in conjunction with the partial map of Fig. 20*b*, which, with due attention to the key of Fig. 20*a*, may be used indifferently for any one of the eight quadrants of the general map.

Accurate maps of this and other hyperbolic functions have been published.[6] They are not common, because they are useful only in connection with line and network problems, and because other maps (such as the k-plane map, Sec. 10.5) can be used to solve many of these problems and are easier to draw.[1]



(a)

FIG. 20a.—*Hyperbolic tangent* (*or cotangent*) *of* $\rho = \sigma + j\tau$.
$A = |\coth \rho|$; $1/A = |\tanh \rho|$; $\varphi = \underline{/\coth \rho}$; $-\varphi = \underline{/\tanh \rho}$.

RANGE SELECTION TABLE

Range of $\tau \pm n\pi$*	Range of σ	Quadrant	Range of φ	Range of A
$0 < \tau^\circ < 45^\circ$	$\sigma > 0$	1	$0 > \varphi^\circ > -90^\circ$	$A > 1$
$0 < \tau^\circ < 45^\circ$	$\sigma < 0$	2	$-90^\circ > \varphi^\circ > -180^\circ$	$A > 1$
$45^\circ < \tau^\circ < 90^\circ$	$\sigma > 0$	3	$0 > \varphi^\circ > -90^\circ$	$A < 1$
$45^\circ < \tau^\circ < 90^\circ$	$\sigma < 0$	4	$-90^\circ > \varphi^\circ > -180^\circ$	$A < 1$
$90^\circ < \tau^\circ < 135^\circ$	$\sigma > 0$	5	$0 < \varphi^\circ < 90^\circ$	$A < 1$
$90^\circ < \tau^\circ < 135^\circ$	$\sigma < 0$	6	$90^\circ < \varphi^\circ < 180^\circ$	$A < 1$
$135^\circ < \tau^\circ < 180^\circ$	$\sigma > 0$	7	$0 < \varphi^\circ < 90^\circ$	$A > 1$
$135^\circ < \tau^\circ < 180^\circ$	$\sigma < 0$	8	$90^\circ < \varphi^\circ < 180^\circ$	$A > 1$

* n is any positive integer. Values of τ° such as 30 and 210° are equivalent.

Example of evaluation of the reflection factor and input impedance with the help of the coth ρ map. Given a reversible network

[1] Maps of the hyperbolic functions should not be regarded as charts for the solution of specific problems, but as mathematical tools, like tables of logarithms or of the circular functions. The solution of line problems without the help of such tools is comparable to the longhand extraction of square roots.

of the following constants

$$\alpha = 0.18 \text{ nep.} \qquad \beta = 62° \qquad Z_0 = 600 \text{ ohms } \underline{/-28°}$$

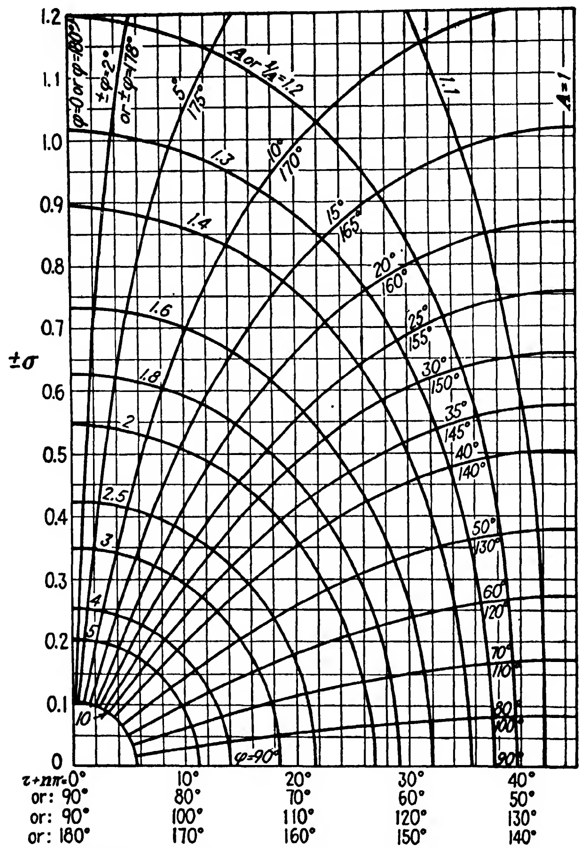


FIG. 20*b*.—*Hyperbolic tangent (or cotangent) of* $\rho = \sigma + j\tau$. *See Fig.* 20*a* *for range selection.* $A = |\coth \rho|;\ \frac{1}{A} = |\tanh \rho|;\ \varphi = \underline{/\coth \rho} = -\underline{/\tanh \rho}$.

The reflection constant of the load impedance

$$Z_r = 350 \text{ ohms } \underline{/42°}$$

is required with respect to the network.

In accordance with the position

$$z_r = \frac{Z_r}{Z_0} = Ae^{j\varphi}$$

we have from the given data

$$A = \frac{350}{600} = 0.583 \qquad \varphi^\circ = -42 - 28 = -70^\circ*$$

At the intersection of the lines $A = 0.583$ and $\varphi^\circ = -70$ deg. on the map (Fig. 20*a*), we locate point P, whose rectangular coordinates are

$$\sigma = 0.15 \qquad \tau^\circ = 60.6^\circ$$

The value of ρ is

$$\rho = \sigma + j\tau = \sigma + j\frac{\pi}{180}\tau^\circ = 0.15 + j0.106$$

Note. The actual values of σ and τ° are found on the map for one quadrant (Fig. 20*b*); there is a choice of two values for each coordinate if this map is used, but the ambiguity is removed by consulting the key tabulation of Fig. 20*a*.

Rather than the numerical value of the reflection factor, the input impedance is usually of interest. From Eqs. (99) and (100) we note that, if ρ is changed to $\rho + \theta$, its hyperbolic cotangent changes from z_r to z_i.

Graphically, the change from ρ to $\rho + \theta$ causes the ρ point to move a distance α along the σ axis and a distance β along the τ axis. Looking upon ρ and θ as *vectors*, we may say that they *add vectorially*. In any event, a new point results; at this point, the map will show new values of A and φ, from which Z_i may be computed.

The preceding example will now be carried on to the determination of the input impedance. Given the values for the network constants

$$\alpha = 0.18 \text{ neper} \qquad \beta^\circ = 62^\circ$$

and the values of σ and τ° (at point P),

$$\sigma = 0.15 \qquad \tau^\circ = 60.6$$

we have, adding α to σ and β° to τ°

$$\sigma' = 0.33 \qquad \tau^{\circ\prime} = 122.6^\circ$$

Thus the new point P' is located (Fig. 20*a*). The directed seg-

* φ° and τ° are the values of φ and τ in degrees. It would be better to use radians throughout, but since tables and slide rules use degrees, this unit is more convenient.

ment PP' represents the *vector* θ. Its horizontal projection is the phase constant, its rise, or vertical projection, the attenuation constant. At P' we read (with the help of Fig. 20*b*)

$$\frac{1}{A'} = 1.43 \qquad \varphi^{\circ\prime} = 52^\circ$$

and since

$$A'e^{j\varphi} = z_i = \frac{Z_i}{Z_0}$$

we have (using the value of Z_0 given in the previous example),

$$Z_i = 600 \text{ ohms } A'\underline{/(28^\circ + 52^\circ)} = 420 \text{ ohms } \underline{/80^\circ}$$

The reader will find another example relating to the use of the hyperbolic tangent map in Sec. 5.10. Applied to lines or cables, the map is particularly useful, although it is valid for any reversible network of known constants.

If the input impedance is known (by measurement), the map may be used to locate the position of a "short" or "open" along the cable. Problems relating to power lines may also be solved in this way.[6] However, the hyperbolic tangent map has been brought to the attention of the reader, not so much because of its value as a mathematical tool, but rather as a preliminary example of the ways in which impedance changes may be represented on the complex plane (Sec. 8.3).

5.8. Input Impedance of Open- and Short-circuited Lines. Fault Detection. For $Z_r = 0$ (short circuit) and $Z_r = \infty$ (open circuit), Eq. (101) becomes identical to Eqs. (25) and (26), respectively, namely

$$Z_{sc} = Z_0 \tanh \theta \qquad Z_{oc} = Z_0 \coth \theta$$

For a line of length l, whose transfer constant per unit length is θ (Sec. 3.4), the above take the form

$$Z_{sc} = Z_0 \tanh (\theta l) \tag{102}$$
$$Z_{oc} = Z_0 \coth (\theta l) \tag{103}$$

In some cases, approximate forms of (102) and (103) may be used to advantage. This is true, for example, when the length l is short compared to the wavelength. Analysis of Z_{sc} and Z_{oc} in this case confirms the validity of the method of line parameter measurement discussed in Sec. 3.12.

Expanding (102) in series, we have

$$Z_{sc} = Z_0 \frac{\sinh(\theta l)}{\cosh(\theta l)} = Z_0 \frac{\theta l + \frac{(\theta l)^3}{3!} + \frac{(\theta l)^5}{5!} + \cdots}{1 + \frac{(\theta l)^2}{2!} + \frac{(\theta l)^4}{4!} + \cdots}$$

$$= \theta l Z_0 \frac{1 + \frac{(\theta l)^2}{3!} + \frac{(\theta l)^4}{5!} + \cdots}{1 + \frac{(\theta l)^2}{2!} + \frac{(\theta l)^4}{4!} + \cdots} \qquad (104)$$

If $|\theta l|$ is small, all second and higher order terms in the above may be neglected. The error introduced will be less than 1 per cent if

$$\frac{|\theta l|^2}{2!} = 0.01 \qquad \text{or} \qquad |\theta l| = 0.0141 \qquad (105)$$

But $|\theta|$ can be written

$$|\theta| = \sqrt{\alpha^2 + \beta^2} = \beta \sqrt{1 + \left(\frac{\beta}{\alpha}\right)^2}$$

For most lines, approximately, $\beta = 2\pi/\lambda$ and $(\beta/\alpha)^2 \ll 1$, hence

$$|\theta| = \frac{2\pi}{\lambda}$$

and (105) becomes

$$\frac{2\pi l}{\lambda} = 0.141 \qquad \text{or} \qquad l = 0.0224\lambda \qquad (106)$$

In conclusion, *if the length of the section is less than about* 2 *per cent of the wavelength*, we may write, neglecting higher order terms of (104),

$$Z_{sc} = \theta l Z_0$$

or, substituting values of the line constants in terms of the series and shunt parameters (Sec. 3.2),

$$Z_{sc} = l Z_s$$

and likewise, expanding (103),

$$Y_{oc} = \frac{1}{Z_{oc}} = l Y_p$$

It is therefore evident that, if the length is below the limit indicated, the Q of the short-circuited section will closely approximate Q_s, and the Q of the open-circuited section will likewise approach Q_p (Sec. 3.12).

Another useful approximate form of (102) and (103) results when it is assumed that α, the attenuation constant, is small compared to the phase constant β and that dissipation in the line is low enough to justify the approximations

$$\beta = \frac{2\pi}{\lambda} \qquad Z_0 = \sqrt{\frac{L}{C}}$$

All the above assumptions are valid, in first approximation, when the frequency is comparatively high (Sec. 7.1). They enable us to write Z_{sc} and Z_{oc} in the form

$$Z_{sc} = j\sqrt{\frac{L}{C}} \tan\frac{\omega l}{\Omega} \qquad Z_{oc} = -j\sqrt{\frac{L}{C}} \cot\frac{\omega l}{\Omega}$$

If the *reactance* of the short-circuited line and the *susceptance* of the open-circuited line are expressed, the two expressions are of the same form, namely

$$X_{sc} = \sqrt{\frac{L}{C}} \tan\frac{\omega l}{\Omega} \tag{107}$$

$$B_{oc} = \sqrt{\frac{C}{L}} \tan\frac{\omega l}{\Omega} \tag{108}$$

The above are periodic functions of ω, and the *period* (frequency interval within which X_{sc} and B_{oc} go through a full cycle) is given by

$$\Delta\omega = \frac{2\pi\Omega}{l}$$

In the last line, Ω is the wave velocity on the line, approaching in this case the velocity in free space, or 186,000 m.p.s. (Sec. 3.6).

$\Delta\omega$ can be obtained from a plot of X_{sc} or B_{oc} against frequency; l, the length of the line, will result. Explicitly, l is given by

$$l = \frac{2\pi\Omega}{\Delta\omega} = \frac{\Omega}{\Delta f} = \frac{186{,}000}{\Delta f} \text{ mile} \tag{109}$$

This relation is used to find the length of line separating the point at which the impedance is measured from a fault in the line, which may be a short across the line or an interruption (open circuit). This procedure locates the breakdown approximately, obviating the necessity of exploring the entire line from terminal to terminal.

Example. After the service has been interrupted due to an unknown cause, impedance measurements across the line are made at one of the terminals. The impedance is measured with a bridge over a range of frequencies, resulting in the plot of Fig. 21*a*. The distance from the terminal to the fault and the nature of the fault are to be deduced from this plot.

The ordinates of the plot obtained are values of the impedance *magnitude* $\sqrt{R^2 + X^2}$. According to either (107) or (108), such a plot should be periodic and go through 0 at intervals of frequency equal to $\Delta f/2$, as shown in dotted line.

Actually, the ordinates fluctuate between a minimum and a maximum; however, the distance between two successive peaks can be taken as $\Delta f/2$. Hence, we have

$$\Delta f = 32{,}000 \text{ c.p.s.}$$

and the unknown distance is

$$l = \frac{186{,}000}{32{,}000} = 5.8 \text{ miles}$$

The nature of the fault (whether open or short) can be argued by observing whether the impedance tends to zero or infinity for $f = 0$. In the example given, the impedance tends to zero, hence the line must be shorted.

★5.9. Input Impedance of Near-matched Networks. Very often small differences between the load impedance and Z_0, the matching impedance of a network or line, cannot be avoided. Filter networks, for example, are sometimes designed under the assumption that they are matched throughout the operating frequencies, whereas actually this is true only at one or two frequencies.(5) At the other frequencies the match is approximate. The same is generally true of parallel cables.

When the mismatch is small, approximate formulas and constructions may be advantageously used. These are all the more helpful because chart methods, notably the hyperbolic tangent map, are not convenient in such cases, as ρ becomes large.

Consider expression (101) for the input impedance. Expanding the hyperbolic functions into sums of exponentials, we have

$$Z_i = Z_0 \frac{Z_r(e^\theta + e^{-\theta}) + Z_0(e^\theta - e^{-\theta})}{Z_0(e^\theta + e^{-\theta}) + Z_r(e^\theta - e^{-\theta})}$$
$$= Z_0 \frac{(Z_r + Z_0)e^\theta + (Z_r - Z_0)e^{-\theta}}{(Z_r + Z_0)e^\theta - (Z_r - Z_0)e^{-\theta}}$$

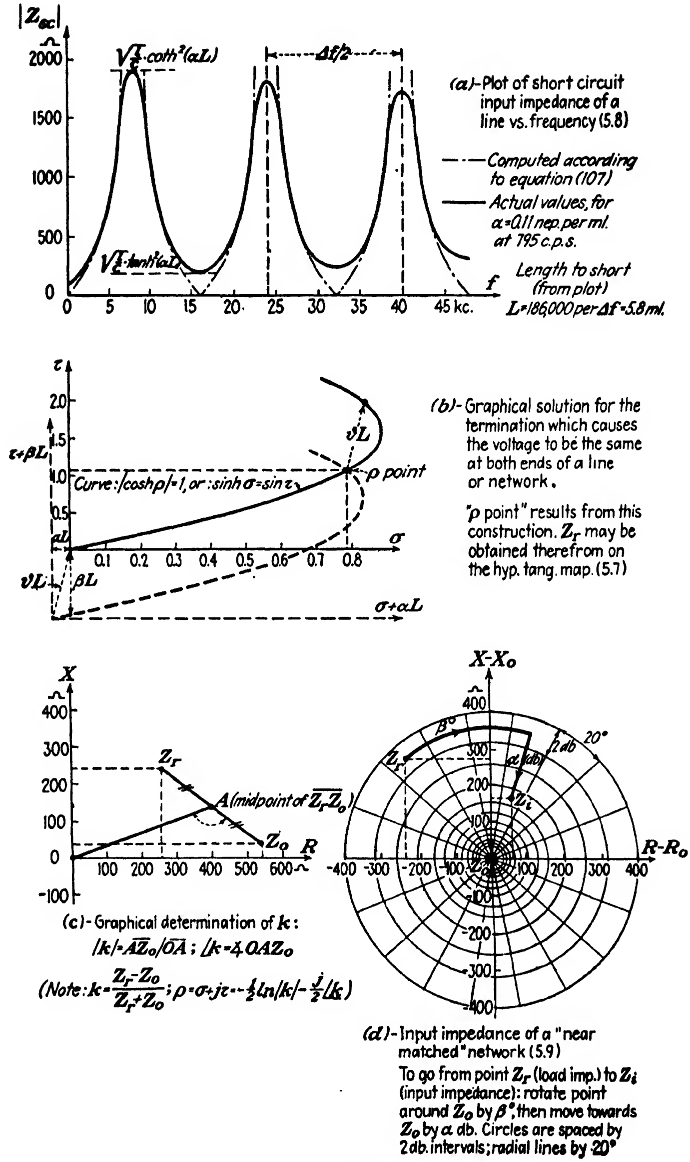


FIG. 21.—*Input impedance.*

In terms of the reflection coefficient k (Sec. 5.4), the above may be written

$$Z_i = Z_0 \frac{1 + ke^{-2\theta}}{1 - ke^{-2\theta}} \tag{110}$$

[*Input impedance in terms of k*]

There is no particular advantage in writing Z_i this way as a rule, because the apparent simplification is offset by the necessity of computing k.* However, when the load is close to the matching value, the reflection coefficient k is small. Now (110) may be evaluated by means of the series expansion

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots$$

Terms of the second and higher orders are negligible, however, if k is small (in absolute value) compared to unity, which we assume to be the case. Accordingly, we may write

$$Z_i = Z_0(1 + ke^{-2\theta})^2$$

Again, in expanding, the second-order term drops out. Hence

$$Z_i = Z_0(1 + 2ke^{-2\theta}) \tag{111}$$

This can be written, evaluating k,

$$Z_i = Z_0 + (Z_r - Z_0)\frac{2Z_0}{Z_0 + Z_r}e^{-2\theta}$$

which, noting that $\dfrac{2Z_0}{Z_0 + Z_r} = 1 + k$, becomes

$$Z_i = Z_0 + (Z_r - Z_0)e^{-2\theta} + \frac{(Z_r - Z_0)^2}{Z_0 + Z_r}e^{-2\theta}$$

The last term may here also be neglected. Hence, finally, we have

$$Z_i = Z_0 + (Z_r - Z_0)e^{-2\theta} \tag{112}$$

[*Input impedance for near-matched networks*]

A construction based on the above expression is given in Fig. 21*d*.

5.10. Illustrative examples.

Input to output voltage ratio. Assuming a line of known length and constants terminated by a resistive load, this question is

* The magnitude and angle of k are obtained most conveniently by the construction of Fig. 21*c*.

asked: For what values of the load resistance is the output voltage higher than the input? The voltage ratio of any reversible network of known constants is given by Eq. (97). Since the load is a pure resistance, we replace y_r by g_r:

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \cosh\theta + g_r \sinh\theta$$

The hyperbolic functions expand as follows:

$$\left\{\begin{array}{l} \cosh\theta = \cosh(\alpha + j\beta) = \cosh\alpha\cos\beta + j\sinh\alpha\sin\beta \\ \sinh\theta = \sinh(\alpha + j\beta) = \sinh\alpha\cos\beta + j\cosh\alpha\sin\beta \end{array}\right. \tag{113}$$

Hence, expressing the *magnitude* of the voltage ratio

$$\left|\frac{\mathbf{V}_1}{\mathbf{V}_2}\right|^2 = (\cosh\alpha\cos\beta + g_r\sinh\alpha\cos\beta)^2 + (\sinh\alpha\sin\beta + g_r\cosh\alpha\sin\beta)^2$$

The problem may be solved by setting this magnitude equal to unity, which yields an equation in g_r. Solving this equation, we find the load value which makes the two voltages equal. Any *higher* value of g_r will make the input voltage greater, and vice versa. Accordingly, we write

$$\left|\frac{\mathbf{V}_1}{\mathbf{V}_2}\right|^2 = 1$$

which, substituting the value for the ratio and arranging terms, becomes

$$g_r^2(\sinh^2\alpha\cos^2\beta + \cosh^2\alpha\sin^2\beta) + 2g_r\cosh\alpha\sinh\alpha + \cosh^2\alpha\cos^2\beta + \sinh^2\alpha\sin^2\beta - 1 = 0$$

We now make use of the identities

$$\left\{\begin{array}{l} |\sinh(\alpha + j\beta)|^2 = \sinh^2\alpha\cos^2\beta + \cosh^2\alpha\sin^2\beta \\ \qquad\qquad = \sinh^2\alpha\sin^2\beta \\ |\cosh(\alpha + j\beta)|^2 = \cosh^2\alpha\cos^2\beta + \sinh^2\alpha\sin^2\beta \\ \qquad\qquad = \sinh^2\alpha + \cos^2\beta \end{array}\right. \tag{114}$$

and obtain

$$g_r^2(\sinh^2\alpha + \sin^2\beta) + 2g_r\cosh\alpha\sinh\alpha + (\sinh^2\alpha - \sin^2\beta) = 0$$

Solving for the positive value of g_r, we have

$$g_r = \frac{\sqrt{\sinh^2\alpha + \sin^4\beta} - \cosh\alpha\sinh\alpha}{\sinh^2\alpha + \sin^2\beta}$$

Numerically, consider an open wire line of length 10 ml., $\alpha = 0.008$ nep./ml., at the frequency 3,000 c.p.s. For such a line we have, approximately,

$$\beta = \frac{\omega}{\Omega} = \frac{2\pi \times 3{,}000}{186{,}000} = 0.1013 \text{ rad./mile}$$

We must, of course, substitute αl and βl for α and β in the general form. We compute

$$\begin{aligned} \sinh \alpha l &= \sinh 0.08 = .08 & \sin \beta l &= \sin 1.013 = 0.53 \\ \cosh \alpha l &= \cosh 0.08 = 1.003 & \sin^2 \beta l &= 0.281 \\ \sinh^2 \alpha l &= 0.0064 & \sin^4 \beta l &= 0.079 \end{aligned}$$

Hence

$$g_r = \frac{\sqrt{0.0064 + 0.079} - 1.003 \times 0.08}{0.0064 + 0.281} = 0.737$$

Under the conditions of the problem, if the characteristic impedance of the line is 600 ohms, the input and output voltages have the same peak or r.m.s. value when the load impedance is

$$|Z_r| = \frac{600}{0.737} = 814 \text{ ohms}$$

Any value higher than this will result in the Ferranti effect, the line acting in this respect as a step-up transformer. We have assumed that both Z_0 and Z_r are resistive; if not, the conclusion is still valid provided their angles are the same.

Note. Variations of the above problem may present formidable mathematical difficulties if solved by the above method. These difficulties may be surmounted with the help of considerations based on the use of the ρ plane (Sec. 5.7). The voltage ratio may be written

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\cosh (\rho + \theta)}{\cosh \rho}$$

Suppose we had drawn, on the ρ plane, the line

$$|\cosh \rho| = 1$$

(This line, shown in Fig. 21*b*, has the equation $\sinh \sigma = \sin \tau$.) The point for any particular value of load can be found by means of the hyperbolic cotangent chart (Sec. 5.7). Given the transfer constant of the network or line section, this is added vectorially to ρ, just as in the determination of input impedance. If the point $\rho + \theta$ lies on the line we have drawn, the voltage ratio is unity. If it lies to the right of it, the output is lower than the input, and vice versa.

Suggested Exercise. Check the result obtained in the previous problem by the ρ plane method.

Fault location on a cable. The impedance of a broken-down cable, measured at one of the terminals at some particular value of frequency, is

$$Z_i = 870 \text{ ohms } \underline{/-34^\circ}$$

The constants of the cable are known for the same frequency. They are

$$Z_0 = 685 \text{ ohms } \underline{/-44^\circ} \qquad \alpha = 0.12 \text{ nep./mile} \qquad \beta^\circ = 9^\circ/\text{mile}$$

The coth ρ chart may be used to locate the breakdown. The values of A and φ° corresponding to the measured input impedance are

$$A' = \left|\frac{Z_i}{Z_0}\right| = \frac{870}{685} = 1.27 \qquad \varphi^{\circ\prime} = \underline{/Z_i} - \underline{/Z_0} = 10^\circ$$

Since $A' > 1$ and $0 < \varphi^\circ < 90^\circ$, the key (Fig. 20*a*) indicates that we are in quadrant 7, therefore we must have $\sigma > 0$;

$$135^\circ < \tau^\circ < 180^\circ.$$

The map for one quadrant (Fig. 20*b*) gives, at the intersection of the lines $A = 1.27$ and $\varphi^\circ = 10^\circ$, the values

$$\sigma = 0.95 \qquad \tau^\circ = 162^\circ$$

(The alternative readings do not agree with the permissible values of σ and τ°, hence must be ruled out.)

We can now locate point Q', corresponding to the input impedance of the cable, on the map of Fig. 20*a*. Each added mile of cable would move this point to the right by 9° and upwards by 0.12 (this movement resulting from the vector addition of the transfer constant of a mile of cable). Hence, if measurements had been made at various lengths from the breakdown, the corresponding points would lie on a straight line of slope

$$\frac{\alpha}{\beta^\circ} = \frac{0.12}{9^\circ}$$

If a line of such a slope is drawn through Q', we see that it passes through the point *SC*. At this point, $A = 0$, corresponding to zero impedance (short circuit).

The total phase shift between Q' and point SC is 72 deg. Hence, the distance in miles

$$l = \frac{72}{9} = 8 \text{ miles}$$

Each mile of cable is graphically represented on the map as the segment marked *mile interval.*

Suggested Exercise. Obtain from the chart the magnitude of the input impedance into the cable for distances of 1, 2, 3 . . . miles from the short and plot against distance. Repeat for an open-circuited cable.

CHAPTER VI

POWER FLOW THROUGH NETWORKS AND LINES

6.1. Criteria of Power Comparison. Fundamental equations governing the performance of a network, or line, when the load impedance has a generic value, have been derived in the preceding chapter. Particular attention has been given to the voltage and current ratios and to the input impedance. The transmitted power, or the power absorbed by the load, will now be considered.

The results of this study will be, necessarily, of a comparative nature. The ratio between the power in question and some other value of power, taken as a standard, will have to be expressed. In some cases, uncertainty may arise as to the particular value to be chosen as reference.

To bring light on this point, let us take a familiar situation where power is involved. Consider, for example, the power available at the shaft of an electric motor. We speak of a "5-kw." motor, thereby associating the motor with a definite value of power, rather than a ratio. However, we mean by that only that 5 kw. is the highest power which the motor will deliver continuously at the rated speed without overheating.

If we rule out the possibility of overload, no *value* of power can be considered as characteristic of the motor, but we can still *compare* the power output with the input. The ratio power output/power input is the *efficiency* of the motor. It varies depending on the speed; but for a given speed, it is characteristic of the system made up of the motor and its mechanical load.

Similarly, the *transmission loss* expresses the comparison between input and output power in a network and is therefore characteristic of the system comprising the network and its load (at a given frequency). If the load has the matching value, then it becomes characteristic of the network alone, and equal to the attenuation constant α (Sec. 2.2).

The transmission loss is the only one of the various losses to be considered which actually represents a "loss" as the power engineer understands it, *i.e.*, power wasted in the form of heat.

The *insertion loss*, for example, has a different meaning, as was explained in Sec. 1.9. It measures the ratio between the power that the load would receive if connected *directly to the source* and the power that it receives *through the network*. Here also, if the load matches, we have identity between the insertion loss, the transmission loss, and the attenuation constant. In general, however, the insertion loss does not indicate any waste of power. A dissipationless network (a filter section, for example) may have a very high insertion loss.

We might say that the transmission loss compares the power flow at two different points of a system, and the insertion loss compares the flow into the same load from two different sources. As a third criterion, we may compare the power flowing from the *same source* into *two different loads*. We do this when we assign power ratings to light bulbs; it is assumed that they all draw power from the same source (the domestic supply). Because in this instance the difference in power arises from a change, or *permutation* of loads at the same source terminals, the corresponding loss, more precisely defined in the following, will be called in general *permutation loss*. This loss, like the insertion loss, may be positive or negative. It is useful in correlating other losses but not significant in itself unless the standard of comparison is carefully selected.

As a particularly significant example of permutation loss, consider the loss in power (positive or negative) resulting when a *matching load* is replaced by a generic impedance. This type of loss will be called *reflection loss*, and its study is particularly important.

As will be explained, the term *reflection loss* has a somewhat different meaning in the literature.

6.2. Loss Terminology. To avoid possible confusion, some care must be taken in the choice of symbols associated with various values of power. The notation used for power in Sec. 1.9, and everywhere for voltages and currents, will serve as a basis, namely, the symbol for power transmitted through a junction j (Sec. 1.8) will have j as a subscript.[1]

In addition, a superscript will denote the impedance into which

[1] Junction 1 is always the input to a network, junction 2, the output; when only source and load are considered, their common terminals will be junction 1.

the power is transmitted (the impedance toward load at junction j). For example, P_2^0 is the power flowing into the impedance Z_0 (characteristic impedance) at junction 2 (output terminals). For other examples, see Fig. 22.

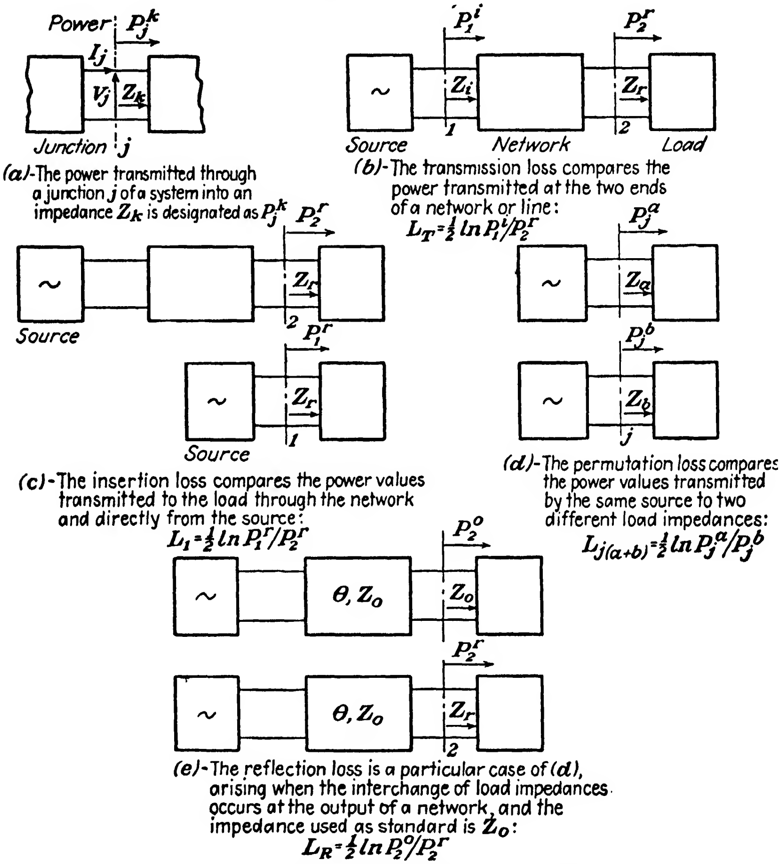


FIG. 22.—*Criteria of power comparison.*

The various losses whose significance has been explained in the preceding section can now be formally defined, in agreement with the above notation and with reference to Fig. 22. Neper values of the losses (Sec. 1.9) are expressed. Thus we have

$$L_T = \frac{1}{2} \ln \frac{P_1^i}{P_2^r} \tag{115}$$

[*Definition of transmission loss*]

$$L_I = \tfrac{1}{2} \ln \frac{P_1{}^r}{P_2{}^r} \tag{116}$$

[*Insertion loss*]

$$\alpha = \tfrac{1}{2} \ln \frac{P_1{}^0}{P_2{}^0} \tag{117}$$

[*Attenuation constant*]

$$L_{J(a \to b)} = \tfrac{1}{2} \ln \frac{P_j{}^a}{P_j{}^b} \tag{118}$$

[*Permutation loss due to change from load Z_a to Z_b at junction j*]

$$L_R = \tfrac{1}{2} \ln \frac{P_2{}^0}{P_2{}^r} \tag{119}$$

[*Reflection loss due to reflection at output of a network*]

The prevalent usage of the term *reflection loss,* as distinct from the above definition, will now be explained, and reasons for not adhering to it will be given. The following definition has been given:

$$\text{Reflection loss} = \ln \left| \frac{\mathbf{I}_2{}'}{\mathbf{I}_2} \right|$$

where $\mathbf{I}_2{}'$ is the matching value of output current, $\mathbf{I}_2$ the value when the load is Z_r. Noting that (88), $\mathbf{I}_2 = \mathbf{I}_2{}' - \mathbf{I}_2{}''$, and that (Sec. 5.4), $\mathbf{I}_2{}''/\mathbf{I}_2{}' = k$, we can write the above as follows:

$$\text{Reflection loss} = -\ln |1 - k|$$

The above quantity depends therefore, like k and ρ, on the relation between load and characteristic impedance and may be used to define the load—not particularly conveniently, however. As an expression of reflection loss, the quantity is certainly no more valid than the corresponding voltage ratio in logarithmic form, or

$$\ln \left| \frac{\mathbf{V}_2{}'}{\mathbf{V}_2} \right| = \ln \left| \frac{\mathbf{V}_2{}'}{\mathbf{V}_2{}' + \mathbf{V}_2{}''} \right| = \ln \left| \frac{1}{1 + k} \right| = -\ln |1 + k|$$

The choice of the term *reflection loss* to indicate either one of the two above-mentioned quantities is therefore quite arbitrary. There is no reason for introducing either quantity in the course of analysis, nor is the evaluation of either quantity necessarily called for. On the other hand, the variation of L_R (119) with frequency is identical, aside from an additive constant, to the variation of transmitted power with frequency; it leads therefore to an evaluation of selectivity (Sec. 9.5). Other applications of the reflection loss will be taken up (6.8).

6.3. Correlation of Losses. For a given network, all the losses and the attenuation are mutually related. Equations expressing these relationships are readily arrived at and are instrumental in computing the losses themselves. We may write, for example,

$$\frac{P_1^r}{P_2^r} = \frac{P_1^i}{P_2^r}\frac{P_1^r}{P_1^i}$$

Taking the logarithms, and using equations (115), (116), (118),

$$L_I = L_T + L_{1(r\rightarrow i)} \tag{120}$$

The insertion loss equals the transmission loss, plus the permutation loss which results when Z_r is replaced by Z_i at the source terminals. This equation may be used to find L_I from L_T.

L_T may be similarly broken down. We have

$$\frac{P_1^i}{P_2^r} = \frac{P_1^o}{P_2^o}\frac{P_2^o}{P_2^r}\frac{P_1^i}{P_1^o}$$

hence, taking logarithms,

$$L_T = \alpha + L_R + L_{1(i\rightarrow 0)} \tag{121}$$

Usually, L_T is determined directly and the last line is helpful in finding L_R.

Finally, a relation between the insertion and reflection losses may be derived, by comparing Eqs. (120) and (121)

$$L_I = \alpha + L_R + L_{1(r\rightarrow 0)} \tag{122}$$

For an output-matched network (Sec. 2.4), Eqs. (120), (121), and (122) reduce to the following:

$$L_T = L_I = \alpha \qquad L_R = 0$$

It is worthy of note that, when the output is matched, the insertion loss equals α independently of the source impedance. In general, however, L_I depends on Z_S, while L_T never does.

6.4. Transmission Loss. To evaluate any loss, we must express a ratio of powers. The power may be written as a function of current or voltage, as follows:

$$P = I^2R = V^2G$$

Expressing current and resistance, we have, for the transmission loss

$$L_T = \ln \left|\frac{\mathrm{I}_1}{\mathrm{I}_2}\right| + \tfrac{1}{2} \ln \frac{R_i}{R_r}$$

Substituting the value of the current ratio, we have (98)

$$L_T = \ln \left|\frac{\sinh (\theta + \rho)}{\sinh \rho}\right| + \tfrac{1}{2} \ln \frac{R_i}{R_r} \tag{123}$$

The second term of the above may be evaluated by means of the input impedance expression (101). Since, however, the transmission loss is seldom computed, this evaluation will not be carried out, and (123) will be used as it stands in the derivation of the insertion loss.

6.5. Insertion Loss. The insertion loss can now be obtained from Eq. (120). To this end, the permutation loss $L_{1(r \to i)}$ must be found. In general, we have

$$\begin{aligned} L_{J(a \to b)} &= \ln \left|\frac{I_j^a}{I_j^b}\right| + \tfrac{1}{2} \ln \frac{R_a}{R_b} \\ &= \ln \left|\frac{Z_s + Z_b}{Z_s + Z_a}\right| + \tfrac{1}{2} \ln \frac{R_a}{R_b} \end{aligned} \tag{124}$$

having let Z_s stand for the impedance toward the source at j. In particular, $L_{1(r \to i)}$ can be written

$$L_{1(r \to i)} = \ln \left|\frac{Z_s + Z_i}{Z_s + Z_r}\right| + \tfrac{1}{2} \ln \frac{R_r}{R_i}$$

hence, we have for the insertion loss

$$L_I = L_T + L_{1(r \to i)} = \ln \left|\frac{\sinh (\theta + \rho)}{\sinh \rho}\right| + \ln \left|\frac{Z_s + Z_i}{Z_s + Z_r}\right|$$

We may now make use of Eqs. (99) and (100), namely

$$z_i = \frac{Z_i}{Z_0} = \coth (\theta + \rho) \qquad z_r = \frac{Z_r}{Z_0} = \coth \rho$$

to which we may add a similar definition for the reflection constant of the source impedance, ρ' (Sec. 5.4)

$$z_s = \frac{Z_s}{Z_0} = \coth \rho' \tag{125}$$

The insertion loss may now be written[1]

$$L_I = \ln \left| \frac{\sinh (\theta + \rho) \coth \rho' + \cosh (\theta + \rho)}{\sinh \rho \coth \rho' + \cosh \rho} \right|$$
$$= \ln \left| \frac{\sinh (\theta + \rho) \cosh \rho' + \cosh (\theta + \rho) \sinh \rho'}{\sinh \rho \cosh \rho' + \cosh \rho \sinh \rho'} \right|$$

Hence, the final form

$$L_I = \ln \left| \frac{\sinh (\theta + \rho + \rho')}{\sinh (\rho + \rho')} \right| \tag{126}$$

[*Insertion loss in terms of the reflection constants*]

Equation (126), like (97), (98), and (99), lends itself well to a geometric interpretation based on the use of the ρ plane (Sec. 5.7). The insertion loss can therefore be computed quite readily with the use of maps of the hyperbolic functions, by the method outlined in Sec. 5.10.

The analytical evaluation is lengthy and can be carried out most conveniently in terms of the reflection coefficients k and k', defined as follows:

$$k = e^{-2\rho} = \frac{z_r - 1}{z_r + 1} \qquad k' = e^{-2\rho'} = \frac{z_s - 1}{z_s + 1}$$

Expanding the hyperbolic sines in (126) and noting that $|e^\theta| = e^\alpha$

$$L_I = \ln \left| \frac{e^{(\theta+\rho+\rho')} - e^{-(\theta+\rho+\rho')}}{e^{(\rho+\rho')} - e^{-(\rho+\rho')}} \right| = \ln \left| e^\alpha \frac{1 - e^{-2(\rho+\rho'+\theta)}}{1 - e^{-2(\rho+\rho')}} \right|$$

Hence

$$L_I = \alpha + \ln \left| \frac{1 - kk'e^{-2\theta}}{1 - kk'} \right| \tag{127}$$

[*Insertion loss in terms of k and k'*]

The case when $k = k'$, or $Z_s = Z_r$, is particularly important, because most transmission systems have equal terminations. In this case, (127) may be written

$$L_I = \alpha + \ln |1 + ke^{-\theta}| + \ln |1 - ke^{-\theta}| - \ln |1 + k| - \ln |1 - k| \tag{128}$$

[*Insertion loss between equal terminations*]

Another important case, that of the near-matched network, will be taken up later (Sec. 6.7). For examples of insertion loss evaluation, see Sec. 6.8.

[1] Note the identity

$$\ln |A| + \ln |B| = \ln |A| \cdot |B| = \ln |AB|$$

6.6. Reflection Loss. To find the reflection loss, Eq. (122) may be used, but $L_{1(0\to r)}$ must be evaluated first. Using the general formula for the permutation loss (124), we obtain

$$L_{1(o\to r)} = \ln \left|\frac{Z_s + Z_r}{Z_s + Z_0}\right| + \tfrac{1}{2} \ln \frac{R_0}{R_r}$$

There is advantage in expressing the above in terms of k and k'. Note that we have, in agreement with the definition of the reflection coefficient (92)

$$z_r = \frac{Z_r}{Z_0} = \frac{1+k}{1-k} \qquad z_s = \frac{Z_s}{Z_0} = \frac{1+k'}{1-k'}$$

Substituting, clearing, and simplifying, we have for $L_{1(0\to r)}$

$$L_{1(o\to r)} = \ln \left|\frac{1-kk'}{1-k}\right| + \tfrac{1}{2} \ln \frac{R_0}{R_r} \tag{129}$$

Equation (122) may now be used in combination with (127) and (129) to find the reflection loss. Thus

$$L_R = L_I - \alpha + L_{1(0\to r)} = \ln \left|\frac{1-kk'e^{-2\theta}}{1-kk'}\right| + \ln \left|\frac{1-kk'}{1-k}\right| + \tfrac{1}{2} \ln \frac{R_0}{R_r}$$

Simplifying, we obtain

$$L_R = \tfrac{1}{2} \ln \frac{R_0}{R_r} - \ln |1-k| + \ln |1-kk'e^{-2\theta}| \tag{130}$$

[*Reflection loss*]

The third term of (130) vanishes when $k' = 0$, or when the source matches the network. It therefore accounts for the effect of a *mismatch at the input* upon the reflection loss due to a *mismatch at the output*. For this reason, the term has been called *interaction loss.* The interaction loss may be neglected when the attenuation is large.

When the interaction loss is negligible, a more convenient form of the reflection loss expression is the following, obtained by expanding k:

$$L_R = \tfrac{1}{2} \ln \frac{R_0}{R_r} + \ln \tfrac{1}{2} \left|\frac{Z_0 + Z_r}{Z_0}\right| \tag{131}$$

[*Reflection loss neglecting interaction*]

The above form will be used in the study of selectivity of impedance transforming networks (Sec. 9.5).

★**6.7. Near-matched Networks.** Near-matched networks have been taken up before, in connection with the input impedance (Sec. 5.9). Because of the frequent occurrence of imperfect matching in communication systems, the approximate formulas valid in such cases should be given attention.

Let us carry out the evaluation of

$$f(w) = \ln|1 + w| + \ln|1 - w|$$

where $w = u + jv$ is a complex number of magnitude small compared to unity. Expanding the first logarithm in a series of which we consider the first and second term only (the remainder being obviously negligible), we have

$$\begin{aligned} 2\ln|1 + w| &= \ln(1 + 2u + u^2 + v^2) \\ &= 2u + u^2 + v^2 - \tfrac{1}{2}(2u + u^2 + v^2)^2 + \cdots \end{aligned}$$

Repeating for the second logarithm,

$$\begin{aligned} 2\ln|1 - w| &= \ln(1 - 2u + u^2 + v^2) \\ &= -2u + u^2 + v^2 - \tfrac{1}{2}(-2u + u^2 + v^2)^2 + \cdots \end{aligned}$$

Adding the last line to the one above, we obtain the required expression

$$\ln|1 + w| + \ln|1 - w| = v^2 - u^2 = -|w|^2 \cos 2\underline{/w} \qquad (132)$$

(Note that terms of the fourth order have been dropped.) The error involved in (132) is quite small, since all second- and third-order terms have been retained; the third-order terms, however, cancel out.

Equation (132) may be used to evaluate the insertion loss in the case of a network between equal terminations, of impedance not far from the characteristic value. Actually, the formula thus obtained continues to be accurate when the match is very poor; when there is a difference of 40 per cent between Z_0 and Z_r, assuming the angles to be the same (which is the most unfavorable case), the error is the order of 0.02 neper.

Using (132), Eq. (128) takes the form

$$L_I = \alpha + |k|^2 \{\cos 2\underline{/k} - e^{-2\alpha} \cos 2(\underline{/k} - \beta)\} \qquad (133)$$

[*Insertion loss between equal near-matching terminations*]

The magnitude and angle of k can be found most conveniently by construction, as shown in Fig. 21*c*. Depending on the angle of

k and on the phase constant β, the insertion loss may be higher or lower than the attenuation.

6.8. Illustrative examples.

Insertion loss. The insertion loss is required for a 10-mile section of the cable whose constants were computed in Sec. 3.11. Both the exact formula (128) and the approximate formula (133) will be used. The terminations are assumed equal, both of 600-ohm impedance, zero angle.

The data for the cable will be repeated below

$$Z_0 = 480 - j490 \text{ ohms} \qquad \text{attenuation } \alpha l = 0.1785 \times 10 = 1.785 \text{ nepers} \qquad \beta^\circ l = 100^\circ$$

Let us now compute k. We have

$$\begin{aligned} Z_r + Z_0 &= 480 + 600 - j490 = 1185 \text{ ohms } \underline{/-24.45^\circ} \\ Z_r - Z_0 &= 600 - 480 + j490 = 504.5 \text{ ohms } \underline{/76.25^\circ} \end{aligned}$$

Hence

$$k = \frac{Z_r - Z_0}{Z_r + Z_0} = 0.426\underline{/100.7^\circ} = -0.0792 + j0.417$$

$ke^{-\theta}$ can now be computed

$$\begin{aligned} ke^{-\theta} &= 0.426 \times 0.168\underline{/100.7^\circ - 100^\circ} \\ &= 0.0715\underline{/0.7^\circ} \qquad \text{or} \qquad 0.0715 \end{aligned}$$

Now we have

$$\begin{aligned} 1 + ke^{-\theta} &= 1.0715 \qquad \ln|1 + ke^{-\theta}| = 0.06907 \\ 1 - ke^{-\theta} &= 0.9285 \qquad \ln|1 - ke^{-\theta}| = \underline{-0.0751} \\ & \qquad\qquad\qquad\qquad\qquad\quad -0.00603 \end{aligned}$$

$$\begin{aligned} 1 + k &= 0.9208 + j0.417 \qquad |1 + k| = 1.01 \\ 1 - k &= 1.0792 - j0.417 \qquad |1 - k| = 1.155 \end{aligned}$$

$$\begin{aligned} \ln|1 + k| &= 0.00995 \\ \ln|1 - k| &= \underline{0.1441} \\ & \quad 0.15405 \end{aligned}$$

Finally,

$$L_I = 1.785 - 0.00603 - 0.15405 = 1.625 \text{ nepers.}$$

Using (133), we obtain

$$L_I = \alpha + (0.426)^2\{\cos(2 \times 100.7^\circ) - (0.168)^2 \cos[2 \times (100.7^\circ - 100^\circ)]\} = 1.611 \text{ nepers}$$

Reflection loss. A long line of 600 ohms characteristic impedance is terminated by a matching receiver. Another 600-ohms receiver can be inserted in parallel. When the second receiver is added the speech power in the first drops to a lower value. Find the corresponding decibel loss.

We must evaluate the reflection loss resulting when a matching load of 600 ohms is replaced by a 300-ohm load. To this we must add a loss of 3.01 db, because only half the load power will be made available at the first receiver after the insertion of the second (3.01 db correspond to a power ratio of 2 to 1).

The interaction loss is assumed negligible. Using Eq. (131), we have for the reflection loss

$$L_R(N) = \tfrac{1}{2} \ln \frac{600}{300} + \ln \tfrac{1}{2} \left| 1 + \frac{300}{600} \right|$$

$$L_R \text{ (db)} = 10(\log 2) - 20 \log (0.75) = 3.01 - 2.5 = 0.51 \text{ db}$$

Hence, the total value of the required loss

$$L \text{ (db)} = 0.51 + 3.01 = 3.52 \text{ db}$$

(1 watt of power is cut to 0.445 watt.)

Suggested Exercise. Consider the alternative scheme of two 600-ohm receivers in series, each provided with a shorting switch. The loss in power at the first receiver when the second is cut in (by opening the shorting switch) will be found to have the value of 3.52 db, as found above. Show why the loss must be equal in the two cases.

CHAPTER VII

BEHAVIOR OF ELECTRICALLY LONG LINES HAVING SMALL DISSIPATION

7.1 Electrically Long Lines. In Chap. V we discussed at some length the *input impedance* of a network with generic termination. By way of illustration, we considered the input impedance of transmission lines or cables and its use in fault location. We found that a mismatch somewhere along the line, and particularly an extreme mismatch such as a short or open circuit, will cause the impedance measured across the input terminals to vary periodically with frequency (Fig. 21*a*).

The telephone engineer attempts to avoid such variations in normal operation, as they are accompanied by changes in the insertion loss and frequency distortion, as well as other anomalies (such as echo and cross talk) arising from the transient nature of signals and the complexity of telephonic systems.

With the advent of much higher frequencies as a means of communication, it gradually became apparent that this peculiar behavior of mismatched lines can sometimes be used to advantage.

Before going further, we should have clearly in mind the meaning of the term *electrical length* of a line. In the equations which define the electrical behavior of lines, the physical length l never appears alone; it is always multiplied by the transfer constant per unit length θ. Two lines of different lengths will behave identically if they are the same in all other respects and if they operate at two different frequencies, so that θl is the same for both.

To satisfy this condition, frequency and physical length must vary inversely, so that the product fl remains constant; in fact, Eq. (40) [see also Eq. (46)] gives θl in the form

$$\theta l = \omega l \sqrt{LC} \sqrt{(d_s + j)(d_p + j)} = \frac{2\pi fl}{\Omega} \sqrt{(d_s + j)(d_p + j)}$$

showing that for a given line θl remains the same if length and frequency have a constant product.

Using λ, the wavelength on the line, the above may be written more simply

$$\theta l = \frac{2\pi l}{\lambda} \sqrt{(d_s + j)(d_p + j)}$$

The product $2\pi l/\lambda$ is often called *line angle.* It is approximately equal to the over-all phase shift caused by the line in matched operation and is a measure of the electrical length of the line.

As we shall see further on, particularly useful properties are exhibited by lines a quarter wavelength long, provided their physical length does not make them unwieldy. A quick numerical check shows that at a frequency of 100 mc./sec. the quarter wavelength in space, or on open wire lines or air-filled cables, is 75 cm. (Expressing the wavelength in meters, the frequency in megacycles per second, they are numerically related as follows: $f = 300/\lambda$.)

We might conclude the foregoing remarks as follows: An electrically long line is one whose length is of the same order of magnitude as the wave length on the line (usually one quarter or more). Except for cables with solid dielectric, the wavelength on the line is simply related to frequency, just as the wavelength in free space. Hence, only at very high frequencies can a line be physically short (comparable to the dimensions of apparatus) and at the same time electrically long. For this reason, the study of *long* lines usually assumes high operating frequencies, at which some factors that are important at telephonic frequencies may be considered negligible.

7.2. Attenuation in Long Lines. At very high frequencies, the parameters R and G of a line (Sec. 4.1) are far from constant. The dissipation factors d_s and d_p which appear in (134) are therefore difficult to evaluate. In any case, however, they are very small compared to unity. As may be readily verified, this makes for small numerical values of the attenuation α in relation to the phase shift β. (We recall that α in nepers and β in radians are the real and imaginary parts of θ.)

It follows that, *for some purposes,* the constants of long lines at very high frequency may be considered the same as those of the hypothetical lossless line (Sec. 3.2). This does not mean that transmission at high frequency can be effected without losses. In fact, although the attenuation is small compared to the phase shift, the latter is so large when considerable distances are involved that the attenuation may have large values.

If, for example, $\beta/\alpha = 1{,}000$, over a distance of 1,000 wavelengths the attenuation totals 2π nepers, or 54.5 db; now, at 100 mc./sec., 1,000 wavelengths correspond to a distance of 1.87 miles. Yet a line for which $\beta/\alpha = 1{,}000$, though obviously "lossy" if used for high-frequency long-distance transmission, is lossless for all practical purposes when only 1 or 2 wavelengths long. In particular, it will have virtually the same *input impedance* as a lossless line with the same termination and of the same small physical length. Its *characteristic impedance* will also be close to that of a lossless line, *i.e.*, a pure resistance. For convenience, formulas giving the attenuation and characteristic impedance at very high frequencies are tabulated below, without derivation. The formulas for α assume that no energy is lost in the dielectric.

	Attenuation α, db/m.	Characteristic impedance R_0, ohms	a equals	b equals
Parallel line in air	$36.2\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\sqrt{f}}{R_0}10^{-6}$ (134)*	$276 \log \frac{b}{a}$ (135)*	radius of wire cm.	distance between centers, cm.
Concentric cable, air- or gas-filled	$18.1\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\sqrt{f}}{R_0}10^{-6}$ (136)*	$138 \log \frac{b}{a}$ (137)*	inner conductor radius, cm.	outer conductor radius, cm.

* This number is the equation number.

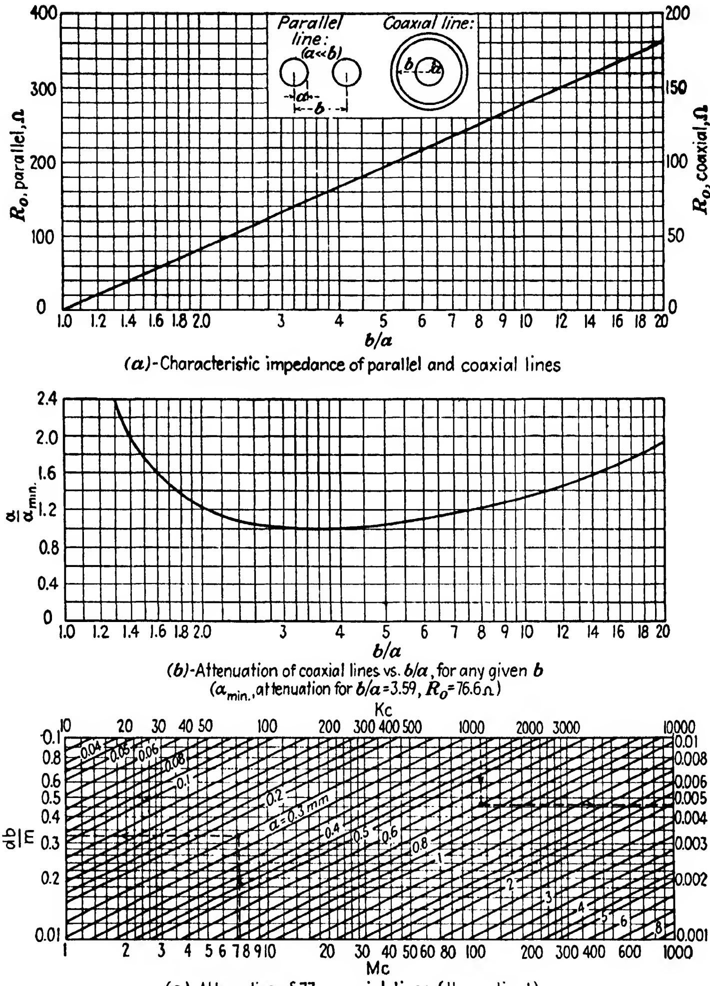
Conductors are assumed to be copper (See also Fig. 23).

A discussion on the optimum dimensions of concentric cables appears in Sec. 7.7.

In line with the above discussion, let us now consider transmission lines having the following properties:

1. Long electrical length (length comparable to λ)
2. High operating frequency
3. Negligible loss per wavelength (because $\alpha \ll \beta$)
4. Generic terminations

Condition (3) enables us to drop the attenuation α from the expressions in most cases. For this reason, and for brevity, the term *lossless* will be used at times to designate transmission lines in the above category.



(a)-Characteristic impedance of parallel and coaxial lines

(b)-Attenuation of coaxial lines vs. b/a, for any given b ($\alpha_{min.}$, attenuation for $b/a = 3.59$, $R_0 = 76.6\Omega$)

(c)-Attenuation of 77Ω coaxial lines (theoretical) ($\alpha = 3.06 \frac{\sqrt{f}}{a}$, where f = frequency, mc. a = inner cond. radius, mm. α = attenuation, db. per km.)

FIG. 23.—*Constants of lines for very high frequencies.*

7.3. Input Impedance and Impedance Transforming Action of Lossless Lines. The premise that α/β is a small number brings considerable simplification to the analysis of line performance, particularly with nonmatching loads, provided the line length is of the order of λ.

In general, the input and load impedances of a line of length L are correlated (see Sec. 5.6) by the equations

$$z_i = \coth(\rho + \theta L)$$
$$z_r = \coth \rho$$
$$z_i = \frac{Z_i}{Z_0} \qquad z_r = \frac{Z_r}{Z_0}$$

The expression $\rho + \theta L$ may be written

$$\sigma + \alpha L + j(\tau + \beta L)$$

where αL is the total attenuation of the line section, in nepers. This is, by assumption, small compared to unity; hence it may be neglected, barring some instances in which the load is purely reactive. We may, therefore, when Z_r has a generic value, use the approximations

$$z_i = \coth(\rho + j\beta L) = \coth\left(\rho + j\frac{2\pi L}{\lambda}\right) \tag{138}$$
$$z_r = \coth \rho \tag{139}$$

Recalling the interpretation given to (99) and (100), we may interpret the above as follows: For every value of impedance, and for a given line, there is a point on the ρ plane; the two ρ points corresponding to the input and load impedance of a reactive line section are separated by a distance $2\pi L/\lambda$ along the axis of imaginaries. Of the two components σ and τ which characterize the impedance toward load, only one, τ, is changed by the insertion of the line. (If the line had high dissipation per wavelength, both σ and τ would be changed appreciably, and in a fixed proportion.)

This impedance transforming action of the lossless line, and the concept of impedance transformation in general may be clarified by the following analogy:

Consider (Fig. 24*a*) an impedance $R + jX$. Imagine this to be connected at one end of an inductive line, consisting of two long, uniformly wound solenoids, each of reactance $x/2$ ohms per unit length at the given frequency. Let all distributed capacities be negligible at this frequency. This nductive line will transform the impedance $R + jX$ into $R + j(X + xL)$, if L is the line length. The transformation is limited in that it affects only X, leaving R unchanged. Furthermore, X can only be increased.

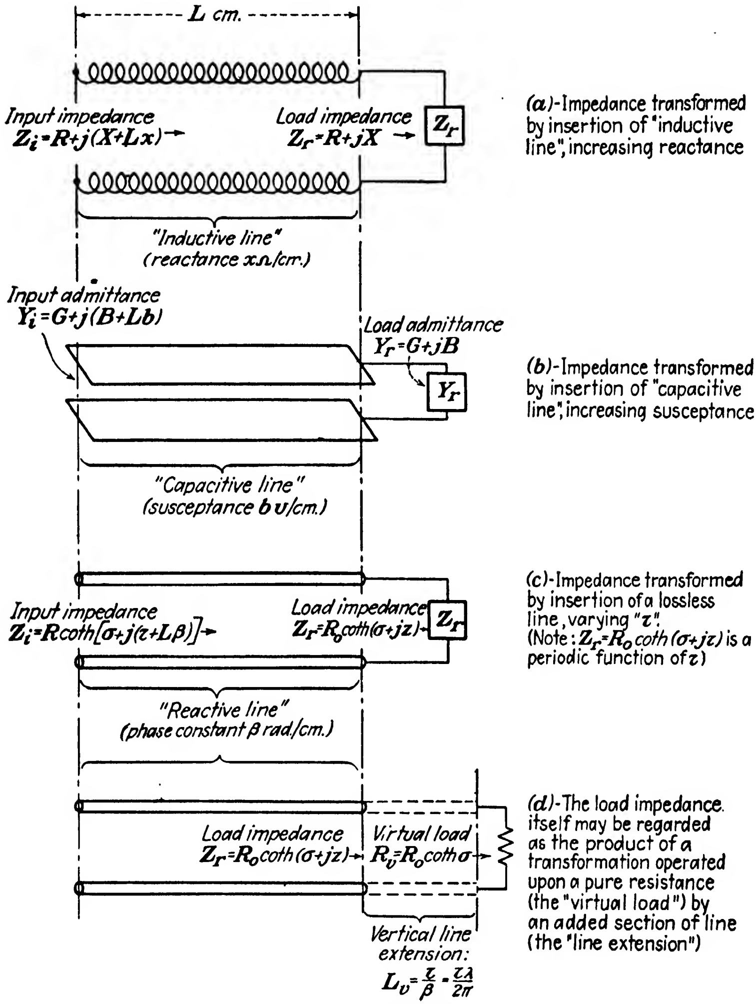


FIG. 24.—*Impedance transforming action of the lossless line.*

Similarly, we might imagine a capacitive line made of two parallel strips held close together, with negligible inductance. Connected to an impedance of value

$$\frac{1}{G + jB}$$

such a line would transform it to the value

$$\frac{1}{G + j(B + bL)}$$

Of the two parameters G and B which have now been used to define the impedance, this transformation has increased B, leaving G unchanged (Fig. 24*b*).

Finally, consider a lossless line of characteristic impedance R_0 and length L (Fig. 24*c*). Connected to an impedance of value

$$R_0 \coth (\sigma + j\tau)$$

this line transforms it into the value,

$$R_0 \coth [\sigma + j(\tau + \beta L)]$$

provided we assume $\alpha \ll \beta$. This time the impedance has been expressed in terms of R_0 and the two parameters σ and τ; of these, the transformation has affected only τ. It should be noted that since the impedance is a periodic function of τ, an indefinite increase of τ brings it through a cycle which keeps repeating itself, whereas an increase in X or B brings it closer and closer to infinity or zero.

All three of the transformations of Fig. 24 are unaccompanied by loss of power, which is one of the requisites of impedance transformation problems (Sec. 8.1). However, none of them can be used, in general, to bring the impedance to a specified value, since each transformation only varies one of the two parameters by which impedance is defined.

Impedance transformation problems and the methods for their solution will be taken up later (Chap. VIII); many such methods will be based on Eqs. (138) and (139) and their geometric interpretation.

The input impedance computation may have to be carried out analytically. Equation (101) can be used for this purpose. It simplifies to the following:

$$z_i = \frac{z_r + j \tan (2\pi L/\lambda)}{1 + jz_r \tan (2\pi L/\lambda)} \tag{140}$$

[*Input impedance of a lossless line*]

From the above we may obtain the rectangular components of the impedance and admittance number:

$$r_i = \frac{r_r\,[\cot^2(2\pi L/\lambda) + 1]}{r_r^2 + [x_r - \cot(2\pi L/\lambda)]^2} \tag{141}$$

$$x_i = \frac{x_r\,[\cot^2(2\pi L/\lambda) - 1] - \cot(2\pi L/\lambda)(r_r^2 + x_r^2 - 1)}{r_r^2 + [x_r - \cot(2\pi L/\lambda)]^2} \tag{142}$$

$$g_i = \frac{g_r\,[\cot^2(2\pi L/\lambda) + 1]}{g_r^2 + [b_r - \cot(2\pi L/\lambda)]^2} \tag{143}$$

$$b_i = \frac{b_r\,[\cot^2(2\pi L/\lambda) - 1] - \cot(2\pi L/\lambda)(g_r^2 + b_r^2 - 1)}{g_r^2 + [b_r - \cot(2\pi L/\lambda)]^2} \tag{144}$$

$$\left[\begin{array}{ll} z_i = r_i + jx_i = \dfrac{Z_i}{R_0} & z_r = r_r + jx_r = \dfrac{Z_r}{R_0} \\ y_i = g_i + jb_i = Y_iR_0 & y_r = g_r + jb_r = Y_rR_0 \end{array}\right]$$

7.4. Virtual Line Extension and Virtual Load. Standing Wave Ratio. It is often convenient to imagine the load, or termination, of a lossless line replaced by a *virtual termination*, consisting of an additional length of the same line extending back of the load terminals and loaded by a pure resistance (Fig. 24*d*).

It can be shown that such a substitution is always possible; for each value of the actual load impedance there is a value of the *virtual line extension* and a value of the *virtual load* such that the virtual and actual terminations are equivalent. This may be surmised from the fact that the impedance depends on two distinct real numbers, σ and τ. Now, if the line is lossless, the virtual line extension will appear to be a function of τ only; the virtual load, a function of σ only.

The load impedance may be written, as usual,

$$Z_r = R_0 \coth(\sigma + j\tau)$$

Let the virtual load have the value

$$R_v = R_0 \coth \sigma$$

and the virtual extension, the value

$$L_v = \frac{\tau\lambda}{2\pi}$$

The impedance of the virtual termination may now be written as an input impedance, using (138),

$$Z = R_0 \coth\left(\sigma + j\,\frac{2\pi L_v}{\alpha}\right) = R_0 \coth(\sigma + j\tau) = Z_r$$

showing that R_v and L_v are the values which bring about the desired equivalence.

The virtual termination illustrates the principle discussed in the preceding section. If a lossless line transforms the load impedance by adding to the value of the parameter τ, then the load impedance itself may be regarded as the product of a transformation, operated on a pure resistance (for which $\tau = 0$) by a line of such length as to add the correct value of τ (Fig. 24*d*).

Expressions for R_v and L_v suitable for numerical work are given below.

$$L_v = \frac{\tau\lambda}{2\pi} = \frac{\lambda}{4}\left\{\frac{\tan^{-1} A}{\pi} + B\right\} \tag{145}$$

[*Virtual line extension*][1]

$$r_v = \frac{R_v}{R_0} = \coth \sigma = C + \sqrt{C^2 - 1} \tag{146}$$

[*Virtual load number, or standing wave ratio*]

where A, B, and C have the following values:

for $x_r < 0$: $B = 0$ — for $x_r = 0$, $r_r < 1$: $B = 1$

for $x_r > 0$: $B = 1$ — for $x_r = 0$, $r_r > 1$: $B = 0$

$$A = \frac{2x_r}{r_r^2 + x_r^2 - 1}$$

$$C = \frac{r_r^2 + x_r^2 + 1}{2r_r}$$

$$\left(Z_r = r_r + jx_r = \frac{Z_r}{R_0}\right)$$

L_v and r_v may be expressed in terms of k, the reflection coefficient (Sec. 5.4), and a modification of the construction used to determine k (Fig. 21*c*) may be used to find L_v and r_v rather more conveniently than by computation. The new construction is shown in Fig. 25 for the two cases of inductive and capacitive load; actually, the construction is the same in both cases, but it is not easy to avoid confusion in measuring the angle. This construction follows from that of Fig. 21*d*, as the reader may verify, because -2τ is the angle of k (94) and because

$$r_v = \coth \sigma = \frac{1 + e^{-2\sigma}}{1 - e^{-2\sigma}} = \frac{1 + |k|}{1 - |k|}$$

The virtual load number r_v is primarily of interest because it is equal to the *standing wave ratio*, as will be shown in the

[1] The term *virtual line angle* is sometimes used for τ.

following section, while L_v determines the position of maxima and minima of voltage along the line with respect to the load terminals. The construction of Fig. 25 is useful, therefore, for the purpose of determining how the voltage (and current) will vary along the line for any particular loading. The opposite problem of evaluating an unknown impedance by measurements taken along a line will be taken up in Sec. 7.6.

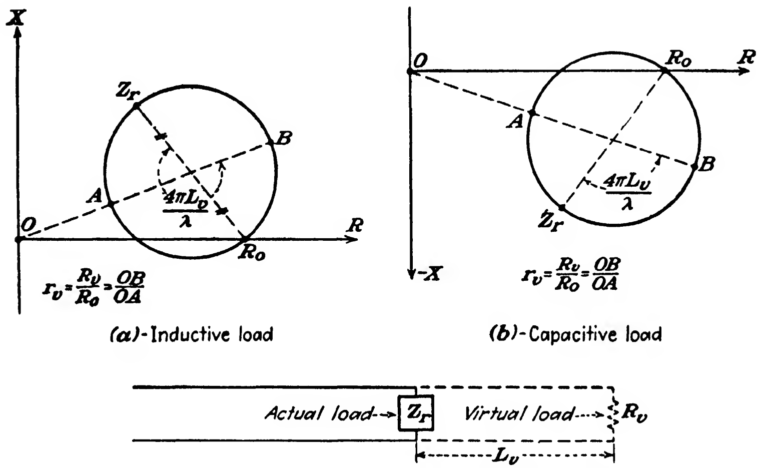


FIG. 25.—*Construction for the virtual termination of a lossless line.*

7.5. Voltage Distribution along Reflecting Lines. Standing Waves. The value taken by the voltage and current as a function of time and position along a section of line with generic load will now be considered, assuming as usual steady-state conditions.

The analysis will first be carried out in general, thereby extending the discussion of Sec. 4.1 to the case when reflection is present. Particular attention will subsequently be given to the distribution of voltage and current in a lossless line.

The ratio of voltages at the two ends of a line is, in general [Eq. (97)],

$$\frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\cosh [(\alpha + j\beta)L + \rho]}{\cosh \rho}$$

and the ratio of currents (98)

$$\frac{\mathbf{I}_1}{\mathbf{I}_2} = \frac{\sinh [(\alpha + j\beta)L + \rho]}{\sinh \rho}$$

We are now concerned, not with the terminal values, but with values all along the line. The above equations must therefore be applied to a section of variable length, whose output we will consider fixed at some reference point conveniently situated on the line, while the input is at a variable distance x from the reference point, counted positively in the direction of the source. By giving x a range of values, negative and positive, we will then obtain ratios of the voltage at various points to the voltage at the reference point and also the distribution, for a given instant of time.

For reasons of convenience, we shall place our reference point *where the reflection factor* ρ *has a real value.* There is a value of ρ (a function of the impedance toward load) for every point along the line; all of these values are represented by points on the ρ plane (Fig. 20*a*) located along a straight path.

This path must cross the axis of reals ($\tau = 0$) somewhere, although the intersection may correspond to a virtual point of the line, beyond either termination.

We will then write σ for ρ in the general expressions. Also, $\mathbf{V}(x, t)$ will be written in place of $\mathbf{V}_1$, to signify the voltage at a distance x from the reference point and at time t; and $\mathbf{V}(0, t)$ will be the new notation for $\mathbf{V}_2$; the currents will be handled in the same way. Finally, x will take the place of L. Accordingly, we have

$$\frac{\mathbf{V}(x, t)}{\mathbf{V}(0, t)} = \frac{\cosh [(\alpha + j\beta)x + \sigma]}{\cosh \sigma}$$

$$\frac{\mathbf{I}(x, t)}{\mathbf{I}(0, t)} = \frac{\sinh [(\alpha + j\beta)x + \sigma]}{\sinh \sigma}$$

Like x, t must be counted from some reference instant. For convenience, let $t = 0$ when the voltage at $x = 0$ is a positive maximum. Accordingly, we may write

$$v(0, t) = \hat{V}_0 \cos \omega t \qquad \mathbf{V}(0, t) = V_0 e^{j\omega t}$$

and we have

$$\mathbf{V}(x, t) = \frac{\hat{V}_0 e^{j\omega t}}{\cosh \sigma} \cosh [\alpha x + \sigma + j\beta x] \tag{147}$$

Taking as the origin of time the instant at which $\mathbf{I}(0, t)$ is a positive maximum, we obtain

$$\mathbf{I}(x, t) = \frac{\hat{I}_0 e^{j\omega t}}{\sinh \sigma} \sinh [\alpha x + \sigma + j\beta x] \tag{148}$$

Note that the origin of time is not the same for the two Eqs. (147) and (148); t stands for two different variables in the two equations, except when the voltage and current are in phase at $x = 0$ (as for the reactive line).

The hyperbolic functions of (147) and (148) may be expanded into sums and differences of exponentials, as follows:

$$\mathbf{V}(x, t) = \frac{\hat{V}_0}{e^\sigma + e^{-\sigma}} [e^{\sigma+\alpha x+j(\omega t+\beta x)} + e^{-(\sigma+\alpha x)+j(\omega t-\beta x)}]$$

$$\mathbf{I}(x, t) = \frac{\hat{I}_0}{e^\sigma - e^{-\sigma}} [e^{\sigma+\alpha x+j(\omega t+\beta x)} - e^{-(\sigma+\alpha x)+j(\omega t-\beta x)}]$$

Noting that the *instantaneous values* of voltage and current are the real part of the corresponding vector expressions and that the real part of $e^{j\varphi}$ is $\cos \varphi$, we have finally

$$v(x, t) = \frac{\hat{V}_0}{e^\sigma + e^{-\sigma}} [e^{\sigma+\alpha x} \cos (\omega t + \beta x) + e^{-(\sigma+\alpha x)} \cos (\omega t - \beta x)] \quad (149)$$

$$i(x, t) = \frac{\hat{I}_0}{e^\sigma - e^{-\sigma}} [e^{\sigma+\alpha x} \cos (\omega t + \beta x) - e^{-(\sigma+\alpha x)} \cos (\omega t - \beta x)] \quad (150)$$

[*Voltage and current among the line in the general case. Direct and reflected waves*]

Voltage and current are expressed above as the resultants of two waves traveling with the same velocity ω/β. This interpretation follows closely along the lines of the discussion of Sec. 4.1.

Eqs. (149) and (150) could have been obtained as the solution of a differential equation (Sec. 13.3). That is the method generally used in problems involving the propagation of disturbances in continuous media.

The general expressions will now be applied to lines in which $\alpha \ll \beta$. Assuming that the line length is of the order of λ, the term αx may be neglected for all values of x corresponding to points of the line proper. The reference point may now be taken at the virtual load for which $\rho = \sigma$ by hypothesis (Sec. 7.4). Moreover, since the virtual load is resistive, voltage and current at the virtual load are in phase, hence t is the same in the voltage and current equations, which can be written

$$v(x, t) = \frac{\hat{V}_0}{e^\sigma + e^{-\sigma}} \left[e^\sigma \cos\left(\omega t + \frac{2\pi x}{\lambda}\right) + e^{-\sigma} \cos\left(\omega t - \frac{2\pi x}{\lambda}\right) \right] \quad (151)$$

$$i(x, t) = \frac{\hat{I}_0}{e^\sigma - e^{-\sigma}} \left[e^\sigma \cos\left(\omega t + \frac{2\pi x}{\lambda}\right) - e^{-\sigma} \cos\left(\omega t - \frac{2\pi x}{\lambda}\right) \right] \quad (152)$$

[*Voltage and current along line, neglecting losses*]

The distribution of voltage or current at any time may be found by "freezing" the variable t at some value, as was done in Sec. 4.1. For $t = 0$, we have

$$v(x, 0) = \hat{V}_0 \cos \frac{2\pi x}{\lambda} \qquad i(x, 0) = \hat{I}_0 \cos \frac{2\pi x}{\lambda}$$

showing that both v and i vary sinusoidally along the line at this particular instant, *i.e.*, when they go through a maximum at the virtual load. The distribution of voltage at this time is shown as the solid line of Fig. 26a. A fraction of a cycle later it will have changed as indicated by the dotted line, which is still sinusoidal because it is the sum of two sinusoidal waves, although of different phase and amplitude.

In general, the distribution is analyzed more conveniently with the help of the *polar diagram* of voltage (or current). The polar diagram was used in the discussion of voltage along the matched line (Sec. 4.1); it is no other than the locus, or path, of the point $\mathbf{V}(x, t)$ for x variable. (For those who are used to thinking of $\mathbf{V}$ as a vector, the *point* $\mathbf{V}$ is best described as the arrow of the vector.)

To obtain the polar diagrams, the expressions for $\mathbf{I}(x, t)$ and $\mathbf{V}(x, t)$ must be used. The value assigned to the parameter t (time) does not matter, because the only change which the polar diagram undergoes from time 0 to time t is a rotation ωt in the positive (counterclockwise) direction. Evidently, therefore, once the diagram is at hand for $t = 0$, it can readily be made available for any value of time. Accordingly we may substitute $t = 0$ in Eqs. (147) and (148), in addition to making $\alpha x = 0$ and $\beta x = 2\pi x/\lambda$ to account for the lossless character of the line. Thus, we have

$$\mathbf{V}(x, 0) = \frac{\hat{V}_0}{\cosh \sigma} \cosh \left[\sigma + j \frac{2\pi x}{\lambda} \right]$$
$$= \hat{V}_0 \left[\cos \frac{2\pi x}{\lambda} + j \tanh \sigma \sin \frac{2\pi x}{\lambda} \right] \quad (153)$$
$$\mathbf{I}(x, 0) = \frac{\hat{I}_0}{\sinh \sigma} \sinh \left[\sigma + j \frac{2\pi x}{\lambda} \right]$$
$$= \hat{I}_0 \left[\cos \frac{2\pi x}{\lambda} + j \coth \sigma \sin \frac{2\pi x}{\lambda} \right] \quad (154)$$

If $\mathbf{V}(x, 0)$ and $\mathbf{I}(x, 0)$ are considered as points of planes, the loci described by these points when x is varied are ellipses, as may be verified by separately equating the real and imaginary parts of the above equations. If we let

$$\mathbf{V} = v + ju \qquad \mathbf{I} = i + jh \quad (155)$$

we have

$$\begin{cases} v = \hat{V}_0 \cos \dfrac{2\pi x}{\lambda} \\ u = \hat{V}_0 \tanh \sigma \sin \dfrac{2\pi x}{\lambda} \end{cases} \quad (156)$$

$$\begin{cases} i = \hat{I}_0 \cos \dfrac{2\pi x}{\lambda} \\ h = \hat{I}_0 \coth \sigma \sin \dfrac{2\pi x}{\lambda} \end{cases} \quad (157)$$

Equations (156) are the parametric equations of an ellipse, whose radius vector has a maximum value $\hat{V}_0$ along the axis of reals, a minimum value $\hat{V}_0 \tanh \sigma$ along the axis of imaginaries. For every x (point of the line) there is a point of the ellipse, and the radius vector at that point is the peak value of voltage at x.

The correspondence of points on the line to points on the voltage ellipse is best understood by graphical construction (Fig. 26*a*). Taking $\hat{V}_0$ as the unit length, two circles of radii *1* and $\tanh \sigma$ are drawn. A line is then traced at angle $2\pi x/\lambda$ with the axis of reals, intersecting the circles at two points. The *real coordinate* of the point on the *outer circle* and the *imaginary coordinate* of the point on the *inner circle* are the components of the voltage vector. This construction follows immediately from Eqs. (156) and is, as the reader will recognize, the standard procedure for constructing an ellipse.

As the angle $2\pi x/\lambda$ is varied, the real component of $\hat{V}_0$ takes a range of values which, plotted against a distance axis, show how the instantaneous voltage is distributed over the line. To obtain the distribution for any particular time t, the voltage ellipse

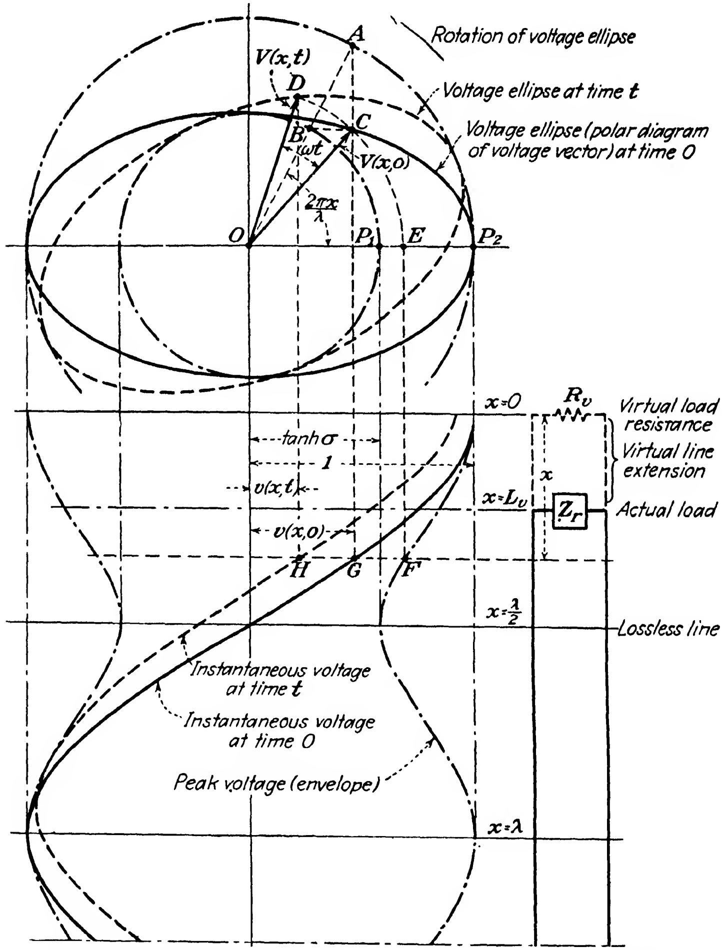


FIG. 26a.—*Voltage distribution along a lossless line with generic termination.*

must be rotated by an angle ωt. Each voltage vector is so rotated and its real component changes accordingly.

If, instead of the real component, we plot the length of the voltage vector (radius vector of the ellipse) against x, we obtain the variation of *voltage amplitude* (or peak voltage) along the line. This plot does not change with time. It is periodic along x, with period $\lambda/2$, but not sinusoidal. Its equation is

$$\hat{V}_x{}^2 = \frac{\hat{V}_0{}^2}{2}\left[1 + \tanh^2\sigma + \left(\cos\frac{4\pi x}{\lambda}\right)(1 - \tanh^2\sigma)\right] \quad (158)$$

The ratio of the maximum amplitude to the minimum is called *standing wave ratio* (s.w.r.). Its value is coth σ, which is also the value of the *virtual load number* (Sec. 7.4). Thus, if the line is loaded by a resistance twice as high as the characteristic resistance, the voltage maxima and minima will be in the ratio of 2 to 1, the load terminals being a point of maximum. If the load resistance were half the characteristic value, the standing wave ratio would be the same but the load terminals would be a point of minimum voltage. In this case, the virtual load would be a quarter wavelength behind the load proper.

If the line is open- or short-circuited, the standing wave ratio is infinity. The polar diagram in either case degenerates into a straight line, and the voltage amplitude is zero at points spaced by intervals of a half wavelength (nodes). This behavior of the line parallels closely the vibrations of elastic bodies "driven" at some point and rigidly held by supports incapable of absorbing energy, such as a taut string fixed at one end and tied to a tuning fork at the other.

The current distribution is entirely similar to that of voltage, except for this, that voltage maxima correspond to current minima, and vice versa. The current ellipse is rotated by a right angle with respect to the voltage ellipse. The standing ratio is the same, whether the voltage or current distribution is considered. At all points of minimum and of maximum (voltage or current) the voltage and current are in phase; hence their product is the power, which must be constant at all points. The maximum to minimum ratio of voltage and current must therefore be the same, in agreement with Eqs. (156) and (157).

The polar diagram has been studied in the two cases of an attenuating line without reflection (Sec. 4.1) and of a *lossless* line when reflection is present. In the first case it was found to have the form of a spiral; in the second, of an ellipse.

In the most general case, that of a line with appreciable attenuation with nonmatching load, which was studied analytically in the first part of this section, the polar diagram has the form of an elliptic spiral.[6] From an engineering standpoint, the general case is not very important because reflection is purposely kept low in long sections of line where the over-all attenuation is large (Sec. 9.2).

7.6. The Line as an Impedance Measuring Device. Most engineers are acquainted with the difficulties experienced in the measurement of impedance at very high frequencies. These difficulties arise, chiefly, because two points not coincident in space are also separated electrically, *i.e.*, they have an impedance between them; and two points not infinitely distant always have an admittance between them. While at audio and broadcast frequencies it is comparatively easy to keep these stray effects so low that the electrical relationships between points of a circuit need not depend on the size and position of its parts, at vhf this is impossible.

If the impedance to be measured is to be connected to a measuring device, the impedance and admittance of the connection will invariably introduce an intolerable error. The only solution lies in the use of the connection itself as a measuring device. The theory of the lossless line, as developed in the preceding sections, indicates how this can be accomplished. The procedure is as follows:

a. The unknown impedance is connected to one end of a section of line, while the other end is driven by a source of the required frequency. The length of the line should be a half wavelength at least.

b. The distances from the unknown impedance to the nearest point of maximum voltage on the line and to the nearest point of minimum voltage are recorded; or, alternatively, maximum and minimum current points are used.

c. The maximum and minimum values of voltage (or current) along the line are measured.

The measurement described under (*b*) determines the value of the virtual line extension (Sec. 7.4), while (*c*) yields the standing wave ratio. The unknown impedance may be computed from these data.

An example will illustrate the procedure. Suppose that the r.m.s. voltage has been read continuously along the line so that a plot of the voltage distribution is available. This is shown in

Fig. 26*b* for two cases: when the unknown is inductive and when it is capacitive.

Expressing the unknown impedance in terms of σ and τ, we have (139)

$$Z_r = R_0 \coth \rho = R_0 \frac{\coth \sigma + j \tan \tau}{1 + j \coth \sigma \tan \tau}$$

Recalling that $\coth \sigma = r_v$ and $\tau = 2\pi L_v/\lambda$, we have the value of the unknown impedance in terms of the virtual parameters (Sec. 7.4)

$$Z_r = R_0 \frac{r_v + j \tan \dfrac{2\pi L_v}{\lambda}}{1 + jr_v \tan \dfrac{2\pi L_v}{\lambda}} \tag{159}$$

[*Impedance of termination in terms of virtual extension and standing wave ratio*]

Letting $\hat{V}_2$, x_2 stand for the maximum value of peak voltage and the corresponding distance from the load end, and $\hat{V}_1$, x_1 for the minimum value and corresponding distance, we have (see Fig. 26*b*)

$$r_v = \frac{\hat{V}_2}{\hat{V}_1} \qquad \frac{2\pi L_v}{\lambda} = \frac{\pi}{2}\left(1 - \frac{x_2}{x_1 - x_2}\right)$$

Substituting these values, the following is obtained:

$$Z_r = R_0 \frac{\hat{V}_2 \sin A + j\hat{V}_1 \cos A}{\hat{V}_1 \sin A + j\hat{V}_2 \cos A} \qquad \left[A = \frac{\pi}{2}\frac{x_2}{x_1 - x_2}\right] \tag{160}$$

[*Impedance of unknown termination in terms of measurements obtained on the line*]

If currents have been measured instead of voltage, the maximum current amplitude should take the place of $\hat{V}_2$, and x_1 should stand for the distance from the unknown impedance to the point of highest current. Peak or effective values may be used indifferently, of course.

For the experimental technique of measuring voltage and current distribution, the reader is referred to the literature.[7]

A geometric construction based on the polar diagram (Sec. 7.5) may be used in place of Eq. (160). Suppose, for example, that the distribution of peak voltage along the line has been obtained

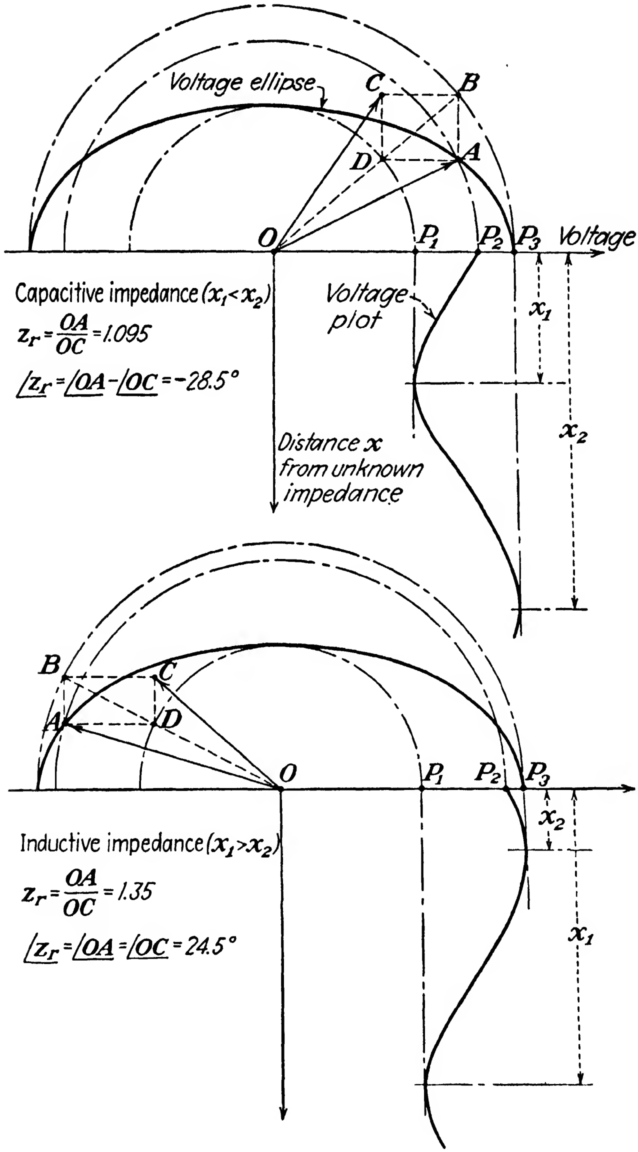


FIG. 26b.—*Graphical determination of an unknown impedance from measurements taken along a line.*

(Fig. 26*b*). Three concentric circles are drawn, with radii V_1, V_r, and V_2 (V_r = voltage at the unknown impedance). Then the voltage ellipse can be constructed—for simplicity, oriented as shown, *i.e.*, corresponding to the instant when the voltage at the virtual load is a positive maximum. Only a small arc of the ellipse need be drawn, so as to locate point A at the intersection of the ellipse with the circle of radius V_r. Since the radius vector of the ellipse represents the voltage vector, point A determines the latter for the position $x = 0$ (unknown impedance terminals), or $\overrightarrow{OA} = \mathbf{V}_r$.

Proceeding with the construction, points B, C, and D are located. $\overrightarrow{OC}$ is the radius vector of a second ellipse, identical to the voltage ellipse except for a right-angle rotation. The new ellipse is therefore similar to the current ellipse (Sec. 7.5) except for the dimensions, the major radius being $\hat{V}_0$ instead of $\hat{I}_0 \coth \sigma$. [$\hat{V}_0$ and $\hat{I}_0$ peak values at the virtual load, see Eqs. (157).] It follows that $\overrightarrow{OC}$ is related to the current vector at the load terminals by the equation

$$\overrightarrow{OC} = \mathbf{I}_r \frac{\hat{V}_0}{\hat{I}_0} \tanh \sigma = R_0 \coth \sigma \tanh \sigma \, \mathbf{I}_r = R_0 \mathbf{I}_r$$

Consequently, we have

$$\frac{\overrightarrow{OA}}{\overrightarrow{OC}} = \frac{\mathbf{V}_r}{R_0 \mathbf{I}_r} = \frac{Z_r}{R_0}$$

and the ratio of the complex numbers represented by the points A and C (or the vectors $\overrightarrow{OA}$ and $\overrightarrow{OC}$) is the *impedance number* for the unknown load

$$\frac{\overline{OA}}{\overline{OC}} = |z_r| \qquad \angle COA = -\underline{/z_r}$$

Figure 26*b* illustrates the difference in voltage distribution for the two cases of inductive and capacitive load. When the load is capacitive, the voltage falls as the point of measurement moves away from the load; the reverse is true of an inductive load.

When the current distribution has been measured, the construction is the same, except for the order in which the two vectors $\overrightarrow{OA}$ and $\overrightarrow{OC}$ are obtained.

7.7. Balanced and Unbalanced Transmission Lines. In our discussion of transmission lines it has not been necessary to introduce distinctions, except for the fact that electrically long lines, when short in physical length, can usually be regarded as free of losses.

For practical purposes, however, it is well to have in mind the fundamental difference between *balanced* and *unbalanced* lines. This subdivision has been made before (Sec. 1.2); a system is said to be balanced when its connections with ground do not carry current; the opposite is true of an unbalanced system.

In order to make the definition general, we should think of *ground* as being, not necessarily the surface of the earth, but any conducting enclosure or shield surrounding the electrical system. (If such an enclosure is not in evidence, its place is taken by the surface of the earth, which, being virtually indefinite in extent, is equivalent to a closed surface.)

A balanced system, then, is one in which no current flows into or out of the shield surrounding the system at any point. If a transmission line is used to connect parts of such a system, there will be no longitudinal flow of current through the part of the shield that surrounds the line (or, if the shield were removed, through the ground). There will be two conductors within the shield, carrying equal currents in opposite directions. As the shield is not needed to carry current, it may be dispensed with without causing ground currents to flow. Most balanced lines are unshielded. Many balanced lines may run close together without interference.

The term *unbalanced* usually refers to those lines which consist of one conductor and a shield, or one conductor and ground. The shield (or ground) is used as a return. A shield is required in the vast majority of cases, since the use of ground as a return causes interference as well as serious losses, particularly at high frequency.

A shielded unbalanced line generally takes the form of a coaxial, or concentric, cable. This is altogether different from telephonic cables, which actually consist of many balanced lines, or pairs, within a common sheath that serves as a mechanical protection and does not have an electrical function.

A coaxial cable is more expensive than an unshielded balanced line of comparable attenuation. However, unshielded lines cannot be used above a certain frequency; hence, the widespread use of coaxial lines in the higher frequency ranges. Shielded balanced lines suitable for very high frequency are also used, although of less simple construction.

When the distance is short, the choice between the two types of line may be affected by the terminations. A balanced line is the logical choice, for example, in connecting a push-pull amplifier to a dipole antenna.[10] It is not impossible, however, to match a balanced line to an unbalanced load, or vice versa. Suppose a coaxial feeder had to be used to energize the dipole antenna (Fig. 27*a*). The antenna is balanced when equal currents flow through its two arms at the junction with the cable. Without attempting a rigorous discussion, which would require field concepts, we will recall the well-known fact that high-frequency currents flow only on the surface of conductors and do not traverse them. There are, in consequence, three possible paths for the current to take as it leaves the antenna: the outside surface of the shield, the inside surface of the shield, and the surface of the inner conductor. The last two are of necessity equal and opposite, because the charge densities on opposite surfaces of a dielectric are equal and opposite. Hence, to ensure balanced operation, we must make it impossible, or at least difficult, for the current to follow the third path (the outer surface of the shield).

The device of Fig. 27*b* accomplishes this purpose. The unbalancing current would have to flow along the surface of a resonant cavity, formed by the hollow space between the shield and an outer *skirt*, a quarter-wave long. This cavity is equivalent to a quarter-wave line terminated in a short circuit, or *stub* (Sec. 10.4). If a given current were to flow in and out of the cavity at its mouth, a very much greater current would have to circulate at the bottom of the cavity. Or we might say that the impedance of the cavity is effectively in series with the path of the unbalancing current, although expressions borrowed from circuit theory are not generally useful in this type of problem. The skirt arrangement may well be described as a *trap* which stops undesirable currents from flowing.

As an alternative means of preventing the flow of unbalancing currents, a magnetic field may be used to convey energy from the balanced to the unbalanced part of the system. This device is very commonly used at all frequencies. It consists of a transformer which has a grounded center tap on the balanced side (Fig. 27c).

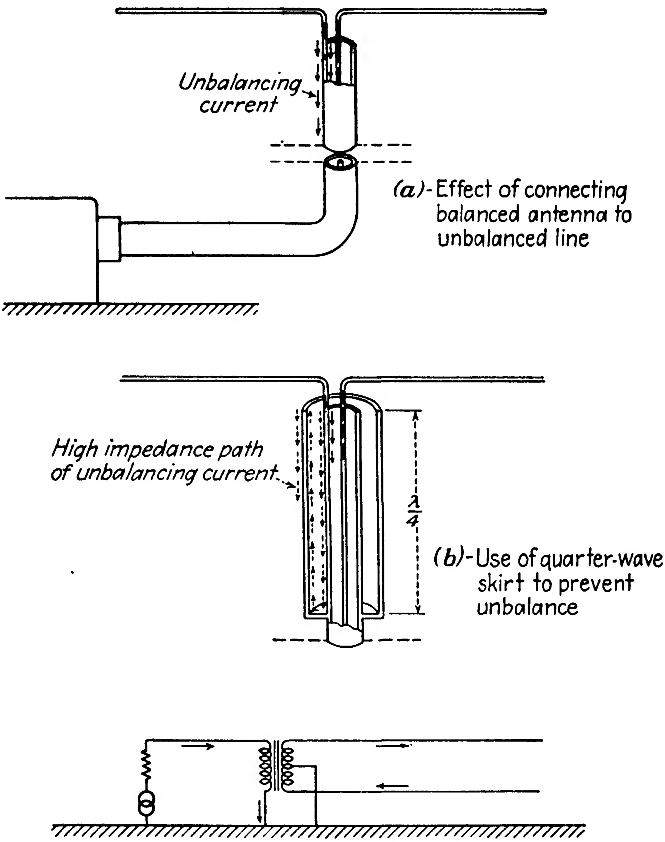


FIG. 27.—*Junction of balanced and unbalanced systems.*

7.8. Illustrative examples.

Optimum proportions of coaxial cables. For a given shield radius b, (Fig. 23), there is a value of core radius a which minimizes the attenuation of the coaxial cable. This is also true of the parallel line (for hf operation), but with this difference:

in the parallel line, the cost of the line is affected by both a (wire radius) and b (wire spacing); in the coaxial, the cost is influenced primarily by the value of b and only slightly by a. Hence, for the coaxial line, a can be selected exclusively on the basis of performance.

(An analogous problem would be that of finding for a 100-gal. drum the diameter which permits the highest pressure in the drum, given the quality and weight of steel to be used in its construction. The diameter does not affect the cost appreciably, while the weight of steel does. For any given weight there is therefore an optimum diameter, which can be determined on the basis of strength only.)

We may write the attenuation of the coaxial line, for a given frequency, in the form (Sec. 7.2)

$$\alpha = \frac{K}{b} \frac{1 + b/a}{\log b/a}$$

where K may be regarded as a constant. If we assign a value to b, there will be a value of a which minimizes α; there is advantage, however, in choosing b/a as the variable, since the optimum value of this must be a numerical constant independent of b, as the form of the above equation indicates. Letting $b/a = x$, we find that α is a minimum when

$$\frac{d\alpha}{dx} = \frac{K'}{b} \frac{\ln x - 1 - 1/x}{\ln^2 x} = 0 \qquad (K' \neq K)$$

which may be written

$$1 + \frac{1}{x} = \ln x$$

The above may be solved by trial and error, or graphically. The solution is

$$x = \frac{b}{a} = 3.59$$

The variation of α with x is shown in Fig. 23*b*.

The characteristic impedance of the coaxial cable is a function of x only (Sec. 7.1). Its optimum value is therefore

$$R = 138 \log 3.59 = 76.6 \text{ ohms}$$

The problem of parallel lines is not essentially different, except that b, the wire spacing, is not economically important unless a shield has to be provided. For shielded pairs in high-frequency operation, there is an optimum value of b/a consistent with the lowest attenuation for a given shield diameter. Such lines are not used unless a balanced circuit is necessary.

Standing wave ratio on a line many wavelengths long. The discussion on voltage and current distribution has been carried out (Sec. 7.5) for a line of short length compared to λ. Long lines in hf operation are generally matched to their terminations, in which case the analysis of matched lines (Sec. 4.1) applies. If this precaution is not taken there will be standing waves on the line; the s.w.r. will be sensibly uniform over a few wavelengths but will decrease progressively from the load to the source.

Consider the following example: A 77-ohm coaxial cable is used to connect a uhf transmitter to its radiator (antenna). The cable is 200 m. long, with $b = 1$ cm. (b = inner radius of shield.) The frequency is 100 mc.; consequently, the attenuation (Fig. 23c) is 1.08 db, or 0.124 nep./100 m. The standing wave ratio at the load (the antenna) is 1.5. Values of the s.w.r. at the transmitter and halfway down the cable are required.

At the load, the s.w.r. is coth σ (146), where σ is the real part of the complex number ρ, a function of the load impedance (Sec. 5.4). A value of ρ may be assigned to any point of the line, as a function of the *impedance toward load* at that point. At a distance x from the load this value is

$$\rho(x) = \rho(0) + \theta x$$

where $\rho(0)$ is the value of ρ at the load ($x = 0$), θ the transfer constant of the line. The real part of the above may be written

$$\sigma(x) = \sigma(0) + \alpha x$$

Hence, the standing wave at distance x from the load is

$$r_v(x) = \coth \sigma(x) = \coth [\sigma(0) + \alpha x] = \frac{r_v(0) + \tanh \alpha x}{r_v(0) \cdot \tanh \alpha x + 1}$$

where $r_v(0)$ is the s.w.r. at the load. Numerically,

for $x = 100$ m. (halfway: $\alpha x = 0.124$ neper):

$$r_v = \frac{1.5 + 0.1233}{1.5 \times 0.1233 + 1} = 1.37$$

for $x = 200$ m. (at the transmitter: $\alpha x = 0.248$ neper):

$$r_v = \frac{1.5 + 0.243}{1.5 \times 0.243 + 1} = 1.28$$

Suggested Exercise. Construct the curve coth σ vs. σ, from the table. Find the value of σ corresponding to coth $\sigma = 1.5$ on the curve. To this value of σ, add $\alpha x = 0.248$. The corresponding value of coth σ is 1.28. This graphical method takes the place of the computation above, and solves the problem for any set of data.

CHAPTER VIII

PROBLEMS INVOLVING IMPEDANCE TRANSFORMATION

8.1. Classification of Impedance Transformation Problems. The problem of *coupling* a generator to its load so as to obtain the highest flow of power and that of *matching* a line to its terminations will be taken up in the following sections. These problems are often solved by the use of transmission lines, and their discussion is a logical sequence to the analysis of electrically long lines.

On the other hand, the problems mentioned above may be considered as particular instances of a wider class of problems, all of which call for the design of an *impedance transforming* network of suitable characteristics and of reactive character.

The characteristics of this network may be specified over a range of frequencies or at a single value of frequency. The coupling problem and the matching problem are essentially single-frequency problems, although the performance in the immediate neighborhood of the operating frequency may be of importance.

A rough classification of all the *impedance transformation problems* may hinge upon the extent of the frequency range involved. The chief object of the problem is essentially the same in all cases, that of ensuring *maximum power flow* between the source and the load; this may be required over a given range of frequencies or at a single frequency. Discrimination against frequencies outside the operating range may also be an object. We may list the various impedance transformation problems in order of increasing complexity, as follows:

a. The coupling problem. This is the problem of obtaining maximum flow of power from a given source to a given load at some particular frequency. It is solved by the insertion of a coupling network which transforms the load impedance so as to obtain the highest power from the source—hence, into the load, since the coupling network does not absorb appreciable power.

In some cases, a high degree of *selectivity* (Sec. 9.4) may be required, to prevent all frequencies outside a narrow band from being transmitted. Often the network is made adjustable, so as to vary the operating frequency over a range; such networks are called *tuners* (Sec. 10.5).

A modification of the coupling problem calls for maximum voltage across the load rather than power, as in the coupling of voltage amplifier stages.

The operating frequency is generally high in this, as in other single-frequency problems. Such problems have to do with the transmission of *carrier* frequencies, while *audio* frequencies require uniform transmission over a relatively wide range. Source and load are not, as a rule, pure resistances.

b. The matching problem. The only difference between this and the coupling problem is that either the source or the load is now replaced by the output (or input) of a long transmission line. The problem calls for the design of reactive networks suitable for matching the line to its terminations. As the characteristic impedance of lines is practically a pure resistance, especially at high frequencies, maximum flow of power is thereby obtained.

The matching problem may be regarded as a particular case of the coupling problem and is handled in the same way.

c. The transformer problem. This problem arises when the transmitted power is to be kept close to the maximum over a wide frequency range, both source and load being resistive in character but of different values. Essentially, this is an audio-frequency problem, solved by the use of close-coupled transformers. However, problems of this type are not uncommon at higher frequencies. The vhf solution, employing many short sections of line, is discussed in Sec. 10.1.

d. The filter problem. This is the most complex of all impedance transformation problems. Filters are generally inserted between equal resistive terminations; within the transmitted range of frequency, therefore, they must have little or no transforming action on the load impedance. Outside this range, they must transform the load impedance into a value approaching a pure reactance, so as to reduce the flow of power below a specified limit.

The design of filters is not generally carried out on the basis of their impedance transforming action; for this reason, the filter problem is seldom associated with other impedance transformation problems. Yet the distinction between a coupling network and a band-pass filter is not a very sharp one.

8.2. Condition for Maximum Power Transfer. The coupling problem, problem 1 in the preceding section, calls for the transformation of the load impedance to that value which, connected to a given source at a given frequency, absorbs from this source the greatest amount of power. This transformation must be accomplished so that only a negligible part of the power drawn from the source is lost in the transforming network.

The value of *transformed impedance*, as defined above, is readily found. Let Z_t stand for this value and Z_g for the source impedance. The power received by Z_t may be written

$$P = \frac{E^2}{|Z_g + Z_t|^2} R_t = \frac{E^2}{(R_g + R_t)^2 + (X_g + X_t)^2} R_t \quad (161)$$

where E is the source e.m.f. This power depends on both R_t and X_t; but if we assign some definite value to R_t, then the maximum of power will occur for $X_t = -X_g$. This conclusion does not depend on the value assigned to R_t; the condition $X_g = -X_t$ must therefore be considered necessary (but not sufficient) if P is to have the highest value possible. If we imagine this condition to be met, then P is written

$$P = \frac{E^2}{(R_g + R_t)^2} R_t$$

This expression depends on R_t only, and we soon find (or remember from elementary circuit theory) that it is a maximum when $R_g = R_t$. This is the second condition. Both conditions together determine the required value of transformed impedance, as follows:

$$Z_t = R_t + jX_t = R_g - jX_g \quad (162)$$

[*Value of transformed impedance for maximum transmitted power*]

We may express the above by this simple statement: A source delivers the highest amount of power when connected to a load whose impedance is the *conjugate* of the source impedance. When such a condition exists there is said to be a *conjugate match* of impedance at the junction of the source to its load;

the condition itself is known as the condition of *maximum power transfer*. (The abbreviation m.p.t. will be used occasionally.)

It will appear (Sec. 9.2) that if a transmission system includes a source, a load, and an intermediate network free of losses, the m.p.t. condition, if it is met at either end of the network, must be met at the other end as well and at all inner junctions of the network. The system then constitutes a *maximum power transfer chain* (Sec. 2.10).

8.3. Representation of Impedance Changes Due to Branch Additions. Having set the goal of the coupling problem as the transformation of a given load impedance Z_r into the value Z_t specified by (162), let us now turn our attention to the means available for carrying out the transformation.

Three types of impedance transformation have already been considered (Fig. 24). While it is easy to see that neither of the first two means (the inductor and the capacitor) solves the problem by itself, we shall find presently that a combination of both is sufficient in general. (We would expect this to be true, since the impedance is defined by two real numbers.)

It is worth while, at this point, to develop a method that will enable us to visualize what happens to the impedance of a given two-pole when circuit elements are added to it in various ways. The method consists essentially of the use of the complex plane, previously discussed in connection with the function ρ (Sec. 5.7). At first, no restriction will be imposed on the type of circuit element added; then, reactive elements will receive particular attention. Their use will lead to a solution of the coupling problem.

The method relies upon the use of one or more points (on as many planes) to represent impedance. It should be noted that the word *impedance* is sometimes used broadly, sometimes in a narrow sense. In the narrow sense, impedance is the complex number

$$Z = \frac{\mathbf{V}}{\mathbf{I}}$$

When used broadly, impedance is that property of a two-pole, or *load*, which may be defined indifferently by Z or by Y, or by other complex numbers, or by a schematic.

In general, a property is not the same as the number used to express it, although both are expressed by the same word. Thus, "hardness" is a property of matter, in the broad sense. In the narrow sense, we might say that something has "hardness 5," using some particular hardness scale. "Hardness" now denotes the number used to measure the property, instead of the property itself.

It is now possible to say, without ambiguity, that two complex numbers, the admittance Y and the impedance Z, may be used (among others) to define or measure the property *impedance* of a given two-pole. This is almost a truism, of course; yet such is the inadequacy of words to express technical thought that much of the basic literature prefers not to make this statement and uses Z everywhere to the exclusion of Y, in spite of the fact that Y is more convenient in the majority of cases.

Now, both Z and Y, like all complex numbers, may be represented by points (not by vectors; only quantities having a direction in space are legitimately called *vectors*, although the phrases *voltage vector* and *current vector* and the corresponding arrows have become so common that a change is hardly warranted).

The admittance Y is represented by the point of coordinates G and B on the Y plane; the impedance Z, by the point of coordinates R and X on the Z plane. A circuit addition corresponds to a "jump" of both points to new positions; we therefore have two distinct ways of visualizing a change of impedance due to circuit additions, by following the movements of either the Z point or the Y point.

We begin by locating the points Z_r and Y_r, representing the load impedance upon which we are effecting the transformation. Now if we add a branch, or two-pole, *in series* with the load, the impedance of the combination will be the sum of the complex numbers Z_r and Z_s (load and branch impedances). On the Z plane, the impedance point will be brought to its new position by moving it a distance R_s along the axis of reals and a distance X_s along that of imaginaries (Fig. 28*a*).

Addition of a shunt branch has a similar effect on the Y point, displacing it horizontally by G_p and vertically by B_p (Fig. 28*b*).

Now, suppose we add first a series and then a parallel or shunt branch. If we consider the Z point, the series addition is readily interpreted (Fig. 28*a*); as for the shunt addition, we have as yet

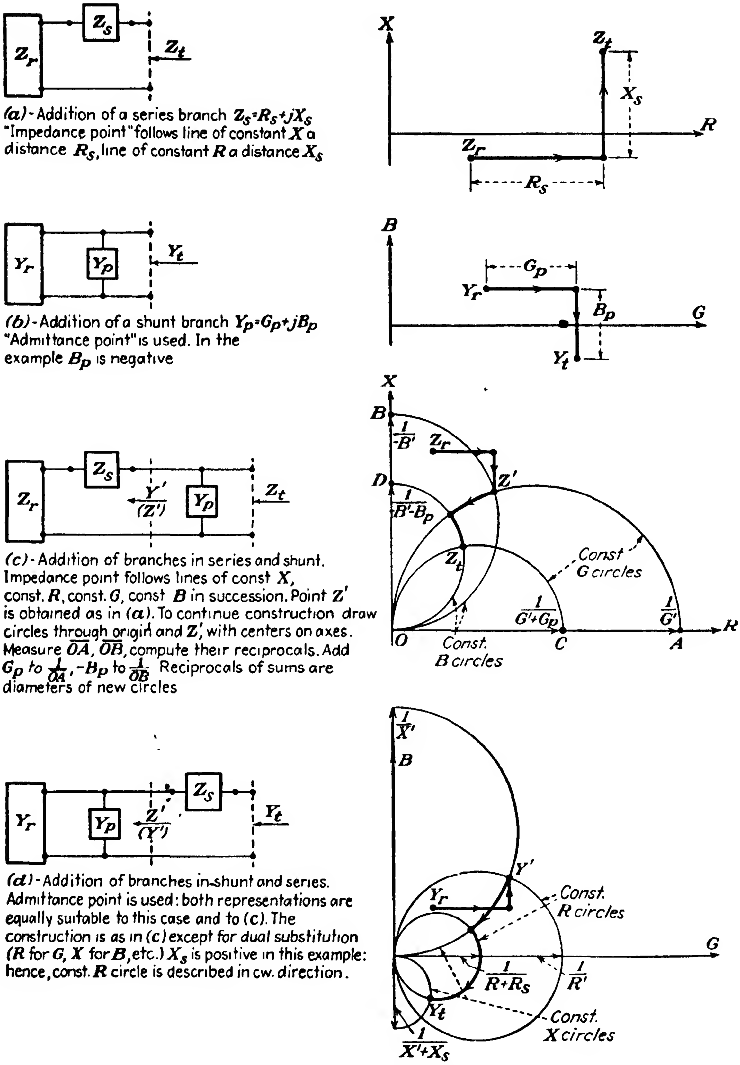


FIG. 28.—*Geometric representation of branch additions.*

no way of representing the movement of the Z point due to such an addition, and it is likely that there will be a difference between this type of movement and those considered up to now.

The difference is formal, however, not fundamental. When we were moving Z a distance R_s along the R axis, we were actually following a *line of constant* X (specifically, the line $X = X_r$) up to its intersection with a *line of constant* R (specifically, the line $R = R_t + R_s$). In general, it is clear that if one parameter only is altered, the line followed will be associated with a constant value of the other parameter. Now, when we add the shunt branch, we are in effect changing the conductance and susceptance of the two-pole by given amounts. The impedance point must therefore follow a constant B line and a constant G line successively (or vice versa).

Such lines on the Z plane are not parallel to the axes; hence they cannot be traced quite as readily as the constant R and constant X lines. They could be readily followed, however, if they were drawn on the paper once for all. A map of the Y function on the Z plane would then be available (Sec. 5.7). The lines of this map would be transformations of parallel orthogonal lines; we shall discuss them in the next section.

For practical purposes, the effect of the series-parallel addition may be obtained without drawing the whole map beforehand; the construction of Fig. 28*c* may be used. A similar construction, applicable when the Y point is considered, is given in Fig. 28*d*. Both constructions are justified by the analysis of the following section.

★8.4. Mapping of the Function Y on the Z Plane, and Vice Versa. We may write

$$Z = |Z|e^{j\varphi} = \frac{1}{|Y|} = \frac{\cos\varphi}{G}\,e^{j\varphi} \qquad [\varphi = \underline{/Z}]$$

Expanding,

$$Z = \frac{1}{G}\,(\cos^2\varphi + j\cos\varphi\sin\varphi) = \frac{1}{2G}\,(1 + \cos 2\varphi + j\sin 2\varphi)$$
$$= \frac{1}{2G}\,(1 + e^{j2\varphi}) \tag{163}$$

It is convenient at this point to think of Z as a vector; we find then that it can be resolved into two components, $\frac{1}{2}G$ and $e^{j2\varphi}/2G$. If G is kept constant, the first component is fixed, the second rotates around the point $(\frac{1}{2}G, 0)$, *i.e.*, around the *tip* of the first component vector. Thus it is seen that the tip of the resultant vector describes a circle around the point; or, in more precise language, the constant G line on the Z plane is a circle through the origin, of radius $\frac{1}{2}G$, with center on the R axis.

This circle cuts the R axis at a distance $1/G$ from the origin, hence it is very easy to construct. For instance, the circle for $G = 1$ millimho passes through the origin, has its center on the R axis, and cuts this axis at the point $R = 1{,}000$ ohms.

Suppose the circle is drawn for a value G and we wish to draw a new circle for the value $G + G_p$. We read the value of the intercept for the G circle, take the reciprocal, add G_p to this, and take the reciprocal again. The last value is the intercept for the new circle.

The constant B lines are similarly obtained. We may write

$$Z = -\frac{\sin\varphi}{B}e^{j\varphi} = \frac{j}{2B}(-1 + e^{j2\varphi}) \tag{164}$$

proving that, if B is kept constant, the Z point must move on a circle passing through the origin, of radius $\frac{1}{2}B$; the center will be on the *negative* X axis for *positive* B (when Z is in the fourth quadrant), and vice versa. Accordingly, we need only draw a circle through the origin and the Z point, with center on the X axis, to obtain the susceptance B for the point, which is the *negative reciprocal* of the circle's intercept.

If, in the foregoing, we replace each quantity by its dual (Y for Z, G for R, B for X, etc.), we obtain identical rules for the inverse construction (Fig. 28*d*).

The maps of Y on the Z plane and of Z on the Y plane are one and the same. The transformation of one line into another resulting when, for each point of the line, we take the *inverse* point (point whose complex number is the reciprocal of the origin value) is called *inversion*. Either the Z map or the Y map may be regarded as an inversion operated on a system of orthogonal lines parallel to the axes (Fig. 29); both must therefore have the same form. We shall make further use of the inversion process in Sec. 8.7.

8.5. Impedance Transformation by Reactive Branch Addition. Coupling Networks. We are now in a position to solve easily, at least on paper, the coupling problem or any other single-frequency impedance transformation problem.

Let us first assume that we have available two branches, or boxes, with two exposed terminals for each; each box has zero resistance and adjustable reactance (or identically, zero con-

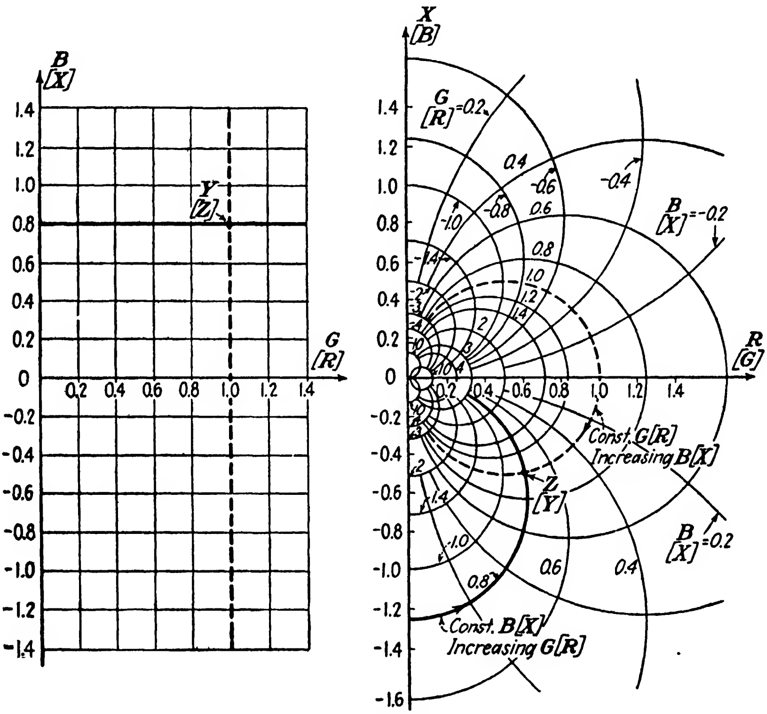


FIG. 29.—*Map of the inverse function, illustrating the inversion of orthogonal straight lines into orthogonal families of circles.*

ductance and adjustable susceptance) for the assigned frequency. In other words, the solution will call for a perfect coil of inductance L and a perfect condenser of capacitance C, where L and C may have any value. In practice, such requirements cannot be met by lumped elements at very high frequencies (Sec. 8.6).

Using the Z plane (the Y plane would do just as well), we locate the points Z_r and Z_g (for the load and source impedances, which are given); then we locate Z_t, symmetrical to Z_g about the R axis (because the two numbers Z_g and Z_t are conjugate as shown in Sec. 8.3).

We must transform Z_r to the value Z_t; hence, the Z point must be brought from Z_r to Z_t in some way. Since we are adding only reactive elements, only lines of constant R and constant G may be followed. There are four such lines (two circles and two straight lines) through the two end points of the transformation. Three possible cases are listed in Fig. 30, depending on the relative values of resistance and conductance. If no more than two branches are used, two distinct solutions are possible in cases 1 and 2 and four solutions in case 3—of which two use a coil and a condenser and two use condensers only (or coils only). A numerical example is given, as well as a key for recognizing readily the branch addition corresponding to each type of path followed (Fig. 30).

Coupling networks of this type are used at audio and radio frequencies, whenever power has to be transmitted at a fixed frequency (see also Secs. 8.8 and 9.7). The design is influenced by the fact that coils have appreciable dissipation. For this reason, other methods of coupling involving mutual inductances are often used. The method we have used can be extended to cover mutual inductances; for the time being, however, we shall extend it in another direction, to help us in solving the coupling problem at high frequencies by means of the transmission line.

8.6. Requisites of Coupling Networks at High Frequency. Use of the Transmission Line. As the frequency is increased, the true value of reactance of a coil deviates more and more from the theoretical value based on the inductance alone, because of the mounting effect of distributed capacitances. This not only makes it difficult to design a coil for a specified reactance, but often results in very high losses. We find, therefore, that the device of linking the same current many times with the same magnetic path, used so successfully at low frequencies, must inevitably be discarded; if we do not want to incur high displacement currents (Sec. 12.2) there must be some relation between the voltage and the distance separating two points.

We are led therefore to the use of transmission lines, singly or in combination. This use has the added advantage that the performance of lines can be predicted, because of their simple geometry.

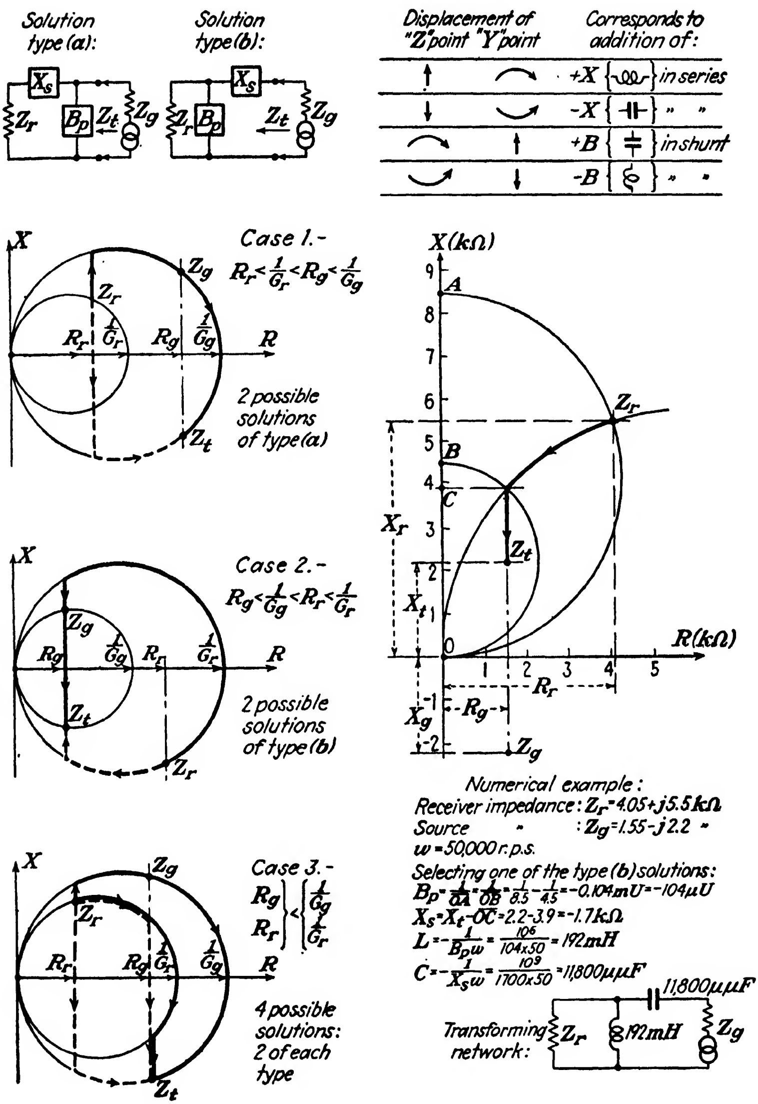


FIG. 30.—*Coupling networks consisting of two reactive elements.*

The impedance transforming action of the line was discussed in Sec. 7.3. We shall now represent this transforming action by the displacement of the Z or Y point due to the insertion of a line, very much as we did when branches were added in series or parallel.

We know that, with regard to a given network or line, an impedance may be represented, in addition to the numbers Z and Y, by any of the numbers z, y, k, and ρ (Sec. 5.4). Of these we have found ρ, the reflection constant, particularly useful because, to express the change from load to input impedance of a network, we can simply say that the transfer constant θ of the network has been added to the reflection constant of the load (Sec. 5.7). On the ρ plane such a change is represented by a straight-line movement (Fig. 20*a*).

On the Z and Y planes, the movement is not yet known to us, but we can determine it, just as we determined the movement of the Z point due to a shunt addition, by noting the following analogies:

1. When a *series branch* is added, we add the *branch impedance* to the *load impedance*.

2. When a *shunt branch* is added, we add the *branch admittance* to the *load admittance*.

3. When a *four-terminal network* (reversible) is inserted, we add the *transfer constant* of the network to the *reflection constant* of the load.

We may regard an addition to the number ρ as the result of two separate additions to its components, σ and τ. Graphically, these correspond to movements along lines of constant τ and constant σ, respectively. It now appears that a map of such lines on the Z and Y planes is all we need in order to represent the movement of the Z or Y point due to the insertion of a network in front of the load, or, in particular, to the insertion of a line.

The construction of this map will be discussed in the following section; we shall find that we can readily construct the required impedance paths, as we did for the constant G and constant B lines, without resort to a predrawn map. This practical consideration makes the Z or Y planes useful in impedance transformation problems, while the ρ plane (Sec. 5.7) is more suitable for long-line problems. Still another method of representation (the k plane, Sec. 10.5) is particularly effective when adjustable elements are involved.

★8.7. Mapping of the ρ Function on the Z and Y Planes. We proceed here much as we did for the map of Y on the Z plane, except that the steps are more tedious. We have in general (Sec. 5.7)

$$Z = Z_0 \coth \rho$$

We need only consider, however, the insertion of networks for which Z_0 is a pure resistance—lossless line sections, in fact. Hence, we may write

$$Z = R_0 \coth \rho = R_0 \frac{\cosh \sigma \cosh j\tau + \sinh \sigma \sinh j\tau}{\sinh \sigma \cosh j\tau + \cosh \sigma \sinh j\tau}$$

Clearing the imaginary arguments, rationalizing the denominator, and simplifying, with the help of (114), we obtain

$$Z = R_0 \frac{\sinh \sigma \cosh \sigma - j \sin \tau \cos \tau}{\sinh^2 \sigma + \sin^2 \tau}$$

which we may write still more simply

$$Z = R_0 \frac{\sinh 2\sigma - j \sin 2\tau}{\cosh 2\sigma - \cos 2\tau} \tag{165}$$

whence the equations

$$\begin{cases} \cosh 2\sigma - \cos 2\tau = \dfrac{R_0}{R} \sinh 2\sigma \\ \cosh 2\sigma - \cos 2\tau = -\dfrac{R_0}{X} \sin 2\tau \end{cases} \tag{166}$$

Solving for τ

$$\begin{cases} \sin 2\tau = -\dfrac{X}{R} \sinh 2\sigma \\ \cos 2\tau = \cosh 2\sigma - \dfrac{R_0}{R} \sinh 2\sigma \end{cases} \tag{167}$$

Solving for σ

$$\begin{cases} \sinh 2\sigma = -\dfrac{R}{X} \sin 2\tau \\ \cosh 2\sigma = \cos 2\tau - \dfrac{R_0}{X} \sin 2\tau \end{cases} \tag{168}$$

Squaring and adding Eqs. (167), τ is eliminated and the equation of the constant σ paths is obtained in the form

$$R^2 + X^2 - 2RR_0 \coth 2\sigma + R_0^2 = 0$$

This, by the expedient of adding and subtracting $(R_0 \coth 2\sigma)^2$, can be written

$$(R - R_0 \coth 2\sigma)^2 + X^2 = \left(\frac{R_0}{\sinh 2\sigma}\right)^2 \tag{169}$$

Similarly, squaring (168), eliminating σ, then adding and subtracting $(R_0 \cot 2\tau)^2$

$$(X + R_0 \cot 2\tau)^2 + R^2 = \left(\frac{R_0}{\sin 2\tau}\right)^2 \tag{170}$$

Equations (169) and (170) are equations of circles. We conclude from inspection of these equations that

1. The *constant σ lines* on the Z plane are circles with center on the R axis at distance $R_0 \coth 2\sigma$ from the origin and radius $R/\sinh 2\sigma$.
2. The *constant τ lines* on the Z plane are circles with center on the X axis at distance $-R_0 \cot 2\tau$ from the origin and radius $R_0/\sin 2\tau$.

A more thorough investigation reveals how these circles may be constructed without having to compute the hyperbolic functions. Consider, in fact, a circle with center in the origin and radius R_0, henceforth referred to as the *characteristic circle* for the reactive line in question. We can show the following to be true:

3. All circles with center on the R axis which cross the characteristic circle at right angles are circles of constant σ;
4. All circles passing through the intersections of the characteristic circle with the R axis are circles of constant τ; and the arc of such circles determined by either intersection and the positive X axis subtends at the center an angle 2τ.

Positions 3 and 4 can be shown to be consequences of positions 1 and 2, which were derived analytically.

Consider (Fig. 31) the circle with center C crossing the characteristic circle at right angles. Let a number σ be defined by the following:

$$\overline{OC} = R_0 \coth 2\sigma$$

Then the radius of the circle is, by inspection,

$$\overline{CO} = R_0 \sqrt{\coth^2 2\sigma - 1} = \frac{R_0}{\sinh 2\sigma}$$

and consequently the circle answers the description given for a constant σ circle, which proves position 3.

The proof of 4, similar to the above, is left to the reader. As a corollary of 3 and 4, it may be shown that

5. The circles of constant σ and of constant τ cross at right angles.

This is generally true of two families of lines that have been obtained from two orthogonal families of straight lines by a conformal transformation (Sec. 5.7).

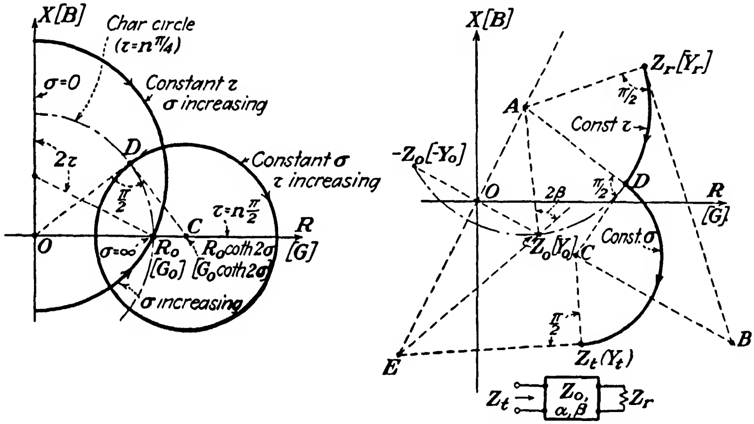


FIG. 31.—*Left, loci of constant σ and constant τ on the Z plane. (For Y plane, use bracketed notation.) Valid when $Z_0 = R_0$. Right, construction for the input impedance of a generic network of constants Z_0 and $\theta = \alpha + j\beta$.*
NOTE: $OB/OZ_0 = \coth 2\sigma$, $OC/OZ_0 = \coth (2\sigma + 2\alpha)$.

We have discussed the map of the ρ function on the Z plane. To obtain the map of the function on the Y plane, inversion may be used (Sec. 8.4). Rules for this inversion are implicit in Fig. 31. Instead of the circle of radius R_0, the circle with radius G_0 must be taken as the characteristic circle; the positive X axis corresponds to the negative B axis; otherwise, the construction is identical.

As a result of the foregoing analysis, we have at our disposal, for use in the next chapter, a method whereby sections of lossless line may be handled just as conveniently as pure inductors or capacitors, as far as their impedance transforming action is concerned.

It should be noted that lines similar to the constant σ and constant τ circles on the Z plane make frequent appearances in problems of applied physics. Such lines represent the flow lines and the lines of equal pressure in the plane movement of a fluid between a source and a sink; or lines of force and equipotentials in the electric (or magnetic) field between parallel conductors (Sec. 11.10).

8.8. Illustrative examples.

Efficiency of coupling networks. Whenever coils are used in a transmission network, some power is lost because of dissipation in the coil. This is also true, but in much smaller degree, of condensers.

Consider, in particular, a network of the type discussed in Sec. 8.5, in the case when both source and load impedances are pure resistances. In this case only two solutions are possible if no more than two elements are to be used. They are shown in (*a*) and (*b*), Fig. 32. The coil is considered to have some resistance, which is represented as a series resistance in Fig. 32*a* and as a shunt resistance in Fig. 32*b*. Both representations are equally valid in single-frequency problems. With reference to the schematics and complex plane diagrams of Fig. 32, we have the following:

SOLUTION *a* (FIG. 32)

Total power absorbed by coil and load, equal to power input

$$P_i = I^2(R_l + R_r)$$

Power dissipated in coil

$$P_d = I^2R_l$$

Power absorbed in load (power output)

$$P_0 = I^2R_r$$

Ratio power input/power output

$$\frac{P_i}{P_0} = 1 + \frac{R_l}{R_r}$$

SOLUTION *b* (FIG. 32)

Total power absorbed by coil and load, equal to power input

$$P_i = V^2G_t$$

Power dissipated in coil

$$P_d = V^2G_l$$

Power absorbed in load (power output, obtained by subtraction)

$$P_0 = V^2(G_t - G_l)$$

Ratio power input/power output

$$\frac{P_i}{P_0} = \frac{1}{1 - G_l/G_t}$$

We shall follow the usual practice of expressing power ratios in the form of losses; it will be assumed that the coil absorbs only a small fraction of the total power. Hence, the following approximations:

Power loss due to coil dissipation (in nepers)

$$L_d = \ln \frac{P_i}{P_0} = \frac{R_l}{R_r} \text{ nepers} \quad (171)$$

Power loss due to coil dissipation (in nepers)

$$L_d = \ln \frac{P_i}{P_0} = \frac{G_l}{G_t} \text{ nepers} \quad (172)$$

The above can be put in more significant form by the use of dimensionless parameters (Sec. 3.5). Let, as usual,

$$d = \frac{1}{Q} = \frac{R_2}{X_2} = \frac{G_l}{B_l}$$

stand for the dissipation factor of the coil, and let

$$N = \frac{R_t}{R_r} = \frac{G_r}{G_t}$$

stand for the ratio of source and load resistances. We may now write (171) and (172) as follows:

$$L_d = d\frac{X_l}{R_r} \qquad\qquad L_d = d\frac{B_l}{G_t}$$

From inspection of the complex plane diagrams, we observe that X_l is the geometric mean of R_r and $R_t - R_r$; likewise, B_l is the mean of G_t and $G_r - G_t$. Hence

$$L_d = d\sqrt{\frac{R_t - R_r}{R_r}} = d\sqrt{N-1}$$

$$L_d = \sqrt{\frac{G_r - G_t}{G_t}} = d\sqrt{N-1} \quad (173)$$

We come, therefore, to this conclusion: when the dissipation is small and the dissipation factor of the coils used may be considered to be independent of the coil impedance, the two solutions are equivalent from the standpoint of efficiency.

When the dissipation is not small, a true analysis is very complicated, because maximum power flow from the source no longer corresponds to maximum flow into the load. Therefore, equality of source conductance with transformed conductance is no longer called for, and the optimum value of N remains to be determined. It may be argued, however, that in such cases

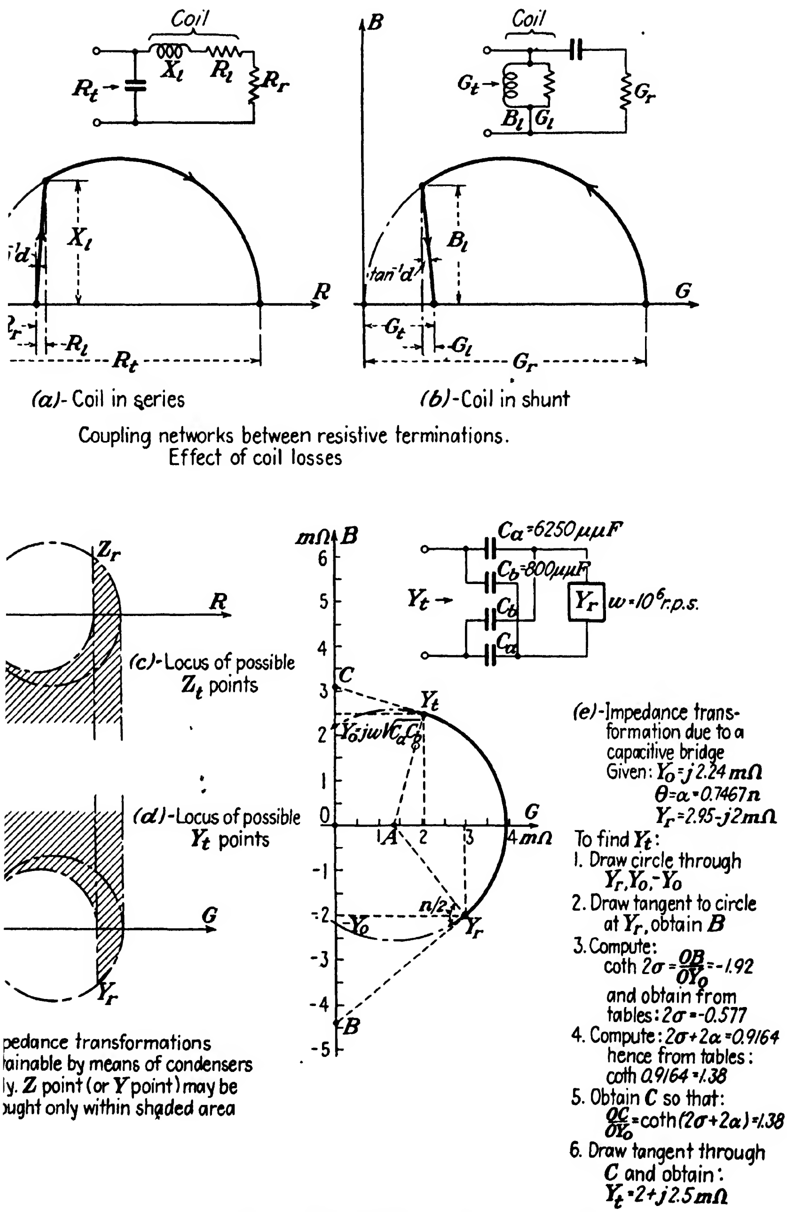


FIG. 32.—*Examples of impedance transformation.*

solution b is less efficient by inspection of the exact expressions for the power ratio; for $G_l \to G_t$, the power ratio in b tends to infinity; for $R_l \to R_r$, the power ratio in a tends to 2 as a limit.

Another important consideration in favor of solution a is this: The value of coil inductance called for in this solution is always less than for the other. From the expressions of X_l and B_l, we have the following:

$$\frac{L_a}{L_b} = 1 - \frac{1}{N}$$

where L_a is the coil inductance for solution a, L_b the inductance for solution b, and N the impedance ratio, taken so as to be always greater than unity.

At radio frequencies, the *effective inductance* of a coil must take into account the distributed capacity; the coil is actually a parallel LC circuit which must be brought near resonance in order to obtain high values of reactance. This can only be accomplished at the expense of the effective Q, since high circulating currents produce dissipation. For this reason solution b, requiring a higher reactance, is unsatisfactory. On the behavior of coils at high frequency, see group 7 in the Bibliography.

Example. An L network of type a, Fig. 32, couples a 10,000-ohm tube to a 70-ohm coaxial cable. The Q of the coil is 80 at the operating frequency. The loss due to the coupling is

$$L_d = \frac{1}{80}\sqrt{\frac{10{,}000}{70} - 1} = 0.149 \text{ neper} = 1.29 \text{ db}$$

Suggested Exercise. Compute the values for the condenser and coil of the above example at 100 kc., both graphically and analytically.

Impedance transformation by condenser networks. Since condensers have a much better Q than coils, it is advantageous, whenever possible, to use coupling networks made up of condensers only for the purpose of impedance transformation at a single frequency. The possible values of transformed impedance are limited, however. Geometrically, for every load impedance point on the complex plane, there is a region of the plane containing all possible transformed impedance points if condensers only are used in the transformation.

The region (shaded area of Fig. 32c) is bounded by the following lines:

The constant G circle through Z_r
The constant G circle through the point $R = R_r$, $X = 0$
The lines $R = \frac{1}{Gr}$, $R = R_r$, $R = 0$

On the admittance plane, the corresponding region is similarly defined (Fig. 32*d*).

It can be readily shown that no number of condensers added in shunt or series can move the Z point anywhere outside the region of Fig. 32*c*. In fact, a condenser added in series moves the Z point down (*i.e.*, in the negative X direction); a condenser in parallel moves it in a circle, clockwise toward the origin. No combination of such displacements can bring the point outside the region, as is immediately evident.

It can also be shown that any symmetrical network composed entirely of condensers will bring the impedance point somewhere within the region. The transforming action of a capacitive bridge network will be taken up in the example below.

Suggested Exercise. Analyze and define by regions of the complex plane the transformations of impedance possible when the following types of circuit elements are used:

1. Condensers and resistors
2. Coils only (neglecting dissipation)
3. Coils and resistors

Impedance transformation by reversible networks. Capacitive bridge. A construction for the input impedance of a generic reversible network (the transformed impedance), given the load impedance and the network constants, is indicated in Fig. 31*b*. It uses the constant σ and constant τ lines of Sec. 8.7. The advantage of this graphical method over those of Chap. V is that no predrawn chart or map is necessary.

As an example, consider the impedance transformation due to a capacitive bridge network (or *lattice network*, Fig. 32*e*). We have for this network

Short-circuit admittance:

$$Y_{sc} = \frac{Y_a + Y_b}{2} = j\omega \frac{C_a + C_b}{2}$$

Open-circuit admittance:

$$Y_{oc} = \frac{2}{Z_a + Z_b} = \frac{2j\omega C_a C_b}{C_a + C_b}$$

Characteristic admittance:

$$Y_0 = \sqrt{Y_{sc}Y_{oc}} = j\omega\sqrt{C_aC_b}$$

Transfer constant:

$$\theta = \frac{1}{2}\ln\frac{\sqrt{Y_{sc}} + \sqrt{Y_{oc}}}{\sqrt{Y_{sc}} - \sqrt{Y_{oc}}} = \frac{1}{2}\ln\frac{1 + \dfrac{2\sqrt{C_aC_b}}{C_a + C_b}}{1 - \dfrac{2\sqrt{C_aC_b}}{C_a + C_b}} = \ln\left|\frac{\sqrt{C_a/C_b} + 1}{\sqrt{C_a/C_b} - 1}\right|$$

(*Note.* $\alpha = \theta$, $\beta = 0$.)

Numerically, take for example

$$\omega = 10^6 \text{ r.p.s.} \qquad C_a = 6{,}250\ \mu\mu\text{f} \qquad C_b = 800\ \mu\mu\text{f}$$

Hence

$$\sqrt{C_aC_b} = 2{,}240\ \mu\mu\text{f} \qquad \sqrt{C_a/C_b} = 2.8$$

and

$$Y_0 = j2{,}240\ \mu\text{mhos} \qquad \alpha = \ln\ (3.8/1.8) = 0.7467 \text{ neper.}$$

The construction (Fig. 32*e*) is carried out on the basis of the above data. As a result, we find for a load admittance:

$$Y_r = 2{,}950 - j2{,}000\ \mu\text{mhos}$$

the transformed admittance

$$Y_t = 2{,}000 + j2{,}500\ \mu\text{mhos}$$

It will be noted that the Y_t point is within the region of Fig. 32*d*.

CHAPTER IX

USE OF LINES AS MATCHING DEVICES

9.1. The Line as a Solution of the Coupling Problem. As a simple application of the geometry developed in the preceding chapter, we may imagine a section of lossless line (of length

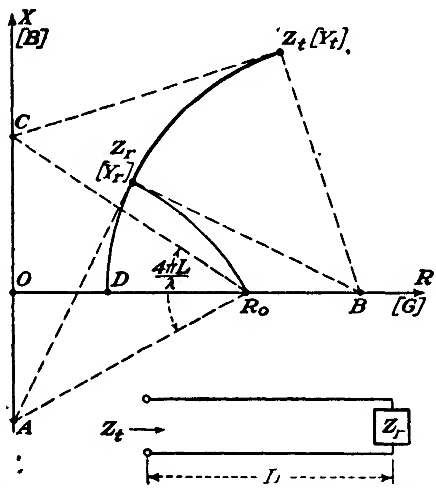


FIG. 33.—*Impedance transformation by a lossless line.*

To find the transformed impedance Z_t,

1. Locate Z_r and R_o on Z plane.
2. Draw circle through R_o and Z_r, with center A on the X axis.
3. Drop normal to AZ_r at Z_r: obtain point B.
4. Measure angle $4\pi L/\lambda$ *cw* from AR_o at R_o: obtain C.
5. Draw circle through Z_r with center in B.
6. Draw tangent to above circle through C: obtain Z_t.

The same construction may be carried out on the Y plane (brackets).
The ratio OR_0/OD is the s.w.r. on the line.

L, characteristic impedance R_0, attenuation $\alpha \ll \beta$, connected to a load Z_r, and proceed to find graphically the impedance at the free end. This problem can be solved in a number of ways; the method of Fig. 33, however, has the advantage that neither computation nor a predrawn map is required. It follows directly from the discussion of Sec. 8.7 and points the way to a solution of the inverse problem, *i.e.*, that of transforming a given impedance to a prescribed value by means of a line section of suitable characteristics—the coupling problem, in fact (Sec. 8.1).

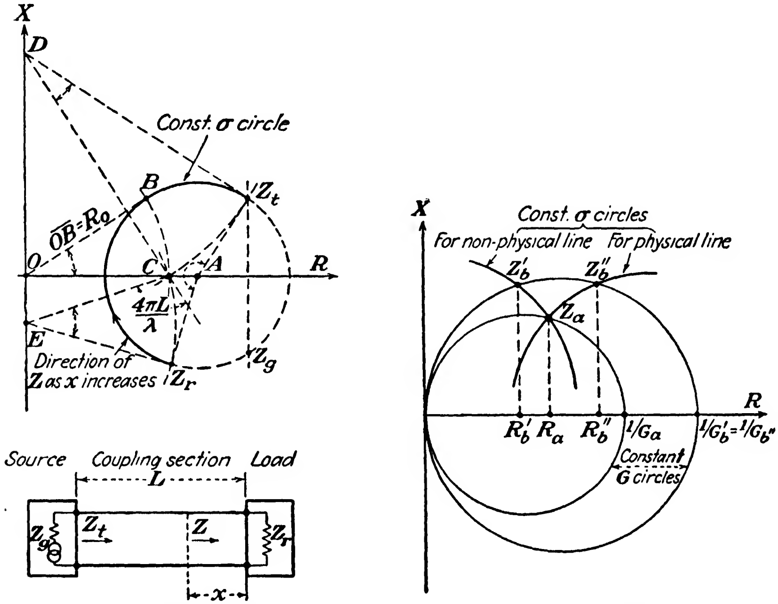


FIG. 34.—*Solution of the coupling problem by means of a section of lossless line.*

The procedure is that of Fig. 33, in reverse order. Locate points Z_r and Z_t, for the load and transformed impedances. (Z_t must be the conjugate of Z_g for max. power transfer.)

Draw a circle through Z_r, Z_t, with center on the R axis. This is the constant σ circle described by the Z point as x, distance from load, increases.

Draw tangents to this circle at points Z_r, Z_t, and from the origin. This locates points B, C, D, E. R_0 and L/λ are measured on the diagram as shown. These data determine the line required for maximum power transfer.

NOTE: The problem can be solved by the use of a single section only if the terminations are so related that whichever has *higher conductance* also has *lower resistance.*

Let Z_a represent the load impedance. No physical line can transform this impedance so that its conductance and resistance both decrease, because, as shown on left, in that case the constant σ circle would surround the origin. An imaginary value of R_0 would result.

Figure 34 is an illustration of this solution of the coupling problem. Given a source impedance Z_g and a load impedance Z_r, the characteristics R_0 and L of the coupling section which, inserted between them, will ensure maximum power transfer, are uniquely determined by retracing the steps of Fig. 33 in the opposite sequence. The lateral dimensions a and b are readily obtained from R_0 by means of the formulas of Sec. 7.1 and corresponding charts (see Sec. 7.2).

This solution of the coupling problem is not always possible. As shown in Fig. 34, the two points Z_r and Z_t can always be joined by a constant σ circle, but some of these circles do not correspond to realizable or *physical* lines. Simple considerations of geometry point to the following rule: Two terminations may be coupled by a single section of line only when the termination having *higher resistance* also has *lower conductance*, and vice versa.

Example. Consider the following values:

$$Z_g = 30{,}000 - j50{,}000 \qquad Z_r = 60{,}000 + j20{,}000$$

We have in this instance

$$R_r > R_g$$

The values of conductance are

$$G_g = \frac{3 \times 10^{-4}}{9 + 25} = 8.82\ \mu\text{mhos}$$

$$G_r = \frac{6 \times 10^{-4}}{36 + 4} = 15\ \mu\text{mhos}$$

Hence

$$G_r > G_g$$

The terminations cannot be coupled by a single line section, as the load *resistance* and *conductance* are both lower than the values for the source.

While theoretically possible, this type of coupling may not be practicable owing to the difficulties of realizing very low or high values of characteristic impedance. In general, it has the serious disadvantage of rigidity; once the cable is built, the characteristic impedance cannot be adjusted; length adjustments are almost as difficult if the source and load terminals have fixed positions. Aside from tuner problems, which require m.p.t. over a range of frequencies (Sec. 10.4), it is always desirable to allow some latitude in the design of high-frequency coupling devices, since the terminating impedances are seldom accurately known. For this reason, tuners are used even when the frequency is never varied, as they can be adjusted to work between terminations of any value. If the terminal impedances do not change and can be measured with accuracy, the fixed coupling consisting of a single line section, when practicable, is the most efficient solution.

9.2. Matching Sections: Maximum Power Transfer in Dissipative Systems. One of the drawbacks of the coupling section, as has been pointed out, is that the length of the section cannot be established independently of the terminal impedances. This is no longer a disadvantage when the coupling section is used to convey power in or out of a long transmission line, or between two such lines. We have previously (Sec. 8.1) classified this as the matching problem, as distinct from the coupling problem, of which the former is a particular case. The distinction hinges on the fact that systems comprising long transmission lines function most efficiently when they are matched at both ends, as will be shown, and the coupling sections are designed with this end in view.

As an example, consider a physically long line of characteristic impedance R_0 connecting a load Z_r to a source of impedance Z_g. The terminations are matched to the line by means of short matching sections of characteristic impedances R_1 and R_2. These values, as well as the lengths L_1 and L_2, are determined graphically (Fig. 35) just as in the previous problem. The first matching section operates as a coupling between Z_r and R_0; the second couples R_0 to Z_t (or transforms R_0 into Z_t, the conjugate of Z_g). A numerical example accompanies the diagram of Fig. 35. It is apparent from a study of the diagram that not all lines can be matched to a given set of terminations in this manner. For given values of Z_g and Z_r, there are values of R_0, the characteristic impedance of the connecting line, which cannot be used. They are the values comprised between the *resistance* and the reciprocal of the *conductance* of either termination. These prohibited intervals are shown more clearly on the plot of impedance vs. length of Fig. 35.

A system such as the one we are considering is perhaps the best illustration of an important general point; it will pay us, therefore, to analyze it in some detail.

We are immediately conscious of the fact that the three line sections of Fig. 35 operate under entirely different conditions. This is brought out by a plot of instantaneous voltage along the system. The center section is matched, and the impedance looking both ways at all its points is identically R_0; therefore, its voltage distribution is patterned after Fig. 17.

The end sections, on the contrary, have reflecting terminations, and the impedance in either direction varies throughout their

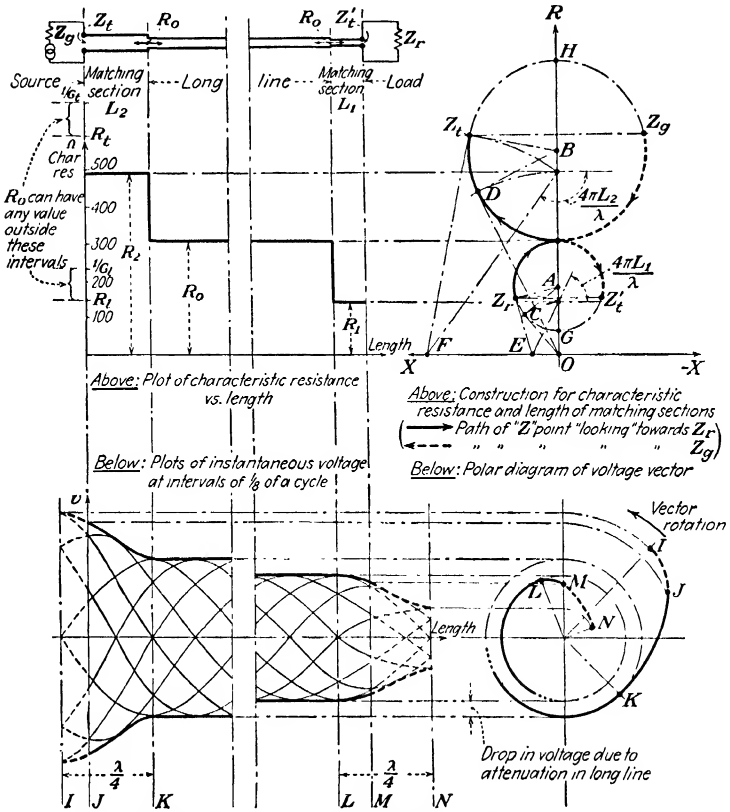


FIG. 35.—*Matching a long line to its terminations by means of short impedance transforming sections.* In the example shown,

$$Z_r = 190\Omega\underline{/+37°};\ Z_g = 640\Omega\underline{/-21°40'};\ R_0 = 305\Omega.$$

The construction of Fig. 34 is carried out twice: first between Z_r and R_0, then between R_0 and Z_t, conjugate of Z_g.

Results: $R_1 = 490\Omega$; $R_2 = 140\Omega$; $L_1/\lambda = 0.177$; $L_2/\lambda = 0.35$.

Relation between polar diagram of voltage vector and impedance diagram. The ratio of max. to min. voltage (r.m.s. or peak), or standing wave ratio is the square root of the max. to min. impedance ratio.

length. The voltage distribution is therefore that of Fig. 26. There are definite relationships between the polar diagram of voltage, on which the distribution plot is based, and the Z-plane diagram.

Stressing what could be termed the *physicist's* point of view, we might say that a wave is propagated continuously through the system from source to load. This wave carries power and is attenuated uniformly within each section. In addition, reflected waves are set up at the two ends by the reflecting terminations. These waves do not enter the center section; their energy remains stored in two oscillating systems, namely, the source with the adjacent matching section and the load with its matching section. Analogous mechanical systems are easily imagined and may provide useful illustrations.

In spite of these obvious differences, the entire system is uniform in one respect: if we cut it at any point, the impedance towards the load and the impedance towards the source are always *conjugates*. This is true, by hypothesis, at the source terminals, but we may verify that the condition holds everywhere. Let us suppose that the direction of energy flow is inverted and let Z_g now represent the load impedance. As we move along the system away from Z_g, the impedance point follows the constant σ circle in the direction of increasing τ, *i.e.*, clockwise. If we carry this out for the whole system, we find that, point for point, the new path is the mirror image of the original one with respect to the R axis (dotted line, Fig. 35). The two paths meet on the R axis at the point which represents R_0, the characteristic impedance of the center section, which is seen in both directions at all points of this section.

Since two points symmetrically placed about the R axis represent conjugate impedances, we reach the conclusion that, for this case at least, the condition of m.p.t. is met through the system if it is met at one point.

The above proposition can be proved for a wide class of systems by a *reductio ad absurdum*. Suppose a dissipationless network is used to transform a load impedance Z_r into a value Z_t such as will ensure maximum power flow from a source of impedance Z_g. Now, imagine the network to be cut at an internal junction J. Let Z_t' stand for the impedance at this junction, looking toward the load, and Z_g' for the impedance toward the source.

Assume that Z_t' and Z_g' are *not* conjugate; by suitable alterations of the interposed network, we can in this event cause them to vary so as to be conjugate and thereby increase the flow of power through J.

It follows that the power flow through J is not the highest possible; now, the same thing can be said of the power from the source, as no power is lost in the network by hypothesis.

This statement contradicts our premise of a conjugate match at the source terminals. The assumption that Z_t' and Z_g' are not conjugate is therefore a fallacy, and we reach the conclusion that, if there is a conjugate match at one junction of a transmission system, this is also true of any other junction, *provided the section of the system included between the two junctions does not dissipate power.*

If a transmission system extends over a considerable distance, its over-all dissipation must inevitably be large. We cannot, therefore, ensure a *conjugate match* at all points of such a system—at least, not in general.

We can, however, realize an *equal match* over the entire system, if every part is terminated by its image impedance. Specifically, in the case of a line, we know that it is possible for the impedance in both direction to be Z_0 at all points.

An *equal match* is also a *conjugate match* when the matching value is real. A matched transmission line at high frequency has a real value of matching impedance; it is therefore an exception to the general rule, in that it ensures a conjugate match over its entire length, although the losses in it may be very high. Distortionless lines have the same property at all frequencies, and all open wire lines come very close to it.

It is now apparent that a long line must be matched for efficient operation. The same conclusion may be reached in other ways. Using the wave concept, we may argue that the energy stored in a reflecting line does not contribute to transmission while it does contribute to the losses. Upon evaluation of the transmission loss in the reflecting line as compared to the attenuation, we would obtain the same result.

Whenever the terminations themselves do not match the line, as in the system of Fig. 35, reflection should be confined within short matching sections, such as have been considered.

9.3. Quarter-wave Transformer. As a particular case of the matching problem, consider the connection of two long lines a and b, of unequal characteristic resistances R_a and R_b, by the intermediary of a short matching section.

It is helpful, although by no means necessary, to think of transmission in a definite direction. We will assume the energy to go from line a into line b.

Evidently, the source and load impedances of the section are not uniquely given, unless the terminations of lines a and b

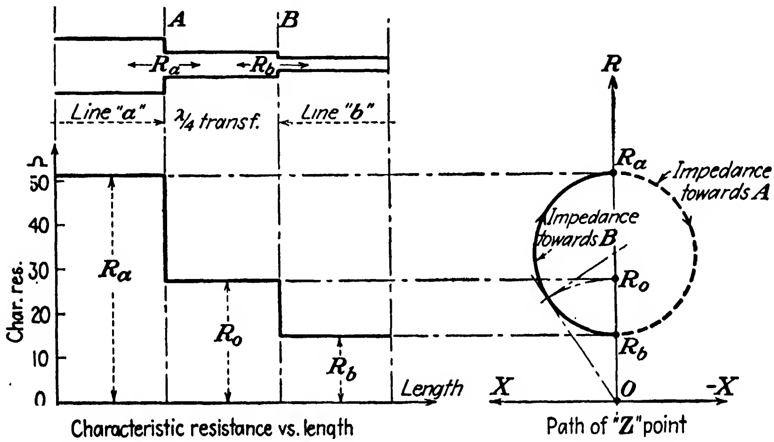


FIG. 36.—*Quarter-wave transformer. In the example shown,* $R_a = 51\,\Omega$; $R_b = 15\,\Omega$. *Length of transformer* $L = \lambda/4$. *Characteristic resistance.*

$$R_0 = \sqrt{R_a R_b} = 27.5\,\Omega.$$

Voltage ratio (s.w.r.) $r_v = \sqrt{\dfrac{R_a}{R_b}} = 1.845.$

are assigned. However, it has been established in the preceding section that a system of this sort can meet the m.p.t. condition only if the impedance in both directions is resistive over all but a small part of the length. Hence, both lines a and b must be in matched operation.

To ensure such operation, the source must be matched to line a and the load to line b by suitable matching networks, which may be sections of line as described in Sec. 9.2.

In addition, the two lines must be mutually matched. This requirement assigns the values of load and source impedance for the connecting section. They are R_a at junction A (source) and R_b at junction B (load), as shown in Fig. 36.

The connecting section should therefore be designed to provide maximum power flow between resistive terminations. This can only be obtained if the impedance at both ends is the same in both directions, or if the section is image connected (Sec. 2.10).

Geometrically, this problem differs from the general coupling problem (Fig. 34) in that the points corresponding to source and transformed impedance, which in the general case are symmetrically placed about the real axis, now fall on the axis and coincide. The load impedance point also falls on the R axis.

The matching section must bring the load impedance point from R_b back to the R axis at R_a. Since the path followed (assuming use of a lossless line) must be a constant σ circle with center on the R axis, we conclude that the path is a semicircle.[1] The source impedance point describes the other half of the circle (Fig. 36).

By the general construction of Fig. 33, we find immediately that the line angle $2\pi L/\lambda$ for the section is $\pi/2$, hence the length must be $\lambda/4$, which justifies the term *quarter-wave transformer*.

The determination of R_0, characteristic resistance of the matching section, also conforms to the general case. We draw the tangent to the semicircle (Fig. 36) from the origin and the characteristic circle through the point of tangency. The radius of the latter is R_0.

It is shown in plane geometry that points O, R_a, R_0, and R_b, as obtained on the R axis in the above construction, are so related that

$$\overline{OR_0}^2 = \overline{OR_a} \cdot \overline{OR_b}$$

Bearing in mind the significance of the segments, we may write the above as follows:

$$R_0^2 = R_aR_b \therefore \frac{R_a}{R_0} = \frac{R_0}{R_b} \tag{174}$$

R_a/R_0 and R_b/R_0 are the load and input resistance numbers of the quarter-wave line; hence, the general statement: *A quarter-wave line transforms a resistance into another resistance, whose number is the reciprocal of the first.* More particularly, *the*

[1] If a line is terminated by a resistance, its input impedance point describes a semicircle as the line length varies from 0 to $\lambda/4$; it completes a full circle as the length is increased to $\lambda/2$; and it goes twice around the same circle for each addition of a wavelength.

For nonresistive loads, the arc of circle described for each $\lambda/4$ increase is not a semicircle, but a $\lambda/2$ increase still corresponds to a full circle.

matching section between two lines (quarter-wave transformer) should have length equal to $\lambda/4$ and characteristic resistance the geometric mean of those for the two lines to be connected.

The action of the quarter-wave transformer is so simple that considerations of a very general nature would have enabled us to reach the same conclusions.

Consider a wave propagated over line *a*. This will be partially reflected at point *A* (Fig. 36). Also, there will be a *refracted component,* to use the terminology of optics, which will keep right on. In turn, this will be partially reflected at *B*.

The two partially reflected waves will coexist in line *a*. They will be out of phase by a number of cycles corresponding to the time of propagation from *A* to *B* and back. If the length *AB* is $\lambda/4$, this time is evidently a half period, so that the phase difference is half a cycle. If, furthermore, the amplitude of the two waves is the same, there will be no reflected wave in line *A*, which will be in matched operation.

Of the two conditions for absence of reflection, one has to do with phase; this is satisfied by making the length of the connecting section equal to $\lambda/4$, as we have seen. The other has to do with amplitude and must be met by adjusting the real part of the reflection factors at the two junctions (Sec. 5.4). This is done by properly selecting the impedance of the matching section. We could reobtain by this method the condition $R_0^2 = R_a R_b$.

Interference phenomena due to thin transparent films are similarly explained. A close analogy to the quarter-wave transformer is supplied by the thin transparent coatings which minimize reflection at the surface of optical lenses. These films have thickness equal to one-quarter of the wavelength of the dominant color; their index of refraction is the geometric mean of the two absolute indices for the lens and for air.

9.4. Selectivity, Tuning, and Resonance. Methods for coupling a generator to its load and matching a line to its terminations, or two lines to each other, have been considered. One type of solution, the same in all cases, has been taken up so far; namely, the use of a short section of line as an impedance transforming network. This solution is possible because two variables, the characteristic resistance and length of the section, are separately subject to choice, although in some cases the solution is not realizable.

It has been pointed out that solutions of this type lack flexibility, as maximum power transfer is obtained only for a single value of the frequency. For example, the quarter-wave transformer meets the m.p.t. condition only if the wavelength is four times the section length. This is true at one value of frequency, the *midband frequency* f_0. At this frequency the transmitted power, for a given source, has the highest possible value. At all the other frequencies the power flow is capable of increase by altering the load impedance. The quarter-wave transformer is said to be *selective* in favor of f_0.

Quantitative considerations regarding selectivity will be taken up later (Sec. 9.5), and the selectivity of the quarter-wave transformer in particular will be evaluated (Sec. 9.6). For the present, let us consider qualitatively the advantages and disadvantages of selectivity and its physical significance.

It is evidently necessary, in some cases, to single out a narrow frequency band, excluding all others. The best known example is the antenna coupling of a radio receiver. A high degree of selectivity is often desirable in such cases; moreover the coupling network may have to be adjustable so that the midband frequency may be selected at will within a range of the spectrum.

Selectivity must be avoided when the signal components are spread over a wide band or when adjustments are not contemplated. As a rule, when a narrow frequency band is transmitted at any one time, its selection should take place at one point, or at a few accessible points of the transmission system. The remainder of the system should be capable of transmitting any one of the selected frequencies with substantially uniform efficiency.

Let us see how the single-section couplings (in particular, the quarter-wave transformer) fit into the above picture. They are not selective enough to single out a narrow band and cannot be tuned to any desired frequency. Hence, we must rule them out whenever selectivity appears as a necessary feature. They may be used, with limitations, as nonselective couplings; only a quantitative analysis will reveal whether or not this use is permissible in any specific case. For the quarter-wave transformer, this analysis will be carried out in Sec. 9.6, by a method based on the variation of input impedance with frequency. To begin with, however, a broad discussion of selectivity based on the wave conception may be of value.

It was shown in Sec. 7.5 that the voltage (or current) on a line, when reflection is present, may be considered as the resultant of two waves traveling in opposite directions, of unequal amplitudes (151).

An alternative subdivision results if we write (151) as follows:

$$v(x, t) = \frac{\hat{V}_0}{e^\sigma + e^{-\sigma}}\left[(e^\sigma - e^{-\sigma}) \cos\left(\omega t + \frac{2\pi x}{\lambda}\right) + e^{-\sigma}\left\{\cos\left(\omega t + \frac{2\pi x}{\lambda}\right) + \cos\left(\omega t - \frac{2\pi x}{\lambda}\right)\right\}\right]$$
$$= \hat{V}_0 \tanh \sigma \cos\left(\omega t + \frac{2\pi x}{\lambda}\right) + \frac{2\hat{V}_0}{1 + e^{2\sigma}} \cos \omega t \cos \frac{2\pi x}{\lambda} \tag{175}$$

The voltage now appears as the resultant of two component distributions, a *standing* wave and a *traveling* wave of the usual type, having the same direction as the flow of energy. The former has zero amplitude at all times for $x = \lambda/4, 3\lambda/4, 5\lambda/4, \ldots$ (at the voltage *nodes*); hence, if it existed alone, no power could be transmitted down the line. (The transmitted power at a junction of zero voltage is zero.) Geometrically, the voltage ellipse would degenerate into a straight line if the standing wave alone were present; the corresponding current ellipse would be an orthogonal line; hence, the voltage and current would be in quadrature everywhere, which confirms the fact that the time average of transmitted power at any point is zero.

While it does not contribute directly to the transmission of power, the standing wave causes energy to be exchanged between the electric and magnetic fields which surround the line. At time $t = \pi/2\omega, 3\pi/2\omega, 5\pi/2\omega, \ldots$ the standing wave voltage vanishes everywhere and the standing wave energy is stored in the magnetic field, which has maximum strength at $x = \lambda/4, 3\lambda/4, 5\lambda/4, \ldots$ (at the voltage nodes). At $t = 0, \pi/\omega, 2\pi/\omega, \ldots$ the standing wave energy is entirely due to electric field, with maximum density at $x = 0, \lambda/2, \lambda, 3\lambda/2, \ldots$ Potential and kinetic energy are similarly exchanged in the vibration of elastic bodies.

In addition, the standing wave affects the transmission of power *indirectly*, as it alters the value of voltage and current at all points, and particularly at the driven end. In fact, the impedance transforming action of the line may be ascribed to the standing wave exclusively.

Bearing in mind the above considerations we can see what causes a coupling section to be selective. The degree of selectivity is closely related to the comparative values of standing wave energy and transmitted power. Frequency has no bearing on the transmitted power *except indirectly*, because of its effect on the standing wave. In particular, consider the case when the standing wave energy (energy of oscillation) is a maximum at the impressed frequency, and assume that the transmitted power is also a maximum for the same frequency. A frequency variation in either direction will cause the oscillation energy to drop sharply and with it the impedance transforming action upon which the transmitted power depends.

When the frequency has one of the values that make the oscillation energy a maximum, the system is said to be in *resonance*. We may then conclude that *selectivity* is accompanied by *resonance*. We are used to this association with regard to lumped *LC* circuits. These, however, do not transmit power, although they dissipate some. In coupling networks, we must think of resonance and transmission as superimposed. Selectivity is high when resonance plays a major part.

9.5. General Validity of the Parameter Q_0 as a Measure of Selectivity: Its Analytical and Geometrical Significance. The reader is probably acquainted with the definition of the Q *of a resonant circuit* as the inverse of the bandwidth for half power, or

$$Q = \frac{1}{\delta'}$$

where, in general,

$$\delta = \frac{2(f - f_0)}{f_0}$$

is the bandwidth and, in particular,

$$\delta' = \frac{2(f' - f_0)}{f_0}$$

f' being the value of frequency at which, under certain conditions, the power taken by the coil-condenser combination is one-half the maximum. (The conditions will be specified at the end of this section.)

Equally familiar is the definition of the Q *of a coil,* namely,

$$Q = \frac{\omega L}{R}$$

where R is the (series) resistance and L the inductance of the coil at the frequency $\omega/2\pi$.

It can be shown readily that if an LC series combination includes a dissipative coil and a nondissipative condenser and receives energy from a constant voltage (zero impedance) source, the Q of the circuit and the Q of the coil, as defined above, are approximately the same, or

$$\frac{1}{\delta'} = \frac{\omega L}{R}$$

Likewise, if a parallel LC system includes a dissipative coil and nondissipative condenser and takes energy from a constant current source (zero admittance source) the above equality holds approximately. This circumstance makes it possible to measure the Q of coils by methods based on selectivity, and conversely, to estimate selectivity of radio amplifier couplings from the Q of the coil.

The two quantities in question have, however, nothing in common except numerical equality in a particular case, and their definition is altogether too restricted to be of use under any other set of conditions.

In order to extend the study of selectivity to any linear system, we must end all confusion between the parameter used to evaluate selectivity and that used to evaluate dissipation. To avoid the introduction of a totally new symbol, we shall use Q_0 for the first purpose, Q for the second. The definitions of Q_0 and Q will be quite general and yet equivalent to the current definitions of Q of a circuit and Q of a coil in the particular cases previously mentioned.

The general definition of Q is well known. Q and *power factor* are used by communication and power engineers for essentially the same purpose. We have, for any linear two-pole

$$Q = \left|\frac{X}{R}\right| = \left|\frac{B}{G}\right| = |\tan \underline{/Z}| = |\tan \underline{/Y}| \tag{176}$$

[*Definition of Q for a two-pole*]

The Q of a coil, $\omega L/R$, evidently is included in the above definition.

Q_0 may be defined indifferently on the basis of the variation of impedance, admittance, voltage, or current with the frequency. Impedance and admittance are better suited for the purpose, because no assumption is required as to the nature of the source. If we seek a definition in terms of current or voltage, or power, we must make some premise with regard to the source.

It is clear that if we wish to describe the behavior of the two-pole at variable frequency by a single real number, the frequency variation must be restricted to a first-order interval in the neighborhood of some particular frequency. Hence any exact expression of Q_0 must include a derivative with respect to frequency taken for this particular frequency value.

Thus, we will write

$$Q_0 = \frac{1}{2}\left|\frac{f}{Z}\frac{\partial Z}{\partial f}\right|_{(f=f_0)} = \frac{1}{2}\left|\frac{f}{Y}\frac{\partial Y}{\partial f}\right|_{(f=f_0)} \tag{177}$$

[*Definitions of Q_0 based on variation of Z and Y with frequency*]

In the above, f_0 is a value of frequency for which Z (or Y) is a real. If there are several such frequencies, there is a value of Q_0 for each. As a rule, we are interested in the behavior of the system in the neighborhood of one particular frequency (midband). It is therefore legitimate to regard Q_0 as single valued. Noting that for $f = f_0$, we have

$$Z = R_0 \qquad Y = G_0$$

we obtain an alternative form of Eqs. (177)

$$Q_0 = \frac{1}{2}\frac{f_0}{R_0}\left|\frac{\partial Z}{\partial f}\right|_{(Z=R_0)} = \frac{1}{2}\frac{f_0}{G_0}\left|\frac{\partial Y}{\partial f}\right|_{(Y=G_0)} \tag{178}$$

Consider now a two-pole receiving energy from a source such that at the frequency f_0 the condition for maximum power transfer (Sec. 8.2) is established. Since for $f = f_0$ the two-pole impedance is R_0, the source impedance must likewise be R_0. Assume, furthermore, the source impedance to be constant with frequency (R_0 at all frequencies) and the source e.m.f. to have constant amplitude, variable frequency. At the generic frequency f, the two-pole impedance is Z and the current is (Fig. 37a)

$$\mathbf{I} = \frac{\mathbf{E}}{R_0 + Z}$$

Differentiating with respect to frequency,

$$\frac{\partial \mathbf{I}}{\partial f} = -\frac{\mathbf{I}^2}{\mathbf{E}}\frac{\partial Z}{\partial f}$$

Letting $\mathbf{I}_0 = \mathbf{E}/2R_0$ stand for the current at f_0 (maximum current), we have from the above

$$\frac{\partial \mathbf{I}}{\partial f}_{(f=f_0)} = -\frac{\mathbf{I}_0}{2R_0}\frac{\partial Z}{\partial f}_{(f=f_0)}$$

Multiplying both sides by $f/\mathbf{I}_0$ and equating the magnitudes,

$$\frac{f}{\mathbf{I}_0}\left|\frac{\partial \mathbf{I}}{\partial f}\right|_{(f=f_0)} = \frac{f}{2R_0}\left|\frac{\partial Z}{\partial f}\right|_{(f=f_0)} = Q_0$$

The last line, and the dual equation similarly obtained, may be used to define Q_0, as follows:

$$Q_0 = \left|\frac{f}{\mathbf{I}}\frac{\partial \mathbf{I}}{\partial f}\right|_{(f=f_0)} = \left|\frac{f}{\mathbf{V}}\frac{\partial \mathbf{V}}{\partial f}\right|_{(f=f_0)} \tag{179}$$

[*Definition of Q_0 based on voltage or current changes in neighborhood of m.p.t. frequency*]

It will be necessary to obtain, from the analytical expressions of Q_0, (177) and (179), practical formulas for finding Q_0 geometrically and for finding the neper (or db) loss of transmitted power, in terms of Q_0 and the frequency departure from midband under representative conditions.

Such formulas must be based on approximations. They result when we substitute finite differences for the differentials of frequency, impedance, and so on.

Let us use the following notation, in agreement with most of the literature:

$$\delta = \frac{2\,\Delta f}{f_0} = \frac{2(f - f_0)}{f_0} \tag{180}$$

∂, the *bandwidth*, should be used only when the departure from midband is small.[1]

Using impedance and admittance numbers, we have also

$$\Delta z = \Delta r + j\,\Delta x = \frac{\Delta Z}{R_0} = \frac{Z - R_0}{R_0} \qquad \Delta y = \Delta g + j\,\Delta b = \frac{Y - G_0}{G_0} \tag{181}$$

[1] If Δf is large, the following variable should replace δ:

$$W = \frac{f}{f_0} - \frac{f_0}{f} \tag{180a}$$

W and δ coincide for small departures from midband.

Replacing the differentials in (178) by finite differences and using (180) and (181), we obtain the following:

$$Q_0 = \frac{|\Delta z|}{\delta} = \frac{|\Delta y|}{\delta} \tag{182}$$

[*Approximate value of* Q_0]

The approximation is better for small values of δ, naturally. Thus, Q_0 may be found by computing a finite change in impedance due to a small frequency variation from midband and dividing by δ. Figure 37 shows this graphically.

To find the loss in transmitted power for a given Q_0 and δ, assume first that m.p.t. is obtained at the midband frequency. In this case, the neper loss in power resulting from a change in load impedance (due to departure of f from midband) is none other than the reflection loss (Sec. 6.6). This is given by (131), as interaction need not be considered in the present case. Equation (131) may be rewritten in the form

$$L_R = \ln \left|\frac{R_0 + Z}{2R_0}\right| \sqrt{\frac{R_0}{R}} \tag{183}$$

having replaced Z_0 by R_0 and Z_r, receiver impedance, by Z, impedance of a generic two-pole. Using (181), we obtain from (183)

$$\begin{aligned} L_R &= \ln \left|1 + \frac{\Delta z}{2}\right| \sqrt{1 + \Delta r} \\ &= \ln \left(\frac{\sqrt{(1 + \Delta r/2)^2 + (\Delta x/2)^2}}{\sqrt{1 + \Delta r}}\right) \\ &= \tfrac{1}{2} \ln \left[1 + \frac{|\Delta z|^2}{4(1 + \Delta r)}\right] \end{aligned}$$

Assuming Δz to be small and expanding in series, we have the simple expression

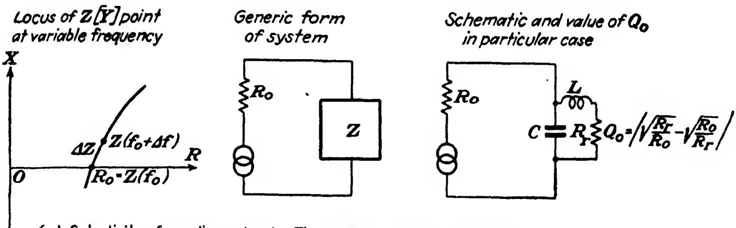
$$L_R = \tfrac{1}{8} |\Delta z|^2 \tag{184}$$

and finally, making use of (182),

$$L_R = \tfrac{1}{8}(Q_0\delta)^2 \tag{185}$$

[*Off-midband loss in terms of* Q_0 *and* δ]

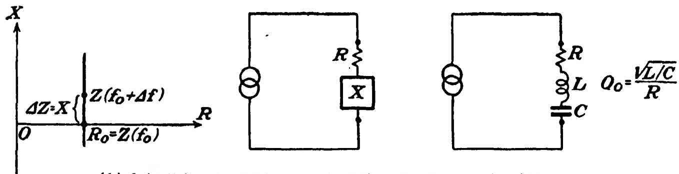
We are now in a position to correlate Q_0 with selectivity in a tangible way. Let δ have a value δ_1 such that $L_R = 1$ db.



(a)-Selectivity of coupling networks. The power loss near midband is:

$$L_R = \frac{1}{8}(Q_0\delta)^2 \text{ nep}$$

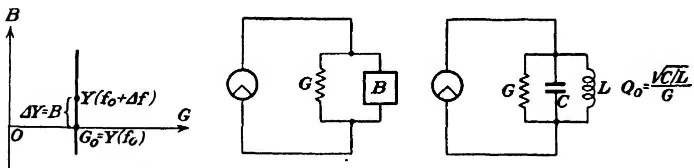
At midband, max. power transfer conditions are satisfied ($Z = R_0$)



(b)-Selectivity of a constant resistance two-pole at constant voltage

$$L_R = \frac{1}{2}(Q_0\delta)^2 \text{nep.}$$

The power loss L_R equals the current drop $\ln \frac{I_0}{I}$



(c)-Selectivity of a constant-conductance two-pole at constant current

$$L_R = \frac{1}{2}(Q_0\delta)^2 \text{nep.}$$

The power loss L_R equals the voltage drop: $\ln \frac{V_0}{V}$ Tuned plate amplifiers approach this

In all cases:

$z = \frac{Z}{R_0}$ $\quad L_R = \frac{1}{2} \ln \frac{P_0}{P}$ $\quad P_0$, power input of two-pole at f_0

$y = \frac{Y}{G_0}$ $\quad \delta = 2\frac{f - f_0}{f_0} = 2(n-1)$ $\quad P$ " " " " " " f

$n = \frac{f}{f_0}$ $\quad Q_0 = \frac{1}{2} \left| \frac{\partial z}{\partial n} \right|_{(n=1)} = \frac{1}{2} \left| \frac{\partial y}{\partial n} \right|_{(n=1)}$ $\quad f_0$, midband frequency for which $Z = R_0$, $Y = G_0$

approx: $Q_0 = \frac{|\Delta z|}{\delta} = \frac{|\Delta y|}{\delta}$

FIG. 37.—*Use of Q_0 in evaluation of selectivity.*

Substituting in (185), we have the identity

$$\frac{1}{8.68} = \tfrac{1}{8}(Q_0\delta_1)^2$$

or

$$Q_0 = \frac{0.96}{\delta_1} \tag{186}$$

Thus, Q_0 *is approximately equal to the reciprocal of bandwidth for a* 1-*db drop in transmitted power,* assuming maximum power transfer at midband.

Equation (186) is useful in the study of coupling networks from the standpoint of selectivity. The coupling network and its load are considered in place of the generic two-pole.

The selectivity of tuned amplifiers requires a different approach. We must now assume that the two-pole is fed from a constant current source (or source of zero admittance; the plate conductance of pentodes is low enough to permit this assumption). We may also consider the two-pole (or tank circuit) to have constant value of conductance in the neighborhood of midband (resonant frequency). Let I stand for the constant r.m.s. value of current and G_0 for the constant conductance, while R will designate the resistance of the tank circuit at some generic frequency and $R_0 = 1/G_0$ will be the resistance at midband. The system is shown in Fig. 37*c*.

We have for the power taken by the tank circuit at frequency f

$$P = I^2R$$

while the power taken at frequency f_0 (midband) is

$$P_0 = I^2R_0 = \frac{I^2}{G_0}$$

Noting that

$$R = \frac{G_0}{G_0^2 + B^2}$$

and that

$$B = Y - G_0 = \Delta Y$$

we have for the power ratio [see Eq. (181)],

$$\frac{P_0}{P} = \frac{G_0^2 + B^2}{G_0^2} = \frac{G_0^2 + (\Delta y)^2}{G_0^2} = 1 + (\Delta y)^2$$

Finally, eliminating Δy in the above, with the help of (182),

$$\frac{P_0}{P} = 1 + (Q_0\delta)^2 \tag{187}$$

Let δ' be the *bandwidth for half power;* in other words, let $\delta = \delta'$ for $P = \frac{1}{2}P_0$. The last line then becomes the identity

$$Q_0 = \frac{1}{\delta'}$$

showing that Q_0, as defined generally by Eqs. (177) and the following, also fits the restricted definition for the Q of a resonant circuit as given in the literature.

Expressing the power ratio in (187) in nepers, and assuming $Q_0\delta$ to be small, we have

$$\frac{1}{2} \ln \frac{P_0}{P} = \frac{1}{2}(Q_0\delta)^2 \tag{188}$$

[*Loss of tank circuit in neighborhood of resonance*]

A similar expression would be obtained with regard to the system of Fig. 37*b* (series resonant circuit with constant voltage source; *i.e.*, the dual of the system in Fig. 37*c*). We conclude that whenever a selective two-pole is energized from a constant current source (for two-poles with constant conductance) or from a constant voltage source (if the two-pole has constant resistance), the selectivity, measured by the loss in power for a given small δ, is *four times greater* than if the source matches the resistance of the two-pole at midband.

In the following section, the value of Q_0 will be computed for the loaded quarter-wave transformer. This will serve as an illustration to the foregoing theory.

9.6. Selectivity of the Quarter-wave Transformer. Reverting to the system of Sec. 9.4, Fig. 36, let us find the value of Q_0 (as defined in the preceding section) for the two-pole consisting of the quarter-wave transformer and all that part of the transmission system which is on the load side of the transformer. We will use the first of Eqs. (177) for the purpose. The two-pole impedance Z is now evidently the input or transformed impedance of the transformer. We have, therefore, (140)

$$Z = R_m \frac{R_b + jR_m \tan (2\pi L/\lambda)}{R_m + jR_b \tan (2\pi L/\lambda)}$$

The symbols are those of Sec. 9.4. Note that L, the length of the transformer, is

$$L = \frac{\lambda_0}{4}$$

where λ_0 is the wavelength at midband. (In Sec. 9.4, the wavelength was considered fixed, hence no subscript was necessary. As we are now dealing with a variable frequency problem, the notation without subscript must indicate a generic value.)

Hence, we may write

$$\frac{2\pi L}{\lambda} = \frac{\pi}{2}\frac{\lambda_0}{\lambda} = \frac{\pi}{2}\frac{f}{f_0} \tag{189}$$

Differentiating Z with respect to frequency, we have

$$\frac{dZ}{df} = j\frac{\pi R_m}{2f_0}\frac{R_m{}^2 - R_b{}^2}{[R_m \cos{(f/f_0)(\pi/2)} + jR_b \sin{(f/f_0)(\pi/2)}]^2}$$

Taking the magnitude of the derivative at $f = f_0$,

$$\left|\frac{dZ}{df}\right|_{(f=f_0)} = \frac{\pi R_m}{2f_0}\left[\left(\frac{R_m}{R_b}\right)^2 - 1\right]$$

Noting that at midband we have

$$Z = R_a$$

Eq. (177) yields the following value of Q_0:

$$Q_0 = \tfrac{1}{2}\left|\frac{f}{Z}\frac{\partial Z}{\partial f}\right|_{(f=f_0)} = \frac{\pi}{4}\frac{R_m}{R_a}\left[\left(\frac{R_m}{R_b}\right)^2 - 1\right]$$

Finally, eliminating R_m (174),

$$Q_0 = \frac{\pi}{4}\left[\sqrt{\frac{R_a}{R_b}} - \sqrt{\frac{R_b}{R_a}}\right] \tag{190}$$

The ratio $\sqrt{R_a/R_b}$ is the *voltage ratio* of the transformer (or inverse of the current ratio). Because it is also equal to the *standing wave ratio* in the transformer, we may denote it by the symbol r_v (146). We may write accordingly

$$Q_0 = \frac{\pi}{4}\left(r_v - \frac{1}{r_v}\right) \tag{191}$$

[*Q_0 of the $\lambda/4$ transformer in terms of r_v, standing wave ratio (voltage ratio)*]

When r_v is large, the above may be written approximately

$$Q_0 = 0.785 r_v$$

Hence, the bandwidth for a 1-db loss is given by (186)

$$\delta_1 = \frac{0.96}{Q_0} = \frac{1.225}{r_v}$$

and the *per cent departure* of frequency from midband for a 1-db loss is

$$100 \frac{\Delta f}{f_0} (1 \text{ db}) = 50\delta_1 = \frac{61.2}{r_v}$$

For example, if a quarter-wave transformer is used to step the voltage up (or down) twenty times, there will be a 1-db loss in power, with respect to midband, when the frequency is 3.06 per cent off midband.

It should be stressed that such computations are less and less accurate as the departure from midband increases. Equation (185), upon which they are based, represents the loss as proportional to the square of δ. Geometrically, the equation represents a parabola (Fig. 38) which osculates the true loss curve at midband frequency. The plot of P/P_0 against f/f_0 is periodic. We note, in fact, that Z, input impedance of the transformer, equals R_a when the transformer length is a quarter-wave plus a whole number of half-waves, or when

$$L = \frac{\lambda}{4} + n\frac{\lambda}{2} = \frac{\lambda}{4}(1 + 2n)$$

Having defined λ_0 as $4L$, we have from the above

$$\frac{\lambda_0}{\lambda} = 1 + 2n = \frac{f}{f_0}$$

In conclusion, maximum power will be transmitted for

$$f = f_0(1 + 2n),$$

or when the frequency is either midband or any *odd harmonic* of midband.

Our analysis has shown that the quarter-wave transformer is satisfactory as a nonselective method of coupling, provided the voltage ratio (square root of the impedance ratio) is below some value which will obviously depend on the requirements of each specific problem and which can easily be arrived at by stating such requirements in the form of a maximum allowable value for Q_0. If the voltage ratio is too high, other means must be used. We must then resort to a truly nonselective transformer, whose design is no longer a single-frequency problem but falls under heading of a transformer problem (Sec. 8.1). It is, essentially, the problem of finding a substitute for the close-coupled inductive

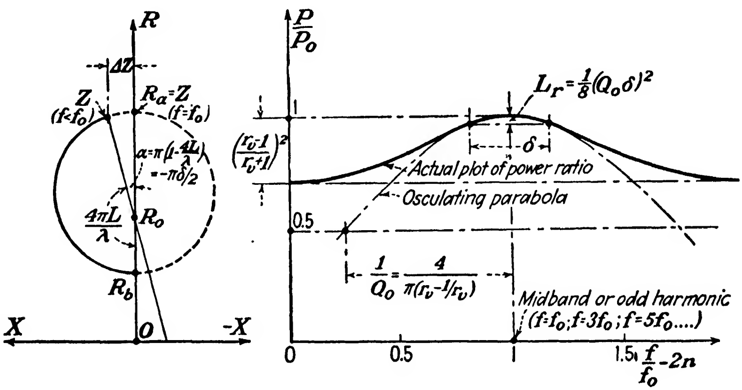


FIG. 38.—*Selectivity of* $\lambda/4$ *transformer.*

The ratio $P/P_0 = \dfrac{\text{transmitted power}}{\text{max. transmitted power}}$ is plotted against $f/f_0 - 2n$ where f = frequency of transmission, f_0 = midband frequency and n = positive integer. Maxima occur when frequency is f_0 or an odd harmonic of f_0.

transformer at high frequencies. A possible solution, the multisection transformer, will be taken up in the following chapter.

9.7. Illustrative examples.

Q_0 of the constant resistance two-pole. The simplest constant resistance two-pole is the series resonant circuit (Fig. 37*b*). For this we have

$$Z = R + j\left(\omega L - \frac{1}{\omega C}\right) = R + j\sqrt{\frac{L}{C}}\left(\omega\sqrt{LC} - \frac{1}{\omega\sqrt{LC}}\right)$$

Using the dimensionless parameters

$$\frac{Z}{R} = z \qquad \omega\sqrt{LC} = \frac{\omega}{\omega_0} = n$$

we may write

$$z = 1 + j\frac{\sqrt{L/C}}{R}\left(n - \frac{1}{n}\right)$$

Differentiating with respect to n,

$$\frac{dz}{dn} = j\frac{\sqrt{L/C}}{R}\left(1 + \frac{1}{n^2}\right)$$

We may write the general expressions for Q_0 (177) in the form

$$Q_0 = \tfrac{1}{2}\left|\frac{dz}{dn}\right|_{(n=1)} = \tfrac{1}{2}\left|\frac{dy}{dn}\right|_{(n=1)} \tag{192}$$

and for the two-pole under consideration

$$Q_0 = \tfrac{1}{2}\left|j\frac{2\sqrt{L/C}}{R}\right| = \frac{\sqrt{L/C}}{R}$$

Note that if we consider the physical coil of the series circuit to include the resistance R, the Q of the coil at the frequency of resonance will be given by

$$Q_{L(f=f_0)} = \frac{\omega_0 L}{R} = \frac{\sqrt{L/C}}{R}$$

from which we see that in the constant resistance two-pole

$$Q_0 = Q_{L(f=f_0)}$$

as previously noted (Sec. 9.5).

The constant G two-pole of Fig. 37c is dual to the above. We may, therefore, write immediately for this two-pole

$$Q_0 = \frac{\sqrt{C/L}}{G}$$

A combination of coil and condenser in shunt can be represented only approximately by Fig. 37c because the conductance of a dissipative coil is not constant. A better representation is that of Fig. 37a. We shall find, however, that, for the same values of L and C, if the coil dissipation is small, Q_0 is approximately the same for Fig. 37a and Fig. 37c.

Selectivity of L networks. L networks as a means of coupling a source to its load have been discussed in Sec. 8.5. We will now evaluate Q_0 for one such network and its load, jointly considered as a two-pole. Resistive terminations will be assumed. We have (Fig. 39a)

$$Z = j\omega L + \frac{1}{G_r + j\omega C}$$

At midband, Z must have the value R_0 equal to the generator resistance. Hence, for $\omega = \omega_0$,

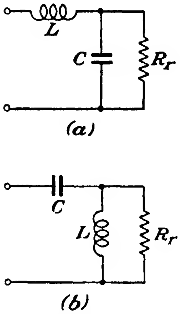


FIG. 39.—*Coupling networks.*

$$Z_{\omega=\omega_0} = R_0 = j\omega_0 L + \frac{G_r - j\omega_0 C}{G_r^2 + \omega_0^2 C^2}$$

We have from the above

$$L = \frac{C}{G_r^2 + \omega_0^2 C^2} \qquad R_0 = \frac{G_r}{G_r^2 + \omega_0^2 C^2}$$
$$(R_0 < R_r)$$

Hence

$$\frac{L}{C} = R_0 R_r \qquad \omega_0^2 C^2 = \frac{C}{L} - G_r^2$$

and the element values

$$L = \frac{R_0}{\omega_0}\sqrt{r_v^2 - 1} \qquad C = \frac{1}{\omega_0 R_r}\sqrt{r_v^2 - 1}$$

having written (in keeping with the notation used for the quarter-wave transformer) $r_v = \sqrt{R_r/R_0}$ for the voltage ratio of the coupling network.

The impedance Z may now be written in ratio form, as follows:

$$z = jn\sqrt{r_v^2 - 1} + \frac{r_v^2}{1 + jn\sqrt{r_v^2 - 1}}$$

where

$$z = \frac{Z}{R_0} \qquad n = \frac{\omega}{\omega_0}$$

Adding and subtracting unity, we may reduce z to the more convenient form

$$z = 1 - \frac{(n^2 - 1)(r_v^2 - 1)}{D}$$
$$[D = 1 + jn\sqrt{r_v^2 - 1}] \tag{193}$$

Differentiating,

$$\frac{dz}{dn} = (r_v^2 - 1)\,\frac{-2nD + (n^2 - 1)\,dD/dn}{D^2}$$

Hence, we have for Q_0

$$Q_0 = \frac{1}{2}\left|\frac{dz}{dn}\right|_{(n=1)} = \left|\frac{r_v^2 - 1}{D}\right| = r_v - \frac{1}{r_v} \tag{194}$$

When Q_0 is large, it is approximately equal to r_v and can be written in terms of L and C as follows:

$$Q_0 \approx r_v = \sqrt{\frac{R_r}{R_0}} = \sqrt{\frac{R_r R_0}{R_0}} = \frac{\sqrt{L/C}}{R_0}$$

Since R_0 is the resistance of the two-pole at midband, the above coincides with (192), obtained for the simple series LC combination. We conclude that if the two configurations of Fig. 39*a* and Fig. 37*b* have the same values of L and C and the same low value of resistance for the frequency of zero reactance, their selectivity is very nearly the same. If, however, in the system of Fig. 39*a*, R_0 and R_r are comparable, this equality no longer holds.

Another point worthy of note comes up when we compare (194) with (191), the expression for Q_0 of the quarter-wave transformer. The form of the two expressions is the same but we find that, in all cases, the L network is more selective by 12.7 per cent.

The system of Fig. 39*a* and that of Fig. 37*a* are dual. Equation (194) gives therefore the value of Q_0 for both systems, except that in the case of Fig. 37*a* we have

$$r_v = \sqrt{\frac{G_r}{G_0}} = \sqrt{\frac{R_0}{R_r}}$$

(G must be written in place of R when we go from a system to its dual). The approximate expression for r_v large may be obtained in the same way as before

$$Q_0 \approx \frac{\sqrt{C/L}}{G_0}$$

We see from the above that Q_0 is the same for the two configurations of Fig. 37*a* and Fig. 37*c* if L and C are the same, and if G_t has the same low value. Parallel combinations of coil and condenser (tank circuits) are more accurately represented by Fig. 37*a*, but the circuit of Fig. 37*c* can be used much more conveniently with approximately the same results.

Suggested Exercise. Find the value of Q_0 for the two-poles of Figs. 39*a* and 39*b*.

CHAPTER X

COMPOSITE LINES AND STUBS

10.1. Multiple Reflection. The solution of the transformer problem by transmission line methods will be discussed at this point, although it would be more logical, in some respects, to defer it until after the treatment of the transformer proper (Chap. XIV).

There are strong reasons in favor of discussing at one time devices for solving different problems by similar means, rather than discussing dissimilar methods used in solving the same problem.

The problem of securing uniform transmission between different impedance levels over a wide frequency band is solved at audio frequencies by the transformer proper. At all frequencies, the flow of power into nonmatching loads may be increased by the use of vacuum tube amplifiers (Chap. XVII).

Wavelengths comparable to apparatus dimensions permit a solution of the transformer problem based on *multiple reflection* taking place at the discontinuities of characteristic impedance in a disuniform transmission line.

Multiple reflection may be explained by the method previously used to help understand the action of the quarter-wave transformer (Sec. 9.3). Another approach, through the values taken by the input impedance, is suitable here, as formerly, for a quantitative discussion.

Consider (Fig. 40) a disuniform transmission line, coaxial or otherwise, consisting of a number of sections of equal length L but of different characteristic impedances. A wave entering the system will be reflected at each junction of two consecutive sections. If we define the phase of any one reflected wave as the angle of the voltage vector due to this wave with the voltage vector due to the wave reflected at the first junction, both voltages being taken at any point of the input line (line feeding into the system), and define the nth reflected wave as that due to reflection at the nth junction, we may find the phase of the nth wave quite readily.

The difference in travel of the two waves (the first and nth wave) is twice the distance between points of reflection, or $2(n - 1)L$, assuming that all reflections are accompanied by

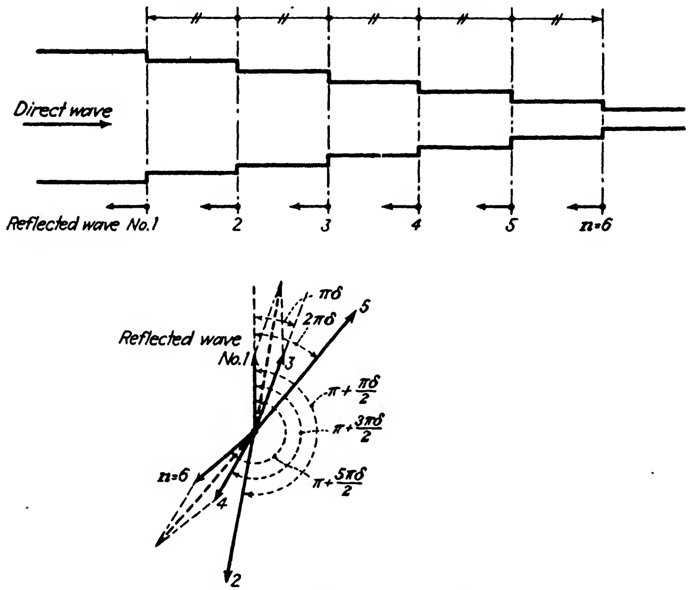


FIG. 40.—*Multiple reflection.*

equal phase shifts (Sec. 5.4). The phase ϕ of the nth wave is to 2π as this distance is to the wave length λ, hence

$$\phi = \frac{2(n - 1)L2\pi}{\lambda} = \frac{4(n - 1)\pi L}{\lambda}$$

Letting $\lambda_0 = 4L$, we have

$$\phi = \pi(n - 1)\frac{\lambda_0}{\lambda} = \pi(n - 1)\frac{f}{f_0}$$

where f_0, midband frequency, is that value at which each section is a quarter wavelength long. There is some advantage in using δ, the bandwidth, in place of frequency. We have from (180)

$$\frac{f}{f_0} = \frac{\delta}{2} + 1$$

Hence

$$\phi = (n - 1)\pi\left(\frac{\delta}{2} + 1\right)$$

Computing the phase difference to the nearest cycle, which is sufficient for the purpose, we have

1. For odd-order waves ($n = 1, 2, 3, \cdots$):

$$\phi = 0, \pi\delta, 2\pi\delta, 3\pi\delta, \cdots$$

2. For even-order waves ($n = 2, 4, 6, \cdots$):

$$\phi = \pi + \frac{\pi\delta}{2}, \pi + \frac{3\pi\delta}{2}, \pi + \frac{5\pi\delta}{2}, \cdots$$

Figure 40 illustrates the case of $n = 6$ (5 sections). It can be seen that if the amplitudes of the reflected waves in the input line are in the correct mutual relation, they all add up to zero. (The first and third cancel the second; the fourth and sixth cancel the fifth.) The same thing is evidently true for any number of sections. It is therefore possible for the input line to be totally free of reflection, in which case the condition of maximum power transfer is satisfied at junction 1, hence anywhere in the system.

The validity of this result does not depend *explicitly* on the value of δ. However, we cannot prove by this type of reasoning whether or not the assumptions we have made regarding the amplitude of the reflected waves and the phase shifts at the points of reflection can be satisfied independently of frequency. We are only justified in concluding that the system in question can probably be made to transmit uniformly over a range of frequencies, whose extent will depend on the number of sections used.

Exact data on the optimum design and performance of the system will be obtained by expressing the input impedance as a function of δ.

10.2.★ The Multisection Transformer. Figure 41*a* reproduces, for convenience, the system of Fig. 40 with this difference, that the sections are now numbered instead of the junctions. Sections 1, 2, 3, . . . , n are arranged in order of increasing characteristic impedance; this is brought from the value R_a for the input line to the value R_b for the output line, in a series of $n + 1$ steps of values $A_0, A_1, A_2, \ldots, A_n$. The generic term A_k in this sequence is the difference in characteristic impedance between the section of order $k + 1$ and the section of order k.

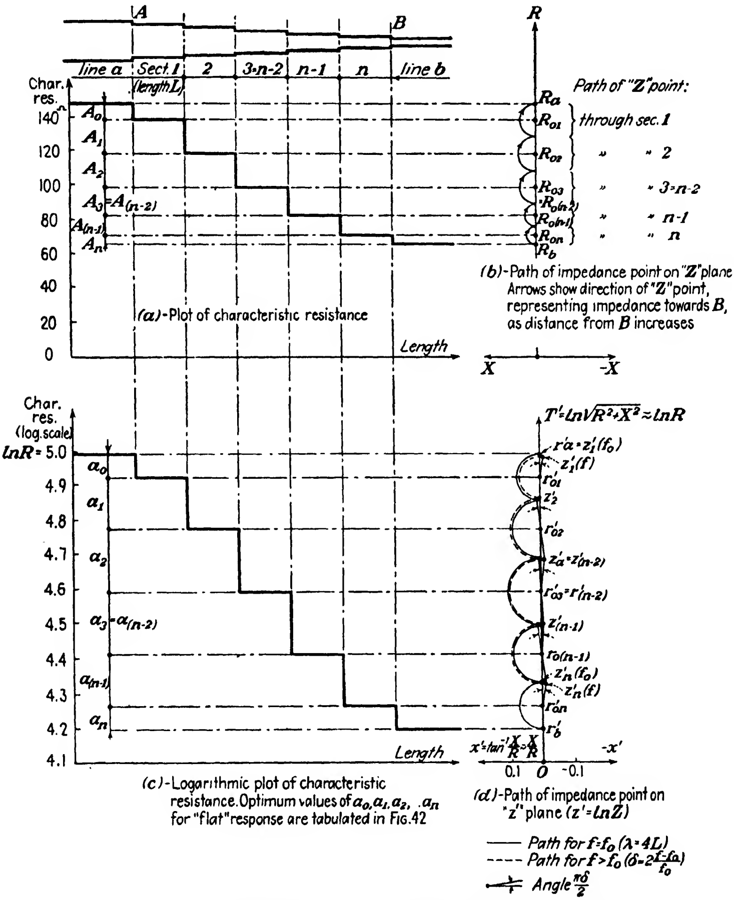


FIG. 41.—*Diagram of the multisection transformer.*

If L is the common length of all sections, when the wavelength is $\lambda = 4L$, each section operates as a quarter-wave transformer at midband, and the impedance toward load (toward R_b) varies through the system as shown in Fig. 41b.

The analysis of the system simplifies materially if we use the transformation of Fig. 41*d*. Consider the complex quantity

$$z' = r' + jx' = \ln Z \tag{195}$$

We may think of a plane, every point of which has coordinate r' and x', representing, in fact, the complex number z'. This plane will be used as a basis of discussion instead of the Z plane.

We must find explicit relations giving r' and x' in terms of R and X. These relations (implicit in Eq. 195) will enable us to locate the point on the z' plane corresponding to a given Z point.

We may write (Sec. 2.2)

$$\begin{aligned} z' &= \ln Z = \ln |z| + j\underline{/Z} \\ &= \ln \sqrt{R^2 + X^2} + j \tan^{-1} \frac{X}{R} \end{aligned}$$

and hence

$$\left.\begin{aligned} r' &= \ln \sqrt{R^2 + X^2} \\ x' &= \tan^{-1} \frac{X}{R} \end{aligned}\right\} \tag{196}$$

Consider now the input impedance of the nth section and the corresponding point in the new representation. Let us agree to use the symbol Z_k for the input impedance of section k and R_{0k} for the characteristic impedance of the same section. At midband, each section has the impedance transformation action of a quarter-wave section (Sec. 9.3). Hence we have, for $\delta = 0$,

$$Z_n = R_n = \frac{(R_{0n})^2}{R_b}$$

Taking logarithms,

$$\ln R_n = 2 \ln R_{0n} - \ln R_b$$

Hence, and from (196), noting that $X_n = 0$,

$$\begin{aligned} r_n' &= 2r_{0n}' - r_b' \\ \therefore\ r_{0n}' &= \tfrac{1}{2}(r_n' + r_b') \end{aligned}$$

Note that, while on the Z plane the characteristic resistance of the section is the *geometric mean* of the input and load impedances at midband, on the z' plane r_{0n}' is the *arithmetic mean* of r_n' and r_b'. Let us now make the position

$$a_k = r_{0k}' - r_{0(k+1)}' \tag{197}$$

and in particular

$$a_n = r_{0n}' - r_b'$$

The sequence $a_0, a_1, a_2, \ldots, a_n$ corresponds to the sequence $A_0, A_1, A_2, \ldots, A_n$ in the logarithmic transformation. We write, accordingly, for $\delta = 0$,

$$r_n' = z_n = r_{0n}' + a_n \tag{198}$$

We now ask for the value of z_n' when $\delta \neq 0$. The answer takes a simple form if a_n is small compared to r_{0n}'. Consider (Fig. 31) the constant σ and constant τ circles (Sec. 8.7), at whose intersection the Z point falls for δ other than zero. Note, in particular, that the constant τ circle crosses the R axis at an angle

$$\pi - \frac{4\pi L}{\lambda} = \pi\left(1 - \frac{\lambda_0}{\lambda}\right) = \pi\left(1 - \frac{f}{f_0}\right) = \frac{\pi\delta}{2}$$

Note, furthermore, that the same constant τ circle crosses the constant σ circle at right angles.

Now, imagine the whole construction transferred to the z' plane. As the transformation is isogonal[4] (Sec. 5.7) the constant σ path is still approximately a circle, provided its radius is small compared with the distance from the origin ($a_n \ll r_{0n}'$). The constant τ path must go through the center of this circle and cut it at right angles; it may therefore be assimilated to a radius of the circle, rotated by an angle $\pi\delta/2$ from the direction of the negative r' axis.[1]

We conclude that the z_n' point may be expressed analytically, using the exponential notation, as follows:

$$z_n' = r_{0n}' + a_n e^{-j\frac{\pi\delta}{2}} \tag{199}$$

To find the input impedance of the next section, we may argue that the corresponding z' point lies diametrally opposite to z_n' about the new center $r_{0(n-1)}'$, except for a negative rotation about $r_{0(n-1)}'$ by an angle $\pi\delta/2$. Analytically,

$$z_{(n-1)}' = r_{0(n-1)}' + (r_{0(n-1)}' - z_n')e^{-j\frac{\pi\delta}{2}}$$

Letting, for brevity, p stand for the operator $e^{-j\frac{\pi\delta}{2}}$, the above

[1] More rigid proof of this argument is not difficult, but tedious. For practical purposes, the error involved is small enough to be negligible.

may be written as follows:

$$z_{(n-1)}' = r_{0(n-1)}' + a_{(n-1)}p - a_n p^2$$
$$\left(p = e^{-j\frac{\pi\delta}{2}}\right)$$

Continuing in the same way, we find for z_1', which corresponds to the input impedance to the first section or to the entire multisection transformer, the following expression:

$$z_1' = r_{01}' + a_1 p - a_2 p^2 + a_3 p^3 - \cdots \pm a_n p^n \tag{200}$$

The plus sign before the last term should be used when n is odd.

Rather than in z' itself, we are interested in the departure of z' from its midband value. Letting $\Delta z'$ stand for this departure, we define $\Delta z'$ as follows (Fig. 41d):

$$\Delta z' = z_1' - r_1' = z_1' - r_{01}' - a_0 \tag{201}$$

We have, therefore

$$\Delta z' = -[a_0 - a_1 p + a_2 p^2 - \cdots \pm a_n p^n] \tag{202}$$

To carry on the analysis, we must express $\Delta z'$ as a series in powers of δ. If we let

$$\Delta z' = \Phi(\delta)$$

the following expansion may be used for the purpose (*Maclaurin's* expansion[4]):

$$\Phi(\delta) = \Phi(0) + \Phi'(0)\delta + \frac{\Phi''(0)}{2!}\delta^2 + \cdots + \frac{\Phi^{(m)}(0)}{m!}\delta^m + \cdots.$$

In the above, for the function under discussion, the first term is zero; the mth derivative with respect to δ of (202) for $p = 1$ $(\delta = 0)$ may be written by inspection

$\left[\text{the } m\text{th derivative of } e^{k\left(-\frac{\pi\delta}{2}\right)} \text{ is } \left(-j\frac{k\pi}{2}\right)^m e^{k\left(-j\frac{\pi\delta}{2}\right)}, \text{ or } \left(-j\frac{k\pi}{2}\right)^m \text{ for } \delta = 0\right]$:

$$\Phi^{(m)}(0) = a_1\left(-j\frac{\pi}{2}\right)^m - 2^m a_2\left(-j\frac{\pi}{2}\right)^m + 3^m a_3\left(-j\frac{\pi}{2}\right)^m - \cdots \pm n^m a_n\left(-j\frac{\pi}{2}\right)^m$$
$$= \left(-j\frac{\pi}{2}\right)^m [a_1 - 2^m a_2 + 3^m a_3 - \cdots \pm n^m a_n]$$

Hence

$$\frac{\Phi^{(m)}(0)}{m!}\,\delta^m = \frac{\left(-j\frac{\pi\delta}{2}\right)^m}{m!}\,[a_1 - 2^m a_2 + 3^m a_3 - \cdots \pm n^m a_n]$$

$$= -\frac{\left(-j\frac{\pi\delta}{2}\right)^m}{m!}\sum_{k=1}^{n}(-1)^k k^m a_k$$

and we have for the entire expansion

$$\Delta z' = -\sum_{m=1}^{\infty}\frac{\left(-j\frac{\pi\delta}{2}\right)^m}{m!}\left\{\sum_{k=1}^{n}(-1)^k k^m a_k\right\} \tag{203}$$

We may expand the double summation in the following form, better suited to the purpose of the discussion:

$$\begin{aligned}
\Delta z' = {} & j\frac{\pi}{2}(-a_1 + 2a_2 - 3a_3 + \cdots \pm na_n)\delta \\
& + \frac{\pi^2}{8}(-a_1 + 4a_2 - 9a_3 + \cdots \pm n^2 a_n)\delta^2 \\
& - j\frac{\pi^3}{54}(-a_1 + 8a_2 - 27a_3 + \cdots \pm n^3 a_n)\delta^3 \\
& \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
& - \frac{\left(-j\frac{\pi}{2}\right)^m}{m!}(-a_1 + 2^m a_2 - 3^m a_3 + \cdots \pm n^m a_n)\delta^m
\end{aligned} \tag{204}$$

10.3. Optimum Design for Wide-band Transmission. The multisection transformer has been analyzed in the preceding section to the extent of obtaining the *departure* of the input impedance from the midband value in terms of the bandwidth δ. We have actually expressed this departure in logarithmic form as the quantity

$$\Delta z' = z' - r_a' = \ln\frac{Z}{R_a} \tag{205}$$

No particular value has been assigned as yet to the *impedance steps* $a_0, a_1, \ldots, a_n$ which (all except a_0) appear in the coefficients of powers of δ in the series expansion of $\Delta z'$.

Inspection of (204) discloses that if enough sections are used (n sufficiently large), we may cause an arbitrary number of terms in the expansion to vanish by a correct choice of the impedance steps. Assuming $\delta \ll 1$, which makes the series convergent, this means that $|\Delta z'|$ can be made as small as we like.

Given the number of sections used, the following points must be taken up:

1. How many terms of (204) can be made to vanish by a proper selection of the impedance steps?
2. How are the steps to be selected?
3. If the steps have been selected correctly (optimum design), what is the power loss referred to midband in terms of δ?

In answer to the first question, observe that two relations exist between the $n + 1$ steps $a_0, a_1, a_2 \ldots, a_n$, independently of the values assigned to them. The total difference of impedance determines their sum, which may therefore be regarded as a known quantity

$$\sum a = r_a' - r_b' = \ln \frac{R_a}{R_b} \tag{206}$$

In addition, if we examine carefully the solid-line diagram of Fig. 41d (representing the locus of the z' point for midband as we go through the system), we find that the above sum may be obtained by adding every other step, beginning with either the first or the second, and doubling the total. In fact, each step includes the radii of two adjoining semicircles; therefore, when we double every other step, we are taking the sum of all the diameters or of all the steps. Symbolically, we may write

$$\Sigma a = 2\Sigma a_{\text{even}} = 2\Sigma a_{\text{odd}}$$

or identically,

$$\Sigma a_{\text{even}} - \Sigma a_{\text{odd}} = 0$$

and expanding,

$$a_0 - a_1 + a_2 - a_3 + \cdots \pm a_n = 0 \tag{207}$$

We need $n - 1$ equations, in addition to (206) and (207), in order to assign the values of the $n + 1$ steps. By setting the first $n - 1$ coefficients of (204) equal to zero, we obtain that many equations. We conclude that, if n sections are used, $n - 1$ terms of (204) can be made to vanish by proper design, leaving only terms of nth and higher order in δ.

Passing now to the second question, combining (207) with the equations obtained by setting the first $n - 1$ terms of (204) equal to zero, we have the system

$$\begin{cases} a_0 - a_1 + a_2 - \cdots \pm a_n & = 0 \\ -a_1 + 2a_2 - \cdots \pm na_n & = 0 \\ -a_1 + 4a_2 - \cdots \pm n^2a_n & = 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots & \\ -a_1 + 2^{n-1}a_2 - \cdots \pm n^{n-1}a_n & = 0 \end{cases} \tag{208}$$

The system above determines not the steps themselves but the ratios between any two of them. When these are known, (205) may be used to find the actual values. If we solve (208) for different values of n, we invariably arrive at the proportion

$$a_0:a_1:a_2: \cdots :a_n = b_0:b_1:b_2: \cdots :b_n \tag{209}$$

where $b_0, b_1, \ldots, b_n$ are the *binomial coefficients* of order $n + 1$.[1] From the above and (205) we obtain

$$a_0 = b_0 \frac{\Sigma a}{\Sigma b}$$

$$a_1 = b_1 \frac{\Sigma a}{\Sigma b}$$

or, in general,

$$a_K = \left[\ln \frac{R_a}{R_b}\right] \frac{b_K}{\sum_{K=0}^{n} b_K} \tag{210}$$

A table of binomial coefficients, which may be extended indefinitely by a very simple rule, is given in Fig. 42. Any coefficient of the table is the sum of the one directly above with the one above and to the left. The sums of the coefficients advance as the powers of 2. A numerical example, for $n = 5$, accompanies the table.

Next comes the question of performance. This must be judged on the basis of power loss, referred to midband, over a wide range of frequencies on either side of midband. A single parameter,

[1] General proof of (209), although not particularly difficult, is beyond the scope of the present discussion. W. W. Hanson is credited with having shown that the steps should be in the same proportion as the binomial coefficients; his work on the subject does not appear to have been published at the time of this writing. (See J. C. Slater, "Microwave transmission," McGraw-Hill Book Company, Inc., p. 60 1941.

such as Q_0, could only serve to establish the loss in the immediate neighborhood of midband, hence would be of no value. The information called for can only be adequately expressed by a set of curves (one for each value of n) of power loss against frequency, obtained under the assumption of *optimum design.*

Number of sections n:	Binomial coefficients:							$\sum_{k=0}^{n} b_k$:	First non-zero term of Δz Multiply by $\ln(Ra/Rb)$
	b_0	b_1	b_2	b_3	b_4	b_5	b_6		
1	1	1						2	$-j\frac{n\delta}{4}$
2	1	2	1					4	$-\frac{(n\delta)^2}{16}$
3	1	3	3	1				8	$j\frac{(n\delta)^3}{64}$
4	1	4	4	4	1			16	$\frac{(n\delta)^4}{256}$
5	1	5	10	10	5	1		32	$-j\frac{(n\delta)^5}{1024}$
6	1	6	15	20	15	6	1	64	$-\frac{(n\delta)^6}{4096}$

FIG. 42.—*Table of binomial coefficients.* For optimum performance of the multisection transformer, the logarithmic impedance steps $a_0\ a_1\ a_2\ .\ .\ .\ a_n$ (Fig. 41) should have the following value:

$$a_k = \left[\ln \frac{R_a}{R_b}\right] \frac{b_k}{\sum\limits_{k=0}^{n} b_k}$$

NOTE: R_a and R_b = terminal impedances; b_k = kth binomial coefficient; n = number of sections.

Numerical example:

R_a = 150 ohms; R_b = 45 ohms ($\ln R_a/R_b$ 1.20387); n = 5
Binomial coefficients $b_0\ b_1\ b_2$ . . . . :
1; 5; 10; 10; 5; 1
Logarithmic steps $a_0\ a_1\ a_2$ . . . . :
0.0376; 0.188; 0.376; 0.376; 0.188; 0.0376
Corresponding impedance ratios:
1.0382; 1.207; 1.456; 1.456; 1.207; 1.0382
Characteristic impedances (ohms):
144.4; 119.6; 82.1; 56.4; 46.8

A number of predictions regarding the variation of transmitted power with frequency may be made at the start. At the frequencies f_0, $3f_0$, $5f_0$, . . . (midband and odd harmonics) the length L of each section equals a quarter wavelength plus a whole number of half wavelengths. The impedance transforming action of each section is therefore that of a quarter-wave section. It follows that, if maximum power transfer exists at midband, it will also exist at the odd harmonics of midband, as in the case of the quarter-wave transformer.

At frequencies 0, $2f_0$, $4f_0$, . . . (zero and even harmonics of midband), the sections have no impedance transforming action, and the loss will depend only on the end values of impedance. More precisely, the loss referred to midband will have the value (183)

$$L_R = \ln \tfrac{1}{2}\left(\sqrt{\frac{R_a}{R_b}} + \sqrt{\frac{R_b}{R_a}}\right) \tag{211}$$

We conclude that, irrespective of the number of sections, the plot of transmitted power will be periodic, with points of minimum at frequencies 0, $2f_0$, $4f_0$. . . . These minimum values of transmitted power will be all the same and will depend only on the total unbalance of impedance.

For practical purposes, the evaluation of L_R need not be accurate except where L_R is relatively small, as this is the condition which delimits the useful operating range of the device. The following approximations are therefore in order:

$$L_R = \tfrac{1}{8}|\Delta z|^2 \tag{184}$$

$$\begin{aligned} \Delta z' &= \ln Z - \ln R_a = \ln\left(1 + \frac{\Delta Z}{R_a}\right) \\ &= \ln(1 + \Delta z) = \Delta z \end{aligned} \tag{212}$$

$$\left[z = \frac{Z}{R_a} \qquad \Delta Z = Z - R_a\right]$$

and combining,

$$L_R = \tfrac{1}{8}|\Delta z'|^2 \tag{213}$$

The evaluation of L_R thus reduces to the evaluation of $\Delta z'$. The series expansion (204), which has been used to determine the rule for optimum design, is not useful for the present purpose except when δ is small, in which case the expansion converges so rapidly that it may be identified with its lowest order term, which as we know is the term of order n. The value of this term is given in the tabulation of Fig. 42.

To obtain values over a wide range, we must resort to (202). Evaluation of this expression, although tedious, is not difficult. It is best carried out graphically, as a vector addition. We find in this manner that as δ increases, the z' point moves out from r' in a spiral (Fig. 43). The radius vector of the spiral is the magnitude of $\Delta z'$; L_R may be obtained from this by squaring and dividing by 8. As a result, the plots of Fig. 44 have been drawn. They are universal; to fit a particular case, the ordinate must be multiplied as indicated.

We learn from the plots that there is little advantage in increasing the number of sections beyond four or five. With five

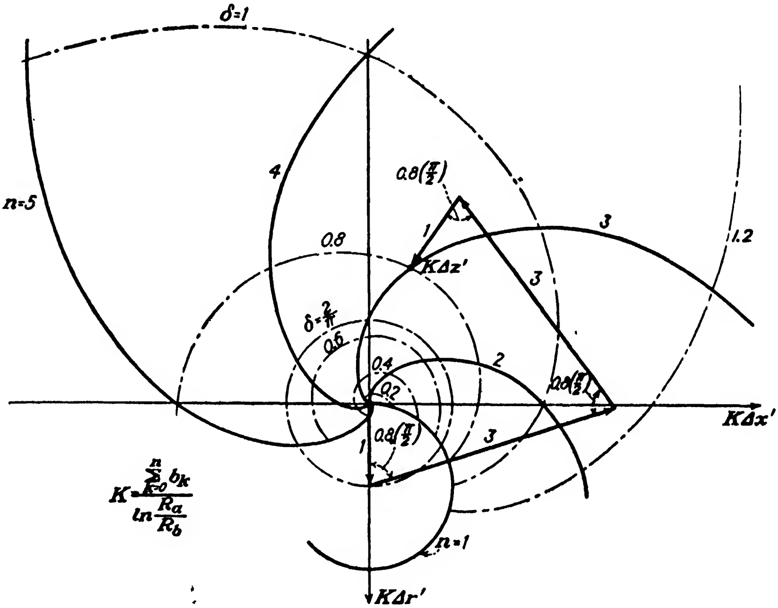


FIG. 43.—*Polar plots of* $K\Delta z' = K \ln \dfrac{Z}{R_a}$, *illustrating the variation of input impedance Z with the frequency for the n-section transformer.* Points are obtained by vector sum according to (202). Construction shown for $n = 3$, $\delta = 0.8$.

sections, transmission is essentially uniform within 50 per cent of midband.

10.4. Tuning Stubs. At the cost of some complication, the selectivity associated with impedance transformation at high frequency can be eliminated for practical purposes by the use of the multisection transformer.

A high value of selectivity, on the other hand, is generally not objectionable in adjustable coupling devices, or *tuners.* The study of these devices will now be taken up in some detail.

Adjustable couplings must take the place of fixed couplings in the following cases:

1. When uniform transmission is required, and the value of midband frequency must vary from time to time over a wide range.

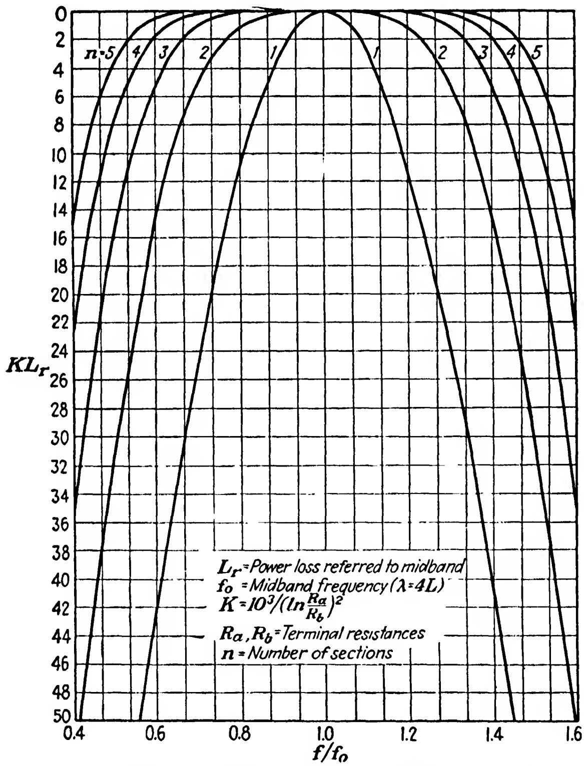


FIG. 44.—*Amplitude distortion of the multisection transformer.*

2. When the load or source impedance (or both) are unknown, or inaccurately known, or subject to variation. (It should be noted that such variations usually accompany variations of frequency.)

If only the frequency is subject to change, a single adjustment may suffice. As an example, consider a quarter-wave transformer inserted between two long lines. A simple change of length will *tune* the transformer to any frequency. The example is not of practical value, because the quarter-wave transformer lacks selectivity.

If the terminal impedances are variable, two adjustments are generally needed to preserve the condition of maximum power transfer. One of the adjustable parameters might be the length of the coupling section, the other, its characteristic impedance. It is more convenient, however, from a mechanical standpoint, to leave both length and characteristic impedance unaltered and to add one or two *adjustable susceptances* in shunt with the line. If only one susceptance is added, its *position* must be made variable to provide the second adjustment. This is the principle of the *sliding-shunt tuner* (Sec. 10.6). Two variable susceptances shunting the line at fixed points constitute a *double-shunt tuner* (Sec. 10.5).

As a preliminary to the study of tuners, we must therefore discuss the method used at high frequencies to provide a variable susceptance. Consider the input admittance of a section of lossless line, short-circuited at the far end. Making the usual substitutions, valid for lossless lines, in the generic expression (102), we have

$$Y_{sc} = -jG_{0s} \cot \frac{2\pi L}{\lambda} \tag{214}$$

It should be noted that the above expression, in which α has been neglected in comparison with β, becomes quite inaccurate for physical lines under particular conditions (Sec. 7.3). The exact expression for the input admittance will be given later (Sec. 10.9); we may now observe that (214) represents a susceptance which takes all values from $-\infty$ to $+\infty$ as L is varied from 0 to $\lambda/2$. Actually, infinite values are never reached and Y_{sc} is a conductance for these values of L.

We may conclude, at any rate, that a line section of length $\lambda/2$, equipped with a movable shorting connection (a *plunger* if the line is coaxial), offers at the free end a susceptance that may be varied between a very large negative value and a very large positive value. The conductance of such a line is small and may be disregarded, except as a loss factor (Sec. 10.9).

Such sections are called *stubs*. They replace lumped element tank circuits at frequencies so high that coils are no longer usable, while the wavelength is short enough to permit the practical use of shunt branches a half wavelength long. The construction and general properties of stubs are illustrated in Fig. 45.

Open-circuited lines of length variable between $\lambda/4$ and $3\lambda/4$ could theoretically be used in place of stubs. In some ways, such a solution is preferable, as it does not require high currents to flow through a sliding contact.

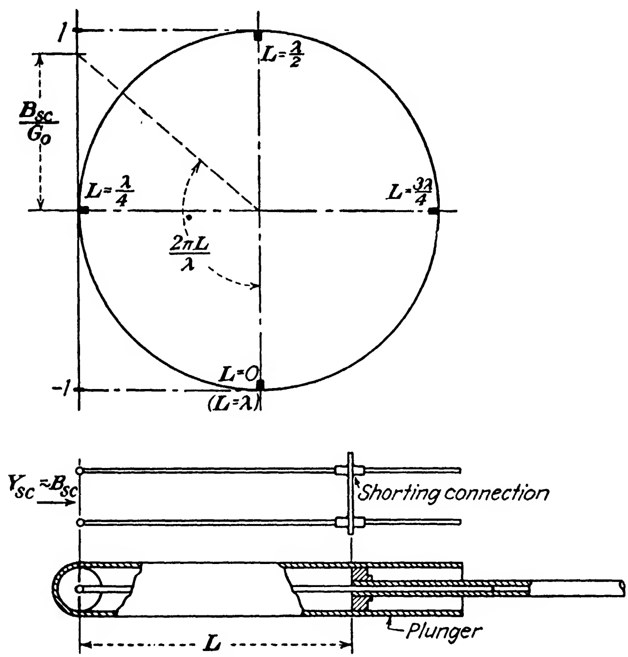


FIG. 45.—*Short-circuited line section (stub).*

10.5. The Double-shunt Tuner: Use of the k Plane. If two stubs are shunted across a source and its load before coupling, the corresponding Y points take some position along the two ordinates whose distances from the origin are, respectively, G_r and G_t. The two points may be moved up and down the ordinates by adjusting the two stubs. Under these conditions, the coupling may be effected by a line section whose characteristics are arbitrary, within limits.

Such a system, known as a *double-shunt tuner*, is shown in Fig. 46. Except for the addition of the stubs, this is again the coupling section of Fig. 34. However, while in Fig. 34 the coupling section was required to move the impedance (or admittance) point from the *point* Z_r to the *point* Z_t, in Fig. 46 the coupling section may bring the Y point from *anywhere on the line* $G = G_r$ to *anywhere on the line* $G = G_t$. Given a line section, it is generally possible to find two points on these loci which represent a possible pair of values for the load and input admittance of the section.

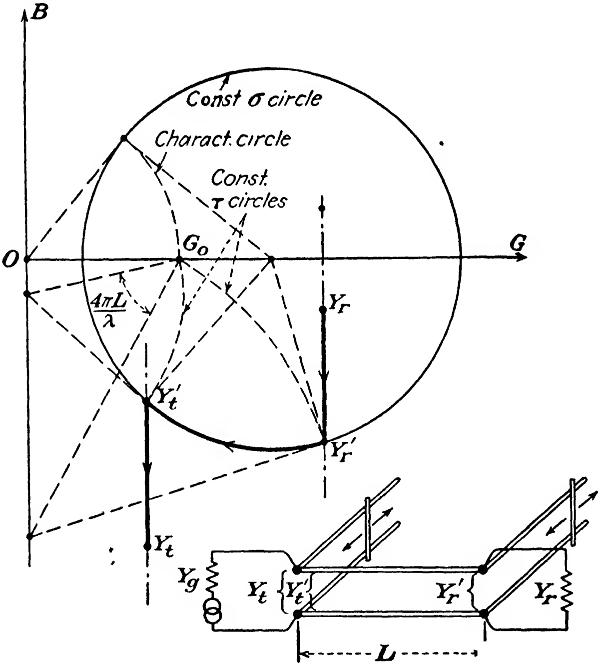


FIG. 46.—*Double-shunt tuner.*

These two points cannot be located very readily on the Y plane. Here, as in previous problems, there is advantage in using the particular method of representation which best fits the conditions of the problem. Since R_0, the characteristic impedance of the coupling section, may now be regarded as a known quantity, we are free to consider, in place of Z or Y, any function of Z and R_0, or Y and G_0. The function ρ has already been used extensively in place of impedance, both geometrically and analytically. In dealing with lossless networks, some of the advantages of the ρ plane cease to be important, and there are good reasons for selecting as a method of representation the k plane. This is based on the function

$$k = e^{-2\rho} = \frac{1-y}{1+y} = \frac{z-1}{z+1}$$

$$\left[y = \frac{Y}{G_0} \qquad z = \frac{Z}{R_0} \right]$$

introduced in Sec. 5.4 as the reflection coefficient.

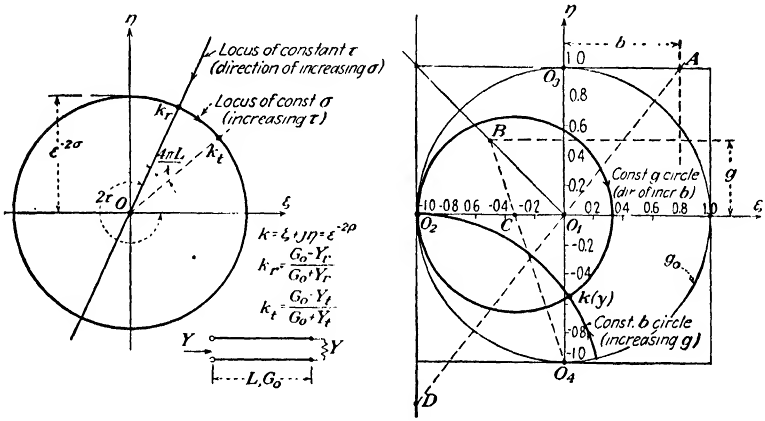


FIG. 47.—*The k-plane diagram.* Left, loci of constant σ and constant τ on the k plane. Right, loci of constant g and constant b on the k plane. k is the reflection coefficient:

$$k = \xi + j\eta = e^{-2\rho} = \frac{1-y}{1+y}.$$

Loci of constant σ, constant τ, constant g and constant b can all be mapped quite readily on the k plane. Such maps, generally spoken of as *hemispheric charts* are in common use and have generally superseded the hyperbolic tangent map (the ρ plane) in high-frequency work.

From the identity

$$k = e^{-2\rho} = e^{-2(\sigma+j\tau)}$$

we see that the loci of constant σ and constant τ on the k plane are also loci of constant $|k|$ and constant $\underline{/k}$, respectively, hence, concentric circles about the origin and straight lines, as illustrated in Fig. 47a. If a load impedance is represented by a value of k, or a point on the k plane, the impedance transformation operated on it by a line of length L is represented by a rotation of this point about the origin, in clockwise direction, by an angle $4\pi L/\lambda$.

For this property to have practical value, we must imagine the z or y function mapped on the k plane, because then it would be easy to go from the k point to the impedance (admittance) or vice versa. Such a map might consist of the two orthogonal families of constant g and constant b loci. We obtain the equations of these loci by the usual method (Secs. 8.4 and 8.7), which consists of the following steps:

1. Expression of k (point of the plane) in terms of y (function to be mapped);
2. Subdivision of the complex equation obtained in (1) into two real equations;
3. Elimination of b and g in turn from the two equations. By definition, we have (step 1)

$$k = \frac{1 - y}{1 + y}$$

In the above, we let as usual

$$y = g + jb$$

and define the coordinates ξ and η of the k plane as follows:

$$k = \xi + j\eta \tag{215}$$

We have, accordingly,

$$\xi + j\eta = \frac{1 - g - jb}{1 + g + jb}$$

Separating real and imaginary parts (step 2), we obtain the system

$$\begin{cases} \xi(1 + g) - b\eta = 1 - g \\ b\xi + \eta(1 + g) = -b \end{cases} \tag{216}$$

Eliminating b from the above (step 3),

$$\xi^2 + \eta^2 + \xi\frac{2g}{1 + g} = 1 - g$$

This is the equation of a circle on the k plane, one of a family of constant g circles. Writing the equation in the form

$$\left(\xi + \frac{g}{1 + g}\right)^2 + \eta^2 = \frac{1}{(1 + g)^2} \tag{217}$$

we see that such circles have center on the real axis and go through the point $\eta = 0$, $\xi = -1$ for all values of g. The circle corresponding to any particular value of g can be constructed geometrically as shown in Fig. 47*b*.

Now, eliminating g from (216),

$$\xi^2 + \eta^2 + 2\xi + 2\frac{\eta}{b} = 0$$

which may be written

$$(\xi + 1)^2 + \left(\eta + \frac{1}{b}\right)^2 = \frac{1}{b^2} \tag{218}$$

indicating that all constant b circles pass through the point $\eta = 0$, $\xi = -1$ and are tangent to the axis of reals at this point. The construction for obtaining such circles is also shown in Fig. 47*b*.

The complete map is shown in Fig. 48. We note that the constant g circle for $g = 0$ has center in the origin and unit radius. This circle bounds the region of the k plane whose points correspond to physical values of admittance (points outside the region corresponding to negative values of g).

Various calculating devices include this map and a radial scale which can be rotated around the center. A circular scale is also provided. A useful arrangement, which the reader can easily assemble, consists of the map (rigidly backed) covered by a circular sheet of tracing paper or other transparent material, fastened at the center so as to permit its rotation. Uniform divisions from 0 to 0.5 should be drawn around the edge of this movable sheet, which will be referred to as the *rotating dial* (Fig. 48).

Such a device is particularly helpful in the analysis of tuners. (Its operation may be visualized easily, providing a concrete basis for thought, even if the device is not physically available.)

Consider, for example, a double-shunt tuner operating between terminations of conductances G_r and G_t. Knowing the characteristic impedance of the coupling section, we may thereby determine the two conductance numbers g_r and g_t and the corresponding constant g circles.

These two circles will be visible through the rotating dial sheet, which the reader will imagine superimposed upon the map of Fig. 48 in such a position that the zero division of the dial falls on the positive real axis (*reference line*) (Fig. 49*a*).

We now take a pencil and make a tracing of the $g = g_r$ circle on the rotating dial sheet; then we rotate this sheet clockwise until the dial reading at the reference line (initially 0) equals L/λ, the length of the coupling section from stub to stub, in wavelengths (Fig. 49*b*)

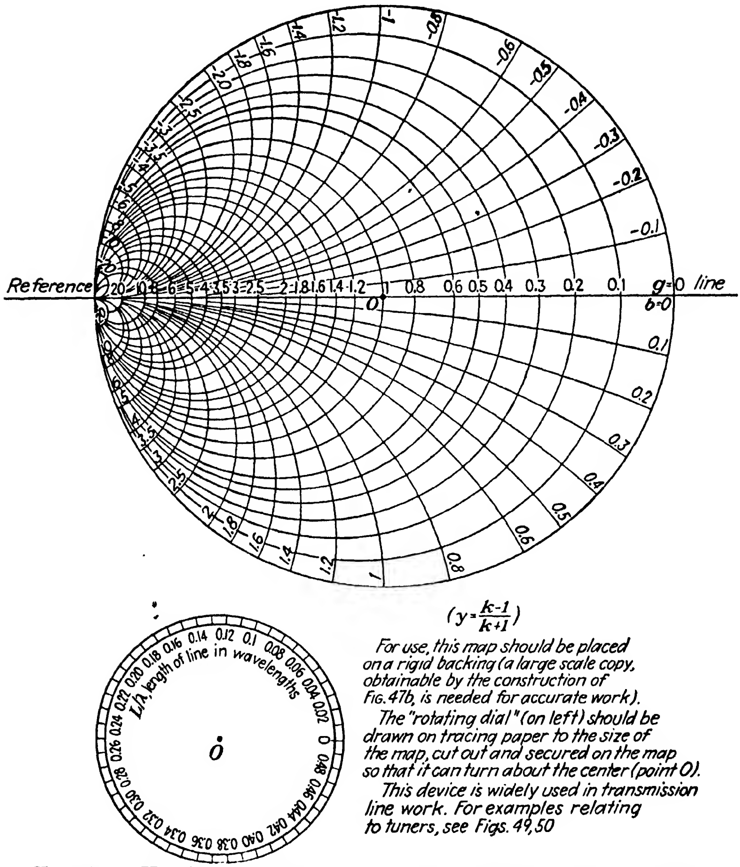
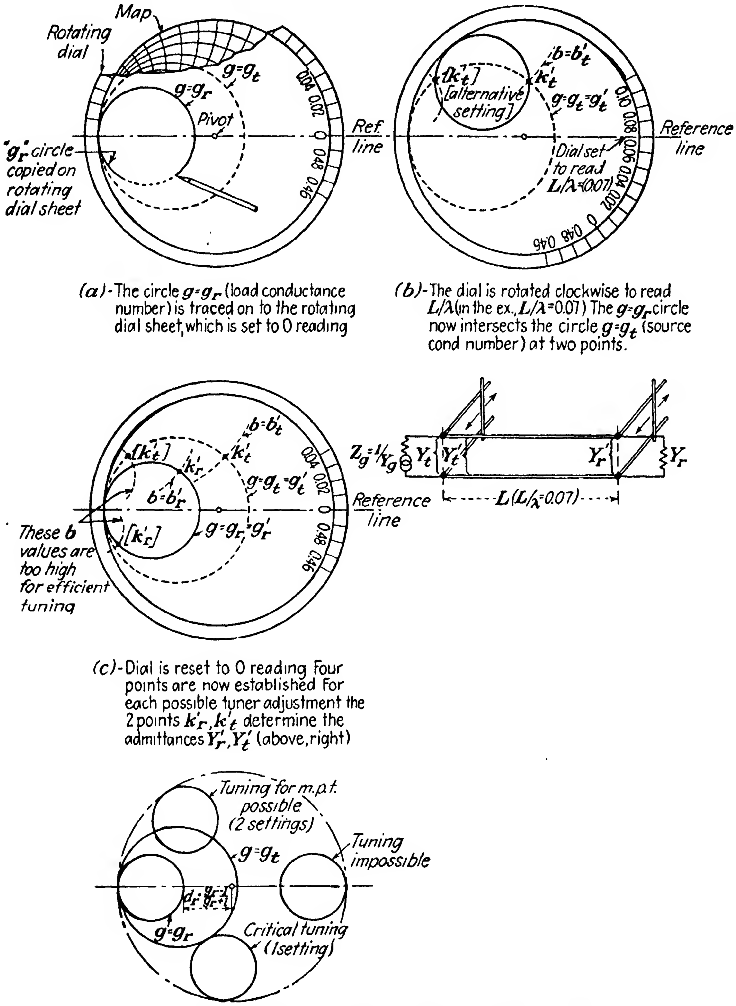


FIG. 48.—*Hemispheric chart (map of the y function on the k plane).*

In this new position, the tracing of the $g = g_r$ circle on the rotating sheet may or may not intersect the $g = g_t$ circle of the map. Assuming that it does, two intersections are located and marked on the rotating sheet, and the corresponding values of g and b are read on the map and recorded.

The moving sheet is now brought back to its original position. The two points marked on it will now designate on the map new values of g and b (Fig. 49c).



(a)-The circle $g=g_r$ (load conductance number) is traced on to the rotating dial sheet, which is set to 0 reading

(b)-The dial is rotated clockwise to read L/λ (in the ex., $L/\lambda=0.07$) The $g=g_r$ circle now intersects the circle $g=g_t$ (source cond number) at two points.

(c)-Dial is reset to 0 reading Four points are now established For each possible tuner adjustment the 2 points k'_r, k'_t determine the admittances Y'_r, Y'_t (above, right)

FIG. 49.—*Graphical analysis of the double-shunt tuner.*

We have, all told, four sets of values for g and b, or four values of admittance. For each intersection of the circles, there are two such values—the load and input admittances of the coupling section when the system is set for maximum power transfer. The two intersections correspond to two distinct settings, both consistent with maximum power transfer.

Consider the values of b obtained for one of the two settings, namely, b_r' and b_t' (Fig. 49c). b_r' is the susceptance number of the load, shunted by the stub; b_t' is the negative of the susceptance number for the source, similarly shunted (for maximum power transfer). From b_r' and b_t', knowing b_r and $b_g = -b_t$ (susceptance numbers for the load and source proper), we can obtain the stub conductances, hence the stub lengths and the stub losses (Sec. 10.9). We should find in this way that some of the settings are such as to make the operation of the tuner very inefficient. This is likely to happen when the two circles intersect very close to the point $k = -1$, or $y = \infty$.

A survey of conditions affecting transmission through the double-shunt tuner would become rather complex because of the number of variables involved. Primarily, however, we are interested in ascertaining under what conditions the device will be capable of tuning for maximum power transfer, without regard to the value of stub losses. Geometrically, we are asking whether or not the two circles ($g = g_r$ and $g = g_t$) will intersect after mutual rotation about the origin by the angle $4\pi L/\lambda$.

The answer depends on the relative values of conductance. Consider, in fact, the minimum distance of a generic constant g circle from the origin, or center of rotation. This is expressed (Fig. 49c) by

$$d = \frac{g - 1}{g + 1} \tag{219}$$

counting d as positive when the origin is outside the circle. If this is true of both the $g = g_r$ and the $g = g_t$ circles, a mutual rotation of these circles about the origin by an angle π (for $L = \lambda/4$) would bring them outside one another, the distance between the nearest points taking the value

$$D = d_r + d_t = \frac{g_r - 1}{g_r + 1} + \frac{g_t - 1}{g_t + 1}$$

It is evident that if D is positive, there will be values of L/λ for which the device cannot be tuned (or more precisely, cannot be tuned for maximum power transfer; there will always be a setting for which the power is highest in a relative sense).

By writing $D = 0$, we may find the *limiting relation* between g_r and g_t

$$g_r g_t = 1 \tag{220}$$

If (220) is satisfied (if the characteristic conductance of coupling section is the geometric mean of source and load conductances), the system can be tuned for all values of L/λ, *except* 0 *and multiples of* 0.5, which must be excluded in all cases, since they require infinite values of susceptance. There will always be two possible settings for m.p.t., except if $L/\lambda = \frac{1}{4}$, when there will be only one setting, the two circles being tangent for this value.

If $g_r g_t < 1$, tuning is also continuously possible (with the exceptions stated above), with the difference that two alternative settings exist at all times, the circles having always two common points.

If $g_r g_t > 1$, suppose L/λ is increased continuously from a small initial value (because of a frequency increase, for example). At first, double tuning will be possible, until L/λ has reached the value

$$\frac{L}{\lambda} = \frac{\cos^{-1}\left(1 - \dfrac{2}{g_r g_t}\right)}{4\pi} \tag{221}$$

For this value the circles will be tangent, as may be verified. Using the terminology of coupled circuits, we might call this condition *critical tuning*. A further increase of L/λ makes tuning impossible, until the critical condition is reached again at the value

$$\frac{L}{\lambda} = \frac{1}{2} - \frac{\cos^{-1}\left(1 - \dfrac{2}{g_r g_t}\right)}{4\pi} \tag{222}$$

The two values (221) and (222) are symmetrically spaced about the value $L/\lambda = \frac{1}{4}$. We conclude that if $g_t g_r > 1$, there are frequency ranges over which the device will not tune, in the sense that, whatever the setting, it will not transmit the maximum possible power. These ranges recur periodically and center about the values f_0, $3f_0$, $5f_0$, . . . where f_0 is the frequency for which $L = \frac{1}{4}\lambda$. The width of the ranges increases with the product $g_r g_t$ and vanishes when this is unity. Again, by analogy with coupled circuits, we might say that the system is *undercoupled* when it cannot be tuned for m.p.t., *overcoupled* when there is double tuning, and, as already mentioned, *critically coupled* when there is only one m.p.t. setting. Considering the frequency as fixed and the length L of the coupling section variable between 0 and $\frac{1}{2}\lambda$, the system is *undercoupled* when

$$\frac{\lambda}{4}\left(1 - \frac{\phi_c}{\pi}\right) < L < \frac{\lambda}{4}\left(1 + \frac{\phi_c}{\pi}\right) \tag{223}$$

where

$$\phi_c = \cos^{-1}\left(\frac{2}{g_r g_t} - 1\right) \tag{224}$$

[*Critical angle for double-shunt tuner*]

may be called the *critical angle.*

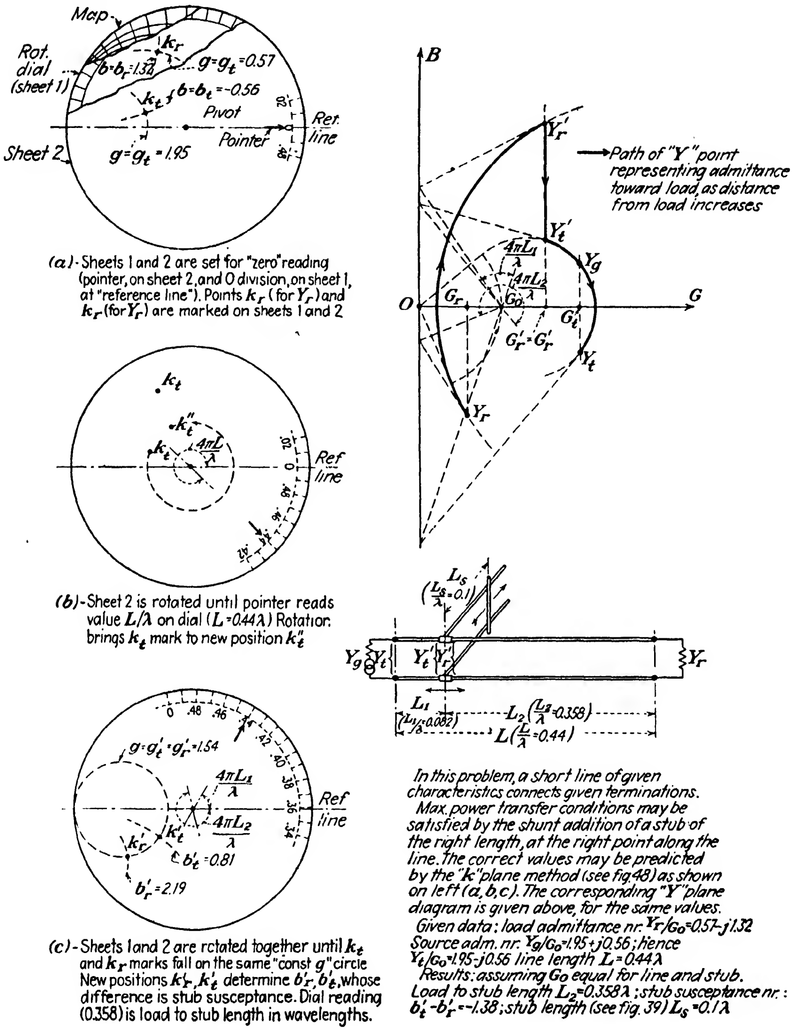
Up to this point, the characteristic impedance of the coupling section has been looked upon as a given quantity. Suppose now that both G_0 and L are open to choice, while G_r and G_t are given, at least approximately. We may wish to find the most favorable values of G_0 and L. We must assume that λ is to vary between given limits and base the discussion upon an "average" value of λ, which we shall call λ_0. For this average operation we would like the stub losses (Sec. 10.9) to be the lowest possible, which will be true, or nearly true, when the susceptance numbers b_r' and b_t' (Fig. 49*c*) are both zero, for which we must have

$$G_r G_t = G_0^2 \qquad L = \frac{\lambda_0}{4}$$

We reach the conclusion that for optimum performance the coupling section of the double-shunt tuner should operate exactly as a quarter-wave transformer section (Sec. 9.3) and be designed accordingly. Under these conditions, the stubs merely serve the purpose of tuning out the susceptance components of source and load.

10.6. Sliding-shunt Tuner. The operation of this device, like that of the double-shunt tuner, may be understood with the help of the Y-plane diagram, although the k plane (Sec. 10.5) is more convenient for a quantitative analysis.

Figure 50 shows the variation of the *admittance toward load* through the sliding-shunt tuner, which consists of a line section of fixed length shunted by a stub (Sec. 10.4) at some intermediate position, which may be adjusted.



(a) - Sheets 1 and 2 are set for "zero" reading (pointer, on sheet 2, and 0 division, on sheet 1, at "reference line"). Points k_r (for Y_r) and k_r (for Y_r) are marked on sheets 1 and 2

(b) - Sheet 2 is rotated until pointer reads value L/λ on dial ($L = 0.44\lambda$) Rotation brings k_t mark to new position k_t''

(c) - Sheets 1 and 2 are rotated together until k_t and k_r marks fall on the same "const g" circle New positions k_r', k_t' determine b_r', b_t', whose difference is stub susceptance. Dial reading (0.358) is load to stub length in wavelengths.

In this problem, a short line of given characteristics connects given terminations. Max. power transfer conditions may be satisfied by the shunt addition of a stub of the right length, at the right point along the line. The correct values may be predicted by the "k" plane method (see fig. 48) as shown on left (a, b, c). The corresponding "Y" plane diagram is given above, for the same values.
Given data: load admittance nr. $Y_r/G_0 = 0.57 - j1.32$ Source adm. nr. $Y_g/G_0 = 1.95 + j0.56$; hence $Y_t/G_0 = 1.95 - j0.56$ line length $L = 0.44\lambda$
Results: assuming G_0 equal for line and stub. Load to stub length $L_2 = 0.358\lambda$; stub susceptance nr.: $b_t' - b_r' = -1.38$; stub length (see fig. 39) $L_s = 0.1\lambda$

FIG. 50.—*Graphical analysis of the sliding-shunt tuner (general case).*

The path of the Y point representing this variation is a constant σ circle as we go from the load to the load side of the stub, a constant G line as we pass from the load side to the source side of the stub, then again a constant σ circle. In fact, the insertion of a reactive line varies only τ (constant σ), and the addition of a shunt susceptance, the stub, varies only B (constant G).

Comparing this diagram with that of Fig. 46, which illustrates the action of the double-shunt tuner, we notice that while previously a constant σ circle joined together two constant G lines, the sequence is now inverted.

The conductance toward load and the conductance toward source must have a common value at the stub, as Fig. 50 clearly indicates. This condition determines one of the two adjustments (the position of the sliding contact of the stub to the line). The other adjustment (length of the stub) is assigned by the second condition for maximum power transfer, expressed in terms of admittances (Sec. 8.2), namely, equal and opposite values of susceptance in the two directions.

The problem of determining whether or not the tuner can be set for maximum power transfer under a given set of conditions is solved by finding the *stub position* along the line and the *stub susceptance* consistent with maximum power transfer.

If we find that the stub position falls on the line proper (between terminations) and the stub susceptance is not excessive, the tuner will work. Among the given conditions we must include the line length L, the characteristic conductance of the line G_0, the load admittance Y_r, the transformed admittance (conjugate of the source admittance) Y_t, and the wavelength λ. (If this is variable, the answer must be obtained for both maximum and minimum λ.) The length L must be short (not large compared to the wavelength), unless one of the terminations matches the line, which is a special case and will be taken up separately.

When Y_r and Y_t have generic values, this problem can be solved exactly only by trial and error. However, with the help of the k-plane map (Sec. 10.5), supplemented by the rotating dial and an additional sheet, also pivoted at the origin, the procedure of trial and error becomes a very simple one. Having located the two points on the map corresponding to Y_r and Y_t, we rotate them *mutually* (relative to one another) by the known angle $4\pi L/\lambda$ (L = length of the coupling line section); then we rotate them *together* until they both fall on the same constant g circle.

The total rotation of each point from the original position, divided by $4\pi/\lambda$, is the distance between the corresponding termination and the stub.

Note that the need for trial and error has not been eliminated, although it has been reduced to the simple operation of moving two points about a third (the three points forming a rigid system) until the two points fall on the same circle. Because there is an immediate visual check, the correct solution is arrived at immediately. The same problem, if carried out analytically, would involve very considerable labor. In such cases the use of trick methods (as, for example, the rotating sheets) is justified. When the analytical solution is straightforward, these methods are not called for.

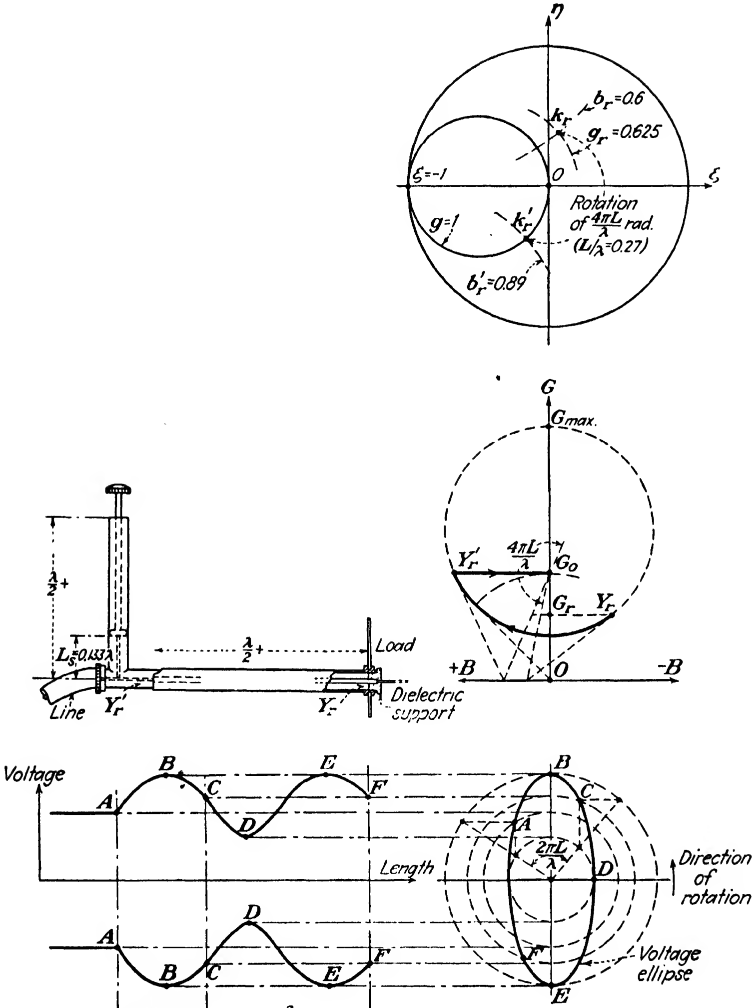
Figure 50 describes the above procedure in detail for a particular case. The Y-plane diagram for the same values is also given.

When the coupling line is matched on one side the problem is very much simpler; the distance between the stub and the matching termination is now immaterial, since over this distance the line is matched and has no impedance transforming action. We need not regard this section of line as part of the tuner. The tuner has the function of matching this line to the nonmatching termination.

Consider, for example, a long line with a matching source and nonmatching load (Fig. 51). In this system, the conductance toward load will vary as we move along the line away from the load. There will be points where this conductance has the matching value G_0, although the susceptance is not zero at these points. If we connect a stub across the line at such a point and adjust its length correctly, this susceptance will be neutralized and, from this point on toward the source, the line will be matched.

The stub position and stub susceptance are easily obtainable by construction, on either the Y or the k plane (Fig. 51). The arrangement commonly used for coaxial cable matching is shown.

The stub position may be found by an experimental procedure if means for measuring the voltage (or current) along the line are available. It can easily be shown that the stub should be placed at a point where the voltage is the *geometric mean* of the maximum and minimum values (Sec. 7.5).



Data of problem: $Y_l/G_0 = y_r = g_r + jb_r = 0.625 - j0.6$ *Results:* L *(distance from stub to load, less* $\lambda/2) = 0.27\lambda$. $-b_r'$ *(stub susceptance nr.)* $= -0.89$ *Hence* $L_s = 0.133\lambda$ *(see fig. 39).*
Other information: r_v *(standing wave ratio)* $= \sqrt{\frac{G_{max}}{G_{min.}}} = 2.41$; $\frac{V_r}{V_0} = \frac{\text{load voltage}}{\text{line voltage}} = \sqrt{\frac{1}{g_r}} = 1.26$

FIG. 51.—*Line-to-load matching by means of the sliding-shunt tuner.*

Consider, in fact, the expression for the power transmitted at any point (junction) of the line

$$P = V^2G$$

where V is the voltage at the junction and G the conductance toward load. P does not vary appreciably over distances along the line comparable with the wavelength; we may, therefore, regard it as a constant and write

$$V^2G = \text{const.}$$

At points where V is maximum (voltage antinodes) the conductance will be a minimum, and vice versa; these extreme values must be related as follows:

$$V_{max}{}^2G_{min} = V_{min}{}^2G_{max} = V^2G$$

where V and G are the values at a generic point. Now consider a point where the voltage is the geometric mean of V_{max} and V_{min}. For such a point we have

$$V = \sqrt{V_{max}V_{min}}$$

Substituting the above in the previous line and writing this as two separate equations,

$$V_{max}V_{min}G = V_{max}{}^2G_{min}$$
$$V_{max}V_{min}G = V_{min}{}^2G_{max}$$

Multiplying the two equations,

$$G^2 = G_{min}G_{max}$$

Now, G_{min} and G_{max} are no other than the terminal admittances (formerly designated by the resistances R_a and R_b, Sec. 9.3) of a quarter-wave section; therefore, we must have

$$G = G_0$$

We conclude that the conductance toward load at the point in question must have the characteristic value, and a stub shunting the line at this point, if properly adjusted, will satisfy the conditions for maximum power transfer.

10.7. Selectivity of the Shunt Tuner. In the preceding sections, high-frequency coupling networks with adjustable elements (tuners) have been discussed, and methods have been worked out for finding whether or not these can be tuned or adjusted for maximum power transfer under given conditions.

To complete the discussion, three additional points should be taken up: *selectivity*, *sharpness of tuning*, and *efficiency*. To avoid excessive complication, only one type of tuning device will be used as the basis of further analysis, the *matching* tuner of Fig. 51. This device is very often used in vhf transmission.

The study of selectivity reduces to the determination of Q_0, as we know (Sec. 9.5). Knowing Q_0, we may obtain the loss in transmitted power L_R due to a departure from midband frequency (185) or use Q_0 as a means of comparison.

Consider a long line coupled to its load by means of a stub (Sec. 10.6), the stub position and stub length being adjusted so that m.p.t. conditions are satisfied for the midband frequency f_0. At this frequency the admittance toward load, Y_t, measured close to the stub on the source side (Fig. 52), must therefore have the matching value G_0. At a generic frequency this admittance will have the value

$$Y_t = Y_r' + Y_s$$

where Y_r' is admittance toward load measured close to the stub on the load side and Y_s the stub admittance. Dividing the above by G_0, we obtain

$$y_t = y_r' + \frac{Y_s}{G_0}$$

where y_t and y_r' are now admittance numbers. y_r' is the input admittance number for the section of length L (Fig. 52), which is given by (140) as follows:

$$y_r' = \frac{y_r + j \tan \phi}{1 + jy_r \tan \phi} \tag{225}$$

In the above, y_r is the load admittance number, ϕ is the line angle

$$\phi = \frac{2\pi L}{\lambda} \tag{226}$$

Y_s is the stub admittance, given by (214)

$$Y_s = -jG_{0s} \cot \phi_s \tag{227}$$

where G_{0s} is the characteristic admittance of the stub line, and

ϕ_s the stub line angle

$$\phi_s = \frac{2\pi L_s}{\lambda} \tag{228}$$

We have, therefore, the following expression for y_t:

$$y_t = y_r' + \frac{Y_s}{G_0} = \frac{y_r + j\tan\phi}{1 + jy_r\tan\phi} - j\frac{G_{0s}}{G_0}\cot\phi_s \tag{229}$$

The value of Q_0 for the tuner is given by (192)

$$Q_0 = \tfrac{1}{2}\left|\frac{dy_t}{dn}\right|_{(n=1)}$$

where $n = f/f_0$ is the *frequency number*. Thus

$$Q_0 = \tfrac{1}{2}\left|j\frac{(1 - y_r y_r')}{\cos^2\phi(1 + jy_r\tan\phi)}\frac{d\phi}{dn} + j\frac{G_{0s}}{G_0\sin^2\phi_s}\frac{d\phi_s}{dn}\right|_{(n=1)}$$

To evaluate the above, note that

$$\phi = \frac{2\pi L}{\lambda} = \frac{2\pi L}{\lambda_0}\frac{\lambda_0}{\lambda} = n\phi_0$$
$$\phi_s = \frac{2\pi L_s}{\lambda} = \frac{2\pi L_s}{\lambda_0}\frac{\lambda_0}{\lambda} = n\phi_{0s}$$

where ϕ_0, ϕ_{0s} are the line angles at midband. Hence

$$\frac{d\phi}{dn} = \phi_0 \qquad \frac{d\phi_s}{dn} = \phi_{0s}$$

Note, in addition, that (225) gives

$$y_r = \frac{y_r' - j\tan\phi}{1 - jy_r'\tan\phi} \tag{230}$$

and that at midband ($n = 1$)

$$y_t = 1$$
$$y_r' = y_t - \frac{Y_s}{G_0} = 1 + j\frac{G_{0s}}{G_0}\cot\phi_{0s} \tag{231}$$

[See Eqs. (227) and (229).]

Substituting (231) into (230), and both (231) and (230) into the expression for Q_0, we obtain, after some algebraic manipulation,

$$
\begin{aligned}
Q_0 &= \tfrac{1}{2}\left| j\phi(1 - y_r'^2) + j\frac{G_{0s}}{G_0}\frac{\phi_s}{\sin^2\phi_s}\right|_{(n-1)} \\
&= \tfrac{1}{2}\left| 2\phi_0 \cot\phi_{0s}\frac{G_{0s}}{G_0} + j\frac{G_{0s}}{G_0}\left(\frac{\phi_{0s}}{\sin^2\phi_{0s}} + \phi_0\frac{G_s}{G_0}\cot^2\phi_{0s}\right)\right| \\
&= \frac{G_{0s}}{G_0}\phi_0\cot^2\phi_{0s}\sqrt{\tan^2\phi_{0s} + \tfrac{1}{4}\left[\frac{G_{0s}}{G_0} + \frac{\phi_{0s}}{\phi_0}(1 + \tan^2\phi_{0s})\right]^2}
\end{aligned}
\tag{232}
$$

[Q_0 *for the shunt tuner*]

The same expression may be arrived at through considerations of plane geometry based on the admittance diagram (Fig. 52) and on Eq. (182).

It should be noted that (232) has been derived on the assumption that the variation of load admittance Y_r with frequency is small compared with the variation of Y_t. This assumption is justifiable in the majority of cases.

Lines of constant Q_0 on a plane of coordinates L/λ and L_s/λ have been plotted (Fig. 52) for the important case of $G_{0s} = G_0$. These lines show immediately how the selectivity depends on the stub setting and position.

It should be borne in mind that these values of Q_0 refer to the whole tuner, comprising stub, line, and load. The value of Q_0 for the stub itself is entirely different (Sec. 10.9). This value should be used if the stub itself is the load, for example, when the stub is used as tank circuit in uhf amplification (Sec. 10.9). The selectivity is much higher in such cases.

The sliding-shunt tuner can be made very selective only by using a low impedance stub ($G_{0s} \gg G_0$). This increases the stub loss, however.

10.8. Sharpness of Tuning. When selectivity is the object of discussion, it must be assumed that the parameters of the selective receiver have fixed values, while the frequency is considered variable.

A parallel analysis considers the frequency as a constant, and one or more parameters of the system as subject to nonsimultaneous variations. These parameters are the adjustments which must be set to some optimum value for maximum power transfer.

The *sharpness of tuning* or *sharpness of adjustment* measures the power loss caused by a variation of any one adjustment from

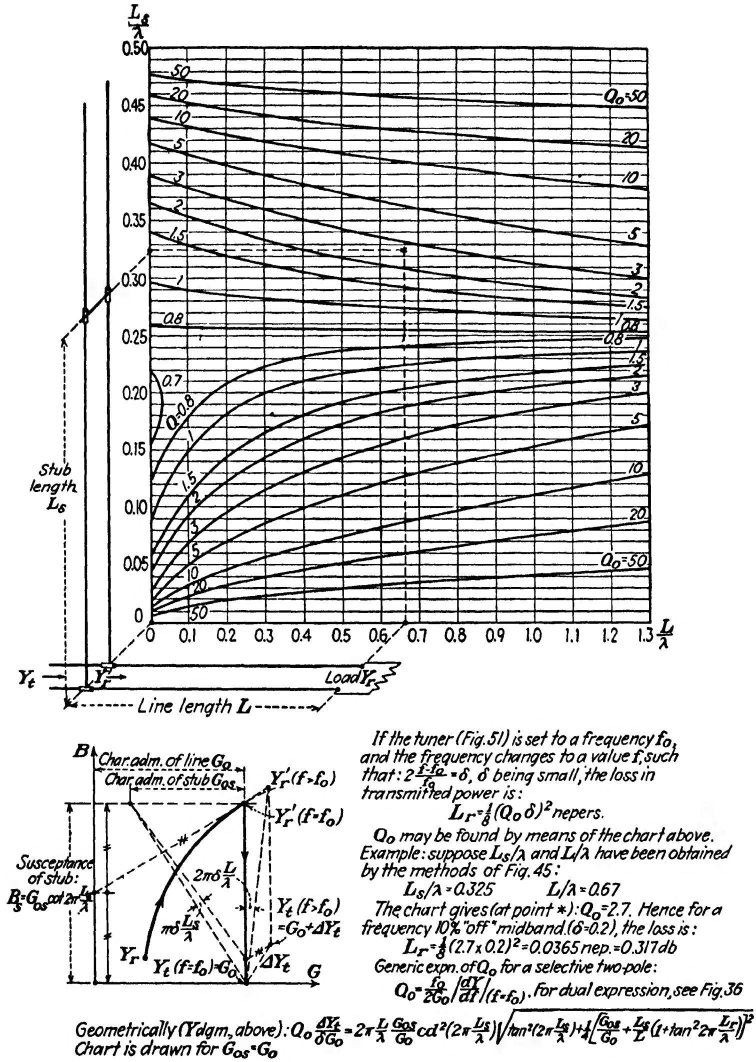


FIG. 52.—*Selectivity of the sliding-shunt tuner.*

the optimum. This evaluation may be carried out, for each adjustment, by means of a coefficient analogous, but not necessarily equal, to Q_0.

Let A stand for one of the adjustable parameters (such as stub length, condenser setting or capacity, mutual inductance, etc.) and A_0 for its optimum value. Under our general assumptions (Sec. 9.5) when all the adjustments have optimum values, the receiver admittance Y is a real. If only A is subject to variation, the other adjustments being set at the optimum, we must have for $A = A_0$

$$y = \frac{Y}{G_0} = 1 \qquad z = \frac{Z}{R_0} = 1$$

Expressing A in ratio form, through the parameter

$$a = \frac{A}{A_0}$$

we may define S_a, *sharpness of the adjustment A*, as follows:

$$S_a = \tfrac{1}{2}\left|\frac{\partial z}{\partial a}\right|_{(a=1)} = \tfrac{1}{2}\left|\frac{\partial y}{\partial a}\right|_{(a=1)} \tag{233}$$

[*Definition of* S_a]

Assuming that maximum power transfer conditions are satisfied for $a = 1$, the power loss due to a variation (detuning) of A is given approximately by

$$L_R = \tfrac{1}{8}(S_a\gamma)^2 \tag{234}$$

[*Power loss due to detuning*]

where

$$\gamma = 2\frac{A - A_0}{A_0} = 2(a - 1) \tag{235}$$

is analogous to the bandwidth δ. S_a may be approximately defined as the reciprocal of the value of γ which causes a power loss of 1 db (see Eq. 186).

Consider, as an illustration, the shunt tuner discussed in the preceding section (Fig. 53). The sharpness of the two adjustments, stub position and stub length, will be evaluated separately. Assuming first that the stub position has the optimum value and the stub length is variable, the *sharpness of stub adjustment* may be defined by the following:

$$S_{Ls} = \frac{L_{0s}}{2}\left|\frac{\partial y_t}{\partial L_s}\right|_{(L_s=L_{0s})} = \frac{\phi_{0s}}{2}\left|\frac{\partial y_t}{\partial \phi_s}\right|_{(\phi_s=\phi_{0s})} \tag{236}$$

where the symbols have the meaning of the preceding section (Sec. 10.7).

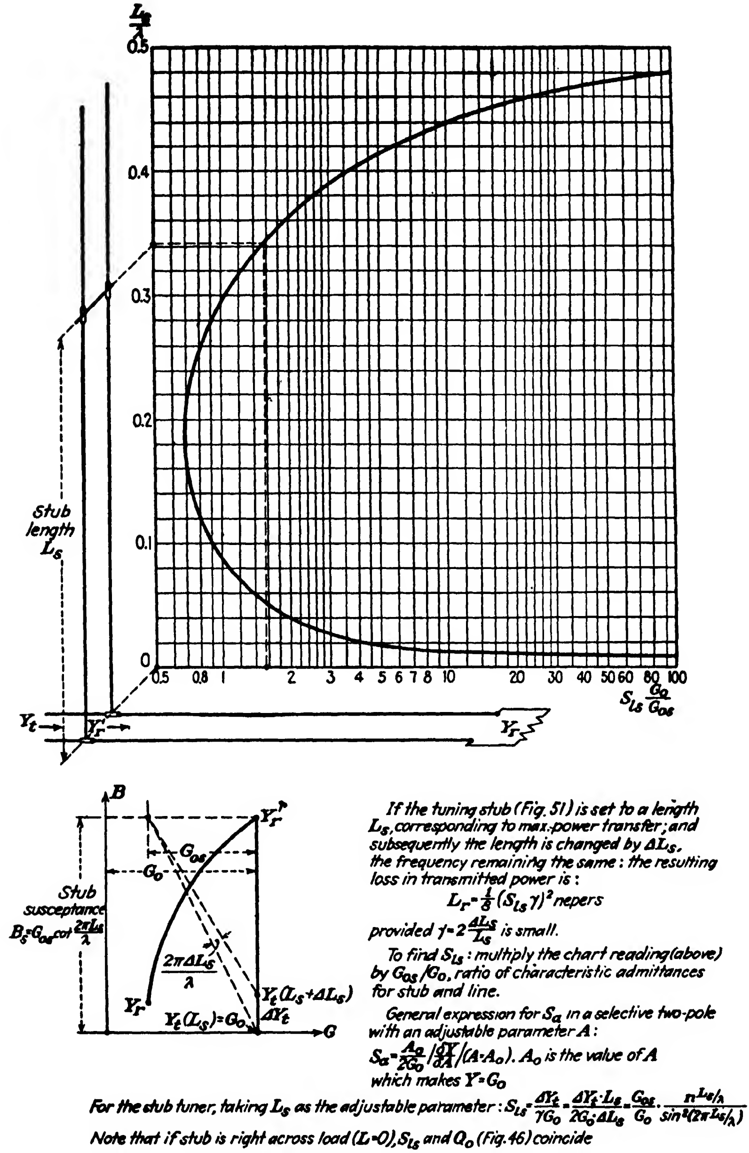


FIG. 53.—*Sharpness of stub tuning.*

Carrying out the differentiation on y_t as expressed in (229), we obtain

$$S_{Ls} = \frac{1}{2}\frac{G_{0s}}{G_0}\frac{\phi_{0s}}{\sin^2\phi_{0s}} \tag{237}$$

[*Sharpness of stub adjustment*]

We find that $S_{Ls} = Q_0$ for $L = 0$ (stub directly across load), showing that when the stub length is the only adjustment, equal relative departures of stub length and frequency from the optimum value have the same effect. This is not always true of systems with only one adjustment. In lumped LC systems, the condenser setting must vary twice as much as the frequency, in a relative sense, for the same drop in voltage. A plot of S_{Ls} against L_s/λ is given in Fig. 53.

To evaluate the *sharpness of slider adjustment,* we must hold the stub length fixed and allow the stub position to vary. The differentiation must now be carried out with respect to ϕ, the stub to load line equal to angle $2\pi L/\lambda$ (Fig. 53), to obtain

$$S_L = \frac{\phi_0}{2}\left|\frac{\partial y_t}{\partial \phi}\right|_{\phi=\phi_0} \tag{238}$$

Accordingly, we find from (229)

$$S_L = \frac{G_{0s}}{G_0}\phi_0 \cot\phi_{0s}\sqrt{1+\left(\frac{G_{0s}\cot\phi_{0s}}{2G_0}\right)^2} \tag{239}$$

[*Sharpness of slider adjustment*]

S_L depends on both ϕ_{0s} and ϕ_0, the optimum stub and line angles. It is proportional to ϕ_0; for a given ϕ_0, it vanishes when $\phi_{0s} = \pi/2$, or when the stub length is $\lambda/4$. Under these conditions, the stub has no effect (except for losses) and could be removed without disturbing the system. The line is matched over its entire length, which can only be true if $Y_R = G_0$.

Example. Let us find the sharpness of the two adjustments in the tuner of Fig. 51 for the values given. Note that for this computation the actual value of L, stub to load distance, must be used, not the value to the nearest half wavelength. Added half wavelengths have no impedance transforming action but affect considerably both selectivity and sharpness.

$$\frac{G_{0s}}{G_0} = 1 \qquad \phi_0 = 2\pi\frac{L_0}{\lambda} = 2\pi(0.5 + 0.27) = 3.58$$

$$\phi_{0s} = 2\pi\frac{L_{0s}}{\lambda} = 2\pi(0.133) = 0.835$$

Hence

$$\cot \phi_{0s} = 0.9057 \qquad \sin \phi_{0s} = 0.7412$$

We have for the *sharpness of stub adjustment*

$$S_{Ls} = \frac{1}{2}\,\frac{0.835}{(0.7412)^2} = 0.76$$

and for the *sharpness of slider adjustment*

$$S_L = 3.58(0.9057)\sqrt{1 + \left(\frac{0.9057}{2}\right)^2} = 3.56$$

The value of Q_0, computed through (232), is $Q_0 = 3.94$.
The same variation in transmitted power would be caused by

A 1 per cent variation in frequency
A 1.1 per cent variation in stub position
A 5.2 per cent variation in stub length

10.9. Q_0 of Stub: Stub Loss and Tuner Efficiency. The concept of efficiency, when applied to coupling devices, must be carefully defined. As an engineering term, *efficiency* generally means the ratio power output/power input (Sec. 6.1); the present case is no exception to this general rule. However, our entire discussion of coupling networks, tuners in particular, has been based on the premise of lossless networks, or 100 per cent efficiency. If this premise does not hold, the conditions for maximum power transfer cannot be met at all junctions of the network simultaneously (Sec. 9.2); we are therefore obliged to distinguish between maximum power transfer *at the source* and maximum power transfer *at the load*. Fortunately, the efficiency is generally so high that this distinction need not be made for practical purposes. We shall therefore continue to assume that, for maximum transmitted power, the input impedance of the coupling network Z_t is the conjugate of the source impedance Z_g.

In particular, with reference to the shunt tuner of Fig. 53, we will continue to assume that, for tuned operation, the input conductance (measured on the source side of the stub) equals G_0, characteristic admittance of the line. If V is the voltage at this point, the input power will be

$$P_I = V^2 G_0$$

A small fraction of this power will be lost in the stub; another smaller fraction, in the portion of line between stub and load; the remainder is the power output. We can, without major error, neglect the loss in the line proper in comparison with the stub loss. As the stub is directly across the voltage V, the stub loss will be given by

$$P_s = V^2 G_s$$

where G_s is the stub conductance. Hence the *efficiency*

$$\eta = \frac{P_l - P_s}{P_l} 100 = 100\left(1 - \frac{G_s}{G_0}\right) \tag{240}$$

Instead of efficiency, the *transmission loss* may be expressed. This is given by

$$L_T = \tfrac{1}{2} \ln \frac{P_l}{P_l - P_s} = \tfrac{1}{2} \frac{G_s}{G_0} \text{ nepers} = 4.34 \frac{G_s}{G_0} \text{ db} \tag{241}$$

The answer to the efficiency question lies therefore in the evaluation of G_s, stub conductance. In first-order approximation, we have considered this to be zero (214), a conclusion based on the simplified expression for the input impedance of reactive lines (140). In the present discussion we are interested in second-order effects; we must therefore go back to the general, exact expression for the input impedance of a short-circuited line and simplify this only after careful consideration.

Equation (102) may be written

$$\begin{aligned} Y_{sc} &= Y_0 \coth \theta L \\ &= Y_0 \frac{\tanh \alpha L(1 + \cot^2 \beta L) - j \cot \beta L(1 - \tanh^2 \alpha L)}{1 + \cot^2 \beta L \tanh^2 \alpha L} \end{aligned}$$

The above notation is general. For stubs in particular, we have been using the symbols

Y_s for Y_{sc} (stub admittance)
L_s for L (stub length)
ϕ_s for βL (stub angle $= 2\pi L_s/\lambda$)
Y_{0s} for Y_0 (characteristic admittance of stub)

We rewrite, accordingly,

$$Y_s = Y_{0s} \frac{\tanh \frac{\alpha}{\beta} \phi_s(1 + \cot^2 \phi_s) - j \cot \phi_s\left(1 - \tanh^2 \frac{\alpha}{\beta} \phi_s\right)}{1 + \cot^2 \phi_s \tanh^2 \frac{\alpha}{\beta} \phi_s} \tag{242}$$

The ratio α/β is small by hypothesis (Sec. 7.1) and has the value (135)

$$\frac{\alpha}{\beta} = \frac{d_s}{2} = \frac{1}{2Q_s}$$

where Q_s has been called the *series* Q of the line and is the limiting value of

$$Q = \left|\frac{X}{R}\right| = \left|\frac{B}{G}\right|$$

for the input impedance of a short-circuited line section (stub) whose length approaches zero (Sec. 3.12). (See Sec. 9.5 on the general definition of Q as opposed to Q_0.)

Since Q_s is always very high (several hundred) and $0 < \phi_s < \pi$ in practical applications, we may assume

$$\tanh \frac{\alpha}{\beta} \phi_s = \tanh \frac{\phi_s}{2Q_s} = \frac{\phi_s}{2Q_s}$$

$$\tanh^2 \frac{\alpha}{\beta} \phi_s \ll 1$$

and simplify Y_s to the following:

$$Y_s = Y_{0s} \frac{(\phi_s/2Q_s)(1 + \cot^2 \phi_s) - j \cot \phi_s}{1 + (\phi_s/2Q_s)^2 \cot^2 \phi_s} \tag{243}$$

Y_{0s} has been considered to be real in reactive lines; its exact value is (39)

$$Y_{0s} = \sqrt{\frac{C}{L}} \sqrt{\frac{d_p + j}{d_s + j}}$$

Because $d_p \ll d_s$ and $d_s = 1/Q_s \ll 1$, the above may be written

$$Y_{0s} = G_{0s}\left(1 + \frac{j}{2Q_s}\right) \tag{244}$$

where

$$G_{0s} = \sqrt{\frac{C}{L}}$$

may be regarded indifferently as the magnitude or the real component of Y_{0s}. Consequently, Y_s may be written

$$Y_s = G_{0s}\left(1 + \frac{j}{2Q_s}\right) \frac{\tan \phi_s + \cot \phi_s - j\dfrac{2Q_s}{\phi_s}}{\dfrac{2Q_s}{\phi_s} \tan \phi_s + \dfrac{\phi_s}{2Q_s} \cot \phi_s} \tag{245}$$

The above expression is accurate over the entire range of ϕ_s. Separate expressions for G_s and B_s may be obtained from it as follows:

$$
\left.
\begin{aligned}
G_s &= G_{0s}\,\frac{\tan\phi_s + \cot\phi_s + \dfrac{1}{\phi_s}}{\dfrac{2Q_s}{\phi_s}\tan\phi_s + \dfrac{\phi_s}{2Q_s}\cot\phi_s} \\
B_s &= G_{0s}\,\frac{\dfrac{1}{2Q_s}(\tan\phi_s + \cot\phi_s) - \dfrac{2Q_s}{\phi_s}}{\dfrac{2Q_s}{\phi_s}\tan\phi_s + \dfrac{\phi_s}{2Q_s}\cot\phi_s}
\end{aligned}
\right\} \tag{246}
$$

[*Components of stub admittance (valid for all values of ϕ_s)*]

The function Y_s is illustrated graphically in Fig. 54.

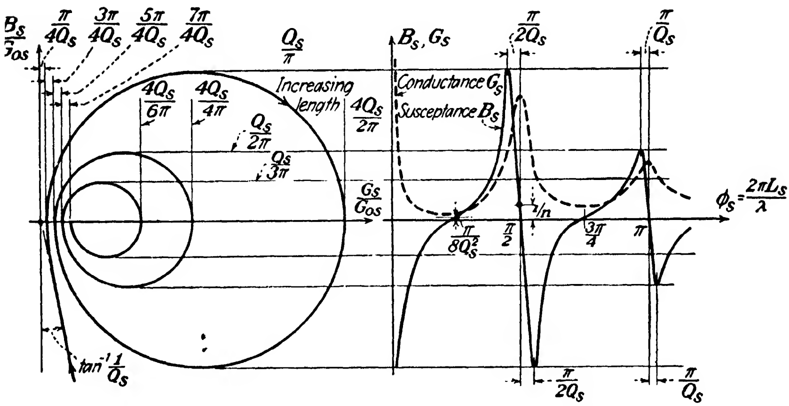


FIG. 54.—*Polar diagram and plots of stub admittance.*

We may verify from the above that Q_s is the limit of the absolute value of B_s/G_s as ϕ_s approaches zero. (The verification is left to the reader.) Another notable fact emerges from (245), namely, Q_s is also equal to Q_0, the selectivity of the stub. (Attention has been called before to the distinction between this and the selectivity of the stub *in combination* with a shunting load.) To prove the point, we will first obtain approximate values for G_s and B_s, valid for *nearly* all values of ϕ_s (except in the neighborhood of $\phi_s = 0$ and $\phi_s = \pi$). These result when we drop all terms of (246) which have Q_s in the denominator. Thus, we have

$$G_s = \frac{G_{0s}}{2Q_s}\left(\frac{\phi_s}{\sin^2 \phi_s} + \cot \phi_s\right) \tag{247}$$
$$B_s = -G_{0s} \cot \phi_s$$

[*Components of stub admittance. Approximate expressions not valid when ϕ_s is close to* 0 *or* $n\pi$]

We see from (247) that Y_s is real for $\phi_s = \pi/2$ and has the value

$$G_s\left(\phi_s = \frac{\pi}{2}\right) = \frac{\pi G_{0s}}{4Q_s} \tag{248}$$

Now, if f_0 is the frequency for which Y_s is a real, we may write

$$n = \frac{f}{f_0} = \frac{\lambda_0}{\lambda} = \frac{\phi_s}{\pi/2}$$

and, according to (191),

$$Q_0 = \tfrac{1}{2}\left|\frac{1}{G_s}\frac{dY_s}{dn}\right|_{(n=1)} = \tfrac{1}{2}\left|\frac{1}{G_s}\frac{dY_s}{d\phi_s}\frac{d\phi_s}{dn}\right|_{(n=1)}$$
$$= \tfrac{1}{2}\left|\frac{\pi/2}{G_s}\frac{dY_s}{d\phi_s}\right|_{(\phi_s=\pi/2)}$$

Substituting the value of G_s for $\phi_s = \pi/2$ given by (248),

$$Q_0 = \frac{Q_s}{G_{0s}}\left|\frac{dY_s}{d\phi_s}\right|_{(\phi_s=\pi/2)} \tag{249}$$

If we differentiate Y_s as given by (247) with respect to ϕ_s and substitute $\pi/2$ for ϕ_s, we obtain jG_{0s} as the result. Hence

$$Q_0 = Q_s|j| = Q_s \tag{250}$$

In Sec. 3.12 a general method of measuring Q_s and Q_p, and hence evaluating the line constants, was indicated. In particular, Q_s may be obtained, according to this method, as the Q of a short-circuited line section of small length compared to λ. This may be done by resonating the section in combination with a capacity with negligible losses and measuring Q_0 for this resonant system. At high frequency, however, this method is not convenient, as the length would have to be exceedingly short. Equation (250) suggests the alternative method, valid when shunt dissipation is negligible, of measuring Q_0 *for the short-circuited line alone*, adjusted to the length $\lambda/4$, which is numerically equal to Q_s and may be used to find the attenuation constant.

Example. To measure the attenuation constant of a coaxial cable at 40 mc., the following arrangement is used: The cable is cut to a length approaching a quarter-wave (1.875 m.) and plugged at one end. The other end is connected to a variable frequency, high impedance source (a pentode, for example). Facilities for accurate readings of frequency and voltage at the input of the cable[11] are provided. Readings are:

Frequency, mc.	Voltage, any unit
$f' = 40{,}430$	$V = 21.3$
f_0 (midband)	$V_0 = 58.2$ (maximum reading)
$f = 40{,}680$	$V = 21.3$

We may consider the shorted cable as a constant conductance two-pole receiving energy from a constant current source (Sec. 9.5). Hence, Eq. (187) may be used. This may be written

$$Q_0 = \frac{1}{\delta}\sqrt{\left(\frac{V_0}{V}\right)^2 - 1}$$

or approximately, in view of the narrow band

$$Q_0 = \frac{f + f'}{2(f - f')}\sqrt{\left(\frac{V_0}{V}\right)^2 - 1}$$

Thus, we have for Q_s

$$Q_s = Q_0 = \frac{81.11}{0.5}\sqrt{7.45 - 1} = 412$$

and for the attenuation.

$$\alpha = \frac{\beta}{2Q_s} = \frac{\pi}{Q_s\lambda} = \frac{\pi}{412 \times 7.5} = 1.035 \times 10^{-3}\text{ nep./m.} = 18.85\text{ db/mile}$$

It can be readily shown that the sharpness of tuning, as defined by Eq. (233), is numerically equal to Q_0, or Q_s, for the stub alone. The above computation may, therefore, be carried out without change if *length readings* (at constant frequency) are available instead of frequency readings at constant length. Such length measurements may present difficulties, however, in view of the high accuracy required. (Frequency readings can be accurate because they depend on the setting of a specially constructed standard.)

Reverting to the chief subject of this section, the determination of shunt tuner efficiency, we may express this on the basis of Eqs. (240) and (247), provided the stub length is not too close to 0 or $\lambda/2$, in which cases (247) is not valid, and the exact equation (246) yields too low a value of efficiency for practical operation.

Accordingly, we have for the efficiency

$$\eta = \left[1 - \frac{2\pi\alpha_s}{\lambda}\frac{G_{0s}}{G_0}\left(\frac{\phi_s}{\sin^2\phi_s} + \cot\phi_s\right)\right]100 \tag{251}$$

[*Efficiency of the shunt tuner*]

and for the transmission loss

$$L_T = \frac{\pi\alpha_s}{\lambda}\frac{G_{0s}}{G_0}\left(\frac{\phi_s}{\sin^2\phi_s} + \cot\phi_s\right) \text{ nepers} \tag{252}$$

where G_0 = characteristic admittance of the line
G_{0s} = characteristic admittance of the stub
$\phi_s = 2\pi L_s/\lambda$ = stub angle when adjusted for m.p.t.
α_s = attenuation of the stub line, nep./m.
λ = wavelength, in meters

Note. $2\pi\alpha_s/\lambda = 1/2Q_s$, where Q_s is the value for the stub line.

Example. Take, as in the preceding example, $Q_s = 412$. Assume the stub angle to be 0.18 radians for m.p.t. The efficiency is

$$\eta = 1 - \frac{1}{2 \times 412}\left(\frac{0.18}{(0.1794)^2} + 5.485\right) = 93.65 \text{ per cent}$$

and the transmission loss

$$L_T = 0.0584 \text{ db}$$

10.10. Illustrative examples.

Effect of dielectric supports on coaxial cable transmission. In many coaxial cables, the inner conductor is kept in position by regularly spaced supports of low-loss dielectric material. Electrically, each support may be considered as a capacity added in shunt to the uniform cable, since substitution of a solid dielectric of constant k for an equal volume of air over a small axial length d of the cable increases the capacity of the cable by the amount

$$C_d = d(k - 1)C$$

which may be considered as a lumped capacity shunting the cable somewhere within the thickness d of the support. The support is assumed to be cylindrical of thickness d; however, for a correct evaluation, an *equivalent* thickness, allowing for edge effects, should be used. This can best be determined by direct measurement. C, in the above, is the distributed capacity of the cable (capacity per unit length).

For a wavelength λ, the shunting admittance due to each support has the value

$$Y_d = j\omega C_d = j\frac{2\pi d}{\lambda}\frac{1}{\sqrt{LC}}(k-1)C = j\frac{2\pi d}{\lambda}\sqrt{\frac{C}{L}}(k-1)$$
$$= j\frac{2\pi d}{\lambda}G_0(k-1)$$

where G_0 is the characteristic admittance of the line. Writing, for brevity, ϕ_d for the angle $2\pi d/\lambda$, y_d for the admittance number Y_d/G_0, the above takes the form

$$y_d = j\phi_d(k-1) \tag{253}$$

Consider (Fig. 55) the variation of *admittance toward load* through the cable in question, assuming the load to have the matching value.[1] This variation may be plotted as the path described by the point denoting the complex quantity

$$y' = \ln Y$$

(The dual quantity z' was used in the analysis of the multisection transformer, Sec. 10.2.) Letting L stand for the length between consecutive supports, hereafter called a *section*, the discussion of Sec. 10.2 has shown that the locus of $z' = \ln Z$ for points of a section is an arc of circle subtending an angle

$$2\phi_L = \frac{4\pi L}{\lambda}$$

with center on the point

$$r_0' = \ln R_0$$

provided Z remains close to R_0. The identical statement may be made with respect to the locus of y', except that the center is

$$g_0' = \ln G_0$$

It follows that, if the load admittance of a section corresponds to a value y_r', the input admittance of the same section will correspond to the value

$$y_i' = g_0' + (y_r' - g_0')e^{-j2\phi_L} \tag{254}$$

We now ask for the path described by the y' point as we go from

[1] By "matching value" is meant that computed from the dimensions without regard to the supports. In practice, when supports are closely spaced, an "average" value of matching impedance is used.

the *load side* to the *source side* of one of the dielectric supports. Again, if y_r' corresponds to the admittance on the load side and y_t' to that on the source side, we will have

$$y_t' = \ln Y_t = \ln (Y_r + Y_d) = \ln Y_r\left(1 + \frac{Y_d}{Y_r}\right)$$
$$= y_r' + \ln\left(1 + \frac{Y_d}{Y_r}\right)$$

Approximately, since $Y_d \ll Y_r$ and $Y_r \approx G_0$,

$$y_t' = y_r' + \frac{Y_d}{G_0}$$

and from (253)

$$y_t' = y_r' + y_d = y_r' + j\phi_d(k - 1) \tag{255}$$

We are now in a position to investigate the path of the y' point as we progress from the load in the direction of the source through a number n of sections. We will assume that the first support is at the load end of the cable and that the load itself has the matching admittance G_0.

We have, accordingly,

At the load:

$$y' = g_0'$$

At the output of the first section:

$$y' = g_0' + j\phi_d(k - 1)$$

At the input of the first section:

$$y' = g_0' + j\phi_d(k - 1)e^{-j2\phi_L}$$

At the output of the second section:

$$y' = g_0' + j\phi_d(1 + e^{-j2\phi_L})(k - 1)$$

At the input of the second section:

$$y' = g_0' + j\phi_d(e^{-j2\phi_L} + e^{-j4\phi_L})(k - 1)$$

At the output of the third section:

$$y' = g_0' + j\phi_d(1 + e^{-j2\phi_L} + e^{-j4\phi_L})(k - 1)$$

At the input of the third section:

$$y' = g_0' + j\phi_d(e^{-j2\phi_L} + e^{-j4\phi_L} + e^{-j6\phi_L})(k - 1)$$

. . . .

At the input of the nth section:

$$y' = g_0' + j\phi_d(k - 1) \sum_{m=1}^{n} e^{-j2m\phi_L}$$

The last branch of the path (locus of the y' point through the nth section) is circular with radius

$$a_n = \phi_d(k - 1) \left| \sum_{m=1}^{n} e^{-j2m\phi_L} \right|$$

If the nth section continued indefinitely, without further supports, in the direction of the source, the y' point would continue to

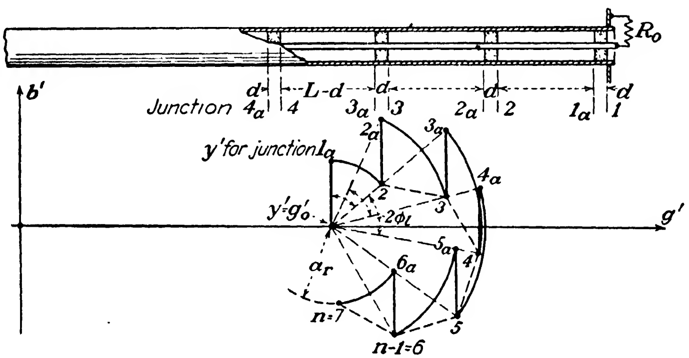


FIG. 55.—*Effect of dielectric supports on coaxial line.*

describe this circle. The maximum and minimum values of conductance toward load in this ideal continuation of the nth section would be related as follows:

$$\ln G_{\max} - \ln G_{\min} = 2a_n$$

hence the *standing wave ratio* is

$$r_{vn} = \sqrt{\frac{G_{\max}}{G_{\min}}} = e^{a_n} = 1 + a_n + \frac{a_n^{\,2}}{2!} + \cdots$$

Because a is small, second and higher order terms of the above may be dropped, and we have approximately

$$r_{vn} = 1 + \phi_d(k - 1) \left| \sum_{m=1}^{n} e^{-j2m\phi_L} \right| \tag{256}$$

[*Standing wave ratio due to supports of coaxial cable, n supports away from matching load*]

In the above, ϕ_d is the angle of line $2\pi d/\lambda$ covered by the thickness

d of the supports, assumed cylindrical, and ϕ_L is the angle between supports. The summation indicated in (256) is best obtained graphically as the closing side of an n-sided polygonal line, with unit sides and angle $2\phi_L$ between adjacent sides.

In the important practical case when the distance between supports is small compared to the wavelength, the analysis may be carried on to an interesting conclusion. In this case, the maximum value of the summation, or of the closing side of the polygonal, corresponds to a value of n such that the polygonal approaches a semicircle. For this value the total line angle, covered by the n sections, is given by

$$\sum_{m=1}^{n} 2\phi_L = 2n\phi_L = \pi$$

hence

$$\frac{4\pi nL}{\lambda} = \pi \qquad \text{or} \qquad nL = \frac{\lambda}{4} \tag{257}$$

from which we conclude that the maximum value of standing wave ratio occurs at a distance $\lambda/4$ from the matching load (or at distances $3\lambda/4$, $5\lambda/4$, . . .). As for the maximum value itself, note that the summation of (256) is now approximated by the diameter of a semicircle, in which may be inscribed a polygonal having n sides of unit length. The summation is therefore approximately given by $2n/\pi$, and we have

$$r_{v\,\max} = 1 + \phi_d(k - 1)\,\frac{2n}{\pi}$$

which may be written, from (257)

$$r_{v\,\max} = 1 + \frac{2\pi d}{\lambda}\,\frac{\lambda}{2\pi L}\,(k - 1)$$

or finally

$$r_{v\,\max} = 1 + \frac{d}{L}\,(k - 1) \tag{258}$$

[*Maximum value of s.w.r. in a coaxial cable with many supports to the wavelength, for matching load*]

where d = thickness of supports (corrected for edge effects)
L = distance between supports
k = dielectric constant of supports

$r_{v\,\max}$ may be obtained directly from measurements of capacity.

The thickness d has been defined through the equation

$$C_d = d(k - 1)C$$

where C_d is the added capacity due to the support, C the capacity of the unsupported cable per unit length. Suppose we measure the capacity of a length L of cable (distance between supports) with and without one of the supports. (Both measurements will actually be obtained by difference, eliminating the effect of additional supports necessary for the measurement.) In this way, we would obtain the ratio

$$\gamma = \frac{CL + C_d}{CL} = 1 + \frac{C_d}{CL} = 1 + \frac{d}{L}(k - 1)$$

which, by comparison with (258), is seen to be equal to the maximum standing wave ratio. In conclusion, *the maximum s.w.r. in a matched cable with many supports to the wavelength is approximately equal to the ratio of capacities measured on a convenient length of the cable, with and without the supports.*

Setting of the shunt tuner for resistive load matching. A graphical solution for the values of stub and line lengths of a shunt tuner matching a line to a generic load has been given (Fig. 51). If the load is resistive, the analytical solution is sufficiently simple and straightforward to warrant its use in preference to graphical methods. We must have in this case

$$y_t = \frac{g_r + j\tan\phi_L}{1 + jg_r\tan\phi_L} + j\frac{G_{0s}}{G_0}\cot\phi_s = 1$$

where $y_t = Y_t/G_0$ is the transformed admittance number which must equal unity to match the line; $g_r = G_r/G_0 = R_0/R_r$ is the conductance number for the load; G_0 and G_{0s} are the characteristic admittances of line and stub, respectively; and ϕ_L and ϕ_s are the line angles for line and stub. Solving the real part of the above, we have

$$\tan\phi_L = \sqrt{\frac{1}{g_r}} = \sqrt{\frac{R_r}{R_0}} \tag{259}$$

Using the above value of $\tan\phi_L$ and solving the imaginary part,

$$\cot\phi_s = \frac{R_{0s}}{R_0}\left(\sqrt{\frac{R_r}{R_0}} - \sqrt{\frac{R_0}{R_r}}\right) \tag{260}$$

Note that $\sqrt{R_r/R_0}$ is the ratio of currents. The above may be used to compute the selectivity, sharpness of tuning, and efficiency. A numerical example is given below.

Example. Suppose the tuner is used to match a 77-ohm cable to a 10,000-ohm load. Assume that the stub and cable have the same characteristic impedance. We have

$$\sqrt{\frac{R_r}{R_0}} = 11.4 \qquad \cot \phi_s = 11.4 - \frac{1}{11.4} = 10.522$$
$$\phi_s = 0.0943$$

Equation (251) for the efficiency gives in this case

$$\tfrac{1}{100}\eta = 1 - \frac{1}{2\phi_s}\{0.0943[1 - (10.522)^2] + 10.522\}$$

Taking the same value of Q_s as in the example of Sec. 10.9 ($Q_s = 412$),

$$\eta = 97.44 \text{ per cent}$$

Note. When $\cot \phi_s$ is a large number, the efficiency may be written more simply

$$\eta = \left\{1 - \frac{\cot \phi_s}{Q_s}\right\} 100 \tag{261}$$

and for resistive loads

$$\eta = \left\{1 - \frac{\sqrt{R_r/R_0} - \sqrt{R_0/R_r}}{Q_s}\right\} 100 \tag{262}$$

The efficiency has the same value if the direction of transmission is reversed.

Lumped equivalent of the line stub. Stub shunted by capacity. Stubs are used in place of lumped LC tank circuits in uhf amplification. The characteristics of the amplifier stage, particularly the gain, depend on the total value of plate circuit admittance, which in this case is the stub admittance added to the input admittance of the following stage. When the system is tuned this total value reduces to a conductance, and this determines the maximum gain. In the following, the problem will be simplified by considering the stub shunted by a capacity C (for the value of C to be used in gain computations, see Fig. 56*b* and Sec. 16.6). The conductance of the combination is therefore the stub conductance, given over the useful range by Eq. (247). This is a function of ϕ_s, the stub angle. When the system is tuned, this angle will differ from the value $\pi/2$, which would produce resonance without the shunting capacity. Hence, the plate circuit conductance for maximum gain and the maximum value of gain itself depend on the value of C and on the frequency. This dependence will now be investigated.

For an exact solution, we would have to solve a transcendental equation, which would require graphical methods. We will compromise by expanding cot ϕ_s, which appears in the stub admittance formula, into an algebraic expression; by so doing,

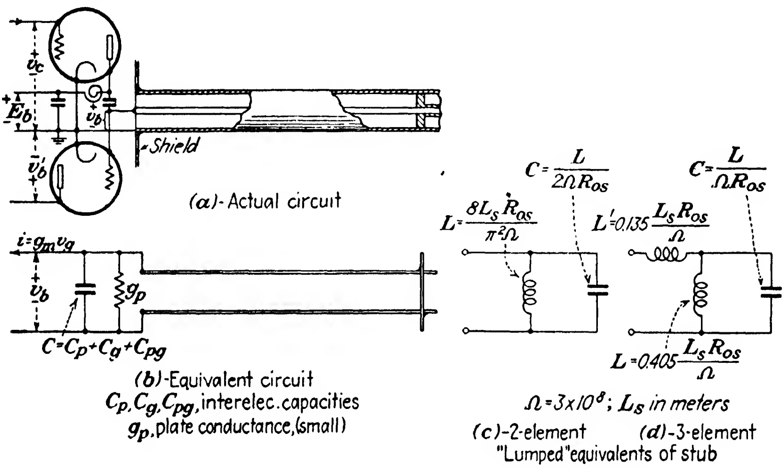


FIG. 56.—*Concentric line used as coupling impedance.*

we will assimilate the stub to a parallel LC combination and treat it as such. We have for the susceptance of the stub (247)

$$B_s = -G_{0s} \cot \phi_s$$

and for that of a parallel LC system

$$B_s' = \sqrt{\frac{C}{L}}\left(\omega\sqrt{LC} - \frac{1}{\omega\sqrt{LC}}\right) \tag{263}$$

For an approximate equivalence of the two expressions over a useful frequency range, we must satisfy two conditions; namely,

1. The two functions must vanish for the same ω.
2. The two functions must have the same derivative at the zero point.

By imposing these conditions, we may obtain values of L and C which may be used to represent the stub in problems where only the frequency is variable. This does not, however, solve the present problem, nor any problem where adjustments of stub length are contemplated. What we need is an expression for B in terms of G_{0s} and ϕ_s but *having the form* of (263). To obtain this, observe that to satisfy the first condition we must have simultaneously

$$\phi_s = \frac{\pi}{2}$$

$$\omega = \frac{1}{\sqrt{LC}}$$

To these we must add the general relation

$$\phi_s = \frac{2\pi L_s}{\lambda} = \frac{\omega L_s}{\Omega}$$

where Ω is the wave velocity in free space (Sec. 3.6). Combining, we have

$$\sqrt{LC} = \frac{1}{\omega} = \frac{L_s}{\phi_s \Omega} = \frac{2L_s}{\pi \Omega} \tag{264}$$

which is one of two equations needed to find the equivalent values of L and C. Furthermore, we find

$$\omega \sqrt{LC} = \omega \frac{2L_s}{\pi \Omega} = \frac{2}{\pi}\left(\frac{\omega L_s}{\Omega}\right) = \frac{2\phi_s}{\pi}$$

which enables us to write (263) in the form

$$B_s' = \sqrt{\frac{C}{L}}\left(\frac{2\phi_s}{\pi} - \frac{\pi}{2\phi_s}\right) \tag{265}$$

We have not yet made use of condition (2). Differentiating the original function (247) we obtain

$$\frac{dB_s}{d\phi_{s\,(B_s=0)}} = \left[\frac{G_{0s}}{\sin^2 \phi_s}\right]_{\left(\phi_s=\frac{\pi}{2}\right)} = G_{0s}$$

and differentiating (265),

$$\frac{dB_s'}{d\phi_{s\,(B_s'=0)}} = \left[\sqrt{\frac{C}{L}}\left(\frac{2}{\pi} + \frac{\pi}{2\phi_s^2}\right)\right]_{\left(\phi_s=\frac{\pi}{2}\right)} = \frac{4}{\pi}\sqrt{\frac{C}{L}}$$

Now, equating the two derivatives,

$$\sqrt{\frac{C}{L}} = \frac{\pi}{4} G_{0s} \tag{266}$$

The above, combined with (264), gives the equivalent values of

L and C shown in Fig. 56c. Substituted in (265), it gives

$$B_s' = \frac{\pi G_{0s}}{4}\left(\frac{2\phi_s}{\pi} - \frac{\pi}{2\phi_s}\right) \tag{267}$$

Comparing the above with (247), we find that by assimilating the stub to a lumped system we have made the approximation

$$\cot \phi_s \approx \frac{\pi}{4}\left(\frac{\pi}{2\phi_s} - \frac{2\phi_s}{\pi}\right) \tag{268}$$

The same approximation enables us to consider in place of G_s, the stub conductance (247), the expression

$$\begin{aligned} G_s' &= \frac{G_{0s}}{8Q_s\phi_s}\left[\phi_s^4 + \phi_s^2\left(2 - \frac{\pi^2}{2}\right) + \frac{\pi^2}{2}\left(1 + \frac{\pi^2}{8}\right)\right] \\ &= \frac{G_{0s}}{8Q_s\phi_s}\left[\phi_s^4 - 2.94\phi_s^2 + 11\right] \end{aligned} \tag{269}$$

We must now write the tuning equation in terms of B_s' and $C\omega$, solve it for ϕ_s, and substitute the solution in G_s'. The tuning equation is

$$B_s' + C\omega = 0$$

Writing in the expression (267) for B_s' and solving for ϕ_s, we obtain the positive root

$$\phi_s = \frac{\pi}{2}\sqrt{1 + \frac{4}{\pi^2}R_{0s}^2C^2\omega^2} - R_{0s}C\omega \tag{270}$$

It is best to solve the above numerically for any specific case, and substitute the result into (269). An example follows.

Example. Consider a stub of 77-ohm characteristic impedance, with $Q_s = 412$ (as in preceding examples). To be computed, the conductance of the stub, when tuned across a 10 $\mu\mu$f capacity, at 200 mc.

We have

$$R_0C\omega = 77 \times 2\pi \times 200 \times 10 \times 10^{-6} = 0.96$$

$$\phi_s = \frac{\pi}{2}\sqrt{1 + \left(\frac{2}{\pi} \times .96\right)^2} - .96 = 0.88$$

Hence

$$G_s' = \frac{10^6}{77 \times 8 \times 0.88 \times 412}\left[(0.88)^4 - 2.94 \times (0.88)^2 + 11\right] = 41.9\ \mu\text{mhos}$$

For a comparison, let us find the conductance of the stub when tuned alone. This is (for $\phi_s = \pi/2$)

$$G_s = \frac{\pi}{4}\frac{G_{0s}}{G_s} = 24.95\ \mu\text{mhos}$$

The shunting capacity is therefore responsible for a 40.5 per cent drop in gain (assuming a constant current source).

Recalling the difficulty, inherent to the use of lumped tank circuits, of obtaining "high Q" coils with inductance small enough to tune out the input susceptance of a tube at very high frequency, we note that this difficulty is not eliminated by the use of distributed systems, such as the stub. Although quantitatively much better results are obtained, the trend is the same.

Suggested Exercise. A more exact equivalent of the stub than that of Fig. 56*c* is the three element two-pole of Fig. 56*d*. The equivalence is based on the coincidence of the first zero and the first pole (barring the origin) of the two admittance functions, and of the derivatives at the first zero. Check the values of the elements given in Fig. 56*d*, which were obtained on this basis.

CHAPTER XI

A BRIEF REVIEW OF ELECTROMAGNETIC THEORY. STATIC FIELDS

11.1. Introduction. At the start of Chap. I the distinction between *field*, *circuit*, and *network* problems was briefly introduced as a preliminary to the study of four-terminal networks. The solution of both network and circuit problems depends on certain basic assumptions that only field theory can justify. In a sense, therefore, every network or circuit problem is a field problem to begin with. A logical presentation should start out with field analysis and continue on to circuits and networks. This book is not designed to cover that much ground and assumes some knowledge of field theory as well as an understanding of alternating currents and complex algebra.

It is expected, however, that some readers, while well acquainted with certain aspects of field theory, may not have a comprehensive picture embracing field, circuit, and network theory as parts of a whole.

If one takes network assumptions and definitions for granted, one can go a very long way without such a picture. Transmission line theory, for example, has been based so far on these assumptions entirely. Eventually, however, these assumptions must be questioned to see just what problems can or cannot be treated in this way. If such a discussion were presented at the start, much of its significance would be lost.

These are the reasons for undertaking a review of field theory at this point. The object is to trace out a continuous thread from the general laws of electricity which relate to the field vectors to those particular laws (having the form of voltage-current relations) which justify the treatment of circuit and line problems as it has been carried out up to this point. There will be an effort to achieve this limited objective with the simplest possible means, without loss of rigor or generality. Hence, only the *integral form* of the field equations will be cited. The only concept that the reader must have clearly in his mind from the outset is that of a vector field. It is anticipated, of course, that many readers will find this review unnecessary and will pass on to subsequent chapters.

11.2. Electric Current and Electric Field Intensity. In defining the fundamental electrical quantities, we must remember that they are known to us only through the observation of phenomena or actions which we ascribe to them (force actions, visible discharges, etc.). If we confine ourselves to the study of currents within conductors (as opposed to space currents), we may consider the *charge* or *quantity* of positive electricity as the cause of all observable phenomena. Of these, force actions are particularly important as they constitute the link between electrical and mechanical quantities. We need not inquire into the nature or even the existence of the positive charge, provided we assume it to be so distributed as to give rise to the existing force actions. The positive charge is correctly thought of as the end point of a *line of force*, and nothing more. It is helpful, however, to think of it as a fluid (to use an old-fashioned word) which is distributed *continuously*, although seldom *uniformly*, within material media, where it may flow or remain stationary.

Consider a small particle of matter upon which lies a charge Q. Imagine it placed at a point P of a region including a system of charged bodies. There will be a force $\mathbf{F}$ on the particle; this will depend on its position in space (the coordinates of P) and on the amount of charge Q. It will not depend appreciably on the distribution of Q over the particle; as this is of small extent. Now, the limit

$$\mathbf{E} = \lim_{Q \to 0} \frac{\mathbf{F}}{Q} \tag{271}$$

[*Definition of* **E**]

has a definite value depending on the coordinates of P; this is a *vector* value, or better still, because it varies from point to point, a *vector point function*, generally known as the *electric field intensity*, or *electric force vector*.

Next, take a system of conductors through which charge is flowing. We shall fix our attention now not on a *point* of the system, but on a *surface*, plane or otherwise, and not necessarily extending over a conductor cross section. Through this surface a certain charge Q will have passed during a time interval τ ending at time t. The limit

$$i = \lim_{\tau \to 0} \frac{Q}{\tau} \tag{272}$$

[*Definition of* i]

is the *current* through the surface at time t. This is not a vector, but a scalar quantity; and not a point function, but a surface function, having a particular value for every surface in space. Alternatively, i may be defined as the time derivative of a time function $Q(t)$ equal to the total amount of charge having traversed the surface from some initial time t_0 to the time t, ($i = dQ/dt$). The current i through a given surface may or may not vary with time.

11.3. Current Density. The Surface Integral. Because the current is a surface function, its value must be obtainable from that of a suitable *point function* integrated over the surface. As a parallel, consider the rate of flow of water in a river; if we know the velocity of water at all points of a cross section of the river bed, we may arrive at the rate of flow. Most methods of flow measurement are based on a measurement of velocity; the alternative consists in measuring the amount of flow over a known time interval, which corresponds to Eq. (272).

If ρ is the *charge density* and $\mathbf{U}$ its velocity of flow, we may write

$$\mathbf{J} = \rho \mathbf{U} \tag{273}$$

[*General*]

Both $\mathbf{J}$ and $\mathbf{U}$ are vector point functions. $\mathbf{J}$ is called the *current density,* or *current flow vector;* if $\mathbf{J}$ is known for all points of a surface S, then we may evaluate the current through S by the *surface integral*

$$i = \int\int_S J_n \, dS \tag{274}$$

[*General*]

In the above, J_n stands for the *component of* $\mathbf{J}$ *along the positive normal* to dS (Fig. 57).

The importance of thoroughly understanding surface (and line) integrals cannot be overemphasized. It is not necessary in most cases to acquire practice in the computation of such integrals, which is very involved in any except the simplest cases. What one should possess is a mental picture, for example, of the *surface integral* or *flow through a surface* in relation to the *vector* which represents the flow at any point, which may be velocity of flow, flux density, current density, stream vector, etc. Figure 57 may be found helpful to this end.

From the form of Eq. (274) we note that, if the current density **J** is known throughout a region, we may evaluate the current

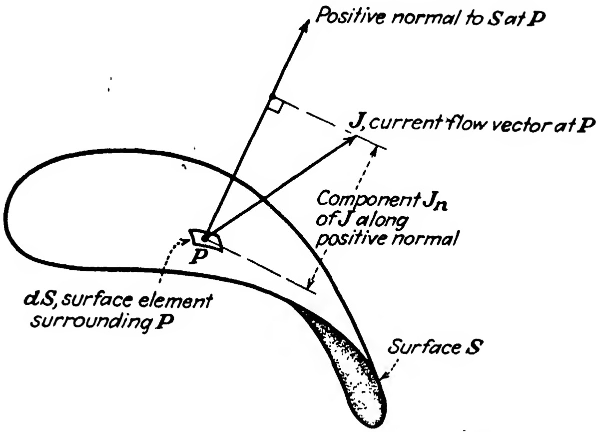


FIG. 57.—*Current through a surface* $\left(i = \int\int_S J_n dS\right)$. *Illustrates surface integral.*

through a surface of the region, provided *we assign a positive direction* to the normal at some point of the surface. Intuition tells us that the positive direction of all the other normals to the surface is determined in consequence. If we do not assign the positive normal, there is no way of telling in which direction the surface is traversed by the current i, since i is a scalar. If the positive normal is assigned, we may conclude that the current flows in the direction of the positive normal if $i > 0$, and vice versa.

When we put an arrow next to a wire in a schematic, this is intended to show not the direction of the current but the direction of the *positive normal* to the cross section at which current is measured. The current flows with the arrow if positive, against it if negative.

We have so far introduced three vector point functions: **E**, **U**, and **J**. The velocity **U** has little physical meaning except for free charges, and we shall not refer to it further. The electric force vector **E** and the current flow vector **J** are important because they lead directly to the concepts of current (274) and of voltage, as the following will show.

11.4. Voltage. The Line Integral. Ohm's Law. The definition of voltage that follows is a general one, like that given for current (4). No restrictions are imposed, not even those that are taken for granted by engineers in most applications. For this reason it may sound unfamiliar. We must first state that voltage is a *line function*, in the sense that it has a value for any particular open line or path that we may trace in a region. This line may be partly or totally within conductors. We might add more precisely that the *direction of the positive tangent* to the path must also be assigned before we say that a

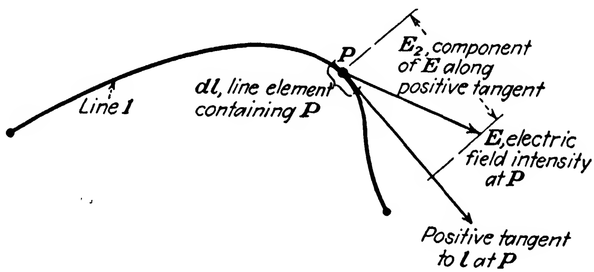


FIG. 58.—*Voltage over a path* $\left(v = \int_l E_l\, dl\right)$. *Illustrates line integral.*

certain voltage is positive or negative. Having assigned a path l and its positive direction, the corresponding voltage is

$$v = \int_l E_l\, dl \tag{275}$$

[*Definition of v*]

where E_l is the component of **E**, the field intensity, along the positive tangent (Fig. 58).

The integral of (275) is called the *line integral* of **E** over l. While the surface integral has to do with the idea of *flow*, the line integral is associated with *work done*. The work W done by the force **F** of the field, in moving a charge Q over a path l, may be written

$$W = \int_l F_l\, dl$$

Dividing by Q and letting Q approach zero,

$$\lim_{Q\to 0} \frac{W}{Q} = \int_l \frac{F_l}{Q}\,dl = \int_l E_l\,dl = v \tag{276}$$

Hence, we have the following definition of voltage: *The voltage over l is the work done by the force of the field in moving a charge Q along l in the positive direction, divided by Q, for Q approaching zero;* or, alternatively, *the voltage over l is the work done by an outside force (electromotive force) in moving a charge Q along l in the negative direction, divided by Q, for Q approaching zero.*

The arrow between terminals (Fig. 1 and others) is intended to show the direction of the outside force, or electromotive force, which would ideally perform the work W of Eq. (276). This arrow is therefore opposite to the positive direction of l. When the voltage is positive at the terminals, W must be positive, hence E and the arrow must be in opposite directions.

As the current is the integral of $\mathbf{J}$ over a surface (4), and the voltage is the integral of $\mathbf{E}$ over a line, we would expect a relation between $\mathbf{E}$ and $\mathbf{J}$ to yield an expression of voltage in terms of current, at least in a particular case. Such a relation exists *for points of a homogeneous solid conductor.* It is simply

$$\mathbf{J} = \gamma \mathbf{E} \tag{277}$$

[*Valid within homogeneous solid conductors*]

where γ is a material constant, the *conductivity*. As an application of (277), consider (Fig. 61*a*) a section of conductor; assume that the current through it is the same for all cross sections at any time and that $\mathbf{J}$ is uniform over each cross section. Then we have for the current

$$i = |\mathbf{J}|S$$

where S is the area of a normal cross section. For the voltage *taken along a current streamline*[1]

$$v = \int_l E_l\,dl = \int_l \frac{|\mathbf{J}|}{\gamma}\,dl = i\int_l \frac{dl}{\gamma S}$$

Now, letting,

$$G = \frac{1}{\int_l dl/\gamma S}$$

[1] A *streamline* or *flow line* is a line whose tangent at every point indicates the direction of the flow vector (in this case, $\mathbf{J}$) at the point. In the case of a force vector, such as E, such a line would be called a *force line.*

we have the *voltage-current relation*

$$i = Gv \tag{278}$$

The above (*Ohm's* law) is valid for conductors when the current is the same in all cross sections and *provided the voltage is measured along a streamline of current.* The restrictions may be removed when **J** and **E** are time constants at all points of the region.

11.5. Definition of a Static Region. Electrostatic Potential. Magnetic Field Intensity. If, within a region of space, **E** and **J** are constant with time at all points, we will refer to it as a *static region;* we will call the vector fields of **E** and **J** in the region *static fields,* in spite of the fact that there will be, in general, a flow of electric charge in the region. The study of static fields is wider than the study of *electrostatics,* if this term is taken to mean that there are no moving charges (no currents). It is essentially the study of direct current systems. A revision of the concepts of static fields becomes necessary when the field vectors, **E** and **J**, change with time; but as long as these vectors are time constants, such a revision is not necessary.

In a static region **E** and **J** obey certain laws, which can be expressed in several ways. We will call them the *laws of static fields;* they are actually no other than the well-known *Kirchhoff* laws, expressed in a general form, as follows:

First law of static fields: *The voltage over a closed path within a static region is zero,* or

$$\oint_l E_l \, dl = 0 \tag{279}$$

[*Static fields only*]

where l is any closed path or loop within the region. The integral

$$\oint_l E_l \, dl$$

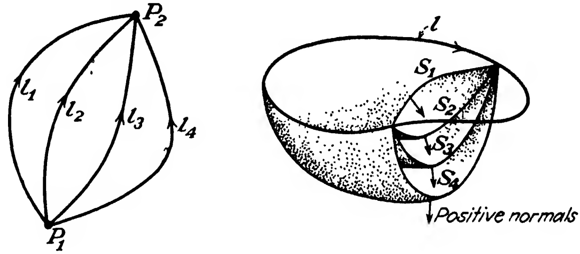
is often called *circuitation* of **E** over l.

Second law of static fields: *The current through a closed surface (surface totally enclosing a volume) within a static region is zero,* or

$$\oint\!\!\oint_S J_n \, dS = 0 \tag{280}$$

[*Static fields only*]

As an important corollary of the first law, we have the following: *The voltages over all the paths having the same end points are* *equal in a static region.* This is illustrated in Fig. 59*a*; the reader may easily verify it by considering two of the paths shown. They form a loop, over which the voltage must be zero (279). Clearly, then, in a static region the voltage is no longer a *line function;* it is a *point function,* in the sense that for each pair of points taken in a definite order there exists one and only one voltage. We may now talk without ambiguity of the *voltage between two points;* in general this is not possible, although in



(a)-In a static region:

$$\oint_{(l)} E_l\, dl = 0$$

Hence the voltages over all paths l_1 l_2 l_3.... with common end points are equal

(b)-In a static region:

$$\iint_{(S)} J_n\, ds = 0$$

Hence the currents through all surfaces bounded by the same contour l are equal. Surfaces are partially shown

FIG. 59.—*Geometric significance of Kirchhoff's laws, generally valid only for static fields.*

many cases this expression, while lacking precision, has an accepted practical significance.

Consider, then, the voltages between each of a set of points P_1, P_2, P_3, . . . and a *reference point* P_0. We may associate a value of voltage to each point of the set, always taking P_0 as the second point. In fact, we may do this for all the points of the region; in this way, voltage becomes a scalar function of the point coordinates, or a *scalar point function.* This is called the *electrostatic potential*

$$U = -\int_{P_0}^{P} E_l\, dl$$

A static field of **E** may be represented in two ways: by mapping the force vector at every point, or by mapping the potential function. The second method is easier, because the potential function, a scalar, is denoted by a single number. A set of equipotential surfaces, each corresponding to a value of the potential, defines tle field. The reader is probably on familiar ground here; if not, he should consult the literature.[9]

The second law of static fields (280) also has an important corollary, as follows: *The currents through all the surfaces having a common boundary are equal in a static region.* This corollary is illustrated in Fig. 59*b*. It is an immediate consequence of (280) and suggests that, in a static region, the current is not a *surface function* but a *loop function;* one and only one value of current traverses every closed path (loop) of the region. It is the current *linked* with the loop. To grasp what this means, imagine the streamlines of current to have substance. If we thread a loop through them, we will catch a definite number within the loop. But this can only be true if each streamline is itself a loop. If we say that two lines *link* each other, the phrase has definite meaning only if both lines are closed. We conclude that, within a static region, the streamlines of current are closed; they do not terminate anywhere. Under such conditions the vector **J** is said to be *solenoidal* (tubular). We can draw further conclusions from the solenoidal quality of **J** in a static region. If the current is a loop function, then it should be possible to obtain its value by the *integration of a point function* over the loop. A new point function has to be invented for this purpose, just as the potential function was introduced when it became possible to associate voltage with a point. Let us call this new function **H** (known as *magnetic field intensity*). It is impossible to define **H** as we defined **E**, as such a definition presupposes a force measurement, and the means for this are lacking where **H** is concerned. An incomplete definition of **H** is implicit in the following:

$$\oint_l H_l \, dl = i \tag{281}$$

[*Static fields only*]

known as *Ampère's* circuitation law. The definition is incomplete because there are many fields of **H** for which (281) would be satisfied; this ambiguity disappears, however, if we further state that **H** must obey (283) at every point, as will be shown.

For Eq. (281) to have exact meaning, we must specify the positive direction of the circuitation. There is a general rule which relates the *positive normal* to a surface with the *positive tangent* to its contour (Fig. 60). The rule is easily remembered if we think of a disklike surface; looking in the direction of the normal, the tangent points clockwise.

The reader will not fail to see a certain similarity between the expressions of voltage in terms of **E** (275) and current in terms of **H** (281). Both voltage and current are given as line integrals. However, current may be so expressed only when it is constant with time, while for voltage the expression is general. Furthermore, for current the line integral must always be taken around a closed loop. Finally, there is this difference, that **E** has

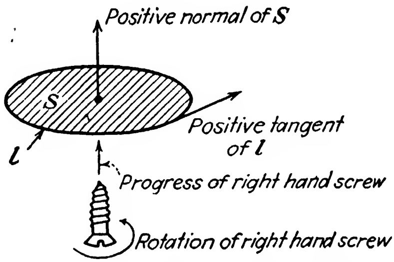


FIG. 60.—*Conventional relation between the positive normal to a surface and the positive tangent to its contour.*

definite meaning apart from voltage, which is a derived quantity; **H,** on the contrary, is an abstraction predicated on current.

11.6. Electric and Magnetic Flux Densities. As the presence of an electric charge is manifested by a force to which the charge (or the charged body) is subject, the presence of a current is revealed by a force which acts upon the current (or the conductor carrying it). It is possible to imagine experimental means for measuring the force acting on a very short and thin *current element* in the presence of other currents. These experiments would show that the force varies in direction and intensity depending on the orientation of the current element; we cannot, therefore, arrive at a vector point function similar to **E** (271) simply by dividing this force by the current and letting this approach zero. However, suppose we let the current approach zero by making the cross section dS of the element smaller and smaller, keeping the current density **J** in the element constant in value and direction and uniform over the cross section. Then we could find the value of

$$\lim_{(dS\to 0)} \frac{\mathbf{F}}{dS\,dl}$$

where dl is the element length, assumed small ($dS\,dl$ is the element of volume). If we computed the above limit, which is a vector quantity, for all possible orientations of the element and current densities, we would find that it satisfies the vector equation

$$\lim \frac{\mathbf{F}}{dS\,dl} = \mathbf{J} \times \mathbf{B} \tag{282}$$

[*Definition of* **B**]

where **B** is the same in all cases, independent of **J**. Interpreting the cross-product notation, the limit of force over volume for small volumes is *normal* to the plane of **J** and of a vector **B**, function only of the position of the current element in space. The value of the limit is equal to the area of the parallelogram constructed on **J** and **B**.

Equation (282) may be taken as the definition of **B**, the *magnetic flux density* or *induction vector*. The definition is complete, although it would take more than one experiment to determine **B** at any point; several orientations of the current element would be necessary.

The value of **B** at any point of a region depends on the distribution of the currents in the region. This dependence is expressed by the equation

$$\mathbf{B} = \mu \mathbf{H} \tag{283}$$

[*General*]

where μ (the *permeability*) is a scalar quantity which depends on the medium, and also, when the medium is ferromagnetic, on **B** itself. Equation (283) is not sufficient to determine **B** from a given distribution of currents, because **H** is not uniquely defined (only the *circuitation* of **H** around any particular loop is assigned by the current distribution in a static region). To (283) must be added a general experimental law relating to **B**, namely, that **B** is solenoidal everywhere under all conditions, or

$$\oint\!\!\oint_S B_n\,dS = 0 \tag{284}$$

[*General*]

Equations (281), (282), and (283) together permit us to find **B** around a given current distribution. **H** may be regarded merely as a step in this procedure. Nevertheless, (283) and (284) jointly determine the vector **H** at all points, and therefore the **H** field is uniquely defined.

There is a formal similarity between the induction vector **B** and the current density in a static region. Both are solenoidal. Hence, in a static region, a definite current links each closed line of magnetic flux, and vice versa. Hence, also, the *magnetic flux through a surface*, a scalar, defined by

$$\Phi = \int\int_S B_n \, dS \tag{285}$$

[*Definition of* Φ]

depends, like the current through a surface in a static region, only on the contour or loop l bounding the surface. It follows that we can invent a new vector point function such that its line integration over a contour gives the magnetic flux through the contour. This new function bears exactly the same relation to **B** as **H** does to **J**. It is called the *magnetic vector potential*, **A**, defined by the equation

$$\Phi = \oint_l A_l \, dl \tag{286}$$

supplemented by the condition that **A** must be *solenoidal* in a static region.[1] Like the electrostatic potential, the vector potential is useful in field-mapping problems.

So far nothing has been said about the electric force vector beyond defining it (171) and stating that its loop integral is zero in a static region (179). There is also, as we would expect, a general law relating **E** to the charges in the region. This relation, like that between **B** and the currents of the region, may be expressed by introducing a new vector point function, serving as a link between charge and **E**, just as **H** serves as a link between current and **B**. Let us call this new vector **D**, the *electric displacement vector* or *electric flux density*. This will be defined by expressing its relation to **E** on the one hand and to electric

[1] Under dynamic conditions **A** is defined as follows:

$$\Phi = \oint_l A_l \, dl \qquad \oint\!\!\oint_S A_n \, dS = -\int\int\int_\tau \mu\epsilon \frac{\partial U}{\partial t} \, d\tau$$

or, vector analytically,

$$\begin{cases} \mathbf{B} = \operatorname{curl} \mathbf{A} \\ -\mu\epsilon \dfrac{\partial U}{\partial t} = \operatorname{div} \mathbf{A} \end{cases}$$

for the significance of U, See Sec. 12.3.

charge on the other. The relations are

$$\mathbf{D} = \epsilon\mathbf{E} \tag{287}$$
[*General*]

$$Q = \oint\!\!\oint_S D_n\,dS \tag{288}$$
[*General*]

In (287), ϵ is a material constant, the *permittivity;* in (288) Q is the charge contained within the closed surface S. Equation (288) is the well-known law of *Gauss*. The *electric flux* through a surface is defined by

$$\Psi = \int\!\!\int_S D_n\,dS \tag{289}$$
[*Definition of* Ψ]

11.7. Inductance. Consider the three equations

$$\mathbf{J} = \gamma\mathbf{E} \tag{277}$$
$$\mathbf{B} = \mu\mathbf{H} \tag{283}$$
$$\mathbf{D} = \epsilon\mathbf{E} \tag{287}$$

We have seen how (277) leads to a relation between the *surface integral* of $\mathbf{J}$ (current) and the *line integral* of $\mathbf{E}$ (voltage) within conductors (Ohm's law). Similar relations may be obtained from (283) and (287). In each case, however, there has to be a unique value for both the line and the surface integral within a certain region.

For example, in the case of a section of conductor (Fig. 61*a*) terminated by equipotential cross sections, there is a unique value for the voltage across the section and the current in the section. We can therefore define (if not compute) the ratio

$$G = \frac{i}{v}$$

or the *conductance* of the section and show that, because of (277), G is constant.

Consider now (283). The surface integral of $\mathbf{B}$ is the *magnetic flux* Φ (285) and the line integral of $\mathbf{H}$, taken over a loop, is the current through the loop, but only in a static region (281). Imagine (Fig. 62*a*) that all the magnetic flux of a region is channeled within a doughnut-shaped core and all the current flow lines are channeled within a ring of wire linked with the core.

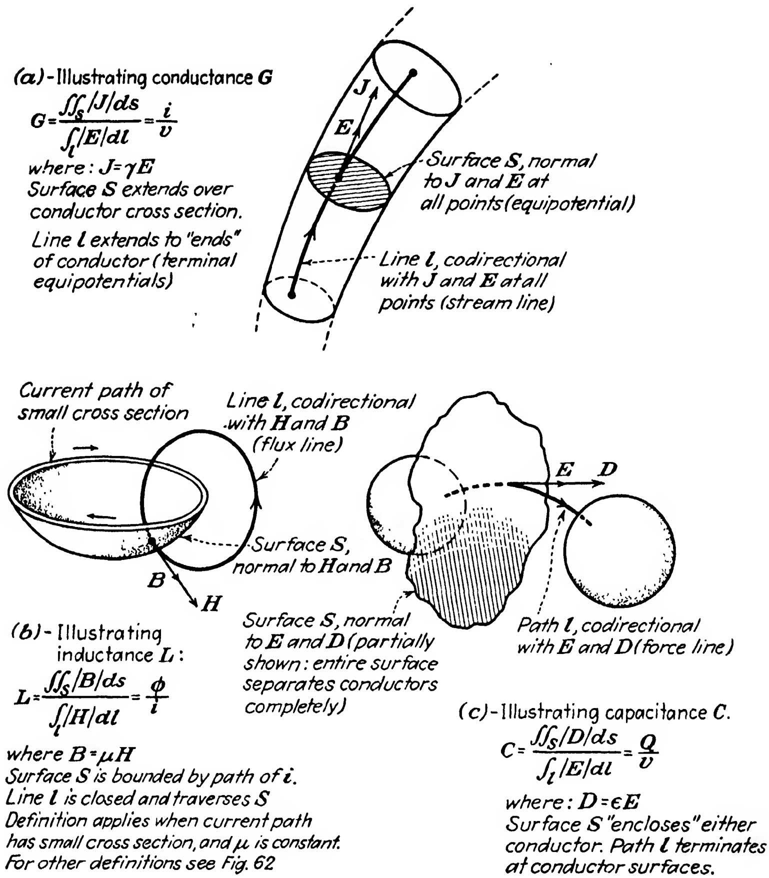


FIG. 61.—*Conductance, inductance, capacity.*

Under these conditions, taking the loop integral of **H** along a line of magnetic flux (inside the core)

$$i = \int_l H_l\, dl = \int_l |\mathbf{H}|\ dl = \int_l \frac{|\mathbf{B}|}{\mu}\, dl$$

Assuming (for simplicity) **B** uniform in the core, and letting S stand for its normal cross section,

$$\Phi = |\mathbf{B}|S$$

Hence

$$i = \int \frac{\Phi}{\mu S}\,dl$$

But as the **B** field is solenoidal, Φ must the the same at all cross sections and can be taken out of the integral sign

$$i = \Phi \int \frac{dl}{\mu S}$$

which can be written

$$\Phi = Li \tag{290}$$

having let

$$L = \frac{\Phi}{i} = \frac{1}{\int dl/\mu S} \tag{291}$$

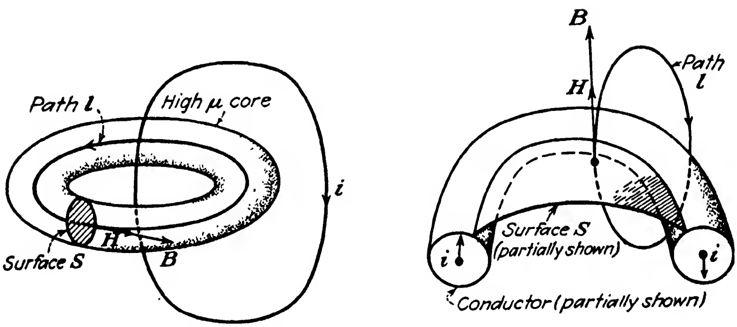


FIG. 62.—*Definition of inductance in special cases.*

L, the inductance, is constant if μ is constant. This, unfortunately, is not true of ferromagnetic materials, for which an *equivalent* inductance may be defined for specific purposes.(9)

Even assuming μ constant, the significance of L is restricted. The example given is a particularly simple one, selected for its similarity with that of Fig. 61*a*.

Inductance could also be rigorously defined if the current made several turns around the core; all the flux lines in this case would still have the same value of linked current (current crossing a surface whose contour is on the flux line). When the magnetic flux is everywhere in space instead of being confined within the core, the inductance can be defined only if the current flows within a wire of negligible cross section (Fig. 61*b*). Suppose, in fact, the current to flow in a sizable conductor (Fig. 62*b*) penetrated by flux. Some flux lines would, in this event, link more current than others. There would be a range of values for $\oint_l H_l \, dl$ and the general definition[1]

$$L = \frac{\Phi}{i} = \frac{\iint_S B_n \, dS}{\oint_l H_l \, dl} \tag{292}$$

[*Definition of inductance*]

would not lead to a unique value. In such cases one can define the *external inductance* by means of (292), with the understanding that the surface S has its boundary on the surface of the conductor, and that this boundary is the loop l (Fig. 62*b*). Thus defined, the external inductance is not rigorously single valued, but very nearly so. Another definition of inductance is based on the energy associated with the flow of current; this definition(9) is more general but not useful for our limited purpose.

In any case, inductance ceases to have exact meaning when $\mathbf{J}$ is not solenoidal everywhere. If the current flow lines are open, each loop l can no longer be associated with a unique value of current. It follows (Sec. 11.5) that inductance has exact meaning only in a static region. This will come as a surprise to many, who may be accustomed to the use of inductance in a-c problems. Actually, the inductance, computed or measured for static (d-c) conditions, may be considered approximately equal to the ratio Φ/i at all times, up to a value of frequency which depends on the approximation required and on the geometry of the system.

If the geometry is particularly simple, one can define a *distributed inductance* which continues to have exact meaning up to extremely high frequencies. Consider a parallel transmission line. The static field configuration may be worked out by the method shown in Fig. 63 (Sec. 11.10). This configuration, *within an orthogonal cross section*, also obeys the laws of dynamic fields (Maxwell's equations), as we shall see (Sec. 12.4). This does not mean that other dynamic configurations are not possible. It can be shown, however, that these do not occur except when the wavelength is comparable with the lateral dimensions.(10)

[1] The absolute values of **B** and **H** may be used instead of the components B_n, H_l, if l is a flux and S cuts all flux lines orthogonally (Fig. 61).

Assuming, then, the configuration of Fig. 67, we may define the distributed inductance as follows:

$$l = \lim_{(\Delta z \to 0)} \frac{\Delta\Phi}{i\,\Delta z} \tag{293}$$

[*Definition of distributed inductance*]

where i is the current through either wire at a given cross section, Δz the thickness of a slab including this cross section, and $\Delta\Phi$ the external flux in the slab (Fig. 67). The quantity $\Delta\Phi/\Delta z$ may be interpreted as a *linear flux density*. l is a constant because this linear density is proportional to the current all along the line.

The distributed inductance defined in this way (more exactly called the *external* distributed inductance, see Fig. 62*b*) varies slightly with the current distribution within the wire, hence with frequency. The reader is referred to the extensive literature on this subject.[9] For the inductance coefficient relating to systems of several currents, see Sec. 14.3.

11.8. Capacity. Going back to the derivation of Ohm's law (278), if, instead of a section of conductor, we consider a section of dielectric limited by two equipotentials, which presupposes static fields (Sec. 11.5), we arrive at a relation between the electric flux Ψ and the voltage v similar to Ohm's law. The only difference is that $\mathbf{D}$ is taken in place of $\mathbf{J}$, ϵ in place of γ.

Consider the dielectric of a condenser in which $\mathbf{D}$ is assumed uniform over each equipotential. We have for the voltage taken over a line of $\mathbf{D}$ between the end equipotentials (the condenser plates)

$$v = \int_l |\mathbf{E}|\,dl$$

For the flux we have

$$\Psi = |\mathbf{D}|S = \epsilon|\mathbf{E}|S$$

hence

$$v = \Psi \int_l \frac{dl}{\epsilon S}$$

Letting

$$C = \frac{1}{\int_l dl/eS}$$

we may write the above in the form

$$\Psi = Cv \tag{294}$$

If we define the capacity by the above equation, as the ratio of electric flux and voltage, we associate this parameter with a portion of the dielectric, traversed by a given Ψ or adjacent to a given portion of the conductors. In the absence of such qualifications, it is understood that capacity is the ratio between the *entire electric flux* and the voltage of the two-conductor system (Fig. 61*c*), or, from Gauss's law, Eq. (287),

$$C = \frac{Q}{v} = \frac{\oint\!\!\oint_S D_n\, dS}{\int_l E_l\, dl} \tag{295}$$

[*Definition of capacity*]

The expressions of capacity and inductance (291) and (295) are dual; however, to obtain inductance, we divide an integral taken over an *open surface* by an integral taken over a *closed line;* for capacity, an integral over a *closed surface* is divided by an integral over an *open line.*

Capacity, like inductance, has no exact meaning outside a static region, although it may be used with good approximation for slowly varying fields, particularly when the spacing is close. Nor can capacity be defined if there is electric charge distributed in the space between conductors. In this case various surfaces enclosing each conductor would also enclose different amounts of space charge and would be traversed by different amounts of electric flux (Gauss's law).

The concept of capacitance, like that of inductance, can be extended to include dynamic conditions if the geometry permits. The *distributed capacity* of a transmission line is an example. This parameter is similar in all respects to distributed inductance. Its definition is (Fig. 67)

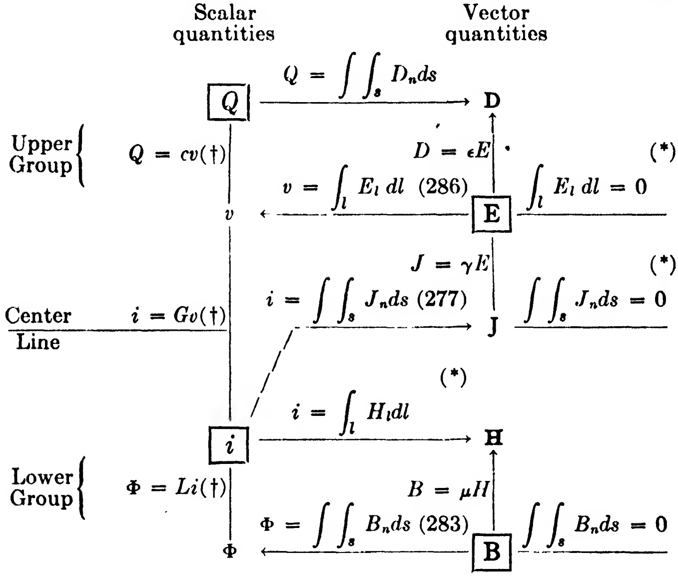
$$C = \lim_{(\Delta Z \to 0)} \frac{\Delta\Psi}{v\, \Delta Z} \tag{296}$$

[*Definition of distributed capacity*]

Partial capacities and capacity coefficients relating to systems of several conductors will be discussed in Sec. 15.1.

11.9. Recapitulation. The Unit System. It is advisable at this point to cast a bird's-eye glance over the ground already covered before proceding further. We may do this with the help of the following table, which brings together the quantities of electromagnetism and their mutual relations. Only those relations involving the quantities themselves and length (or powers of length) have been included in the tabulation. Relations involving time, or time, length and mass (force or energy), are best considered separately. Of these, some have already appeared as definitions of i, **E**, and **B**. Others (Maxwell's

MUTUAL RELATIONS BETWEEN THE QUANTITIES OF ELECTROMAGNETISM



Notes. Two quantities symmetrically placed about the center line are dual. Two quantities similarly placed in their respective groups are of the same order.

Quantities shown thus: $\boxed{Q}$, serve to define the others. Arrow points to defined quantity.

Relations marked thus: (*) are valid only in static regions.

Relations marked thus: (†) have significance only in particular cases.

For relations involving time (dynamic relations) see Sec. 12.1.

equations) will be taken up later. Many more do not come within the scope of our analysis.

Looking at the tabulation, we observe two groups of four quantities, thus:

$$\begin{array}{cc} Q & \mathbf{D} \\ v & \mathbf{E} \\ \hline i & \mathbf{H} \\ \Phi & \mathbf{B} \end{array}$$

These eight quantities may be paired together in two distinct ways. We may, for example, take each quantity and its dual, which has been arranged symmetrically about the center line. For example, Q and Φ are dual; **D** and **B** are dual. Dual quantities are formally similar; similar relations exist between dual sets (the relation of **D** to Q is similar to that between **B** and Φ). But no general relation exists between a quantity and its dual.

Another way of pairing quantities together is by what may be called their *order*. Each member of the upper group (Q, **D**, v, **E**) occupies a definite order or position within the group, and so does each member of the lower group. Thus, for example, **E** is defined in terms of Q and force (271). Likewise, **B** is defined in terms of i and force (282). **D** is defined in terms of Q by the law of Gauss (287); the definition is not unique, however, and has to be supplemented by a relation with **E**. Similarly, **H** is defined in terms of i by Ampere's circuital law, but not uniquely; a relation with **B** must be added.

If we move the upper group downward until Q coincides with i, each quantity will coincide with the quantity of the same order. In this way, **D** may be paired with **H**, **E** with **B**, v with Φ, and Q with i.

This second type of correspondence (by order) differs from the first (by duality) in that there are general relations between terms of the same order, and they are most important. These relations involve time and will be taken up in the next chapter.

For the sake of continuity, no mention of units has been made so far. If the equations which we have taken as definitions are to be free of numerical constants, which is highly desirable, the definitions themselves must serve to establish the dimensions of the defined quantities, hence their units. The system of units thus evolved, starting from the international units of length, mass, and time (meter, kilogram, and second), and the coulomb, is called the *rationalized m.k.s.* or *Giorgi* system. The following table gives for each quantity the equation serving to define it (by number), and, accordingly, the dimensions and the units. Dimensions and units will be found to check if one bears in mind the dimensions of the *derived* units (ampere, volt, weber, henry, farad).

DIMENSIONS AND UNITS

Symbol	Name	Defined by equation	Dimensions as defined	Units
Q	Electric charge	——	Q	Coulombs
D	Electric displacement	287	$L^{-2}Q$	Coulombs/sq. m.
i	Current	272	$T^{-1}Q$	Amp.
J	Current density	274	$L^{-2}T^{-1}Q$	Amp./sq. m.
H	Magnetic field intensity	281	$L^{-1}T^{-1}Q$	Amp./m.
v	Voltage	275	$L^2T^{-2}MQ^{-1}$	Volts
E	Electric field intensity	271	$LT^{-2}MQ^{-1}$	Volts/m.
Φ	Magnetic flux	285	$L^2T^{-1}MQ^{-1}$	Webers
B	Magnetic induction	282	$T^{-1}MQ^{-1}$	Webers/sq. m.
G	Conductance	278	$L^{-2}TM^{-1}Q^2$	Mhos
γ	Conductivity	277	$L^{-3}TM^{-1}Q^2$	Mhos/m.
L	Inductance	289	L^2MQ^{-2}	Henry
μ	Permeability	283	LMQ^{-2}	Henry/m.
C	Capacity	294	$L^{-2}T^2M^{-1}Q^2$	Farads
ϵ	Permittivity	286	$L^{-3}T^2M^{-1}Q^2$	Farads/m.

Values of ϵ and μ in free space:

$$\epsilon_0 = 8.854 \times 10^{-12} \text{ farad/m.}$$
$$\mu_0 = 1.257 \times 10^{-16} \text{ henry/m.}$$
$$\frac{1}{\sqrt{\epsilon_0\mu_0}} = 2.998 \times 10^8 \text{ m./sec.} \qquad \text{(velocity of light)}$$

★11.10. Application to Plane Fields. Static Fields of the Parallel Line. As an illustration and elaboration on the concepts of this chapter and for purposes of reference, we shall consider the static fields surrounding a parallel line (Fig. 63*a*). The line will be assumed to have infinite length; consequently, the static field distribution is the same in all orthogonal cross sections, of which Fig. 63*a* represents one; the results will apply, in practice, to all cross sections of a long line at a distance from the terminations large compared to the lateral dimensions.

Within the orthogonal cross section, all point functions may be expressed in terms of only two coordinates, x and y, instead of three. The problem of mapping the field reduces therefore to a two-dimensional or plane problem; fields of this type are often called plane fields.

Consider (Fig. 63*a*) a point P on the xy plane, and a line l from P to 0, center of symmetry. Assume P to be outside the conductors. We may associate with point P four distinct integral functions, namely

(a) $$-\int_l E_l\,dl = U(x, y) \tag{297}$$

This is the *electrostatic potential* at P, equal to the voltage (Sec. 11.4) between P and 0; point 0 is chosen arbitrarily as reference point. U is single valued for all possible paths from P to 0.

(b) $$\int_l E_n\,dl = U'(x, y) \tag{298}$$

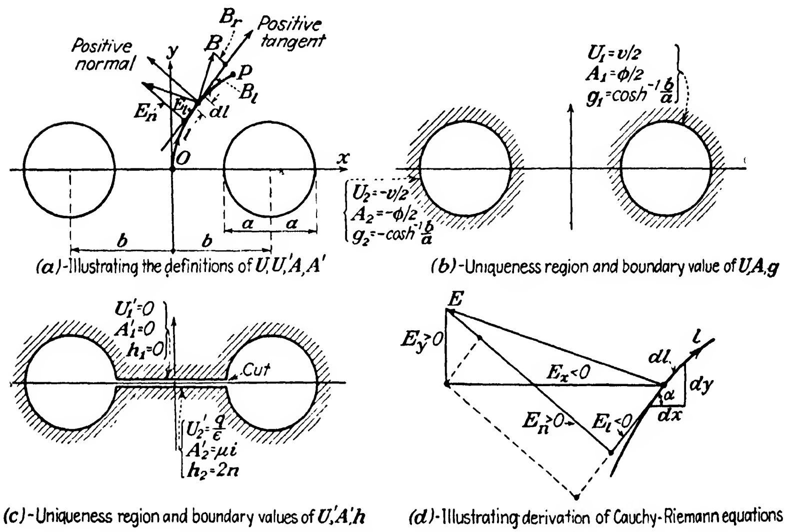


FIG. 63.—*Potentials of a plane field.*

This function is equal to Ψ/ϵ, where Ψ is the *electric flux* across a surface S obtained by translating l in the z direction by the unit length, ϵ the permittivity of the dielectric. In order to establish the uniqueness of U' for all possible paths l, we must make a cut along the x axis and limit l to the region of Fig. 63*c*.

(c) $$-\int_l B_l\,dl = A'(x, y) \tag{299}$$

This function, like U', is single valued only if l is restricted to the region of Fig. 63*c*. A' is sometimes called *magnetostatic potential.*

(d) $$\int_l B_n\,dl = A(x, y) \tag{300}$$

This function is equal to the magnetic flux Φ through the surface

S defined under (298). A is single valued for all possible paths l. The vector $A_z = A$, $A_x = A_y = 0$ is the *magnetic vector potential* (Sec. 11.6).

The functions U and A are continuous and single valued[4] over the region of the xy plane outside the conductors (Fig. 63b); at the boundaries (conductor surfaces) we have

$$\begin{cases} U_1 = \dfrac{v}{2} & U_2 = -\dfrac{v}{2} \\ A_1 = \dfrac{\Phi}{2} & A_2 = -\dfrac{\Phi}{2} \end{cases} \tag{301}$$

where v is the voltage between conductors, Φ the external magnetic flux (Sec. 11.7). The above boundary conditions are based on the premise that **E** and **B** are directed along the normal and the tangent to the surface, respectively. The second premise is rigorously justified only when the current flow is limited to the conductor surface.

The functions U' and A' are continuous and single valued over the region of Fig. 63c. If U' and A' are arbitrarily made equal to zero at the upper edge of the cut, their boundary values will be (Fig. 63c)

$$\begin{cases} U_1' = 0 & U_2' = \dfrac{q}{\epsilon} \\ A_1' = 0 & A_2' = \mu i \end{cases} \tag{302}$$

where q is the charge per unit length on the positive conductor, i the current in either conductor. These boundary values arise from the definitions (299) and (300) and from the laws (281) and (287) (Ampere's circuital law and law of Gauss).

The physical laws of static fields have helped us establish the *boundary values* of the four functions and the facts that the functions are continuous and single valued in the regions defined. This information, apparently very limited, is actually sufficient to determine the functions at all points of the region by mathematical methods. This is the mapping process. There are short cuts to this process, but the general method is of interest to us here.

Let us express the *differentials* of the four functions, or increments of the functions due to an infinitesimal displacement dl of P along l. We have, by definition,

$$\begin{cases} dU = -E_l\,dl & dU' = E_n\,dl \\ dA = B_n\,dl & dA' = -B_l\,dl \end{cases} \tag{303}$$

and in terms of the partial derivatives of the functions along x and y

$$dU = \frac{\partial U}{\partial x}\,dx + \frac{\partial U}{\partial y}\,dy \tag{304}$$

and similarly for the other three functions. dx and dy are related to dl as shown in Fig. 63d. From this figure we also derive the following relations:

$$\begin{cases} E_l = E_x \cos\alpha + E_y \sin\alpha = E_x \dfrac{dx}{dl} + E_y \dfrac{dy}{dl} \\ \therefore\ E_l\,dl = E_x\,dx + E_y\,dy \end{cases} \tag{305}$$

$$\begin{cases} E_n = -E_x \sin\alpha + E_y \cos\alpha = -E_x \dfrac{dy}{dl} + E_y \dfrac{dx}{dl} \\ \therefore\ E_n\,dl = -E_x\,dy + E_y\,dx \end{cases} \tag{306}$$

Similar expressions are obtained for $-B_l\,dl$ and $B_n\,dl$. Substituting these expressions for the right-hand terms of Eqs. (303) and the expansions in terms of partial derivatives (304) for the left-hand terms, we obtain

$$\begin{cases} \dfrac{\partial U}{\partial x}\,dx + \dfrac{\partial U}{\partial y}\,dy = -E_x\,dx - E_y\,dy \\ \dfrac{\partial U'}{\partial x}\,dx + \dfrac{\partial U'}{\partial y}\,dy = -E_x\,dy + E_y\,dx \\ \dfrac{\partial A}{\partial x}\,dx + \dfrac{\partial A}{\partial y}\,dy = -B_x\,dy + B_y\,dx \\ \dfrac{\partial A'}{\partial x}\,dx + \dfrac{\partial A'}{\partial y}\,dy = B_x\,dx + B_y\,dy \end{cases} \tag{307}$$

From (307) we obtain *differential relations* between the potential functions and the field vectors, by comparing the coefficients of dx and dy. Thus

$$\begin{cases} -E_x = \dfrac{\partial U}{\partial x} = \dfrac{\partial U'}{\partial y} & E_y = \dfrac{\partial U'}{\partial x} = -\dfrac{\partial U}{\partial y} \\ B_x = \dfrac{\partial A'}{\partial x} = -\dfrac{\partial A}{\partial y} & B_y = \dfrac{\partial A}{\partial x} = \dfrac{\partial A'}{\partial y} \end{cases} \tag{308}$$

Such differental relations can be derived, more generally, in three dimensions, with the help of the theorems of *Green* and *Stokes*. The relation between U and $\mathbf{E}$ is then written

$$\mathbf{E} = -\mathbf{grad}\ U \tag{309}$$

and that between $\mathbf{A}$ and $\mathbf{B}$

$$\mathbf{B} = \mathrm{curl}\ \mathbf{A} \tag{310}$$

Expanding the above in cartesian coordinates, we would reobtain (308) for the case in question ($A_z = A$, $A_x = A_y = 0$).

Equations (308) may be broken up into two systems of differential equations

$$\begin{cases} \dfrac{\partial U}{\partial x} - \dfrac{\partial U'}{\partial y} = 0 \\ \dfrac{\partial U}{\partial y} + \dfrac{\partial U'}{\partial x} = 0 \end{cases} \tag{311}$$

and

$$\begin{cases} \dfrac{\partial A'}{\partial x} - \dfrac{\partial A}{\partial y} = 0 \\ \dfrac{\partial A'}{\partial y} + \dfrac{\partial A}{\partial x} = 0 \end{cases} \tag{311a}$$

From these systems it is possible to obtain differential equations of the second order in each of the four variables; for example, by differentiating the first of (311) with respect to x, the second with respect to y, and eliminating the cross derivatives, we obtain

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = 0 \tag{312}$$

Equations (311) or (311a) are the equations of *Cauchy-Riemann*, and (312) is the equation of *Laplace*. The Laplace equation is satisfied by all the potential functions of static fields; its solution, combined with the boundary conditions, is the most general method of field mapping in three dimensions.

Two-dimensional fields are conveniently studied by conformal transformations similar to those of network theory. It may be shown, in fact, that if two point functions

$$g = g(x, y) \qquad h = h(x, y)$$

satisfy the Cauchy-Riemann equations,

$$\frac{\partial g}{\partial x} - \frac{\partial h}{\partial y} = 0 \qquad \frac{\partial g}{\partial y} + \frac{\partial h}{\partial x} = 0$$

for all points of a region or domain D of the xy plane; and if the complex number

$$z = x + jy$$

represents a point of D (Sec. 5.7), then the complex number

$$f = g + jh \tag{313}$$

is an *analytic function* of z, or

$$f = f(z)$$

A function is said to be *analytic* within a domain if it possesses a single-valued derivative within the domain; this condition may be shown to follow from the Cauchy-Riemann equations.[4] All algebraic and transcendental functions of complex variables are analytic within specific domains.

Consider in particular the function

$$f = \ln \frac{z + r}{z - r} \tag{314}$$

where $r = \sqrt{b^2 - a^2}$ ($r = \overline{OR}$, Fig. 64). This function is analytic except at the points $z = \pm r$. Its real part g has constant value over the circles

$$(x - b)^2 + y^2 = a^2 \tag{315}$$
$$(x + b)^2 + y^2 = a^2 \tag{316}$$

representing the conductor surfaces. We have, in fact

$$g = \ln \left|\frac{z + r}{z - r}\right| = \tfrac{1}{2} \ln \frac{(x + r)^2 + y^2}{(x - r)^2 + y^2}$$

Replacing y by its value obtained from (315), we have

$$g_1 = \tfrac{1}{2} \ln \frac{b + \sqrt{b^2 - a^2}}{b - \sqrt{b^2 - a^2}}$$

for the first conductor surface, and

$$g_2 = -g_1$$

for the second. These values may be put in the form

$$g_1 = -g_2 = \cosh^{-1} \frac{b}{a} \tag{317}$$

The imaginary part h of f may be zero or any multiple of π

for $y = 0$. However, as z moves continuously from the upper to the lower edge of the cut of Fig. 63c without crossing the cut, h increases by 2π. Letting $h = 0$ at the upper edge, we have the boundary values of h

$$h_1 = 0 \qquad h_2 = 2\pi \tag{318}$$

Now consider the complex function

$$\frac{v}{2g_1}(g + jh) = \frac{v}{2g_1} \ln \frac{z + r}{z - r} \tag{319}$$

This function is analytic in z and its real part has the value $\pm v/2$ at the conductor surfaces; further, it is constant over the x axis. These are also properties of the complex potential function $U + jU'$, and it may be shown that they are sufficient to establish the identity of the two functions. We may, therefore, write

$$U + jU' = \frac{v}{2g_1}(g + jh) \tag{320}$$

We may use the above to find a relation between v and q. In general,

$$U' = \frac{v}{2g_1} h$$

and in particular at the boundary,

$$U_2' = \frac{q}{\epsilon} = \frac{v}{2g_1} h_2 = \frac{v}{2g_1} 2\pi = \frac{v\pi}{g_1}$$

From this and from (317) we have

$$C = \frac{q}{v} = \frac{\pi\epsilon}{\cosh^{-1} b/a} \tag{321}$$

[*Distributed capacity of the parallel line*]

C, the distributed capacity, is equal in this particular case to the capacity per unit length. (q is the charge per unit length.)

Expressing the complex potential in terms of C and z,

$$U + jU' = \frac{vC}{2\pi\epsilon} \ln \frac{z + r}{z - r} \tag{322}$$

The *complex magnetic potential* $A + jA'$ may be arrived at in a similar way. First, we consider the function

$$\frac{\mu i}{h_2}(g + jh)$$

and verify that its imaginary part coincides with A' over one of the boundaries. Next, we verify that this function is analytic, and thirdly, that its real part is constant over the conductor surfaces. These facts establish the identity of the function with $A + jA'$, which may be written

$$A + jA' = \frac{\mu i}{h_2}(g + jh) \tag{323}$$

A relation between Φ and i results from the above, because if

$$A = \frac{\mu i g}{h_2}$$

then on the conductor surface we must have

$$\frac{\Phi}{2} = \frac{\mu i g_1}{h_2} = \frac{\mu i \cosh^{-1} \dfrac{b}{a}}{2\pi}$$

hence

$$l = \frac{\Phi}{i} = \frac{\mu \cosh^{-1} \dfrac{b}{a}}{\pi} \tag{324}$$

[*Distributed inductance of the parallel line*]

In terms of i and z, the complex magnetic potential is

$$A + jA' = \frac{\mu i}{2\pi} \ln \frac{z + r}{z - r} \tag{325}$$

From the expressions of L and C we may verify that

$$\frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\epsilon\mu}} = 3 \times 10^8 \text{ m./sec.} \tag{326}$$

a relation formerly given without proof (Sec. 3.6). Approximate value of L and C for small wire diameters may also be readily obtained. They have been given in Sec. 7.2.

A graphical procedure for mapping the fields, based on the theory developed in this section, accompanies Fig. 64. For other examples of plane field mapping and a complete statement of the theory, the literature[4] should be consulted.

What little has been covered in this section serves to establish some important general facts about plane fields, which we shall have occasion to use later, namely,

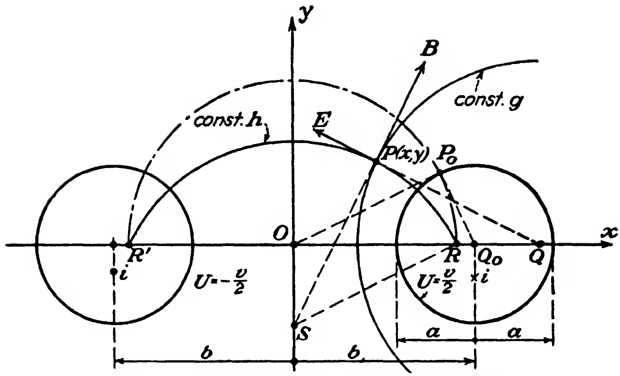


FIG. 64.—*Static fields of the parallel line.*

To map the fields:

1. Draw tangent OP_0 to conductor surface; drop arc P_0R about O, locating R. Linear conductors through R, R' would reproduce external fields of cylindrical wires (this is true for B field only approximately).
2. Draw constant h circle through P and R with center on y axis; tangent to circle at P locates Q.
3. Draw constant g circle through P with center Q.
4. Obtain values of g for points P and P_0:

$$g_0 = \cosh^{-1}\frac{OQ_0}{Q_0P_0} = \cosh^{-1}\frac{b}{a}; \qquad g = \pm\cosh^{-1}\frac{OQ}{QP}\ (+ \text{ if } x > 0)$$

5. Obtain electrostatic potential U at P:

$$U = \frac{v}{2}\frac{g}{g_0} \qquad (v = \text{line voltage})$$

6. Obtain magnitude of vector potential A at P:

$$A = \frac{i}{2}\frac{g\mu}{\pi}\ (i = \text{line current: } A \text{ has direction of current in nearest wire}).$$

7. Repeat for other points on the same constant h circle; plot U and A against distance measured along circle; slope of U plot is **E**; slope of A plot is **B**; directions of **E** and **B** are as shown.
8. Obtain distributed parameters:

$$L = \frac{\mu g_0}{\pi} \qquad C = \frac{\pi\epsilon}{g_0}$$

$$\frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\epsilon\mu}} = 3 \times 10^8 \text{ m/sec. in air}$$

$$R_0 = \sqrt{\frac{L}{C}} = \frac{g_0}{\pi}\sqrt{\frac{\mu}{\epsilon}} \doteq 120 \cosh^{-1}\frac{b}{a} \text{ ohms in air.}$$

NOTE: In the above, b is ½ the center to center spacing.

1. Since the two complex potentials are both analytic functions of $z = x + jy$, lines of *constant* U and *constant* U' may quite generally be obtained from lines of constant x and constant y (parallels to the axes) by an *isogonal transformation*(4) (Sec. 15.7); the same is true of lines of *constant* A and *constant* A'.

2. Since $U + jU'$ and $A + jA'$ are identical functions, except for a real multiplier, lines of *constant* U (equipotentials of the **E** field) are also lines of *constant* A (lines of flow of the **B** field); hence, **E** and **B** are orthogonal everywhere in the space between conductors. This second conclusion is less general than the first, as it is subordinate to the fact that the lines in question must coincide at the conductor boundaries; and this in turn is only exactly true for surface currents.(10) The conclusion is not limited, however, by the shape of the conductors (provided they are parallel, of course).

Another important property of plane fields, that of retaining their configuration under dynamic conditions, will be discussed in Sec. 12.4.

Suggested Exercise. Show that in a coaxial cable, the complex potentials result from the function

$$f = \ln z$$

multiplied by appropriate real coefficients. Find the values of these coefficients from the boundary conditions and obtain values for the distributed parameters.

CHAPTER XII

MAXWELL'S EQUATIONS AND THEIR APPLICATION TO CIRCUIT ELEMENTS

12.1. General Form of the Field Laws. The laws of static fields, as we have seen, are expressed by the equations

$$\oint_l E_l \, dl = 0 \tag{279}$$

$$\oint\!\!\!\oint_S J_n \, dS = 0 \tag{280}$$

[*Valid for static fields*]

From these relations we can derive the statements that, in a static region (Sec. 11.5), the sum of the *voltages* (Sec. 11.4) taken over sections of a closed path and the sum of the *currents* flowing out of a closed surface are zero for all paths and all surfaces. These statements, due to Kirchhoff, continue to be approximately valid in dynamic regions (regions where **E** and **J** are time variables) provided the paths and surfaces are suitably chosen. In this sense, Kirchhoff's laws are often considered valid for a-c as well as d-c circuits.

In a general sense, however, we must modify equations (279) and (280) if we wish them to be rigorously valid under all conditions. The modified equations are

$$\oint_l E_l \, dl + \frac{d\Phi}{dt} = 0 \tag{327}$$

$$\oint\!\!\!\oint_S J_n \, dS + \frac{dQ}{dt} = 0 \tag{328}$$

[*Generally valid*]

where Φ is the *magnetic flux* linked with the closed path or loop l, and Q, the *electric charge* enclosed within the closed surface S, all the quantities being time variables.

Upon equations (327) and (328) is based the theory of lumped systems (circuit theory) under generic conditions, and particularly that of harmonically driven systems (a-c circuits). The transition from the field equations (327) and (328) to the branch equations (voltage-current relations) for lumped elements will be carried out later; it depends, as we shall see, upon the possibility of disregarding second-order effects with engineering approximation.

Should we require an *exact* solution of *any* dynamic problem, or should we inquire into phenomena due chiefly or entirely to the second-order effects mentioned above (radiation, for instance), then we would find that equations (327) and (328), while valid, are insufficient. In such problems we would need a relation between magnetic flux and current density, and so far such a relation is available for *static fields only*. [See Eqs. (281) and (283) in the tabulation of Sec. 11.9. Note that the corresponding relation between **E** and **D** is not restricted to static fields.]

Historically, Eq. (328) was established first, as it simply states that the total current leaving a volume must equal the rate of time decrease of the charge in the volume. This law may be identified with the definition of current (272) and was confirmed experimentally as soon as it became possible to measure currents and charges independently.

Equation (327) represents a later development, due to Faraday's interpretation of much experimental evidence on induced currents. Maxwell supplied the missing equation relating the flux and current densities in general (for both static and dynamic fields) by his theoretical work, later supported by the observations of Hertz.

12.2. Maxwell's Equations in the Integral Form. We may rewrite (328), expressing Q in terms of the electric flux density, (Eq. 287), thus

$$\oint\!\!\oint_S J_n\,dS + \frac{d}{dt}\oint\!\!\oint_S D_n\,dS = 0$$

If the medium is at rest, *i.e.*, stationary, with respect to the surface S, we can carry out the *surface integration* and the *time differentiation* in inverse order with the same result. Hence, the alternative form[1]

$$\oint\!\!\oint_S \left(J_n + \frac{\partial D_n}{\partial t}\right) ds = 0 \tag{329}$$

[1] The time derivative of **D** and of other point functions is usually written as a partial derivative to imply that the space variables on which the function also depends (x, y, a in the cartesian system) are considered constant.

Comparing the above with (280) and recalling the definition of a solenoidal vector (Sec. 11.5), it appears that, while **J** is solenoidal only in static regions, $\mathbf{J} + \partial\mathbf{D}/\partial t$ has this property in dynamic regions as well. The flow lines of this composite vector are always closed; over part of each flow line, the vector may be equal to **J** and represent an actual current density, while elsewhere it may be entirely due to the time variation of electric flux density.

Consider the charging of a condenser. The streamlines of the charging current terminate at the surface of the condenser plates. Just outside the surface there is a value of electric flux density increasing with time; its rate of increase, $\partial\mathbf{D}/\partial t$, is a vector quantity which continues the current streamline outside the conductor across the dielectric.

When discussing static fields, we argued (Sec. 11.5) that if **J** is solenoidal, its integral over a surface, which is the current through the surface, must depend only on the *contour* of the surface and be expressible as the integral of an appropriate vector point function over this contour. Experiment shows (as observed by Ampere) that $\mathbf{H} = \mathbf{B}/\mu$ is one such point function.

Going now to the general case, $\mathbf{J} + \partial\mathbf{D}/\partial t$ is solenoidal. Its integral over a surface must once again be a function of the contour and expressible by contour integration of an appropriate point function. Maxwell inferred that the same point function, $\mathbf{H} = \mathbf{B}/\mu$, integrated over a contour l, gives, in a *static region*, the current through a surface S, bounded by l:

$$\oint_l H_l\,dl = i = \iint_S J_n\,dS$$

and in a *dynamic region*, the summation

$$\oint_l H_l\,dl = i + \frac{d\Psi}{dt} = \iint_S \left(J_n + \frac{\partial D_n}{\partial t}\right) dS$$

where Ψ is the electric flux (285) through the surface S. $d\Psi/dt$ was called by Maxwell the *displacement current* through S. Thus, Maxwell extended Ampere's circuital law to dynamic fields by adding the displacement current to that due to the flow of charge, or, in symbols,

$$\oint_l H_l\,dl = i + \frac{d\Psi}{dt} \tag{330}$$

Maxwell developed the modern theory of dynamic fields, using as the starting point equation (330), jointly with equation (317), which expresses in general form the results of Faraday's observations. This system of two equations may be regarded as the basis of all electromagnetic theory, as all the other laws (287), (279), (280), (281), (284), and (328) can be derived from it. The two equations are brought together below for convenience

$$\oint_l E_l\,dl = -\frac{d\Phi}{dt} \tag{327}$$

$$\oint_l H_l\,dl = i + \frac{d\Psi}{dt} \tag{330}$$

[*Maxwell's equations in the integral form*]

In the first equation, l is the contour of the surface through which Φ is taken; in the second, l is the contour of the surface through which both i and Ψ are taken. Aside from this condition, there is no restriction on the choice of loops and surfaces.

12.3. The Scalar Potential. The reader will find in the literature[(10)] the development of electromagnetic wave theory from Maxwell's equations. The equations will be used here only for the purpose of justifying the voltage-current relationships currently used in the study of transmission lines as well as lumped systems. It will appear that in the case of transmission lines, such relations can be considered theoretically sound, while for lumped circuits they are only useful approximations.

We shall require for this purpose an extension of the concept of electrostatic potential (Sec. 11.4) from the static to the dynamic case. In a static region the potential is

$$U = -\int_{P_0}^{P} E_l\,dl \tag{331}$$

[*Valid in static regions*]

U is a single valued function of P because loop integral of **E** vanishes for any loop of a static region, or

$$\oint_l E_l\,dl = 0 \tag{279}$$

which may be expressed by stating that **E** is a *conservative* vector. (The word *conservative* implies that the potential energy of a particle subject to the force of the field is conserved, or remains unaltered, whenever the particle describes a closed path beginning and ending at the same point.)

In a dynamic region **E** is no longer conservative, as (279) is no longer valid; hence, the potential as defined by (331) is no longer a single-valued point function. Loci of constant U, or *equipotential surfaces*, cease to have meaning unless we contrive to define U in a more general way.

Clearly, this general definition of U can only be written in terms of a conservative vector. **E** is not such a vector, generally speaking. We have to look for a *dynamic addition* to **E** which will make it conservative in all cases. We traveled a similar road before, when we found a dynamic addition ($\partial \mathbf{D}/\partial t$) which made the vector **J** solenoidal in all cases. We started then from the general equation (328), brought the derivative within the integral sign, and expressed the sum of the integrals as the integral of a single vector. We can do exactly the same thing now, taking the other general equation (327) as the starting point. This runs:

$$\oint_l E_l \, dl + \frac{d\Phi}{dt} = 0 \tag{327}$$

We can express Φ as a loop integral, because it represents the flow of a solenoidal vector **B** across a surface. This was done before, in defining the *vector potential* **A**

$$\Phi = \oint_l A_l \, dl \tag{293}$$

Taking the differentiation within the integral sign and combining the integrals, we may now write (327) in the form

$$\oint_l \left(E_l + \frac{\partial A_l}{\partial t} \right) dl = 0 \tag{332}$$

signifying that $\mathbf{E} + \partial \mathbf{A}/\partial t$ is conservative in dynamic as well as in static regions. The *scalar potential* may now be defined as follows:

$$U = -\int_{P_0}^{P} \left(E_l + \frac{\partial A_l}{\partial t} \right) dl \tag{333}$$

[*General*]

In a static region, the derivative $\partial \mathbf{A}/\partial t$ vanishes and the scalar potential becomes identical with the electrostatic potential, as defined by (331). The electrostatic potential is actually no other than the value which the scalar potential takes in a static region and need not be considered to have separate meaning.

In general, it is evident that the equipotential surfaces take different configurations in going from static to dynamic conditions. There are exceptional cases when this is not true. It would be more accurate to say that, exceptionally, in certain systems, the scalar potential may be defined in general without recourse to a dynamic term. Hence, the dynamic study of such systems is comparatively very simple.

12.4. The Invariance of Plane Fields. The scalar potential U may be defined without recourse to a dynamic term by restricting the path l of the integration (333) so that it is perpendicular to **A**, the vector potential, at all points.

Take, for example, the parallel line of Sec. 11.10. (Any other system of parallel conductors would do.) Assume the path of integration l to be confined within a plane orthogonal to the conductors, as we did for the static field analysis.

If the current flows parallel to the conductor axes, it can be shown vector analytically and otherwise that **A** is likewise oriented.[1] Then it is clear that **A** and the path l must be mutually orthogonal ($A_l = 0$) everywhere. The definition of U valid for plane fields under static conditions (79) *continues to be valid in the dynamic case.*

As a consequence, all four potential functions continue to be defined as before, and the entire analysis of plane fields remains valid. All the quantities will of course vary with time, but *at any particular instant* the fields will depend on v and i, line voltage and current, exactly as if these were constant. This is an important principle, and for easy reference we shall call

[1] Consider a straight linear current in empty space and a cylindrical surface coaxial with it, both indefinite. If **A** at any point of this surface had a radial component, this would be true of all such points, due to the axial symmetry; the cylinder being a closed surface, **A** in this case would not be solenoidal, which contradicts its definition (Sec. 11.6). Next take a circle on this cylinder. If, at some point, **A** had a component tangent to the circle, this would be true of all points of the circle and the loop integral of **A** around the circle would not be zero. Since this is equal to the flux and there are no axial flux lines, such components cannot exist. Hence, **A** must be parallel to the current. By the superposition principle, we conclude that **A**, due to a system of parallel currents, has the direction of these currents.

it the *principle of invariance of plane fields.* It must be understood that *plane fields* are defined as the fields of a system of parallel currents; if the conductors are parallel but the current does not flow axially, we cannot regard the resulting fields as plane.

We cannot fully analyze the performance of parallel lines under all conditions without recourse to much more complex methods of analysis than those used so far. Thanks to the principle stated above, we can, however, justify the simpler methods by showing that they are in accord with the laws of electromagnetic theory, provided we assume the current to flow axially at all points. This matter will be taken up in Sec. 13.1.

12.5. Branch Equation for the Coil. Going back to Eqs. (327) and (328), let us see how the branch equations of the basic circuit elements or two-poles, the *coil* and the *condenser*, follow from them.

A coil is shown in Fig. 65. The exact meaning of the term *coil* is implicit in the following assumptions relating to it, which permit the derivation of the branch equation. We must first imagine a boundary separating the coil from the remainder of the electrical system of which it is a part. This may be a material boundary (in a shielded coil) or an ideal one. In any case it must be a closed surface through which pass the two leads joining the coil to the rest of the system; in Fig. 65 it is shown as an indefinite plane surface.

Next, we must imagine another surface (surface S, Fig. 65), also closed, contained entirely within the coil boundary and crossed by the coil wire twice, once at the boundary and once at an arbitrary point.

The first assumption is that the total displacement current through S is negligible compared to the conduction current at the boundary (the coil current, i). This can be written

$$\frac{d\Psi}{dt} \ll i \tag{334}$$

The above must be valid for any S within the definition given; hence, the current in the coil wire may be considered to be constant throughout the coil (330).

Next, consider a closed path l as shown (Fig. 65). This path follows the surface of the coil wire everywhere except on the boundary surface. Let Φ stand for the magnetic flux linked by l (Sec. 11.5) when the current is i. Now imagine the coil to be short-circuited along the entire path l; we would then have

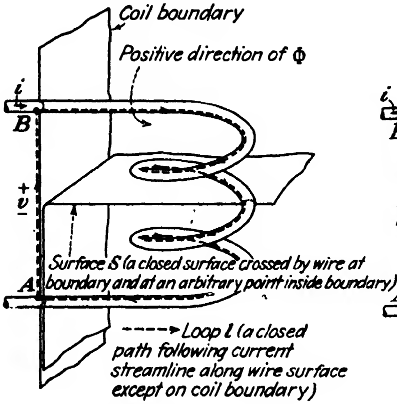


FIG. 65*a*.—*Voltage-current relation for the coil.*

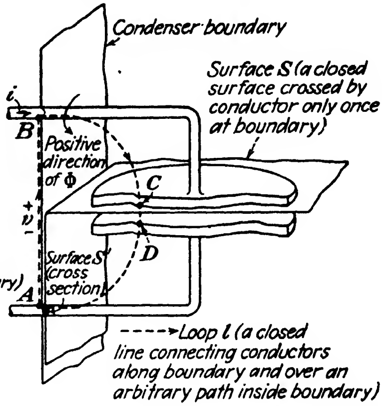


FIG. 65*b*.—*Voltage-current relation for the condenser.*

Assumptions:

$$\frac{d\Psi}{dt} \ll i \qquad \Phi \approx Li$$

where Ψ = electric flux through any surface S
Φ = magnetic flux through loop l
L = inductance of coil short-circuited along boundary (see Fig. 61*b*)

Assumptions:

$$\frac{d\Phi}{dt} \ll v \qquad Q \approx cv$$

where Φ = magnetic flux through any loop l
Q = electric flux through surface S
C = capacity of condenser open-circuited at boundary.

Derivation: for coil:

$$\oint_l E_l\,dl = \underbrace{-\int_B^A E_l\,dl}_{v} + \underbrace{\int_B^A E_l\,dl}_{Ri} = \underbrace{-\frac{d\Phi}{dt}}_{}$$

hence: $$v = Ri + L\frac{di}{dt}$$

For condenser:

$$\oiint_S J_n\,dS = \underbrace{-\iint_{S'} J_n\,dS}_{i} + \iint_{S-S'} J_n\,dS = -\frac{dQ}{dt}$$

hence: $$i = Gv + c\frac{dv}{dt}$$

a closed current path, for which there is an inductance L, as defined in Sec. 11.7 (Fig. 62*b*). It is clear that, as long as the coil is part of a system, it does not possess a definable inductance; inductance has no exact meaning when it refers to a section of a current circuit, with one exception—the distributed inductance of a parallel line. Having thus defined L as the inductance of the coil in short circuit, and noting that i is constant over the coil length, we can formulate the second assumption

$$\Phi \approx Li \tag{335}$$

The above implies that, *for a given coil current*, the magnetic field *inside the coil boundary* does not change appreciably, whether the coil is short-circuited across its terminals or connected to an outside source. (If the reader finds it difficult to imagine a current flowing in the coil without external connections, it might be helpful to imagine that the short-circuiting connection includes a small battery.)

A third, rather obvious assumption should be added for completeness. No conduction current is assumed to flow in the space surrounding the wire.

We may now obtain the required voltage-current relation from (327). The loop integral of E may be written for the path l and split in two as follows (Fig. 65):

$$\oint_l E_l\,dl = -\int_B^A E_l\,dl + \int_B^A E_l\,dl = -\frac{d\Phi}{dt} \tag{336}$$

The first integral $\int_B^A E_l\,dl$ is, by definition, the coil voltage v. We cannot associate a voltage with the coil unless we agree on a path over which the voltage is taken (Sec. 11.4), unless we assume static conditions. This path must be considered to lie along the coil boundary and to form part of the loop l along which we have imagined the coil to be short-circuited when defining its inductance. To be sure, in most cases the exact choice of a path will not make much difference, but our premises must be precise if we want our mathematical statements to have meaning.

The second integral, taken entirely over a path within the conductor, can be expressed in terms of current with the help of Eq. (277). We must consider the current to be constant over the whole path, which can be done with small error, thanks to our first assumption (334). Then, following the procedure of Sec. 11.4, we may write

$$\int_B^A E_l\,dl = Ri$$

where R is the *coil resistance*.

Replacing the integrals of (336) by their values, as above, and using (335) to express Φ in terms of current, we have the branch equation for the coil[1]

$$v = Ri + L\frac{di}{dt} \tag{337}$$

The importance of associating the coil voltage with a specific path is brought out by the following observation. Suppose no path were specified; then we could take the coil voltage, for example, over a path following the coil wire. In this case the voltage would be the second integral of (336) instead of the first, and we would write the coil equation as follows:

$$v = Ri$$

which evidently contradicts (337). The reason for specifying that the voltage path lie along the boundary, for the coil as well as for other two-poles, is this: If we consider the two-pole in question and the source from which it receives power to have a common boundary, the voltage taken along the boundary is common to both.

It is wrong to think of voltage as a difference of potential, except in a static region, and in any case it is not necessary to explain voltage in this way. There should be a better word than *voltage* to denote the line integral of the field (in some languages, *tension* or its equivalent is used); but the *concept* of voltage is more general and less abstract than that of electrostatic potential; the latter is widely used to explain voltage, mainly because it lends itself to the gravitational analogy and thus eliminates the necessity of introducing the line integral. This method leads to inevitable confusion, however, when dynamic effects are studied.

12.6. Branch Equation for the Condenser. The assumptions relating to the condenser are as follows:

1. For any loop l entirely within the condenser boundary (Fig. 65*b*) and extending along the boundary between conductors

$$\frac{\partial \Phi}{\partial t} \ll v \tag{338}$$

[1] A lossless or ideal coil is one for which $R = 0$. In some circuit diagrams actual coils are represented by an ideal coil in series with a resistor. This is an equivalent representation.

where Φ is the flux linked by the loop l. From this assumption we conclude that the voltage between the two conductors, over any path entirely within the boundary, may be considered equal to v, the condenser voltage (defined as the voltage along a path lying on the boundary).

2. If Q is the electric flux through a surface S as shown (Fig. 66), we assume

$$Q \approx Cv \tag{339}$$

where C is the capacity of the condenser, open-circuited at the boundary. Capacity, like inductance, cannot *in general* be associated with a part of a system; it relates to a closed system of two conductors.

3. The electric field within the condenser plates and leads is negligibly small (small resistivity). As a consequence of this assumption and of assumption (1), the voltage over l from C to D equals the condenser voltage.

We can now obtain the required relation from (328). This can be written in the form

$$\oint\!\!\!\!\oint_S J_n\,ds = -\iint_{S'} J_n\,ds + \iint_{S-S'} J_n\,ds = -\frac{dQ}{dt} \tag{340}$$

where S' is the cross section of the conductor at the boundary (Fig. 66), $S - S'$ the remainder of the surface S. The two integrals represent respectively the coil current i and the leakage current in the condenser dielectric, Gv (Fig. 61a). Expressing Q in terms of C and v (339), we have

$$i = Gv + C\frac{dv}{dt} \tag{341}$$

which is the *branch equation* for the condenser.[1]

12.7. Lumped Linear System. Systems made up of coils and condensers, for which the voltage-current relations are expressed by (23) and (27), connected through their terminals, have been called *lumped linear systems* or *linear circuits* or *networks*. (In the present book, however, the term *network* is used exclusively to

[1] A lossless or ideal condenser is one for which $G = 0$. In circuit diagrams, actual condensers are sometimes shown as ideal condensers shunted by a conductance (an equivalent representation); usually however, G can be neglected for practical purposes.

denote a *part* of a system, lumped or otherwise, joined to the rest at more than two points, as defined in Sec. 1.1.)

The definition of a *lumped linear system* includes, of course, systems in which some of the two-poles are resistors. A resistor may be considered as a coil for which $L = 0$ and R is finite, or a condenser for which $C = 0$ and G is finite. The definition does not include, however, mutual inductances, unless they are considered replaced by their T equivalents (Fig. 77*c*).

It is not the purpose of this book to go systematically into lumped circuit theory. The reader is assumed to have a working knowledge of the more fundamental phases of it. Specific problems which involve the use of lumped circuits have been treated from time to time, and in doing this it has never been considered necessary to explain the meaning of the quantities and symbols currently used in steady-state circuit work (a-c circuits). However, the present chapter has the purpose of effecting a connection, in the mind of the reader, between the basic laws of electromagnetism and the methods used to solve circuit and network problems; the remaining link in this connection, as far as circuit problems are concerned, consists in showing how the concept of *impedance* originates from the voltage-current relations (337) and (341). This will be done in the following section, the subject of which will not be pursued further. At appropriate points bibliographical notes will direct the reader to the literature on elementary as well as advanced phases of circuit theory.

Chapter XIII will go back to the laws of electromagnetism and their application not to lumped but to distributed or continuous systems.

★12.8. Complex Voltage and Current. Impedance. Oscillation Constant. The solution of any circuit problem ultimately reduces to the simultaneous solution of a number of equations similar to (337) or (341) (branch equations), to which must be added a sufficient number of *conditions*, *i.e.*, the voltage or current at some particular junction (Sec. 1.5) as a function of time (boundary value problems); or the voltage and current at some junction and at some instant of time, after which the system is allowed to oscillate freely, none of the variables being any longer conditioned except by their mutual relationships (initial value problems).

Steady-state a-c circuit problems are boundary value problems; the solution for transients is an initial value problem. This terminology is general rather than typical of circuit theory.

The branch equations upon which is based the solution of either kind of circuit problem are, in general, linear differential equations in v and i. (If we think of a single branch as a series combination of R, L, and C, as is frequently done, the branch equation is of the *integrodifferential* type.) Such equations may, however, reduce to linear *algebraic* equations of the form

$$v = Zi \tag{342}$$

if both v and i are a special type of time function, which we may readily identify by checking (342) against either (337) or (341). Let us ask, for example, what functions $i(t)$ and $v(t)$ are compatible with (342) and with the general form (337) of the coil equation. Eliminating v between (342) and (337), we have an equation in $i(t)$

$$Zi - Ri - L\frac{di}{dt} = 0 \tag{343}$$

This may be integrated by inspection; we have for the unknown function

$$i(t) = \hat{I}e^{pt} \tag{344}$$

where p stands for

$$p = \frac{Z - R}{L} \tag{345}$$

and the ratio Z, in terms of p and branch parameters, has the value

$$Z = R + pL \tag{346}$$

We find, therefore, that (337) may be written in the form of a linear algebraic equation only if i varies exponentially with time. The same is true of (341).

The family of functions $i(t)$ which satisfy both (342) and (337) might be extended very considerably if we were to include *complex functions.* What physical significance would such functions have, however? A current is a scalar quantity, definable by a single number. A complex number cannot represent the instantaneous value of a current.

The difficulty may be surmounted by requiring that the instantaneous value of current be equal, not to the function which satisfies (342), but to its *real part* (or to its imaginary part; we will take the first alternative). Thus we may write[1]

$$i = Re(\mathbf{I})$$

and likewise

$$v = Re(\mathbf{V})$$

where **I** and **V** are complex functions, which must be so chosen that they satisfy Eq. (342), while their real parts satisfy Eq. (337). These functions are called the *complex current* and the *complex voltage*.

If **I** and **V** satisfy (342), we may write

$$\mathbf{V} = Z\mathbf{I} \tag{347}$$

We may now write (337) in terms of **I** only, as follows

$$Re(Z\mathbf{I}) = R[Re(\mathbf{I})] + L\frac{d}{dt}[Re(\mathbf{I})]$$

Noting that

$$\frac{d}{dt}[Re(\mathbf{I})] = Re\left[\frac{d\mathbf{I}}{dt}\right]$$

the above may be written

$$Re\left\{Z\mathbf{I} - R\mathbf{I} - L\frac{d\mathbf{I}}{dt}\right\} = 0 \tag{348}$$

Any root of the equation

$$Z\mathbf{I} - R\mathbf{I} - L\frac{d\mathbf{I}}{dt} = 0$$

satisfies (348). The complex function $\mathbf{I}(t)$ may now be obtained by integration, as follows:

$$\mathbf{I}(t) = \mathbf{I}_0 e^{pt} \tag{349}$$

The above is the same as (344), except that p may now be complex. As before, we have

$$Z = \frac{\mathbf{V}}{\mathbf{I}} = R + pL \tag{350}$$

Z is no longer the ratio of instantaneous values of i and v; it is the ratio of the *complex values* (also called *current vector* and *voltage vector*). Z is called *impedance* and is, in general, a complex number. It is a function of p and of the parameters of the two-

[1] $Re\ (Z)$ stands for real part of Z.

pole. The reciprocal of Z, the *admittance* Y, is often simpler to express in terms of these parameters; for example, we have for the condenser (Eq. 341)

$$Y = \frac{\mathbf{I}}{\mathbf{V}} = G + pC \tag{351}$$

Both Z and Y have meaning only when the instantaneous current (or voltage) varies with time in such a manner as to be identifiable with the real part of the complex function (349). Let us see what time variations are included in this group.

Without loss of generality, we may assume the constant $\mathbf{I}_0$ to be a real $\hat{I}_0$ (this merely assigns the time origin). Then, letting

$$p = \alpha + j\omega \tag{352}$$

we find

$$i = \hat{I}_0 e^{\alpha t} \cos \omega t \tag{353}$$

For the impedance of a two-pole to have the meaning of the ratio between *complex voltage* and *complex current*, the *instantaneous* current must vary with time as indicated by (353). As a particular case of (353) (for $\omega = 0$), we have exponentially decreasing or increasing currents, in which case, as we have seen, the complex current becomes a real and coincides with the instantaneous value. For $\alpha = 0$, (353) becomes the equation of an alternating or harmonic current.

In general, (353) describes an oscillation whose amplitude increases or decreases exponentially. The two numbers α and ω fully define the character of the oscillation; for this reason, the complex number $p = \alpha + j\omega$ is called *oscillation constant.* For harmonic oscillations, the oscillation constant reduces to $j\omega$ and the familiar expressions for the steady-state impedance of a coil and admittance of a condenser result from the general expressions (347) and (351).

It can be shown, by solving the system of branch equations,[8] that the currents and voltages in a freely oscillating lumped linear system (Sec. 12.7) obey (353), α and ω having particular values depending on the parameters of the system. We may in consequence write a branch equation of the form

$$\mathbf{V} = Z\mathbf{I}$$

for each branch or two-pole of such a system. For a closed system, the resulting system of equations will show that all voltages and currents must be zero, unless the branch impedances are suitably related to one another. This condition, when each impedance is written as a function of p and the branch parameters, becomes an equation in p; the roots of this equation determine the frequency and decrement of the free oscillations of which the system is capable.

12.9. Applications and examples.

Zeros and poles of a lumped linear system. Consider in particular the closed systems consisting of a branch Z (the impedance of a branch may be used in a wider sense to denote the branch itself) in open and short circuit (Fig. 66). Considering the open circuit as a branch of zero admittance and the short circuit as a branch of zero impedance, we have

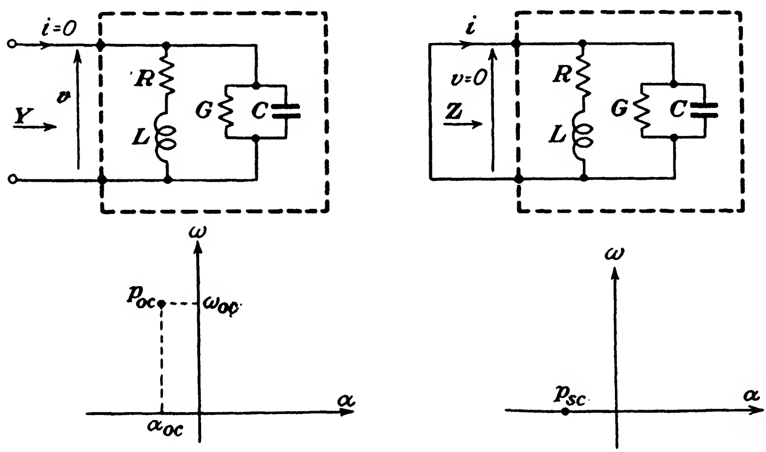


FIG. 66.—*Poles and zeros.*

a. Two-pole in open circuit.
The voltage v must have the form

$$v = Re(\mathbf{V}) = Re(\hat{V}e^{pt}) = \hat{V}e^{\alpha t}\cos\omega t$$
$$(p = \alpha + j\omega)$$

For $\mathbf{V} \neq 0$, we must have

$$Y = Y(p) = \frac{\mathbf{I}}{\mathbf{V}} = 0$$

since $\mathbf{I} = 0$ (open circuit). Solving the above for p, we obtain the value

$$p_{oc} = \alpha_{oc} + j\omega_{oc}$$

p_{oc} is the zero of Y (Y may have more than one zero in general). α_{oc} and ω_{oc} are the characteristics of the open-circuit oscillations.

Two-pole in short circuit.
For $\mathbf{I} \neq 0$, we must have

$$Z = Z(p) = \frac{\mathbf{V}}{\mathbf{I}} = 0$$

since $\mathbf{V} = 0$ (short circuit). Solving for p we obtain

$$p_{sc} = \alpha_{sc} + j\omega_{sc}$$

p_{sc} is the zero of Z
Note: p_{oc} is also a root of the equation

$$Z(p) = \infty$$

and may be considered a pole of Z. Likewise, p_{sc} is a pole of Y.

BRANCH IN OPEN CIRCUIT

For open branch

$$\mathbf{I} = 0$$

For branch of admittance Y

$$\mathbf{I} = \mathbf{V}Y$$

Combining,

$$\mathbf{V}Y = 0$$

First solution:

$$V = 0$$

(no oscillation)

Second solution:

$$Y = 0$$

(condition for oscillation)

BRANCH IN SHORT CIRCUIT

For shorted branch

$$\mathbf{V} = 0$$

For branch of impedance Z

$$\mathbf{V} = \mathbf{I}Z$$

Combining,

$$Z\mathbf{I} = 0$$

First solution:

$$I = 0$$

(no oscillation)

Second solution:

$$Z = 0$$

(condition for oscillation)

Let the branch in both cases be, for example, a parallel combination of coil and condenser (Fig. 66). The well-known rules for obtaining the impedance or admittance of parallel and series combinations apply, of course, in the general case when p is not a pure imaginary. Hence, we have

$$Y = G + pC + \frac{1}{R + pL} = \frac{1 + (R + pL)(G + pC)}{R + pL} \tag{354}$$

$$Z = \frac{R + pL}{1 + (R + pL)(G + pC)} \tag{355}$$

The two conditions for oscillation (for open and short circuit) are, respectively,

Open circuit ($Y = 0$):

$$(R + pL)(G + pL) + 1 = 0$$

Solving, we have the following roots:

$$p_{oc} = -\frac{1}{2}\left(\frac{G}{C} + \frac{R}{L}\right) \pm \sqrt{\frac{1}{4}\left(\frac{G}{C} + \frac{R}{L}\right)^2 - \frac{1 + RG}{CL}}$$

$\therefore$

$$\begin{cases} \alpha_{oc} = -\frac{1}{2}\left(\frac{G}{C} + \frac{R}{L}\right) \\ \omega_{oc} = \dfrac{1}{\sqrt{CL}} \sqrt{1 + RG - \frac{1}{4}\left[G\sqrt{\frac{L}{C}} + R\sqrt{\frac{C}{L}}\right]^2} \end{cases}$$

Short circuit ($Z = 0$):

$$R + pL = 0$$

$$p_{sc} = -\frac{R}{L}$$

$\therefore$

$$\begin{cases} \alpha_{sc} = -\dfrac{R}{L} & (356) \\ \omega_{sc} = 0 & (357) \end{cases}$$

The values above are those of the decrement and angular velocity of free oscillation for the two systems. The oscillation constant p_{oc} for oscillation of the branch in open circuit is a *zero* of the branch admittance (it is the value of p for which $Y = 0$). Similarly, p_{sc} is a *zero* of the branch impedance. Inversely, p_{sc} is a *pole* of the branch admittance and p_{oc} a *pole* of the impedance. (When p is at a pole of impedance, the impedance is infinite.)

Geometrically, the Z plane and the p plane are correlated, point for point, by an equation in the two complex variables Z and p, for example, Eq. (355). When the Z point is at the origin, the corresponding p point (point p_{sc}) occupies on the p plane a *zero* of the function Z. When Z is at infinity, the corresponding p point (p_{oc}) occupies a *pole* of the function (Fig. 67).

In the example chosen, there is only one pole and one zero for either Z or Y. This is not always the case. The function Z for any lumped linear two-pole is fully determined by its zeros and poles.

The impedance function theory and its applications to network design are treated extensively in the sources cited in the bibliography under group 5. A short bibliography on the theory of transients in linear systems will be found under group 8.

Definition of Q_0 in terms of the oscillation constant. In Sec. 9.5, Q_0 was defined as follows:

$$Q_0 = \frac{1}{2}\frac{\omega_o}{G_0}\left|\frac{dY}{d\omega}\right|_{(\omega=\omega_o)} = \frac{1}{2}\frac{\omega_o}{R_0}\left|\frac{dZ}{d\omega}\right|_{(\omega=\omega_o)}$$

The above definition made it possible to obtain Q_0 from the impedance or admittance function by differentiation; it also relates Z_0 directly to the loss in power, or to the drop in voltage, caused by a frequency departure from midband.

We are now in a position to show that Q_0, as defined above, is also given by

$$Q_0 = -\frac{\omega_{oc}}{2\alpha_{oc}}$$

when it refers to a zero of susceptance, and by

$$Q_0 = -\frac{\omega_{sc}}{2\alpha_{sc}}$$

with reference to a reactance zero, provided that, in either case, $Q_0 \gg 1$.

Consider a selective branch, such as that of Fig. 66. For this we have

$$p_{oc} = \alpha_{oc} + j\omega_{oc}$$

meaning that, if we impressed across the branch an e.m.f. of angular frequency ω_{oc} and decrement $-\alpha_{oc}$(α_{oc} is an intrinsically negative number for any passive branch), no current would flow in the branch.

Let us ask the following: If a *harmonic* e.m.f. of frequency ω_{oc} (and decrement 0, of course) is impressed, what will be the admittance of the branch? This will be a value of harmonic (a-c) admittance as distinct from generalized admittance.

This question may be answered in two separate ways, and a comparison of the answers yields the desired definition of Q_0.

First, let us imagine that the p point is moved from p_{oc} to $j\omega_{oc}$ on the p plane (Fig. 66*a*). This interprets graphically the change from the oscillation of constant p_{oc} to the harmonically driven condition. At the same time, the Y point will have moved out of the origin ($Y = 0$) to some point not far from the origin, and the path of this movement will be an isogonal transformation[4] of the path followed by p. Since p moved in a direction orthogonal to the axis of imaginaries, such must be the direction of Y (the axis of imaginaries transforms into itself). Hence, Y moves from the origin along the axis of reals; its value for $p = j\omega_{oc}$ (impressed harmonic e.m.f. of frequency ω_{oc}) is a real number G_0. Hence ω_{oc}, imaginary component of the branch admittance zero, is itself a zero of susceptance, and we have

$$Y(p = j\omega_{oc}) = G_0$$

Secondly, we may express $Y(p = j\omega_{oc})$ by a *MacLaurin* expansion of Y about the origin. Thus

$$Y_{(p=j\omega_{oc})} = Y_{(p=p_{oc})} + \frac{dY}{dp}_{(p=p_{oc})}(j\omega_{oc} - p_{oc}) + \cdots + \frac{1}{2!}\frac{d^2Y}{dp^2}_{(p=p_{oc})}(j\omega_{oc} - p_{oc})^2$$

Noting that

$$j\omega_{oc} - p_{oc} = -\alpha_{oc}$$

is small, the expansion may be limited to the first-order term and written

$$Y_{(p=j\omega_{oc})} = -\frac{dY}{dp}_{(p=p_{oc})}\alpha_{oc}$$

The derivative dY/dp does not change appreciably as p moves from p_{oc} to $j\omega_{oc}$; furthermore, as Y is an analytic function,[4] its value is unaffected by the direction of the infinitesimal displacement dp on the complex plane. Hence the equality

$$\frac{dY}{dp}_{(p=p_{oc})} = \frac{dY}{d(j\omega)}_{(p=p_{oc})} = \left|\frac{dY}{d\omega}\right|_{(\omega=\omega_{oc})}$$

Finally, comparing the two expressions of $Y(p = j\omega_{oc})$,

$$G_0 = -\alpha_{oc}\left|\frac{dY}{d\omega}\right|_{(\omega=\omega_{oc})}$$

from which we have, as we set out to prove

$$Q_0 = \frac{1}{2}\frac{\omega_{oc}}{G_0}\left|\frac{dY}{d\omega}\right|_{(\omega=\omega_{oc})} = -\frac{\omega_{oc}}{2\alpha_{oc}}$$

This equality—as well as its counterpart in terms of the components of the impedance zero p_{sc}—is significant in that it brings out the physical aspect of selectivity. We may consider the *power* stored in the open-circuited branch to decay as the square of the voltage (or current); hence, expressing the ratio of the power lost during a period T to the power still in store, we have

$$\frac{P_{\text{lost}}}{P_{\text{stored}}} = e^{-2\alpha_{oc}T} - 1$$

The exponent $-2\alpha_{oc}T$ may be rewritten $-4\pi(\alpha_{oc}/\omega_{oc})$ and because this is a small fraction of unity, we have, expanding the exponential,

$$\frac{P_{\text{lost}}}{P_{\text{stored}}} = -\frac{4\pi\alpha_{oc}}{\omega_{oc}} = \frac{2\pi}{Q_0}$$

Here is, therefore, yet another interpretation of Q_0. This parameter, when large compared to unity, equals the power dissipated during the unit electrical angle, expressed as a fraction of the total stored power. For a coupling element, instead of power dissipated, we could say more properly power transmitted. We are thus brought back to the considerations of Sec. 9.4.

CHAPTER XIII

EXPONENTIAL LINES

13.1. Application of the Field Equations to the Transmission Line. The last chapter was chiefly devoted to the application of the laws of electromagnetism to lumped systems. Two types of branches having lumped parameters were discussed: the coil and the condenser, with the resistor as the limiting case of either type. It was postulated that, in a coil, the magnetic field obeys the laws of static fields for practical purposes, in spite of not having constant intensity; and the same is true for the electric field in a condenser. In both cases the field is called *quasi-stationary.*

These postulates were found useful, as they enabled us to write the product Cv for Q, the total electric flux of the condenser, and Li for Φ, the total magnetic flux of the coil. In making these substitutions, we relied implicitly upon the validity of the laws of static fields, Eqs. (279) and (280), failing which C and L would have no general meaning (Secs. 11.7 and 11.8). Thanks to these substitutions, the field equations took the form of voltage-current relations (the branch equations).

In a mathematical sense, this simplification of the field equations was obtained at the cost of neglecting one of the two correction terms which mark the change from static to dynamic conditions. These correction terms, being time derivatives, increase in importance as time variations become more and more rapid; if we are concerned with harmonic (a-c) variations, this importance may be said to grow with the frequency. This explains in general terms why lumped circuit theory has to be discarded at very high frequencies.

There is a type of structure to which the field equations may be applied without neglecting either correction term, so that the resulting voltage-current relations are not restricted to low frequencies.

This is the plane-field structure in which the current flow vector has uniform direction.

Application of the field laws to such a structure, limited to the case of two conductors (bifilar transmission line) will be considered in the following; the assumptions from which transmission line theory has been developed in earlier chapters will thus be found to be in accord with field theory.

Figure 67 represents a system of two parallel conductors. Two circular cylinders of equal radius are considered, but the

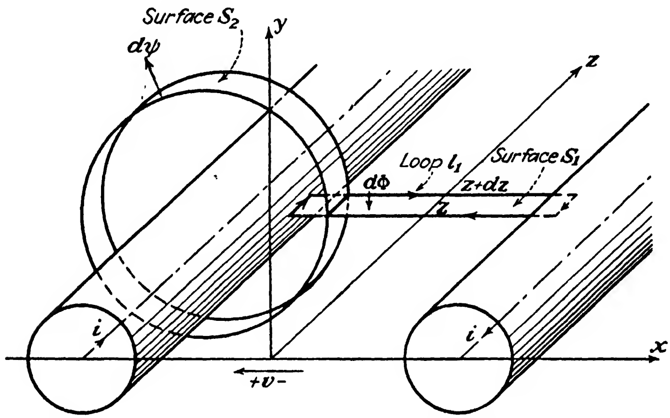


FIG. 67.—*Application of the field equations to the transmission line.*

conclusions will be applicable to cylinders of any cross section, including coaxial structures.

The static fields of such a system have already been studied; by the principle of invariance of plane fields (Sec. 12.4), we are able to extend the results of this study to the dynamic condition, provided the current is assumed to flow in the axial direction. A number of these results will be used in the derivation of the line equations, namely, the following:

1. The law of static fields

$$\oint_l E_l \, dl = 0$$

holds within the plane xy (plane of the cross section). Hence, the voltage between conductors, or *line voltage* v, is single valued, *provided it is obtained by integration along a path lying on the xy plane.*

2. There is no electric flux in the z direction (E is in the xy plane everywhere). Hence there can be no displacement current

(Sec. 12.2) in the z direction. This confirms the model of the transmission line in which the shunting branches connect *only opposite points.* Moreover, since an indefinite plane may be assimilated to a closed surface, and the total (conduction plus displacement) current through a closed surface is zero, the currents in the two conductors must be equal and opposite. If they are not, the cross-section plane must intersect a third conductor somewhere; and this must be taken into consideration.

3. The electric flux $d\Psi$ through a strip S_2 of width dz completely surrounding either conductor at z (Fig. 67) is proportional to the line voltage v at z, and the coefficient of proportionality is the same as for static conditions. Hence, if C is the distributed capacity (Sec. 11.8)

$$d\Psi = Cv\,dz \tag{358}$$

4. The magnetic flux $d\Phi$ through a strip S_1 of width dz ending at the conductor surfaces (Fig. 67) is likewise given by

$$d\Phi = Li\,dz \tag{359}$$

where L is the distributed external inductance (Sec. 11.7).

To proceed with the analysis, the first of Maxwell's equations (13) may be applied to the rectangular loop l_1, bounding S_1 (Fig. 67). Note that here, as in general when the circuitation of E is taken wholly or partly over a current path, the positive direction of l_1 must be *with* the current, so that the positive normal to S may be *with* the flux. Thus

$$\oint_{l_1} E_l\,dl = -\frac{\partial}{\partial t}(d\Phi) = -L\,dz\,\frac{di}{dt}$$

We now break up the loop integral into four parts, for the four sides of the rectangle, noting that the contribution to the integral due to the x-directed sides is dv, difference between the line voltages at z and $z + dz$. The contribution of the z-directed sides may be evaluated in terms of J_0, the current density at the conductor surface, by Ohm's law (278). Thus

$$dv + \frac{2J_0}{\gamma}\,dz = -L\,dz\,\frac{di}{dt}$$

We will base subsequent steps on the assumption that the current density J varies with time proportionally at all points of the conductor cross section. The error due to this assumption is negligible at very high frequencies. To correct for this error, an internal inductance should be added to the external value L. We let

$$J_0 = K\frac{i}{a}$$

$$R = \frac{2K}{a\gamma_1} \tag{360}$$

where K is a coefficient depending on the current distribution within the conductor, a the conductor cross-sectional area, γ_1 the conductivity, and R the resistance per unit length. Thus

$$-\frac{\partial v}{\partial z} = Ri + L\frac{\partial i}{\partial t} \tag{361}$$

We have obtained from Maxwell's equation a voltage-current relationship, the first of two differential equations which constitute the classical approach to transmission line theory. [The performance of lines in steady state may be analyzed without recourse to these equations with the help of the network assumptions (Chap. III).]

The second line equation emerges when we apply (328). Considering the pillbox-shaped volume whose lateral surface is the strip S_2, the net outgoing conduction current is di, the differential of current over dz, plus some leakage current across S_2. Hence

$$di + \iint_{S_2} J_n\,dS + \frac{\partial}{\partial t}(d\Psi) = 0$$

The second term is the leakage current; the third, the displacement current. Using Ohm's law and assuming the dielectric to have uniform leakage conductance γ_2,

$$\iint_{S_2} J_n\,dS = \iint_{S_2} E\gamma_2\,dS = \frac{\gamma_2}{\epsilon}\iint_{S_2} \epsilon E\,dS = \frac{\gamma_2\,d\Psi}{\epsilon}$$

Thus the equation becomes

$$di + \frac{\gamma_2}{\epsilon}\,d\Psi + \frac{\partial}{\partial t}(d\Psi) = 0$$

Substituting the value for $d\Psi$ given by (358), and letting

$$\frac{C\gamma_2}{\epsilon} = G \tag{362}$$

we have the final form

$$-\frac{\partial i}{\partial z} = Gv + C\frac{\partial v}{\partial t} \tag{363}$$

Actually G, the leakage conductance per unit length, is seldom given by (362) because many imperfect dielectrics do not obey Ohm's law. The actual value of G is difficult to determine except by measurement.

Regarding the signs of (361) and (363), the sign of the left-hand terms depends on the choice of a positive direction for z. This has been taken to coincide with the direction of instantaneous power flow arising from the positive values of v and i.

13.2. The Uniform Line. Wave Character of the Solution. Further manipulation of Eqs. (361) and (363) depends on the conditions of the problem. In a great many practical cases the parameters L, C, R, G are virtually constant, both with time and with the distance, henceforth designated by x. We have then a *uniform line*, and the mathematical solution of (361) and (363), considered jointly as a system of partial differential equations, is straightforward. In outline, the procedure is as follows:

1. Differentiating (361) with respect to x (now taking the place of z), (363) with respect to t, and eliminating the crossed derivative $\partial^2 i/\partial x\,\partial t$ between the two equations, we obtain

$$\frac{\partial^2 v}{\partial t^2} + 2a\frac{\partial v}{\partial t} + (a^2 - b^2)v = \Omega^2\frac{\partial^2 v}{\partial x^2} \tag{364}$$

[*Telegrapher's equation*]

where

$$a = \tfrac{1}{2}\left(\frac{R}{L} + \frac{G}{C}\right) \qquad b = \tfrac{1}{2}\left(\frac{R}{L} - \frac{G}{C}\right) \tag{365}$$

$$\Omega = \frac{1}{\sqrt{LC}} \tag{366}$$

An identical equation in i would be obtained by carrying out the preliminary differentiation in inverse order. In either case, we now have an equation in only one variable and its time and space derivatives, known as the *telegrapher's equation*.

2. It can be shown that the solution of (364) must be a function of x and t of the following form:

$$v(x, t) = X_1T_1 + X_2T_2 + X_3T_3 + \cdots + X_mT_m + \cdots \tag{367}$$

where the mth member of the summation is the product of a *function of x only*, X_m, by a *function of t only*, T_m.

The mth member of the summation itself satisfies (364); this requires X_m and T_m to have *the same form* in all the terms of the summation, the only difference arising from the values of a number of constants of integration, which vary from one term to the next.

3. The form of X_m and T_m will now be determined. Substituting

$$v = X_m T_m$$

in Eq. (364), we have

$$X_m \frac{d^2T_m}{dt^2} + 2aX_m \frac{dT_m}{dt} + (a^2 - b^2)X_m T_m = T_m \Omega^2 \frac{d^2X_m}{dx^2}$$

and dividing through by $X_m T_m$,

$$\frac{1}{T_m}\left(\frac{d^2T_m}{dt^2} + 2a \frac{dT}{dt}\right) + a^2 = \frac{\Omega^2}{X_m} \frac{d^2X_m}{dx^2} + b^2$$

The left-hand term of the above cannot be a function of anything except t, or the right-hand term of anything except x. However, we immediately see that the left-hand term cannot be a function of t because if it were, and t were to change while x remained constant, the equation would no longer be satisfied. Hence both terms are constants and equal, and we may write

$$\frac{1}{T_m}\left(\frac{d^2T_m}{dt^2} + 2a \frac{dT_m}{dt}\right) + a^2 = -k_m{}^2 \tag{368}$$

$$\frac{\Omega^2}{X_m} \frac{d^2X_m}{dx^2} + b^2 = -k_m{}^2 \tag{369}$$

4. We have succeeded in writing two homogeneous differential equations with constant coefficients in T_m and X_m. These can be solved by familiar methods. Observe, however, that $-k_m{}^2$ may be positive, negative, or complex;[1] we only know that it is a constant. Hence, k_m (the *separation constant*) may be real or complex. The solutions are

$$T_m = [A_m e^{jk_m t} + B_m e^{-jk_m t}]e^{-at} \tag{370}$$

$$X_m = C_m e^{jh_m x} + D_m e^{-jh_m x} \tag{371}$$

[1] If $-k_m{}^2$ is complex, T_m must be complex also. Our premise (367) does not imply that the individual terms must be real, even though we may require the summation to be real.

having let

$$h_m = \frac{\sqrt{k_m^2 + b^2}}{\Omega} \tag{372}$$

If the conditions of the problem require h_m and k_m to be real, we may expand (370) and (371) into sums of sines and cosines of real arguments

$$\begin{aligned} T_m &= [A_m' \cos k_m t + B_m' \sin k_m t]e^{-at} \\ X_m &= C_m' \cos h_m x + D_m' \sin h_m x \end{aligned}$$

and the mth term of (367) may be written

$$\begin{aligned} v_m(x,t) = [&A_m'C_m' \cos k_m t \cos h_m x + B_m'D_m' \sin k_m t \sin h_m x \\ &+ A_m'D_m' \cos k_m t \sin h_m x + B_m'C_m' \cos h_m x \sin k_m t]e^{-at} \end{aligned}$$

or, using the relations of trigonometry,

$$\begin{aligned} v_m(x, t) = [&A_m'' \cos (h_m x + k_m t) + B_m'' \sin (h_m x + k_m t) \\ &- C_m'' \cos (h_m x - k_m t) - D_m'' \sin (h_m x - k_m t)]e^{-at} \end{aligned} \tag{373}$$

We may interpret the above as follows: The voltage v_m, particular solution of (367), is the sum of two waves traveling in the direction of positive x with the velocity

$$\frac{k_m}{h_m} = \frac{k_m \Omega}{\sqrt{k_m^2 + b^2}} = \Omega \frac{1}{\sqrt{1 + (b/k_m)^2}} \tag{374}$$

and two waves traveling in the opposite direction with the same velocity, all four attenuated as they travel.

The velocity of propagation will be different for each particular solution unless $b = 0$. We realize without difficulty that if the total voltage is made up of component waves traveling with different velocities, the resultant wave will be distorted as it travels. In this sense, we may define a distortionless line as one for which [Eq. (365)]

$$b = \tfrac{1}{2}\left(\frac{R}{L} - \frac{G}{C}\right) = 0 \qquad \text{or} \qquad \frac{R}{L} = \frac{G}{C}$$

We thus reobtain the condition (73) of Sec. 4.3, which defines a distortionless line for the particular standpoint of steady-state operation.

It should be stressed that (373) is not the answer to all possible line problems. We have, in fact, assumed k_m to be a real, which may not actually be the case; furthermore, the number of terms making up the entire value of v and the constants appearing in each term should be determined. All this calls for some statement regarding the conditions of the problem, for example:

5. The *initial value* of v may be known all along the line. In this event

$$v(o, x) = V(x)$$

is a known function of x. Comparing this with (367), in which the constants and the values of h have been written in as unknowns and t has been made equal to zero, we would then arrive at some conclusion regarding the number of terms and the corresponding values of h. In addition, two *boundary conditions* (one for each end of the line) would have to be assigned.

6. The *boundary value* of v may be known for all values of time, the line being driven at one end by the application of a known time-variable voltage. Again comparing with (367), in which x has been made equal to zero (taking the origin of x at the driven end) we may find the number and values of the k coefficients. An additional boundary condition (for the receiving end) must be given.

7. As a particular case of boundary value problem, the line may be driven by a *steady-state harmonic voltage*. The procedure (3) may be used in this case; we would find that the summation of (367) must consist of two terms, corresponding to two complex conjugate values of $k(k_1 = \omega + ja;\ k_2 = \omega - ja)$. By a proper choice of the constants of integration, this two-term summation reduces to the required cosine function

$$v(o, t) = \hat{V} \cos \omega t$$

However, it is not necessary to go through this procedure. The line equations (361) and (363) are similar to the coil and condenser equations (337) and (341). We may, therefore, as in Sec. 12.8, let

$$\begin{aligned} v(x, t) &= Re[\mathbf{V}(x, t)] \\ i(x, t) &= Re[\mathbf{I}(x, t)] \end{aligned} \tag{375}$$

Now, if v and i are solutions of (361) and (363), we may verify that $\mathbf{V}$ and $\mathbf{I}$ are solutions of the system

$$\begin{cases} -\dfrac{\partial \mathbf{V}}{\partial x} = R\mathbf{I} + L\dfrac{\partial \mathbf{I}}{\partial t} \\ -\dfrac{\partial \mathbf{I}}{\partial x} = G\mathbf{V} + C\dfrac{\partial \mathbf{V}}{\partial t} \end{cases} \tag{376}$$

obtained by substituting **V** and **I** for v and i in the original system. Conversely, if we solve (375), the real part of the solution will be the required voltage (or current) function. There is advantage in the use of this device if the boundary value is a cosine (or sine) function, as the corresponding complex value is an exponential. In the following section the steady-state solution will be worked out in this way.

13.3. Steady-state Solution: Reappearance of Network Concepts. The steady-state solution of the line equations is generally carried out by substituting for **V** and **I** in Eqs. (375) the functions

$$\begin{aligned} \mathbf{V}(x, t) &= \mathbf{V}_x e^{j\omega t} \\ \mathbf{I}(x, t) &= \mathbf{I}_x e^{j\omega t} \end{aligned} \tag{377}$$

where $\mathbf{V}_x$ and $\mathbf{I}_x$ are functions of x only. Since it is possible to satisfy (376) by this position and at the same time meet the *boundary condition* of a steady-state harmonic signal, *i.e.*,

$$\mathbf{V}(0, t) = \mathbf{V}e^{j\omega t}$$

we are justified in concluding that the position (377) is valid. It is important, however, to realize the full meaning of this position. It means that because we *impress* a harmonic voltage at one point of the line, both voltage and current at all other points will be harmonic functions of time. It was stated without proof in Sec. 1.7 that this was true of all linear systems. Now we have proof of this all-important premise in the case of the transmission line. It should be understood that the premise is not self-evident; it is not true of any other periodic function of time. If a square wave is impressed at one end of the line, it will not, in general, appear as a square wave at the other end; it will be distorted, unless condition (73) is satisfied. Variations of α and β/ω with the frequency (Sec. 4.2) cause distortion only when *several frequencies are transmitted at once*, in which case the impressed signal is not harmonic. Harmonic signals are the only *periodic* signals for which all lines are distortionless. This is true also of those nonperiodic signals which can be written

$$v(t) = Re(\mathbf{V}e^{pt})$$

where p may be real, complex, or imaginary. (The harmonic signal corresponds to an imaginary value of p.)

If the values (377) are written into the system (376), the time function $e^{j\omega t}$ drops out and the equations reduce to[1]

$$\begin{aligned} -\mathbf{V}_x' &= Z\mathbf{I}_x \\ -\mathbf{I}_x' &= Y\mathbf{V}_x \end{aligned} \tag{378}$$

where

$$Z = R + j\omega L \qquad Y = G - j\omega C \tag{379}$$

are the *series impedance* and the *shunt admittance* (Sec. 3.2). Taking the x derivative of the first line and comparing with the second, we reobtain (369) in the form

$$\mathbf{V}_x'' - ZY\mathbf{V}_x = 0 \tag{380}$$

An equation in $\mathbf{I}_x$, otherwise identical to the above, is obtained by proceding in inverse order.

It should be noted that (380) does not contain the arbitrary separation constant k_m which appears in (369). This constant has already been assigned by the position (377). In fact, comparing this position with the general solution (367), we would find that all the terms drop out except one, for which we must have

$$T_m = e^{j\omega t}$$

Hence, comparing with (370)

$$B_m = 0 \qquad A_m = 1 \qquad k_m = \omega - ja \tag{381}$$

The solution of (380) is

$$\mathbf{V}_x = Ae^{\sqrt{ZY}x} + Be^{-\sqrt{ZY}x} \tag{382}$$

Comparing this with (371), solution of (369), we find

$$h_m = \frac{\sqrt{k_m^2 + b^2}}{\Omega} = -j\sqrt{ZY} = -\sqrt{-ZY}$$

Recalling positions (365), (366), and (379), it may be verified that $k_m = \omega - ja$ satisfies the above. We may put (382) in a convenient form by the following positions:

[1] As there are no more time derivatives, we may use without ambiguity the notation

$$f' = \frac{df}{dx} \qquad f'' = \frac{d^2f}{dx^2} \cdots$$

$$K = \sqrt{AB} \qquad \rho' = \tfrac{1}{2} \ln \frac{B}{A} \tag{383}$$
$$\theta = \sqrt{ZY}$$

noting that θ has been defined earlier (Sec. 3.4) as the *transfer constant per unit length* of the line. Thus

$$\mathbf{V}_x = K \cosh (\rho' - \theta x) \tag{384}$$

$\mathbf{I}_x$ may now be obtained by differentiation, using the first of Eqs. (378)

$$\mathbf{I}_x = \frac{K\theta}{Z} \sinh (\rho' - \theta x) \tag{385}$$

Dividing (384) by (385), and letting

$$\frac{Z}{\theta} = \sqrt{\frac{Z}{Y}} = Z_0 \qquad \text{(characteristic impedance)} \tag{386}$$

we have

$$\frac{\mathbf{V}_x}{\mathbf{I}_x} = Z_0 \coth (\rho' - \theta x)$$

Finally, letting,

$$\rho = \rho' - \theta l$$
$$\left[\frac{\mathbf{V}_x}{\mathbf{I}_x}\right]_{(x=l)} = Z_r \quad \text{(receiver impedance)}$$
$$\left[\frac{\mathbf{V}_x}{\mathbf{I}_x}\right]_{(x=0)} = Z_i \quad \text{(input impedance)}$$

we reobtain (99) and (100), namely

$$Z_r = Z_0 \coth \rho$$
$$Z_i = Z_0 \coth (\rho + \theta l)$$

(*Note.* For $l = \infty$, $Z_i = Z_0$, characteristic impedance)

We need not go any further, because the two methods of approach, based on the assumptions of network theory and on Maxwell's equations, respectively, have brought us to the same point; from here on they are no longer distinct. We will therefore leave the subject of the uniform transmission line and go on to lines with variable parameters.

13.4. The Exponential Line. The most general case in which all four distributed parameters of a line vary with x cannot be conveniently handled. We must simplify the problem by assuming a lossless line ($R = G = 0$); losses may be evaluated approximately on the basis of the lossless current distribution (Sec. 7.5).

This preliminary assumption leaves only two variable parameters, L and C, of which only one may vary independently. We know that in a lossless two-dimensional system (326)

$$\Omega = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\epsilon\mu}} \tag{387}$$

Consequently, the equation for the lossless disuniform line need not include more than one variable coefficient. To write the equation, we use (378) as the starting point. Differentiating the first line and noting that Z and Y are now functions of x,

$$-\mathbf{V}_x'' = Z'\mathbf{I}_x + Z\mathbf{I}_x'$$

or, substituting for $\mathbf{I}_x$ and $\mathbf{I}_x'$,

$$\mathbf{V}_x'' - \frac{Z'}{Z}\mathbf{V}_x' - ZY\mathbf{V}_x = 0 \tag{388}$$

We may readily evaluate the coefficients, making at the same time the following positions:

$$\frac{Z'}{Z} = \frac{j\omega L'}{j\omega L} = \frac{L'}{L} = 2\gamma$$

$$ZY = (j\omega L)(j\omega C) = -\omega^2 LC = -\left(\frac{\omega}{\Omega}\right)^2 = -\beta^2$$

The quantity β was defined previously as the phase constant, which for a lossless line is equal to the transfer constant θ divided by j and may also be written $\beta = 2\pi/\lambda$. The new quantity γ, in general a function of x, may be defined in more than one way. Recalling (387), the following may be easily verified:

$$2\gamma = \frac{L'}{L} = -\frac{C'}{C} = \frac{R_0'}{R_0} \tag{389}$$

where, as usual

$$R_0 = \sqrt{\frac{L}{C}}$$

is the characteristic impedance, now a real and variable from point to point. Equation (388) and a similar equation for current may now be written in the form

$$\left\{\begin{aligned} \mathbf{V}_x'' - 2\gamma(x)\mathbf{V}_x' + \beta^2\mathbf{V}_x &= 0 \\ \mathbf{I}_x'' + 2\gamma(x)\mathbf{I}_x' + \beta^2\mathbf{I}_x &= 0 \end{aligned}\right. \tag{390}$$

[*Equations for the lossless disuniform line*]

We will not carry through the integration of (390) in general. Only one type of disuniform line has been found to have useful properties—the *exponential line,* for which γ is a constant and Eq. (390) may be integrated immediately.

An exponential line may be defined as one in which the characteristic impedance increases or decreases exponentially with x, the distance from a given point of the line. We have, accordingly,

$$L(x) = L(0)e^{2\gamma x} \qquad C(x) = C(0)e^{-2\gamma x} \qquad R_o(x) = R_o(0)e^{2\gamma x} \tag{391}$$

where γ may be positive or negative. The above is evidently in agreement with the definition of γ given by (389). γ may now be called the *decrement* of the distributed capacity (or the *increment* of the inductance).

Let us now integrate the second of Eqs. (390), which may be written simply

$$\mathbf{I}_x'' + 2\gamma\mathbf{I}_x' + \beta^2\mathbf{I}_x = 0 \tag{392}$$

This is a homogeneous differential equation, similar to (368). Its solution is

$$\mathbf{I}_x = [Ae^{\sqrt{\gamma^2-\beta^2}\,x} + Be^{-\sqrt{\gamma^2-\beta^2}\,x}]e^{-\gamma x} \tag{393}$$

Note that for $\gamma = 0$ (uniform line), the above reduces to the sum of a sine and a cosine of the real argument βx, or of the hyperbolic sine and cosine of the imaginary argument $j\beta x = \theta x$. The most convenient form of (393) for further manipulation is

$$\mathbf{I}_x = Ke^{-\gamma x} \sinh\left(\rho' - x\sqrt{\gamma^2 - \beta^2}\right) \tag{394}$$

obtained by making

$$K^2 = -AB \qquad \rho' = \tfrac{1}{2}\ln\left(-\frac{A}{B}\right)$$

The integration constants K and ρ' must be determined by assigning two boundary conditions. Usually these include a relation between $\mathbf{I}_x$ and $\mathbf{V}_x$ for $x = l$ (determined by the load impedance), and the value of $\mathbf{I}_x$ or $\mathbf{V}_x$ for $x = 0$ (the signal value). We find, however, that the *ratio* of the two variables is determined everywhere by a single boundary condition, the load impedance. We may therefore proceed as in the case of the uniform line, by computing $\mathbf{V}_x$ from $\mathbf{I}_x$, hence their ratio, and setting this equal to the load impedance for $x = l$. Using (394) and the second equation of (378),

$$\mathbf{V}_x = -\frac{\mathbf{I}_x'}{Y} = j\frac{\mathbf{I}_x'}{\omega C}$$

we obtain

$$\mathbf{V}_x = \frac{jK}{\omega C}[-\sqrt{\gamma^2 - \beta^2}\, e^{-\gamma x} \cosh(\rho' - x\sqrt{\gamma^2 - \beta^2}) \\ - \gamma e^{-\gamma x} \sinh(\rho' - x\sqrt{\gamma^2 - \beta^2})] \quad (395)$$

Now, dividing the above by (394), we have

$$\frac{\mathbf{V}_x}{\mathbf{I}_x} = \frac{1}{j\omega C}[\sqrt{\gamma^2 - \beta^2} \coth(\rho' - x\sqrt{\gamma^2 - \beta^2}) + \gamma] \quad (396)$$

Noting that

$$\frac{1}{j\omega C} = \frac{\sqrt{L/C}}{j\omega\sqrt{CL}} = \frac{R_o}{j\beta}$$

$$\sqrt{\gamma^2 - \beta^2} = j\beta\sqrt{1 - \left(\frac{\gamma}{\beta}\right)^2}$$

(396) may be put in the form

$$\frac{\mathbf{V}_x}{\mathbf{I}_x} = R_o\left[\sqrt{1 - \left(\frac{\gamma}{\beta}\right)^2} \coth\left\{\rho' - j\beta_x\sqrt{1 - \left(\frac{\gamma}{\beta}\right)^2}\right\} - j\frac{\gamma}{\beta}\right] \quad (397)$$

where it should be noted that

$$\frac{\gamma}{\beta} = \frac{\gamma\lambda}{2\pi}$$

depends only on the frequency, while R_o is a function of x. We may now make $x = l$ (length of line) in (397) and equate to Z_r, load impedance; then make $x = 0$ and equate to Z_i, input impedance. We thus obtain the following equations, characterizing the impedance transforming action of the exponential line:

$$Z_r = R_{or}\left[A \coth \rho - j\frac{\gamma}{\beta}\right] \quad (398)$$

$$Z_i = R_{or}\left[A \coth(\rho + j\beta l A) - j\frac{\gamma}{\beta}\right] e^{-2\gamma l} \quad (399)$$

[*Impedance transforming action of the exponential line*] where $A = \sqrt{1 - (\gamma/\beta)^2}$ $\rho = \rho' - j\beta l A$ (defined by Z_r)

R_{or} = characteristic impedance of line at load

R_{oi} = characteristic impedance of line at input

$\gamma = 1/2l \ln R_{or}/R_{oi}$ = line decrement

l = line length

$\gamma/\beta = \gamma\lambda/2\pi$

Z_r = load impedance

Z_i = input impedance

Equations (398) and (399) enable us to devise an *equivalent circuit* for the exponential line. Any two lossless networks having the same impedance transforming action are equivalent in all respects.[1] Hence, the required equivalent network may be obtained by analyzing the steps involved in the transformation due to the exponential line and reproducing them one by one.

Let

$$Z_r' = R_{or}A \coth \rho = Z_r + j\frac{\gamma}{\beta}R_{or}$$

This impedance may be obtained by adding a reactance

$$X = \frac{\gamma}{\beta}R_{or}$$

to the load impedance. X is the reactance of a condenser of capacity

$$C = \frac{1}{\gamma R_{or}\Omega} = -\frac{10^{-2}}{3\gamma R_{or}}\ \mu f$$

This capacity has a *negative* value when γ is positive, in which case and for a particular ω it may be replaced by an inductance of value

$$L = \frac{\gamma R_{or}\lambda}{4\pi^2\Omega}$$

If we consider Z_r' as the load impedance of a *uniform* lossless line of characteristic impedance

$$Z_o' = AR_o$$

and of length

$$l' = Al$$

[1] The two networks will receive, hence transmit, the same power. Therefore, the load voltage and current will be the same in the two cases. The input values are also equal, as the input impedances are equal. Thus, all four values of voltage and current are the same in both cases.

then the input impedance of this line will be

$$Z_i' = AR_o \coth (\rho + j\beta Al)$$

Adding to Z_i' the reactance

$$-X = -\frac{\gamma}{\beta} R_{or}$$

due to the condenser

$$C = \frac{1}{\gamma R_{or}\Omega}$$

we obtain, as may be verified by inspection of (399), the impedance

$$Z_i'' = Z_i e^{2\gamma l}$$

Finally, if Z_i'' is connected across the output of a perfect transformer (Sec. 14.6) of turn ratio

$$t = \frac{N_2}{N_1} = e^{\gamma l}$$

the input impedance of the transformer is Z_i.

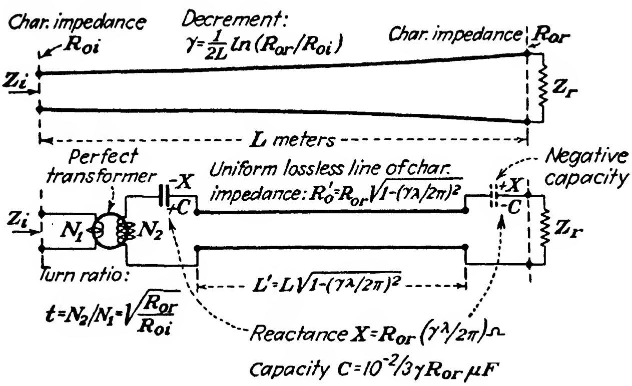


FIG. 68.—*Equivalent circuit of a lossless exponential line for the wavelength λ. The circuit shown applies to a divergent line ($\gamma > 0$). For $\gamma < 0$, the positive and negative capacities change places and the turn ratio N_2/N_1 becomes less than unity.*

A simpler circuit, equivalent to the line only with regard to the input impedance, is given in Fig. 69.

The equivalent circuit (Fig. 68) consists of two condensers, one positive and one negative, a lossless uniform line whose characteristics change with frequency, and a perfect transformer.

This circuit cannot be physically realized (or there would be no point in using the exponential line), but it may be used conveniently in analyzing the performance of the actual system.

Consider the case of a matching load. We must interpret this to mean that the load impedance is equal to the characteristic value *at the load end*, or

$$Z_r = R_{or}$$

Let us see what happens when the frequency is so high that

$$\gamma\lambda \ll 2\pi \tag{400}$$

In this event the two series reactances are negligible and the uniform line is matched, hence without effect on the impedance; the equivalent circuit reduces to the perfect transformer alone.

We conclude that, if the condition (400) holds, the exponential line solves the transformer problem (Sec. 8.1). Sections of exponential line may therefore be used to couple together lines of different characteristic impedances. Transmission through this system will be comparatively uniform for wavelengths below the value

$$\lambda_{\max} = \epsilon \frac{4\pi l}{\ln R_{or}/R_{oi}} \tag{401}$$

where ϵ is a small number which depends on the allowable fluctuation of the insertion loss. This dependence will be investigated (Sec. 13.7). Taking 0.1 as a representative value for ϵ, we find

$$\lambda_{\max} = \frac{1.25l}{\ln R_{or}/R_{oi}}$$

or

$$\frac{l}{\lambda_{\max}} = 0.8 \ln \frac{R_{or}}{R_{oi}} \tag{402}$$

and conclude that the length of the exponential section, measured in wavelengths for the lowest operating frequency, must be of the same order of magnitude as the natural logarithm of the impedance ratio.

★13.5. Analysis of Exponential Line Performance. By resolving the exponential line into its equivalent circuit (Fig. 68), we have been able to show that such a line approaches the characteristics of a perfect transformer (Sec. 14.6), if the frequency is sufficiently high. Let us now follow up that preliminary survey by a more detailed analysis.

The performance of a lossless exponential line, as that of any lossless network, is fully described by the value of its input impedance, since this, with the known source impedance, determines the transmitted power (Sec. 5.6). Expressions for the input impedance of the exponential line, suitable for numerical computation, may be derived from (398) by elimination of coth ρ. These expressions are given below. For convenience, a new notation, reducing to a minimum the number of variables, will be used. The new symbols are

$z_r = Z_r/R_{or}$	load impedance number (ratio of the load impedance to the characteristic impedance of the exponential line *at the load*)	
$z_i = Z_i/R_{oi}$	input impedance number (as above, referred to the input)	(403)
$\Gamma = \gamma l = \ln \sqrt{R_{or}/R_{oi}}$	nominal voltage or current ratio, in nepers (the voltage or current ratio has this value when the line operates as a perfect transformer)	
$\phi = \beta l = 2\pi l/\lambda$	line angle	

A single general expression could be written. However, this expression becomes indeterminate at the cutoff frequency, for which $\phi = \Gamma$. The limit of z_i for $\phi \rightarrow \Gamma$ has been evaluated and is given separately. A distinction is also made between the two cases $\phi > \Gamma$ and $\phi < \Gamma$ in order to avoid functions of imaginary arguments. Thus

Above cutoff ($\phi > \Gamma$):

$$z_i = \frac{z_r \sqrt{\phi^2 - \Gamma^2} + (z_r\Gamma + j\phi) \tan \sqrt{\phi^2 - \Gamma^2}}{\sqrt{\phi^2 - \Gamma^2} - (\Gamma - j\phi z_r) \tan \sqrt{\phi^2 - \Gamma^2}} \tag{404}$$

At cutoff ($\phi = \Gamma$):

$$z_i = \frac{z_r + z_r\Gamma + j\Gamma}{1 - \Gamma + jz_r\Gamma} \tag{405}$$

Below cutoff ($\phi < \Gamma$):

$$z_i = \frac{z_r \sqrt{\Gamma^2 - \phi^2} + (z_r\Gamma + j\phi) \tanh \sqrt{\Gamma^2 - \phi^2}}{\sqrt{\Gamma^2 - \phi^2} - (\Gamma - jz_r\phi) \tanh \sqrt{\Gamma^2 - \phi^2}} \tag{406}$$

[*Expressions for the input impedance of a lossless exponential line*]

Graphical evaluation of z_i is also possible; it can be based on the methods developed for the uniform line (Fig. 33) and on the equivalent circuit (Fig. 68). The latter may be simplified

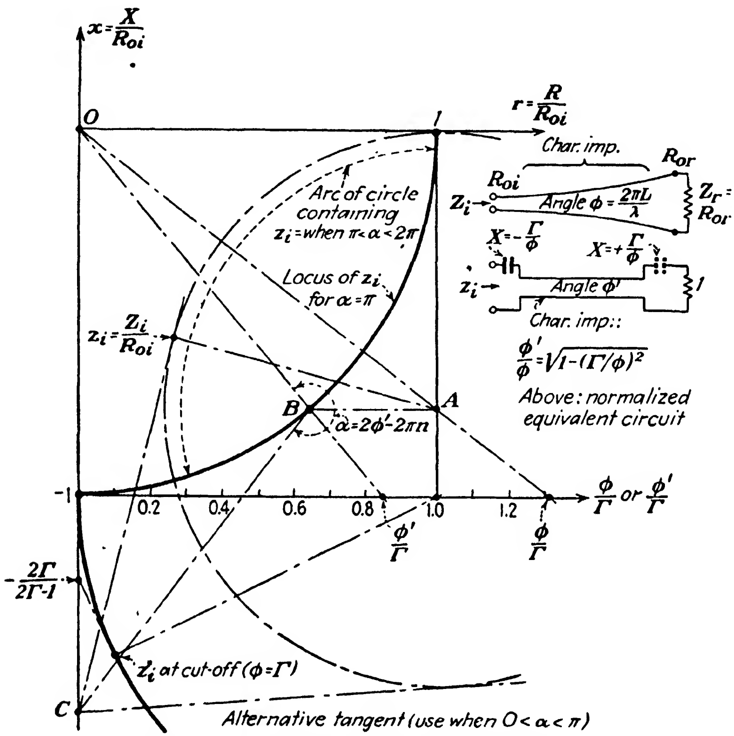


FIG. 69.—*Impedance transforming action of the lossless exponential line: divergent line with matching load.* Construction steps:

1. Measure ϕ/Γ on ϕ/Γ scale. In example, $\phi = 3.51$ radius, $\Gamma = 2.68$.
2. Draw OA, AB ($=$ to r axis), OB. Obtain ϕ'/Γ. Draw circle with center A tangent to r axis.
3. Compute ϕ' and $\alpha = 2\phi' - 2\pi n$ (n = number of times 2π is contained in $2\phi'$). Measure α clockwise from BO. Obtain C.
4. Draw tangent from C to circle with center A. Point of tangency is z_i, input-impedance number. Sector of z plane inside circle of unit radius about origin contains z_i when $\pi < \alpha < 2\pi$. This removes ambiguity in choice of tangents. z_i is always 1 for $\alpha = 0$.

somewhat if we do not require it to be equivalent to the line in every respect, but only as far as the input impedance is concerned. We may then use the device of eliminating the perfect transformer and multiplying all impedances on the secondary or load side of this by the square of the transformer ratio, or R_{oi}/R_{or}. The resulting network is that of Fig. 69.

The constructions of Fig. 69 (for $\Gamma > 0$) and Fig. 70 (for $\Gamma < 0$) are based on this network. They can be used to determine the operation of the line with a matching load ($z_r = 1$) above cutoff and at cutoff (below cutoff the operation is of no practical interest). Above cutoff, the characteristic impedance and

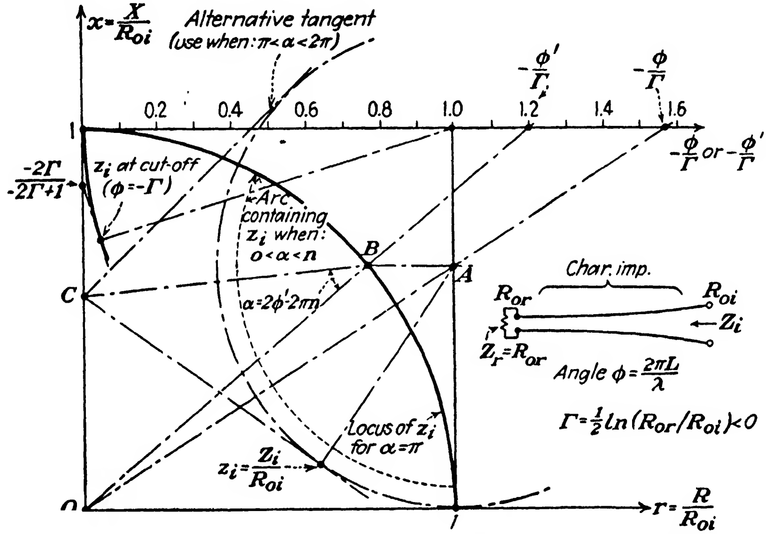


FIG. 70.—*Impedance transforming action of the lossless exponential line: convergent line with matching load* (construction steps as in Fig. 69).

the angle of the equivalent uniform line are real numbers. The constructions determine these values and apply the principle of Fig. 33 with some modifications due to the addition of the series capacities.

Polar plots of the input impedance of a step-up ($\Gamma > 0$) and a step-down ($\Gamma < 0$) exponential line have been traced out by such graphical methods and are shown in Fig. 71. As the frequency is decreased, the impedance point describes ever-widening circles, all of which pass through the nominal value $Z_i = R_{oi}$, until finally the impedance becomes very nearly a pure reactance before swinging back to the real axis at the point $Z_i = R_{or}$ (for d-c).

The fluctuation of transmitted power is correlated with the frequency of transmission in Sec. 13.7 at the end of this chapter.

13.6. The Exponential Stub Line. A section of uniform line, short-circuited at one end, may be used as a reactive two-pole whose susceptance may vary from large positive to large negative values, depending on the length or on the wavelength. This use was discussed in Sec. 10.4; sections of line used in this manner are often called *transmission line resonators* or *stub lines.* They may be of the parallel or coaxial type.

While no mention of the use of disuniform lines as resonators appears to have been made in the literature as yet, it may be

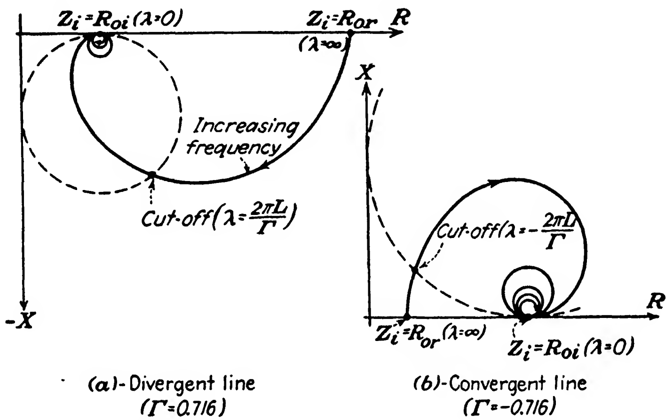


FIG. 71.—*Polar plots of input impedance for the matched exponential line.*

shown that the exponential line offers some advantages over the uniform line in this connection.

Any valid criterion of comparison must be based primarily on the value of *resonant impedance,* as it is generally called, although the value in question is a resistance and its inverse, the conductance, is actually more significant because the stub is connected in shunt with other branches. Secondarily, such aspects as physical dimensions, freedom from interference and radiation, and structural difficulties require consideration.

The conductance of the uniform stub line at resonance was evaluated in Sec. 10.9. We must now extend this evaluation to the case when the decrement γ has a generic value, thus including in the analysis both the convergent and divergent types of exponential stub lines, as well as the uniform stub line.

The evaluation will be based on a method often used to determine the losses of resonators or oscillating systems. The current and voltage distributions are worked out on the assumption that the system is lossless; the loss, based on these current and voltage values, is then obtained by integration over the entire system.

The general expression of I_x for the lossless exponential line (Eq. 394) is taken as the starting point. We further specify that $I_x = 0$ for $x = 0$. This means that x, distance along the line in the direction of energy transmission, is counted from a current node. Thus the general expression

$$\mathbf{I}_x = Ke^{-\gamma x} \sinh (\rho' - x\sqrt{\gamma^2 - \beta^2}) \tag{394}$$

becomes

$$\mathbf{I}_x = -Ke^{-\gamma x} \sinh x\sqrt{\gamma^2 - \beta^2} \tag{407}$$

having set $\rho' = 0$ in accordance with the condition $\mathbf{I}_x = 0$ at $x = 0$.

Only one constant of integration remains to be determined. To obtain it, we assign the value of $\mathbf{I}_x$ at the first antinode in the direction of increasing x. This antinode, in a stub line, corresponds to the shorting connection. We will let

$$x = l_s \qquad \mathbf{I}_x = \mathbf{I}_s$$

at the first antinode. Of these values, the second is set arbitrarily, the first is a function of γ and β.

Hence we have

$$I_s = -Ke^{-\gamma l_s} \sinh L_s\sqrt{\gamma^2 - \beta^2}$$

and, eliminating K between (407) and the above,

$$I_x = I_s e^{\gamma(l_s - x)} \frac{\sinh x\sqrt{\gamma^2 - \beta^2}}{\sinh l_s\sqrt{\gamma^2 - \beta^2}} \tag{408}$$

The next step is to express $(\sinh l_s\sqrt{\gamma^2 - \beta^2})$ in terms of γ and β only, thus simplifying the remaining manipulation and obtaining at the same time useful information on the length of the resonant stub. On a uniform line this is a quarter wavelength, or an odd multiple of it; on the exponential stub line, the length depends on γ.

To obtain the desired relation we must specify that the point $x = l_s$ is actually a current antinode or that the voltage at this point must be zero. An expression for the voltage $\mathbf{V}_x$, consistent with (408), may be obtained by differentiation just as the generic voltage distribution (407) was obtained from (394). Thus

$$\mathbf{V}_x = -\mathbf{I}_s \frac{R_o e^{\gamma(l_s - x)}}{j\beta} \frac{\sqrt{\gamma^2 - \beta^2} \cosh x \sqrt{\gamma^2 - \beta^2} - \gamma \sinh x \sqrt{\gamma^2 - \beta^2}}{\sinh l_s \sqrt{\gamma^2 - \beta^2}} \tag{409}$$

The denominator of the above must vanish for $x = l_s$, according to our premise. Hence

$$\sqrt{\gamma^2 - \beta^2} \cosh l_s \sqrt{\gamma^2 - \beta^2} = \gamma \sinh l_s \sqrt{\gamma^2 - \beta^2}$$

which may be written in the following forms

$$\cosh l_s \sqrt{\gamma^2 - \beta^2} = \frac{\gamma}{\beta} \tag{410}$$

$$\sinh l_s \sqrt{\gamma^2 - \beta^2} = \frac{\sqrt{\gamma^2 - \beta^2}}{\beta} \tag{411}$$

and, introducing the notation of Eqs. (403),

$$\cosh \sqrt{\Gamma_s^2 - \phi_s^2} = \frac{\Gamma_s}{\phi_s} \qquad (\Gamma_s > 1)$$

$$\cos \sqrt{\phi_s^2 - \Gamma_s^2} = \frac{\Gamma_s}{\phi_s} \qquad (\Gamma_s < 1) \tag{412}$$

[*Relation between exponent and line angle in the exponential stub*]

where $\Gamma_s = \gamma l_s = \ln \sqrt{R_{or}/R_{oi}}$

$\phi_s = 2\pi l_s/\lambda$ (413)

R_{os} = characteristic impedance at shorted end

R_{oi} = characteristic impedance at open end

l_s = length of stub

From Eqs. (408) and (411), we now obtain the following expression for $\mathbf{I}_x$:

$$\mathbf{I}_x = \beta I_s e^{\gamma(l_s - x)} \frac{\sinh x \sqrt{\gamma^2 - \beta^2}}{\sqrt{\gamma^2 - \beta^2}} \tag{414}$$

The above determines the current distribution in a lossless line when the current at the shorted end is I_s. If we hold $\mathbf{I}_s$ fixed and imagine the line to have distributed resistance, the current values will be affected slightly. A small current will now flow at the current node ($x = 0$) to make up for the dissipation, but obviously the loss due to this *added* current is a negligible fraction of the entire value. To clarify the point, assume the current to be 10 amp. at the short-circuited end at all times. If the line were lossless, the current would vary from 0 to 10 amp. as we go from $x = 0$ to $x = l_s$. In the actual dissipative line, the current will vary from, say, 0.1 ma. at the input to 10 amp, holding the latter value deliberately constant. The values of joule loss due to the two current distributions are evidently the same for all practical purposes.

In evaluating the losses, we will neglect dissipation in the dielectric and assume the distributed resistance R to have uniform value (Sec. 7.2). The loss is then given by the integral

$$P_s = R \int_0^{l_s} \frac{I_x^2}{2}\, dx \tag{415}$$

Carrying out the integration, we obtain

$$P_s = -\frac{RI_s^2}{8(\gamma - \beta^2)} \Big[2\sqrt{\gamma^2 - \beta^2} \sinh l_s \sqrt{\gamma^2 - \beta^2} \cosh l_s \sqrt{\gamma^2 - \beta^2} + \gamma\,(\cosh^2 l_s \sqrt{\gamma^2 - \beta^2} + \sinh^2 l_s \sqrt{\gamma^2 - \beta^2}) - \frac{\beta^2}{\gamma} - \left(\gamma - \frac{\beta^2}{\gamma}\right) e^{2\gamma l_s} \Big]$$

Using (410) and (411), the above simplifies to

$$P_s = \frac{RI_s^2}{8\gamma}\left(e^{2\gamma l_s} - 4\frac{\gamma^2}{\beta^2} - 1\right) \tag{416}$$

This value of the loss in the resonant stub will be used to find the stub conductance. We will again make use of the fact that the voltage and current distributions are not affected by the losses if the maximum values are assigned. The voltage distribution has a point of maximum (antinode) at $x = 0$; this maximum value may be obtained from (409) and (411), as follows:

$$\mathbf{V}_0 = jI_sR_{oi}e^{\gamma l_s}$$

Noting that (413)

$$e^{\gamma l_s} = \sqrt{\frac{R_{or}}{R_{oi}}}$$

the above may be written

$$\mathbf{V}_0 = jI_sR_{om} \tag{417}$$

where

$$R_{om} = \sqrt{R_{oi}R_{or}}$$

is the geometric mean of the end values of characteristic resistance or, which is the same thing, the value of R_o at the center point. Now, the conductance is given by

$$G_s = \frac{2P_s}{\hat{V}_0^2} = \frac{R}{4R_{om}^2\gamma}\left(e^{2\gamma l_s} - 4\frac{\gamma^2}{\beta^2} - 1\right) \tag{418}$$

To express G_s in terms of easily recognized quantities, we may define the *mean* Q_s of the exponential stub as the Q_s of a uniform stub for which both r and R_o are those of the center point of the exponential stub. By this definition, we have

$$Q_{sm} = \frac{1}{d_{sm}} = \frac{\omega L_m}{R} = \frac{\omega\sqrt{L_m/C_m}\sqrt{L_mC_m}}{R} = \frac{\omega R_{om}}{\Omega r} \tag{419}$$

(The subscript m denotes geometric mean value, or value at the center point.) Recalling that at high frequencies (Sec. 7.2)

$$\frac{\omega}{\Omega} = \beta$$

we have finally

$$Q_{sm} = \frac{\beta R_{om}}{R} \tag{420}$$

Using the above, as well as Eqs. (413), the expression for the stub conductance takes the form

$$G_s = \frac{1}{Q_{sm}R_{om}}\left[-\frac{\Gamma_s}{\phi_s} + \frac{\phi_s}{\Gamma_s}\frac{e^{2\Gamma} - 1}{4}\right] \tag{421}$$

[*Conductance of the exponential stub at resonance*]

where Q_{sm} = value of Q_s at center point (geometric mean value)
R_{om} = characteristic impedance at center point

Equation (421) is valid for all values of Γ_s, including $\Gamma_s = 0$, for which value we should reobtain the conductance of the uniform stub at resonance, or (248)

$$G_s = \frac{\pi}{4Q_sR_o}$$

We cannot check this by simple substitution; the limit of (421) when Γ_s tends to zero must be evaluated. This may be done by expanding $e^{2\Gamma_s}$ in a series; for Γ_s small this expansion reduces to

$$e^{2\Gamma_s} = 1 + 2\Gamma_s$$

Noting, moreover, that for the uniform line at resonance

$$\phi_s = \frac{\pi}{2}$$

we have from (105), in accordance with previous results,

$$G_s = \frac{1}{Q_s R_o}\left[\frac{\pi}{2}\frac{1 + 2\Gamma_s - 1}{4\Gamma_s}\right] = \frac{\pi}{4Q_s R_o}$$

The results of the foregoing analysis are implicit in Eqs. (412) and (421). For each value of the exponent Γ_s, positive or negative, (412) gives the corresponding line angle ϕ_s for resonance (more precisely, the shortest line angle, assuming that the stub includes only one voltage node and one current node, with a node at each end). A plot of ϕ_s vs. Γ_s appears on Fig. 72 and indicates that convergent stubs, for which the characteristic impedance is lowest at the short, are longer than uniform stubs for the same wavelength; with an exponent of -2.12 the resonant stub length is $\lambda/2$ instead of $\lambda/4$. On the other hand, a *divergent* stub with $\Gamma_s = 1.44$ has a resonant length of $\lambda/8$.

Substituting for each value of Γ_s the corresponding ϕ_s in Eq. (421), a plot of G_s vs. Γ_s has been obtained, and this also appears on Fig. 72. It shows that if we compare stubs whose mean characteristics (Q_s and R_o at the center point) are the same, the conductance is lower for divergent than for uniform stubs. The resonant impedance of a divergent stub with $\Gamma_s = 2.02$ is twice that of a uniform stub with the same mean values.

The characteristic impedance of stubs shunting a coaxial line is usually made equal to that of the line itself. If it were made greater, undesirable discontinuities would result at the junction. Assuming the line to have a 70-ohm characteristic impedance, we may therefore make a valid comparison between a uniform 70-ohm stub shunting the line and a divergent stub having a characteristic impedance of 70 ohms at the input and 210 ohms at the short (for corresponding dimensions, see Fig. 73). Let us find the resonant length and conductance in the two cases.

Let $Q_s = 1{,}000$ for both stubs at the input end.

We have, for the divergent stub,

$$\Gamma = \tfrac{1}{2} \ln \tfrac{210}{70} = \ln \sqrt{3} = 0.55$$
$$Q_{sm} = Q_s \sqrt{3} = 1{,}730$$
$$R_{om} = R_o \sqrt{3} = 121 \text{ ohms}$$

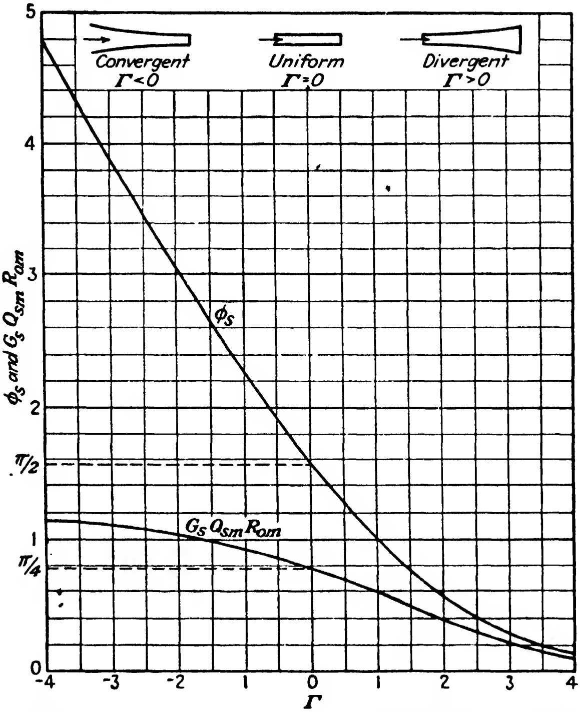


FIG. 72.—*Plots of resonant angle and resonant conductance of stubs against stub exponent* Γ.

Note:

$\Gamma = \frac{1}{2} \ln \frac{R_{or}}{R_{oi}}$ (R_{or}, characteristic impedance at short; R_{oi}, at input)

$\phi_s = 2\pi l_s/\lambda$ (l_s, length of stub at resonance; λ, wavelength)

G_s = conductance of stub at resonance

Q_{sm} = mean $Q_s = \omega L/R$ at center of stub

R_{om} = mean $R_o = \sqrt{L/C}$ at center of stub

From the plot (Fig. 72)

$$G_s Q_{sm} R_{om} = 0.69$$

Hence
$$\phi_s = 1.24 \text{ rad.}$$
$$G_s = \frac{0.69 \times 10^6}{121 \times 1{,}730} = 3.3\ \mu\text{mhos}$$

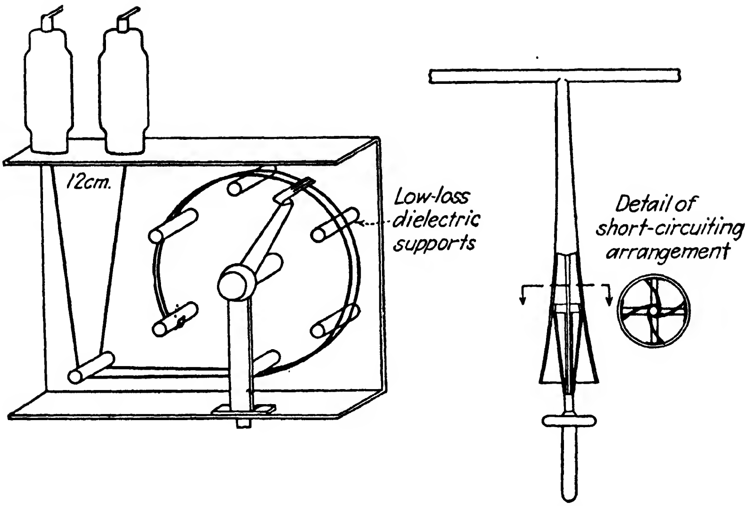


FIG. 73.—*Examples of exponential-stub application.*

a. Parallel convergent stub for

$\lambda = 5$ m.

(to be connected across widely spaced terminals). Numerical examples:

(1) Wire radius (uniform):

$$a = 4 \times 10^{-3} \text{ m.}$$

(2) Half distance between centers:
At input, $b_i = 6 \times 10^{-2}$ m.
At short, $b_r = 5 \times 10^{-3}$ m.

(3) Characteristic impedance:

$$R_{oi} = 408 \text{ ohms}$$
$$R_{or} = 82.8 \text{ ohms}$$

(4) Characteristic impedance at center point:

$$R_{om} = \sqrt{408 \times 82.8} = 184 \text{ ohms}$$

(5) Q_s at center point

$$Q_{sm} = 4{,}280 \frac{R_{om}}{(1/a + 1/b)\sqrt{\lambda}} = 1{,}990$$

(6) Exponent Γ:

$$\Gamma = -\frac{1}{2}\ln\frac{408}{82.8} = -0.7976$$

(7) From plot (Fig. 72):
$\phi_s = 2.12 \quad G_sQ_{sm}R_{om} = 0.9$

(8) Resonant length:
$L_s = 2.12\lambda/2\pi = 1.69$ m.

(9) Resonant impedance:
$R_s = Q_{sm}R_{om}/0.9 = 406{,}000$ ohms

b. Coaxial divergent stub for $\lambda = 5$ m. (to be connected across 70-ohm cable). Numerical example:

(1) Inner conductor radius:
(2) Outer conductor radius:
(3) Characteristic impedance:

At input, $a_i = 3.11 \times 10^{-3}$ m.
$b_i = 10^{-2}$ m.
$R_{oi} = 70$ ohms

At short, $a_r = 2.42 \times 10^{-3}$
$b_r = 8 \times 10^{-2}$
$R_{oi} = 210$ ohms

(4) Characteristic impedance at center point:

$$R_{om} = \sqrt{70 \times 210} = 121.1$$

(5) Q_s at center point:

$$Q_{sm} = 8{,}560 \frac{R_{om}}{(1/a + 1/b)\sqrt{\lambda}} = 2{,}620$$

(6) Exponent Γ:

$$\Gamma = \frac{1}{2}\ln\frac{210}{70} = 0.55$$

(7) From plot (Fig. 72):
$\phi_s = 1.24 \quad G_sQ_{sm}R_{om} = 0.69$

(8) Resonant length:

$$L_s = 1.24\frac{\lambda}{2\pi} = 0.99 \text{ m.}$$

(9) Resonant impedance:

$$R_s = Q_{sm}\frac{R_{om}}{0.69} = 460{,}000 \text{ ohms}$$

The corresponding values for the uniform stub are

$$\phi_s = 1.57 \text{ rad.}$$
$$G_s = \frac{\pi \times 10^6}{4Q_sR_0} = \frac{\pi \times 10^6}{4 \times 1{,}000 \times 70} = 11.2\ \mu\text{mhos}$$

By using the divergent stub, the length may be cut by 21 per cent and the conductance (or the stub loss, which is proportional to it), by about 70 per cent.

It may be said generally, if without precision, that the divergent stub has advantages as a resonator across closely spaced terminals. On the other hand, the convergent stub should have useful applications when the terminals need not be close together.

Consider, for example, the tank circuit of an amplifier (Fig. 73). If the conventional coil and condenser are used, it is impossible to obtain a high value of resonant impedance at very high frequencies, as for example, 60 Mc. (5 m.). If a uniform coaxial resonator were used, assuming $Q_s = 1{,}000$ and $R_0 = 200$ ohms, a resonant impedance of

$$R_s = \frac{4Q_sR_0}{\pi} = 255{,}000 \text{ ohms}$$

could be obtained. Such a resonator would have to have a resonant length of 1.25 m. and a diameter of 7.5 cm., and these dimensions would prohibit its practical use.

An open wire uniform stub would be equally impractical and would present serious interference problems. However, a *convergent* open wire stub would minimize these disadvantages. Over most of its length, this stub would have such a low value of center-to-center spacing that it could be rolled on a drum (Fig. 73). In addition, the spacing would be lower where the current is larger, which would reduce radiation to a negligible value. A numerical example is given in Fig. 73.

13.7. Applications and examples.

Design of the exponential transformer. It was pointed out in Sec. 13.4 that the exponential line approaches the performance of a perfect transformer (Sec. 14.6) as the wavelength tends to zero. Considering in particular the use of such a line as a means for securing uniform transmission between lines or cables with different values of characteristic impedance, or other unequal resistive terminations, we will now investigate how the degree of uniformity, or the freedom from amplitude distortion (Sec. 4.2), is related to the design parameters. It will be recalled that another solution of the same problem (the transformer problem), also suitable for very high frequencies, is based on the use of several uniform line sections (Sec. 10.2). A comparison between the two solutions will be made.

The exponential transformer has a characteristic impedance which varies exponentially between a value R_{oi} at the input end, equal to the source resistance, and a value R_{or} at the output end, equal to the load resistance.

The input impedance (impedance looking from the source into the exponential line) has been derived, both graphically and analytically. Its nominal value is R_{oi}, equal to the source resistance. This value is, however, realized only when the wavelength has certain definite values in relation to the line length. The nominal condition is represented in Fig. 69 by the point $R/R_{oi} = 1$. The input impedance point is in this position only when

$$\phi' = \phi \sqrt{1 - \left(\frac{\Gamma}{\phi}\right)^2} = n\pi \tag{422}$$

where ϕ' is the line angle of the equivalent uniform line (Fig. 69), $\phi = 2\pi l/\lambda$ the line angle proper, $\Gamma = \frac{1}{2} \ln R_{or}/R_{oi}$ the exponent, and n any integer.

We may write (422) in the form

$$\frac{\phi}{\Gamma} = \sqrt{1 + \left(\frac{n\pi}{\Gamma}\right)^2} \tag{423}$$

which lends itself to a convenient interpretation, based on the relationship between Γ and the cutoff frequency. We have defined the cutoff frequency as the value for which

$$\phi = \Gamma$$

At this value of frequency, a marked change in the exponential line operation takes place (Fig. 71), and it is convenient to refer all frequencies to this value. In terms of the cutoff frequency f_c we may write

$$\frac{\phi}{\Gamma} = \frac{f}{f_c} \tag{424}$$

and the cutoff frequency itself is given by

$$f_c = \frac{\Omega\Gamma}{2\pi l} = \frac{55}{l} \log \frac{R_{or}}{R_{oi}} \text{ Mc./sec.} \tag{425}$$

[*Cutoff frequency of the exponential line (l in meters)*]

Equation (423) now takes the form

$$\frac{f}{f_c} = \sqrt{1 + \left(\frac{\Omega n}{2lf_c}\right)^2} \tag{426}$$

Any frequency that satisfies the above, where n is an integer, is transmitted by the exponential line under maximum power transfer conditions (Sec. 8.2) or, which is the same thing in the case of resistive terminations, without reflection.

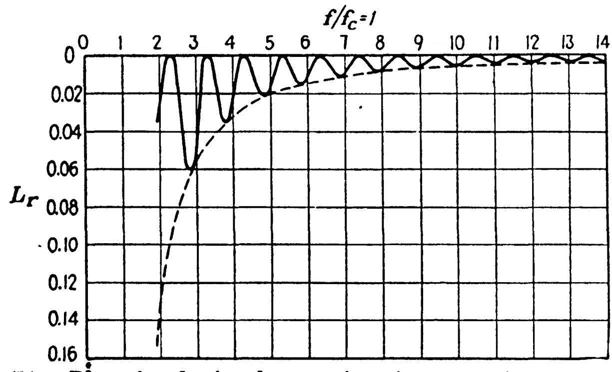


FIG. 74.—*Plot of reflection loss against frequency for an exponential line transformer.* L_R = *reflection loss in nepers;* f_c = *cut-off frequency; plot drawn for* $\Gamma = 0.8$*; envelope shown as dotted line.*

If the reflection loss (Sec. 6.6) is plotted against f/f_c (Fig. 74), the plot will go through zero at the points given by (426). Between these points there will be departures of the input impedance from the nominal value and of the reflection loss from zero. These departures will increase in magnitude as f decreases. The envelope curve of Fig. 74 may be interpreted as a plot of amplitude distortion vs. the minimum operating frequency limit. The distortion is given as the neper ratio L_R of the maximum to the minimum value of transmitted power within any frequency range having $f_{\min}$ as the lower limit and including at least one point of maximum ($L_R = 0$).

The equation of the envelope may be obtained with the help of the general relation (184) and of the polar plots of Z_i (Fig. 71). The general relation is

$$L_R = \frac{1}{8}|\Delta z|^2 \text{ neper}$$

where Δz is the relative departure of input impedance from the nominal or m.p.t. value. Δz is assumed to be small.

From the polar plots we see that within the range of useful operation the impedance point describes a series of nearly circular paths starting and ending at the nominal point, where they are tangent to the R axis. Hence, the departure is greatest when the impedance point is directly below or above the nominal point.

We also see from the constructions (Figs. 69 and 70) that the impedance point lies at all times upon a circle tangent to the R/R_{oi} axis at $R/R_{oi} = 1$ and having radius $\Gamma/\phi = f_c/f$. It follows from this and the above considerations that the maxima of Δz have, for all practical purposes, the value

$$\Delta z = 2\frac{f_c}{f}$$

Hence, the equation of the envelope

$$L_R = \frac{1}{2}\left(\frac{f_c}{f}\right)^2 \tag{427}$$

It is clear that if we substitute for f in the above equation the value $f_{\min}$ of the lowest operating frequency, the maximum loss over the operating range will be equal to or slightly less than the value of L_R given by the same equation, or

$$L_{R\max} \leq \frac{1}{2}\left(\frac{f_c}{f_{\min}}\right)^2 \tag{428}$$

We shall be on the safe side if we consider the last line to be an equation rather than an inequality. In this case, combining with (425) and solving for the minimum length of the exponential line,

$$l = \frac{117}{F_{\min}\sqrt{L_{R\max}}}\left(\log\frac{R_{or}}{R_{oi}}\right) \tag{429}$$

[*Minimum length of the exponential line*]

In the above, l is in meters, $F_{\min}$ in megacycles, $L_{R\max}$ in db, and the logarithm is to the base 10.

Comparison of the exponential and multisection lines as solutions *of the transformer problem.* In order to compare the two solutions, graphical means must be used because no convenient analytical

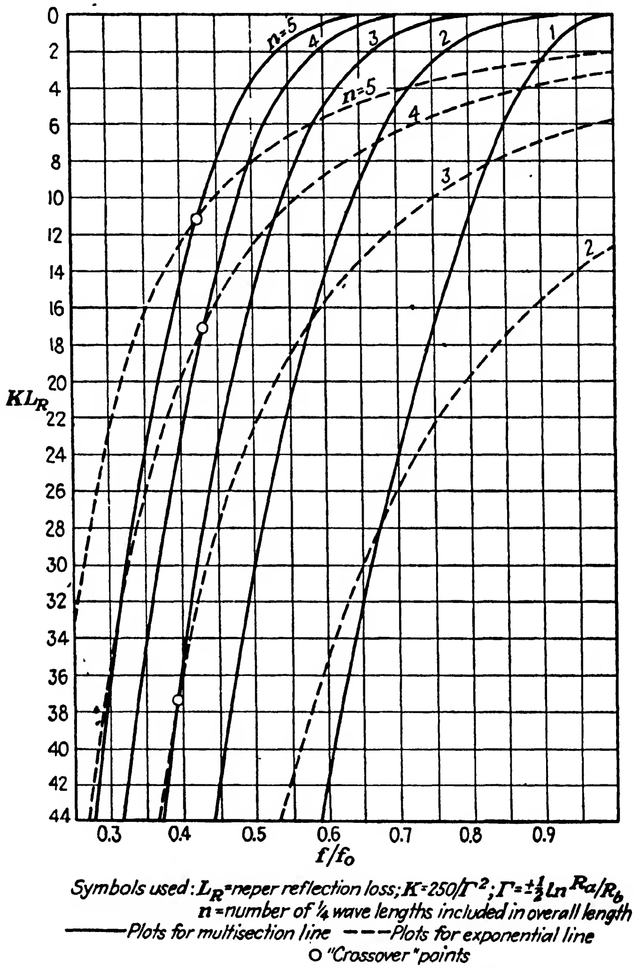


FIG. 75.—*Comparison of the exponential and multiaction solutions of the vhf transformer problem.*

expression comparable to (428) has been derived for the multisection transformer (Sec. 10.2). Instead, plots of KL_R against f/f_0 have been obtained (Fig. 44) for various values of n. Similar plots must be drawn for the exponential line for a valid comparison.

The symbols used in Fig. 44 have the following significance:

$$K = \frac{10^3}{(\ln R_a/R_b)^2} = \frac{250}{\Gamma^2} \tag{430}$$

$$n = \frac{4l}{\lambda_o} = \frac{4lf_o}{\Omega} \tag{431}$$

$$f_o = \sqrt{f_{\min} f_{\max}} \tag{432}$$

f_o is the midband value of frequency (geometric mean of the extremes of the useful range), n the number of quarter wavelengths comprised in the length of the line at midband.

We now must express KL_R in terms of f/f_c and n for the exponential line. From Eq. (428) we have

$$KL_R = \frac{250}{2\Gamma^2}\left(\frac{f_c}{f}\right)^2 = \frac{250\, f_c^2}{2\Gamma^2 f_o^2}\left(\frac{f_o}{f}\right)^2$$

and from (426)

$$KL_{R\max} = 500\left(\frac{\Omega}{4\pi l f_o}\right)^2\left(\frac{f_o}{f}\right)^2$$

Now, recalling (431),

$$KL_R = \frac{500}{\pi^2 n^2}\left(\frac{f_o}{f}\right)^2 \tag{433}$$

Figure 75 shows plots of (433) for $n = 2, 3, 4, 5$ superimposed to the plots of Fig. 44. For each value of n the two solutions, multisection and exponential, are thus made directly comparable. There is this difference: noninteger values of n are possible for the exponential but not for the multisection line; on the other hand, the plots for the exponential line are to be interpreted as *envelopes*, actual values of L_R fluctuating between zero and the envelope.

A numerical example will bring out the significance of the plots. Assume the following requirements:

Frequency range: 5 to 50 mc.

Impedance ratio: 10 to 1

Allowable distortion (maximum reflection loss): 1 db

We have from the above

$$\frac{f_{\min}}{f_o} = \frac{5}{\sqrt{5 \times 50}} = 0.316$$

$$KL_{R\max} = \frac{10^3}{(\ln 10)^2 \times 8.68} = 21.74$$

The plots (Fig. 75) show for $f/f_o = 0.316$ and for $n = 5$, the following values of $KL_{R\max}$:

Exponential line:

$$KL_{R\max} = 20$$

Multisection line:

$$KL_{R\max} = 32$$

In both cases the total length would have to be

$$l = \frac{n\Omega}{4f_o} = \frac{300 \times 5}{4 \times 15.8} = 23.7 \text{ m.}$$

hence the exponential line is better suited.

Whether one or the other type of line should be used depends on the value of $f_{\min}/f_o$, *i.e.*, on the bandwidth. For each n there is a value of f/f_o at which the two plots, for the exponential and multisection lines, cross over. All these cross-over points, as it happens, lie at or near the abscissa

$$f/f_o = 0.425$$

This value corresponds to the bandwidth

$$W = \frac{f_o}{f_{\min}} - \frac{f_{\min}}{f_o} = \frac{f_{\max} - f_{\min}}{f_o} = 2.35 - 0.425 = 1.925$$

Unless other factors (in addition to amplitude distortion and over-all lengths) are taken into consideration, we may conclude that the exponential line is preferable as the solution of the transformer problem whenever the bandwidth is greater than 2.

CHAPTER XIV

INDUCTIVE COUPLING AND TRANSFORMERS

14.1. Linkage of Flux and Current. The laws governing electric and magnetic fields under static conditions were reviewed in Chapter XI. Inductance and capacity were defined there (Secs. 11.7 and 11.8). Both these concepts have significance only when the fields change relatively slowly with time. Moreover, their definition presupposes that only two charged conductors, or only one current circuit, have effect upon the field. In spite of these restrictions, these parameters play an all-important part in the analysis of lumped systems.

Distributed values of L and C continue to have meaning when the fields are not slowly varying, because of the geometry of the system to which they refer; but the second restriction, that only one current and one voltage have effect on the field at any particular point, embraces distributed as well as lumped values of L and C.

We will now go on to consider a system of *two* currents, stipulating that because of the geometry of the system and the slowness of current variations, displacement currents (Sec. 12.2) are negligible, and in consequence the current in each circuit is the same at all points. This is true, for practical purposes, at power and lower audio frequencies, and in many cases at higher frequencies.

We will ultimately be interested in relations between the currents i_p and i_s (Fig. 76), designated hereafter for convenience as *primary* and *secondary*, and the corresponding linked values of magnetic flux. First, however, let us discuss the subject of linkage.

The flux or flow of a vector linked with a loop or closed line was illustrated in Fig. 57. The magnetic flux linked with a current is included in the general definition, the loop being in this case coincident with the current path and in the same direction. The linked flux is positive when to an observer looking in the direction of the flux lines the current (or loop) appears to have *clockwise* direction.

It should be noted that, just as we have defined the *linked flux* with respect to a *current path*, we can also define the *linked current* with respect to a *flux path*, more commonly called *line of flux* or *line of force*. In fact, current and flux are formally similar; both may be defined as the flow of a vector through a surface. However, when the current is confined within wires of small cross section, the linked flux is the same for all current paths of the same circuit. The reverse is never quite true; the

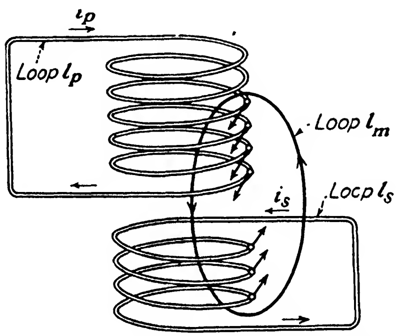


FIG. 76.—*Linkage of flux and current.*

Loop l_m, taken along any line of force in the magnetic flux tube Φ_m, has +4 linkages with loop l_p (path of primary current), −3 linkages with l_s (path of secondary current).

Number of linkages is number of penetrations of linking loop through a surface bounded by linked loop or vice versa.

The sign of each penetration is determined by the rule of the right-handed screw (Fig. 60).

linked current is not the same for all the flux lines of the same magnetic field, except in very special cases.

It is therefore necessary, for purposes of evaluation, to subdivide the magnetic field existing at any particular instant into *tubes* or *partial fluxes*, such that all lines of each tube have the same linked current, or, to use the established phrase, the same number of *current linkages*.

For example, the partial flux Φ_m (Fig. 76) has −3 secondary and +4 primary current linkages. To determine these numbers, we must obtain the current linked with any flux line of Φ_m, or, according to the definition, draw a surface bounded by this flux line and count the number of times each current traverses this surface. The convention with regard to the sign is the usual one; if, looking in the direction of current, the flux line appears to rotate clockwise, the linkage is positive. All flux lines may be considered to have the same direction; any convention regarding the direction of flux and currents may be used, provided we respect the rule concerning the relationship between each current and the corresponding linked flux.

We may, in general, write the *linked current* for each partial flux Φ_m as follows:

$$i_m = p_m i_p + s_m i_s \tag{434}$$

where p_m and s_m (primary and secondary linkages) may be positive or negative integers. The linked current is variously referred to as *magnetomotive force* or *ampere turns*. When the so-called *practical* system is used, the magnetomotive force is defined as 0.4π times the linked current. It is obvious that the concept of turns, useful as it is in many applications, lacks generality.

Next, we must assume that at any particular instant of time, each partial flux is equal to the corresponding linked current multiplied by a positive coefficient. We will write, therefore,

$$\Phi_m = L_m(p_m i_p + s_m i_s) = L_m i_m \tag{435}$$

The positive coefficient L_m has the dimensions of an inductance (Sec. 11.7); it is not, however, a constant parameter.[1]

14.2. Self and Mutual Inductances. Coefficient of Coupling. We may use (435) to express the fluxes linked with the primary and secondary currents, respectively.

If Φ_p is the flux linked with the primary, consider the contribution to Φ_p due to the partial flux Φ_m. It will be clear from an inspection of Fig. 76 that if a surface bounded by any flux line of Φ_m is crossed p_m times by i_p, then a surface bounded by i_p is crossed p_m times by Φ_m. Generally speaking, the number of mutual linkages of two closed paths may be defined in two ways, by associating a surface with either closed path and counting the positive penetrations of the other path through this surface.

[1] The boundaries of the partial flux tubes have been shown to change during each cycle under harmonic excitation.

As a result, we may write

$$\Phi_p = \sum_{m=1}^{n} L_m p_m (p_m i_p + s_m i_s)$$

$$\Phi_s = \sum_{m=1}^{n} L_m s_m (p_m i_p + s_m i_s)$$

and, collecting terms in i_p and i_s

$$\begin{aligned} \Phi_p &= L_p i_p + M i_s \\ \Phi_s &= L_s i_s + M i_p \end{aligned} \tag{436}$$

having let

$$L_p = \sum_{m=1}^{n} L_m p_m^2 \qquad L_s = \sum_{m=1}^{n} L_m s_m^2 \qquad M = \sum_{n=1}^{n} L_m p_m s_m \tag{437}$$

L_p and L_s are the *coefficients of self-inductance*, M the *coefficient of mutual inductance*. Experiment shows that all three of these coefficients are constant[1] when the media have constant permeability.

We may define L_p through Eq. (436) as the ratio Φ_p/L_p for $i_s = 0$; we may define L_s in the same manner. M may be defined identically by the following:

$$M = \frac{\Phi_p^{(i_p=0)}}{i_s} \qquad \text{or} \qquad M = \frac{\Phi_s^{(i_s=0)}}{i_p}$$

The subscripts s and p, used in this section for reasons of clarity, will be replaced later by the subscripts 1 and 2 in agreement with standard usage.

Equations (437) enable us to draw some generic conclusions with regard to the self- and mutual-induction coefficients. It is principally for this purpose that the foregoing discussion on linkages has been carried out.

The values $L_1, L_2, \ldots L_n$; $p_1, p_2, \ldots p_n$; and $s_1, s_2, \ldots s_n$ are not constant with time, but refer to the field configuration at a given instant of time. We know, however, that $L_1, L_2, \ldots$

[1] In the case of large flux variations in ferromagnetic cores, the coefficients, as defined, are not constant. In such cases and for sinusoidal excitation, harmonics are generated and the problem becomes much more complex. It is possible, however, to apply the transformer equations based on (436), using equivalent values of the coefficients and restricting the analysis to fundamental components.

L_n are always positive, while the p and s coefficients may be positive or negative.

Hence, we see immediately that the self-inductances are always positive, while the mutual inductance may be positive or negative.

The particular case for which

$$\begin{array}{ll} p_2 = p_2 = p_3 = \cdots = N_p & \text{(primary turns)} \\ s_1 = s_2 = s_3 = \cdots = N_s & \text{(secondary turns)} \end{array} \qquad (438)$$

is approached when the two current circuits are wound around the same closed ferromagnetic core and the flux density in the core is very large compared to that of the surrounding space. In this case we have, evidently,

$$L_p = L_0 N_p^2 \qquad L_s = L_0 N_s^2 \qquad M = L_0 N_p N_s \qquad (439)$$

where now L_0 is the ratio of the core flux to the linked current, or the current crossing a surface whose edge is within the core. L_0 is the inverse of the core *reluctance*. If the core links with the current only once (Fig. 62*a*), L_0 becomes the inductance of the current circuit.

It may be shown quite generally that

$$L_p L_s - M^2 > 0$$

by substituting the values of Eqs. (437). Thus

$$\begin{aligned} L_p L_s - M^2 &= \sum_{0<m<n} (L_m p_m s_m)^2 + \sum_{h \neq k} L_h p_h^2 L_k s_k^2 - \sum_{0<m<n} (L_m p_m s_m)^2 \\ &\qquad - \sum_{h \neq k} L_h p_h s_h L_k p_k s_k \\ &= \sum_{h>k} L_h L_k (p_h^2 s_k^2 + p_k^2 s_h^2 - 2 p_h s_h p_k s_k) \\ &= \sum_{h>k} L_h L_k (p_h s_k - p_k s_h)^2 \end{aligned} \qquad (440)$$

The difference in question is evidently *positive* and vanishes only when

$$p_h s_k = p_k s_h \qquad (441)$$

for any combination of h and k. The ratio

$$\tau = \left| \frac{M}{\sqrt{L_p L_s}} \right| . \qquad (442)$$

is therefore never greater than unity. τ is the *coefficient of coupling.* That unity coupling is physically impossible may be shown by applying (440) to two partial fluxes, of which Φ_k links only with the primary. We would have in this case

$$p_h s_k = 0 \qquad p_k s_h \neq 0$$

showing that (440) cannot be satisfied for any combination of h and k unless *all* the flux lines link both currents.

The ideal case described by Eqs. (438) is, of course, consistent with unity coupling, but this condition is never fully realized. On the other hand, unity coupling is not subject to the condition that all the flux lines have the same number of linkages.

In the discussion on the transformer, we will find that this is equivalent to a T network of coils, of inductances $L_p - M$, $L_s - M$, and M. Let us express the first two inductances with the help of (437). We obtain

$$\begin{aligned} L_p - M &= \sum_{m=1}^{n} L_m p_m (p_m - s_m) \\ L_s - M &= \sum_{m=1}^{n} L_m s_m (s_m - p_m) \end{aligned} \tag{443}$$

In closely coupled transformers, owing to the geometry of the windings, all the primary linkages must have the same sign, and likewise all the secondary linkages. Hence, if M is positive (437), p_m and s_m have the same sign for any m and $L_p - M$, $L_s - M$ have opposite signs. This is also true for M negative.

14.3. Systems of Several Currents. Linear relations similar to (436) may be written between currents and linked fluxes when more than two currents have effect on the field. We may arrive at such relations by assuming that the magnetic flux density, at any point in space, due to a current I flowing in a circuit element of infinitesimal length dl, is given by [9]

$$\mathbf{B} = \frac{\mu I \, \mathbf{dl} \times \mathbf{r}}{r^3} \tag{444}$$

The flux density due to each entire current circuit may be obtained by integrating the above over the length of the circuit; the resulting value will have I as a multiplier. Proceeding likewise for all currents, we conclude that the flux density at any point is a linear function of all the currents. The same may be said of the flux linked with each current; hence the system

$$\begin{aligned}
\Phi_1 &= L_{11}I_1 + L_{12}I_2 + \cdots + L_{1n}I_n \\
\Phi_2 &= L_{21}I_2 + L_{22}I_2 + \cdots + L_{2n}I_n \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\Phi_n &= L_{n1}I_1 + L_{n2}I_2 + \cdots + L_{nn}I_n
\end{aligned} \tag{445}$$

14.4. The Transformer Equation. A transformer differs from the two-current system of Fig. 76 because it does not contain two complete circuits. A transformer is a four-terminal net-

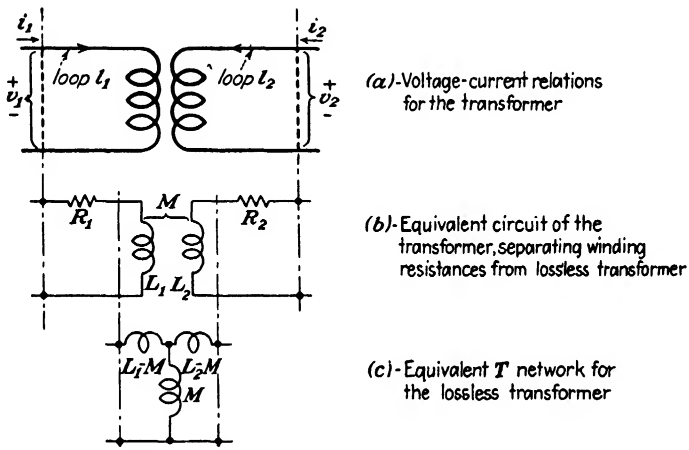


FIG. 77.—*Equivalent "T" of the transformer.*

work, not a closed system (Sec. 1.1). In applying to the transformer the conclusions reached in the preceding sections, we run up against the same difficulty as in applying the concept of inductance to a coil (Sec. 12.5).

We must here, as in the case of the coil, run a closed path along each winding and across the terminals (Fig. 77*a*). These paths will not coincide with the current paths over their entire length. In order to apply (436) we must assume that the flux linked with l_1, for example, is substantially the same as if i_1 actually followed l_1 across the terminals.

Subject to this condition, we obtain the induced e.m.f. values, or loop integrals of the electric field around l_1 and l_2, by taking the time derivatives of Eqs. (436). The subscripts 1 and 2 will now be used in place of p and s. Thus

$$\begin{cases} -e_1 = -\oint_{l_1} E_l\,dl = L_1 \dfrac{\partial i_1}{\partial t} + M \dfrac{\partial i_2}{\partial t} \\ -e_2 = -\oint_{l_2} E_l\,dl = L_2 \dfrac{\partial i_2}{\partial t} + M \dfrac{\partial i_1}{\partial t} \end{cases} \tag{446}$$

Each loop integral may now be broken up, as in (Sec. 12.5), into the voltage across the terminals and the contribution due to the electric field within the wire of the windings, which is the product of current and resistance. Hence, expressing the terminal voltages,

$$\begin{cases} v_1 = i_1R_1 - e_1 = i_1R_1 + L_1 \dfrac{\partial i_1}{\partial t} + M \dfrac{\partial i_2}{\partial t} \\ v_2 = i_2R_2 - e_2 = i_2R_2 + L_2 \dfrac{\partial i_2}{\partial t} + M \dfrac{\partial i_1}{\partial t} \end{cases} \tag{447}$$

(The positive directions are those of the generic network, Fig. 1*a*.) It is clearly possible to consider a lossless transformer in series with the resistances R_1 and R_2 as an equivalent circuit for the transformer (Fig. 77*b*). The transmission properties of the lossless transformer will therefore be discussed in the following.

14.5. Network Coefficients for the Lossless Transformer. Input Impedance and Voltage Ratio. The equations for the lossless transformer may be written in complex form for steady-state harmonic values, as follows:

$$\begin{cases} \mathbf{V}_1 = j\omega(L_1\mathbf{I}_1 + M\mathbf{I}_2) \\ \mathbf{V}_2 = j\omega(L_2\mathbf{I}_2 + M\mathbf{I}_1) \end{cases} \tag{448}$$

Comparing the above with the generic network equations (14), we obtain the network coefficients for the transformer

$$z_{11} = j\omega L_1 \qquad z_{22} = j\omega L_2 \qquad z_{12} = z_{21} = j\omega M \tag{449}$$

It may be easily verified that the T network of Fig. 77*c* has the same coefficients, which makes it equivalent to the transformer. We have seen, however, that in absence of stray flux, the two series inductances of the T must have opposite signs if the shunt inductance M is to be positive. Hence, one of the three must always be negative. It is therefore impossible to reproduce in this way any of the useful transmission properties of the transformer.

The transformer performance may be summarized by the values of input impedance and voltage or current ratio. The insertion loss may be obtained in terms of the input impedance.

We will first obtain generic expressions for these values in terms of the network coefficients, then the particular ones for the transformer. Finally, a useful equivalent circuit for the lossless transformer will be obtained; this in turn will provide us with a simple graphical interpretation.

From the system (14) we obtain the following general expressions:

$$\frac{\mathbf{I}_2}{\mathbf{I}_1} = -\frac{z_{21}}{Z_r + z_{22}} \tag{450}$$

[*Current ratio for generic network*]

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{z_{21}Z_r}{z_{11}Z_r + D_z} \tag{451}$$

[*Voltage ratio*]

$$Z_i = \frac{\mathbf{V}_1}{\mathbf{I}_1} = \frac{z_{11}Z_r + D_z}{Z_r + z_{22}} \tag{452}$$

[*Input impedance*]

where $z_{11}\ z_{22}\ z_{12}\ z_{21}$ = network impedance coefficients (Sec. 2.7)
$D_z = z_{11}z_{22} - z_{21}z_{12}$ = determinant of network coefficients
Z_r = load impedance

(*Note.* for passive linear networks, $z_{12} = z_{21}$.)

We will now rewrite the above for the lossless transformer, using the coefficients as given by (449) and noting that

$$D_z = -\omega^2(L_1L_2 - M^2) = -\omega^2L_1L_2(1 - \tau^2)$$

$$\frac{\mathbf{I}_2}{\mathbf{I}_1} = -\frac{\tau}{t}\frac{1}{1 - jz_r} \tag{453}$$

[*Current ratio for lossless transformer*]

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = -\tau t\,\frac{jz_r}{1 - \tau^2 - jz_r} \tag{454}$$

[*Voltage ratio*]

$$Z_i = j\omega L_1\,\frac{1 - \tau^2 - jz_r}{1 - jz_r} \tag{455}$$

[*Input impedance*]

where

$$\tau = \frac{M}{\sqrt{L_1L_2}} = \text{coefficient of coupling} \tag{456}$$

$$t = \sqrt{\frac{L_2}{L_1}} = \text{turn ratio} \tag{457}$$

$$z_r = \frac{Z_r}{\omega L_2} = \frac{Z_r}{\omega L_1 t^2}$$

(Z_r includes secondary resistance of dissipative transformer.)

It appears from the above that the voltage ratio will be uniformly close to t, the turn ratio of the transformer, for any value of load impedance, provided the coupling is near unity. The current ratio is nearly equal to the inverse of the voltage ratio only if z_r is small (load impedance small compared to the open-circuit impedance from the output side). The input

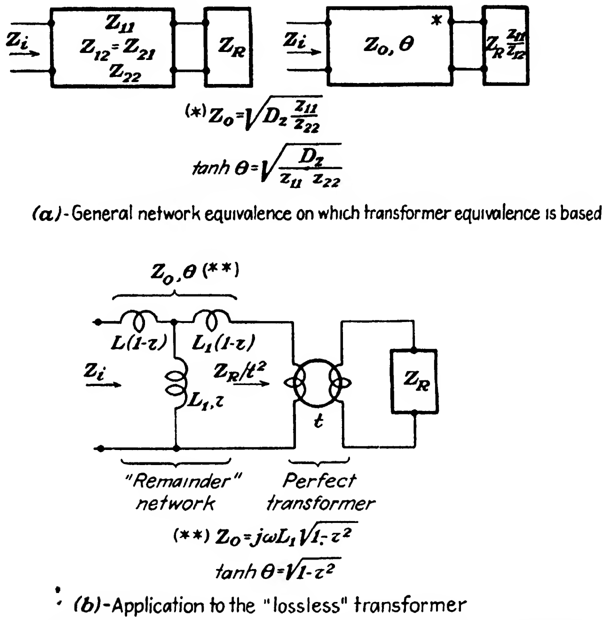


FIG. 78.—*Equivalent circuit for the lossless transformer, including a perfect transformer and a reversible T network.*

impedance approaches the open-circuit value when z_r is large or when the coupling is loose; for close coupling and z_r small it approaches $1/t^2$ times the load impedance.

14.6. Equivalent Circuit of the Lossless Transformer. Given a *nonreversible* lossless network (Sec. 2.9), it will now be shown that its impedance transforming action may always be duplicated by a *reversible* network, provided the load impedance is multiplied by the ratio z_{11}/z_{22} (Fig. 78*a*). This is a useful equivalence as it permits us to extend the graphical methods applicable to reversible networks (Fig. 31*b*). To obtain the constants of the reversible network for generic values of the coefficients, let us express the input impedance of the reversible network, loaded by

$Z_r(z_{11}/z_{22})$, and compare it with Eq. (452). Using Eq. (101), we have

For the reversible network of unknown constants Z_0, θ loaded by $Z_r(z_{11}/z_{22})$:

$$Z_i = Z_0 \frac{Z_r(z_{11}/z_{22}) + Z_0 \tanh \theta}{Z_0 + Z_r(z_{11}/z_{22}) \tanh \theta}$$
$$= Z_0 \tanh \theta \frac{1 + (Z_r z_{11}/Z_0 z_{22}) \coth \theta}{1 + (Z_r z_{11}/Z_0 z_{22}) \tanh \theta}$$

For the given network (452):

$$Z_i = \frac{z_{11}Z_r + D_z}{Z_r + z_{22}} = \frac{1 + (z_{11}Z_r/D_z)}{1 + (Z_r/z_{22})} \frac{D_z}{z_{22}}$$

Comparing the coefficients of Z_r in the two expressions,

$$\frac{z_{11}}{D_z} = \frac{z_{11}}{z_{22}Z_0} \coth \theta \qquad \therefore\ Z_0 = \frac{D_z}{z_{22}} \coth \theta$$
$$\frac{1}{z_{22}} = \frac{z_{11}Z_r}{z_{22}Z_0} \tanh \theta \qquad \therefore\ Z_0 = z_{11} \tanh \theta$$

Hence

$$Z_0 = \sqrt{D_z \frac{z_{11}}{z_{22}}} \tag{458}$$
$$\tanh \theta = \sqrt{\frac{D_z}{z_{11}z_{22}}} \tag{459}$$

are the constants of the reversible network in the equivalence of Fig. 78*a*.

Let us now apply the equivalence to the lossless transformer. The load impedance must be changed to the value

$$Z_r \frac{z_{11}}{z_{22}} = Z_r \frac{L_1}{L_2} = \frac{Z_r}{t^2}$$

In other words, it must be divided by the *square of the turn ratio.* If we inserted a *perfect transformer* before the load (Fig. 78*b*), it would account for this change in the value of load impedance. The perfect (or ideal) transformer (not to be confused with the lossless transformer) is characterized by this relation of the terminal values

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{\mathbf{I}_1}{\mathbf{I}_2} = t$$

which in physical transformers is satisfied at best only approximately. The remainder of the equivalent circuit of Fig. 78*b* is wholly responsible for the departure of the real transformer from its ideal behavior, and, as such, is particularly significant.

This remainder is a reversible network, whose constants may be obtained from (458) and (459) by substitution of the transformer coefficients (449). Thus

$$Z_0 = \sqrt{-\omega^2 L_1 L_2(1 - \tau^2)\frac{L_1}{L_2}} = jmX_1 \tag{460}$$

$$\tanh\theta = \sqrt{\frac{-\omega^2 L_1 L_2(1 - \tau^2)}{-\omega^2 L_1 L_2}} = m \tag{461}$$

where

$$m = \sqrt{1 - \tau^2} \qquad X_1 = \omega L_1 \tag{462}$$

Z_0 and θ completely define this network. However, its T equivalent affords a more concrete picture.

In general, T equivalents of reversible networks having known constants can be constructed by solving the system

$$\begin{cases} Z_{0c} = Z_a + Z_b \\ Z_{sc} = Z_a + \dfrac{Z_b Z_a}{Z_a + Z_b} = Z_a \dfrac{Z_a + 2Z_b}{Z_a + Z_b} \end{cases} \tag{463}$$

where Z_a is either series arm, Z_b the shunt arm. The roots are

$$Z_a = Z_{0c}\left[1 \pm \sqrt{1 - \frac{Z_{sc}}{Z_{0c}}}\right] = Z_{0c}\,[1 \pm \sqrt{1 - \tanh^2\theta}]$$

$$Z_b = \pm Z_{0c}\sqrt{1 - \frac{Z_{sc}}{Z_{0c}}} = \pm Z_{0c}\sqrt{1 - \tanh^2\theta}$$

Noting that (26)

$$Z_{0c} = \frac{Z_0}{\tanh\theta}$$

the above may be written

$$\begin{cases} Z_a = Z_0\,(\coth\theta \pm \sqrt{\coth^2\theta - 1}) \\ Z_b = \pm Z_0\sqrt{\coth^2\theta - 1} \end{cases} \tag{464}$$

Using the above, we have for the transformer remainder network (Fig. 78*b*) the following values:

$$\begin{cases} Z_a = j\omega L_1\,[1 \pm \sqrt{1 - 1 + \tau^2}] = j\omega L_1(1 - \tau) \\ Z_b = \pm j\omega L_1\sqrt{1 - 1 + \tau^2} = j\omega L_1\tau \end{cases} \tag{465}$$

having chosen the signs consistent with positive values of the inductances.

14.7. Graphical Determination of Transformer Performance. The equivalent network of Fig. 78 permits us to apply to the transformer the generic construction of Fig. 31*b* for the input impedance of a reversible network.

The reversible network is, in this case, the T network of Fig. 78. For this network, the characteristic impedance Z_0 is a pure imaginary (460) and the transfer constant θ a pure real (461), because $M^2/L_1L_2 < 1$ in all cases.

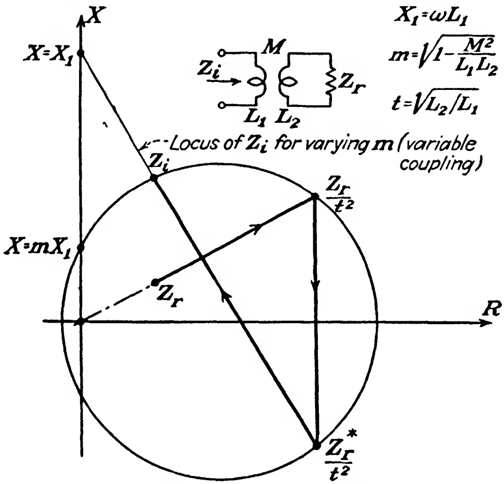


FIG. 79.—*Input impedance of the lossless transformer.*

Procedure: Locate Z_r/t^2 on complex plane. Locate points $X = X_1$ and $X = mX_1$ on X axis. Draw circle with center on R axis through $X = mX_1$ and Z_r/t^2. Draw straight line through $X = X_1$ and Z_r^*/t^2. Z_i is at the intersection of straight line and circle.

It follows that the *load impedance point* Z_r/t^2 and the *input impedance point* Z_i (Fig. 31*b*) must lie on the same constant τ circle.[1] In fact, insertion of the T network of Fig. 78 before the load does not add to the imaginary part τ of the reflection constant of the load, but only to the real part, σ. By comparison with Fig. 31*b*, we see that the constant τ circle in this case has center on the R axis and cuts the X axis at $X = mX_1$ (Fig. 79).

The construction of Fig. 79 shows that Z_i is located at the

[1] The reader is cautioned against confusion between τ, imaginary part of the reflection constant (Sec. 5.4), and τ, coefficient of coupling. The expression $M/\sqrt{L_1L_2}$ will be written in full for the coefficient of coupling wherever it is felt that there could be ambiguity.

intersection of the above circle with a straight line drawn through the *conjugate point* (symmetrical about the R axis) of Z_r/t^2, and the point $X = L_1$ on the X axis. This line is the locus of Z for *variable* m, as may be argued from the fact that it may be drawn independently of $m = \sqrt{1 - (M^2/L_1L_2)}$.

To show that Z_i must lie on the above-mentioned straight line, we must go back to Eq. 455, expressing the input impedance as the ratio of two complex numbers. Such ratios may always be evaluated graphically by intersecting two straight lines, as will be shown in the following. One of the straight lines and the constant r circle are used in the construction of Fig. 79.

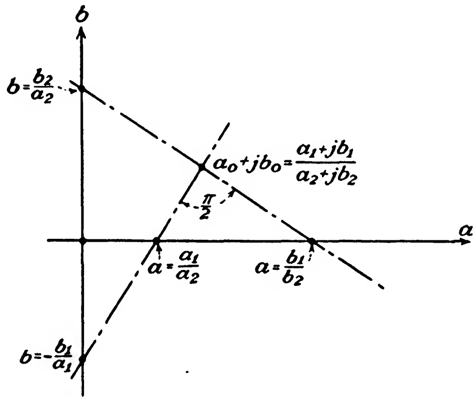


FIG. 80.—*Construction for the ratio of two complex numbers.*

To locate the ratio of two complex numbers on the complex plane of coordinates a, b, let

$$a_0 + jb_0 = \frac{a_1 + jb_1}{a_2 + jb_2} \tag{466}$$

Separating real and imaginary terms, we have the identities

$$\begin{aligned} a_0a_2 - b_0b_2 &= a_1 \\ a_0b_2 + b_0a_2 &= b_1 \end{aligned}$$

a_0 and b_0 must satisfy simultaneously the following equations in a and b:

$$aa_2 - bb_2 = a_1 \tag{467}$$
$$ab_2 + ba_2 = b_1 \tag{468}$$

The above are represented on the ab plane by the straight orthogonal lines of Fig. 80.

Let us find the lines in question, in the particular case when the input of impedance of the transformer is taken in place of $a_0 + jb_0$. We rewrite

(455) as follows:

$$z_i = j\frac{m^2 - jz_r}{1 - jz_r} \tag{469}$$

having let for brevity

$$m = \sqrt{1 - \frac{M^2}{L_1L_2}} \tag{462}$$

$$z_i = \frac{Z_i}{\omega L_1} \tag{470}$$

$$z_r = \frac{Z_r}{\omega L_1 t^2} \tag{471}$$

Comparing (469) with (466), we find that the generic coefficients now have the values

$$a_1 = r_r \qquad b_1 = n^2 + x_r \qquad a_2 = 1 + x_r \qquad b_2 = -r_r$$

Equation (467) is now written

$$r(1 + x_r) + xr_r = r_r \tag{472}$$

As may be easily verified, the above represents a straight line on the rx plane which cuts the x axis at

$$x = 1 \qquad r = 0$$

and the values

$$r = r_r \qquad x = -x_r$$

satisfy (472). To draw the line on the impedance plane proper, all these values must be multiplied by ωL_1. On the impedance plane (of coordinates R, X) the line must go through the points

$$X = \omega L_1 = X_1 \qquad \text{and} \qquad R = 0$$

$$X = -\frac{X_r}{t^2} \qquad \text{and} \qquad R = \frac{R_r}{t^2}$$

as in Fig. 79.

The impedance transforming action of the transformer for various values of load, primary reactance, and coupling coefficient will be revealed very clearly to the reader by means of the construction of Fig. 79 carried out in a number of cases. The frequency response of a transformer with resistive load will be investigated in Sec. 14.11.

In some applications the primary current (or the primary voltage) is practically independent of the transformer input impedance. In such cases the complex current ratio (or voltage ratio) of the transformer fully describes its performance. Graphical constructions for these ratios, which can be used in power as well as other applications, are given in Fig. 81. They are based on the principle of Fig. 80, applied to Eqs. (453) and (454).

14.8. The Autotransformer. If two coils of self-inductances L_1, L_2, and mutual inductance M (defined as in Sec. 14.2) are connected to source and load as shown in Fig. 82a, the resulting

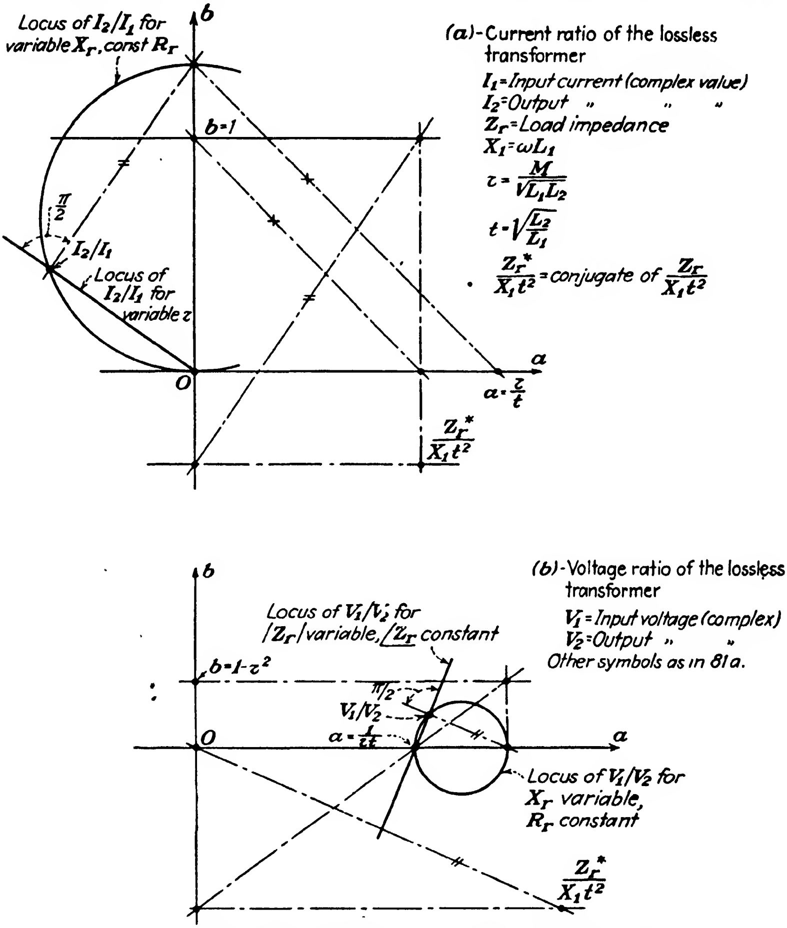


FIG. 81.—*Voltage and current ratios of the lossless transformer.*

four-pole, known as the *autotransformer*, is equivalent to a transformer connected as in Fig. 82b, having values L_1', L_2', and M', which may be expressed in terms of L_1, L_2, and M. The autotransformer (tapped coil) is often used in radio circuits, when primary and secondary do not have to be at different d-c potentials.

To find the parameters of the equivalent transformer, we may express the *linked flux* for the primary and secondary meshes of Fig. 82*a*, noting that the current through coil 1 is $i_1 + i_2$ and that through coil 2 is i_2. Thus

$$\begin{aligned}\Phi_1 &= L_1(i_1 + i_2) + Mi_2 = L_1i_1 + (L_1 + M)i_2 \\ \Phi_{1+2} &= \Phi_1 + \Phi_2 = L_1i_1 + (L_1 + M)i_2 + L_2i_2 + M(i_1 + i_2) \\ &= (L_1 + L_2 + 2M)i_2 + (L_1 + M)i_1\end{aligned}$$

Φ_1 and Φ_{1+2}, differentiated, give the primary and secondary values of e.m.f.; hence, we may consider these values of flux

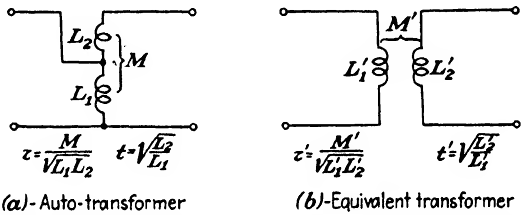


FIG. 82.—*Equivalent values of coupling and turn ratio for the autotransformer:*

$$\tau' = \frac{1 + \tau t}{\sqrt{1 + t^2 + 2\tau t}} \qquad t' = \sqrt{1 + t^2(1 + 2\tau^2)}$$

to be the values of the equivalent transformer. By comparison with Eqs. (446) we may write for the constants of the equivalent transformer

$$\left\{\begin{array}{ll} L_1' = L_1 & L_2' = L_1 + L_2 + 2M) \\ & M' = L_1 + M \end{array}\right. \tag{473}$$

The expression for L_2' is, of course, the well-known expression for the inductance of a series combination of mutually coupled coils. M may in general have positive or negative value; in the autotransformer, however, M is not negative as the two coils are generally part of the same winding.

The values τ' and t' for the coefficient of coupling and turn ratio of the equivalent transformer may be obtained from (473). Thus

$$\tau' = \frac{M'}{\sqrt{L_1'L_2'}} = \frac{L_1 + M}{\sqrt{L_1^2 + L_1L_2 + 2L_1M}} = \frac{1 + \tau t}{\sqrt{1 + t^2 + 2\tau t}} \tag{474}$$

$$t' = \sqrt{\frac{L_2'}{L_1'}} = \sqrt{\frac{L_1 + L_2 + 2M}{L_1}} = \sqrt{1 + t^2(1 + 2\tau^2)} \tag{475}$$

[*Equivalent parameters of the autotransformer*]

We deduce from (474) that the equivalent coupling coefficient can be made to approach unity in two ways: by making τ as near unity as possible, *i.e.*, by closely coupling the coils; or by making t small. In other words, if the tap is near the high end of the coil, close coupling may effectively be obtained even if the mutual induction between the two coil sections is small. This is a distinct advantage of the autotransformer connection; however, the advantage vanishes for high equivalent turn ratios.

As may be verified from (474), if $m = \sqrt{1 - \tau^2}$ is small (as it always is in magnetic core transformers), the equivalent value m' for the autotransformer connection is approximately

$$m' = m \frac{t}{\sqrt{1 + t^2}}$$

In this case the equivalent turn ratio may be written approximately

$$t' = \sqrt{1 + 3t^2}$$

Hence, substituting,

$$m' = m \sqrt{\frac{(t')^2 - 1}{(t')^2 + 2}} \tag{476}$$

For example, if the required equivalent turn ratio (voltage ratio) is 2, the relation is as follows:

$$m' = \frac{m}{\sqrt{2}}$$

It will be shown in Sec. 14.11 that the parameter $m = \sqrt{1 - \tau^2}$ has a direct bearing on the frequency response of the transformer and may be taken as a direct measure of the effect of leakage flux. The above shows, therefore, that the improvement which may be secured in this respect by the autotransformer connection is rather limited.

14.9. Coupled Circuits. Most radio and intermediate frequency amplifiers are transformer coupled. As distinct from audio-frequency or power transformers, the types used at radio and intermediate frequencies have loose coupling and small self-inductances. The theory and constructions developed in the preceding sections are valid for any transformer in linear operation and can be applied to these types. However, the function of the radio or *I-F* transformer is radically different and cannot be adequately studied aside from the associated circuit.

Figure 83*a* shows the schematic of an *I-F* transformer stage in which both windings of the transformer are tuned by shunting condensers. An equivalent circuit is shown in Fig. 83*b*. The pentode is shown replaced by a current of value $-g_m\mathbf{V}_1$, where $\mathbf{V}_1$ is the input signal voltage applied to the stage. This representation is justified by the Norton equivalence (Fig. 92*a*), and by the fact that the plate conductance g_p of the pentode

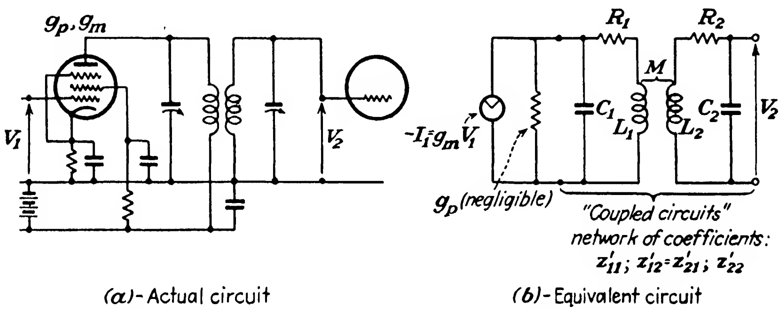


FIG. 83.—*Application of the transformer to an amplifier (coupled circuits). C_1 and C_2 include tube capacities.*

is negligibly small. The fundamentals of linear amplifier theory will be reviewed in Chap. XVI.

Evidently, it is the over-all performance of the stage of Fig. 83, not that of the transformer alone, which is of interest. The performance may be summed up entirely by the *complex voltage ratio* $\mathbf{V}_2/\mathbf{V}_1$. It should be noted that the capacities C_1, C_2 of the equivalent circuit include the stray and interelectrode values (Sec. 15.4) as well as the distributed capacities of the windings. If the frequency is very high, the distributed capacity of the windings, at best an artificial and approximate concept, is no longer significant and the entire lumped system representation must be used with caution.

The entire four-terminal network of Fig. 83*b* has the output terminals open-circuited; we wish to determine its open-circuit output voltage when the input current is known. Therefore, we require the coefficient z_{12}' for the entire network (Sec. 2.7), (prime symbols referring to the whole network as distinct from the transformer without shunting capacities). From z_{12}', the voltage ratio may be readily determined; we have, in fact

$$z_{12}' = \frac{\mathbf{V}_2}{\mathbf{I}_1} = \frac{-\mathbf{V}_2}{g_m \mathbf{V}_1}$$

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = -g_m z_{12}' \tag{477}$$

To find z_{12}', we must make use of expressions which relate the admittance and impedance parameters of a four-terminal network (Sec. 2.7). These can be obtained by comparison of the systems of linear equations (13) and (14) whose coefficients are the parameters in questions. Thus we obtain

$$z_{11} = \frac{y_{22}}{D_y} \qquad z_{22} = \frac{y_{11}}{D_y} \qquad z_{12} = -\frac{y_{12}}{D_y} \tag{478}$$

and the dual expressions

$$y_{11} = \frac{z_{22}}{D_z} \qquad y_{22} = \frac{z_{11}}{D_z} \qquad y_{12} = -\frac{z_{12}}{D_z} \tag{479}$$

In the above, $D_y = y_{11}y_{22} - y_{12}^2$ and $D_z = z_{11}z_{12} - z_{12}^2$ are the determinants of the two sets of parameters.

We have therefore for z_{12}'

$$z_{12}' = -\frac{y_{12}'}{D_y} = -\frac{y_{12}'}{y_{11}'y_{22}' - (y_{12}')^2} \tag{480}$$

The prime admittance parameters may be obtained easily from those of the transformer proper; for example, we have by inspection

$$y_{11}' = y_{11} + j\omega C_1$$

observing that addition of C_1 in shunt with the transformer input must add $j\omega C_1$ to its short-circuit input admittance. Similarly,

$$y_{22}' = y_{22} + j\omega C_2$$

The transfer admittance is not affected by the shunting capacities (see its definition, Sec. 2.7). Hence

$$y_{12}' = y_{12}$$

As a result, we may write (480) as follows:

$$z_{12}' = -\frac{y_{12}}{(y_{11} + j\omega C_1)(y_{22} + j\omega C_2) - y_{12}^2} \tag{481}$$

The admittance parameters of the transformer proper, which appear in the above, must in turn be expressed in terms of the known impedance parameters, using (479). Thus

$$
\begin{aligned}
z_{12}' &= \frac{z_{12}/D_z}{(z_{22}/D_z + j\omega C_1)(z_{11}/D_z + j\omega C_2) - z_{12}^2/D_z^2} \\
&= \frac{z_{12}D_z}{-\omega^2 C_1 C_2 D_z^2 + j\omega D_z(z_{22}C_1 + z_{11}C_2) + z_{11}z_{22} - z_{12}^2} \\
&= \frac{z_{12}}{1 + j\omega(C_2 z_{22} + C_1 z_{11}) - \omega^2 C_1 C_2 D_z} \qquad (482)
\end{aligned}
$$

The impedance coefficients for the dissipative transformer may be obtained from those of the lossless transformer (Sec. 14.5) by simply adding the winding resistances to z_{11} and z_{22}. In expressing them, however, we shall make use of the *dissipation factor* d defined in Sec. 4.1. This factor is the reciprocal of the coil Q (Sec. 9.5) and mathematically better suited to allow for the effect of dissipation in many cases. Thus we have

$$
\begin{aligned}
z_{11} &= R_1 + j\omega L_1 = j\omega L_1(1 - jd_1) \\
z_{22} &= R_2 + j\omega L_2 = j\omega L_2(1 - jd_2) \\
z_{11} &= j\omega M \\
D_z &= z_{11}z_{22} - z_{12}^2 = -\omega^2 L_1 L_2(1 - jd_1)(1 - jd_2) + \omega^2 M^2 \\
&= -\omega^2 L_1 L_2[1 - d_1 d_2 - \tau^2 - j(d_1 + d_2)] \qquad (483)
\end{aligned}
$$

where

$$
d_1 = \frac{R_1}{j\omega L_1} \qquad d_2 = \frac{R_2}{j\omega L_2} \qquad \tau = \frac{M}{\sqrt{L_1 L_2}} \qquad (484)
$$

Substituting into (482)

$$
z_{12}' = \frac{j\omega M}{1 - \omega^2 C_2 L_2(1 - jd_2) - \omega^2 C_1 L_1(1 - jd_1) + \omega^4 C_1 C_2 L_1 L_2[1 - d_1 d_2 - \tau^2 - j_1(d_1 + d_2)]}
$$

At this point, by the suitable choice of parameters, the expression may be put into more significant and convenient form. We shall use the *frequency number*, introduced generically in Sec. 4.1; in this case, the reference frequency will be the geometric mean of the two resonant frequencies, defined as follows:

$$
\omega_1 = \frac{1}{\sqrt{L_1 C_1}} \qquad \omega_2 = \frac{1}{\sqrt{L_2 C_2}} \qquad (485)
$$

Thus we have for the frequency number

$$
n = \frac{\omega}{\sqrt{\omega_1 \omega_2}} \qquad (486)
$$

We further note that

$$\omega^2 C_2 L_2 = \frac{\omega^2}{\omega_2^{\,2}} = \frac{\omega^2}{\omega_1\omega_2}\,\frac{\omega_1}{\omega_2} = n^2\,\frac{\omega_1}{\omega_2}$$

Similarly

$$\omega^2 C_1 L_1 = n^2\,\frac{\omega_2}{\omega_1}$$

and

$$\omega M = \omega\tau\sqrt{L_1L_2} = \frac{\omega^2\tau\sqrt{L_1L_2C_1C_2}}{\omega\sqrt{C_1C_2}} = \frac{\tau n^2}{\omega\sqrt{C_1C_2}}$$

Substituting into (484) and rearranging terms

$$z_{12}' = \frac{\tau}{\omega\sqrt{C_1C_2}}\;\frac{1}{n^2(d_1+d_2) - \left(d_1\dfrac{\omega_2}{\omega_1} + d_2\dfrac{\omega_1}{\omega_2}\right) + j\left[\dfrac{1}{n^2} - \left(\dfrac{\omega_1}{\omega_2} + \dfrac{\omega_2}{\omega_1}\right) + n^2(1 - d_1d_2 - \tau^2)\right]}$$

Finally, expressing the input and output voltages as from (477), we obtain

$$-\frac{k}{d_1+d_2}\frac{\mathbf{V}_1}{\mathbf{V}_2} = n^2 - a + j\frac{1}{d_1+d_2}\left[\frac{1}{n^2} - 2c + n^2(1-\epsilon)\right] \tag{487}$$

where

$$\left\{\begin{aligned} K &= \frac{\tau g_m}{\omega\sqrt{C_1C_2}} \\ a &= \frac{d_2(\omega_2/\omega_1) + d_2(\omega_1/\omega_2)}{d_1+d_2} \\ c &= \tfrac{1}{2}\left(\frac{\omega_1}{\omega_2} + \frac{\omega_2}{\omega_1}\right) \\ \epsilon &= \tau^2 + d_1d_2 \end{aligned}\right. \tag{488}$$

Up to this point, no term has been neglected in the analysis. There is danger in dropping terms, however small, in expressions including a large number of variables. Equation (487), such as it is, is subject to graphical interpretation, which in turn may be simplified with visibly small error.

To interpret (487), let us find the locus of $-\dfrac{K}{d_1+d_2}\dfrac{\mathbf{V}_1}{\mathbf{V}_2}$ for variable n. Small relative variations of frequency are assumed. This locus shows, except for an essentially constant multiplier, how $\mathbf{V}_1$ must vary in phase and magnitude when $\mathbf{V}_2$ is constant (except for frequency). This locus is studied rather than its inverse (locus of V_2 for constant V_1) because of its greater simplicity.

To obtain the locus we let

$$-\frac{K}{d_1 + d_2}\frac{\mathbf{V}_1}{\mathbf{V}_2} = x + jy \tag{489}$$

thus associating each value of the complex ratio with a point of the xy plane, in the usual way. In this plane, circles about the origin are therefore lines of *constant gain;* lines through the origin are lines of *constant phase angle.* We have, from (487),

$$x = n^2 - a$$
$$y = \frac{1}{d_1 + d_2}\left[\frac{1}{n^2} - 2c + n^2(1 - \epsilon)\right]$$

Eliminating n,

$$y = \frac{1}{d_1 + d_2}\left[\frac{1}{x + a} - 2c + (x + a)(1 - \epsilon)\right] \tag{490}$$

which is the equation of a hyperbola. The asymptotes of this hyperbola have equations

$$x = -a$$
$$y = \frac{x(1 - \epsilon)}{d_1 + d_2}$$

The first is parallel to the y axis; the second is inclined with respect to the y axis by the angle whose tangent is $\dfrac{d_1 + d_2}{1 - \epsilon}\cdot$ Since the dissipation factors are always small with respect to unity, the asymptotes are nearly parallel and the hyperbola may be assimilated to a parabola in the neighborhood of its vertex. Analytically, we obtain this result by expanding the term $1/(x + a)$ of Eq. (490) in a series of powers of

$$(x + a - 1) = n^2 - 1.$$

Since practical operation is confined to the neighborhood of the midband frequency (for which $n = 1$), high-order terms of this power expansion may be neglected. Letting for brevity

$$x' = x + a - 1 = n^2 - 1 \tag{491}$$

we may therefore write (490) in the form

$$y = \frac{1}{d_1 + d_2}\,[x'^2 - x'\epsilon + 2(1 - c) - \epsilon] \tag{492}$$

Adding and subtracting $\epsilon^2/4$ within brackets,

$$y = \frac{1}{d_1 + d_2}\left[\left(x' - \frac{\epsilon}{2}\right)^2 + 2(1 - c) - \epsilon\left(1 - \frac{\epsilon}{4}\right)\right] \tag{493}$$

Let us now introduce two new variables

$$\begin{cases} \xi = x' - \dfrac{\epsilon}{2} = x + a - \left(1 + \dfrac{\epsilon}{2}\right) = n^2 - \left(1 + \dfrac{\epsilon}{2}\right) \\ \eta = y + \dfrac{2(c - 1) + \epsilon\left(1 - \dfrac{\epsilon}{4}\right)}{d_1 + d_2} \end{cases} \tag{494}$$

In terms of ξ and η, the locus of $-\dfrac{K}{d_1 + d_2}\dfrac{\mathbf{V}_1}{\mathbf{V}_2}$, given by Eq. (493), reduces to a parabola in the *canonic form* of equation:

$$\eta = \frac{\xi^2}{d_1 + d_2} \tag{495}$$

This parabola can be easily drawn on the $\xi\eta$ plane. Along the parabola, a frequency scale could be marked with the help of the relation between ξ and n (494).

Considering $-\dfrac{K}{d_1 + d_2}\dfrac{\mathbf{V}_1}{\mathbf{V}_2}$ as a vector, or better, as a directed segment, one end of this segment will be on the parabola in question. The other end will be at point $x = 0$, $y = 0$, as implied in Eq. (489). This point must now be located in the $\xi\eta$ coordinate system, *i.e.*, with reference to the parabola we have drawn (Fig. 84). The coordinates for this point will be called ξ_0, η_0 and may be obtained from (494), in which x and y are made equal to zero. Thus

$$\xi_0 = a - \left(1 + \frac{\epsilon}{2}\right)$$
$$\eta_0 = \frac{2(c - 1) + \epsilon\left(1 - \dfrac{\epsilon}{4}\right)}{d_1 + d_2} \tag{496}$$

If ϵ is varied [because of changes in τ, Eq. (488)], the point

ξ_0, η_0 will describe some path. Evidently, if we could draw this path on the $\xi\eta$ plane on which we have already drawn the parabolic variable frequency path (495), we would have the desired information in very compact and accessible form. For each set of values for n and ϵ (frequency and coupling), two points could be located, one on each path. The directed segment joining the two points would give the voltage ratio in phase and magnitude.

The new path, as it happens, turns out to be another parabola identical to the variable frequency path, except that it is upside down and the vertices do not coincide. The steps may be verified by the reader; after eliminating ϵ in the system (496), the resulting second-degree equation may be put into the form

$$\eta_0 - \frac{2c - 1}{d_1 + d_2} = - \frac{(\xi_0 - a + 2)^2}{d_1 + d_2} \tag{497}$$

For the vertex of this variable τ parabola, we have the coordinates

$$\eta_0 = \frac{2c - 1}{d_1 + d_2} \qquad \xi_0 = a - 2 \tag{498}$$

The entire double-ended polar diagram is shown in Fig. 84 for the general case in which the tuning frequencies are not equal. The two end points are obtained by laying off appropriate functions of n and τ on the ξ scale, as shown. For convenience, distances along n are multiplied by $d_1 + d_2$; the ξ component of the resultant vector is likewise multiplied after carrying out the construction. Thus, the entire vector remains multiplied by $d_1 + d_2$, giving $-K \dfrac{\mathbf{V}_1}{\mathbf{V}_2}$. A simpler construction for the important case of $\omega_1 = \omega_2$ is given in Fig. 85a. This case is taken up in some detail in the next section.

If we hold the T point fixed on the diagram of Fig. 84 and let the F point sweep the variable n curve, we can see how the gain and phase shift vary with frequency, noting that the gain is inversely proportional to the length of the TF' segment. There will be two unequal gain maxima, or one maximum, depending on the position of T (or on the coupling coefficient). Plots of gain and phase shift against n for any τ and ω_1/ω_2 can be obtained from Fig. 84.

It should be noted that the quantity

$$a - 2 + \sqrt{2c - 1}$$

is a measure of the lack of symmetry in the characteristics. When this quantity is zero, the parabolas go through each other's vertices, as in the symmetrical case $\omega_1 = \omega_2$.

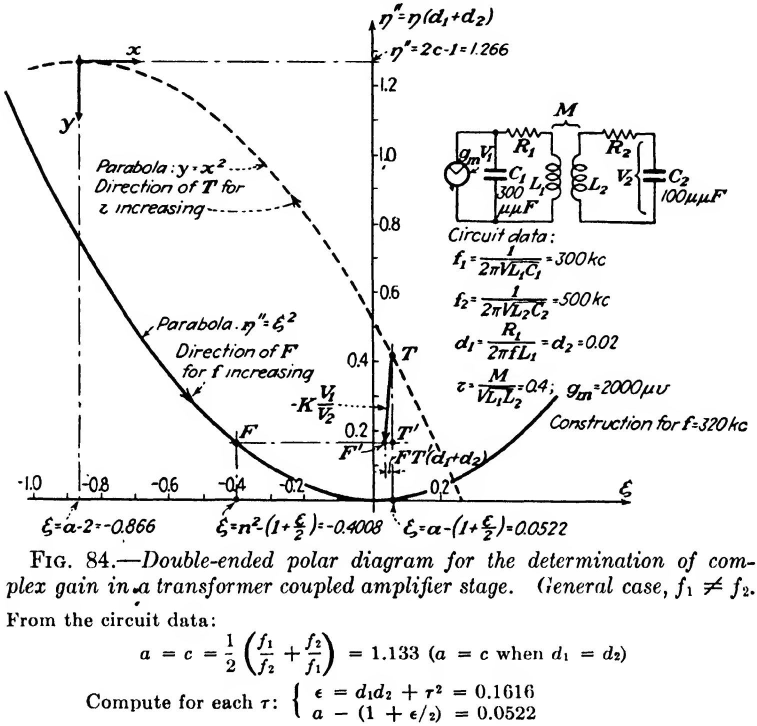


FIG. 84.—*Double-ended polar diagram for the determination of complex gain in a transformer coupled amplifier stage. General case,* $f_1 \neq f_2$.

From the circuit data:

$$a = c = \frac{1}{2}\left(\frac{f_1}{f_2} + \frac{f_2}{f_1}\right) = 1.133 \quad (a = c \text{ when } d_1 = d_2)$$

$$\text{Compute for each } \tau\text{:} \quad \begin{cases} \epsilon = d_1 d_2 + \tau^2 = 0.1616 \\ a - (1 + \epsilon/2) = 0.0522 \end{cases}$$

$$\text{Compute for each } f\text{:} \quad \begin{cases} n^2 = \dfrac{f^2}{f_1 f_2} = 0.68 \\ n^2 - (1 + \epsilon/2) = -0.4008 \\ K = \dfrac{\tau g_m}{2\pi f \sqrt{C_1 C_2}} = \dfrac{736{,}000}{f} = 2.3 \end{cases}$$

From the construction:

$$\frac{V_2}{V_1} = \frac{K}{TF'} = \frac{2.3}{0.25} = 9.2$$

$$\underline{/V_2} - \underline{/V_1} = -\widehat{TF'T'} = -1.49 \text{ radians}$$

Note: This method is accurate, provided $(n^2 - 1)^3 \ll 1$; the error is smaller near midband ($n = 1$).

When $d_1 \neq d_2$, $\quad a = \dfrac{d_1 f_1/f_2 + d_2 f_1/f_2}{d_1 + d_2} \quad c = \dfrac{1}{2}\left(\dfrac{f_1}{f_2} + \dfrac{f_2}{f_1}\right)$

Use of the double-ended polar diagram in connection with coupled circuits raises the question: What class of problems can be handled in this way? Consider a complex quantity $z = x + jy$, function of a number of parameters α, β, γ, and suppose we wish to chart or analyze the variations of z with both α and β. The most general method consists of drawing, on the xy plane, a family of polar diagrams or loci of z for constant values of β and variable α; then repeating for constant α and variable β, (γ being constant for both families). This, however, requires the tracing of a double family of curves, all of which have to be redrawn each time γ is changed. Suppose, however, the locus of z for variable α can be written in the form

$$F(x + B_1, y + B_2, \gamma) = 0 \tag{499}$$

which means that F is an algebraic or transcendental expression in $x + B_1$, $y + B_2$ (where B_1 and B_2 are functions of β and γ), the coefficients of F being functions of γ only. In this case we can proceed as we did after writing Eq. (493) of the preceding discussion, *i.e.*, adopt a new coordinate system

$$\xi = x + B_1$$
$$\eta = y + B_2$$

On the $\xi\eta$ plane, the single curve

$$F(\xi, \eta, \gamma) = 0$$

takes the place of the family of variable α loci. It now becomes necessary to trace out the path described by the origin of the xy axes upon the $\xi\eta$ plane for variable β. This is done by writing

$$\xi_0 = B_1$$
$$\eta_0 = B_2$$

and eliminating β from these two equations. In conclusion, the system is applicable whenever the polar diagram of $z(\alpha\beta\gamma)$ for variable α can be put in the form (499). It should be noted that if this may often be accomplished by suitably selecting α: α may be a function of several among the original parameters used in the analysis.

14.10. Synchronous Coupled Circuits: Critical Coupling. When the two tuning frequencies

$$\omega_1 = \frac{1}{\sqrt{L_1C_1}} \qquad \omega_2 = \frac{1}{\sqrt{L_2C_2}}$$

of the coupled circuits network are equal (we shall call such circuits *synchronous* for brevity), the diagram takes the form of Fig. 85*a*. It is often possible (when $\tau \ll 1$) to draw only the lower end of the variable frequency locus, together with that part of the variable τ locus which corresponds to low values of τ. This section of the variable τ locus may be assimilated to a straight line. Hence the simpler construction of Fig. 85*b*,

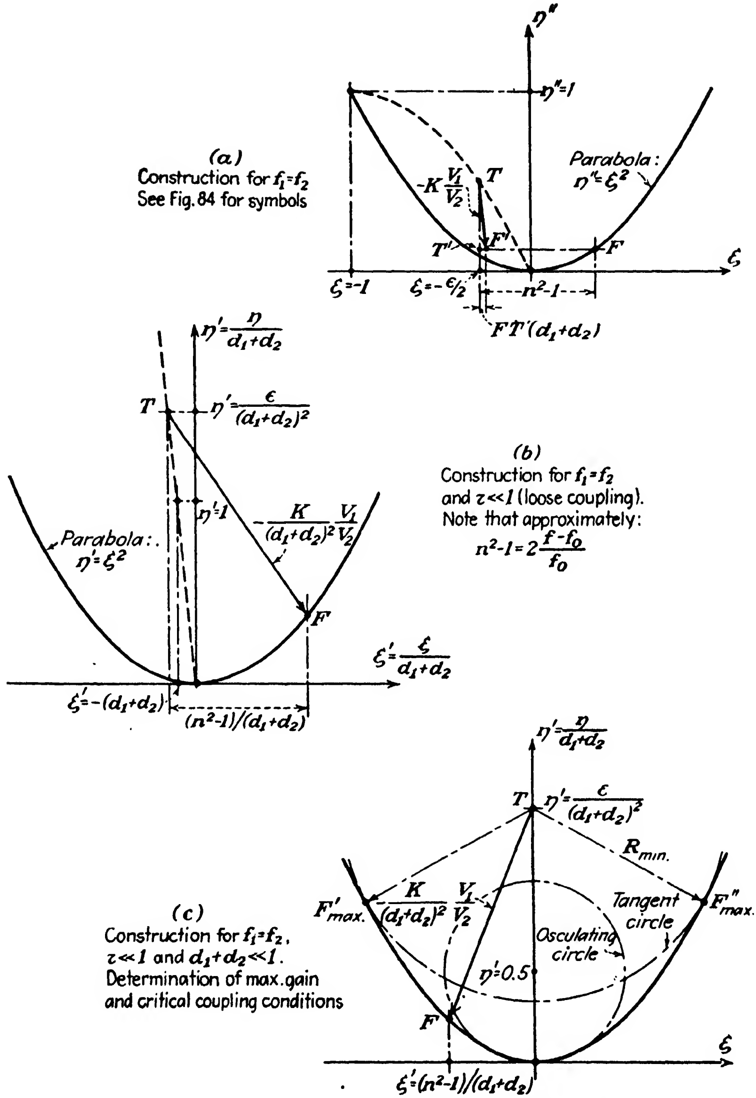


FIG. 85.—*Particular cases of the construction of Fig. 84.*

which may be used effectively to analyze synchronous coupled circuits under all practical operating conditions.

The construction may be simplified still further (Fig. 85c) when the dissipation factors are so small that the variable τ locus may be assimilated with the η axis.

Plots of over-all gain and phase angle, obtained from Fig. 85c, are reproduced in Fig. 86. It will be noted that the plots are based on a common value, d, of the dissipation factors for the two coils. They can, however, be used when d_1 and d_2 are not equal; in this case the values of d and τ to be used in selecting the appropriate plot and computing the constants must be *equivalent* values, obtained in the following manner:

$$\text{Equivalent } d = \tfrac{1}{2}(d_1 + d_2) \tag{500}$$

$$\text{Equivalent } \tau = \sqrt{\tau^2 - \tfrac{1}{4}(d_1 - d_2)^2} \tag{501}$$

To justify this, we note that τ has effect on the diagram of Fig. 85c only through the value of

$$\epsilon = d_1 d_2 + \tau^2 \tag{488}$$

In the case $d_1 = d_2 = d$, the above becomes

$$\epsilon = d^2 + \tau^2$$

Now, if for d we substitute the equivalent value $\frac{1}{2}(d_1 + d_2)$, we must substitute an equivalent value for τ also, so that ϵ may have the correct value for the case $d_1 \neq d_2$, as given by (488) above. Hence

$$(\tau_{eq})^2 + \tfrac{1}{4}(d_1 + d_2)^2 = \tau^2 + d_1 d_2$$

from which we obtain (501).

We also note from the plots, as from the polar diagram, that the gain characteristic is symmetrical for small values of τ, in the sense that the gain is the same for any two frequencies whose geometric mean is the midband, or common tuning frequency. This is, in general, the meaning of *symmetry* as applied to network characteristics. The characteristics are not actually symmetrical unless plotted against the logarithm of frequency or over a narrow frequency range.

The location and height of the peaks in the gain characteristic may be determined analytically in terms of τ on the basis of

Fig. 85c. This could be done by writing an expression for the length TF in terms of n and τ. This expression would actually give us the equation of the gain curve; by taking its derivative

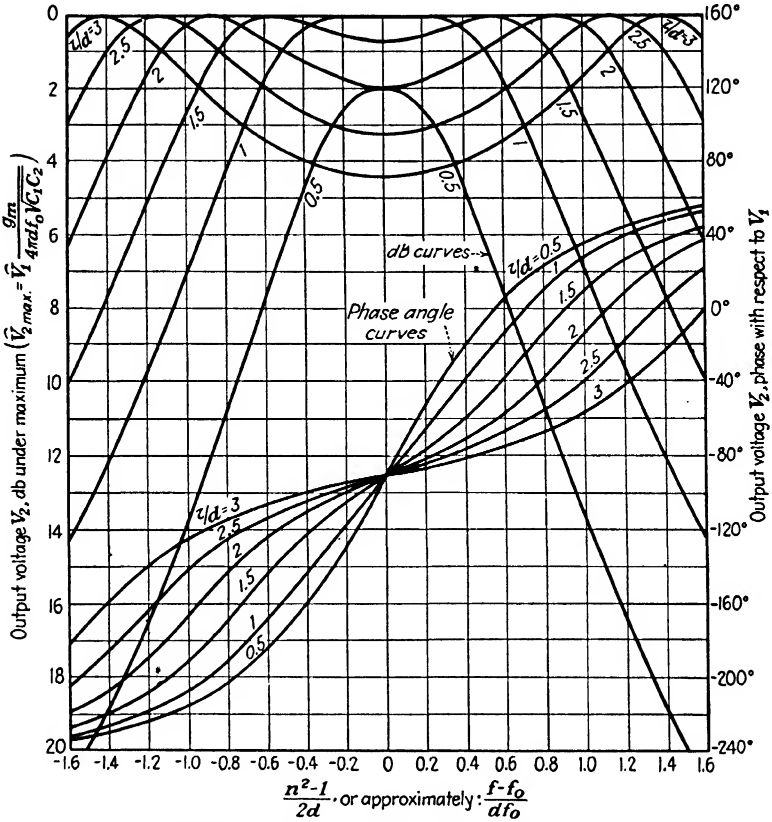


FIG. 86.—*Universal plots of gain and phase shift for synchronous coupled circuits. Plots are drawn for the case* $f_1 = f_2 = f_0$, $d_1 = d_2 = d$ *(see Fig. 84). They are based on the construction of Fig. 85c.*

If $d_1 \neq d_2$, plots can be used if based on equivalent values:

$$\tau_{eq} = \sqrt{\tau^2 - \frac{(d_1 - d_2)^2}{4}} \cdot \quad d_{eq} = \frac{d_1 + d_2}{2} \qquad g_{meq} = g_m \frac{\tau}{\tau_{eq}}$$

with respect to n and setting it equal to 0, we could obtain the desired results. However, a simpler method consists in writing the equation of a circle with center in T (Fig. 85c) and generic radius R.[1] This done, we eliminate ξ from this equation and that of the parabola; the resulting equation in η will have two roots (not four as would be the case for a parabola and a circle in generic mutual relationship). This is because the four intersections of the circle and the parabola are symmetrically placed about the axis of the parabola; a single value of η corresponds to two intersections. Finally, we impose the condition that the two solutions for η at the intersections reduce to one, thus specifying that the circle and the parabola must be tangent. This assigns a value to R, the radius of the circle, and to η for the points of tangency, hence, respectively, for the maximum gain and frequency of maximum gain.

The equation of the circle is

$$\left[\eta - \left(\frac{\epsilon}{d_1 + d_2}\right)\right]^2 + \xi^2 = R^2 \tag{502}$$

and that of the parabola

$$\eta = \frac{\xi^2}{d_1 + d_2}$$

Eliminating and expanding,

$$\eta^2 + \eta\left[d_1 + d_2 - 2\frac{\epsilon}{d_1 + d_2}\right] + \left(\frac{\epsilon}{d_1 + d_2}\right)^2 - R^2 = 0$$

Hence the values of η at the intersections

$$\eta = \frac{\epsilon}{d_1 + d_2} - \frac{d_1 + d_2}{2} \pm \sqrt{\left[\frac{\epsilon}{d_1 + d_2} - \frac{d_1 + d_2}{2}\right]^2 + R^2 - \left(\frac{\epsilon}{d_1 + d_2}\right)^2}$$

The discriminant vanishes, and the intersections coincide into a point of tangency, when

$$R = R_{\min} = \sqrt{\epsilon - \left(\frac{d_1 + d_2}{2}\right)^2} \tag{503}$$

At the same time η takes the value, for the point of tangency,

$$\eta = \frac{\epsilon}{d_1 + d_2} - \frac{d_1 + d_2}{2} \tag{504}$$

[1] The coordinates of Fig. 85c are $\eta' = \eta/(d_1 + d_2)$ and $\xi' = \xi/(d_1 + d_2)$. This change does not affect the configuration, yet makes it possible to draw a single parabola $y' = \xi'^2$ for all possible values of d_1 and d_2.

From (503), substituting (488), and noting that

$$R = \overline{TF} = \frac{K}{(d_1 + d_2)^2} \frac{V_1}{V_2}$$

$$\therefore \left(\frac{V_2}{V_1}\right)_{(\max)} = \frac{K}{R_{\min}(d_1 + d_2)^2}$$

we have the maximum value of gain for the condition when the gain curve has *two separate peaks* (overcoupling)

$$\left(\frac{V_2}{V_1}\right)_{\max} = \frac{g_m}{\omega_0 \sqrt{C_1 C_2}(d_1 + d_2)\sqrt{1 - \left(\frac{d_1 - d_2}{2\tau}\right)^2}} \tag{505}$$

[*Maximum value of gain for overcoupled circuits* $(\omega_1 = \omega_2 = \omega_0)$]

$$\left(\frac{V_2}{V_1}\right)_{\max} = \frac{g_m}{2d\omega_0 \sqrt{C_1 C_2}} \tag{506}$$

[*As above, when* $d_1 = d_2 = d$]

where $d_1 = R_1/\omega L_1$, $\omega_1 = 1/\sqrt{L_1 C_1}$, etc.

In the above, ω_0 has been put in place of ω with small error. It is instructive to compare (506) with the maximum gain obtainable by tuning the joint capacity $C_1 + C_2$ with a single coil of dissipation d (gain of the tuned output amplifier). The comparison is valid if C_1 and C_2 are considered to be limiting factors of the gain, not if their values have to be adjusted by means of variable condensers in order to tune out given inductances.

The two gains compare as follows:

$$\frac{\text{Maximum gain (transformer coupled)}}{\text{Maximum gain (tuned output)}} = \frac{1}{2}\left(\sqrt{\frac{C_1}{C_2}} + \frac{C_2}{C_1}\right) \tag{507}$$

This expression is greater than unity when the two capacities are not equal, in which case the turn ratio of the transformer should be, noting that $\sqrt{L_1 C_1} = \sqrt{L_2 C_2}$

$$t = \sqrt{\frac{L_2}{L_1}} = \sqrt{\frac{C_1}{C_2}}$$

As an example, consider a tube having plate to cathode capacitance (Sec. 15.4) of 20 $\mu\mu f$, to be coupled into a grid to cathode capacitance of 3 $\mu\mu f$ at a single fixed frequency. The ratio (507) would be in this case

$$\frac{2.968}{2} \approx 1.482$$

showing that there is a limited advantage in the use of transformer coupling, in certain cases, from the standpoint of gain as well as bandwidth.

From Eq. (504) we may obtain the frequencies of maximum gain. We have, for the points of tangency F_{max}', F_{max}'' (Fig. 85c)

$$\eta = \frac{\epsilon}{(d_1 + d_2)} - \tfrac{1}{2}(d_1 + d_2)$$

Hence, and from the equation of the parabola $\left(\eta = \frac{\xi^2}{d_1 + d_2}\right)$

$$\xi = \sqrt{\epsilon - \tfrac{1}{2}(d_1 + d_2)^2}$$

Noting that, approximately, $\xi = n^2 - 1$ [Eq. (494)], comparing with the above, solving for n, and expanding ϵ (488),

$$n^2 = 1 \pm \sqrt{\tau^2 + d_1 d_2 - \frac{(d_1 + d_2)^2}{2}}$$

$$\therefore n = \sqrt{1 \pm \sqrt{\tau^2 - \frac{d_1^2 + d_2^2}{2}}}$$

Finally, expanding by means of the binomial theorem,

$$\omega_{max} = \omega_0 \left[1 \pm \tfrac{1}{2}\sqrt{\tau^2 - \frac{d_1^2 + d_2^2}{2}}\right] \tag{508}$$

[*Frequencies of maximum gain for overcoupled circuits* ($\omega_1 = \omega_2 = \omega_0$)]

$$\omega_{max} = \omega_0[1 \pm \tfrac{1}{2}\sqrt{\tau^2 - d^2}] \tag{509}$$

[*As above, when* $d_1 = d_2 = d$]

If the coupling coefficient is fairly large compared to the dissipation factors, the *separation* of the peaks, expressed as a fraction of midband frequency, is sensibly equal to τ itself, thus providing means for measuring this quantity.

The separation vanishes when, in the general case

$d_1 \neq d_2$ (508)

$$\tau = \sqrt{\frac{d_1^2 + d_2^2}{2}} \tag{510}$$

[*Value of* τ *for critical coupling* ($\omega_1 = \omega_2 = \omega_0$)]

$$\tau = d \tag{511}$$

[*As above, when* $d_1 = d_2 = d$]

In a geometric sense, critical coupling occurs when the circle of Fig. 85c osculates the parabola; in this case the gain curve has zero curvature at its peak, which is now at the center of symmetry. Under these conditions, a finite frequency interval is transmitted with low values of amplitude and phase distortion (Fig. 86). The maximum gain value for critical coupling is obtained by substituting (510) into (505). Thus

$$\left(\frac{V_2}{V_1}\right)_{\max} = \frac{g_m}{\omega_0 \sqrt{C_1 C_2}} \frac{\sqrt{2(d_1^2 + d_2^2)}}{(d_1 + d_2)^2} \tag{512}$$

[*Maximum gain for critical coupling* $(\omega_1 = \omega_2 = \omega_0)$]

For the condition $d_1 = d_2 = d$, maximum gain for critical coupling has the value of Eq. (506). Uneven distribution of dissipation in the two coils (inequality of Q values) has a slight effect on the gain at critical coupling. Comparing the gain for $d_1 = d_2 = d$ with the gain for $d_1 d_2 = d^2$, $d_2 = 4d_1$, we find

$$\frac{\text{Gain } (d_1 = \frac{1}{4}d_2)}{\text{Gain } (d_1 = d_2)} = 0.933 \qquad \text{(the constant } d_1 d_2\text{)}$$

When τ is below the critical value, the gain curve goes through a maximum at midband and has a curvature there. We have under these conditions

$$K \frac{\mathbf{V}_1}{\mathbf{V}_2} = \frac{\epsilon}{d_1 + d_2}$$

from which we obtain

$$\left(\frac{V_2}{V_1}\right)_{\max} = \frac{\tau g_m}{\omega_0 \sqrt{C_1 C_2}(\tau^2 + d_1 d_2)} \tag{513}$$

[*Maximum gain for undercoupled circuits* $(\omega_1 = \omega_2 = \omega_0)$]

When the coupling is very loose ($\tau^2 \ll d_1 d_2$), the gain is approximately proportional to τ.

14.11. Applications and examples.

Amplitude distortion due to the transformer inserted between resistive terminations. The transformer (magnetic coupling) provides a solution for the transformer problem in the audio range. As defined in Sec. 8.1, this problem calls for uniform transmission over a wide band when the terminations have resistive values. The performance of a transformer in this respect may be appraised by the variation of transmitted power with frequency (amplitude distortion). Instead of computing, or plotting, transmitted power, we may conveniently compute the db or neper *loss* in transmitted power (Sec. 1.9) referred to the maximum amount of power that could flow between the given terminations. This method has been used before (Sec. 9.5);

the loss in question is called the *reflection loss* (Sec. 6.6). It is obtained (when small) by the approximate formula (184)

$$L_R = \tfrac{1}{8}\left(\frac{|\Delta Z|}{R_g}\right)^2$$

where $|\Delta Z|$ is the distance, on the Z plane, from the input impedance point Z_i to the generator resistance R_g.

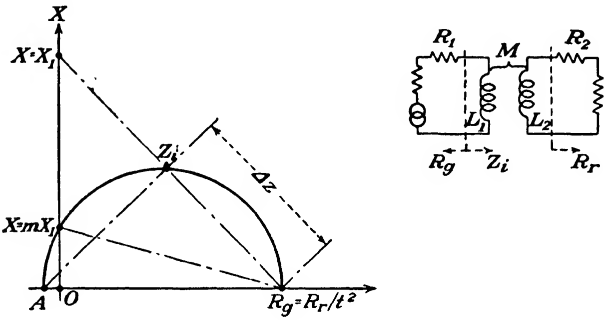


FIG. 87.—*Transformer with resistive load.*

We can evaluate ΔZ, hence L_R, by means of the input impedance construction of Fig. 79. We must assume that R_g, generator resistance, includes the resistance of the primary and R_r, load resistance, that of the secondary, and that the turn ratio of the transformer has the value

$$t = \sqrt{\frac{R_r}{R_g}}$$

Under these conditions the diagram takes the form of Fig. 87. We deduce, by comparing similar triangles, the proportions:

$$\frac{|\Delta Z|}{\overline{AO} + R_g} = \frac{R_g}{X_1} \qquad \frac{\overline{AO}}{mX_1} = \frac{mX_1}{R_g}$$

from which we obtain

$$\frac{|\Delta Z|}{R_g} = \frac{R_g}{X_1} + m^2\frac{X_1}{R_g} \tag{514}$$

Hence, the reflection loss

$$L_R = \tfrac{1}{8}\left(\frac{R_g}{X_1} + m^2\frac{X_1}{R_g}\right)^2 \text{ neper} \tag{515}$$

$$X_1 = \omega L_1 \qquad m^2 = 1 - \tau^2$$

[*Reflection loss for the transformer between resistive terminations*]

The reflection loss is a minimum, as may be verified by differentiating, when frequency is such that the terms under brackets are equal, or

$$X_1 = \frac{R_g}{m} \tag{516}$$

in which case the loss takes the minimum value

$$L_{R\min} = \frac{m^2}{2} \text{ neper} \tag{517}$$

As an example, assume $R_g = 10{,}000$ ohms (including primary resistance, which may be estimated approximately). We wish to determine the minimum values of primary inductance and coupling coefficient for a transformer which is to operate between 10 and 1,000 c.p.s. with less than 2 db amplitude distortion.

Since the minimum reflection loss, given by (517), is always exceedingly small for closely coupled transformers, we may identify the distortion with the maximum L_R within the useful range. For optimum utilization we will therefore set $L_R = 2$ db at both extremes of the range. Thus

$$\frac{2}{8.68} = \tfrac{1}{8}\left(\frac{1}{x'} + m^2x'\right) = \tfrac{1}{8}\left(\frac{1}{x''} + m^2x''\right) \tag{518}$$

having let

$$x' = \frac{X_1}{R_g}_{(f=10)} = \frac{2\pi \times 10 \times L_1}{R_g}$$
$$x'' = \frac{X_1}{R_g}_{(f=1{,}000)} = 100x'$$

Eliminating m^2 from (518), we have

$$x''\left(1.356 - \frac{1}{x'}\right) = x'\left(1.356 - \frac{1}{x''}\right)$$

hence

$$x' = \frac{100 - (1/100)}{1.356(100 - 1)} = 0.745$$

and for the primary inductance

$$L_1 = \frac{R_gx'}{2\pi \times 10} = \frac{10{,}000 \times 0.745}{2\pi \times 10} = 118.5 \text{ henrys}$$

Now m^2 may be obtained from (518)

$$m^2 = \frac{1.356 - (1/x')}{x'} = 0.0188$$

or

$$\tau = \sqrt{1 - m^2} = 0.9906$$

The values obtained are independent of the turn ratio. This does not affect amplitude distortion for given values of τ and L_1. If the turn ratio is high, however, a large primary inductance may be difficult to obtain. It should be noted that the winding capacitance has been neglected in the discussion. This is therefore applicable only to low frequencies.

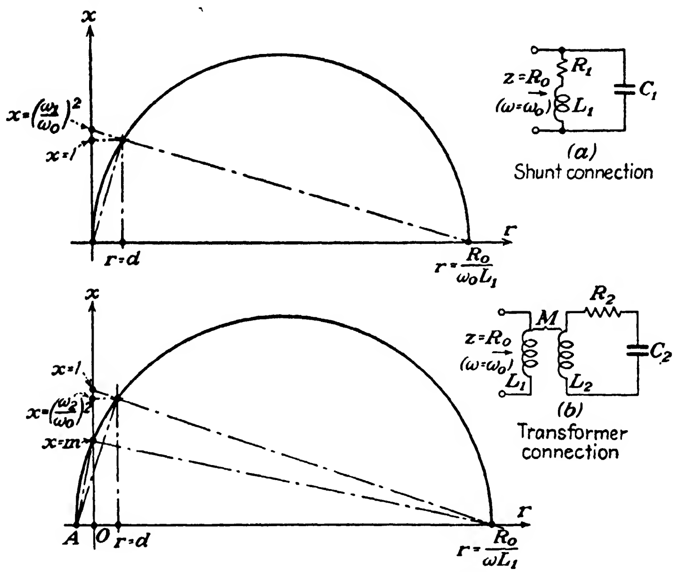


FIG. 88.—*Comparison of resonant systems.*

Comparison of resonant impedance of a tank circuit and of a transformer with capacitive load. Selective two-poles of the antiresonant type, having a high value of resonant impedance (or resonant resistance R_0), are frequently called for. The simplest of such poles is the parallel LC combination (Fig. 88*a*); a possible alternative is offered by a transformer with capacitive load (Fig. 88*b*). A third alternative might be the autotransformer or tapped coil connection. This may be reduced to an equivalent transformer connection by the relations (474) and (475). We will now compare the first and second alternatives, from two standpoints, namely,

1. Assuming that the *same coil* is resonated, first by a condenser across it (Fig. 88*a*), then by a condenser through a magnetic coupling (Fig. 88*b*), at the same frequency in both cases;
2. Assuming that the *capacity* and the frequency are the same in both cases.

The comparison may be carried out on the basis of the Z-plane diagrams. These are given, for the two alternatives, in Fig. 88. The diagram for the tank circuit is based on the principles discussed in Sec. 8.3; that for the transformer circuit, on the construction of Fig. 79. For purposes of comparison, both diagrams have been *normalized*, by making the scale unit equal to ωL_1 in both cases.

Broken down into similar triangles, Fig. 88*a* gives

$$\frac{d}{1} = \frac{(\omega_1/\omega_0)^2}{R_0/\omega_0 L_1}$$

$$\frac{(\omega_1/\omega_0)^2 - 1}{d} = \frac{d}{1}$$

where

$$d = \frac{R}{\omega_0 L_1} \qquad \omega_1 = \frac{1}{\sqrt{L_1 C_1}} \qquad \omega_0 = \text{resonant frequency}$$

From the above we obtain

$$\begin{cases} \omega_0 = \dfrac{\omega_1}{\sqrt{1 + d^2}} & (519) \\ R_0 = \omega_0 L_1 \dfrac{1 + d^2}{d} \approx \dfrac{\omega_0 L_1}{d} & (520) \end{cases}$$

These conclusions are well known, of course. Dissipation causes a slight shift of the resonant frequency, and the resonant impedance is the reactance of the coil at resonance multiplied by the coil Q.

Now consider Fig. 88*b*. Although more complicated, this may be analyzed in a similar manner. Comparing pairs of similar triangles, we have the proportions

$$\frac{R_0/\omega_0 L_1}{m} = \frac{m}{\overline{AO}} \qquad (521)$$

$$\frac{(R_0/\omega_0 L_1) - d_2}{(\omega_2/\omega_0)^2} = \frac{(\omega_2/\omega_0)^2}{AO + d_2} \tag{522}$$

$$\frac{R_0/\omega_0 L_1}{1} = \frac{d_2}{1 - (\omega_2/\omega_0)^2} \tag{523}$$

From (523) we have

$$\left(\frac{\omega_2}{\omega_0}\right)^2 = 1 - \frac{d_2}{R_0/\omega_0 L_1} \tag{524}$$

and with very small error

$$\left(\frac{\omega_2}{\omega_0}\right)^4 = 1 - 2\frac{d_2}{R_0/\omega_0 L_1}$$

From the above, (521) and (524),

$$\left(\frac{1 - \tau^2}{r_0} + d_2\right)(r_0 - d_2) = 1 - 2\frac{d_2}{r_0} \tag{525}$$

having let

$$\tau^2 = 1 - m^2 \qquad r_0 = \frac{R_0}{\omega_0 L_1}$$

Solving (525) for r_0

$$r_0 = \frac{\tau^2 + d_2^2}{2d_2}\left[1 \pm \sqrt{1 - 4\frac{(1 + \tau^2)d_2^2}{\tau^2 + d_2^2}}\right]$$

Discarding the negative sign which does not correspond to the configuration of Fig. 88b, clearing and expressing R_0,

$$R_0 = \omega_0 L_1 \frac{\tau^2 + d_2^2 + \sqrt{(\tau^2 - d_2)^2 - 4d_2^2}}{2d_2} \tag{526}$$

In common cases when $d_2 \ll \tau$, this and (524) become approximately

$$\begin{cases} R_0 = \dfrac{\omega_0 L_1 \tau^2}{d_2} & (527) \\[2ex] \omega_0 = \dfrac{\omega_2}{\sqrt{1 - (d_2/\tau)^2}} & (528) \end{cases}$$

[*Resonant impedance and resonant frequency of transformer with capacitive load* $(d_2 \ll \tau)$]

If L_1 is held the same in the two circuits (Figs. 88a and 88b), we find that the resonant impedance is the same in both cases only for tight coupling, assuming the dissipation factor to be the same for coils L_1 and L_2.

If the capacity is the same (C for both circuits), we have, approximately, taking the resonant frequency to have the lossless value

For Fig. 88*a* (from Eq. 520):

$$R_0 = \frac{1}{d_1\omega_0 C} \tag{529}$$

For Fig. 88*b* (from Eq. 527):

$$R_0 = \frac{1}{d_2\omega_0 C}\left(\frac{\tau}{t}\right)^2 \tag{530}$$

$$\left[t = \sqrt{\frac{L_2}{L_1}}\right]$$

It may be possible, therefore, to obtain a higher resonant impedance, using a given tuning condenser, by the method of Fig. 88*b* than by the direct parallel connection. It is necessary for this purpose to use a turn ratio smaller than the coefficient of coupling.

CHAPTER XV

CAPACITIVE COUPLING

15.1. Systems of Several Charged Bodies: Partial Capacities. In the preceding chapter, an abstract system of two neighboring currents was considered first; this led to general relations between the currents and the magnetic fluxes linked with each, which in turn enabled us to derive the network equations for the transformer. Mutual relationships between flux and current values in systems of many currents were also touched upon briefly. From a practical standpoint, this general case is not as important as the particular case of two currents.

As the dual of a multicurrent system, in a general sense, we may consider a multivoltage system. In place of a number of closed paths, or loops, in which flow well-defined values of current (one value for all cross sections of the same loop), we can imagine a number of closed surfaces, any two of which define a value of voltage (the same for all paths between the two surfaces).

A system of charged conductors immersed in a dielectric medium fills these requirements, provided the charges do not vary too rapidly with time (Sec. 12.6).

If only two conductors are present, the displacement current flowing across the dielectric can be considered due to a condenser connected between the conductor surfaces. In this equivalent representation, the geometry and mutual position of the two conductors no longer have effect on the field of the system, which is now localized between the condenser plates. This idealization, which, in place of a physical system identified by its geometrical configuration, considers a circuit branch identified by a value of capacity, can be carried over to a system of many conductors. In this case the equivalent systems include as many condensers as there are voltages in the system or combinations of conductors taken two at a time.

This equivalence is shown in Fig. 89. The condensers of the equivalent system are called *partial capacities*, and we shall call the electric flux in each condenser the *partial flux* associated with the pair of conductors across which the condenser is ideally connected. Thus, Ψ_{jk} is the partial flux between conductors j and k; C_{jk}, the partial capacity, may be written

$$C_{jk} = \frac{\Psi_{jk}}{v_{jk}} \tag{531}$$

where v_{jk} is the voltage between j and k.

The equivalence of Fig. 89 may be justified in a roundabout manner, as will be shown in the next section.

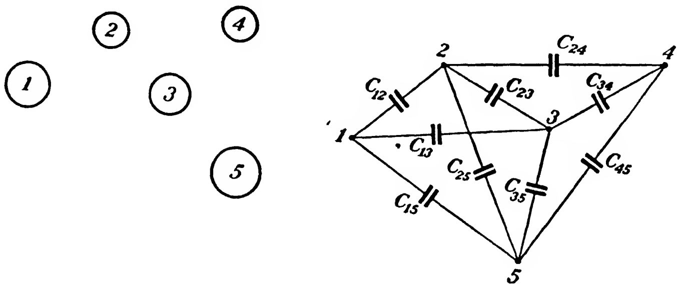


FIG. 89.—*Equivalent representation of a system of n conductors ($n = 5$) illustrating partial capacities.*

15.2. Capacity and Potential Coefficients. The total electric flux leaving conductor j, or the charge Q of this conductor, is given by the sum of the partial fluxes

$$Q_j = \Psi_{j1} + \Psi_{j2} + \cdots + \Psi_{jn}$$

or, using (531),

$$Q_j = C_{j1}v_{j1} + C_{j2}v_{j2} + \cdots + C_{jn}v_{jn}$$

where v_{jk} is the voltage over a path from conductor j (first subscript) to k (second subscript), and $v_{jk} > 0$ when the field and the path have the same direction or when j has higher potential. We could obtain in this way the system

$$\begin{aligned} Q_0 &= \sum_{k \neq 0} C_{0k}v_{0k} \\ Q_1 &= \sum_{k \neq 1} C_{1k}v_{1k} \\ &\cdots\cdots\cdots \\ Q_n &= \sum_{k \neq n} C_{nk}v_{nk} \end{aligned} \tag{532}$$

for the conductors 0, 1, 2, . . . n.[1]

[1] The notation $\sum_{k \neq j}$ designates a sum of terms, in which the subscript k takes all values from 1 to n successively, except the value j.

Let us now consider, rather than the voltages between conductors, the voltage of each conductor to conductor 0, considered as a *ground*, or the potential U of the conductor. In terms of potentials, we have in general

$$v_{jk} = v_{j0} + v_{0k} = U_j - U_k$$

Hence, the expression for Q_j becomes

$$Q_j = \sum_{k \neq j} [C_{jk}(U_j - U_k)] = \left[\sum_{k \neq j} C_{jk}\right] U_j - \sum_{k \neq j} [C_{jk}U_k]$$

The right-hand side of the above is a homogeneous polynomial in the n potentials $U_1, U_2, U_3 \ldots U_n$ and can be written

$$Q_j = \sum_{k=1}^{n} q_{jk}U_k$$

where

$$\begin{aligned} q_{j1} &= -C_{j1} \\ q_{j2} &= -C_{j2} \\ &\cdots\cdots\cdots \\ q_{jj} &= \sum_{k \neq j} C_{jk} \end{aligned}$$

The q coefficients are called *coefficients of capacity*. They are equal and opposite to the partial capacities (hence negative) except when they have two equal indices. In view of the significance of the partial capacities (Fig. 88), it is evident that

$$C_{jk} = C_{kj}$$

hence also

$$q_{jk} = q_{kj}$$

The differentiation between partial capacities and capacity coefficients has been stressed in order to avoid confusion. For practical engineering purposes it is much better to avoid the q coefficients entirely, using only the C parameters, which have concrete physical meaning.

From a mathematical standpoint, however, the q coefficients are convenient. In terms of these, the relation between *charges* and *potentials* is as follows:

$$\left\{ \begin{aligned} Q_1 &= q_{11}U_1 + q_{12}U_2 + \cdots + q_{1n}U_n \\ Q_2 &= q_{21}U_1 + q_{22}U_2 + \cdots + q_{2n}U_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Q_n &= q_{n1}U_1 + q_{n2}U_2 + \cdots + q_{nn}U_n \end{aligned} \right. \tag{533}$$

From this we may derive a similar set of equations, expressing the *potentials* in terms of the *charges*

$$\begin{cases} U_1 = p_{11}Q_1 + p_{12}Q_2 + \cdots + p_{1n}Q_n \\ U_2 = p_{21}Q_1 + p_{22}Q_2 + \cdots + p_{2n}Q_n \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ U_n = p_{n1}Q_1 + p_{n2}Q_2 + \cdots + p_{nn}Q_n \end{cases} \tag{534}$$

where p_{11}, p_{12}, p_{13} . . . are the *potential coefficients.*

Equations (534) may be regarded as a consequence of the *principle of superposition* applied to the potential due to several superimposed distributions of charge density. Suppose a given distribution σ_j causes the total charge to have a value Q_j on conductor j and 0 on all other conductors and causes the potential on conductor k to have the value U_{kj}. It is evident that if σ_j were multiplied by a factor n at every point, the charge Q_j would likewise be multiplied by n and so would the potential U_{kj}, by the superposition principle. Hence, U_{kj} and Q_j must be proportional, or

$$U_{kj} = p_{kj}Q_j$$

Now consider a new distribution σ_h which causes a charge Q_h to appear on conductor h and no charge on any other conductor. By virtue of σ_h, the potential at k will have the value

$$U_{kh} = p_{kh}Q_h$$

and by virtue of both σ_j and σ_h existing together, it will have (again because of superposition) the value

$$U_{kj} + U_{kh} = p_{kj}Q_j + p_{kh}Q_h$$

Continuing this way, we would obviously arrive at Eqs. (534). From these, retracing our steps, we would finally come to the conclusion that the partial capacities C_{11}, C_{12}, . . . are constants.

A full discussion of this subject can be found in the literature.[9] In the following, systems of three conductors (or two conductors and ground) will be taken up in particular detail, as the electrostatic system formed by the electrodes of vacuum tubes falls within this category.

15.3. Systems of Three Conductors: the Capacitive Π Network. Consider a system of three conductors, of which one is grounded, as in Fig. 90*a*. Particular attention should be given to the word *grounded.* It would be more correct to say that the third conductor *shields* the others. Whether or not this conductor is in contact with earth is immaterial. The important thing is that no fourth conductor must have appreciable effect on the field of the system; and this can evidently be true only if one of the three conductors shields the remaining two. The term *grounded* used in this connection actually means that a *closed system* is considered.

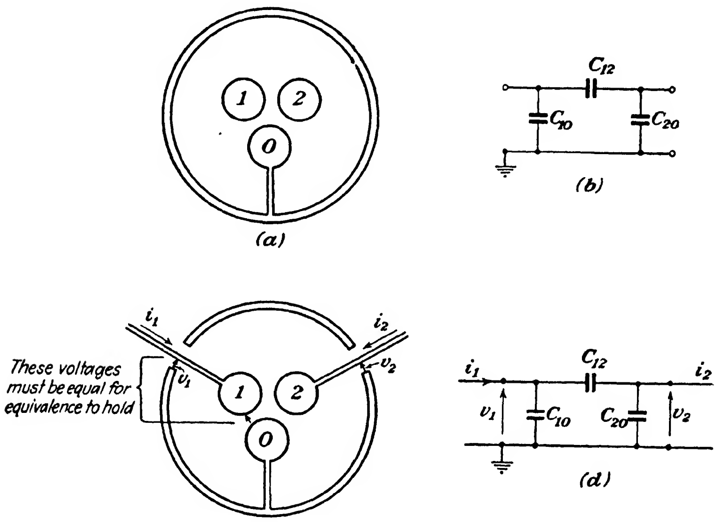


FIG. 90.—*Three-conductor system and equivalent* Π *network.*

In some cases the closed system is actually delimited by the earth's surface, which then constitutes the third conductor. An example in point is a two-conductor open wire line. If the line spacing is small compared to the height from ground, the field of the line is not appreciably affected by ground, and the line is treated as a two-conductor system, as has been done (Chap. III).

Returning to the system of Fig. 90*a*, we may write expressions for the charges on conductors 1 and 2 in terms of the voltages of the system, as we did for the *n*-conductor in Eqs. (532)

$$
\begin{aligned}
Q_1 &= C_{10}v_{10} + C_{12}v_{12} \\
Q_2 &= C_{20}v_{20} + C_{21}v_{21}
\end{aligned}
\tag{535}
$$

We can interpret these equations by the equivalent system of Fig. 90*b*, in which the electric field is considered localized within the condensers C_{10}, C_{20}, C_{12}, and the conductors themselves shrink into points.

This equivalent system has the form of a Π network and can be used to examine the performance of the original system when electrically connected to the outside.

We must use caution, however. When leads are brought out through the shield (conductor 0), the parameters C_{10}, C_{20}, C_{12} no longer describe the system accurately, unless the section of each lead contained inside the shield has been considered as part of the corresponding conductor in the original closed system. Furthermore, the aperture through which the leads are brought out should be so small as not to impair the effectiveness of the shield. If the three conductors are close together, as in Fig. 90*a*, the fields are essentially localized and the above conditions become unimportant.

As the next step, we express the currents through the leads at their entrance through the shield, in terms of the voltages. Of the latter, only two need be expressed, namely, v_{10} and v_{20}. The third may be written

$$v_{12} = v_{10} + v_{02} = v_{10} - v_{20}$$

Since v_{10} and v_{20} are the input and output voltages of the system, considered as a four-pole, the standard notation v_1 will be used in place of v_{10} from this point on, and v_2 for v_{20}.

Thus, referring to Fig. 90*d*

$$\begin{aligned} i_1 = \frac{dQ_1}{dt} &= \frac{d}{dt}[C_{10}v_1 + C_{12}(v_1 - v_2)] \\ &= \frac{d}{dt}[(C_{10} + C_{12})v_1 - C_{12}v_2] \\ i_2 &= \frac{d}{dt}[(C_{20} + C_{21})v_2 - C_{12}v_1] \end{aligned} \tag{536}$$

Going to the harmonic case, we have by means of the usual substitutions (**I** for i, **V** for v, and $j\omega$ for d/dt)

$$\begin{cases} \mathbf{I}_1 = j\omega[(C_{10} + C_{12})\mathbf{V}_1 - C_{12}\mathbf{V}_2] \\ \mathbf{I}_2 = j\omega[(C_{20} + C_{12})\mathbf{V}_2' - C_{12}\mathbf{V}_1] \end{cases} \tag{537}$$

Hence, the admittance coefficients of the four-pole

$$y_{11} = j\omega(C_{10} + C_{12}) \qquad y_{22} = j\omega(C_{20} + C_{12})$$
$$y_{12} = y_{21} = -j\omega C_{12} \tag{538}$$

bearing in mind, however, that the "y" coefficients are not mutually independent and that y_{12} is always negative. The sums $(C_{10} + C_{12})$ and $(C_{20} + C_{12})$ are dual to the *self-inductances*, and C_{12} is dual to the *mutual inductance*. The lossless transformer and the three-conductor system are themselves dual, except that the equivalent T network for the transformer is not always physically realizable, while the Π network equivalent to the three-conductor system always is.

The most important practical example of a three-conductor system is a vacuum tube, either a triode or a multielectrode tube in which all electrodes but two are effectively grounded to the shield (by-passed). In vacuum tubes the currents given by Eqs. (537) are not functional, in the sense that the tube would work better without them. However, they cannot be neglected except, in general, at audio frequencies. A convenient, if somewhat unorthodox, approach consists in regarding the high-frequency electronic currents of the vacuum tube simply as additions to the currents of Eqs. (537) and thus arriving at values for the admittance coefficients of the vacuum tube in linear operation, considered as a four-pole. This line of reasoning will be followed in Chap. XVI; the next section will be devoted to applications of the foregoing theory.

15.4. Applications and examples.

Electrostatic coupling. It has been pointed out that the three-conductor system of Fig. 90 and the lossless transformer are dual structures, as can be verified at once by comparing Eqs. (448) and (537). However, while *magnetic coupling* (by means of the transformer) is in frequent use, *electrostatic coupling* (by means of the three-conductor system, equivalent to a Π condenser network) is seldom resorted to. We can see why this is if we write down the expression for the input admittance of the electrostatic coupling network; this expression results immediately from that of the transformer, by dual substitution. We had for the transformer

$$Z_i = j\omega L_1 \frac{1 - \tau^2 - jz_r}{1 - jz_r} \tag{453}$$

where

$$\tau^2 = \frac{M^2}{L_1L_2}$$

$$z_r = \frac{Z_r}{\omega L_2} \qquad (Z_r = \text{load impedance})$$

Hence, for the electrostatic coupling network (Fig. 90d),

$$Y_i = j\omega(C_{10} + C_{12}) \frac{1 - \tau^2 - jy_r}{1 - jy_r} \tag{539}$$

where

$$\tau^2 = \frac{C_{12}^2}{(C_{10} + C_{12})(C_{20} + C_{12})} \tag{540}$$

$$y_r = \frac{Y_r}{\omega(C_{20} + C_{12})} \qquad (Y_r = \text{load admittance})$$

The transformer is useful chiefly because its input and load impedances (for close coupling) differ essentially only by a real multiplier (the square of the turn ratio t) which, within broad limits, can be given any required value.

Thus, if we make

$$|1 - \tau^2| \ll |z_r|$$
$$|z_r| \ll 1$$

(*i.e.*, for tight coupling and sufficiently low load impedance), the transformer input impedance reduces to

$$Z_i = j\omega L_1(-jz_r) = j\omega L_1\left(-j\frac{Z_r}{\omega L_2}\right) = Z_r\frac{L_1}{L_2} = \frac{Z_r}{t^2}$$

where

$$t = \sqrt{\frac{L_2}{L_1}}$$

Under similar conditions, the input admittance to the electrostatic coupling network would be

$$Y_i = \frac{Y_r}{t^2}$$

where

$$t = \sqrt{\frac{C_{20} + C_{12}}{C_{10} + C_{12}}}$$

but since we have assumed

$$\tau^2 = \frac{C_{12}^2}{(C_{10} + C_{12})(C_{20} + C_{12})} \approx 1$$

we must necessarily have

$$C_{10} \ll C_{12} \qquad C_{20} \ll C_{12}$$

Hence

$$t \approx 1$$

In other words, for close coupling, the shunting capacities of the Π network (Fig. 90*d*) must be negligibly small, in which case the multiplier $1/t^2$ is near unity and the coupling has practically no impedance transforming action.

Measurement of partial capacities (particularly interelectrode capacities). The partial capacities C_{10}, C_{20}, C_{12} are specially important in the case of triodes when the three conductors 0, 1, 2 (Fig. 89*c*) are the three electrodes (cathode, grid, and plate). Of these, the cathode is usually by-passed to an external shield (grounded), hence we shall consider it in place of conductor 0, although other connections are used occasionally (Sec. 16.3).

We shall, therefore, in this example adopt the notation commonly in use for triodes, but the conclusions will be of a general character.

Thus

$$\begin{aligned} C_{10} &= C_{gk} \\ C_{20} &= C_{pk} \\ C_{12} &= C_{gp} \end{aligned}$$

These three values are called *interelectrode capacities.* It is well to remember that they are not, however, capacities in the ordinary sense as applied to two-conductor systems (Sec. 11.8), but *partial* capacities or, which is the same thing, equivalent capacities which would reproduce the three-electrode system if connected in the Π configuration. None of the three interelectrode capacities can be measured directly by impedance measurements (on a bridge). The Π network is an equivalent representation of the system, but it cannot be taken apart for separate measurements of its branches; it must be considered as a box with three exposed terminals.

The interelectrode capacities may be obtained *indirectly* from the following three measurements:

1. Cathode and plate connected together; capacity measured on bridge between grid and remaining electrodes

$$C_g' = C_{gk} + C_{gp}$$

2. Cathode and grid connected together; capacity measured between plate and remaining electrodes

$$C_p = C_{pk} + C_{gp}$$

3. Plate and grid connected together; capacity measured between cathode and remaining electrodes

$$C_k = C_{gk} + C_{pk}$$

During all three measurements, the shield should remain connected to whichever electrode is connected, or by-passed, to the shield in actual operation.

From the expressions of C_g, C_p, C_k, constituting a system in the three unknowns C_{gk}, C_{pk}, C_{gp}, we readily obtain expressions for the latter values. They are

$$\left\{\begin{aligned} C_{gk} &= \frac{C_g - C_p + C_k}{2} \\ C_{pk} &= \frac{-C_g + C_p + C_k}{2} \\ C_{gp} &= \frac{C_g + C_p - C_k}{2} \end{aligned}\right. \tag{541}$$

Another method of obtaining the partial capacities is based on measurements of transfer admittance. The connections are as in the three measurements (1), (2), (3) above, but the procedure is as follows:

1. Cathode shorted to plate. A high-frequency source is connected between grid and cathode. The voltage V_{gk} across the source and the current I_{pk} in the shorting connection (r.m.s. values) are measured. This gives

$$C_{gp} = \frac{I_{pk}}{\omega V_{gk}}$$

2. Cathode shorted to grid; source connected between plate and grid

$$C_{pk} = \frac{I_{gk}}{\omega V_{gp}}$$

3. Plate shorted to grid; source connected between plate and cathode

$$C_{gk} = \frac{I_{gp}}{\omega V_{pk}}$$

The partial capacities must be measured when the tube is "cold" (filament not turned on), or the electronic current would introduce large errors. This is a serious handicap, as cold and hot capacities differ appreciably.

CHAPTER XVI

FOUR-POLE THEORY APPLIED TO THE VACUUM TUBE

16.1. The Vacuum Tube in Linear Operation. A triode differs in operation from the generic three-conductor system (Fig. 89*c*) in that a *space current*, carried by free electrons,[1] flows to the cathode from the plate, or from both grid and plate.

If we add to the space or electron-borne current leaving each electrode, the displacement current due to the changes of electric flux and the conduction current leaving the electrode through the lead which connects it to the outside, the sum must vanish (Sec. 12.1); hence, the current *entering* the electrode through the lead must equal the sum of the space and displacement currents *leaving* the electrode.

We will, with the help of simplifying assumptions, obtain simple expressions for these *lead currents* in terms of the voltages; these expressions will enable us to consider the vacuum tube as a four-terminal network of known coefficients and investigate its transmission characteristics, much as we have done for other networks. We must keep in mind, however, that many of the functions performed by tubes (modulation, detection, etc.) cannot be handled as simply as that of *linear amplification*, for which alone the preliminary assumptions can be considered valid.

Let us consider a vacuum tube in which only two electrodes (the control grid and the anode, or plate) are subject to voltage variations with respect to the cathode. Multielement tubes (pentodes particularly) are included, as the screen and suppressor grid are kept at constant voltages. The number of independent variables, therefore, reduces to two, the instantaneous grid and plate voltages v_g and v_p (Fig. 91*a*).

First assumption. The space currents from the grid and plate are continuous, single-valued functions of v_g and v_p. This

[1] The electrons, being negative charges, move in the opposite direction to the space current.

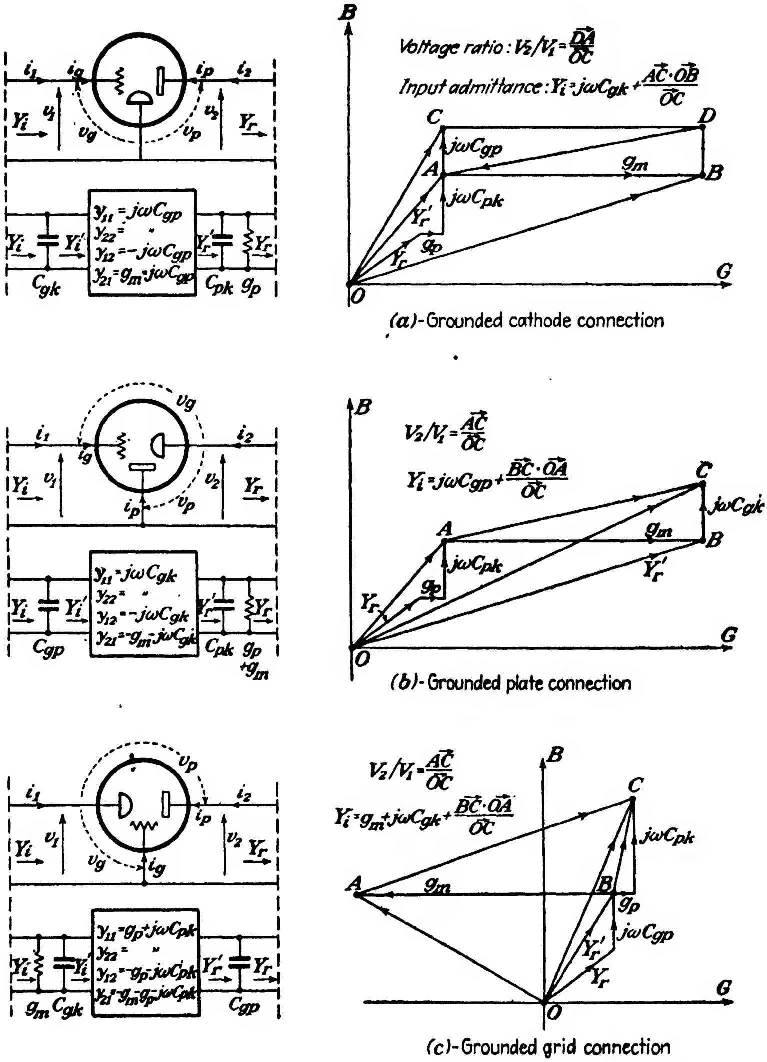


FIG. 91.—*Network equivalents of the vacuum tube in the three connections. Semigraphical solutions for voltage ratio and input admittance.*

assumption is seriously in error only at extremely high frequencies, when the space current leaving the plate and grid is not equal to that entering the cathode at the same instant, because the field changes appreciably during the electron transit time.[11] Barring this exception, we may write

$$\begin{cases} i_g = i_g(v_g, v_p) \\ i_p = i_p(v_g, v_p) \end{cases} \tag{542}$$

We could learn more about these functions through an analysis based on the field equations and the conservation of energy, assuming some simple geometry, continuous distribution of the electrons in the interelectrode space, and negligible energy of emission. In practice, the functions (542) are made available in the form of plots obtained by measurement (the static characteristics). The continuity of the functions enables us, furthermore, to expand them in series about an arbitrary value.

In operation, the voltages v_g and v_p fluctuate about their *quiescent* values. Thus, we may write

$$\begin{cases} v_g = \bar{V}_g + \Delta v_g \\ v_p = \bar{V}_p + \Delta v_p \end{cases} \tag{543}$$

$\bar{V}_g$ and $\bar{V}_p$ are the quiescent values, such as can be measured on the system at rest, with no signal impressed. These values are not necessarily equal to the *average* values nor to the battery voltages.

Δv_g and Δv_p are the *variational* or *signal* values of voltage.

The total values of current may be subjected to a similar breakdown, and the variational values of i_g and i_p may be expressed in terms of Δv_g and Δv_p by expanding the functions (542) about the quiescent values. Thus, we have for i_p

$$\begin{aligned} i_p = i_p(\bar{V}_g + \Delta v_g, \bar{V}_p + \Delta v_p) &= i_p(\bar{V}_g, \bar{V}_p) + \Delta v_g \frac{\partial i_p}{\partial v_g}(\bar{V}_g, \bar{V}_p) \\ &+ \Delta v_p \frac{\partial i_p}{\partial v_p}(\bar{V}_g, \bar{V}_p) + \frac{1}{2!}\left[(\Delta v_g)^2 \frac{\partial^2 i_p}{\partial v_g^2}(\bar{V}_g, \bar{V}_p)\right. \\ &\left. + 2\,\Delta v_g\,\Delta v_p \frac{\partial^2 i_p}{\partial v_g\,\partial v_p}(\bar{V}_g, \bar{V}_p) + (\Delta v_p)^2 \frac{\partial^2 i_p}{\partial v_p^2}(\bar{V}_g, \bar{V}_p)\right] \cdots \end{aligned} \tag{544}$$

and a similar expression for i_g.

Two additional assumptions may be introduced at this point.

Second assumption. All terms of (544) except the first two are *negligibly small* in actual operation. The significance of this point is made clear by considering the characteristic surface:

$$i_p = i_p(v_g, v_p)$$

A set of values of v_g, v_p, characterizes the operation of the tube and defines at the same time a point of this surface, which is called the *point of operation.* If excursions of the point of operation are limited to a region of the characteristic surface which can be assimilated to a plane, then the second assumption is valid.

Third assumption. All terms of the expansion for i_g, space current from the grid, are negligible at all times. This is true when v_g never takes positive values.

When assumptions (2) and (3) are respected, the tube is said to be in *class A* operation. The space current, which flows from the plate only, is then a *linear* function of the two voltages, as (544) reduces to

$$i_p = i_p(\bar{V}_g, \bar{V}_p) + \Delta v_g \frac{\partial i_p}{\partial v_g} (\bar{V}_g, \bar{V}_p) + \Delta U_p \frac{\partial i_p}{\partial v_p} (\bar{V}_g, \bar{V}_p)$$

Hence the *variational* value of current (the *signal* value) is given by

$$\Delta i_p = i_p - \bar{I}_p = g_m \Delta v_g + g_p \Delta v_p$$

having made the positions

$$g_m = \frac{\partial i_p}{\partial v_g} (\bar{V}_g, \bar{V}_p) = \text{mutual conductance} \tag{545}$$

$$g_p = \frac{\partial i_p}{\partial v_p} (\bar{V}_g, \bar{V}_p) = \text{plate conductance} \tag{546}$$

The mutual and plate conductances, together with the interelectrode capacities, characterize the tube in linear operation. The values of g_m and g_p depend on the quiescent values of voltage, since g_m and g_p are defined as the partial derivatives of current taken at the quiescent point. Because of the particular form which the function

$$i_p = i_p(v_g, v_p)$$

takes over a large range of values, for each value of $\bar{I}_p$ (quiescent current) there is only one value of $g_m(\bar{V}_g, \bar{V}_p)$ (mutual conductance at the quiescent point), and of $g_p(\bar{V}_g \bar{V}_p)$.

We may write, in fact[11]

$$i_p = k(v_p + \mu v_g)^n \tag{547}$$

where μ, k, n are constants depending on the tube geometry.[1]

[1] It may be shown that

$$\mu = \frac{C_{gk}}{C_{pk}}$$

for an indefinite plane triode.

Hence

$$g_p(\bar{V}_g, \bar{V}_p) = \frac{\partial i_p}{\partial v_p}(\bar{V}_g, \bar{V}_p) = kn\left(\frac{\bar{I}}{k}\right)^{\frac{n-1}{n}} \tag{548}$$

and

$$\mu = \frac{g_m}{g_p} \tag{549}$$

The reciprocal of g_p, called *plate resistance*, is very commonly used in place of g_p. In the following, however, g_p and g_m will be used exclusively, for the sake of uniformity and convenience. μ is called the *amplification factor*.

16.2. Admittance Coefficients of the Tube, Regarded as a Four-terminal Network. The variational or signal components of space current flowing out of the grid and plate in linear operation have been found to be

$$\begin{aligned} \Delta i_g &= 0 \\ \Delta i_p &= g_m\,\Delta v_g + g_p\,\Delta v_p \end{aligned}$$

subject to the three assumptions of the preceding section.

As far as the space currents are concerned, the tube is therefore a *linear system*. Hence, the principle of superposition (Sec. 1.7) holds for any system made up of the tube and passive linear elements. The performance of such a system under harmonic excitation of variable frequency is therefore significant, even though in actual operation the excitation may be not harmonic but periodic, with components occupying a definite frequency range. We shall, therefore, replace the variational values $\Delta i_g, \Delta i_p, \ldots$ by the corresponding complex values and write

$$\begin{aligned} \mathbf{I}_g{}^e &= 0 \\ \mathbf{I}_p{}^e &= g_m\mathbf{V}_g + g_p\mathbf{V}_p \end{aligned} \tag{550}$$

The superscript e signifies that the values in question are due exclusively to the flow of electrons (space currents). To these we must add the *displacement* currents, which we readily obtain from Eqs. (537) relating to a generic three-conductor system, taking the grid as conductor 1, the plate as 2, and the cathode as 0:

$$\begin{cases} \mathbf{I}_g{}^d = j\omega[(C_{gk} + C_{gp})\mathbf{V}_g - C_{gp}\mathbf{V}_p] \\ \mathbf{I}_p{}^d = j\omega[(C_{pk} + C_{gp})\mathbf{V}_p - C_{gp}\mathbf{V}_g] \end{cases} \tag{551}$$

The total currents into the electrode leads are obtained by adding the electron and displacement values. Thus

$$\begin{aligned} \mathbf{I}_g &= \mathbf{I}_g^e + \mathbf{I}_g^d = j\omega(C_{gk} + C_{gp})\mathbf{V}_g - j\omega C_{gp}\mathbf{V}_p \\ \mathbf{I}_p &= \mathbf{I}_p^e + \mathbf{I}_p^d = [g_m - j\omega C_{gp}]\mathbf{V}_g + [g_p + j\omega(C_{pk} + C_{gp})]\mathbf{V}_p \end{aligned} \qquad (552)$$

So far, no assumption has been made as to the manner in which the tube is connected to sources or receivers of energy. We will now consider three possibilities, as illustrated in Fig. 91. In each case one of the three electrodes (cathode, grid, and plate) is at ground potential; the input and output voltages are those between the grounded electrode and the remaining two. It is always understood that d-c values, as well as other harmonic values, actually exist in addition to the harmonic values of the frequency considered; and that grounding connections consist of by-pass condensers of relatively large values joining the grounded electrode to the tube shield.

In high-frequency operation it may be necessary to distinguish between the interelectrode voltages which appear in Eqs. (552) and the voltages between the electrode leads at the points at which these are brought out through the tube shielding. The difference is the time derivative of the magnetic flux linking the loop delimited by the two leads (Fig. 90*c*). This effect may be minimized by correct design and need not be taken into account, as a rule, unless the frequency is so high that circuit methods of analysis have to be abandoned altogether.

The following table correlates input and output voltages and currents with electrode voltages (referred to cathode) and electrode currents. From these relations and Eqs. (552), the *admittance coefficients* of the three four-poles, consisting of the tube in the three different connections, may be obtained. Their values are listed in the table. Instead of using these values directly, it is more convenient, in practice, to take out from the tube those admittances which appear effectively in shunt with the input or output terminals. The tube is then considered equivalent to a residual four-pole shunted at both ends by these shunting admittances. These equivalent representations appear in Fig. 91, and the admittance coefficients of the residual four-pole are given there and on the table. The shunting admittances are, in practice, lumped with the load and source admittances, following the procedure outlined in the next section.

Admittance Coefficients of the Vacuum Tube in Linear Operation, for Various Connections

Connection		Cathode to ground (Fig. 90a)	Plate to ground (Fig. 90b)	Grid to ground (Fig. 90c)
Input voltage....		$\mathbf{V}_1 = \mathbf{V}_g$	$\mathbf{V}_1 = \mathbf{V}_g - \mathbf{V}_p$	$\mathbf{V}_1 = -\mathbf{V}_g$
Output voltage...		$\mathbf{V}_2 = \mathbf{V}_p$	$\mathbf{V}_2 - \mathbf{V}_p$	$\mathbf{V}_2 = \mathbf{V}_p - \mathbf{V}_g$
Input current....		$\mathbf{I}_1 = \mathbf{I}_g$	$\mathbf{I}_1 = \mathbf{I}_g$	$\mathbf{I}_1 = -(\mathbf{I}_p + \mathbf{I}_g)$
Output current...		$\mathbf{I}_2 = \mathbf{I}_p$	$\mathbf{I}_2 = -(\mathbf{I}_p + \mathbf{I}_g)$	$\mathbf{I}_2 = \mathbf{I}_p$
y coefficients of tube four-pole*	y_{11}	$j\omega(C_{gk} + C_{gp})$	$j\omega(C_{gk} + C_{gp})$	$g_m + g_p + j\omega(C_{gk} + C_{pk})$
	y_{22}	$g_p + j\omega(C_{pk} + C_{gp})$	$j\omega(C_{gk} + C_{pk}) + g_m + g_p$	$g_p + j\omega(C_{pk} + C_{gp})$
	y_{12}	$-j\omega C_{gp}$	$-j\omega C_{gk}$	$-g_p - j\omega C_{pk}$
	y_{21}	$g_m - j\omega C_{gp}$	$-g_m - j\omega C_{gk}$	$-g_m - g_p - j\omega C_{pk}$
Admittance taken out on input side.		$j\omega C_{gk}$	$j\omega C_{gp}$	$g_m + j\omega C_{gk}$
Admittance taken out on output side...........		$g_p + j\omega C_{pk}$	$g_p + g_m + j\omega C_{pk}$	$j\omega C_{gp}$
y coefficients of residual four-pole	y_{11}	$j\omega C_{gp}$	$j\omega C_{gk}$	$g_p + j\omega C_{pk}$
	y_{22}	$j\omega C_{gp}$	$j\omega C_{gk}$	$g_p + j\omega C_{pk}$
	y_{12}	$-j\omega C_{gp}$	$-j\omega C_{gk}$	$-g_p - j\omega C_{pk}$
	y_{21}	$g_m - j\omega C_{gp}$	$-g_m - j\omega C_{gk}$	$-g_m - g_p - j\omega C_{pk}$

* Obtained by writing Eqs. (552) in terms of input and output voltages and current.

16.3. Voltage Ratio and Impedance Transforming Action Due to the Three Vacuum-tube Connections. Having obtained the admittance coefficients for the residual four-pole equivalent to the vacuum tube in the three connections of Fig. 91 (less the admittances shunting input and output, which are more conveniently looked upon as part of the load and source systems), the *voltage ratio* and *impedance transforming action* of these four-poles may be worked out.

We may regard these quantities as descriptive of the linear performance of the tube in the three connections, provided the signal energy is fed into the input side at all times. One more complex value, the output impedance, would be required to establish the behavior of the tube for transmission in both directions. We know in fact that six real numbers, or three complex numbers, are needed to describe fully the action of a four-terminal network (Sec. 2.1).

Transmission in both directions need not be contemplated, however. A glance at the values of the coefficients shows that the principle of reciprocity ($y_{12} = y_{21}$, Sec. 2.3) *does not hold* in the case of vacuum tubes. The transfer admittance from output to input, y_{12}, is due entirely to the interelectrode capacity, except for the grounded grid connection (Fig. 91*c*). Tubes cannot, therefore, be used effectively for transmission in two directions.

The input admittance and voltage ratio expressions are tabulated below for the three residual four-poles of Fig. 91. Semigraphical constructions for these values are given in Fig. 91. An example, illustrating the breakdown of an amplifier circuit, will follow (Sec. 16.5). It should be borne in mind that amplifier analysis simplifies very considerably at low frequencies or when the interelectrode capacity is very small.

INPUT ADMITTANCE AND VOLTAGE RATIO OF VACUUM TUBES IN LINEAR OPERATION, FOR VARIOUS CONNECTIONS

Connection	Cathode to ground (Fig. 90*a*)	Plate to ground* (Fig. 90*b*)	Grid to ground (Fig. 90*c*)
Input admittance (residual four-pole). $Y_i' = \dfrac{D_y + y_{11}Y_r'}{y_{22} + Y_r'}$	$j\omega C_{gp} \dfrac{Y_r' + g_m}{Y_r' + j\omega C_{gp}}$ (553)	$j\omega C_{gk} \dfrac{Y_r' - g_m}{Y_r' + j\omega C_{gk}}$ (554)	$\dfrac{(g_p + j\omega C_{pk})(Y_r' - g_m)}{Y_r' + g_p + j\omega C_{pk}}$ (555)
Voltage ratio (tube or residual four-pole) $\dfrac{\mathbf{V}_2}{\mathbf{V}_1} = -\dfrac{y_{21}}{y_{22} + Y_r'}$	$\dfrac{j\omega C_{gp} - g_m}{Y_r' + j\omega C_{gp}}$ (556)	$\dfrac{j\omega C_{gk} + g_m}{Y_r' + j\omega C_{gk}}$ (557)	$\dfrac{j\omega C_{pk} + g_p + g_m}{Y_r' + g_p + j\omega C_{pk}}$ (558)
where $Y_r' =$	$Y_r + g_p + j\omega C_{pk}$	$Y_r + g_p + g_m + j\omega C_{pk}$	$Y_r + j\omega C_{gp}$

* This is often called the *cathode follower* connection. The corresponding expression reduces to a very simple form when the load admittance has a particular value (Sec. 16.6).

In the above

$D_y = y_{11}y_{22} - y_{12}y_{21}$ = determinant of residual four-pole coefficients
Y_r' = load admittance, plus admittance taken out of tube on output side
Y_i' = input admittance of residual four-pole. To obtain input admittance of tube, add admittance taken out on input side

Note. In practice, it is often unnecessary to compute the actual load and imput admittances (see Fig. 92).

16.4. Amplifier Classification. It is common practice to classify amplifiers according to the following subdivisions:

a. By frequency:

- d-c
- Audio
- Intermediate
- Radio
- Video

b. By loading:

- Voltage amplifier (unloaded)
- Power amplifier (loaded)

c. By class:

- *A*
- *B*
- *C*

d. By method of coupling:

- Direct coupled
- Resistance coupled
- Inductance coupled
- Transformer coupled

e. Depending on the use of ground as a signal current return (Sec. 1.2):

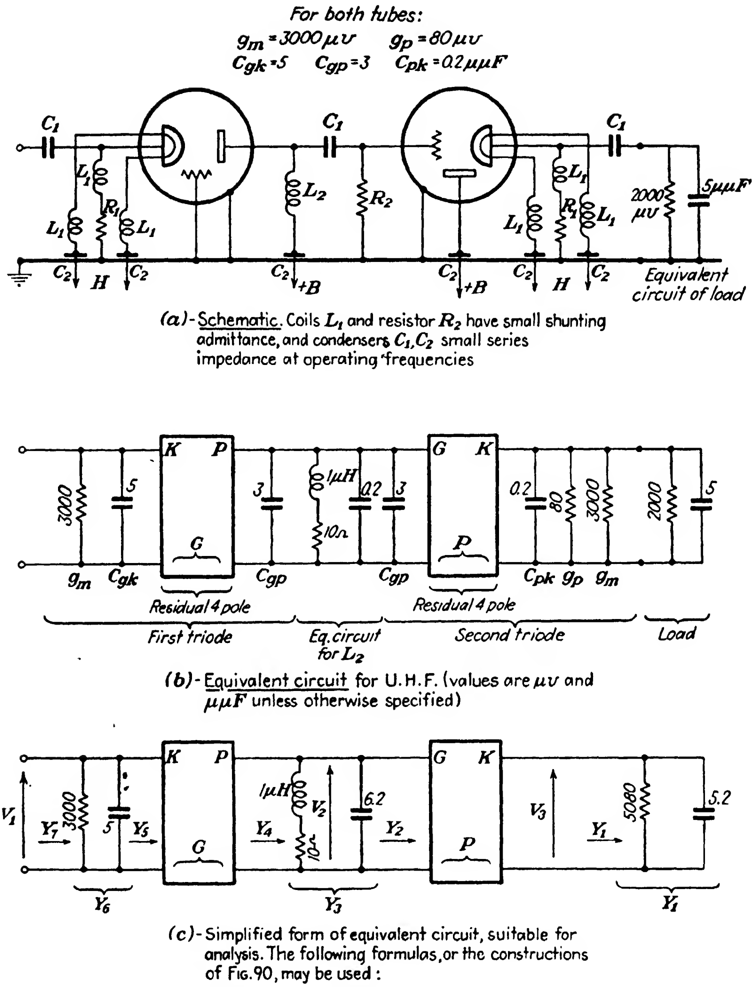
- Balanced
- Unbalanced

f. By presence or absence of regeneration (Sec. 17.1):

- Regenerative
- Neutral
- Degenerative (inverse feedback amplifiers)

These distinctions are useful when discussing amplifiers in general. We must assume here that the reader is familiar with the use and attributes of each type, at least in a qualitative sense. If this is not the case, the literature on the subject(11) should be consulted before proceeding.

The *quantitative* problem of investigating the performance of an amplifier of given design has limited but definite usefulness. Solving this, or any other problem of analysis, does not lead directly to design values. Such a solution does, however, provide a check on tentative values, and if it is not too laborious and time consuming, it can lead to the correct values by trial



$$Y_2 = j\omega C_{gk}\frac{Y_1 - g_m}{Y_1 + j\omega C_{gk}}; \quad Y_4 = Y_2 + Y_3; \quad Y_5 = \frac{(g_p + j\omega C_{pk})(Y_4 - g_m)}{Y_4 + g_p + j\omega C_{pk}}$$

$$\frac{V_3}{V_2} = \frac{j\omega C_{gk} + g_m}{Y_1 + j\omega C_{gk}} \qquad \frac{V_2}{V_1} = \frac{j\omega C_{pk} + g_p + g_m}{Y_4 + g_p + j\omega C_{pk}}$$

FIG. 92.—*Reducing an amplifier schematic to equivalent form.*

and error. Hence the importance of using the simplest possible methods of analysis.

These methods must vary, of course, depending on the type of amplifier. Many of the usual distinctions are not sufficiently precise for the purpose of analysis. The significant distinctions are the following:

a. Whether the amplifier is linear or nonlinear in operation. As a rule, circuit and network analysis can only be applied to the linear case. An exception, relating to tuned amplifiers, will be considered briefly later (Sec. 16.6).

b. Whether the interstage capacity effects can be neglected. At audio frequency, and at higher frequencies if screen grid tubes are used, the susceptance due to this capacity is so small that its effects on performance are negligible. In this event, the analytical problem simplifies very materially, as the performance of each stage reduces to a calculation of voltage ratio and can be worked out independently (Sec. 16.5).

c. Whether or not the amplifier is tuned to a particular frequency in normal operation. A tuned amplifier operates under particular conditions. This simplifies the analysis, as will be seen in Sec. 16.6.

d. Whether the stages are coupled by four-poles or by shunting branches.

e. Whether or not the amplifier stage is the only transmission channel between source and load. In amplifiers with external feedback there is an added transmission channel that may be regarded as a four-pole in parallel with the amplifier or part of it.

16.5. Equivalent Circuit of the Amplifier. The amplifier of Fig. 92 would be classified, according to the preceding criteria, as follows:

a. Linear operation
b. Interphase capacity not negligible
c. Untuned
d. Shunt coupled
e. No other transmission channel

Any other amplifier similarly classified would be subject to a similar method of analysis, although its characteristics and design might be entirely different. A rather unorthodox design has been deliberately selected. The method of analysis is not affected by the type of tube connection (see Fig. 91 for the three alternatives).

The first step in any amplifier analysis is the drawing of an *equivalent circuit* in which the physical circuit elements are replaced by parameters. It is important to recognize the difference between a *schematic diagram* and an *equivalent diagram.* The schematic, or diagram of connections, shows the parts and their metallic connections. It should indicate the commercial ratings of the parts and, especially in high-frequency systems, the exact manner in which they are joined together. Leads can be straightened and their length altered, but all junctions of the schematic should be physical junctions of the system. Schematics are principally used as a guide in wiring, and also, in conjunction with mechanical layout drawing, may be used as the basis of equivalent circuits on which the performance analysis may be carried out.

An equivalent circuit is an entirely different matter. It shows not circuit parts, but parameters, and may differ depending on the frequency for which it is drawn. On the equivalent circuit, a coil rated at 50 mh may appear as a 10μf condenser. By-pass condensers are left out of the equivalent circuit whenever their effect may be considered negligible.

Drawing the equivalent circuit is perhaps the most difficult part of circuit analysis. The rest requires careful numerical work and an orderly procedure. Semigraphical methods (Fig. 91) permit considerable saving of time.

16.5. The Tube as a Generator. When the interstage capacitance does not have appreciable effect, either because of its own low value or because of the low value of the operating frequency, the *mutual conductance* of the tube constitutes the only link between the input and output sides of the tube (except for the grounded grid connection, Fig. 91*c*, in which the plate conductance g_p provides an additional link).

It is more convenient in such cases to represent it as a *generator* (Sec. 2.6) than as a four-pole. The Thévenin and Norton equivalences (Sec. 2.6) can both be used, but the second is generally more convenient. Consider, for example, the tube in the cathode to ground connection, which is by far the most common. If we neglect ωC_{gp}, the voltage ratio (plate voltage/grid voltage) becomes, from Eq. (556),

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = -\frac{g_m}{Y_r'} \tag{559}$$

[*Voltage ratio for the tube with grounded cathode, interstage capacity negligible*]

where $Y_r' = Y_r + g_p + j\omega C_{pk}$

Regarding Y_r' as the load, we can write the load current as follows:

$$I_r' = V_2 Y_r' = -V_1 g_m$$

which means that a driving current (or equivalent generator current) of value $-V_1 g_m$ is impressed by the tube upon the load Y_r. Hence the Norton equivalence of Fig. 93*a*. Similar equivalences for the grounded plate and grounded grid connections are shown in Figs. 93*b* and 93*c*.

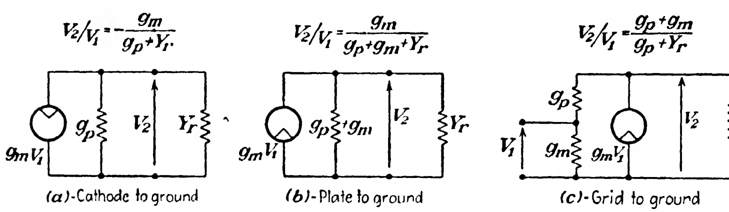


FIG. 93.—*Equivalent circuits for the vacuum tube at low frequency.*

The exclusive use of the grounded cathode connection in multistage amplifiers is readily explained if the interstage capacity effects are neglected. This connection is the only one for which the output voltage is higher than the input *when a chain of similar stages is used.* In both the grounded plate and grounded grid stages, the voltage ratio is less than unity when the load is the input admittance of a similar stage.

For the grounded plate connection, this is evident from the expression of V_2/V_1, which in no case can have value greater than 1 (557). The same is true of the voltage ratio of the grounded grid stage, if we replace load admittance Y_r by the input admittance of a similar stage, which includes g_m.

Voltage amplification may be obtained without recourse to the grounded cathode connection by alternating grounded grid and grounded cathode stages. Such a chain may be thought of as made up of similar stages, each including two tubes connected as in Fig. 91. Large voltage gains are possible by this device, which has the advantage that it works well between terminations of relatively low impedance, comparable to $1/g_m$, which is generally about 500 ohms. Such an impedance, connected across the output of a grounded cathode stage, would reduce its voltage gain to small or negative values. Another advantage of the grounded grid connection is that it has low interstage capacity when triodes are used; the grid, which is at ground potential, prevents capacitive coupling from the cathode to the plate. This function must be discharged by an additional electrode (the screen) when the grounded cathode connection is used, if a low interstage capacity is desired.

16.6. Applications and examples.

Tuned amplifiers. Consider the amplifier stage shown in Fig. 94. Let us obtain values for the input admittance and voltage ratio, assuming that

a. The *tuning condenser* C_t is set for the highest possible value of voltage gain;[1] and

b. The interstage capacity is so low that

$$\omega C_{gp} \ll g_m$$

The second condition is made possible even at uhf by the use of screen grid tubes, particularly pentodes, which have almost universally replaced triodes in voltage amplification.

Because of condition (2), the voltage ratio may be written (557)

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = -\frac{g_m}{Y_r' + j\omega C_{gp}} = -\frac{g_m}{G_r + g_p + j[B_r + \omega(C_{pk} + C_{gp})]} \tag{560}$$

We will regard the load conductance G_r as constant with frequency. This is legitimate by comparison with B_r. Then the highest voltage gain will occur when

$$B_r + \omega(C_{pk} + C_{gp}) = 0$$

and will be

$$\frac{V_2}{V_{1(\max)}} = \frac{g_m}{g_p + G_r} \tag{561}$$

[*Maximum gain of tuned amplifier*]

Now, the input admittance (554) reduces to

$$\begin{aligned} Y_i' &= j\omega C_{gp}\frac{g_p + G_r - j\omega C_{gp} + g_m}{g_p + G_r} \\ &= \frac{\omega^2 C_{gp}^2}{g_p + G_r} + j\omega C_{gp}\left(1 + \frac{g_m}{g_p + G_r}\right) \end{aligned} \tag{562}$$

[*Input admittance of tuned amplifier*]

[1] The voltage gain, expressed in db, is given by

$$20 \log \frac{V_2}{V_1}$$

The term *gain* is very often used for *voltage gain.*

The above value does not include the susceptance ωC_{gk}, which is best considered lumped with the preceding stage.

It will be observed that the real part of (562) increases with the square of ωC_{gp}. This is the chief reason why it is important to reduce C_{gp}.

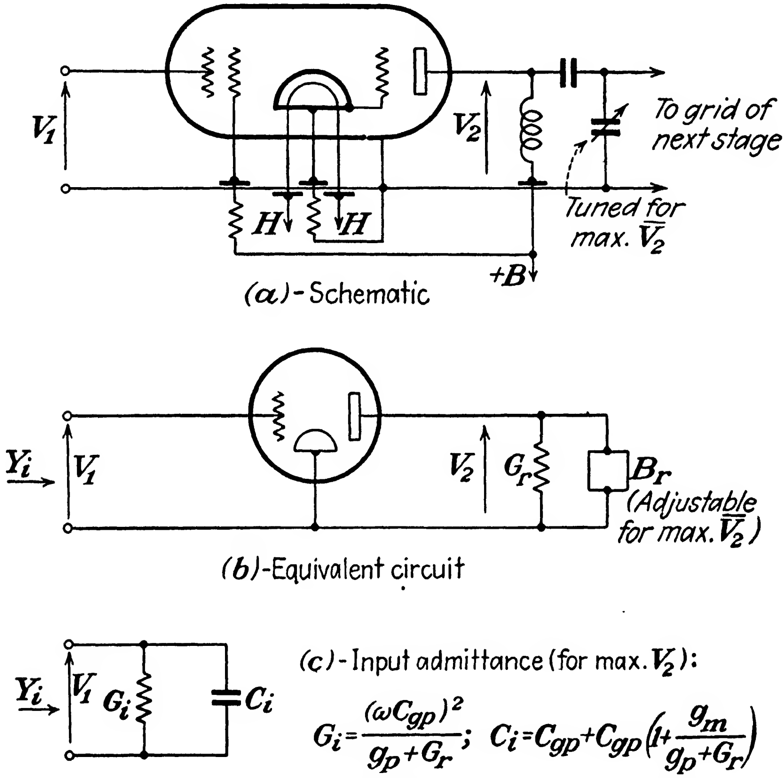


FIG. 94.—*Gain and input admittance of the tuned amplifier for* $g_m \gg wC_{gp}$.

Constant gain operation of the cathode follower stage. Under a particular set of conditions, the performance of the grounded plate, or cathode follower stage, becomes particularly simple. The voltage ratio of such a stage is given, in general, by (558)

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{g_m + j\omega C_{gk}}{Y_r' + j\omega C_{gk}}$$

where Y_r' is the load admittance increased by the admittance taken out of the tube on the output side (Fig. 94); or

$$Y_r' = Y_r + g_m + g_p + j\omega C_{pk} \tag{563}$$

Let us suppose that the load admittance Y_r consists of a resistive branch shunted by a variable condenser, or (Fig. 94),

$$Y_r = G_r + j\omega C_r$$

Then we can write the voltage ratio as follows:

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{g_m + j\omega C_{gk}}{g_m + g_p + G_r + j\omega(C_{pk} + C_r + C_{gk})}$$

This ratio becomes a real number, independent of frequency, when

$$\frac{C_{gk} + C_{pk} + C_r}{g_m + g_p + G_r} = \frac{C_{gk}}{g_m}$$

for which we must have

$$C_r = C_{gk}\frac{g_p + G_r}{g_m} C_{pk} \tag{564}$$

When C_r has the above value, the voltage ratio becomes

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{g_m}{g_m + g_p + G_r} \tag{565}$$

The input admittance of the stage can be written (554)

$$Y_i' = j\omega C_{gk}\frac{Y_r' - g_m}{Y_r' + j\omega C_{gk}}$$

Including the admittances previously taken out of the tube, and expressing Y_r, this becomes

$$Y_i = j\omega\left[C_{gp} + C_{gk}\frac{G_r + g_p + j\omega(C_{pk} + C_r)}{G_r + g_p + g_m + j\omega(C_{pk} + C_{pk} + C_r)}\right]$$

Noting that (564)

$$C_r + C_{pk} = C_{gk}\frac{g_p + G_r}{g_m}$$

the above can be written

$$Y_i = j\omega \left[C_{gp} + C_{gk} \frac{G_r + g_p + j\omega \dfrac{C_{gk}}{g_m}(G_r + g_p)}{G_r + g_p + g_m + j\omega \left[C_{gk} + \dfrac{C_{gk}}{g_m}(G_r + g_m) \right]} \right]$$

$$= j\omega \left[C_{gp} + C_{gk} \frac{G_r + g_p}{G_r + g_p + g_m} \frac{1 + j\omega \dfrac{C_{gk}}{g_m}}{1 + j\omega \dfrac{C_{gk}}{g_m}} \right]$$

or simply

$$Y_i = j\omega \left[C_{gp} + C_{gk} \frac{G_r + g_p}{G_r + g_p + g_m} \right]$$

This shows that if the capacity C_r has the value given by (564), the voltage ratio of the stage is a constant real number, and its input admittance is that of a condenser of value

$$C_i = C_{gp} + C_{gk} \frac{G_r + g_p}{G_r + g_p + g_m} \tag{566}$$

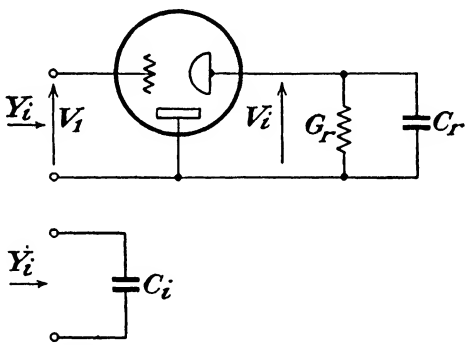


FIG. 95.—*Cathode follower stage.*

This is illustrated in Fig. 95. Under these conditions the stage receives no driving power, since its admittance is purely reactive. Yet it delivers power. Hence, the transmission gain (Sec. 17.1) through the tube is infinite at all frequencies, although the voltage gain is negative (voltage ratio less than unity). Such a stage can be connected across a high Q tank circuit without loading it at any frequency.[1] A large voltage will therefore be permitted to develop across the tank circuit, and a constant large fraction of this will appear across the load conductance G_r.

[1] Except for transit time effects.

CHAPTER XVII

FLOW OF POWER THROUGH HIGH-FREQUENCY AMPLIFIERS

17.1. Transmission of Power through Linear Amplifiers. It has been shown in the preceding chapter that amplifiers in linear operation may be regarded as chains of four-terminal networks and handled analytically by network methods.

In the general discussion on transmission networks (Chaps. II, V, and VI), *power relations* occupied a prominent place. Passive networks (as distinct from those capable of generating power) are used for the purpose of securing maximum power transfer, either at a single frequency or over a range (Sec. 8.1). When maximum power transfer is not enough, an amplifier has to be used. Under these conditions the amplifier has the job of supplying the receiver with more power than the source is capable of delivering, and this increase in available power is a significant quantity. Hence the importance, at least in principle, of power relations in connection with amplifiers.

In a number of cases, however, voltage rather than power is stressed. The ultimate receiver of the amplifier chain may absorb no energy, requiring only an electric field (hence, a voltage) for its operation. A notable example is the cathode-ray oscilloscope; each pair of deflection plates presents a purely capacitive load. In this case we should compare the *amplifier output voltage* to the voltage which the source would impress *directly across* the deflection plates. The latter is not necessarily equal to the amplifier input voltage, because of the difference between the amplifier input admittance and the admittance of the deflection plates. However, as a rule, the source has such low impedance, relatively, that the voltage across it, when connected directly to either the plates or the amplifier input, is the open-circuit voltage. Hence, we may take in this case as a criterion of amplifier performance simply the output to input voltage ratio of the amplifier, or its real value expressed in db (the *voltage gain*).

The same reasoning applies to each stage of an amplifier chain, whatever the nature of the ultimate receiver, when the operating frequency or the tube characteristics are such that the *power input* to each stage is negligible.

We can generalize and say that the increase in load power due to the insertion of an amplifier depends only on the *load* and *source admittances* and on the *over-all voltage gain* (hence, on the stage gains), whenever the *input admittance* of the amplifier is small compared to the *source admittance.*

To show this, we may express the *insertion gain* of the amplifier, which is the negative of the *insertion loss* (Sec. 6.4). Expressed in db, this is

$$\Gamma_I \text{ (db)} = 10 \log \frac{P_r}{P_0} \tag{567}$$

where P_r is the power actually received by the load through the amplifier and P_0 the value that would flow into the load if this were connected across the source directly. If Y_s and Y_r are the source and load admittances, we may write

$$P_0 = \frac{U^2}{|Y_s + Y_r|^2} G_r$$

and

$$P_r = V_r^2 G_r = V_s^2 \left(\frac{V_r}{V_s}\right)^2 G_r$$

V_s and V_r are the voltages at source and receiver, their ratio is the over-all voltage ratio. V_s may be expressed as follows:

$$V_s = \frac{U}{|Y_s + Y_i|}$$

where Y_i is the input admittance to the amplifier. We can now write the insertion gain

$$\Gamma_I = 10 \log \frac{\dfrac{U^2}{|Y_s + Y_i|^2} \left(\dfrac{V_r}{V_s}\right)^2 G_r}{\dfrac{U^2}{|Y_s + Y_r|^2} G_r} = 20 \log \frac{V_r}{V_s} + 20 \log \left|\frac{Y_s + Y_r}{Y_s + Y_i}\right|$$

or, letting Γ_V stand for the *over-all voltage gain,*

$$\Gamma_I \text{ (db)} = \Gamma_V \text{ (db)} + 20 \log \left|\frac{Y_s + Y_r}{Y_s + Y_i}\right| \tag{568}$$

[*Insertion gain of the amplifier*]

If we assume

$$|Y_i| \ll |Y_s|$$

the insertion gain becomes

$$\Gamma_I \text{ (db)} = \Gamma_V \text{ (db)} + 20 \log \left| 1 + \frac{Y_r}{Y_s} \right| \tag{569}$$

The above depends only on the over-all voltage gain and on the ratio of source and load admittances; the latter does not have very much effect, as a rule, since Γ_V is generally large. The insertion gain is affected favorably, all other things being equal, by a decrease in source admittance (higher source impedance).

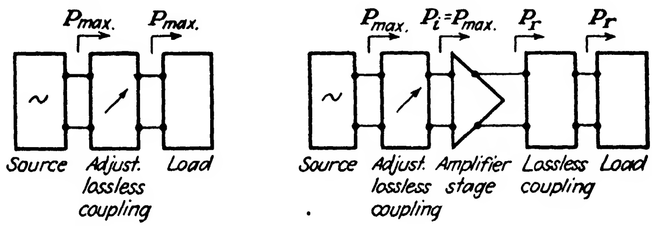


FIG. 96.—*Power increase due to amplifier insertion.*

We may conclude that the voltage gain is the only important factor in amplifier performance as long as the input admittance remains small. At high values of frequency, as we know, this is no longer a safe assumption. Voltage ratio and input admittance both play an important part in high-frequency operation, and the general expression for the insertion gain becomes very complicated. However, in the case of high-frequency amplifiers which are tuned to a particular frequency, the problem may be attacked from a different angle.

Consider a dissipative four-terminal network coupled to a source and load by lossless coupling networks (Fig. 96). Suppose that the coupling on the source side is adjusted for maximum power transfer (Sec. 8.2). We can easily show that the logarithmic ratio of the power P_r flowing to the load under these conditions to the maximum power P_{max} that the source is capable of supplying is the *transmission loss* (Sec. 6.3) of the dissipative network.

In fact, the dissipative network under the conditions given (source to network coupled for m.p.t.) receives power P_{max}.

It transmits power P_r. Hence (115)

$$L_T\text{ (db)} = 10\log\frac{P_i}{P_r} = 10\log\frac{P_{\max}}{P_r}\text{ db}\qquad \begin{matrix}(P_i = \text{power input})\\(P_r = \text{power output})\end{matrix}$$

Likewise, if a vacuum tube (considered as a four-terminal network) is inserted between couplings, the available power at the load will increase by

$$\Gamma_T\text{ (db)} = 10\log\frac{P_r}{P_i}$$

assuming that the maximum power transfer condition is satisfied at the source, both with and without the tube. In other words, if a transmission system includes adjustments for securing m.p.t. at the input end of a given amplifier network, we must consider the *transmission gain* rather than the insertion gain for a valid evaluation of the amplifier's effectiveness.

The transmission gain is a function of the amplifier characteristics and of the load admittance. There is no difficulty in obtaining very large values of the ratio P_r/P_i; in fact, this ratio may become negative, showing that the amplifier puts out power at both ends. When this happens, the amplifier is *regenerative* and the transmission gain ceases to have significance. In designing amplifiers for high-frequency use, regeneration is avoided, as it leads to instability.[1] Hence, the practical problem consists in designing the coupling network on the load side in such a manner that the ratio P_r/P_i is positive at all frequencies and becomes large at the operating frequency. Methods for solving this problem will be considered in the following.

17.2. The Transmission Gain as a Function of Load Admittance. Let us obtain the transmission gain for a tube in the grounded cathode connection. By definition, this is equal to

$$\Gamma_T\text{ (db)} = 10\log\frac{P_r}{P_i} = 10\log\left(\frac{V_2}{V_1}\right)^2\frac{G_r}{G_i} \tag{570}$$

where the symbols have reference to Fig. 96.

We have for the voltage ratio $\mathbf{V}_2/\mathbf{V}_1$ the expression (557)

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{j\omega C_{gp} - g_m}{G_r + g_p + j\omega(C_{gp} + C_{pk}) + jB_r}$$

having replaced Y_r' by its value (Fig. 90*a*).

[1] Regeneration should not be confused with oscillation (see Sec. 17.2).

Hence

$$\left(\frac{V_2}{V_1}\right)^2 = \frac{g_m^2 + \omega^2 C_{gp}^2}{(G_r + g_p)^2 + [B_r + \omega(C_{gp} + C_{pk})]^2} \tag{571}$$

The input admittance (553) is

$$Y_i = Y_i' + j\omega C_{gk} = j\omega C_{gp}\frac{G_r + g_p + g_m + j[B_r + \omega C_{pk}]}{G_r + g_p + j[B_r + \omega(C_{gp} + C_{pk})]} + j\omega C_{gk}$$

Hence, its real part G_i may be written

$$G_i = \omega C_{gp} \frac{(G_r + g_p + g_m)[B_r + \omega(C_{gp} + C_{pk})] - (G_r + g_p)(B_r + \omega C_{pk})}{(G_r + g_p)^2 + [B_r + \omega(C_{gp} + C_{pk})]^2}$$

$$= \omega C_{gp} \frac{g_m[B_r + \omega(C_{gp} + C_{pk})] + (G_r + g_p)\omega C_{gp}}{(G_r + g_p)^2 + [B_r + \omega(C_{gp} + C_{pk})]^2}$$

and the power ratio P_r/P_i goes into the form

$$\frac{P_r}{P_i} = \left(\frac{V_2}{V_1}\right)^2 \frac{G_r}{G_i} = \frac{g_m^2 + \omega^2 C_{gp}^2}{g_m[B_r + \omega(C_{gp} + C_{pk})] + (G_r + g_p)\omega C_{gp}} \frac{G_r}{\omega C_{gp}}$$

Hence, arranging terms,

$$\Gamma_T \text{ (db)} = 10 \log \frac{1 + (g_m/\omega C_{gp})^2}{1 + \dfrac{g_p}{G_r} + \dfrac{g_m}{G_r}\left[1 + \dfrac{C_{pk}}{C_{gp}} + \dfrac{B_r}{\omega C_{gp}}\right]} \tag{572}$$

This ratio, hence the transmission gain, depends on G_r and B_r, components of the load admittance. It should therefore be possible to associate graphically a value of *transmission gain* to each point of the complex plane of coordinates G_r, B_r (the *load admittance plane*). This correspondence will appear to be remarkably simple, in contrast to the complexity of the original expressions. Its usefulness is evident; given the load admittance, we can promptly find by this method *what increase of load power* can be secured by the use of the tube, provided an impedance matching network is inserted between the tube input and the source.

What we require is a map of the transmission gain on the Y plane. This map, in general, will consist of lines of constant Γ (constant transmission gain). Most significant, and easiest to obtain, are the lines

$$\Gamma_T = 0 \qquad \Gamma_T = \infty$$

The transmission gain is infinite when the denominator of (572) is zero. For this we must have

$$-G_r = g_p + g_m\left[1 + \frac{C_{pk}}{C_{gp}} + \frac{B_r}{\omega C_{gp}}\right] \tag{573}$$

which is the equation of a straight line on the G_r, B_r plane. The intercept of this line on the G_r axis is

$$\begin{aligned} G_r(B_r = 0) &= -g_p - g_m\left(1 + \frac{C_{pk}}{C_{gp}}\right) \\ &= -g_m\left(1 + \frac{g_p}{g_m} + \frac{C_{pk}}{C_{gp}}\right) \end{aligned} \tag{574}$$

and the intercept on the B_r axis

$$B_r(G_r = 0) = -\omega C_{gp}\left(1 + \frac{g_p}{g_m} + \frac{C_{pk}}{C_{gp}}\right) \tag{575}$$

Hence the slope

$$-\frac{B_r(G_r = 0)}{G_r(B_r = 0)} = \frac{\omega C_{gp}}{g_m} \tag{576}$$

This line (line $\Gamma_T = \infty$) is shown in Fig. 97.

To locate the line $\Gamma_T = 0$, we observe that, for zero transmission gain, we must have

$$P_r = P_i$$

or, equating the numerator and denominator of (572),

$$G_r\left(\frac{g_m}{\omega C_{gp}}\right)^2 = g_p + g_m\left(1 + \frac{C_{pk}}{C_{gp}} + \frac{B_r}{\omega C_{gp}}\right) \tag{577}$$

By working out the intercepts in the usual way, we find that the straight line represented by (577) is at right angles with the $\Gamma_T = \infty$ line and crosses it at its B-axis intercept (Fig. 97).

The two lines are easily drawn, once the frequency and tube characteristics are known. A practical construction for this purpose is shown in Fig. 97. On this drawing, the shaded region represents the portion of the Y plane for which the transmission gain has positive real values. If Y_r falls within this region, the tube operates as an amplifier, its power output being in excess of the input.

Below the $\Gamma_T = \infty$ line, the power ratio is negative; the source as well as the load receives power from the amplifier, which is then said to be in *regenerative* operation. One must be careful to distinguish between regeneration and generation, or oscillation. The question of whether or not the system oscillates, and at what frequency, depends on the value of the *oscillation constant*, which must take into account the parameters of the entire system, including the source (Sec. 12.8). A regenerative

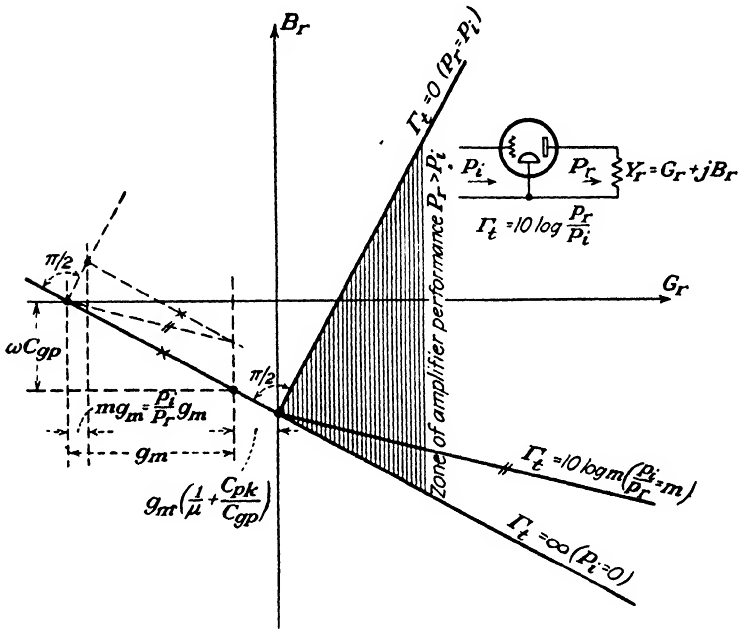


FIG. 97.—*Transmission gain as a function of load admittance (at high frequencies). Cathode to ground connection.*

amplifier will enter into oscillation only if the source admittance components are within given limits.

Above the $\Gamma_T = 0$ line, the tube absorbs signal power in excess of the output and is equivalent to a dissipative network.

Along the $\Gamma_T = \infty$ line, the power ratio is infinite, because of the fact that the input admittance of the tube is imaginary and the tube does not absorb any power. If this condition is approached, the assumption that maximum power transfer takes place between the tube and the source is obviously in error. This assumption is based on the premise (Sec. 9.2) that the efficiency of the coupling networks is high. When the power transmitted through a coupling is small, the losses in the coupling cannot be neglected. Hence the present method does not permit a valid appreciation of the amplifier's effectiveness when its input impedance is entirely, or almost entirely, reactive. On the other hand, such a condition should be avoided for stable high-frequency operation, because it is near the borderline of the regenerative region (Fig. 97). This is not objectionable at low frequencies, or for very low values of C_{pg}, because then the power which the tube is capable of feeding back into the source is minute and incapable of causing sustained oscillations (mathematically, the input conductance G_i must be equal and opposite to G_s, source conductance, for oscillations to take place: at low frequency G_i is invariably small).

Barring the low-frequency case, which is adequately handled by simpler methods (Sec. 16.5), we will take up in the following section the problem of designing a stable amplifier capable of a given increase of load power over the maximum power transfer value.

17.3. Criteria for Stable High-frequency Amplifier Design. Cathode to Ground Connection. Suppose we require an increase of power of Γ_0 db over the m.p.t. value at the angular frequency ω_0. The output to input power ratio will have to be

$$\frac{P_r}{P_i} = \log^{-1}\frac{\Gamma_0}{10}$$

Corresponding to this ratio there is a line on the GB plane which we will now locate. We must have (572)

$$\log^{-1}\frac{\Gamma_0}{10} = \frac{1 + (g_m/\omega_0 C_{gp})^2}{1 + (g_p/G_r) + (g_m/G_r)[1 + (C_{pk}/C_{gp}) + (B_r/\omega_0 C_{gp})]}$$

The last line is a linear equation in B_r and G_r, representing a straight line on the G_rB_r plane. The B_r-axis intercept of this line is the same as for the lines $\Gamma = \infty$ and $\Gamma = 0$ previously discussed. The slope of the new line is

$$\frac{1}{\log^{-1}\Gamma_0/10}\left(\frac{g_m}{\omega_0 C_{gp}} + \frac{\omega_0 C_{gp}}{g_m}\right) - \frac{\omega_0 C_{gp}}{g_m} \qquad (578)$$

The construction of Fig. 97 is based on the above values for the slope and the intercept. This construction can be carried out in reverse direction. Given a value of the load admittance, the transmission gain (hence the increase in transmitted power due to the tube being inserted) can be found in this way.

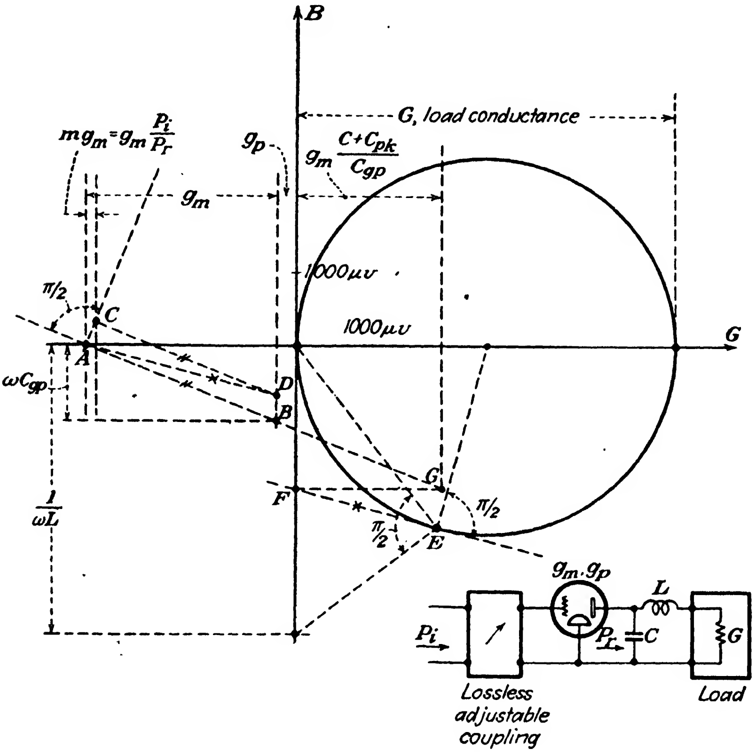


FIG. 98.—*Example of amplifier-output coupling designed to ensure a specified maximum of transmission gain at the operating frequency. Construction yields correct values of L and C.*

Data used: $g_m = 2{,}500\mu\mho$; $g_p = 250\mu\mho$; $C_{gp} = 4\mu\mu F$; $C_{pk} = 0.6\mu\mu F$; load conductance $G = 5{,}000\mu\mho$ (200Ω); frequency 40 *mc.*/sec.

Desired transmission gain: 30 *db*, for which $m = Pi/Pr = 0.05$

From construction (points are obtained in alphabetical order):

$$g_m \frac{C + C_{pk}}{C_{gp}} = 1{,}900 \qquad \frac{1}{\omega L} = 3{,}900\mu\mho$$

Hence:

$$C = 2.44\mu\mu F \qquad L = 1.37\mu H$$

The problem thus reduces to the design of a coupling network, such that the admittance Y_r looking into the network, towards the load, plotted as a point on the complex plane, lies on the constant Γ line corresponding to the required gain, which we assume to have been drawn by the above construction.

This must be true at the operating frequency. At other frequencies, the point Y_r should always lie *above* the line (for stability). Normally, Y_r will move rapidly with frequency, so that by comparison the constant Γ line may be considered stationary. Under these conditions, the polar plot of Y_r for variable frequency must be tangent to the constant Γ line. We can secure this condition by correct proportioning of the coupling elements.

Take, for example, the coil-condenser coupling of Fig. 98. The admittance point, for a passive two-pole, as we know from a-c circuit theory, describes with increasing frequency a path which *curves toward the right.*[1] The same is true of the impedance point. For a coil and resistance in series, this path is a circle through the origin. The presence of the shunting condenser complicates matters; however, we can regard this condenser as part of C_{pk}, and observe that a change in C_{pk} simply shifts the entire construction of Fig. 97 in relation to the B_r axis.

The construction of Fig. 98 is based on this observation and leads directly to the values of inductance and capacity to be used in the coupling for the desired gain consistent with stable operation. A numerical example is given.

As a general rule, to avoid instability, tank circuits or stubs should not be placed directly across the tube output without additional load, unless their Q_0 is relatively poor. The admittance plot of a stub (Fig. 54) will inevitably cross the $\Gamma = \infty$ line at some frequency, and oscillations will result when coupling adjustments are made on the input side.

17.4. Plate to Ground Connection. The voltage ratio of a tube in the plate to ground connection (cathode follower) was found to be (558)

$$\frac{\mathbf{V}_2}{\mathbf{V}_1} = \frac{j\omega C_{gk} + g_m}{Y_r' + j\omega C_{gk}} = \frac{g_m + j\omega C_{gk}}{g_m + G_r + j[B_r + \omega C_{gk} + \omega C_{pk}]}$$

Evidently, this ratio cannot exceed unity; it is therefore commonly understood that the gain of the cathode follower is always negative. This is true only of the voltage gain, however. The *transmission gain* (logarithmic ratio of power output to power input) can have any value, depending on the load admit-

[1] Toward the right of an observer following the Y point in its movement on the complex plane.

tance (admittance between cathode and ground or cathode and plate). It is possible for the power ratio to be negative, in which case, if the proper value of admittance is connected on the input side, the tube will oscillate without external feedback.

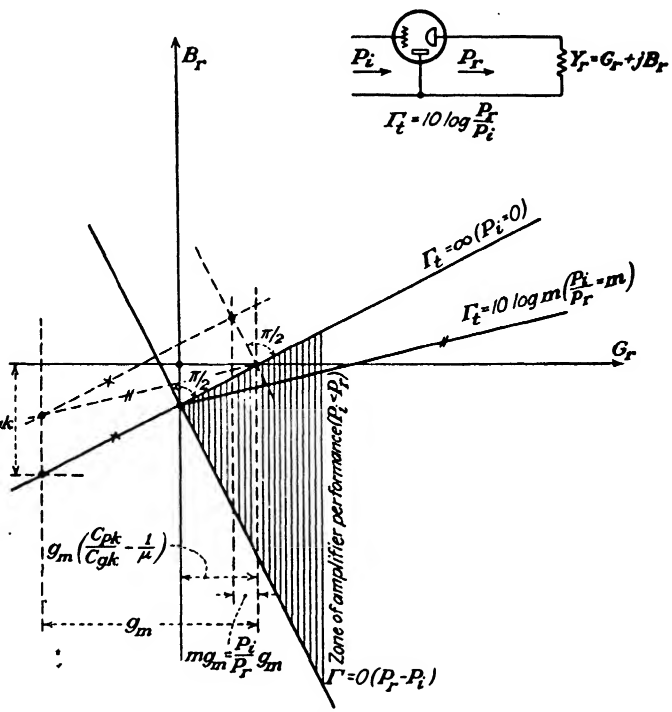


FIG. 99.—*Transmission gain as a function of load admittance (at high frequencies). Plate to ground connection.*

The dependence of transmission gain on load admittance can be analyzed by the method already used for the cathode to ground connection. The algebraic steps are very similar and will be omitted. The final expression for the transmission gain is as follows:

$$\Gamma_T \text{ (db)} = 10 \log \frac{P_r}{P_i} = 10 \log \frac{1 + (g_m/\omega C_{gk})^2}{1 + (g_p/G_r) - (g_m/G_r)[(C_{pk}/C_{gk}) + (B_r/\omega C_{gk})]} \quad (579)$$

[*Transmission gain of the cathode follower stage*]

On the Y_r plane (Fig. 97) may be drawn, as before, loci of constant Γ_T. The loci $\Gamma_T = 0$ and $\Gamma_T = \infty$ are, as before, straight orthogonal lines; however, the $\Gamma_T = \infty$ line now crosses the axis of reals. This means that cathode follower stages may be regenerative when the load susceptance is positive (capacitive output). This is impossible if the grounded cathode connection is used. Figure 99 also shows the construction for obtaining the line corresponding to any particular value of transmission gain. This construction is similar to that of Fig. 97.

It is worth noting that, if the condition

$$\frac{g_p}{g_m} = \frac{C_{pk}}{C_{gk}}$$

were met, all the constant Γ_T lines would meet at the origin; or, in other words, the transmission gain would depend only on the angle of the load admittance. The above condition may be written

$$\mu = \frac{C_{gk}}{C_{pk}}$$

and we know that this is satisfied in the theoretical case of an indefinite plane parallel triode. In practice, the condition cannot be realized because of end effects; however, it may be approached.

The cathode follower connection differs markedly from the grounded cathode connection in one respect. As the frequency is varied, the transmission gain of the grounded cathode stage cannot be kept continuously at a high value; this would require a negative load susceptance directly proportional to the frequency, which cannot be physically realized. To obtain such high values we must rely on a selective load, which must be tuned to the correct frequency; moreover, as was shown, the value of Q_0 for this selective load must be limited or instability will result.

If the cathode follower connection is used, we can, on the contrary, ensure an infinite value of transmission gain at all frequencies; this can be deduced from an inspection of Fig. 99, or as a consequence of the discussion of Sec. 16.6. It was shown there that if the load of a cathode follower is a conductance, shunted by a suitable capacity, the input admittance of the tube is that of a pure capacity, independent of frequency. Under these conditions, the transmission gain must be infinite because no power is absorbed by the tube, while some power is certainly delivered to the load conductance.

In practice, of course, a finite insertion gain will be realized, depending on the efficiency of the input coupling, as a practical example will show. This design has the advantage of greater stability and reduction in the number of necessary adjustments.

17.5. Extension of Linear Concepts to Amplifiers in Nonlinear Operation. The foregoing discussion on amplifiers has been based upon the equation

$$\Delta i_p = g_m \, \Delta v_g + g_p \, \Delta v_p \tag{545}$$

expressing the *variational value* of plate current (departure from rest) as a *linear function* of the variational values of grid and plate voltages.

This linear relationship, as was pointed out, is approached only for small signals. It is not realized when large signals are used, particularly if the grid voltage exceeds the cutoff value at any time during the cycle.

This restriction is not quite so serious as might be supposed. While it is true that (545) holds only for small signals, the complex equation

$$\mathbf{I}_p = g_m \mathbf{V}_g + g_p \mathbf{V}_p \tag{550}$$

which follows from (545) when we assume harmonic variations, *continues to be valid* for large signals under one condition—that both grid and plate voltage vary harmonically.

Consider, in fact, the expansion (544) from which (545) was obtained by dropping all terms except those of the first order. When variations are large, these terms cannot be dropped. We must then use the expansion in its original form, or

$$\Delta i_p = g_m \, \Delta v_g + g_p \, \Delta v_p + \frac{1}{2!} \frac{\partial^2 i_p}{\partial v_g^2} \Delta v_g^2 + \frac{1}{2!} \frac{\partial^2 i_p}{\partial v_p^2} \Delta v_p^2 + \frac{2}{2!} \frac{\partial^2 i_p}{\partial v_g \, \partial v_p} \Delta v_g \, \Delta v_p$$

Now, if we assume the voltages to be harmonic time functions, we may substitute as follows:

$$\Delta v_g = \Delta v_g \sin (\omega t + \alpha_1)$$
$$\Delta v_p = \Delta v_p \sin (\omega t + \alpha_2)$$

and the expansion, after some manipulation, takes the form

$$
\begin{aligned}
\Delta i_p = {} & g_m\,\Delta v_g \sin(\omega t + \alpha_1) + g_p\,\Delta v_p \sin(\omega t + \alpha_2) \\
& + \tfrac{1}{4}\frac{\partial^2 i_p}{\partial v_g^2}[1 - \cos 2(\omega t + \alpha_1)]\,\Delta v_g^2 \\
& + \tfrac{1}{4}\frac{\partial^2 i_p}{\partial v_p^2}[1 - \cos 2(\omega t + \alpha_2)]\,\Delta v_p^2 \\
& + \frac{\partial^2 i_p}{\partial v_g\,\partial v_p}[\cos(\alpha_1 - \alpha_2) - \cos(2\omega t + \alpha_1 + \alpha_2)]\,\Delta v_g\,\Delta v_p
\end{aligned}
$$

The constant terms of this summation stand for variations in average value; their sum constitutes the zero frequency component of Δi_p. Terms contributing to the *fundamental frequency* ω and to each *harmonic frequency* are easily recognized. To each of these harmonic components of Δi_p we may associate a complex current, or current vector. Thus, we may write, for the fundamental component,

$$\mathbf{I}_{p1} = g_m\mathbf{V}_g + g_p\mathbf{V}_p$$

for the second harmonic,

$$
\begin{aligned}
\mathbf{I}_{p2} = \Bigg[& \tfrac{1}{4}\frac{\partial^2 i_p}{\partial v_g^2}\,\Delta v_g^2 e^{j(\alpha_1-\pi/2)} + \tfrac{1}{4}\frac{\partial^2 i_p}{\partial v_g^2}\,\Delta v_p^2 e^{j(\alpha_2-\pi/2)} \\
& + \frac{\partial^2 i_p}{\partial v_g\,\partial v_p}\,\Delta v_g\,\Delta v_p e^{j(\alpha_1+\alpha_2-\pi/2)}\Bigg]
\end{aligned}
$$

and so on. Equation (550) therefore still holds, except that it no longer expresses the entire current variation, but only its fundamental component. This omission does not matter so far as the voltages and currents outside the tube are concerned; we have, in fact, postulated that the voltages at the input and output terminals (plate and grid voltages) are sinusoidal at the fundamental frequency.

According to this assumption, the output voltage is sinusoidal while the output current is not. This is not a difficult condition to realize; it is fulfilled, for practical purposes, whenever the load impedance is sharply resonant to the fundamental frequency. All tuned amplifiers are in this category, particularly class *B* and class *C* amplifiers, which operate with large voltages.

We conclude that, in principle, linear analysis can be extended to all tuned amplifiers. In practice, the values of g_m and g_p to be used for large-signal operation differ from the small-signal values and will vary somewhat with the signal amplitude. These values should be considered as equivalent or effective values, much as the values of mutual and self-inductances of a power transformer.

We cannot, of course, apply linear analysis when grid and plate voltage are sinusoidal but of different frequency, or when either is the sum of two frequencies.

17.6. Application: Increase in Load Power Made Possible by Amplifier Insertion. Bringing the analysis of Sec. 17.4. to a practical level, consider the two transmission systems of Fig. 100.

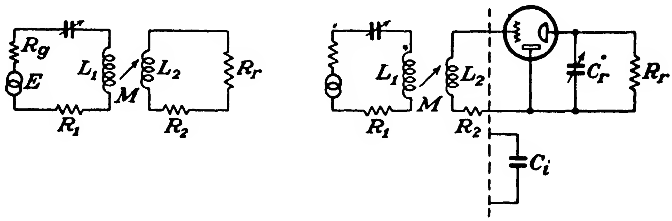


FIG. 100.—*Use of grounded plate connection for power amplification.*

In the first system (Fig. 100*a*), a coupling link, consisting of a transformer and condenser in series with the primary side, is used to transfer power from a resistive source, of resistance R_g, to a resistive load R_r. These values might stand for the characteristic impedances of transmission lines or cables; they may or may not be equal.

In the second system the same link is used to drive a grounded plate triode or pentode, connected on the output side to R_r in shunt with an adjustable capacity C_r. We will assume that C_r is adjusted to the value given by Eq. (564), so that the tube input looks like a capacity C_i, whose value is given by (566).

In the first system the loss resistances of the transformer windings may be neglected (in practice, these windings at high frequencies will consist of a few turns of heavy wire). Hence the power in the load can, by the proper adjustment of the coupling, reach the highest value that the source can put out, or

$$P_0 = \frac{E^2}{4R_g} \tag{580}$$

In the second system, all the power leaving the source (except a negligible fraction lost in R_1) must end up in the winding resistance R_2. Since this power can again, by adjusting the coupling, be made equal to the value given above, we can compute the current through the secondary under maximum power transfer conditions. This is

$$I_{\max} = \sqrt{\frac{P_0}{R_2}} = \frac{E}{\sqrt{R_g R_2}}$$

The voltage developed across C_i is therefore

$$V_{\max} = \frac{I}{\omega C_i} = \frac{E}{2\omega C_i \sqrt{R_g R_2}}$$

If we wrote the condition of maximum power transfer in terms of the coupling parameters, we would find that in order to meet this condition, when the load is capacitive, the secondary must be at or near resonance. We must, therefore, imagine the secondary inductance to be adjusted until it meets this requirement. We will then have

$$\omega C_i = \frac{1}{\omega L_2}$$

and in consequence, letting as usual $Q_2 = \omega L_2/R_2$,

$$V_{\max} = \frac{E}{2}\sqrt{\frac{Q_2}{\omega C_i R_g}}$$

Now the voltage ratio across the tube has a constant value, given by (565). Hence, the voltage across the load will be

$$V_r = \frac{E}{2}\,\frac{g_m}{g_m + g_p + (1/R_r)}\sqrt{\frac{Q_2}{\omega C_i R_g}}$$

From this we readily obtain the power in the load

$$P_r = \frac{V_r^2}{R_r} = \frac{E^2}{4}\,\frac{g_m^2 Q_2}{R_r R_g \omega C_i [g_m + g_p + (1/R_r)]^2}$$

and the ratio between this value and the power obtainable without the tube (580)

$$\frac{P_r}{P_0} = \frac{g_m^2 Q_2}{\omega C_i R_r [g_m + g_p + (1/R_r)]^2}$$

We can make this expression more significant by substituting the value for C_i, as given by (566), or (neglecting from now on g_p

in comparison with g_m and $1/R_r$)

$$C_i = C_{gp} + C_{gk}\frac{1}{1 + g_m R_r} \tag{581}$$

Thus

$$\frac{P_r}{P_0} = \frac{R_r g_m{}^2 Q_2}{\omega C_{gk}\,[(C_{gp}/C_{gk}) + (1/1 + g_m R_r)](1 + g_m R_r)^2}$$

Hence the decibel gain in power

$$\Gamma\ (\text{db}) = 10 \log Q_2 \frac{g_m}{\omega C_{gk}} \frac{1}{[1 + (G_r/g_m)]\{1 + (C_{gp}/C_{gk})[1 + (g_m/G_r)]\}} \tag{582}$$

$$[G_r = (1/R_r)]$$

As an example, for the following data:

$$\begin{aligned} g_m &= 3{,}000\ \mu\text{mhos} \\ C_{gk} &= C_{gp} = 2\ \mu\mu\text{f} \\ \omega &= 3.14 \times 10^8 (50\ \text{Mc./sec.}) \\ Q_2 &= 250 \\ G_r &= 10{,}000\ \mu\text{mhos (100 ohms)} \end{aligned}$$

the gain has the value

$$\Gamma\ (\text{db}) = 10 \log \frac{3{,}000 \times 250}{2 \times 314(1 + 3.33)(1 + 1.3)} = 20.8$$

This example shows how the grounded plate connection may, under certain conditions, be used to advantage as an amplifier. It is important to realize that voltage gain is only one aspect of amplifier performance. In the above example, the voltage gain of the tube has the negative value -12.76 db.

BIBLIOGRAPHY

The references listed below cover fields bordering upon the subject matter of this book. They are suggested as sources of general information. The reader is referred to the subject indices of periodicals for articles dealing with particular phases of the literature, as these are far too numerous to list adequately.

1. Fourier series and harmonic analysis

GUILLEMIN, E. A.: "Communications Networks," vol. I, Chap. X, John Wiley & Sons, Inc., New York.

GARDNER, M. F. and J. L. BARNES: "Transients in Linear Systems," vol. I, Chap. III, John Wiley & Sons, Inc., New York.

BYERLY, W. E.: "Fourier Series and Spherical Harmonics," Ginn and Company, Boston.

2. Electromechanical analogies

OLSON, H. F.: "Dynamical Analogies," D. Van Nostrand Company, Inc., New York.

JOHNSON, W. C.: "Mathematical and Physical Principles of Engineering Analysis," Chap. IV, McGraw-Hill Book Company, Inc., New York.

GUILLEMIN, E. A.: "Communication Networks," vol. I, Chap. I, John Wiley & Sons, Inc., New York.

GARDNER, M. F. and J. L. BARNES: "Transients in Linear Systems," John Wiley & Sons, Inc., New York.

3. Speech and hearing

FLETCHER, H.: "Speech and Hearing," D. Van Nostrand Company, Inc., New York.

CRANDALL, I. B. and D. MACKENZIE: "Analysis of the Energy Distribution in Speech," *Bell System Technical Journal*, Vol. IV, p. 586.

4. Functions of complex variables

MCLACHLEN, N. W.: "Complex Variable and Operational Calculus," Cambridge University Press, London.

PIERPONT, J.: "Functions of a Complex Variable," Ginn and Company, Boston.

BURINGTON, R. S. and C. C. TORRANCE: "Higher Mathematics with Applications to Science and Engineering," Chap. V, McGraw-Hill Book Company, Inc., New York.

TOWNSEND, E. J.: "Functions of a Complex Variable," Henry Holt and Company, Inc., New York.

5. Wave filters and network synthesis

SHEA, T. E.: "Transmission Networks and Wave Filters," D. Van Nostrand Company, Inc., New York.

GUILLEMIN, E. A.: "Communication Networks," vol. II, John Wiley & Sons, Inc., New York.

McILWAIN, K. and J. C. BRAINERD: "High Frequency Alternating Currents," Chap. IX, John Wiley & Sons, Inc., New York.

ZOBEL, O. J.: "Theory and Design of Uniform and Composite Electric Wave Filters," *Bell System Technical Journal*, January, 1923, Vol. 2, pp. 1–46.

JOHNSON, K. S.: "Transmission Circuits for Telephonic Communication," D. Van Nostrand Company, Inc., New York.

6. Telephonic lines and cables

HILL, J. G.: "Telephonic Transmission," Longmans, Green and Company, New York.

WRIGHT, C. A. and A. F. PUCHSTEIN: "Telephone Communication," McGraw-Hill Book Company, Inc., New York.

FLEMING, J. A.: "The Propagation of Electric Currents in Telephone and Telegraph Conductors," Constable and Company, Ltd., London.

WEINBACH, M. P.: "Principles of Transmission in Telephony," The Macmillan Company, New York.

KENNELLY, A. E.: "Electric Lines and Nets," McGraw-Hill Book Company, Inc., New York.

KENNELLY, A. E.: "The Application of Hyperbolic Functions to Electrical Engineering Problems," McGraw-Hill Book Company, Inc., New York.

FRANKLIN, W. S. and S. E. TERMAN: "Transmission Line Theory," Franklin & Charles, Lancaster, Pa.

7. Electrical measurements

CURTIS, H. L.: "Electrical Measurements," McGraw-Hill Book Company, Inc., New York.

CAMPBELL, A. and E. C. CHILDS: "Measurement of Inductance, Capacitance, and Frequency," D. Van Nostrand Company, Inc., New York.

SMYTHE, W. R. and W. C. MICHELS: "Advanced Electrical Measurements," D. Van Nostrand Company, Inc., New York.

HUND, A.: "High-frequency Measurements," McGraw-Hill Book Company, Inc., New York.

BRAINERD, J. G., G. KOEHLER, H. J. REICH, and I. WOODRUFF: "Ultra-high-frequency Techniques," D. Van Nostrand Company, Inc., New York.

8. Transients in networks and lines; oscillation of linear systems

GARDNER, M. F. and J. L. BARNES: "Transients in Linear Systems," John Wiley & Sons, Inc., New York.

PIERCE, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill Book Company, Inc., New York.

Bewley, L. V.: "Traveling Waves on Transmission Systems," John Wiley & Sons, Inc., New York.

Guillemin, E. A.: "Communication Networks," vol. I, John Wiley & Sons, Inc., New York.

9. Electric and magnetic fields

Jeans, J. H.: "Electricity and Magnetism," Cambridge University Press, London.

Attwood, S.: "Electric and Magnetic Fields," John Wiley & Sons, Inc., New York.

Smythe, W. R.: "Static and Dynamic Electricity," McGraw-Hill Book Company, Inc., New York.

Starling, S. G.: "Electricity and Magnetism," Longmans, Green and Company, New York.

10. Electromagnetic waves

Shelkunoff, S. A.: "Electromagnetic Waves," D. Van Nostrand Company, Inc., New York.

Skilling, H. H.: "Fundamentals of Electric Waves," John Wiley & Sons, Inc., New York.

Slater, J. C.: "Microwave Transmission," McGraw-Hill Book Company, Inc., New York.

Ramo, S. and J. R. Whinnery: "Fields and Waves of Modern Radio," John Wiley & Sons, Inc., New York.

Sarbacher, R. I. and W. A. Edson: "Hyper and Ultrahigh Frequency Engineering," John Wiley & Sons, Inc., New York.

11. Electronics

Reich, H. J.: "Theory and Application of Electron Tubes," McGraw-Hill Book Company, Inc., New York.

Chaffee, L. E.: "Theory of Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., New York.

Millman, J. and S. Seely: "Electronics," McGraw-Hill Book Company, Inc., New York.

M.I.T. Staff, "Applied Electronics," John Wiley & Sons, Inc., New York.

Terman, F. E.: "Radio Engineering," 2d ed., McGraw-Hill Book Company, Inc., New York.

INDEX

S

T

07