NETWORK ANALYSIS

M. E. VAN VALKENBURG

NETWORK ANALYSIS

By

M. E. VAN VALKENBURG

Professor of Electrical Engineering
University of Illinois

PRENTICE-HALL, INC.

Englewood Cliffs, N.J.

PRENTICE-HALL ELECTRICAL ENGINEERING SERIES
W. L. EVERITT, Ph.D., *Editor*

LIBRARY OF CONGRESS
CATALOG CARD NO.: 56-5609

First printing.............*November, 1955*
Second printing...............*June, 1956*
Third printing...............*August, 1957*
Fourth printing................*July, 1958*
Fifth printing.............*November, 1958*
Sixth printing.................*June, 1959*

PRINTED IN THE UNITED STATES OF AMERICA
61082

PREFACE

This book is designed for use as an introduction to the study of electric networks from the so-called pole and zero approach. The subject matter may be divided into four parts. (1) Chapters 1 through 3 are concerned with definitions and with the formulation of equilibrium equations. The first chapter contains a discussion of approximation as it relates to networks, the relationship between the network abstraction and the physical system. Here the reader is given the opportunity to plant his feet firmly on the ground before he becomes involved in the myriad details of analysis. The elements are introduced, their laws formulated, their combination into networks discussed. Writing of equilibrium equations for networks is treated in Chapter 3.

(2) Chapters 4 through 8 have to do with the solution of equilibrium equations, integrodifferential equations in general, by both classical and the Laplace transform method. Chapter 8 amplifies the relationship between the time domain and the frequency domain.

These first eight chapters encompass topics classified under the heading of transient analysis of electric circuits. In the remaining chapters, this background is exploited in unifying concepts of transient response and sinusoidal steady-state response by the use of the poles and zeros of network functions.

(3) In Chapters 9 through 11, the reader is introduced to complex frequency, impedance functions, transfer functions, and poles and zeros.

(4) The remaining chapters are devoted to applications of the pole and zero approach to network analysis. Chapters 12 and 13 relate to reactive networks and include Foster's reactance theorem and filters studied from the image parameter point of view. Chapter 14 is an introduction to stagger-tuned amplifier-networks. In the last two chapters, the representation of systems by block diagrams and the stability of feedback systems are studied. References are given at the end of each chapter for those interested in a more advanced or more detailed study.

The literature relating to most of the contents of the book dates back to the 1920's. It has not been until recent years, however, that the pole and zero approach has been widely taught in graduate schools and extensively used by electrical engineers in industry in such areas as circuit design, electronic circuits, and automatic control. This pole and zero approach is now finding its way into the undergraduate curriculum in many different areas of study and in many ways.

The material of *Network Analysis* has been developed by using it in the form of classroom notes for a course for junior and senior students at the University of Utah, which has been offered since 1949. The objective of this course has been to provide background material for the study of such subjects as communications engineering, pulse techniques, power system analysis, and servomechanisms.

I am deeply indebted to my students at the University of Utah—and at Stanford University for some chapters—whose questions and classroom discussions have left many imprints on the book. Indebtedness is also acknowledged to Dr. Glen Wade of Stanford University and Dr. Don A. Baker and Doran Baker of the University of Utah for reading parts of the manuscript and offering helpful suggestions, and to Dean W. L. Everitt, editor of this series.

It was my good fortune to study network synthesis under Professor David F. Tuttle, Jr., at Stanford University. In preparing this book, I have been influenced by his method of approach and his teaching techniques. I also acknowledge the friendly cooperation of the electrical engineering staff at the University of Utah, particularly Professors L. Dale Harris and Philip Weinberg.

M. E. Van Valkenburg

Urbana, Illinois

CONTENTS

CHAPTER 1

DEVELOPMENT OF THE CIRCUIT CONCEPT

1-1. Introduction

One of the methods of science is the continual bringing together of a wide variety of facts to fit into a simple, understandable theory that will account for as many observations as possible. The name *conceptual scheme* has been used by the American chemist and university president James Conant for the theory or picture that results.* Perhaps the most familiar conceptual scheme to students of science and engineering is that of the atomic theory from which we take our picture of the electron and of electric charge. Other important conceptual schemes are conservation of energy and conservation of charge.

Although electricity and magnetism were recognized early in the history of man—the charging of amber by friction, the use of the lodestone in navigation—it was not until the nineteenth century that significant progress was made in developing a conceptual scheme. The discovery by Galvani and Volta about 1800 that electricity could be produced by chemical means greatly simplified experimentation. Important discoveries were made in a relatively short interval of time after Volta. In 1820, Oersted identified the magnetic field with current, and Ampere measured the force caused by the current. In 1831 Faraday, and independently Henry, discovered electric induction. These and other experiments were brought together to form a successful conceptual scheme by the English physicist James Clerk Maxwell in 1873. In Maxwell's equations, as the scheme has come to be known, all electric and magnetic phenomena are explained in terms of fields resulting from charge and current. The success of Maxwell's conceptual scheme is evidenced by the persistent agreement of results deduced from Maxwell's equations with observation for a period of over 100 years.

In view of Maxwell's success, why do we now embark upon a study of *another* conceptual scheme for the same phenomena, the electric circuit? Equally important as a question, how are the two concepts related? The answer to the first of our questions is the practical utility of the circuit concept. As a practical matter, we are not often interested

* James B. Conant, *Science and Common Sense* (Yale University Press, New Haven, 1951).

in fields so much as we are in voltages and currents. The circuit concept favors analysis in terms of voltage and current from which other quantities such as charge, fields, energy, power, etc. can be computed if desired. The answer to our second question will require a longer answer and justification. Briefly, circuit concepts arise from the same basic experimental facts as do Maxwell's equations. However, the circuit involves approximations that are not included in the more general concept of field theory. It is important that we understand the nature of these approximations—the limitations of circuit theory—before we develop our subject.

It will be helpful to define the function of the circuit in terms of two basic building blocks: charge and energy. We regard charge and energy as the least common denominators in describing electrical phenomena, the primitive quantities in terms of which we can build our conceptual scheme of the electric circuit. A physical circuit is a system of interconnected apparatus. Here we use the word *apparatus* to include sources of energy, connecting wires, components, loads, etc. A circuit functions to transfer and transform energy. Energy transfer is accomplished by charge transfer. In the circuit, energy is transferred from a point of supply (the source) to a point of transformation or conversion called the load (or sink). In the process, the energy may be stored.

1-2. Electric charge

Thales of Greece is credited with the discovery about 600 B.C. that amber when briskly rubbed with a piece of silk or fur becomes "electrified" and is capable of attracting small pieces of thread. This same technique for producing electricity was used centuries later by Coulomb in France (and independently by Cavendish in England) in establishing the inverse square law of attraction of charged bodies.

Our present-day understanding of the nature of charge is based on the conceptual scheme of the atomic theory. We picture the atom as composed of a positively charged nucleus surrounded by negatively charged electrons. In the neutral atom, the total charge of the nucleus is equal to the total charge of the electrons. When electrons are removed from a substance, that substance becomes positively charged. A substance with an excess of electrons is negatively charged.

The basic unit of charge is the charge of the electron. Because this charge is so small, the practical unit of the *coulomb* is used. The electron has a charge of 1.601×10^{-19} coulomb.

1-3. Electric current

The phenomenon of transferring charge from one point in a circuit to another is described by the term *electric current*. An electric current may be defined as the time rate of net motion of electric charge across a cross-sectional boundary. A random motion of electrons in a metal does not constitute a current unless there is a net transfer of charge with time.

In equation form, the current* is

$$i = \frac{dq}{dt} \tag{1-1}$$

If the charge q is given in coulombs and the time t is measured in seconds, the current is measured in *amperes* (after the French physicist André Ampère). Since the electron has a charge of 1.601×10^{-19} coulomb, it follows that a current of 1 ampere corresponds to the motion of $1/(1.601 \times 10^{-19}) = 6.25 \times 10^{18}$ electrons past any cross section of a conducting path in 1 sec.

In terms of the atomic theory conceptual scheme, all substances are pictured as made up of atoms. In a solid, some electrons are relatively free of the nucleus; the attractive forces on these electrons are exceedingly small. Such electrons are distinguished by the name *free electrons*. An electric current is the time rate of flow of these free electrons, passing from one atom to the next as pictured in Fig. 1-1.

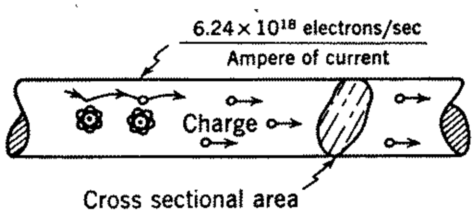


Fig. 1-1. Motion of charge in a conductor.

In some materials, there are many free electrons, so that large currents are easily attained. Such materials are known as *conductors*. Most metals and some liquids are good conductors. Materials with relatively few free electrons are known as *insulators*. Common insulating materials include glass, mica, plastics, etc. There is no sharp dividing line between conductors and insulators. Conduction is possible in other materials than solids. In vacuum tubes, for example, electrons pass through a partial vacuum from one metallic plate to another.

There is a common misconception that since some electric waves propagate at approximately the speed of light the electrons in a conductor travel with this same velocity. The actual mean velocity of free electron drift is but a few millimeters per second! (See Prob. 1-2 for a numerical example.)

1-4. Sources of energy; electric potential

Another conceptual scheme upon which our thinking is based is the *conservation of energy*. By our training in the methods of science, we

* The symbol i for current is taken from the French word *intensité*.

immediately become suspicious of any scheme that creates energy. The law of conservation of energy states that energy cannot be created, nor destroyed, but that it can be converted in form. Electric energy is energy converted from some other form. There are relatively few ways that this can be accomplished. In order of economic importance, some of these methods are the following:

(1) *Magnetic induction.* The familiar rotating generator invented by Faraday in 1831 produces electric energy from mechanical energy of rotation. Often the mechanical energy is converted from thermal energy by a turbine, and in turn, the thermal energy is converted from chemical energy by burning coal.
(2) *Voltaic methods.* Electric batteries produce electric energy by converting chemical energy.
(3) *Electrostatic methods.* The friction machines used by Coulomb and other early experimenters produce electric energy by converting mechanical energy. This method is little used at present, an exception being the Van de Graaff generator used to produce x rays and used in research in nuclear physics.
(4) *Other methods.* Thermal electricity is produced by heat at a junction of dissimilar metals such as bismuth and copper. Light energy can be converted into electric energy by photoelectric devices.

The function of each of these different sources of electric energy is the same in terms of energy and charge. In one form of battery, for example, two metallic electrodes—one of zinc and one of copper—are immersed in dilute sulfuric acid. The formation of zinc and copper ions causes negative charge to accumulate at the electrodes. Energy is supplied to the charge by the difference in the energy of ionization of zinc and copper in the chemical reaction. Once the battery circuit is closed by an external connection, as shown in Fig. 1-2, the chemical energy is expended as work for each unit of charge in transporting the charge around the external circuit. The quantity "energy per unit charge" or identically, "work per unit charge," is given the name *potential* (or the more commonly used term *voltage*). In the form of an equation,

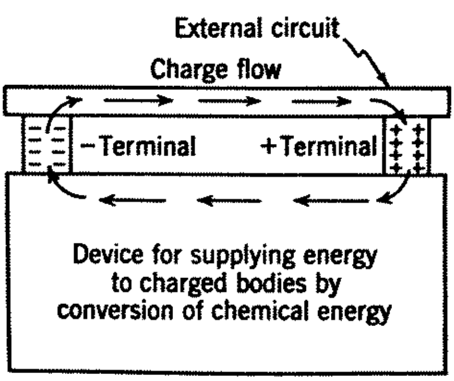


Fig. 1-2. Representation of a battery showing electron flow.

$$v = \frac{w}{q} \tag{1-2}$$

If w is the work (or energy) in joules and q is the charge in coulombs, the potential v is in *volts* (after Alessandro Volta). The potential of an energy source is sometimes described by the term *electromotive force*, abbreviated *emf*, in the electrical literature. We will avoid designating potential as a force because it is misleading and instead use the terms *voltage* or *potential*.

If a differential amount of charge dq is given a differential increase in energy dw, the potential of the charge is increased by the amount

$$v = \frac{dw}{dq} \tag{1-3}$$

If this potential is multiplied by the current, dq/dt as

$$\frac{dw}{dq} \times \frac{dq}{dt} = \frac{dw}{dt} = p \tag{1-4}$$

the result is seen to be a time rate of change of energy, which is *power* p. Thus power is the product of potential and current,

$$p = vi \tag{1-5}$$

and energy is given by the integral equation

$$w = \int p\,dt = \int vi\,dt \tag{1-6}$$

1-5. The relationship of field and circuit concepts

In developing the circuit conceptual scheme, we will follow the same three steps for each of three parameters. These steps are the following:

(1) *The physical phenomenon.* We will discuss in a quantitative manner an electrical phenomenon which is observed by experiment. We will do this in terms of charge and energy.
(2) *Field interpretation.* We will next discuss the interpretation of the phenomenon in terms of a field quantity.
(3) *Circuit interpretation.* Finally, we will introduce a circuit parameter to relate voltage and current in place of the field relationship.

1-6. The capacitance parameter

(1) *Physical phenomenon.* The presence of *charge* on two spatially separated substances—for example, those shown in Fig. 1-3—causes an "action at a distance" in the form of a force between the two substances. This phenomenon we regard as a property of nature, a basic experimental fact. Coulomb found that this force was of such nature

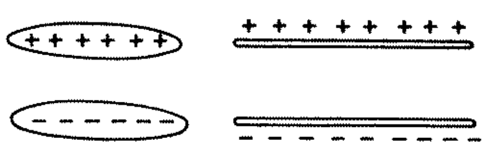


Fig. 1-3. Charged bodies.

that "like charges repel" and "unlike charges attract" and that the force varied according to the equation

$$F = \frac{q_1 q_2}{4\pi\epsilon r^2} \tag{1-7}$$

In this equation, F is the force in newtons directed from point charge to point charge, r is the separation of the point charges in meters, ϵ is the permittivity, having the free-space value of 8.854×10^{-12} farad per meter in the mks system, and q_1 and q_2 are the charges measured in coulombs. It should be understood that this equation applies strictly to point charges only. However, the equation may be applied to any geometry of known charge distribution by vectorially adding all forces.

(2) *Field interpretation.* This phenomenon can be described in terms of a force on a unit charge placed between the two charged bodies. This force per unit charge, a vector quantity since force is a vector quantity, is called an *electric field* of value

$$E = \frac{F}{q} \tag{1-8}$$

As a conceptual aid, this field may be represented by lines drawn in the direction of the force that would be exerted on the unit positive exploring charge at each point. Such lines are illustrated in Fig. 1-4. These lines are conceptual aids: they should not be thought of as actually being present. Using Eqs. 1-7 and 1-8, the electric field may be evaluated for a particular problem.

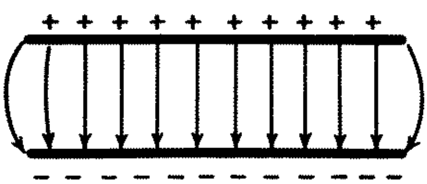


Fig. 1-4. Electric field lines or "lines of force."

(3) *Circuit interpretation.* The work necessary to move a charge from one plate to the other of Fig. 1-4 may be found from the equation

$$w = \int F \cos\theta \, dr \tag{1-9}$$

where dr is an increment of distance between the plates and θ is the angle between the force and the direction of movement dr. An expression for the force has been given by Coulomb's law, Eq. 1-7, which may be substituted into the equation above to give

$$w = \int \frac{q_1 q_2}{4\pi\epsilon r^2} \cos\theta \, dr \tag{1-10}$$

We are more interested in the quantity *work per unit charge,* which is the voltage between the plates. When $q_1 = -q_2 = q$ (equal but opposite charges on the plates), the expression for potential becomes

$$v = \frac{w}{q} = \left(-\int \frac{\cos\theta}{4\pi\epsilon r^2}\,dr\right) q \tag{1-11}$$

The integral of this equation can be evaluated for simple geometry, or in any case can be measured by measuring q and v. For any fixed geometry, the integral is a constant which is given the name *elastance*, symbolized by the letter S. With this definition,

$$v = Sq \tag{1-12}$$

The reciprocal of S is the *capacitance*, which is represented by the letter C. Equation 1-12 may be written

$$q = Cv \tag{1-13}$$

in terms of capacitance. In these equations, if q is measured in coulombs and v in volts, then the unit of C is the *farad* (in honor of Michael Faraday), and the unit for S bears the colorful name *daraf* (farad spelled backwards). The quantity C (or the quantity S) which characterizes the system under study and permits the simple relationship between v and q to be written is known as a *circuit parameter*, the capacitance of a system.*

To reach our objective, a relationship between voltage and current in a capacitive system, there remains the task of studying the relationship of charge and current given by the equation

$$i = \frac{dq}{dt} \tag{1-14}$$

If there is an initial charge on a system, q_0 and the charge increases linearly with time, the charge at any time may be written

$$q = q_0 + kt \tag{1-15}$$

The current is found by differentiating the charge with respect to time, giving the value

$$i = \frac{dq}{dt} = k \tag{1-16}$$

Thus we see that the current in the system is independent of initial charge on that system. In going the other direction, computing charge,

* It should be noted that the circuit parameter described by this equation holds only for the case of two charged bodies with equal and opposite charges. The capacitance concept can be extended, however, to the case of several conductors with any charge distribution. For example, see Fowler, *Introduction to Electric Theory* (Addison-Wesley Publishing Co., Cambridge, 1953), pp. 73 ff.

given the current, we rearrange Eq. 1-14 as

$$dq = i\,dt \tag{1-17}$$

which can be integrated as

$$\int_{q_0}^{q} dq = \int_0^t i\,dt$$

to give

$$q = q_0 + \int_0^t i\,dt \tag{1-18}$$

The total charge on the system of plates is seen to be equal to the sum of the initial charge and the charge deposited by the current.

Returning once more to the relationship, $q = Cv$, current and voltage are related by the equation

$$\frac{dq}{dt} = i = \frac{d}{dt}(Cv) \tag{1-19}$$

If the capacitance C does not vary with time (or with charge), then

$$i = C\frac{dv}{dt} \tag{1-20}$$

If, however, C is not constant but varies as a function of time, the current must be found from the general relationship

$$i = \frac{d}{dt}(Cv) = C\frac{dv}{dt} + v\frac{dC}{dt} \tag{1-21}$$

Similarly, starting with the equation $v = Sq$, we find that

$$v = S\int i\,dt = \frac{1}{C}\int i\,dt \tag{1-22}$$

Equations 1-19 and 1-22 relate the voltage and current in the capacitive system through the circuit parameter C.

Example 1

The sketch of Fig. 1-5(a) shows two plates, one of which is driven by a constant-speed motor so that the capacitance between the two plates varies according to the equation

$$C(t) = C_0(1 - \cos\omega t) \tag{1-23}$$

If the battery potential remains constant at V volts, the current as a function of time may be found from Eq. 1-21 as

$$i = \frac{d}{dt}(Cv) = V\frac{dC}{dt} = \omega C_0 V \sin \omega t \tag{1-24}$$

This time variation of current is shown in Fig. 1-5(c).

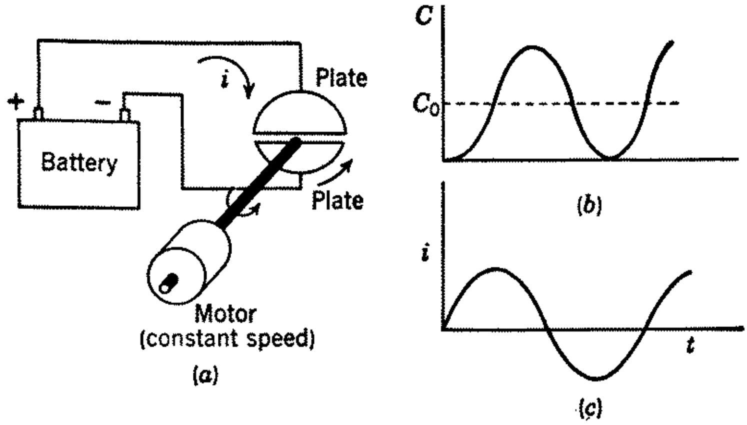


Fig. 1-5. Variable capacitance system.

From the equation $q = Cv$, it is seen that the product Cv cannot change instantaneously, since an instantaneous change in q would mean an infinite current, which is ruled out as a possibility in a physical system. In terms of the time interval $\Delta t = t_2 - t_1$, in which q or Cv changes a finite amount shown in Fig. 1-6, Δt cannot be zero. Instantaneous change of Cv shown as curve 1 is thus ruled out. Typical changes of Cv or q which are permitted are shown as curves 2 and 3.

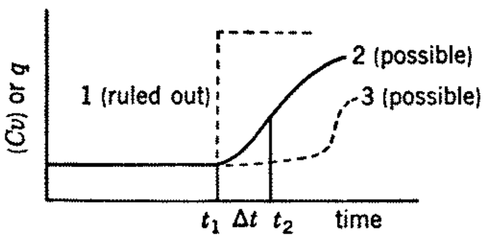


Fig. 1-6. Change of Cv with time.

From another approach, the charge is given as

$$q(t) = q_0 + \int_0^t i\,dt \tag{1-25}$$

by Eq. 1-18. The integral portion of this equation cannot have a finite value in zero time with finite i; that is,

$$\lim_{t\to 0}\int_0^t i\,dt = 0, \qquad i \neq \infty \tag{1-26}$$

The integration process is illustrated in Fig. 1-7 as the summation of infinitesimal areas, i in height and dt in width. The interval from $t = 0$ to t_1 must be greater than zero for any area to be summed.

These mathematical equations aid in visualizing the requirement that the charge in a capacitive system cannot increase or decrease in zero time. However, either capacitance or voltage can change instantaneously so long as the product of the two quantities remains constant, as

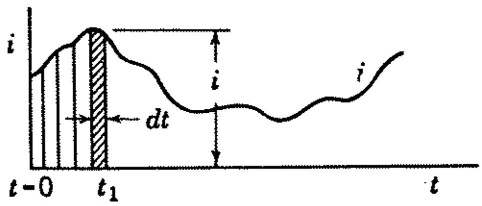


Fig. 1-7. Integration of current.

$$C_1v_1 = C_2v_2 \tag{1-27}$$

where the subscripts 1 and 2 refer to conditions existing at times a vanishingly small interval apart (such as before and after a switch is closed).

In most cases to be considered, the capacitance of a network does not change with time. Under this condition, the above discussion simplifies to the important conclusion that the *voltage of a capacitive system cannot change instantaneously.*

1-7. The inductance parameter

(1) *Physical phenomenon.* Oersted made the important discovery in 1820 that the force between two charged substances depended on the *time rate of flow of charge* (the current). In Oersted's experiment, the needle of a compass was deflected by the presence of a current-carrying conductor, indicating that the effect was related to *magnetism.* In the same year, Ampere measured the force caused by the current and expressed the relationship in equation form. This magnetic effect is an "action at a distance" just as in the case of the force between charged bodies. This "action at a distance" is a basic observational fact; it is not deduced from other knowledge.

(2) *Field interpretation.* The phenomenon described above can be interpreted in terms of the force per unit magnetic pole at all points

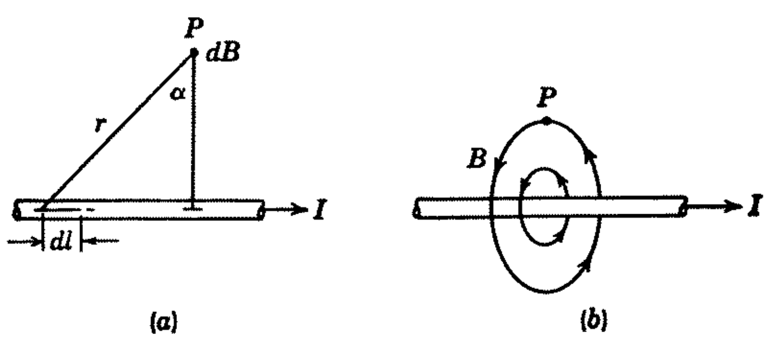


Fig. 1-8. The magnetic field.

in space. Oersted discovered that this force was directed at right angles to the current-carrying conductor. In terms of the geometry of Fig. 1-8(a), Ampere described a *magnetic field density B*, the force per unit magnetic pole, of value

$$dB = \frac{\mu i \cos \alpha \, dl}{4\pi r^2} \tag{1-28}$$

where μ is the magnetic permeability, which is a function of the medium in which the magnetic field exists, i is the current in amperes, and other quantities are defined on the figure. Figure 1-9(a) shows the cross sec-

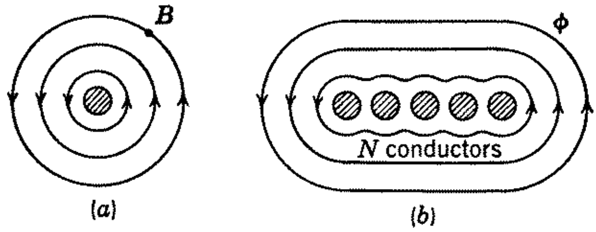


Fig. 1-9. Magnetic field and flux conventions.

tion of a current-carrying conductor. By Eq. 1-28, the magnetic field density will be constant at a constant distance from the conductor. Continuous lines with arrows may be drawn to indicate the direction of B—as a conceptual aid. These are magnetic field density lines or "lines of force." For more complicated geometries than that shown in Fig. 1-9(a), the position of the lines can be found by integrating Eq. 1-28 or by experimentally moving a "point" magnetic pole (if one existed) from place to place in space. A magnetic compass would give an approximate measure of directions.

It is sometimes convenient to replace the lines of magnetic field density by lines of *magnetic flux* defined by the integral equation

$$\phi = \int_s B \cos \theta \, dS \tag{1-29}$$

where θ is the angle between the surface of integration and the field density B. If the currents in each of N conductors, represented in Fig. 1-9(b), are in such a direction that the fluxes add, then $N\phi$ flux linkages* are said to exist. If, however, ϕ_1 lines of flux link N_1 conductors, ϕ_2 lines link N_2 conductors and so forth, the total number of flux linkages is found by algebraic summation as

$$\psi = \sum_{j=1}^{n} N_j \phi_j \tag{1-30}$$

* For a discussion of some of the problems encountered in the use (and misuse) of the concept of flux linkages, see Joseph Slepian, "Lines of force in electric and magnetic fields," *Am. J. Phys.*, **19,** 87 (1951), and Keith McDonald, "Topology of steady current magnetic fields," *Am. J. Phys.*, **22,** 586 (1954).

Assuming that all lines link all conductors, Eq. 1-29 may be modified to give flux linkages, as

$$\psi = N \int_s B \cos \theta \, dS \tag{1-31}$$

To Faraday goes credit for the next basic experimental discovery. Faraday experimented with two conducting circuits in spatial proximity. He found that a *changing* magnetic field produced by one circuit *induced* a voltage in the other circuit. The changing magnetic field could be caused by (1) a conductor moving in space or (2) a current changing with time.

Faraday did not envision this method of inducing voltage in terms of "action at a distance" but in terms of changes in flux linkages. A conductor moving in a magnetic field (as in the case of a generator) is thought of as "cutting flux and hence reducing the flux linkages"; the voltage induced in a stationary conductor (as in a transformer) is thought of as caused by "changing flux linkages" with time. Such pictures are valuable as conceptual aids so long as we do not attach physical significance to flux linkages which are, after all, only a means for accounting for action at a distance. Faraday's law is

$$v = k \frac{d\psi}{dt} \tag{1-32}$$

where k is a proportionality constant. In the mks system the units are selected to make k have unit value: when ψ is in weber-turns, t is in seconds, and $k = 1$, then v is in volts.

(3) *Circuit interpretation.* To derive the circuit relationship between voltage and current in the system described in (2), we begin with Faraday's law,

$$v = \frac{d\psi}{dt} \tag{1-33}$$

or the equivalent integral form

$$\psi = \int v \, dt \tag{1-34}$$

Note, incidentally, the similarity of this expression and the one for charge in terms of current,

$$q = \int i \, dt \tag{1-35}$$

We see that ψ is to voltage as charge is to current, by comparing the two equations. Now flux linkages are related to the magnetic field by Eq. 1-31, and in turn, the magnetic field density is related to the current by Ampere's law, Eq. 1-28. Making these substitutions, with the assumption that i can be removed from the integral,* we have

$$\psi = \left[N \int \left(\int \frac{\mu \cos \alpha \, dl}{4\pi r^2} \right) dS \right] i \tag{1-36}$$

The integral term, which may be evaluated mathematically for simple geometries or may be found by measuring ψ and i, is defined as the *inductance parameter* (or the coefficient of inductance). If ψ and i refer to the same physical system, the parameter is defined as self-inductance, symbolized by the letter L as

$$\psi = Li \tag{1-37}$$

However, if a current i_1 produces flux linkages ψ_2 in another circuit, the parameter is one of *mutual inductance*, and the letter symbol is changed to M as

$$\psi_2 = M_{21} i_1 \tag{1-38}$$

(Again, note the similarity of these equations and the relationship $q = Cv$.) Substituting Eq. 1-37 into Faraday's law gives an equation relating voltage and current in a magnetic circuit,

$$v = \frac{d}{dt} (Li) \tag{1-39}$$

(where M replaces L in appropriate cases). If the inductance does not vary with time, Eq. 1-39 becomes

$$v = L \frac{di}{dt} \tag{1-40}$$

* If the magnetically coupled system is nonlinear, containing some saturating medium, we may say that the flux linkages in circuit k is a function of the currents in all other linked circuits,

$$\psi_k = \psi_k(i_1, i_2, i_3, \ldots, i_n)$$

By Eq. 1-32, the voltage in circuit k is given by Faraday's law as

$$v_k = \frac{d\psi_k}{dt} = \frac{\partial \psi_k}{\partial i_1} \frac{di_1}{dt} + \frac{\partial \psi_k}{\partial i_2} \frac{di_2}{dt} + \cdots + \frac{\partial \psi_k}{\partial i_k} \frac{di_k}{dt} + \cdots + \frac{\partial \psi_k}{di_n} \frac{di_n}{dt}$$

Each partial derivative term is evaluated with all other currents held constant. These terms may be defined as coefficients of inductance so that the voltage becomes

$$v_k = M_{k1} \frac{di_1}{dt} + M_{k2} \frac{di_2}{dt} + \cdots + L_{kk} \frac{di_k}{dt} + \cdots + M_{kn} \frac{di_n}{dt}$$

where M is used for mutual inductance and L for self-inductance. When a system is linear, this equation reduces to one which will later be written as Eq. 1-51.

Equation 1-39 can be integrated to give

$$i = \frac{1}{L}\int v\,dt \tag{1-41}$$

The quantity $(1/L)$ is sometimes symbolized by the upper case Greek letter gamma. The henry (after the American scientist Joseph Henry) is the mks unit for inductance.

In the case of the capacitive system, we found that charge and the product Cv could not change instantaneously. We might be led to suspect that there is a similar relationship for an inductive system in view of the analogies that have been pointed out. Indeed there is such a relationship, which may be found with the help of Eq. 1-34, in definite integral form.

$$\psi = \psi_0 + \int_0^t v\,dt \tag{1-42}$$

From arguments given in the last section about capacitance, the integral in this equation has zero value for $t = 0$. Thus, in a system altered instantaneously—say by the closing of a switch—the flux linkages must be the same before and after the system is altered, but only for a very small interval of time. In terms of Eq. 1-42,

$$\psi = \psi_0 = \text{a constant} \tag{1-43}$$

which is to say that the flux linkages cannot be changed instantaneously in a given system. This conclusion is described as the *principle of constant flux linkages.* If we let the subscript 1 refer to the time just before the system is altered and 2 refer to the same system after it is altered, our statements can be summarized by the equations

$$\psi_1 = \psi_2 \quad \text{or} \quad L_1 i_1 = L_2 i_2 \tag{1-44}$$

The principle of constant flux linkages is similar to the principle of conservation of momentum in mechanics. The analogy is helpful since it is sometimes easier to visualize changes in a mechanical system than in an electric circuit. Newton's force law is

$$F = \frac{d}{dt}Mv \tag{1-45}$$

where F is force, M is mass, and v is velocity. The product Mv is known as *momentum;* the momentum of a system cannot change instantaneously. In a system such as a rocket where mass is lost as a function of time, velocity must change in such a way that momentum remains constant. We see that there are a number of analogous conservation laws:

(1) The conservation of charge:

$$q_1 = q_2 \quad \text{and} \quad C_1v_1 = C_2v_2$$

(2) The conservation of flux linkages:

$$\psi_1 = \psi_2 \quad \text{and} \quad L_1i_1 = L_2i_2$$

(3) The conservation of momentum:

$$p_1 = p_2 \quad \text{and} \quad M_1v_1 = M_2v_2$$

When inductance remains constant, an important specialization of the principle of constant flux linkages results. *In a fixed inductive system, the current cannot change instantaneously.*

Example 2

In a certain inductive system, the current waveform shown in Fig. 1-10 exists. We are required to find the voltage that produces this

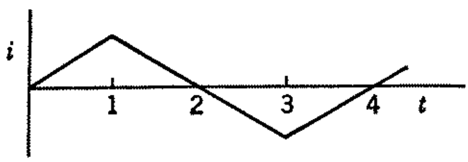


Fig. 1-10. Current waveform.

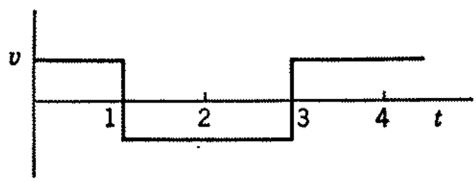


Fig. 1-11. Voltage waveform.

current waveform and the associated charge, both as functions of time. We will assume that L remains constant. The relationship $v = L(di/dt)$ indicates the voltage can be found by differentiation of the current and multiplication by a constant. The result is shown in Fig. 1-11. Charge may be found by integration of the current to give the result shown in Fig. 1-12.

It is important that we be able to apply the concept of inductance to several systems which are magnetically coupled. A set of three coupled

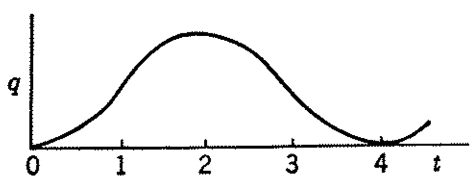


Fig. 1-12. Charge waveform.

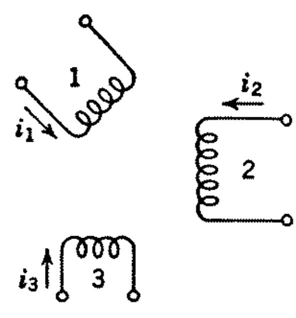


Fig. 1-13. Set of magnetically coupled coils.

coils is shown in Fig. 1-13. To simplify the system for the moment, let i_2 and i_3 be zero and consider the effect of the current i_1. The current i_1 produces ψ_1 flux linkages, found from Eq. 1-37 to be

$$\psi_1 = L_1i_1 \qquad (i_2 = i_3 = 0) \tag{1-46}$$

where L_1 is the self-inductance parameter (usually called just the inductance). In each other circuit, i_1 will produce some number of flux linkages by the proportionality of the mutual inductance parameter. For the particular system under study,

$$\psi_2 = M_{21} i_1 \quad \text{and} \quad \psi_3 = M_{31} i_1 \qquad (i_2 = i_3 = 0) \tag{1-47}$$

The order of subscripts for M requires some further attention. From the two equations, it should be clear that the first subscript refers to the flux linkages and the second to the current. This particular convention is chosen to give a desired symmetry to the general equations, our next topic of study. A crutch for remembering this particular convention is that the subscripts are in the order "effect, cause," if we assume for our conceptual scheme that current produces flux.

In the general case, there will be sources or loads connected to each of the coils shown in Fig. 1-13 and no current will be zero. We will assume for the time being that the current directions and winding senses of the coils are such that all flux linkages are *additive*, postponing the more general case for Chapter 2. The total flux linkages in coil 1 will be made up of flux linkages produced by the current in coil 1 *plus* flux linkages produced by currents i_2 and i_3. In equation form,

$$\psi_1 = L_1 i_1 + M_{12} i_2 + M_{13} i_3 \tag{1-48}$$

and similarly for the other two coils,

$$\psi_2 = M_{21} i_1 + L_2 i_2 + M_{23} i_3 \tag{1-49}$$

$$\psi_3 = M_{31} i_1 + M_{32} i_2 + L_3 i_3 \tag{1-50}$$

The symmetry discussed in the preceding paragraph is now apparent. The mutual inductance coefficients have subscripts designating row and column in the above array of equations.

We are interested in flux linkages only as a stepping stone to voltage. The voltage induced in each coil is given by Faraday's law as the time rate of change of flux linkages. If the inductance parameters are constant, these voltages are readily found by differentiation to be

$$v_1 = L_1 \frac{di_1}{dt} + M_{12} \frac{di_2}{dt} + M_{13} \frac{di_3}{dt} \tag{1-51}$$

$$v_2 = M_{21} \frac{di_1}{dt} + L_2 \frac{di_2}{dt} + M_{23} \frac{di_3}{dt} \tag{1-52}$$

$$v_3 = M_{31} \frac{di_1}{dt} + M_{32} \frac{di_2}{dt} + L_3 \frac{di_3}{dt} \tag{1-53}$$

In Chapters 2 and 3, we will consider the conditions under which some terms in these equations will be negative.

1-8. The resistance parameter

(1) *Physical phenomenon.* The passage of electrons through a material is not accomplished without collisions of the electrons with other atomic particles. Moreover, these collisions are not elastic, and energy is lost in each collision. This loss in energy per unit charge is interpreted as a drop in potential across the material. The amount of energy lost by the electrons is related to the physical properties of a particular substance.

(2) *Field interpretation.* The German physicist Georg Simon Ohm found experimentally that there is a relationship between the current in a substance and the potential drop. In terms of the field concept, the change in energy per unit of charge causes a change in the force per unit charge—or electric field. This effect may be interpreted in terms of a field in the direction of current through the conducting substance. Ohm's experiment may be stated in terms of this field and the current per unit cross-sectional area as

$$J = \sigma E \tag{1-54}$$

where, in mks units, J is the current density in amperes per square meter, E is the field along the conducting substance in volts per meter, and σ is the conductivity of the substance, which is a constant for each particular material.*

(3) *Circuit interpretation.* If the substance which carries the current has an idealized geometry, as that shown in Fig. 1-14(b), it is possible

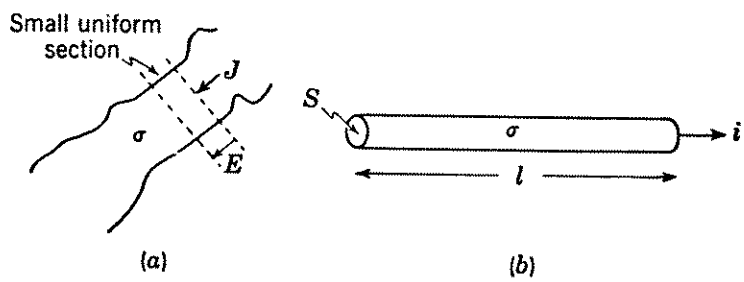


Fig. 1-14. Conductors illustrating Ohm's law.

to reduce the field form of Ohm's law to relate current and voltage. If the cross section of the conductor is *uniform*, the current and current density are related by the equation

$$i = \int J \cos \theta \, dS = JS \tag{1-55}$$

where S is the cross-sectional area. For the same simple geometry, the electric field is uniform and directed along the length of the wire;

* Strictly speaking, Eq. 1-54 is a special case valid only for isotropic substances. Similarly σ is independent of the magnitude of E only for linear substances.

that is,

$$v = El \tag{1-56}$$

as a special case of the more general relationship

$$v = \int E \cos \theta \, dl \tag{1-57}$$

Substituting Eqs. 1-55 and 1-56 into the field form of Ohm's law, Eq. 1-54, gives

$$v = \left(\frac{l}{\sigma S}\right) i \tag{1-58}$$

The quantity $(l/\sigma S)$, which is a constant for constant geometry of the conductor, is given the name the *resistance parameter*—or simply the *resistance*, and is symbolized by the letter R. For geometries other than the simple one of Fig. 1-14(b), computation of the coefficient relating current and voltage for a substance will be more difficult. However, measurement of current and voltage can establish the value of the resistance parameter and by-pass the computation problem. Ohm's law may be written

$$v = Ri \tag{1-59}$$

or, in terms of charge,

$$v = R \frac{dq}{dt} \tag{1-60}$$

The equation $v = Ri$ is sometimes written in the form

$$i = Gv \tag{1-61}$$

where $G = 1/R$ is known as the *conductance*. In the mks system, the unit for resistance is the *ohm* and for conductance is the *mho*.

As well known as Ohm's law is (school children are taught the law and remember it by association with "Vermont = Rhode Island"), Ohm was ridiculed by his fellow scientists when he first announced his law in 1826, and it was some 30 years before his ideas were finally accepted. We must remember, of course, that the concepts of current and voltage were not well understood in his day, the first distinction between the two quantities having been made by Ampere in 1820. Reading newspaper statements such as "10,000 volts passed through his body" can convince one that the distinction is not well understood by laity today.

1-9. Approximation of a system as a circuit

We have discussed the manner in which three electrical phenomena observed experimentally can be described in terms of circuit parameters. A problem that we must eventually face in making use of the circuit concept is that of representing a physical system in terms of these parameters. For example, can we draw a circuit that will represent an electric motor, a piezoelectric crystal, a coil of wire, a transmission line, or an antenna, to name but a few systems?

Suppose we examine some arbitrary physical system, looking for portions of the system to be replaced by equivalent parameters. Possibly the resistive effects would be most easily recognized. A part of the system made of material of high resistivity, with small cross-sectional area and appreciable length, would be recognized as equivalent to large resistance and could easily be distinguished from another part of the system of small resistance. We have found that there is a capacitive effect between any two parts of a system. If the two parts constitute a system capable of concentration of charge, producing a high electric field—say, large area for charge storage and small distance from part to part—the capacitance of that portion of the system is large. Finally, an inductive effect is associated with every current-carrying conductor, and an effect of mutual inductance between every pair of conductors at least one of which is carrying current. If the conductors are located in space in such a way that the magnetic fields reinforce each other, then the inductance, self or mutual, of that portion of the system is large.

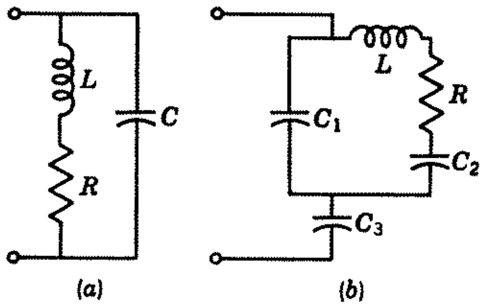


Fig. 1-15. One form of equivalent circuit for (a) an end-excited antenna, and (b) a piezoelectric crystal.

So much for large effects. What about smaller or secondary effects that can be recognized in much the same manner? Just how many effects must be taken into account in representing a system by equivalent parameters?

We can answer our questions only by asking another: just how good do we expect the results to be? The accuracy of our results will be determined by how many separate electrical effects we can take into account by a parameter. We must stop somewhere. We must, at some point, make an *approximation*.

Approximation requires engineering judgment. An approximation which is valid in one case will not be in another. In many practical cases, the resistance and inductance of connecting wires are so small that they may be neglected. Likewise, in most cases of commercial capacitors, the inductive and resistive parameters may be ignored. Much less frequently the resistance and capacitance of coils can be neglected.

In the discussions to follow in other chapters, we will assume that when a schematic of a system is given, all significant parameters have been taken into account. Engineering judgment has been exercised by the individual who made up the problem. But when the student finally applies the techniques of analysis to a problem that *he* makes up, these questions associated with approximation must be answered. It is difficult to write answers to such questions in textbooks; experience is usually the best teacher.

Approximation is not unique to circuit analysis by any means. In solving problems by computing the electric and magnetic fields for all positions in space, there will assuredly be approximations, either in representing the physical system by mathematical equations or in solving the equations. Approximation and analysis are bound together. To ignore the problem of approximation is to lack understanding of the results of analysis.

In many cases, we do not start with an unknown system to be represented by a circuit, but instead with commercial components in combination forming a circuit. A component labeled *inductor*, however, will not behave as a pure inductance. It will, under some circumstances, exhibit capacitive and resistive effects. Such unwanted effects are commonly distinguished by the name *parasitic*. The decision of which effects must be taken into account involves the same engineering judgment as discussed earlier. The parasitic effects can be ignored only as long as the approximation is useful.

In all cases we have assumed that the magnetic and electric fields are isolated and that there is no interaction between the two fields. If there is such an interaction, part of the energy is lost by *radiation.* This will be discussed in the next section.

1-10. Other approximations in circuit representation

In arriving at equations for the circuit parameters, Eqs. 1-11, 1-36, and 1-58, it was necessary to make simplifying approximations: (1) that the charge did not vary with dimensions, and (2) that the current varied with neither the length of the conductor nor the cross-sectional area. If these assumptions do not hold, the values for the parameters are different and difficult to compute.

To illustrate how current and charge might vary with space, suppose that the current is made to flow for but a brief interval of time, and that this pulsed flow is repeated at a periodic rate, a very large number of times each second. Under such conditions, the current and the charge will not be uniform throughout the system. We can imagine some portions of the system with charge and other portions without charge. This being the case, the general expressions must be used in evaluating the capacitance, inductance, and resistance parameters. These new parameter values, computed or measured, will be different from those found with uniform current and charge in the system. Must the parameters of a system be computed for every different current?

The answer to this question is, again, a practical one of engineering judgment. Certainly, there will be conditions requiring some effective value of the parameters—computed for a particular time waveform—to be used. But in many cases, the *approximation* that parameter values are equal to those found for nonvarying or static conditions gives usable results This approximation is strictly valid only in the cases in which the variation of current and charge is slow, the so-called *quasi-stationary state*. We will assume that we are operating in this state in chapters to follow. We thus assume *constant parameters* for changing variations of current and charge.

We further assume that the parameters are constant with the variation of the *magnitude* of charge and current. This is a good approximation for most elements in their nominal operating range. A system composed of such elements is said to be *linear*. We will assume that all systems to be considered (unless otherwise specified) are linear. We thereby exclude *nonlinear* elements and systems. Some systems containing dielectrics change capacitance with the quantity of charge in the system. When iron is used in a magnetic system, the flux produced is not linearly related to current because of saturation. Such resistive materials as the carbon filament in a lamp bulb change resistance as a function of magnitude of current. It should be noted, however, that some nonlinear systems can be considered linear under certain conditions. Vacuum tubes are nonlinear, but for certain analyses may be considered linear over a restricted range of operation.

Besides the assumption of linearity we will include the requirement that all elements in a system be *bilateral*. In a bilateral system, the same relationship between current and voltage exists for current flowing in either direction. In contrast, a *unilateral* system has different laws relating current and voltage for the two possible directions of current. Examples of unilateral elements are vacuum diodes, germanium diodes, crystal detectors, selenium rectifiers, etc.

Many electric systems are physically distributed in space. A transmission line, for example, may extend for hundreds of miles. When a source of energy is connected to the transmission line, energy is transported at nearly the velocity of light. Because of this finite velocity, all electrical effects do not take place at the same instant of time. This being the case, the restrictions discussed earlier apply in the computation of the circuit parameters. When a system is so concentrated in space that the assumption of simultaneous actions through that system is a good approximation, the system is said to be *lumped*. We will consider only lumped systems.

Our circuit approach to the approximation of a system has obscured an effect usually described in terms of the interaction of electric and magnetic fields. As an approximation, we have assumed that the magnetic field is associated only with an inductive system and that the electric field is associated only with a capacitive system. Fields cannot actually be so lumped. The consequence of interaction of the fields is the *radiation* of electromagnetic energy. Open a switch in an inductive system, and the effects will be observed as a noise in nearby radio receivers. Similarly, the ignition spark of an automobile may affect nearby television receivers. Under many conditions, however, the amount of energy lost by radiation is small, and as an approximation can be ignored. We will make this approximation.

The systems we shall study will thus be *lumped*, *linear*, and *bilateral* and will have negligible radiation.

Resistive, capacitive, and inductive elements are identified as *passive elements*. Sources of electric energy are identified as *active elements*. The physical elements themselves are distinguished by different names as *resistors*, *inductors*, and *capacitors*.

1-11. Energy and power

Energy and power are given in terms of voltage and current by Eqs. 1-5 and 1-6, which are

$$p = vi \quad \text{and} \quad w = \int vi\,dt$$

In an inductive system, energy has the value (see Prob. 1-8)

$$W_L = \tfrac{1}{2}Li^2 = \tfrac{1}{2}\frac{\psi^2}{L} \qquad \text{joules} \tag{1-62}$$

and is spoken of as "stored in the magnetic field." In a capacitive system, the energy is given by the relationship (see Prob. 1-9)

$$W_C = \tfrac{1}{2}Cv^2 = \tfrac{1}{2}Sq^2 \qquad \text{joules} \tag{1-63}$$

which is spoken of as "stored in the electric field." Current in a resistor causes energy to be transformed into heat or light. The amount of energy is

$$W_R = R\int i^2\,dt \qquad \text{joules} \tag{1-64}$$

The equivalence of this energy to mechanical energy was first demonstrated by Joule. The inductor and capacitor are *energy storage elements*, while the resistor is an *energy sink*.

Energy is a scalar quantity, always positive. The sum of the energy in a given system can be found by algebraic summation. The power in a system is given in terms of energy by the relationship

$$P = \frac{dW}{dt} = P_r' + \frac{d}{dt}(W_L' + W_C') \tag{1-65}$$

where P_r' is the total power of all resistive elements given as

$$P_r' = \frac{dW_R'}{dt} = \sum_{j=1}^{n} R_j i_j^2 \tag{1-66}$$

and W_L' is the total energy stored in all inductive elements,

$$W_L' = \tfrac{1}{2} \sum_{j=1}^{m} L_j i_j^2 \tag{1-67}$$

and W_C' is the total energy stored in all capacitive elements,

$$W_C' = \tfrac{1}{2} \sum_{j=1}^{p} C_j v_j^2 \tag{1-68}$$

where n, m, and p are the total number of elements of each of the three kinds.

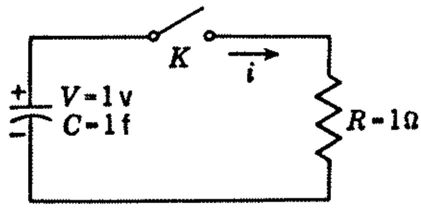
FURTHER READING

The student interested in further reading on the subjects of this chapter should consult *Elementary Electric-Circuit Theory* (McGraw-Hill Book Co., New York, 1945) by Richard H. Frazier, pp. 17-41. More advanced treatments of these concepts are given in *Electric Circuits* (John Wiley & Sons, Inc., New York, 1940) by the MIT Electrical Engineering Staff, pp. 1–8, and in *Linear Transient Analysis* (John Wiley & Sons, Inc., New York, 1954) by Ernst Weber, pp. 1–14. A discussion of physical systems in general is to be found in Chaps. 1 and 2 of *Response of Physical Systems* (John Wiley & Sons, Inc., New York, 1950) by John D. Trimmer. Students interested in further study of electric and magnetic fields as basic concepts should read *The Fundamentals of Electro-Magnetism* (The Macmillan Company, New York, 1939) by Geoffrey Cullwick, starting on p. 1. Maxwell's original writings are found in many libraries under the title, *A Treatise on Electricity and Magnetism* (Oxford Press, New York, 1892).

PROBLEMS

1-1. A solid copper sphere 10 cm in diameter is deprived of 10^{13} electrons by a charging scheme. (a) What is the charge of the sphere in the inductor? *Answer:* 0.23 watt. (h) At what rate is energy being dissipated as heat? (i) At what rate is the battery supplying energy?

1-16. In the circuit shown below, the capacitor is charged to a voltage of 1 volt, and at $t = 0$ the switch K is closed. The current in the circuit is known to be of the form $i(t) = e^{-t}$ amp, $(t > 0)$. At a certain time the current has a value of 0.37 amp. (a) At what rate is the voltage



Prob. 1-16.

across the capacitor changing? (b) What is the value of the charge on the capacitor? (c) What is the time rate of change of the product Cv? (d) What is the voltage across the capacitor? *Answer:* 0.37 volt. (e) How much energy is stored in the electric field of the capacitor? (f) What is the voltage across the resistor? (g) At what rate is energy being taken from the electric field of the capacitor? *Answer:* 0.137 watt. (h) At what rate is energy being dissipated as heat?

1-8. From the defining equation for energy, $W = \int vi\, dt$ show that, for the inductance, $W_L = \frac{1}{2}Li^2$ and $W_L = \frac{1}{2}\psi^2/L$.

1-9. From the equation for energy in Prob. 1-8, show that for the capacitance, $W_C = \frac{1}{2}Cv^2$ and $W_C = \frac{1}{2}Sq^2$.

1-10. Assume that the inductance parameter is defined as the constant relating stored energy and the current squared by the equation $W_L = \frac{1}{2}Li^2$. Making use of the relationship $p = vi$, show that for constant inductance the voltage across the inductor is $v_L = L(di/dt)$.

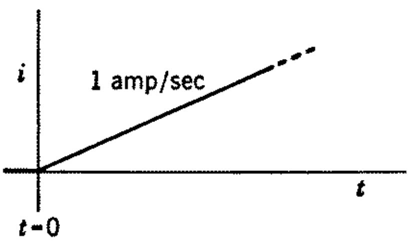
1-11. Carry out a similar derivation to the one suggested in Prob. 1-10 starting with energy for a capacitive system, $W_C = \frac{1}{2}Sq^2$ to show that for constant S,

$$v_C = Sq = S \int i\, dt$$

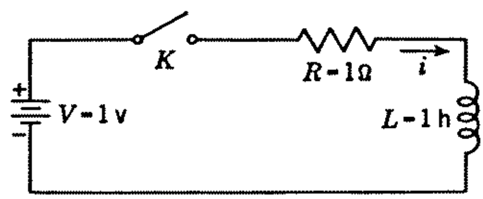
1-12. Show that the following quantities all have the dimension of time: (a) RC; (b) L/R; (c) $\sqrt{LC}$.

1-13. Show that (a) R^2C has the dimension of inductance, (b) $\sqrt{L/C}$ has the dimension of resistance, (c) L/R^2 has the dimension of capacitance.

1-14. The current in a 1-henry inductor follows the variation shown in the accompanying figure. The current increases from $t = 0$ at the rate of 1 amp/sec (for several seconds, at least). Find: (a) the flux linkages in the system after 1 sec, (b) the time rate of change of flux



Prob. 1-14.



Prob. 1-15.

linkages in the system after 2 sec, (c) the quantity of charge having passed through the inductor after 1 sec. *Answer:* (a) 1 weber-turn, (b) 1 weber-turn/sec, (c) 0.5 coulomb.

1-15. In the circuit shown above, the switch K is closed at $t = 0$ (the reference time). The current flowing in the circuit is given by the equation $i(t) = (1 - e^{-t})$ amp, $t > 0$. At a certain time the current has a value of 0.63 amp. (a) At what rate is the current changing? (b) What is the value of the total flux linkages? (c) What is the rate of change of flux linkages? (d) What is the voltage across the inductor? *Answer:* 0.37 volt. (e) How much energy is stored in the magnetic field of the inductor? *Answer:* 0.20 joule. (f) What is the voltage across the resistor? (g) At what rate is energy being stored in the magnetic field of

The photoelectric cell and the pentode vacuum tube amplifier are examples of practical generators that can be approximated as ideal current sources with series or parallel passive elements. The current source is also represented by the symbol of a circle but with an associated arrow rather than polarity markings indicating the positive direction of current flow as shown in Fig. 2-1(*c*).

For both sources, the associated symbol of a lower case letter v or i infers a time-varying source, while the upper case letter V or I is used to denote a time-invariant source. The approximate time variation of output is sometimes sketched within the circle. For example, the symbol $\sim$ placed in the circle indicates a source of sinusoidal voltage or current.

2-2. Current and voltage conventions

A voltage source causes current to flow within the source in the direction from the negative to the positive terminal—or out of the positive terminal and into the negative terminal. This particular convention follows a decision made by Benjamin Franklin in 1752. Franklin's choice was made before electricity was identified with the electron, before the electron or the nature of charge were known. Actually, electrons flow from the negative terminal to the positive terminal, which is in the opposite direction to that established by Franklin. To distinguish the two conventions, the flow of electrons is termed *electron current* and current assumed positive in the direction of Franklin's convention is called *conventional current* (or simply current, since this is the current we will use).

If the negative terminal is used as a reference in measuring the potential of the positive terminal of a potential source, that voltage is considered *positive* and is spoken of as a *voltage rise*. Conversely, if the positive terminal is considered to be the reference in measuring the potential of the negative terminal of the voltage source, the voltage is considered negative and is spoken of as a *voltage drop*.

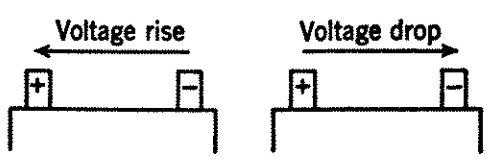


Fig. 2-2. Sign convention for voltage sources.

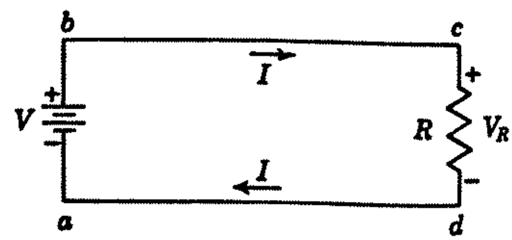


Fig. 2-3. Direction of current and voltage polarities.

In terms of the simple circuit of Fig. 2-3, the voltage source causes current to flow from a to b and around the circuit a-b-c-d-a. Current flowing in the passive resistive element is identified with energy loss and

CHAPTER 2

CONVENTIONS FOR DESCRIBING NETWORKS

2-1. Active element conventions

Active circuit elements are classified by their voltage-current characteristics. Most practical generators maintain approximately constant terminal voltage with increasing load current. Still other types of sources of electric energy maintain approximately constant output of current with increasing terminal voltage. Rather than take actual voltage-current relationships into account, practical energy sources are approximated as either *ideal voltage sources* or *ideal current sources.* These ideal sources are defined as having the following properties:

The Ideal Voltage Source. The ideal voltage source generates voltage with a given time variation. Neither voltage magnitude nor time variation changes with magnitude of output current. Thus the terminal voltage is assumed to be maintained under all conditions from open circuit to short circuit. If in some manner the terminal voltage is made equal to zero, the source behaves as a short circuit. The ideal voltage source has no resistive, inductive, or capacitive effects. Most actual generators may be approximated as an ideal voltage source in series with a resistor, and in some cases, an inductor. The ideal voltage source is represented by the symbol of a circle as shown in Fig. 2-1(a). One exception to this practice is the symbol for a battery shown in Fig. 2-1(b). The polarity marks + and − denote positive and negative terminals of the source.

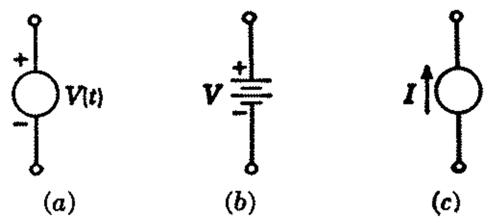


Fig. 2-1. Symbols for active circuit elements.

The Ideal Current Source. The ideal current source generates current with a given time variation. Neither current magnitude nor time variation changes with load. This output current is maintained for any load including zero resistance and infinite resistance.* If the output current of the ideal source is adjusted to be zero, the source is equivalent to an open circuit. As in the case of the voltage source, the ideal current source has no resistive, inductive, or capacitive effects.

* Of course, this property of the ideal source to pump amperes into an open circuit is a poor approximation for actual sources.

The junction formed by two or more elements being connected together is given the name *node*. Elements in series such that identically the same current flows in them form a *branch*. A branch may include active elements. A *network* or *circuit* is formed by interconnection of a number of branches or by coupling a number of separate parts together. We will use the words *network* and *circuit* interchangeably, except that the network is usually more complex, involving more elements than the circuit. A *loop* or *mesh* is a closed contour drawn on the schematic, around one or more window panes of the graph of Fig. 2-4, for example. Any two nodes in a network may be considered a *node pair*. Parts of the network not directly connected by wires but magnetically coupled are called *separate parts* of the network.

Our objective in circuit analysis is to find the currents in the different branches and the voltages at the different nodes. Two quantities of importance in this analysis are the number of *independent loops* and

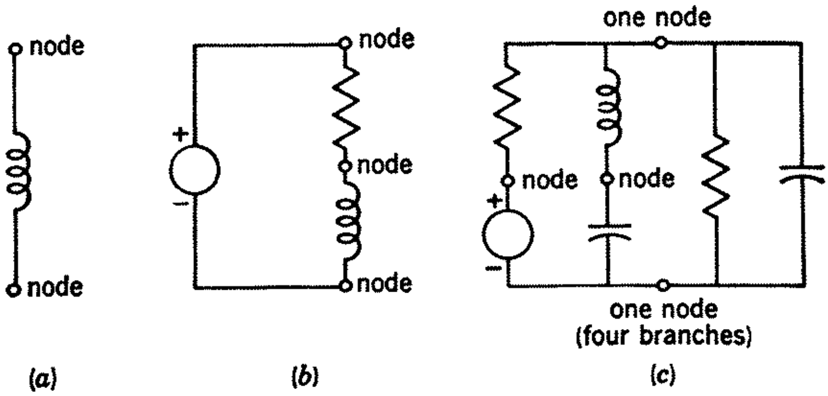


Fig. 2-5. Identification of nodes.

the number of *independent node pairs*. The meaning inferred by the word independent will be discussed in the next chapter in more detail. Briefly, there are as many unknown currents in a system as there are independent loops, and as many unknown voltages as there are independent node pairs.

Let E be the number of elements in the network (counting both active and passive elements, but not mutual inductance). The number of branches in the network is designated B. Quantities relating to nodes will be identified by N_t for the number of nodes, J for the number of different node pairs, and N for the number of independent node pairs. Also, L is the number of independent loops, and finally, S is the number of separate parts of the network.

The number of different node pairs is given by the topological equation

$$J = \tfrac{1}{2}N_t(N_t - 1) \qquad (2\text{-}1)$$

We are usually not interested in all combinations of nodes, but in so there is a drop in potential from c to d. Voltage is a scalar quantity defined as the quotient of the two scalar quantities work and charge. Voltage drops and voltage rises are distinguished by a sign, positive for voltage rise and negative for voltage drop. For this polarity sign to have meaning, the voltage reference must be given or inferred. A certain point in a circuit may have a potential of +50 volts with respect to point a and −30 volts with respect to point b. Should the potential of a point be given without reference, it is assumed to be with respect to a point of zero potential which will be called the *ground* or the *datum node*.

2-3. Network topology (or geometry)

The two-dimensional graph shown in Fig. 2-4 is made up of circles with interconnecting lines. Suppose that we wish to make a study of such graphs, say the relationship of the number of circles and lines, with the following rules imposed: (1) all lines must terminate on circles, (2) at least two lines must join every circle, and (3) the graph will be in two dimensions; lines will not cross lines. By definition, all closed contours having no lines within will be called window panes.

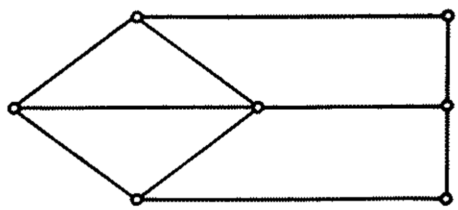


Fig. 2-4. A graph.

Now in any given graph, the number of lines, circles, and window panes are not independent of each other. For example, if the number of circles remains constant and the number of lines is increased by one, there will be one more window pane. Similarly, if the number of window panes is maintained constant and one more circle is added, the number of lines must increase by one, a line having been divided in two by the added circle. There must exist some general law relating the three quantities we have studied.

Our discussion has been a homely example of an elegant branch of mathematics known as *topology*. The general law just mentioned is known from studies in topology. How can we exploit these topological laws or facts in circuit analysis?

By approximating a physical system by ideal circuit elements, as discussed in the first chapter, we have eliminated consideration of three-dimensional systems in favor of a system of interconnected elements usually described as a *wiring diagram* or a *schematic*. The schematic is equivalent to the topological graph of Fig. 2-4 with elements replacing lines. Thus the laws of topology are directly applicable to network schematics. Before giving these laws, we will define terms used in network topology.

discussed in the next chapter). It is, however, not necessary that the loops be chosen in this manner, but there is the risk that the loops will not be independent.* In simple networks, the number of independent loops is equal to the number of "window panes" of our earlier discussion and so can be determined at a glance.

Example 1

In the network shown in Fig. 2-7, there are five nodes, seven elements, and one part. Hence the number of independent loops is three,

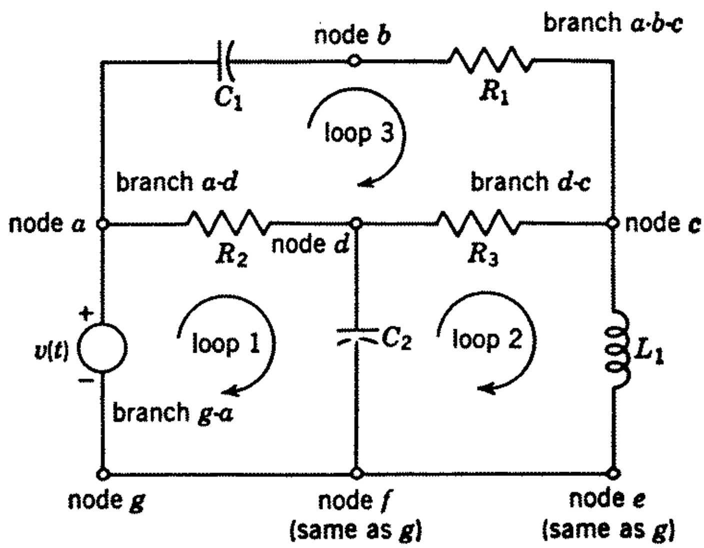


Fig. 2-7. Network of Example 1.

since

$$L = E - N_t + S = 7 - 5 + 1 = 3$$

and the number of independent node pairs is four, since

$$N = N_t - S = 5 - 1 = 4$$

The loops may be assigned as shown on the figure, following paths *a-d-f-g-a*, *d-c-e-f-d*, and *a-b-c-d-a*. If node *g* is selected as the datum node, a suitable choice of the independent node pairs is *a-g*, *b-g*, *c-g*, and *d-g*.

Example 2

Examination of Fig. 2-8 shows that $E = 6$ (M does not count, of course), $N_t = 6$, and $S = 2$. It follows that

$$L = E - N_t + S = 6 - 6 + 2 = 2 \text{ independent loops}$$

$$N = N_t - S = 6 - 2 = 4 \text{ independent node pairs}$$

* The test to determine independence requires that the system determinant, to be discussed in Chap. 3, be nonzero. It is usually better to follow the von Helmholtz rule than to experiment.

the number of independent node pairs, which is equal to the number of unknown voltages, given as

$$N = N_t - S \tag{2-2}$$

or $N = N_t - 1$ for a network with only one part.

The number of independent loops is given in terms of the number of elements and number of independent node pairs as

$$L = E - N \tag{2-3}$$

If we count only nodes at the ends of branches and let that number of branches be B', then the number of independent loops and branches are related as

$$L = E - B' \tag{2-4}$$

Finally, if Eq. 2-2 is substituted into Eq. 2-3, there results

$$L = E - N_t + S \tag{2-5}$$

From these equations, we may determine for a given network which quantity, L or N, is smaller, enabling us to decide which of the two possible approaches to analysis should be taken.

Once the number of independent node pairs is known, there remains the problem of selecting which N node pairs in the network will be used in analysis. Analysis is simplified if one node is used as one member of each of the N node pairs. It is conventional to select the node of zero potential as this common node and to designate it the *datum* or *reference* node. Usually, the negative terminal of one of the active sources is so selected.

There is a similar problem of choice in the case of assigning the independent loops once the number of such loops is known from Eq.

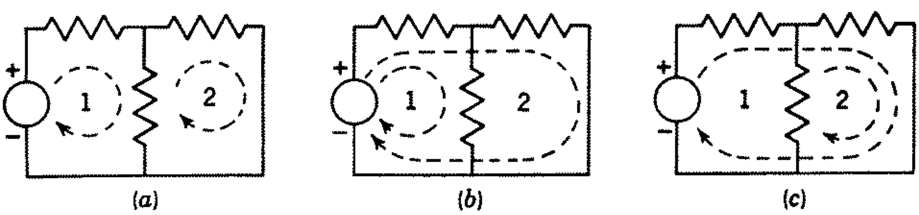


Fig. 2-6. Different independent loops in the same system.

2-5. The rule of von Helmholtz may be used to advantage in designating the paths of the loops. The rule requires that the loops be successively chosen such that each new loop includes at least one new branch not previously included in the path of a loop. If the loops of number required by Eq. 2-5 are so chosen, this process is sufficient to insure the independence of the loops (and of the voltage equations, as will be of the winding is shown marked with a *dot*. Let us assume that the current flows into this dot. We will outline, step-by-step, our conceptual scheme of what happens as a consequence of this current.

(1) Current in winding 1-1 causes a magnetic field ("action at a distance") to exist, which is concentrated along the axis of the coil. The magnitude of the field can be computed from Ampere's law

$$dB = \frac{\mu i_1 \, dl \cos \alpha}{4\pi r^2} \tag{2-6}$$

(These symbols are defined in Chapter 1 in Eq. 1-28.)

(2) There is a magnetic flux ϕ associated with the magnetic field having a value

$$\phi = \int_A B \cos \theta \, dA \tag{2-7}$$

and having a direction determined experimentally and given by the right-hand rule: if the thumb of the right-hand indicates the direction of current, the fingers wrap around the current-carrying conductor in the direction of flux. This flux is assumed confined to the magnetic core, which has the property of being a preferred path for the flux. Applying the right-hand rule, the flux is seen to have the direction indicated by the arrow (clockwise).

(3) Since winding 2-2 is on the same magnetic core as winding 1-1, the flux produced in winding 1-1 *links* winding 2-2. This linking flux can be described as ϕ_{21}, where the subscripts have the order "effect, cause." The number of flux linkages in winding 1-1 is

$$\psi_1 = N_1 \phi_{21} \tag{2-8}$$

In terms of Faraday's law ψ_1 can be computed from the voltage at terminal 1-1 as

$$\psi_1 = \int v_1 \, dt \tag{2-9}$$

Combining Eqs. 2-8 and 2-9 gives the value of flux in terms of the voltage v_1.

$$\phi_{21} = \frac{1}{N_1} \int v_1 \, dt \tag{2-10}$$

(4) Because ϕ_{21} is changing with time, a voltage is induced in winding 2-2 according to Faraday's law. The flux linkages in winding 2-2 are

$$\psi_2 = N_2 \phi_{21} \tag{2-11}$$

and v_2 has the magnitude

$$v_2 = \frac{d\psi_2}{dt} \tag{2-12}$$

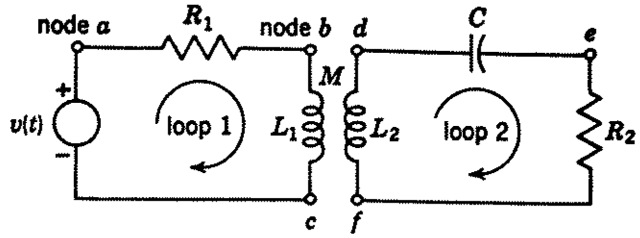


Fig. 2-8. Network of Example 2.

2-4. The dot convention for coupled circuits

When the magnetic field produced by a changing current flowing in one coil induces a voltage in other coils, the coils are said to be *coupled*, and the windings constitute a *transformer*. If the details of transformer construction are known, then for a current changing in one coil, it is possible to compute the magnitude and direction of the voltages induced in all other windings. The necessity for cumbersome blueprints showing construction is eliminated by two characterizing factors. The value of the coefficient of mutual inductance, M (discussed in Chapter 1) is equivalent to details of construction in computing *magnitude* of induced voltage. Most manufacturers mark one end of each transformer winding with a dot (or some such symbol). The dot is equivalent to details of construction as far as *voltage direction* is concerned. In this section, we will discuss the meaning of dot markings, how they are experimentally established, and their significance in circuit analysis.

Two windings are shown on a magnetic core in Fig. 2-9. In this figure, the winding sense is indicated for two windings, winding 1-1

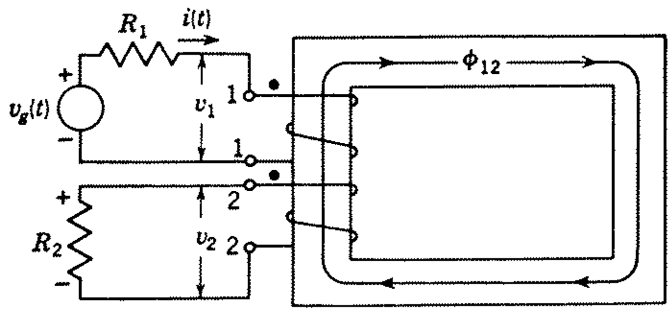


Fig. 2-9. A two-winding magnetic circuit used to establish meaning of the dot convention.

(which might be called the *primary* winding) and winding 2-2 (the *secondary* winding). A time-varying source of voltage, $v_g(t)$ is connected to winding 1-1 in series with resistor R_1. At a given instant, the voltage source has the polarity shown, and the current $i_1(t)$ is flowing in the direction shown by the arrow and is increasing with time. The + end minal of a battery, connecting the negative terminal to the remaining end of the winding. The end of winding 2-2 that momentarily goes positive, as measured with a voltmeter, is the terminal to be dotted in winding 2-2.

Of what value are the dots, which we can now establish, in circuit analysis? Figure 2-11 shows the transformer of Fig. 2-9, including

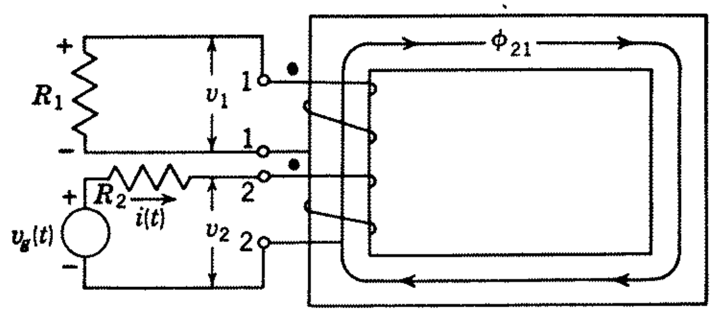


Fig. 2-11.

dots, with the generator and resistor load interchanged. The positive terminal of the voltage source is connected to the dotted end of winding 2-2. A step-by-step analysis of this transformer will show that an increasing current flowing into the dotted terminal of winding 2-2 causes the upper end of winding 1-1 to be positive and so to be the dotted terminal. We would expect, after all, that dots established from 1-1 to 2-2 should agree with those established from 2-2 to 1-1.

Now suppose that the voltage source of Fig. 2-11 has reverse polarity to that shown and that an increasing current flows *out of the dot.* Another step-by-step analysis or simply intuitive reasoning will show that the dotted terminal of winding 1-1 becomes negative under such conditions.

We conclude that, for a transformer with polarity markings (dots), current flowing *into* the dot on one winding induces a voltage in the second winding which is positive at the dotted terminal; conversely, current flowing *out* of a dotted terminal induces a voltage in the second winding which is positive at the undotted terminal. This important rule will be applied in Chapter 3 in formulating circuit equations.

Thus far our discussion has been limited to a transformer with two windings. In a system with several windings, the same type of analysis can be carried on for each pair of windings providing some variation in the form of the dots is employed (such as ● ■ ▲ ◆) to identify the relationship between each pair of windings. In Chapter 3, after the concept of assumed positive direction of current is introduced, it will be shown that the information given by the pair of dots can be given in the sign of the coefficient of mutual inductance. For a system with many windings, this scheme avoids the confusion of a large num-

In the discussion of Chapter 1, the coefficient of mutual inductance was introduced to relate flux linkages with current as $\psi = Mi$. For the system under study

$$\psi_2 = M_{21}i_1 \tag{2-13}$$

and Eq. 2-12 may be written in equivalent (but more useful) form as

$$v_2 = M_{21}\frac{di_1}{dt} \tag{2-14}$$

if M_{21} does not vary with time. Equation 2-14 tells us that a voltage is induced in winding 2-2 having a magnitude of M_{21} volts per unit time rate of change of current i_1. There remains the problem of the direction of this voltage.

(5) The direction of voltage in winding 2-2 can be found with the aid of a law given by the German physicist Lenz in 1834. In terms of the transformer, *Lenz's law* states that the voltage induced in a coil by a change of flux establishes a current in the coil in a direction to oppose the change in flux that produced the voltage. The flux ϕ_{21} is directed upward in Fig. 2-9 and is increasing. To produce a flux ϕ_{12} to oppose this increase in ϕ_{21} requires (by the right-hand rule) that the current flow in the direction shown by the arrow (right to left). Lenz's law is really an application of conservation of energy, since if i_2 produced a flux to aid ϕ_{21}, another increasing current would be induced in 1-1 and so on in a vicious cycle to produce infinite current.

Now that the direction of current in winding 2-2 is established, the top end of the winding is seen to be positive and so is marked with a dot. With a time-varying voltage, the dotted terminals are positive at the same time (and, of course, negative at the same time). This action is illustrated in Fig. 2-10. As shown, $v_g(t)$ increases from zero to a constant value at time t_1. The current i_1 and so flux ϕ_{21} increase with time as shown in (b). Note, incidentally, that Eq. 2-10 does not apply directly, since it gives ϕ_{21} in terms of v_1 rather than v_g. The induced voltage v_2 is proportional to the time rate of change of i_1 and so has the time variation shown in (c). This example suggests a simple experimental method for establishing the dotted ends of transformer windings. On the winding selected as 1-1, arbitrarily mark one end of the winding with a dot and to this terminal connect the positive ter-

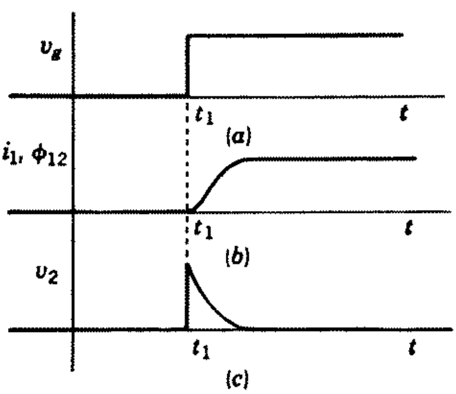
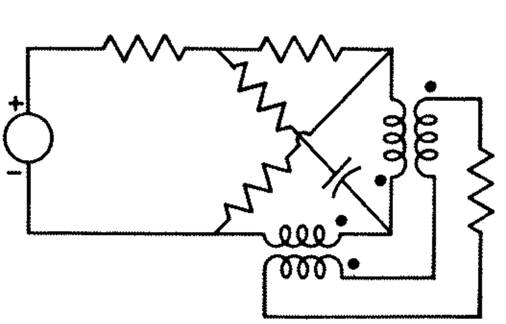
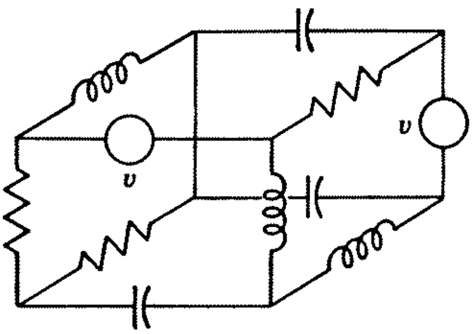


Fig. 2-10. Waveforms in the magnetic circuit of Fig. 2-9.

2-2. Repeat Prob. 2-1 for the network shown in the accompanying figure.



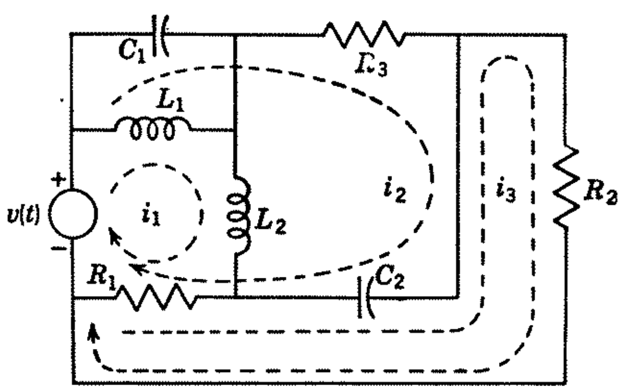
Prob. 2-2.



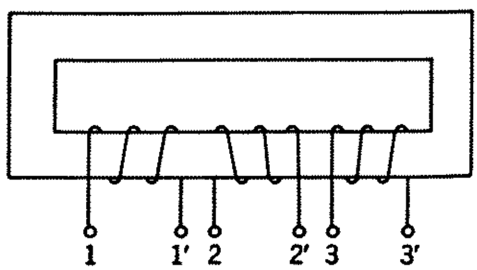
Prob. 2-3.

2-3. In the accompanying figure, a number of elements are arranged on the edges of a cube. For this network, determine L, the number of independent loops, and N, the number of independent node pairs.

2-4. In the network of the figure, the paths for three loops have been selected as shown. Are these three loops independent? Why? (What are the currents in L_1 and L_2 in terms of i_1?)



Prob. 2-4.

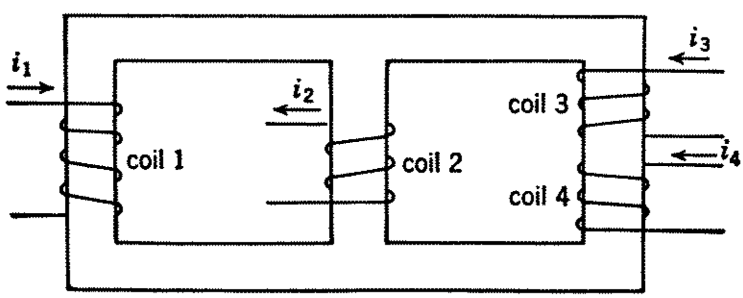


Prob. 2-5.

2-5. The magnetic system shown in the figure has three windings marked 1-1′, 2-2′, and 3-3′. Using three different forms of dots, establish polarity markings for these windings.

2-6. Place three windings on the core shown for Prob. 2-5 with winding senses selected such that the following terminals (placed in the order shown in the figure for Prob. 2-5) have the same mark: (a) 1 and 2, 2 and 3, 3 and 1, (b) 1′ and 2′, 2′ and 3′, 3′ and 1′.

2-7. The figure shows four windings on a magnetic flux-conducting



Prob. 2-7.

ber of similar dots. Both schemes have advantages for particular problems, and both are used in electrical engineering literature.

Example 3

In the system shown in Fig. 2-12, the winding sense of each coil of the transformer is indicated. The polarity markings for each set of

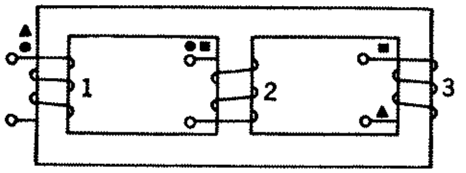


Fig. 2-12. Magnetic circuit for Example 3.

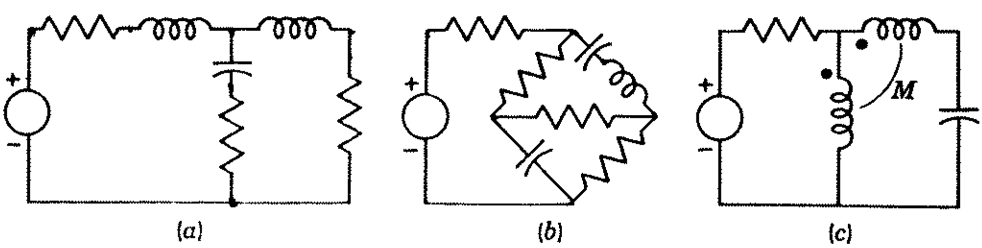
coils are shown on the figure. In each case, one of the dots for each winding-pair was arbitrarily selected and the position of the other dot was then determined.

FURTHER READING

The subject of network topology (or network geometry) is treated in detail in Guillemin's *Introductory Circuit Theory* (John Wiley & Sons, Inc., New York, 1953) in Chap. 1. Other suggested references include *Matrix Analysis of Electric Networks* by LeCorbeiller (John Wiley & Sons, Inc., New York, 1950) and the article "On the foundations of electrical network theory" by Ingram and Cramlet which appears in *J. Math. Phys.*, **23,** 134–155 (1944). See also the article "IRE standards for network topology," *Proc. IRE*, **39,** 27 (1951). On the subject of the dot convention for mutual inductance, see *Principles of Electric and Magnetic Circuits* by Boast (Harper & Brothers, New York, 1950), especially Chaps. 15 and 16.

PROBLEMS

2-1. For each of the circuits shown in the figure: (a) determine the value of the quantities E, S, N_t, N, L, and J; (b) select the N independ-



Prob. 2-1.

ent node pairs, using one datum node; and (c) following the rule of von Helmholtz, draw L independent loops.

CHAPTER 3
NETWORK EQUATIONS

3-1. Kirchhoff's equations

Most network equations are formulated from two simple laws first given by Kirchhoff in 1845.* The first law relates to the sum of the instantaneous voltages of the elements in a loop. It states that *in any loop the sum of the voltage drops must equal the sum of the voltage rises.*

This law follows from the scalar nature of voltage. To illustrate the concept involved, consider another scalar quantity, elevation. Suppose that we make a trip in an airplane, visiting a number of cities but eventually returning to our place of origin. At each stop, we will determine the elevation and record the elevation increase or decrease. When the trip is completed, we can be sure if we are sufficiently accurate that the sum of the elevation increases will just equal the sum of the elevation decreases. Otherwise, we would not be back at our starting elevation.

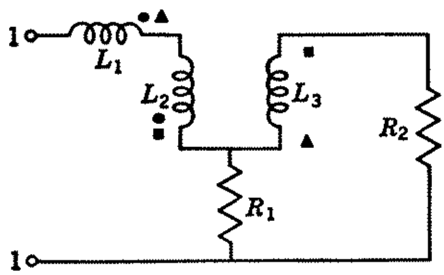
Next, suppose that we make a tour around some loop in a network at some fixed instant of time. At each node, we will measure the voltage with respect to the previous node and record voltage increases and voltage decreases. Once the loop is completely traversed, the sum of the voltage decreases (or drops) must equal the sum of the voltage increases (or rises). The fact that there are other loops in the same network has no effect on the sum of the drops and rises, just as the existence of alternate airline routes does not affect the altitude change summations.

Kirchhoff's second law relates to the sum of instantaneous currents at a node. It states that *the sum of currents flowing into the node equals the sum of currents flowing out.* In an analogous hydraulic system, the sum of water flowing out of a junction of pipes must equal the water flowing in, assuming no storage capacity at the junction. If we assume no charge storage capacity at the nodes of a network, then, just as in the hydraulic system, the currents into that node must equal those out.

* Historically, the work of Kirchhoff closely followed the pioneer works of Faraday in describing electric induction, of Oersted in relating magnetism and electricity in 1820, of Ampere in relating force and current in 1820–25 and of Ohm in relating voltage and current in 1826. At the time Kirchhoff published the work containing these laws, he was 23 years of age. He made contributions in several sciences—there are other Kirchhoff laws in other fields.

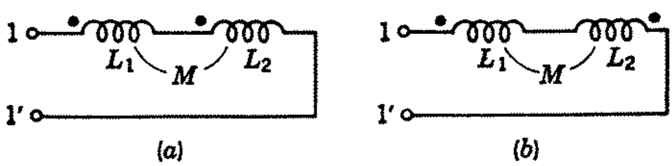
core. Using different shaped dots, establish polarity markings for the windings.

2-8. The accompanying schematic shows the equivalent circuit of a system with polarity marks on the three coupled coils. Draw a transformer with a core similar to that shown for Prob. 2-7 and place windings on the legs of the core in such a way as to be equivalent to the schematic. Show connections between the elements in the same drawing.



Prob. 2-8.

2-9. The accompanying schematics each show two inductors with coupling but with different dot markings. For each of the two systems, determine the equivalent inductance of the system at terminals 1-1′ by combining inductances.



Prob. 2-9.

2-10. A transformer has 100 turns on the primary (terminals 1-1′) and 200 turns on the secondary (terminals 2-2′). A current in the primary causes a magnetic flux which links all turns of both the primary and the secondary. The flux decreases according to the law $\phi = e^{-t}$ weber, when $t \geqq 0$. Find: (a) the flux linkages of the primary and secondary, (b) the voltage induced in the secondary.

With the loop currents as shown in Fig. 3-3, the Kirchhoff voltage laws are

$$R_1I_1 + R_3(I_1 - I_2) = V \tag{3-8}$$

$$-R_3(I_1 - I_2) + R_2I_2 = 0 \tag{3-9}$$

The two sets of Kirchhoff equations, Eqs. 3-6 and 3-7 and Eqs. 3-8 and 3-9, are identical if $I_1 = I_a$ and $I_2 = I_b$. These currents are identical, of course, since they are the currents flowing in R_1 and R_2, respectively. By Eq. 3-5, we find that I_c is expressed in terms of I_1 and I_2 as

$$I_c = I_1 - I_2 \tag{3-10}$$

In analysis, we are ultimately interested in determining currents in the elements—the branch currents. Our example has shown that there are two routes to determine the branch currents: (1) write equations directly in terms of the branch currents, or (2) write equations in terms of loop currents from which the branch currents can be found by addition or subtraction. Since the number of branches is equal or greater than the number of loops, the advantage of simplicity is usually in the second choice.

3-3. Positive directions for currents

Suppose that we were assigned the problem of counting cars traveling each direction on a busy street in a large city. Our first step would be to distinguish cars traveling in the two directions. We would accomplish this by deciding on a positive direction of flow. With this decision made, each car could be considered as moving in the positive direction or opposite to the direction considered positive (although handier terms such as north and south would likely be used).

Similarly, before writing network equations based on Kirchhoff's voltage law, a positive direction of the loop (or branch) currents must be assigned for each loop (or for each branch) and identified with an arrow. Such a decision establishes a positive or reference direction. Currents in the direction opposite to that considered positive are marked with a negative sign. The direction to be assumed positive is arbitrary, of course, but for uniformity, loop currents will usually be assigned a *clockwise* positive direction.

Once the positive direction for the loop current is assigned, the loop may be traversed in either direction in applying the Kirchhoff voltage law. If the loop is traversed in a direction opposite to that assigned for positive loop current, the Kirchhoff equation is not changed, since this is equivalent to multiplying all terms by -1.

Consider the series circuit shown in Fig. 3-1. We see that there are voltage drops across the three passive elements and a voltage rise due to the battery. Kirchhoff's voltage law states that the voltage drops must equal the voltage rises, or

$$V_1 + V_2 + V_3 = V_g \tag{3-1}$$

A part of a network is shown in Fig. 3-2 with the direction of current

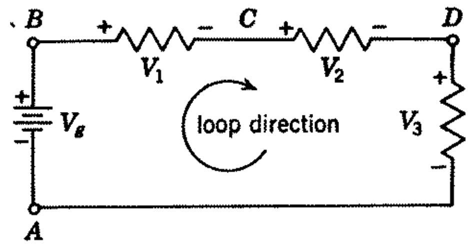


Fig. 3-1. The sum of voltage drops equals the battery voltage by Kirchhoff's voltage law.

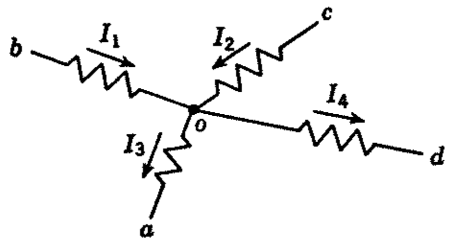


Fig. 3-2. Currents into the node equal those out of the node by Kirchhoff's current law.

shown for each branch attached to a particular node. At that node, the currents flowing into the node must equal those flowing out, or

$$I_1 + I_2 = I_3 + I_4 \tag{3-2}$$

3-2. Branch currents and loop currents

Kirchhoff's voltage law may be applied by using either branch currents or loop currents. To show the equivalence of branch and loop currents, consider the network shown in Fig. 3-3. With currents as assigned, the Kirchhoff voltage equations are

$$R_1I_a + R_3I_c = V \quad \text{(for loop 1)} \tag{3-3}$$

$$-R_3I_c + R_2I_b = 0 \quad \text{(for loop 2)} \tag{3-4}$$

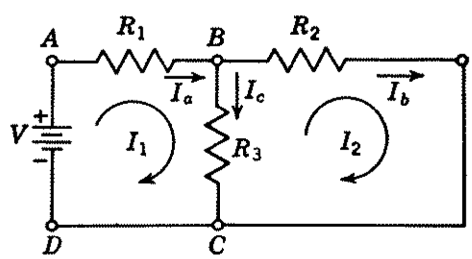


Fig. 3-3. Two-loop network.

At node B, the Kirchhoff current equation is

$$I_a = I_b + I_c \quad \text{or} \quad I_c = I_a - I_b \tag{3-5}$$

This equation may be used to eliminate I_c from the Kirchhoff voltage equations of Eqs. 3-3 and 3-4 as

$$R_1I_a + R_3(I_a - I_b) = V \tag{3-6}$$

$$-R_3(I_a - I_b) + R_2I_b = 0 \tag{3-7}$$

Example 3

A three-loop network is shown in Fig. 3-6 with loop currents i_1, i_2, and i_3 assigned positive directions as shown. Traversing the three

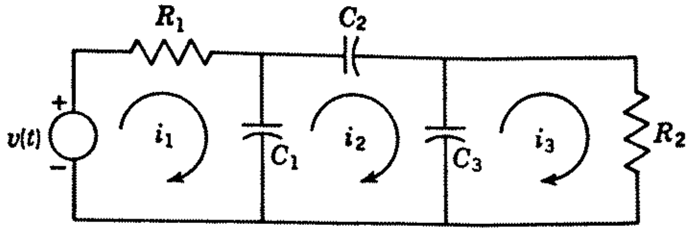


Fig. 3-6. Three-loop network.

loops in turn gives the three Kirchhoff voltage equations

$$R_1 i_1 + \frac{1}{C_1}\int (i_1 - i_2)\,dt = v(t) \tag{3-15}$$

$$\frac{1}{C_1}\int (i_2 - i_1)\,dt + \frac{1}{C_2}\int i_2\,dt + \frac{1}{C_3}\int (i_2 - i_3)\,dt = 0 \tag{3-16}$$

$$\frac{1}{C_3}\int (i_3 - i_2)\,dt + R_2 i_3 = 0 \tag{3-17}$$

3-5. Loop analysis of circuits with coupled coils

The rule developed in Art. 2-4 regarding polarity of induced voltage and current direction with respect to dots can be used to advantage in analysis of circuits with coupled coils. To apply the rule, (1) the polarity markings (dots) for each pair of coupled coils—or equivalent information—must be given, and (2) the positive direction of current flow must be assumed for each loop. A part of a circuit fulfilling these two requirements is shown in Fig. 3-7(a). By the rule, current i_1

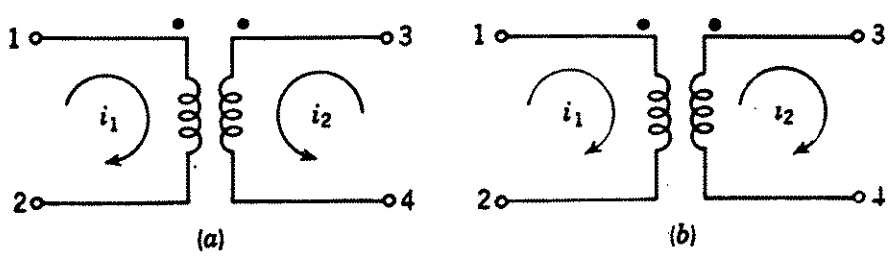


Fig. 3-7. Coupled coils illustrating the relationship of assumed direction of current, polarity markings, and polarity of induced voltage.

enters the dotted terminal of winding 1-2 and so will induce a voltage in winding 3-4 positive at the dotted terminal, terminal 3. The current i_1 thus induces a voltage drop from 3 to 4, or a voltage rise from 4 to 3. Similarly, i_2 induces a voltage in winding 1-2 with terminal

3-4. Formulating equations on the loop basis

A number of examples will illustrate the formulation of equations of equilibrium using Kirchhoff's voltage law.

Example 1

Figure 3-4 shows a series RLC circuit. It is quite clear by inspection that there is but one loop, while there are 3 independent node pairs. The voltage drops across the passive elements must equal the voltage rise due to the active element. Expressions for the voltages across the

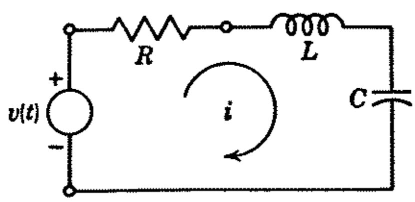


Fig. 3-4. Series circuit.

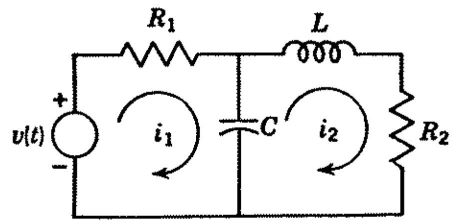


Fig. 3-5. Two-loop network.

passive elements were derived in Chapter 1. In terms of these expressions, Kirchhoff's voltage law requires that

$$Ri + L\frac{di}{dt} + \frac{1}{C}\int i\,dt = v(t) \tag{3-11}$$

at all times. This is an *integrodifferential equation*, which may be changed to a differential equation by differentiation to give

$$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i = \frac{dv(t)}{dt} \tag{3-12}$$

where the derivatives have been arranged in descending order.

Example 2

The network of Fig. 3-5 has two independent loops, since $L = E - N_t + S = 5 - 4 + 1 = 2$, and the two loop currents, i_1 and i_2, have been assigned positive directions as shown. The equilibrium equations of the voltages, based on Kirchhoff's law, are

$$R_1i_1 + \frac{1}{C}\int (i_1 - i_2)\,dt = v(t) \tag{3-13}$$

$$\frac{1}{C}\int (i_2 - i_1)\,dt + L\frac{di_2}{dt} + R_2i_2 = 0 \tag{3-14}$$

Example 4

The winding sense of three coils on a flux-conducting material is shown in Fig. 3-9. We are required to write the Kirchhoff voltage equations, taking into account mutual inductance. With the aid of dots, the system of Fig. 3-9 can be replaced by the equivalent circuit

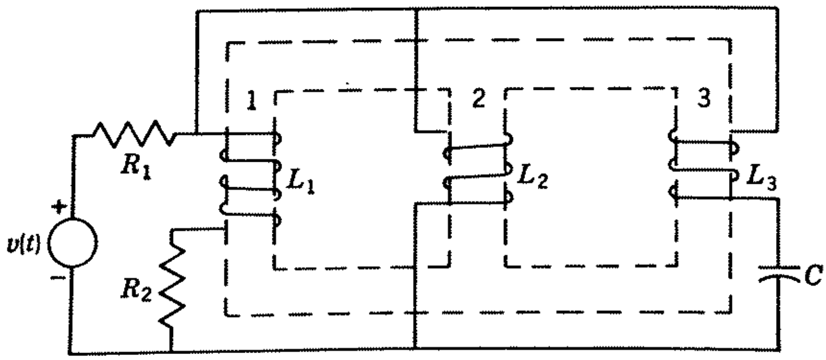


Fig. 3-9. Magnetic circuit of Example 4.

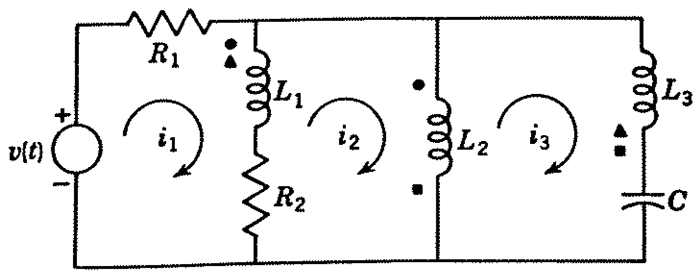


Fig. 3-10. Magnetic system of Fig. 3-9 showing polarity markings and assumed positive direction of current.

of Fig. 3-10. If we use a double subscript notation for mutual inductance to indicate the two coils being considered, the Kirchhoff voltage equations are

$$R_1 i_1 + L_1 \frac{d(i_1 - i_2)}{dt} + M_{12} \frac{d(i_2 - i_3)}{dt} - M_{13} \frac{di_3}{dt} + R_2(i_1 - i_2) = v(t) \tag{3-21}$$

$$R_2(i_2 - i_1) + L_1 \frac{d(i_2 - i_1)}{dt} - M_{12} \frac{d(i_2 - i_3)}{dt} + M_{13} \frac{di_3}{dt} + L_2 \frac{d(i_2 - i_3)}{dt} + M_{21} \frac{d}{dt}(i_1 - i_2) + M_{23} \frac{d}{dt} i_3 = 0 \tag{3-22}$$

$$L_2 \frac{d}{dt}(i_3 - i_2) - M_{23} \frac{di_3}{dt} - M_{21} \frac{d}{dt}(i_1 - i_2) + L_3 \frac{di_3}{dt} + M_{32} \frac{d}{dt}(i_2 - i_3) - M_{31} \frac{d}{dt}(i_1 - i_2) + \frac{1}{C} \int i_3\, dt = 0 \tag{3-23}$$

In this particular problem, the equations would have had simpler form

1—the dotted terminal—positive, and so with a voltage drop from terminal 1 to 2.

In Fig. 3-7(b), the current i_2 has a positive direction reversed from that shown in Fig. 3-7(a). This current is positive when it *leaves* the dotted terminal and hence induces a voltage in winding 1-2 with terminal 2 positive, and so with a voltage *rise* from terminal 1 to 2.

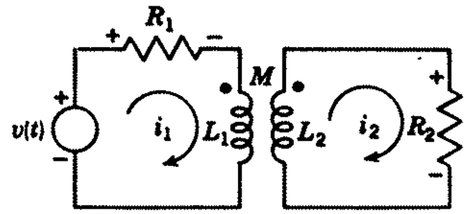


Fig. 3-8. A two-loop coupled circuit illustrating the sign convention for M.

Figure 3-8 shows the coupled coils of Fig. 3-7(b) incorporated into a two-loop coupled circuit. Applying the rules just discussed, the Kirchhoff voltage law applied to the first loop gives the equilibrium equation

$$R_1 i_1 + L_1 \frac{di_1}{dt} - M \frac{di_2}{dt} = v(t) \tag{3-18}$$

The current i_1 produces a voltage drop across L_1, but the current i_2, when positive, with the polarity markings as shown induces a voltage rise across the same terminals. In the second loop, the equilibrium equation is

$$+L_2 \frac{di_2}{dt} - M \frac{di_1}{dt} + R_2 i_2 = 0 \tag{3-19}$$

In these equations the sign before a term of the form $M(di_2/dt)$ indicates a voltage rise if negative and a voltage drop if positive. As long as polarity dots are given along with the direction of positive current, there is no ambiguity, and the rule of Art. 2-4 can be applied successively to all coupled coils. If the number of coils is large, the use of dots of various shapes may become cumbersome. In this case, it is more convenient to assign a plus or minus sign to M, to let M carry the sign in the equation formulation rather than letting the sign be specified by the nature of the induced voltage—drop or rise. These two systems are equivalent, and both will be used in the discussion to follow, just as both are used in the literature.

Consider the circuit of Fig. 3-8, described by Eqs. 3-18 and 3-19. We observe in these equations that the voltages induced by means of the coupled coils are voltage rises of opposite polarity to the voltage drops of either loop. With the current directions given, the dots can be erased, provided a negative sign is identified with mutual inductance as $-M$. By this system, Eq. 3-18 is written

$$R_1 i_1 + L_1 \frac{di_1}{dt} + (-M) \frac{di_2}{dt} = v(t) \tag{3-20}$$

Let us now turn our attention to node c. There the current in R_4 is marked as I_4. Kirchhoff's current law requires that

$$I_3 = I_4 \quad \text{or} \quad I_4 - I_3 = 0 \tag{3-29}$$

Now since
$$I_3 = \frac{1}{R_3}(V_b - V_c) \quad \text{and} \quad I_4 = \frac{1}{R_4}V_c \tag{3-30}$$

we have the second equilibrium equation,

$$-\frac{V_b}{R_3} + \frac{V_c}{R_3} + \frac{V_c}{R_4} = 0 \tag{3-31}$$

Equations 3-28 and 3-31 must be solved simultaneously to give the unknown values of V_b and V_c.

Have we any flexibility in choosing positive directions of current for the different nodes? On the network under consideration, a new current I_3' is marked with an arrow such that $I_3' = -I_3$. In terms of this new current, Kirchhoff's current law is $I_3' + I_4 = 0$. But since $I_3' = -I_3$, this equation is identical with Eq. 3-29, and so with Eq. 3-31. In other words, the positive direction of the branch currents may be assumed at each node independent of previous designations. We thus have two options: (1) Assume positive directions for branch currents once and for all. (2) Assume new positive directions at each node, for example that currents flow out of the node for all passive elements and in the marked direction for active current sources.

As a result of this discussion, we see that the steps to be followed in node analysis are the following:

(1) Select a datum node and identify all unknown node voltages.
(2) Assume a positive direction for all branch currents.
(3) Apply Kirchhoff's current law to each node of unknown voltage, writing each branch current in terms of a node-to-node voltage and appropriate circuit parameters.

It is sometimes convenient to change a voltage source into a mathematically equivalent current source for analysis. In Fig. 3-12(a), let $v(t)$ be the potential of the voltage source and $v_1(t)$ be the potential of the node located between the resistor R_1 and the rest of the network. The current $i(t)$ flows through the resistor R_1. Kirchhoff's voltage law for the circuit of Fig. 3-12(a) is

$$v(t) = R_1 i(t) + v_1(t) \tag{3-32}$$

Solving this equation for $i(t)$ gives

$$i(t) = \frac{v(t)}{R_1} - \frac{v_1(t)}{R_1} \tag{3-33}$$

if the generator $v(t)$ and R_1 had been part of each of the three loops. (See Prob. 3-14.)

3-6. Formulating equations on the node basis

The node basis for formulating the equilibrium equations for circuits makes use of the Kirchhoff law that the sum of currents leaving a node is equal to the sum of currents entering that node. To illustrate the procedures used in node analysis, consider the simple resistive network shown in Fig. 3-11. For this network there are four nodes, marked a, b, c, and d. Following convention, the negative terminal of the active element, node d, is selected as the datum node. There are then three node-pair voltages, the potentials of nodes a, b, and c with respect to node d. However, the potential from node a to node d is known to be equal to the battery voltage. There are thus but two unknown voltages in the network: the voltages of node b and node c with respect to the datum node.

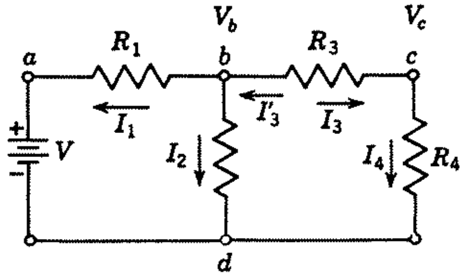


Fig. 3-11. Network illustrating procedures in node analysis.

Having identified the unknown voltages, our next task is to write network equations in terms of these unknown node voltages. This is accomplished in terms of *branch currents* (never loop currents). Each branch current must be assigned a direction considered positive and so marked with an arrow, just as in the case of loop analysis. At node b in the network of Fig. 3-11, the branch currents are marked as I_1, I_2, and I_3, all directed out of the node. By Kirchhoff's current law, we know that

$$I_1 + I_2 + I_3 = 0 \tag{3-24}$$

What are these branch currents in terms of the node voltages? By Ohm's law, they are

$$I_1 = \frac{1}{R_1}(V_b - V) \tag{3-25}$$

$$I_2 = \frac{1}{R_2}(V_b - 0) \tag{3-26}$$

$$I_3 = \frac{1}{R_3}(V_b - V_c) \tag{3-27}$$

Substituting these three equations into Eq. 3-24 gives

$$\frac{V_b}{R_1} - \frac{V}{R_1} + \frac{V_b}{R_2} + \frac{V_b}{R_3} - \frac{V_c}{R_3} = 0 \tag{3-28}$$

of the generator is $v(t)$ in Fig. 3-13(a), we may write

$$\frac{1}{R}[v_1 - v(t)] + \frac{1}{L}\int v_1\,dt + C\frac{dv_1}{dt} = 0 \tag{3-35}$$

or

$$\frac{1}{R}v_1 + \frac{1}{L}\int v_1\,dt + C\frac{dv_1}{dt} = \frac{v(t)}{R}$$

which is identical with Eq. 3-34. Analysis may be carried out with either the voltage source or the equivalent current source.

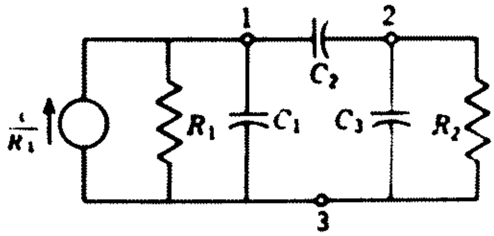


Fig. 3-14. Equivalent circuit of three-loop network of Fig. 3-6.

Example 6

The network shown in Fig. 3-14 is the current source equivalent to the three-loop network shown in Fig. 3-6. Node 3 is the datum node, and the unknown voltages at nodes 1 and 2 are designated v_1 and v_2. At node 1, setting $1/R_1 = G_1$ and $1/R_2 = G_2$,

$$G_1 v_1 + C_1\frac{dv_1}{dt} + C_2\frac{d}{dt}(v_1 - v_2) = \frac{v}{R_1} \tag{3-36}$$

and at node 2,

$$C_2\frac{d}{dt}(v_2 - v_1) + C_3\frac{dv_2}{dt} + G_2 v_2 = 0 \tag{3-37}$$

In this example, formulation on the node basis has resulted in fewer differential equations than on the loop basis in Example 3. Ordinarily it requires less work in solving two simultaneous differential equations than in solving three. The choice of method of formulation, loop or node, also depends on the objective of analysis. In this example, if the voltage at node 2 is desired, the node method has the advantage over the loop method. But if it is the current flowing in capacitor C_3 that is to be found, we must weigh the relative advantages of the two methods. The loop currents can be assigned so that only one loop current flows in C_3, but three simultaneous equations must be solved. Using the node method, we might find the voltage at node 2 first and then determine the current in the capacitor from the equation

$$i_{C_3} = C_3\frac{dv_2}{dt} \tag{3-38}$$

The second method appears to involve less computation in this particular example.

In this current equation, we will identify each separate term. The equation tells us that the current $i(t)$ flowing into the network is equal to a current $v(t)/R_1$ minus a current $v_1(t)/R_1$. This equation may be interpreted in terms of the new network of Fig. 3-12(b) by means of

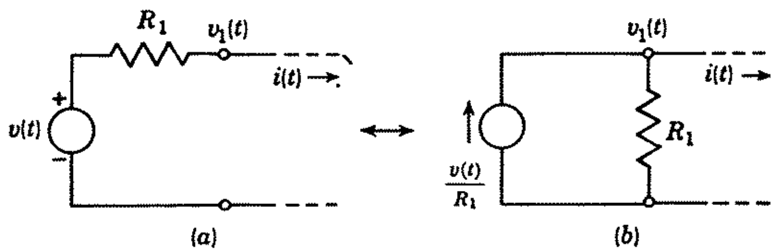


Fig. 3-12. Interchange of sources.

Kirchhoff's current law. The current $v(t)/R_1$ is from an equivalent current source. The current $v_1(t)/R_1$ is the current flowing through the resistor R_1 connected in shunt with the current source. The difference of these two currents is the current flowing into the network. Since the two networks of Fig. 3-12 are described by the same equations, Eqs. 3-32 and 3-33, they are equivalent.

Example 5

Consider the network shown in Fig. 3-13(a). The voltage source may be converted into an equivalent current source by the procedure just described, giving the network of Fig. 3-13(b). Node 2 is designated the

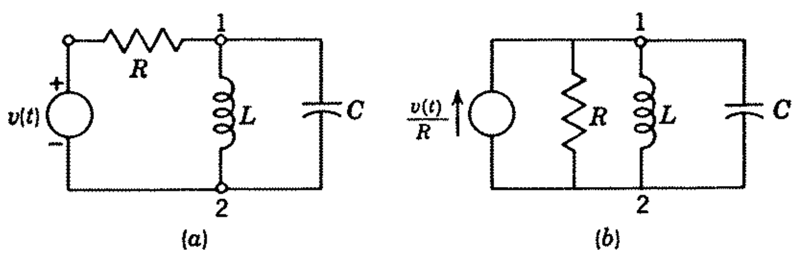


Fig. 3-13. Network for node analysis.

datum node and all branch currents are assigned to flow out of node 1. The expressions for the currents in each of the elements in terms of the voltage were given in Chapter 1. By Kirchhoff's current law, the current equation is

$$\frac{1}{R} v_1 + \frac{1}{L} \int v_1 \, dt + C \frac{dv_1}{dt} = \frac{v(t)}{R} \tag{3-34}$$

Of course, it is not necessary to make the conversion to the current source before analyzing the network. Since the voltage of the + terminal of the generator is $v(t)$ in Fig. 3-13(a), we may write

$$\frac{1}{R}[v_1 - v(t)] + \frac{1}{L}\int v_1\,dt + C\frac{dv_1}{dt} = 0 \tag{3-35}$$

or

$$\frac{1}{R}v_1 + \frac{1}{L}\int v_1\,dt + C\frac{dv_1}{dt} = \frac{v(t)}{R}$$

which is identical with Eq. 3-34. Analysis may be carried out with either the voltage source or the equivalent current source.

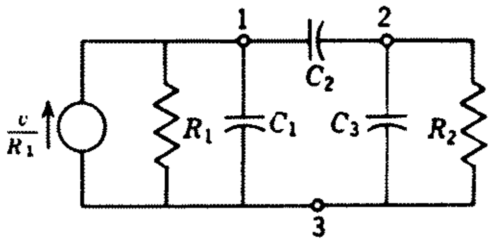


Fig. 3-14. Equivalent circuit of three-loop network of Fig. 3-6.

Example 6

The network shown in Fig. 3-14 is the current source equivalent to the three-loop network shown in Fig. 3-6. Node 3 is the datum node, and the unknown voltages at nodes 1 and 2 are designated v_1 and v_2. At node 1, setting $1/R_1 = G_1$ and $1/R_2 = G_2$,

$$G_1v_1 + C_1\frac{dv_1}{dt} + C_2\frac{d}{dt}(v_1 - v_2) = \frac{v}{R_1} \tag{3-36}$$

and at node 2,

$$C_2\frac{d}{dt}(v_2 - v_1) + C_3\frac{dv_2}{dt} + G_2v_2 = 0 \tag{3-37}$$

In this example, formulation on the node basis has resulted in fewer differential equations than on the loop basis in Example 3. Ordinarily it requires less work in solving two simultaneous differential equations than in solving three. The choice of method of formulation, loop or node, also depends on the objective of analysis. In this example, if the voltage at node 2 is desired, the node method has the advantage over the loop method. But if it is the current flowing in capacitor C_3 that is to be found, we must weigh the relative advantages of the two methods. The loop currents can be assigned so that only one loop current flows in C_3, but three simultaneous equations must be solved. Using the node method, we might find the voltage at node 2 first and then determine the current in the capacitor from the equation

$$i_{C_3} = C_3\frac{dv_2}{dt} \tag{3-38}$$

The second method appears to involve less computation in this particular example.

Example 7

The network shown in Fig. 3-15 differs from the networks of other examples in that there is no series resistance with the voltage source. Although this network has four independent loops, there is but one unknown node voltage, that at node 2. From Kirchhoff's current law, we write

$$C_3 \frac{d}{dt}(v_2 - v_1) + \frac{1}{L}\int (v_2 - v_1)\,dt + Gv_2 + C_2 \frac{dv_2}{dt} = 0 \quad (3\text{-}39)$$

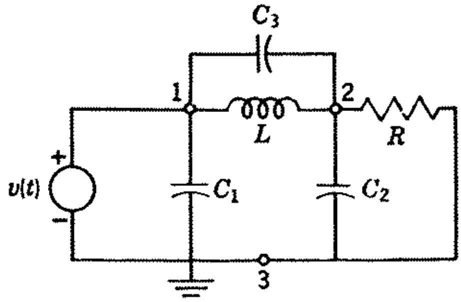


Fig. 3-15. Network of Example 7.

where, as before, $G = 1/R$. Note that C_1 does not appear in the equation. This is because the voltage at node 1 is independent of the capacitor C_1 or any other shunt element. Capacitor C_1 is an extraneous element. The voltage source must maintain terminal voltage for any load (or it is not an ideal element), and so C_1 may be removed without affecting the network equations.

3-7. Duality

Several analogous situations will have been noted in the preceding discussions of this chapter. The statements of the two Kirchhoff laws were almost word for word with voltage substituted for current, independent loop for independent node pair, etc. Likewise, the integrodifferential equations that resulted from the application of the Kirchhoff laws have been similar in appearance. This repeated similarity is only part of a larger pattern of identical behavior patterns in the roles played by voltage and current in network analysis. This similarity, with all the implications, is termed the principle of *duality*.

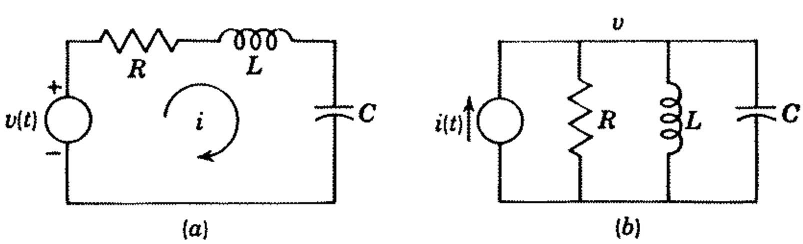


Fig. 3-16. Dual networks.

Consider the two networks completely different in physical appearance shown in Fig. 3-16. Inspection shows that the first might be analyzed to advantage on the loop basis and the other on the node basis.

The resulting equations are

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = v(t) \tag{3-40}$$

$$C\frac{dv}{dt} + Gv + \frac{1}{L}\int v\,dt = i(t) \tag{3-41}$$

These two equations specify identical mathematical operations, the only difference being in letter symbols. The solution of one equation is also the solution of the other. The two networks are *duals*. The roles of current and voltage in the two networks have been interchanged. As a word of caution, one network is not the equivalent of the other in the sense that one can replace the other.

An inspection of the terms of Eqs. 3-40 and 3-41 shows that the following are analogous quantities.

$$Ri \quad \text{and} \quad Gv$$

$$L\frac{di}{dt} \quad \text{and} \quad C\frac{dv}{dt}$$

$$\frac{1}{C}\int i\,dt \quad \text{and} \quad \frac{1}{L}\int v\,dt$$

Evidently the following pairs are dual quantities.

R	and	G
L	and	C
loop current, i	and	v, node-pair voltage
$\left.\begin{matrix} q \text{ or} \\ \int i\,dt \end{matrix}\right\}$	and	$\left\{\begin{matrix} \psi \text{ or} \\ \int v\,dt \end{matrix}\right.$
loop	and	node pair
short circuit	and	open circuit

A simple graphical construction* may be followed in finding the dual of a network.

(1) Inside each loop place a node, giving it a number for convenience. Place an extra node, the datum node, external to the network. Arrange the same numbered nodes on a separate space on the paper for construction of the dual.

* Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 46ff.

(2) Draw lines from node to node through the elements in the original network, traversing only one element at a time. For each element traversed in the original network, connect the dual element—from the chart above—on the dual network being constructed.
(3) Continue this process until the number of possible paths through single elements is exhausted. (Should you slip and go through a connecting wire which is assumed to be a short circuit, the dual element is an open circuit.)
(4) The network constructed in this manner is the dual network. This construction may be checked by writing the differential equations for the two systems, one on the loop basis and the other on the node basis.

This graphical construction is illustrated in Fig. 3-17. Networks that are not planar (that is, cannot be shown schematically in one plane with no wires crossing) do not have duals.

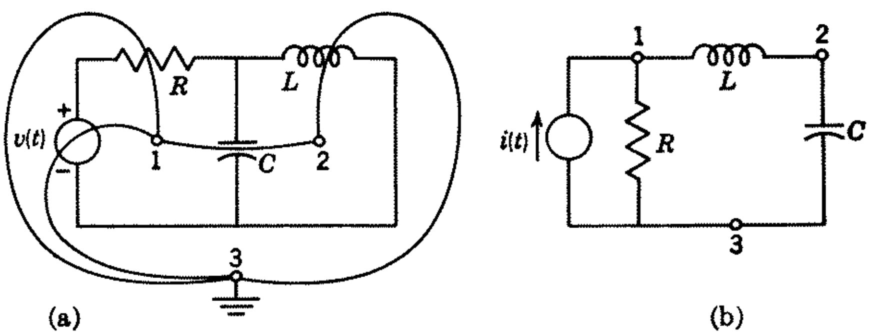


Fig. 3-17. Graphical procedures for finding dual of network: (a) original; (b) dual.

3-8. General network equations

Thus far we have progressed from analysis of very simple networks to successively more complex network configurations. To systematize our approach to the analysis of networks, consider an L-loop network, where L is any number. A representation of such a network is shown in Fig. 3-18. In this diagram, the circuit elements in each branch have been replaced by a straight line for simplicity. The effects of mutual inductance are not indicated but are assumed to be present.

Consider loop 1. This loop may contain resistance, inductance, and capacitance in any one or all of the branches that make up the loop. Let

R_{11} be the total resistance in loop 1.
L_{11} be the total inductance in loop 1.
S_{11} be the total elastance of loop 1.

We use elastance instead of capacitance here because elastance terms add directly for a series circuit, while capacitance terms combine as

$$\frac{1}{C_{11}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots + \frac{1}{C_n}$$

There will be voltage drops in loop 1 produced by current flow in loop 2, in loop 3, loop 4—in fact, all loops in the general case. Rather than specialize on loop 1, consider the effect of currents in the jth loop on

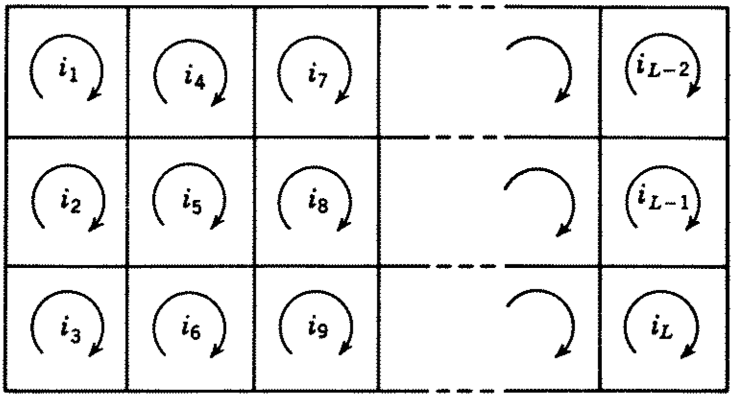


Fig. 3-18. *L*-loop network.

voltage in loop k, where j and k are any integers from 1 to L. For these two loops, let R_{kj} = the total resistance common to loops k and j; L_{kj} = the total inductance (including mutual) common to loops k and j; S_{kj} = the total elastance common to loops k and j. The voltage drop in loop k produced by current i_j is

$$R_{kj}i_j + L_{kj}\frac{di_j}{dt} + S_{kj}\int i_j\,dt \tag{3-42}$$

At this point, we will adopt a special notation for equations of this form by letting the following equation be the equivalent of Eq. 3-42.

$$\left(R_{kj} + L_{kj}\frac{d}{dt} + S_{kj}\int dt\right) i_j = a_{kj}i_j \tag{3-43}$$

This symbolism implies that the variable i_j is operated upon by multiplication by R_{kj}, multiplication by L_{kj} and differentiation, and finally, multiplication by S_{kj} and integration. All three operations are summarized in the symbol a_{kj}.

The total voltage drop in loop k will be found by successively considering loop k and the currents flowing in every other loop. Mathematically this is done by letting j have all values from 1 to L. This total voltage drop must be equal to the total voltage rise from active sources within loop k, which we write as v_k. Then by Kirchhoff's voltage law, we have

$$\sum_{j=1}^{L} a_{kj} i_j = v_k \tag{3-44}$$

There remains only to repeat this process for all loops, by letting k have all values from 1 to L. Thus the most general form for Kirchhoff's voltage law for an L-loop network is

$$\sum_{j=1}^{L} a_{kj} i_j = v_k, \qquad k = 1, 2, \ldots, L \tag{3-45}$$

The expansion of this concise equation is the following set of equations.

$$\begin{array}{l} a_{11}i_1 + a_{12}i_2 + a_{13}i_3 + \ldots + a_{1L}i_L = v_1 \\ a_{21}i_1 + a_{22}i_2 + a_{23}i_3 + \ldots + a_{2L}i_L = v_2 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{L1}i_1 + a_{L2}i_2 + a_{L3}i_3 + \ldots + a_{LL}i_L = v_L \end{array} \tag{3-46}$$

It is helpful to arrange these equations given above in the form of a *chart* (or *schedule*) in which the a-coefficients are emphasized. Such a chart is shown below.

		Coefficient of						
Eq.	Voltage	i_1	i_2	i_3	i_4	i_5	...	i_L
1	v_1	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	...	a_{1L}
2	v_2	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	...	a_{2L}
...	...	...	...	...	...	...	...	...
L	v_L	a_{L1}	a_{L2}	a_{L3}	a_{L4}	a_{L5}	...	a_{LL}

If the loop currents are all assumed positive in the same path direction, clockwise for example, then all a_{jj} terms are positive and all $a_{jk}(j \neq k)$ terms are negative. In actual problems, of course, many of the a-coefficients are zero.

Example 8

A two-loop network is shown in Fig. 3-19. In this network there are two sources of voltage and no mutual inductance. The Kirchhoff voltage law is

$$\sum_{j=1}^{2} a_{kj} i_j = v_k, \qquad k = 1, 2$$

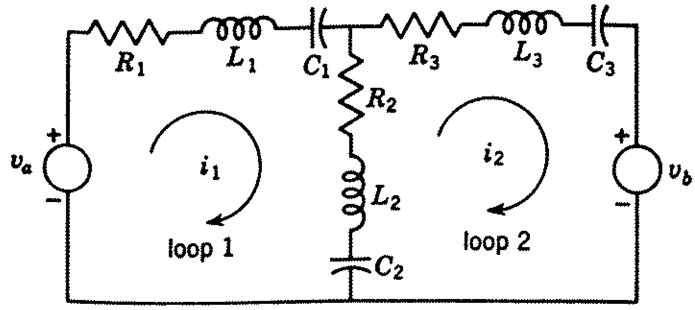


Fig. 3-19. Two-loop network.

or in expanded form,

$$a_{11}i_1 + a_{12}i_2 = v_1, \qquad a_{21}i_1 + a_{22}i_2 = v_2$$

The a-coefficients are found by inspection of the network as follows.

$$a_{11} = (R_1 + R_2) + (L_1 + L_2)\frac{d}{dt} + (S_1 + S_2)\int dt$$

$$a_{22} = (R_2 + R_3) + (L_2 + L_3)\frac{d}{dt} + (S_2 + S_3)\int dt$$

$$a_{12} = a_{21} = -R_2 - L_2\frac{d}{dt} - S_2\int dt$$

Similarly, the voltage terms are recognized to be

$$v_1 = v_a, \qquad v_2 = -v_b$$

The general equations for node analysis will be similar to those found for loop analysis, as might be expected from the principle of duality. Consider a network with N_t nodes and only one part such that there are $N = N_t - 1$ independent node pairs. Now each of the N equations is written from Kirchhoff's current law in terms of current directed into and out of the node. Currents into the node, in turn, are written in terms of node-to-node potentials and the parameters of the elements connected between the nodes being considered. For elements connected as shown in Fig. 3-20, the elements may be replaced by an equivalent system made up as follows: (1) all parallel capacitances replaced by an equivalent capacitance of value $C_{kj} = C_1 + C_2 + \ldots$; (2) an equivalent resistance found by adding

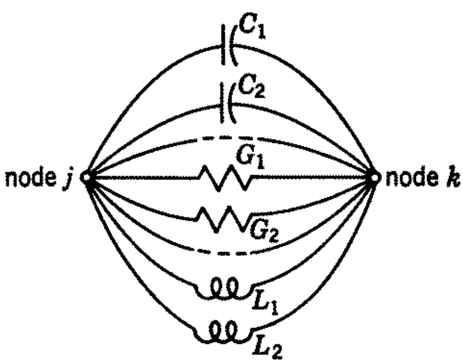


Fig. 3-20. Elements appearing between two nodes, j and k. The C's, G's, and L's may be combined to give an equivalent system.

conductances as $G_{kj} = 1/R_{kj} = G_1 + G_2 + \ldots$; and (3) an equivalent inductance of value L_{kj}, where $1/L_{kj} = 1/L_1 + 1/L_2 + \ldots$. Applying this network simplification to the circuits from node k to all other nodes from $j = 1$ to $j = N$, we have the equation

$$\sum_{j=1}^{N} \left(G_{kj} + C_{kj}\frac{d}{dt} + \frac{1}{L_{kj}}\int dt \right) v_j = i_k, \qquad k = 1, 2, \ldots, N \quad (3\text{-}47)$$

which may be written concisely as

$$\sum_{j=1}^{N} b_{kj}v_j = i_k, \qquad k = 1, 2, \ldots, N \quad (3\text{-}48)$$

by letting b_{kj} summarize the operations

$$\left(G_{kj} + C_{kj}\frac{d}{dt} + \frac{1}{L_{kj}}\int dt \right) = b_{kj} \quad (3\text{-}49)$$

The expansion of Eq. 3-48 has the same form as the expansion for the loop case, Eq. 3-46, with a's replaced by b's, i's by v's, and v's by i's.

In applying this equation to networks, it is not necessary to simplify the network by combining elements. At node j, the capacitance C_{jj} is the sum of the capacitance *connected to* node j. The value of C_{kj} is the sum of the capacitances *connected between* node j and node k. Similar instructions hold for inverse inductance $1/L$ and for conductance $G = 1/R$. Coefficients can thus be found by inspection by simply noting which elements are "hanging on" or "hanging between" the various nodes.

If the same convention for positive current is maintained in formulating all node equations for a network, the sign of b_{kj} will be positive when $k = j$, and negative when $k \neq j$.

Example 9

A network with two independent node pairs is shown in Fig. 3-21. For this network, Kirchhoff's current law is

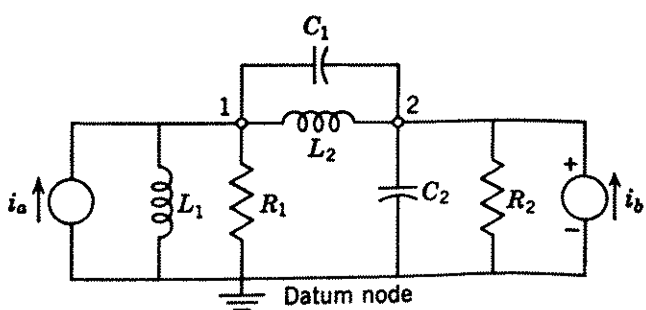


Fig. 3-21. Two-node network.

$$\sum_{j=1}^{2} b_{kj}v_j = i_k, \qquad k = 1, 2$$

or $$b_{11}v_1 + b_{12}v_2 = i_1, \qquad b_{21}v_1 + b_{22}v_2 = i_2$$

Values for the b-coefficients and the i's may be summarized in chart form as follows.

Eq.	Current	Coefficient of v_1	Coefficient of v_2
1	i_a	$G_1 + C_1\frac{d}{dt} + \left(\frac{1}{L_1} + \frac{1}{L_2}\right)\int dt$	$-C_1\frac{d}{dt} - \frac{1}{L_2}\int dt$
2	i_b	$-C_1\frac{d}{dt} - \frac{1}{L_2}\int dt$	$+G_2 + (C_1 + C_2)\frac{d}{dt} + \frac{1}{L_2}\int dt$

3-9. The solution of equations by determinants

Determinants are the mathematical tools we will use for systematic solution of simultaneous equations of the type derived in the last section. The array of quantities with straight line brackets on either side,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} \tag{3-50}$$

is known as a *determinant of order n*. Quantities in horizontal lines form *rows*, and quantities in vertical lines form *columns*. Such a determinant is square, having n rows and n columns. Each of the n^2 quantities in the determinant is known as an *element*. Element position in the determinant is identified by a double subscript, the first subscript indicating row and the second indicating column (numbered from the upper left-hand corner). Elements along the sloping line extending from a_{11} to a_{nn} form the *principal diagonal* of the determinant.

A determinant has a value which is a function of the values of its elements. In finding this value, we must make use of rules for expansion of the determinant in terms of the elements. Second- and third-order determinants have expansions that are familiar from studies in elementary algebra. Expansions for determinants of order higher than the third are conveniently made in terms of *minors*.

The *minor* of any element of a determinant a_{jk} is the determinant which remains when the column and row containing a_{jk} are deleted. In terms of the third-order determinant,

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \tag{3-51}$$

the minor for a_{11}, for example, is

$$\Delta_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} \tag{3-52}$$

A minor of the element a_{jk} multiplied by $(-1)^{j+k}$ is given the name *cofactor*. The cofactor sign is thus found by raising (-1) to the power found by adding the row and the column, $j + k$ as

$$(\text{cofactor}) = (-1)^{j+k}(\text{minor}) \tag{3-53}$$

Since, according to this rule, the cofactor signs alternate along any row or column, the proper cofactor sign can be determined by "counting" (plus, minus, plus, etc.) from a positive a_{11} position to any element, proceeding along any horizontal or vertical path.

Expansion of a determinant in terms of minors (or cofactors) consists of successive reduction of determinant order. A determinant of order n is equal to the sum of the product of the elements of any row or column multiplied by their corresponding $(n - 1)$ order cofactors. Applying this rule to the expansion of the determinant of Eq. 3-51 along the first column gives

$$\begin{aligned} \Delta &= a_{11}\Delta_{11} - a_{21}\Delta_{21} + a_{31}\Delta_{31} \\ &= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \end{aligned} \tag{3-54}$$

There are $2n$ equivalent expansions of the determinant about the n rows or n columns. The minor determinants can, in turn, be expanded by the same rule and the process continued until the value of Δ is given as the sum of $n!$ product factors.

The facts about determinants that we have just reviewed are essential in solving simultaneous equations of the form

$$\begin{gathered} a_{11}i_1 + a_{12}i_2 + a_{13}i_3 + \dots + a_{1L}i_L = v_1 \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{L1}i_1 + a_{L2}i_2 + a_{L3}i_3 + \dots + a_{LL}i_L = v_L \end{gathered}$$

that have resulted from application of Kirchhoff's voltage law (and similar equations from the Kirchhoff current law). The solution to such simultaneous equations is given by *Cramer's rule* as

$$i_1 = \frac{D_1}{\Delta}, \qquad i_2 = \frac{D_2}{\Delta} \dots i_L = \frac{D_L}{\Delta} \tag{3-55}$$

where Δ is the *system determinant* given as

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1L} \\ a_{21} & a_{22} & \dots & a_{2L} \\ \cdot & \cdot & \cdot & \cdot \\ a_{L1} & a_{L2} & \dots & a_{LL} \end{vmatrix} \tag{3-56}$$

which must be different from zero for the solutions $i_1, i_2, \dots, i_n$ to be unique, and D_j is the determinant formed by replacing the jth column of a-coefficients by the column $v_1, v_2, \dots, v_n$.

With Cramer's rule and the method of expansion by minors, simultaneous equations of the form of Eq. 3-46 can be solved. For a third-order equation, the solution for i_1 is

$$i_1 = \frac{D_1}{\Delta} = \frac{v_1\Delta_{11} - v_2\Delta_{21} + v_3\Delta_{31}}{\Delta} \tag{3-57}$$

or

$$i_1 = \frac{\Delta_{11}}{\Delta} v_1 - \frac{\Delta_{21}}{\Delta} v_2 + \frac{\Delta_{31}}{\Delta} v_3 \tag{3-58}$$

Similarly,

$$i_2 = -\frac{\Delta_{21}}{\Delta} v_1 + \frac{\Delta_{22}}{\Delta} v_2 - \frac{\Delta_{23}}{\Delta} v_3 \tag{3-59}$$

and so on. The form of these equations is greatly simplified if all v's except one are zero, corresponding to only one driving voltage source.

Example 10

For a certain three-loop network, the following equations are given.

$$\begin{aligned} 5i_1 - 2i_2 - 3i_3 &= 10 \\ -2i_1 + 4i_2 - 1i_3 &= 0 \\ -3i_1 - 1i_2 + 6i_3 &= 0 \end{aligned}$$

From Cramer's rule we write the solution for i_1 as

$$i_1 = \frac{D_1}{\Delta} = \frac{10\begin{vmatrix} 4 & -1 \\ -1 & 6 \end{vmatrix} - 0\begin{vmatrix} -2 & -3 \\ -1 & 6 \end{vmatrix} + 0\begin{vmatrix} -2 & -3 \\ 4 & -1 \end{vmatrix}}{\begin{vmatrix} 5 & -2 & -3 \\ -2 & 4 & -1 \\ -3 & -1 & 6 \end{vmatrix}} = \frac{230}{43}$$

Similarly,

$$i_2 = \frac{-(+10)\begin{vmatrix} -2 & -1 \\ -3 & 6 \end{vmatrix}}{\Delta} = \frac{150}{43}; \quad i_3 = \frac{+(10)\begin{vmatrix} -2 & 4 \\ -3 & -1 \end{vmatrix}}{\Delta} = \frac{140}{43}$$

3-10. Resistive network analysis

For networks restricted to contain only resistive elements, the a-coefficients in Eq. 3-46 become resistance terms as

$$a_{kj} \longrightarrow R_{kj}$$

and the b-coefficients become conductance terms as

$$b_{kj} \longrightarrow G_{kj}$$

Under this restriction, we can postpone our questions relating to the manipulation of a-coefficients and b-coefficients which include the operations of integration and differentiation. The general form of Kirchhoff's voltage law equations for the resistive case is

$$\sum_{j=1}^{L} R_{kj} i_j = v_k, \qquad k = 1, 2, \ldots, L \tag{3-60}$$

where R_{jj} is the total resistance in loop j, and R_{kj} is the total resistance in common to loop j and loop k. If the loops are all drawn in the same

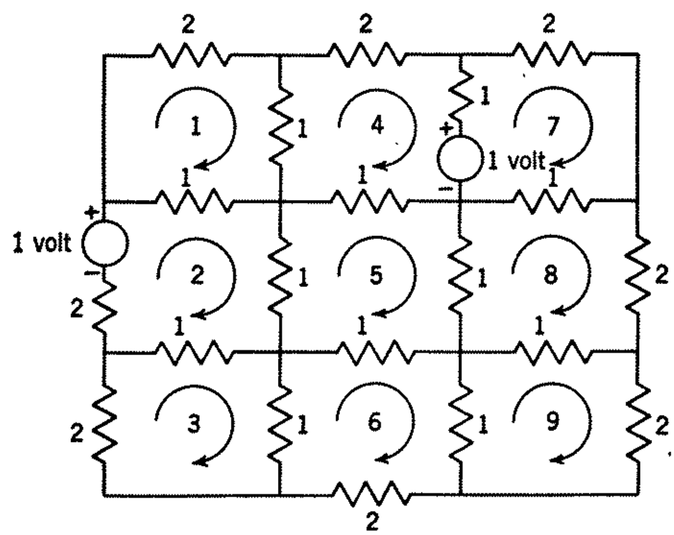


Fig. 3-22. Resistive network analyzed in example: values of resistance in ohms.

direction (say clockwise), then R_{jj} is positive and R_{kj} is negative. As an example, consider the network of Fig. 3-22. For this example, the Kirchhoff voltage equations are summarized in the following chart, where the first row is the equivalent of the equation

$$0 = 4i_1 - i_2 + 0i_3 - i_4 + 0i_5 + 0i_6 + 0i_7 + 0i_8 + 0i_9$$

Eq.	Voltage		Coefficient of i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9
1	0	=	4	−1	0	−1	0	0	0	0	0
2	1	=	−1	5	−1	0	−1	0	0	0	0
3	0	=	0	−1	4	0	0	−1	0	0	0
4	−1	=	−1	0	0	5	−1	0	−1	0	0
5	0	=	0	−1	0	−1	4	−1	0	−1	0
6	0	=	0	0	−1	0	−1	5	0	0	−1
7	1	=	0	0	0	−1	0	0	4	−1	0
8	0	=	0	0	0	0	−1	0	−1	5	−1
9	0	=	0	0	0	0	0	−1	0	−1	4

Several observations of importance can be made from this chart. (1) The elements of the chart are the elements of the system determinant. (2) The elements of the principal diagonal are positive; all others are negative or zero. (3) There is a *symmetry* about the principal diagonal. This symmetry and the sign rule always apply when loops are drawn in a common direction. This characteristic is of value in checking equations.

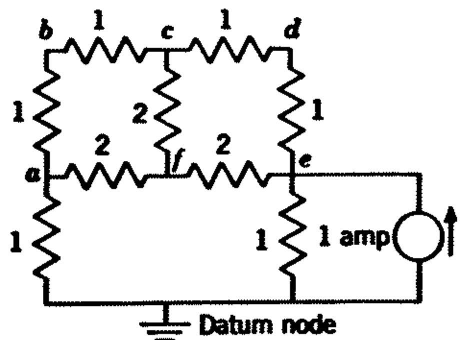


Fig. 3-23. Resistive network analyzed on node basis in example: values of resistance in ohms.

The second example will illustrate formation of node equations for resistive networks. A network to be analyzed on the node basis is shown in Fig. 3-23. The chart equivalent of the six node-pair voltage equations is shown below.

Eq. for node:	Current	Coefficient of v_a	v_b	v_c	v_d	v_e	v_f
a	0	5/2	−1	0	0	0	−1/2
b	0	−1	2	−1	0	0	0
c	0	0	−1	5/2	−1	0	−1/2
d	0	0	0	−1	2	−1	0
e	1	0	0	0	−1	5/2	−1/2
f	0	−1/2	0	−1/2	0	−1/2	3/2

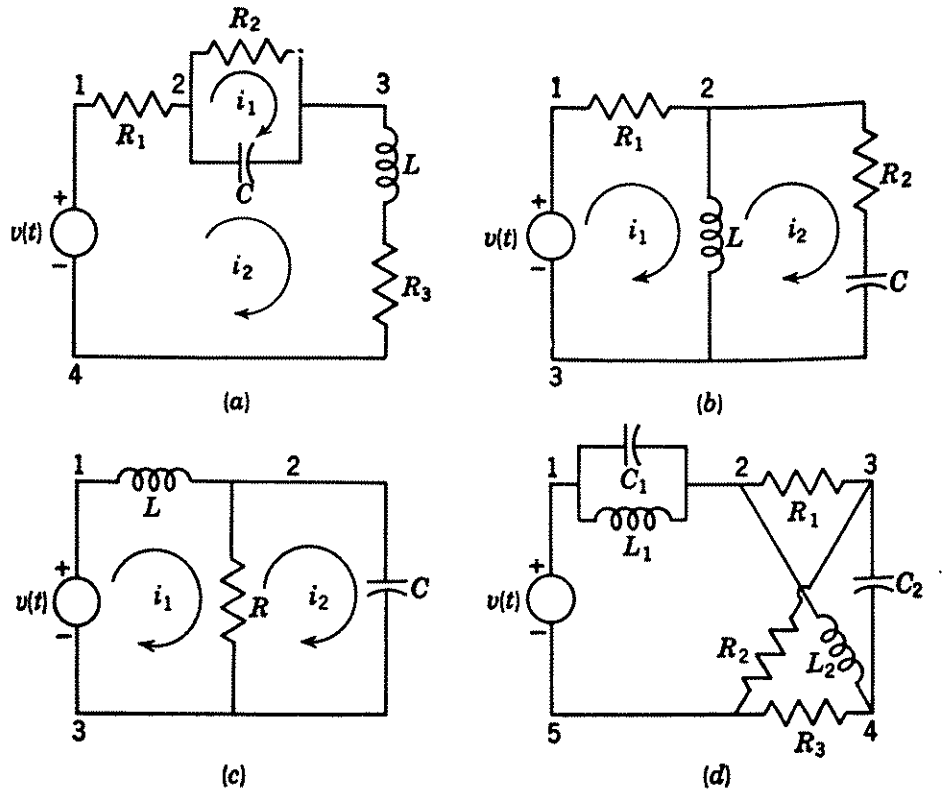
FURTHER READING

For discussions of the formulation of equilibrium equations for networks, see Johnson's *Mathematical and Physical Principles of Engineering Analysis* (McGraw-Hill Book Co., Inc., New York, 1944), pp. 45–67, or *Electric Circuits* by the Electrical Engineering Staff at MIT (John Wiley & Sons, Inc., New York, 1940), pp. 112–138. More advanced treatments are contained in Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 25–49, and in Weber's *Linear Transient Analysis* (John Wiley & Sons, Inc., New York, 1954), Chap. 2. The principle of duality is discussed in many texts, for example those by Johnson and by Gardner and Barnes cited above.

On the subject of writing circuit equations for magnetically coupled circuits, see Kerchner and Corcoran, *Alternating-Current Circuits* (John Wiley & Sons, Inc., New York, 1951), pp. 222–230, or LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 102–110. Further discussion of determinants may be found in many texts in mathematics, for example in Pipes' *Applied Mathematics for Engineers and Physicists* (McGraw-Hill Book Co., Inc., New York, 1946), pp. 69–76, in Wylie's *Advanced Engineering Mathematics* (McGraw-Hill Book Co., Inc., New York, 1951), pp. 573–579, and in Guilleman's *The Mathematics of Circuit Analysis* (John Wiley & Sons, Inc., New York, 1949), Chap. 1.

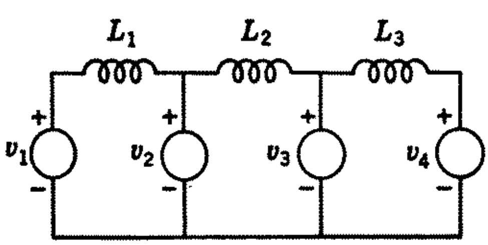
PROBLEMS

3-1. For the four networks shown in the figures, formulate the Kirchhoff voltage equations. For parts a, b, and c, use the loops indicated; for part d, select four appropriate loops.

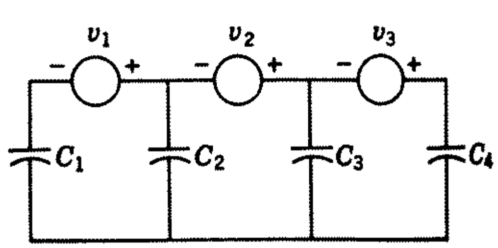


Prob. 3-1.

3-2. In the network shown, we are to write equations that will permit the currents in the inductors to be found. How many *simultaneous* differential equations are required to describe the system? Write the equilibrium equations on the loop basis. Discuss.



Prob. 3-2.

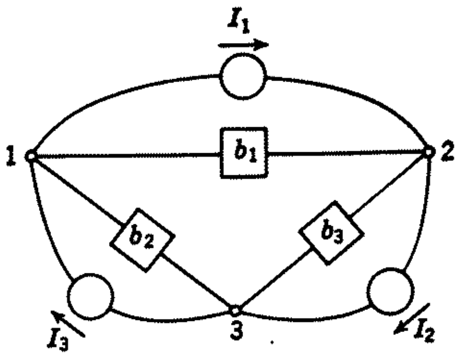


Prob. 3-3.

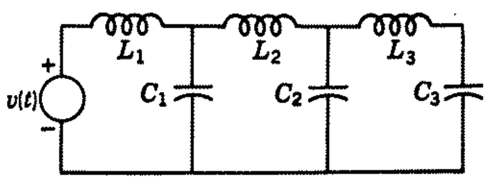
3-3. In the network shown, the problem is to find the current through the capacitors. How many *simultaneous* differential equations are required to describe the system? Write the equilibrium equations on the loop basis. Compare conclusions with those found for Prob. 3-2.

3-4. Formulate a set of node equations to describe the network shown in the figure.

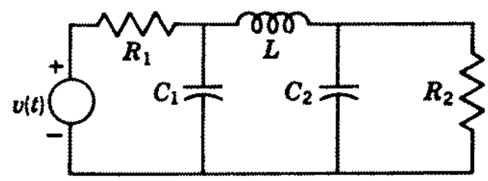
3-5. The network shown in the figure is known as a ladder network (because of its physical appearance). Formulate a set of differential equations on the loop basis. Suppose that the ladder is extended indefinitely by alternately adding inductors and capacitors. Compare the number of loops and nodes for each addition to the ladder.



Prob. 3-4.



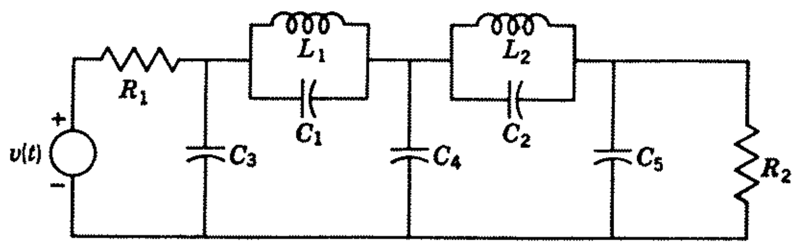
Prob. 3-5.



Prob. 3-6.

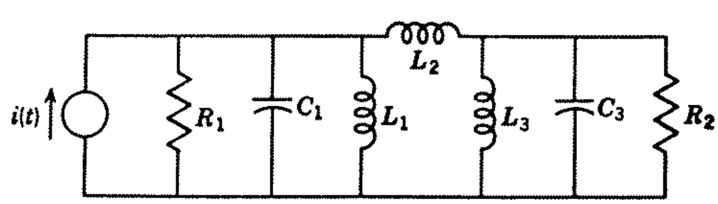
3-6. When the values of the parameters are properly selected, the network shown above is called a Butterworth low-pass filter. Formulate a set of differential equations on the basis (loop or node) that results in the smaller number of simultaneous differential equations.

3-7. The network shown in the figure is of a type designed by the Darlington insertion-loss method. Repeat Prob. 3-6 for this network.



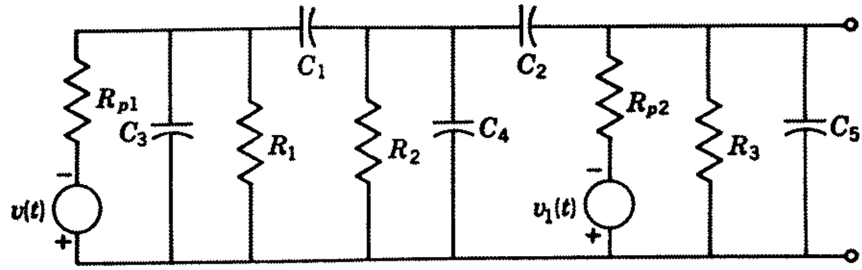
Prob. 3-7.

3-8. The network shown in the figure represents the *interstage network* of some vacuum tube amplifiers. Repeat Prob. 3-6 for this network.



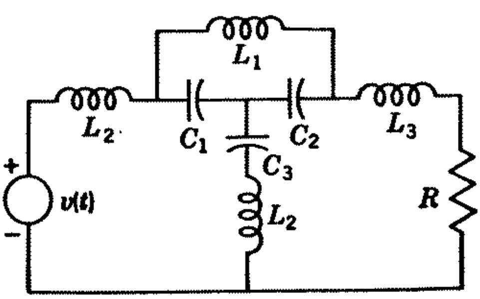
Prob. 3-8.

3-9. The network shown in the figure represents the *equivalent network* of a *two-stage vacuum tube amplifier*. Repeat Prob. 3-6 for this network.

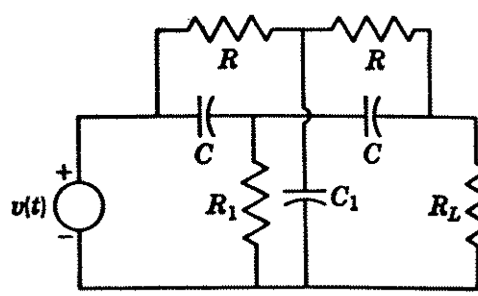


Prob. 3-9.

3-10. The network of this problem represents a *bridged-T filter* network (the inductor forms the bridge across the T). Repeat Prob. 3-6 for this network.



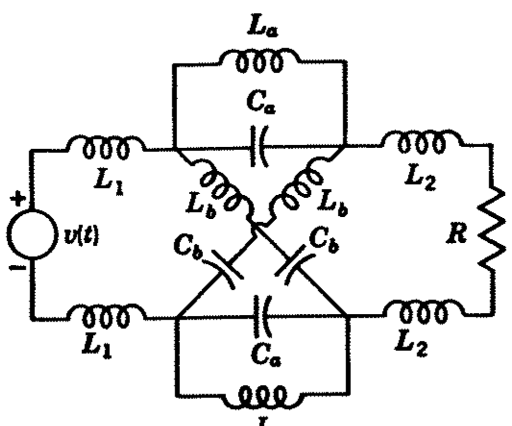
Prob. 3-10.



Prob. 3-11.

3-11. For the *double-T* network shown in the figure, repeat Prob. 3-6.

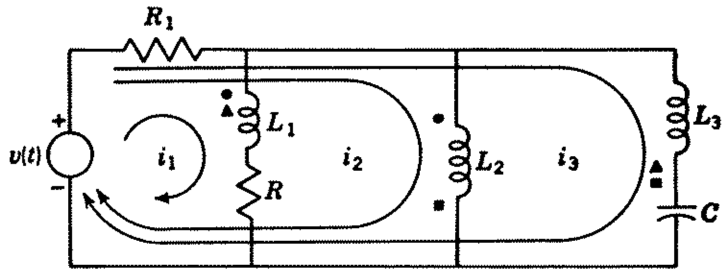
3-12. The network shown in the figure is a *symmetrical lattice filter.* Repeat Prob. 3-6 for this network.



Prob. 3-12.

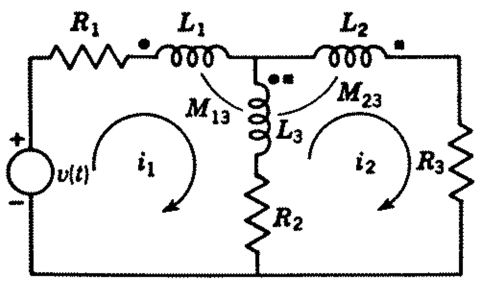
3-13. Consider two magnetically coupled coils with current in each coil. Show that if the currents are in such a direction that the two fluxes *aid*, the sign of M is positive, while if the fluxes *oppose*, the sign of M is negative.

3-14. The circuit shown below is identical to that used in Example 4, Fig. 3-10, but the loops are chosen differently than in the example. For this network formulate the differential equations on the loop basis. Compare the number of terms in the equations that result with those found in Example 4.



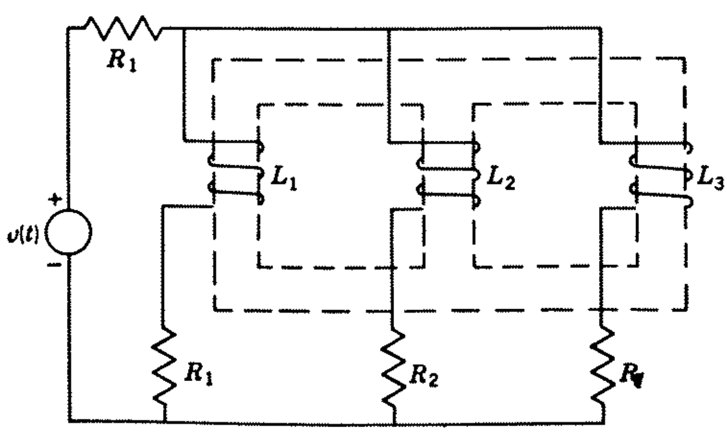
Prob. 3-14.

3-15. A network with mutual inductance is shown below, with the coil winding sense indicated by dots. Write the Kirchhoff voltage equations for this network. Note that $M_{12} = 0$.



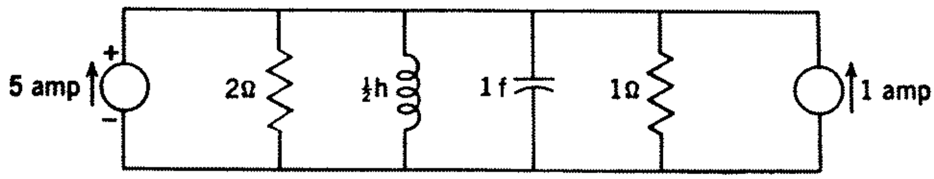
Prob. 3-15.

3-16. Write the loop basis network equations for the system shown in the accompanying figure.



Prob. 3-16.

3-17. For the network shown, (a) find the dual and (b) give element values on the schematic to make the network equations for the duals have identical coefficients.



Prob. 3-17.

3-18. Find the dual of the network of Prob. 3-5.

3-19. Find the dual of the network of Prob. 3-8.

3-20. Find the dual of the network of Prob. 3-9.

3-21. Find the dual of the network of Prob. 3-10.

3-22. The network of Prob. 3-12 appears to be nonplanar (in which case it does not have a *dual*). For this particular network, however, the crossover point can be removed so that the network is planar. (a) Draw the equivalent planar network. (b) Find the dual of the planar network.

3-23. Solve the following system of equations for i_1, i_2, and i_3, using determinants.

$$\begin{aligned} 3i_1 - 2i_2 + 0i_3 &= 5 \\ -2i_1 + 9i_2 - 4i_3 &= 0 \\ 0i_1 - 4i_2 + 9i_3 &= 10 \end{aligned}$$

Answers. $i_1 = 405/159$, $i_2 = 210/159$, $i_3 = 270/159$.

3-24. Solve the following system of equations for the three unknowns, i_1, i_2, and i_3 by determinants.

$$\begin{aligned} 8i_1 - 3i_2 - 5i_3 &= 5 \\ -3i_1 + 7i_2 - 0i_3 &= -10 \\ -5i_1 + 0i_2 + 11i_3 &= -10 \end{aligned}$$

Answers. $i_1 = -295/342$, $i_2 = -615/342$, $i_3 = -445/342$.

3-25. Evaluate the following determinants by minors.

$$\text{(a)} \begin{vmatrix} 2 & -1 & 0 & 0 \\ -1 & 3 & -2 & 0 \\ 0 & -2 & 3 & -1 \\ 0 & 0 & -1 & 2 \end{vmatrix} \qquad \text{(b)} \begin{vmatrix} 1 & -2 & 0 & 3 & 4 \\ -1 & 4 & -1 & 1 & 0 \\ 2 & 0 & 1 & 1 & 3 \\ 4 & -2 & 4 & 2 & -1 \\ 3 & 1 & 3 & -2 & 1 \end{vmatrix}$$

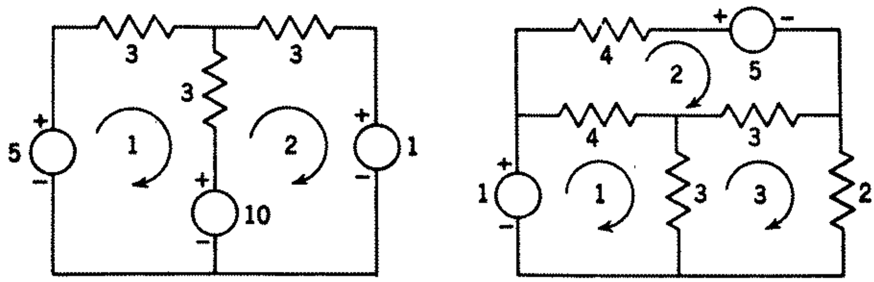
Answers. (a) $+9$, (b) $+133$.

3-26. Consider the equations

$$
\begin{aligned}
3x - y - 3z &= 1\\
x - 3y + z &= 1\\
4x + 0y - 5z &= 1
\end{aligned}
$$

(a) Is (4, 2, 3) a solution? Is $(-1, -1, -1)$ a solution? (b) Can these equations be solved by determinants? Why? (c) What can you conclude regarding the geometry represented by these equations?

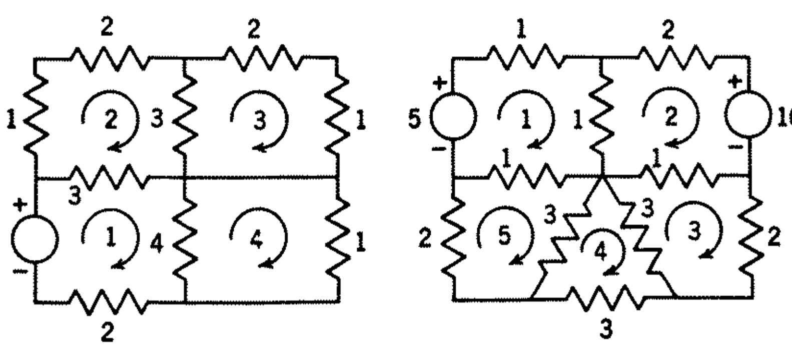
3-27. By inspecting the networks in the accompanying figure (without writing the circuit equations), write the loop basis system determinant. Element values are in ohms and volts.



Prob. 3-27.

Answer. $\begin{vmatrix} 6 & -3 \\ -3 & 6 \end{vmatrix}$; *Answer.* $\begin{vmatrix} 7 & -4 & -3 \\ -4 & 11 & -3 \\ -3 & -3 & 8 \end{vmatrix}$

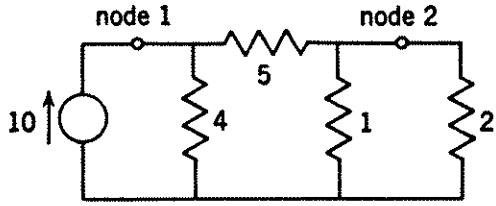
3-28. Repeat Prob. 3-27 for the networks shown in the figure.



Prob. 3-28.

Answers. $\begin{vmatrix} 9 & -3 & 0 & -4 \\ -3 & 9 & -3 & 0 \\ 0 & -3 & 6 & 0 \\ -4 & 0 & 0 & 5 \end{vmatrix}$; $\begin{vmatrix} 3 & -1 & 0 & 0 & -1 \\ -1 & 4 & -1 & 0 & 0 \\ 0 & -1 & 6 & -3 & 0 \\ 0 & 0 & -3 & 9 & -3 \\ -1 & 0 & 0 & -3 & 6 \end{vmatrix}$

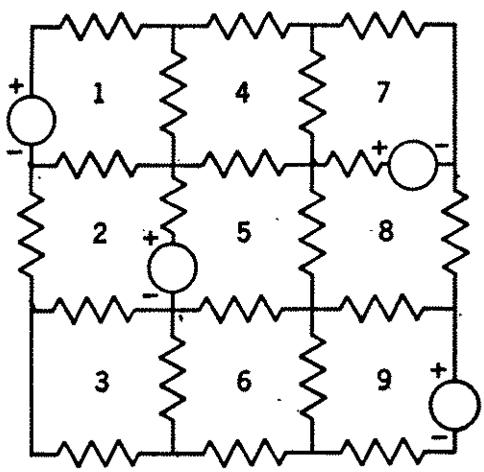
3-29. By inspecting the network shown (without writing the circuit equations), write the node basis system determinant.



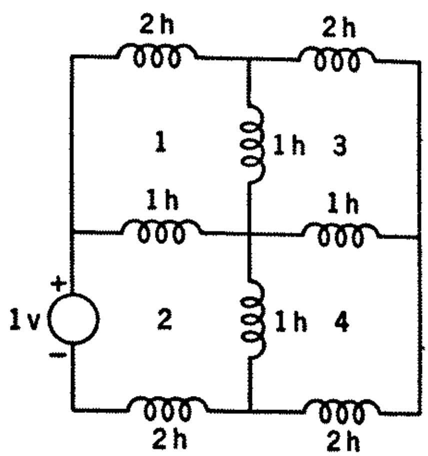
Prob. 3-29.

Answer. $\begin{vmatrix} \dfrac{9}{20} & \dfrac{-1}{5} \\ \dfrac{-1}{5} & \dfrac{17}{10} \end{vmatrix}$

3-30. In the network graph shown in the figure, each branch contains a 1-ohm resistor. Four branches, as shown, contain a 1-volt voltage source. Analyze this network on the loop basis to obtain a set of equations. Simplify by combining like terms in any one equation. The number inside each square is the loop number.



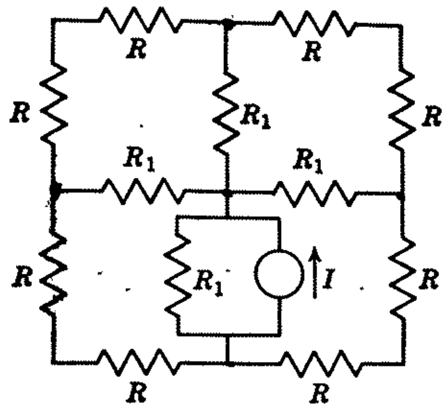
Prob. 3-30.



Prob. 3-31.

3-31. In the network graph shown in the figure, each interior branch contains an inductor of 1 henry and each exterior branch an inductor of 2 henrys. A 1-volt source is located in one branch as shown. The number inside each square is the loop number. Analyze the network on the loop basis to obtain a set of equations. Simplify by combining like terms in any one equation.

3-32. In the network graph shown, $R = R_1 = 1$ ohm, and one branch contains a current source of 1 amp. Analyze the network on the *node* basis to obtain a set of equations. Simplify by combining like terms in any one equation.



Prob. 3-32.

3-33. Repeat Prob. 3-32 with external branch resistors of $R = 2$ ohms and interior branch resistors of $R_1 = 1$ ohm.

CHAPTER 4

FIRST-ORDER DIFFERENTIAL EQUATIONS

4-1. Definitions for differential equations

In this chapter, we will study a number of techniques for the solution of the simplest differential equations, those of first *order* such as

$$a_0 \frac{di}{dt} + a_1 i = 0 \tag{4-1}$$

This equation is of first order because the highest-ordered derivative is the first. Thus differential equations are classified by the highest-ordered derivative they contain. An nth order differential equation may be written

$$a_0 \frac{d^n i}{dt^n} + a_1 \frac{d^{n-1} i}{dt^{n-1}} + \dots + a_{n-1} \frac{di}{dt} + a_n i = v(t) \tag{4-2}$$

for equations of the first *degree.* The *degree* of an equation is the power to which the highest-ordered derivative appears after all possible algebraic reduction.

In Eq. 4-2, i is the *dependent* variable and t is the *independent* variable. When $v(t)$ represents an energy source, it is known as the *forcing function.* The dependent variable i, which is to be found, is called the *response* or *solution.* The differential equation is *linear* if the dependent variable and all its derivatives are of first degree. All other equations are *nonlinear.* A differential equation is said to be *ordinary* if it contains only *total* (and not *partial*) derivatives. For the type of circuits assumed in Chapter 1, the differential equations that describe networks will all be *ordinary, linear differential equations with constant coefficients.* It should be remembered that the techniques we will discuss will not, in general, apply to nonlinear differential equations.

Equation 4-2 is said to be *homogeneous* when $v(t) = 0$; if $v(t)$ is not zero the equation is *nonhomogeneous.*

4-2. General and particular solutions

In electrical problems, the network is assumed to be initially in a known state with all voltages and currents fixed. At an instant of time designated $t = 0$, the system is altered in a manner that can be represented by the opening or closing of one or more switches. The objective of analysis is to obtain mathematical equations for current, voltage, charge, etc. in terms of time measured from the instant equilibrium was altered by the switching.

In the network shown in Fig. 4-1, the switch K is changed from position 1 to position 2 at the reference time $t = 0$.* After the switching has taken place, the Kirchhoff voltage equation is

$$L\frac{di}{dt} + Ri = 0 \tag{4-3}$$

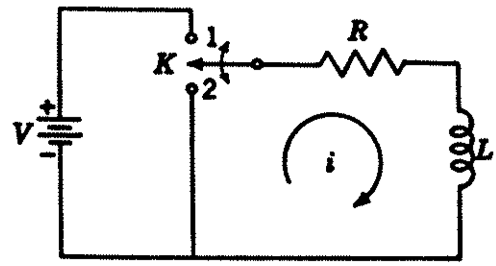


Fig. 4-1. RL circuit.

This is a first-order linear differential equation with constant coefficients. It can be solved if the variables can be separated. This may be accomplished by rearranging Eq. 4-3 in the form

$$\frac{di}{i} = -\frac{R}{L}\,dt \tag{4-4}$$

With the variables separated, the equation can be integrated to give

$$\ln i = -\frac{R}{L}t + K \tag{4-5}$$

where ln designates that the logarithm is to the base $e = 2.718 \ldots$. To simplify the form of this equation, the constant K is redefined in terms of the logarithm of another constant as

$$K = \ln k \tag{4-6}$$

Equation 4-5 may then be written

$$\ln i = \ln e^{-Rt/L} + \ln k \tag{4-7}$$

since, by the definition of a logarithm, $\ln e^x = x$, or $\log_{10} 10^x = x$. Also, from logarithms we know that

$$\ln y + \ln z = \ln yz \tag{4-8}$$

so that Eq. 4-7 may be written

$$\ln i = \ln (ke^{-Rt/L}) \tag{4-9}$$

With the equation in this form, the antilogarithm may be taken to give,

$$i = ke^{-Rt/L} \tag{4-10}$$

* It is assumed that the switch is a "make-before-break" type and that the transition from position 1 to position 2 does not cause an interruption of the current i.

Redefining the constant K as the logarithm of another constant has simplified the form of the solution. Equation 4-10 is the network response or solution. This solution is free of derivatives and expresses the relationship between the dependent and independent variables. That it is the solution can be verified by substituting Eq. 4-10 into Eq. 4-3.

In the form of Eq. 4-10, the solution is known as the *general solution.* If the constant of integration is evaluated, the solution is a *particular solution.* The general solution applies to any number of situations. A particular solution fits the specifications of a particular problem.

To evaluate the constant k, we must know something new about the problem, such as any pair of values of i and t. In this particular problem, we know that the current after switching has taken place must be just the same as before switching because of the inductor in the circuit.* Thus at $t = 0$, we know that the current has the value

$$i(0) = \frac{V}{R} \tag{4-11}$$

This value is known as the *initial condition* of the circuit. Substituting this required condition into Eq. 4-10 gives

$$\frac{V}{R} = ke^0 = k \tag{4-12}$$

The particular solution of this example becomes

$$i = \frac{V}{R}\,e^{-Rt/L} \tag{4-13}$$

4-3. The integrating factor

Consider a nonhomogeneous equation written

$$\frac{di}{dt} + Pi = Q \tag{4-14}$$

where P is a constant and Q may be a function of the independent variable t or a constant. The equation is not altered if every term is multiplied by the same factor. Suppose that we multiply Eq. 4-14 by the quantity e^{Pt}, which will be known as an *integrating factor.*† There results

$$e^{Pt}\frac{di}{dt} + Pie^{Pt} = Qe^{Pt} \tag{4-15}$$

* This is an application of the principle of constant flux linkages discussed in Art. 1-7.

† If P is a function of time, the proper integrating factor is $e^{\int P\,dt}$. See Prob. 4-5.

That this multiplication by a factor "pulled out of the hat" has made possible the solution of Eq. 4-14 can be recognized by recalling the equation for the derivative of a product:

$$d(xy) = x\,dy + y\,dx \tag{4-16}$$

By letting $x = i$ and $y = e^{Pt}$, we have

$$\frac{d}{dt}(ie^{Pt}) = e^{Pt}\frac{di}{dt} + ie^{Pt}P \tag{4-17}$$

which is the left-hand side of Eq. 4-15; thus we have

$$\frac{d}{dt}(ie^{Pt}) = Qe^{Pt} \tag{4-18}$$

This equation may be integrated to give

$$ie^{Pt} = \int Qe^{Pt}\,dt + K \tag{4-19}$$

or

$$i = e^{-Pt}\int Qe^{Pt}\,dt + Ke^{-Pt} \tag{4-20}$$

The first term in Eq. 4-20 is known as the *particular integral;* the second is known as the *complementary function.* Note that the particular integral does not contain the arbitrary constant, and the complementary function does not depend on the forcing function Q.

For any network problem, P will be a positive constant determined by the network parameters, and Q will be either the forcing function or a derivative of the forcing function. In the limit, the complementary function must approach zero, because P is a positive constant; that is

$$\lim_{t\to\infty} Ke^{-Pt} = 0 \tag{4-21}$$

Thus the value of i as time approaches infinity is

$$i(\infty) = \lim_{t\to\infty} i(t) = \lim_{t\to\infty} e^{-Pt}\int Qe^{Pt}\,dt \tag{4-22}$$

When the particular integral does not approach zero in the limit, its value at $t = \infty$ is spoken of as the *steady-state value.* For this case, the particular integral must contain no exponential factor or otherwise it would reduce to zero. In electrical engineering, the steady-state values most frequently encountered have the forms

$$i = A\sin(\omega t + \phi) \quad \text{and} \quad i = \text{a constant} \tag{4-23}$$

Let the general solution of Eq. 4-20 be written as the sum of the two parts of the solution, letting i_P be the particular integral and i_C be the complementary function; thus

$$i = i_P + i_C \tag{4-24}$$

If i_P has either of the forms of Eq. 4-23, it may be written as a steady-state value, designated i_{ss}. A convention has been established for calling the remaining term i_C the *transient* portion of the solution. By this convention, the response is made up of two separate parts:

$$i = i_{ss} + i_t \tag{4-25}$$

The steady-state value is regarded as having been established at $t = 0$, and the transient must adjust itself, mathematically, to account for the response at $t = 0$ and all other times. This is an arbitrary division of the solution which nevertheless has utility as a conceptual aid. The division of solution is made purely by convention; the individual electron in a current has no way of knowing whether it is in the transient or the steady-state division of the current.

Example 1

To illustrate the transient and steady-state portions of the solution to a problem, consider the network of Fig. 4-1 with the switch moved from position 2 to position 1 at $t = 0$. The Kirchhoff voltage equation is, after division by L,

$$\frac{di}{dt} + \frac{R}{L} i = \frac{V}{L}$$

Comparing this equation to Eq. 4-14, we see that

$$P = \frac{R}{L} \quad \text{and} \quad Q = \frac{V}{L}$$

The solution to this equation is given as Eq. 4-20 which becomes for this problem,

$$i = e^{-Rt/L} \int \frac{V}{L} e^{Rt/L}\, dt + Ke^{-Rt/L}$$

Evaluating the integral, we obtain

$$i = \frac{V}{R} + Ke^{-Rt/L}$$

as the general solution. If the current in the network being considered is zero before the switching action, it must be zero afterward because of the inductor. The requirement that $i(0) = 0$ leads to the particular solution

$$i = \frac{V}{R}(1 - e^{-Rt/L}) \tag{4-26}$$

The steady-state and transient divisions of this current are shown in

Fig. 4-2 along with their sum or the actual current. The steady-state portion (V/R) is established at $t = 0$, and the transient term is adjusted such that there is zero current at $t = 0$.

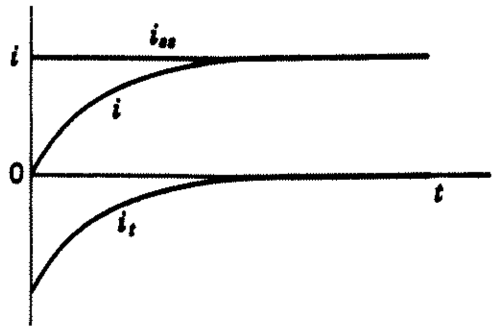


Fig. 4-2. Transient and steady-state parts of the solution of Example 1.

4-4. Time constants

The particular solution of Eq. 4-3 given by Eq. 4-13 may be written in a nondimensional form as

$$\frac{i}{I_0} = e^{-t/T} \tag{4-27}$$

where I_0 is the initial value of current at $t = 0$ and $T = (L/R)$ is the *time constant* of the system. The form of Eq. 4-27 is the solution of all homogeneous first-order differential equations, where I_0 and T have different values for different problems. The physical significance attached to the time constant is of great importance in electrical engineering. When $t = T$, by Eq. 4-27,

$$\frac{i(T)}{I_0} = e^{-1} = 0.37 \tag{4-28}$$

or

$$i(T) = 0.37 I_0 \tag{4-29}$$

In other words, the current decreases to 37% of its initial value in one time constant. By a similar computation, it can be shown that the current decreases to approximately 2% of its initial value in four time constants. A plot of i/I_0 against t/T is shown in Fig. 4-3.

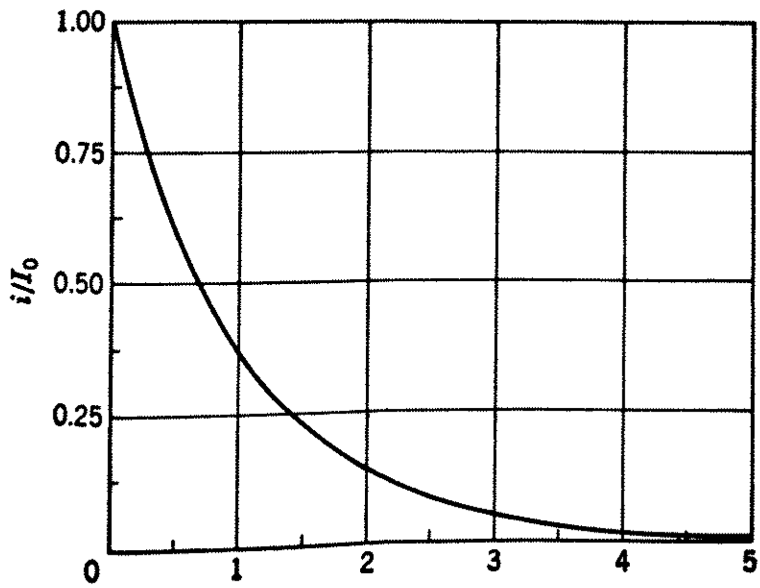


Fig. 4-3. Normalized exponential curve for $e^{-t/T}$.

The solution of the first-order nonhomogeneous differential equation with a constant forcing function is of the form of Eq. 4-26, which may be written in nondimensional form as

$$\frac{i}{I_0} = 1 - e^{-t/T} \tag{4-30}$$

This function is plotted in Fig. 4-4. When $t = T$,

$$i(T) = (1 - 0.37)I_0 = 0.63I_0 \tag{4-31}$$

or the current has reached 63% of its steady-state value in one time constant. Similarly, the current will increase to approximately 98% of its final value in four time constants.

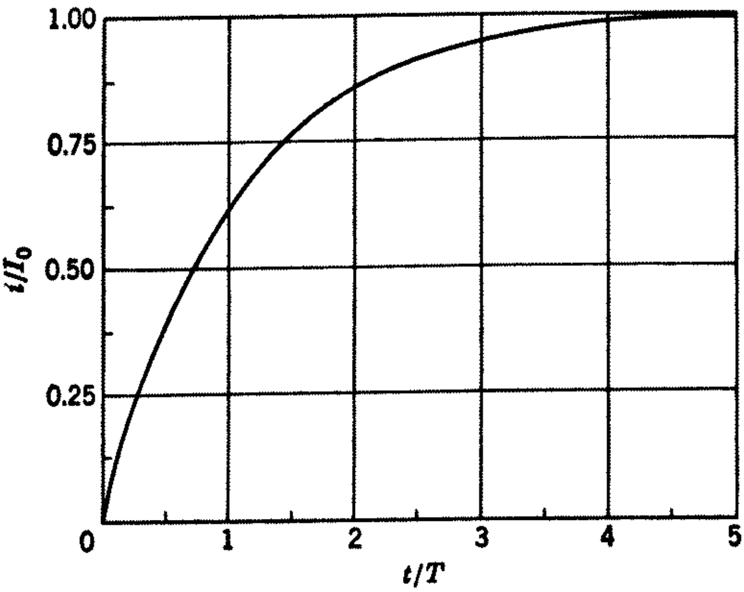


Fig. 4-4. Normalized exponential curve for $(1 - e^{-t/T})$.

The time constant is useful in comparing the behavior of one system with that of another. It is not possible to compare times at which the transient disappears (or reaches its steady state) since, mathematically at least, this requires infinite time. However, the time interval for an exponential function to decrease to 37% of its initial value (or increase to 63% of its final value) is conveniently measured and used as a standard for comparison. As an example, consider a series RC circuit which has a general solution,

$$i = I_0 e^{-t/RC}$$

The time constant for the circuit is $T = RC$. Suppose that R has the value of 100 ohms and C is 1000 $\mu\mu$f; then $T = 100 \times 1000 \times 10^{-12} = 0.1$ μsec. However, if $R = 1000$ megohms and $C = 1$ μf, then $T = 1000$ sec, or 17 min. For one combination of R and C, the current would decrease to 37% of the initial value in the small time of 0.1 μsec; for the other, the current would require about 17 min to decrease to 37% of the initial value.

In experimentally recording a transient, the accuracy of measurement is often of the order of 1 or 2%. For this reason, a transient is sometimes assumed to have disappeared when it reaches 2% of the final value (as accurately as can be determined). Since the time to reach 2% of the final value (or 98% in the case of an increasing exponential) is *four time constants*, it is often assumed that a transient disappears in four time constants. This basis is sometimes used to measure the time constant of a system.

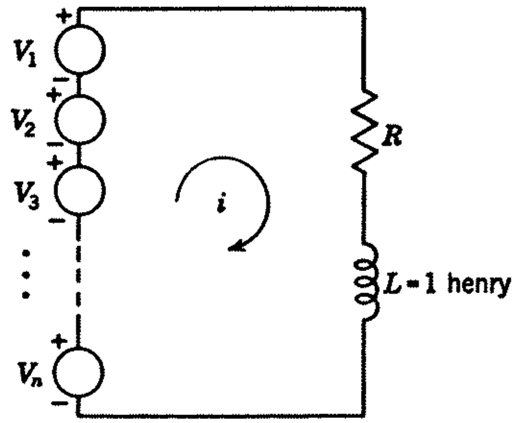


Fig. 4-5. *RL* series circuit with n voltage sources illustrating the superposition theorem.

4-5. The principle of superposition

A series *RL* circuit with n series voltage sources is shown in Fig. 4-5. To simplify the form of the differential equation, L is taken to be 1 henry. By the Kirchhoff voltage law, we have

$$\frac{di}{dt} + Ri = v_1(t) + v_2(t) + \ldots + v_n(t) \tag{4-32}$$

In terms of the general solution of a first-order nonhomogeneous differential equation, given by Eq. 4-20, $P = R$, a constant, $Q = v_1 + v_2 + \ldots + v_n$ and the solution is

$$i = e^{-Rt} \int (v_1 + v_2 + \ldots + v_n)e^{Rt}\, dt + Ke^{-Rt} \tag{4-33}$$

As we have observed before, the particular integral depends on the nature of the forcing function voltages and, for this reason, is given the name *forced response*. On the other hand, the complementary function does not depend on the forcing function (except that K is fixed by the magnitude of the forcing function at $t = 0$ and circuit conditions existing at that time). The complementary function is given the name *free response*. The total response can be thought of as made up of two parts—forced and free. We now have three sets of terms defining the two parts of the solution of the differential equation.

particular integral	and	complementary function
steady-state solution	and	transient solution
forced response	and	free response

All three sets are used in electrical engineering literature, and will be used throughout the text although our preference will be for the last set.

Next, assume that all voltage sources except $v_1(t)$, are removed and replaced by short circuits. The response under this condition is

$$i_1 = e^{-Rt} \int v_1 e^{Rt}\, dt + k_1 e^{-Rt} \tag{4-34}$$

If this experiment is repeated for each generator of the circuit of Fig. 4-5, the response will be similar to that given in Eq. 4-34. Suppose that the currents found in this manner are added together. This sum may be written

$$\begin{aligned} i_1 + i_2 + \ldots + i_n = e^{-Rt} [\textstyle\int v_1 e^{Rt}\, dt + \int v_2 e^{Rt}\, dt \\ + \ldots + \textstyle\int v_n e^{Rt}\, dt] + (k_1 + k_2 + \ldots + k_n) e^{-Rt} \end{aligned} \tag{4-35}$$

Since each k is so far an arbitrary constant, we may set

$$K = k_1 + k_2 + \ldots + k_n \tag{4-36}$$

Because R is a constant, the integral terms in Eq. 4-35 may be combined to give

$$i_1 + i_2 + \ldots + i_n = e^{-Rt} \int (v_1 + v_2 + \ldots + v_n) e^{Rt}\, dt + Ke^{-Rt}$$

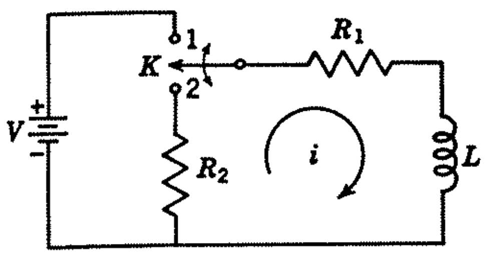
This equation is identical with Eq. 4-33 which was found for the combined forcing functions. In summary, the total response of a linear network is identical to that found by considering each voltage source alone with all other sources removed and replaced by short circuits and then summing the individual responses. This is the application of a general rule known as the *principle of superposition.* This principle holds for voltage sources arbitrarily located in more complex networks and is of great importance in linear network theory. The fact that it does not hold for nonlinear systems is the root of the great difficulty in analyzing such systems.

FURTHER READING

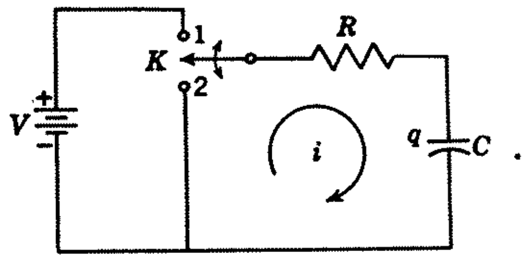
First-order differential equations are discussed by Fich in *Transient Analysis in Electrical Engineering* (Prentice-Hall, Inc., New York, 1951) under the heading of "Classical Solution of Single-Energy Transients," pp. 36–66. See also Kurtz and Corcoran's *Introduction to Electric Transients* (John Wiley & Sons, Inc., New York, 1945), pp. 15–30.

PROBLEMS

4-1. In the circuit shown in the figure, the switch is changed from position 1 to position 2 at $t = 0$, a steady-state current having previously been established in the RL circuit. Find the particular solution for the current in the circuit. *Answer.* $i = (V/R_1)e^{-(R_1+R_2)t/L}$



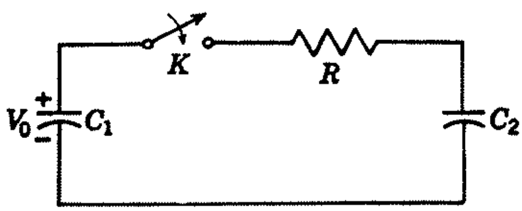
Prob. 4-1.



Prob. 4-2.

4-2. Replace the inductor in Fig. 4-1 with a capacitor. (a) Write the integral equation for the current in the system after the switch is in position 2, assuming that the capacitor was charged to a voltage equal to that of the source while the switch was in position 1. (b) Write the differential equation for the charge under the same conditions as (a). (c) Solve for the charge as a function of time and evaluate the arbitrary constant. *Answer to (c).* $q = CVe^{-t/RC}$

4-3. In the circuit shown, the capacitor C_1 is charged to a voltage V_0 and at $t = 0$ the switch is closed. Solve for the charge as a function of time.



Prob. 4-3.

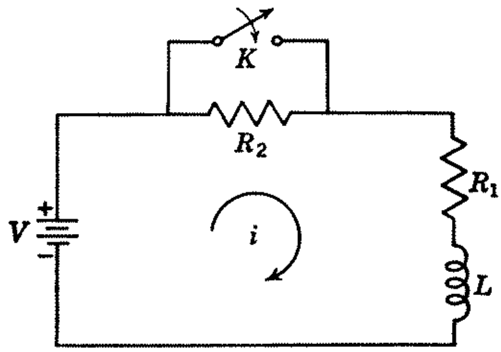
4-4. In the circuit of Prob. 4-2, suppose that the switch is changed from position 2 to position 1 at $t = 0$ and that while in position 2 there was no charge on the capacitor. Find the charge as a function of time.

4-5. We wish to multiply the differential equation

$$\frac{di}{dt} + P(t)i = Q(t)$$

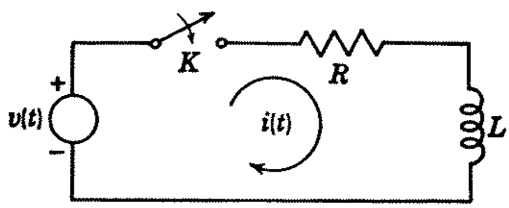
by an "integrating factor" R such that the left-hand side of the equation equals the derivative $d(Ri)/dt$. (a) Show that the required integrating factor is $R = e^{\int P\,dt}$. (b) Using this integrating factor, find the solution to the differential equation that corresponds to Eq. 4-20.

4-6. In the circuit shown in the accompanying figure, the switch K is closed at $t = 0$, a steady-state having previously been attained. Solve for the current in the circuit as a function of time.



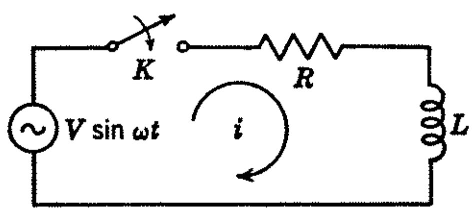
Prob. 4-6.

4-7. In the circuit shown, the voltage source follows the law $v(t) = Ve^{-\alpha t}$, where α is a constant. The switch is closed at $t = 0$. (a) Solve for the current assuming that $\alpha \neq R/L$. (b) Solve for the current when $\alpha = R/L$. Suggestion: Make use of l'Hospital's rule for indeterminant forms.



Prob. 4-7.

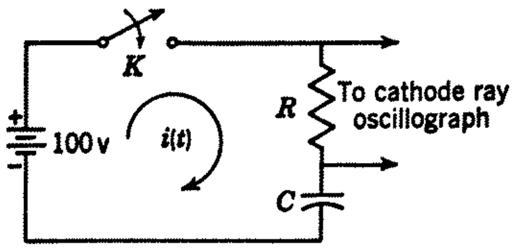
4-8. In the circuit shown, the switch is closed at $t = 0$ connecting a voltage source $v(t) = V \sin \omega t$ to a series RL circuit. For this system, solve for the response $i(t)$.



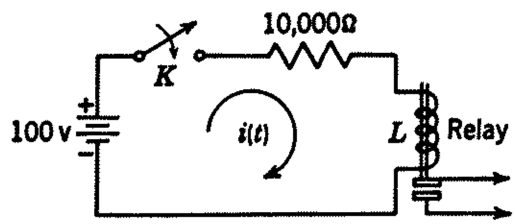
Prob. 4-8.

4-9. Show that the tangent to the curve $i = I_0 e^{-t/T}$ at $t = 0$ intersects the time axis at $t = T$. This will show that if the current decreased at the initial rate, it would be reduced to zero value in one time constant. Similarly, show that the tangent to the curve $i = I_0(1 - e^{-t/T})$ at $t = 0$ intersects that line $i = I_0$ at time $t = T$.

4-10. In the network shown, the switch K is closed at $t = 0$. The current waveform is observed with a cathode ray oscillograph. The initial value of the current is measured to be 10 ma. The transient disappears in 0.1 sec. Find (a) the value of R, (b) the value of C, and (c) the equation of $i(t)$. *Answers.* (a) $R = 10^4$ ohms, (b) $C = 2.5\ \mu f$, (c) $i = 10^{-2}e^{-40t}$.



Prob. 4-10.



Prob. 4-11.

4-11. The circuit shown in the accompanying figure consists of a resistor and a relay with inductance L. The relay is adjusted so that it is actuated when the current through the coil is 0.008 amp. The switch K is closed at $t = 0$, and it is observed that the relay is actuated when $t = 0.1$ sec. Find: (a) the inductance L of the coil, (b) the equation of $i(t)$ with all coefficients evaluated. *Answers.* (a) $L = 620$ henrys, (b) $i = 0.01(1 - e^{-16t})$ amp.

4-12. A switch is closed at $t = 0$, connecting a battery of voltage V with a series RC circuit. (a) Determine the ratio of energy delivered to the capacitor to the total energy supplied by the source as a function of time. (b) Show that this ratio approaches 0.50 as $t \to \infty$.

CHAPTER 5

INITIAL CONDITIONS IN NETWORKS

5-1. Initial conditions in individual elements

In the last chapter, we found that the general solution of a first-order differential equation contained an unknown designated an arbitrary constant. For differential equations of higher order, the pattern will develop that the number of arbitrary constants equals the equation order. If the unknown arbitrary constants are to be evaluated for particular solutions, other things must be known about the network described by the differential equation. We must form a set of simultaneous equations, one of which is the general solution, with additional equations to total the number of unknowns. The additional equations are conveniently given as values of voltage, current, charge, etc., or derivatives of these quantities at the instant network equilibrium is altered by switching action, $t = 0$. Conditions existing at this instant are known as *initial conditions*.

Before the switching action that alters network equilibrium, the elements of the network might have voltages across their terminals or currents through them as a consequence of past history of the driving forces in the network. To evaluate initial voltages or currents, we must determine how each voltage and current changes when the network is altered.

In many problems, conditions assumed to exist before switching action takes place were, in turn, established by switching action at some remote time in the past. Such voltages and currents in the network are said to be in the *steady state*.

We assume that switches act in zero time. To differentiate between the time immediately before and immediately after the closing of a switch, we will use $-$ and $+$ signs. Thus conditions existing just before the switch is operated will be designated as $i(0-)$, $v(0-)$ etc., conditions after as $i(0+)$, $v(0+)$, etc.

Before analyzing initial conditions in networks, we will study the action of each different element at the instant equilibrium is altered.

The Resistor. In the ideal resistor, current and voltage are related by Ohm's law, $v = Ri$. If a step input of voltage, shown in Fig. 5-1, is applied to a resistor network, the current will have the same waveform, altered by the scale factor $(1/R)$. The current through a resistor may change instantaneously if the voltage changes instantaneously. Similarly, voltage may change instantaneously if current changes instantaneously.

The Inductor. It was concluded in Art. 1-7, that the current cannot change instantaneously in a system of constant inductance. Thus closing a switch to connect an inductor to a source of energy will not cause current to flow at the initial instant, and the inductor will act as if it were an *open circuit* independent of the voltage at the terminals. If a current of value I_0 flows in the inductor at the instant switching takes place, that current will continue to flow. For the initial instant, the inductor can be thought of as a current source of I_0 amp.

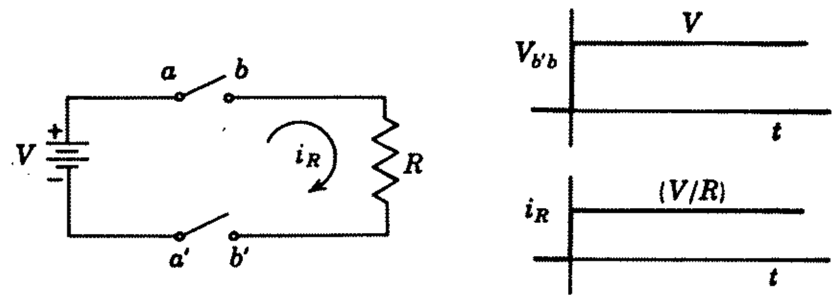


Fig. 5-1. Current and voltage relationships in a purely resistive element.

The Capacitor. In Art. 1-6, proof was offered that the voltage cannot change instantaneously in a system of fixed capacitance. If an uncharged capacitor is connected to an energy source, a current will flow instantaneously, since the capacitor will be equivalent to a *short circuit.* This follows because voltage and charge are proportional in a capacitive system, $v = q/C$, so that zero charge corresponds to zero voltage (or a short circuit). With an initial charge in the system, the capacitor is equivalent to a voltage source of value $V_0 = q_0/C$, where q_0 is the initial charge.

These conclusions are summarized in Fig. 5-2. A similar chart of final conditions for the special case of constant voltage sources is shown in Fig. 5-3. These equivalent circuits are derived from the relationships

$$v_L = L\frac{di}{dt} \quad \text{and} \quad i_C = C\frac{dv}{dt}$$

the derivatives having zero value in each case for invariant voltage sources. The equivalent circuits for final conditions for L and C are opposite to those for the initial conditions for these elements.

It is not always possible to interrupt a current instantaneously in a network by opening a switch. If an attempt is made to open a switch to disconnect an inductor from a voltage source, an arc will be established across the switch to permit the current to flow until the energy of the magnetic field is spent.

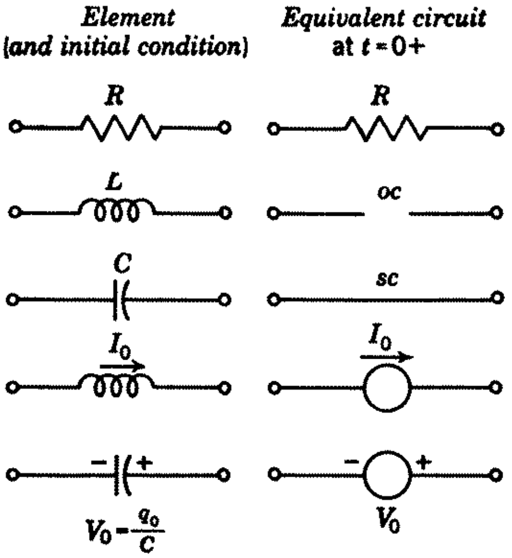


Fig. 5-2. Initial condition equivalent circuits for the elements.

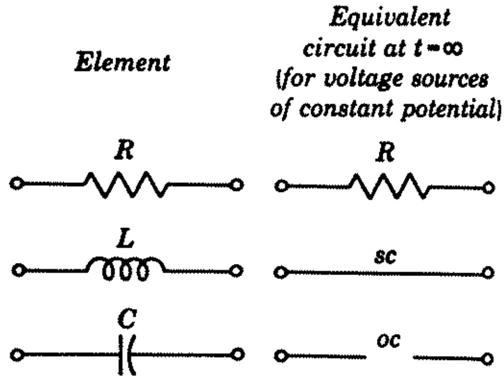


Fig. 5-3. Final condition equivalent circuits for the elements for voltage sources of constant potential.

5-2. Geometrical interpretation of derivatives

Consider the differential equation that describes an *RL* circuit connected to a constant voltage source:

$$L\frac{di}{dt} + Ri = V \tag{5-1}$$

This equation may be arranged in the form

$$\frac{di}{dt} = \frac{1}{L}(V - iR) \tag{5-2}$$

to show the relationship that must exist between current and the time derivative of current. If the switch connecting the voltage source to the circuit is closed at $t = 0$, the current in the system at $t = 0$ must be zero. From Eq. 5-2, the initial value of the derivative is

$$\frac{di}{dt}(0+) = \frac{V}{L} \tag{5-3}$$

Now the quantity di/dt is the slope of the required plot of current as a function of time. Equation 5-3 tells us that this slope is positive and has a magnitude V/L. For some small interval of time, this slope must approximate the actual curve found by solving Eq. 5-1. Assume that the current increases linearly at the rate V/L to a new value i_1 at time t_1. A second approximation to the curve of current as a function of time may be made at this point by using Eq. 5-2 as

$$\frac{di}{dt}(t_1) = \frac{1}{L}(V - i_1R) \qquad (5\text{-}4)$$

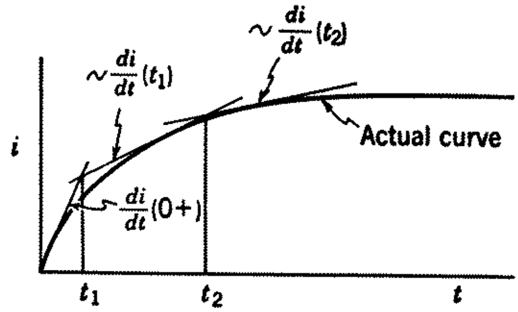


Fig. 5-4. Approximation of an actual curve by tangents to the curve.

Continuation of this process, illustrated in Fig. 5-4, provides a graphical interpretation of the solution of a differential equation. The smaller the time intervals are chosen, the more closely will the approximate curve approach the actual curve.

Just as the first derivative represents slope so the second derivative represents curvature or the rate of change of the slope with time. Figure 5-5 shows several combinations of initial conditions, with the corresponding initial slope and curvature.

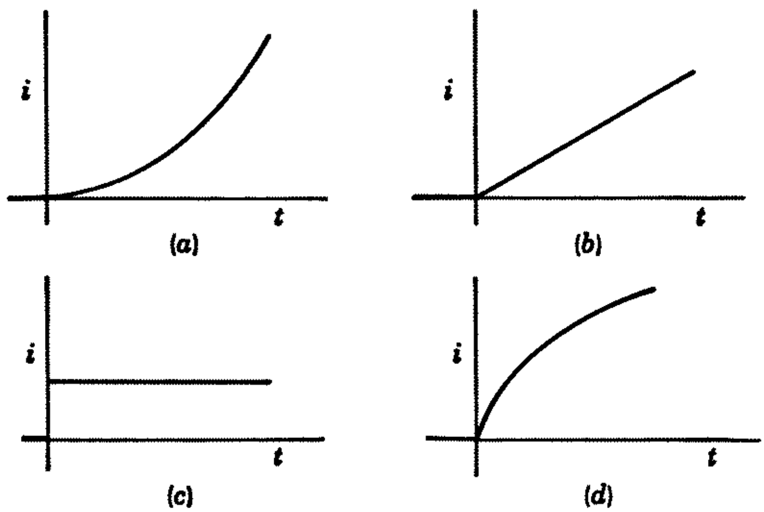


Fig. 5-5. Curves corresponding to typical initial conditions: (a) $i(0+) = 0$, $di/dt(0+) = 0$, $d^2i/dt^2(0+) = K$; (b) $i(0+) = 0$, $di/dt(0+) = +K$, $d^2i/dt^2(0+) = 0$; (c) $i(0+) = K$, $di/dt(0+) = 0$, $d^2i/dt^2(0+) = 0$; (d) $i(0+) = 0$, $di/dt(0+) = +K_1$, $d^2i/dt^2(0+) = -K_2$.

5-3. A procedure for evaluating initial conditions

There is no unique procedure that must be followed in solving for initial conditions. However, it is usually wise practice to solve first for the initial values of the variables—currents or voltages—and then solve for derivatives. The first step is essentially routine and based on the equivalent circuits for $t = 0+$ given in Fig. 5-2. In the second, the details and order of manipulation will be different for each different network. A successful approach will not be obvious at all, a fact that adds interest and offers a challenge in the solution of initial value problems.

Initial values of current or voltage may be found directly from a study of the network schematic. For each element in the network, we must determine just what will happen when the switching action takes place. From this analysis, a new schematic of an equivalent network for $t \doteq 0+$ may be constructed according to these rules:

(1) Replace all inductors by open circuits or by current generators having the value of current flowing at $t = 0+$.
(2) Replace all capacitors by short circuits or by a voltage source of value $V_0 = q_0/C$ if there is an initial charge.
(3) Resistors are left in the network without change.

Consider the two-loop network shown in Fig. 5-6(a). Suppose that the switch is closed at $t = 0$, no voltage having been applied to the

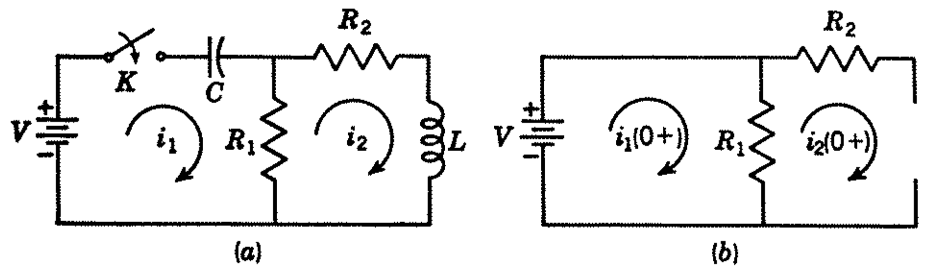


Fig. 5-6. Network illustrating solution for initial conditions: (a) two-loop network; (b) equivalent network at $t = 0+$.

passive network prior to that time. Since there is no initial voltage on the capacitor, it may be replaced by a short circuit; similarly, the inductor may be replaced by an open circuit, there being no initial value of current. The resulting equivalent network is shown as (b) in the figure. In this particular case, there is no need to write equations for the resistor network. By inspection the initial values of the currents are $i_1(0+) = V/R_1$ and $i_2(0+) = 0$ because the second loop is open.

The first step in solving initial values of derivatives is to write the integrodifferential equations from Kirchhoff's laws, employing either the loop or node basis as will give the required quantities more directly. In terms of the network of Fig. 5-6(a), the Kirchhoff voltage equations are

$$\frac{1}{C}\int i_1\,dt + R_1(i_1 - i_2) = V \tag{5-5}$$

$$R_1(i_2 - i_1) + R_2 i_2 + L\frac{di_2}{dt} = 0 \tag{5-6}$$

Since these equations hold in general, they hold at $t = 0+$. Now the values of i_1 and i_2 are known at $t = 0+$. Also the term $(1/C) \int i_1\, dt$ has a known value at $t = 0+$, since this term is the voltage across the capacitor, which is known to be zero since the capacitor acts as a short circuit. (On the node basis, $(1/L) \int v\, dt$ similarly represents current through the inductor.)

We observe that Eq. 5-6 contains a derivative term in addition to terms involving only i_1 and i_2, which are known at $t = 0+$. Algebraically solving for (di_2/dt) gives

$$\frac{di_2}{dt} = \frac{1}{L}\left[R_1 i_1 - (R_1 + R_2) i_2\right] \qquad \text{(general)} \qquad (5\text{-}7)$$

$$\frac{di_2}{dt}(0+) = \frac{1}{L}\left[R_1 \frac{V}{R_1} - (R_1 + R_2)0\right] = \frac{V}{L} \qquad (t = 0+) \qquad (5\text{-}8)$$

The precaution of marking equations as (general) or $(t = 0+)$ is suggested as a safeguard against differentiating equations that hold only for $t = 0+$.

Neither Eq. 5-5 nor Eq. 5-6 contain a (di_1/dt) term. However, if Eq. 5-5, which holds in general, is differentiated and manipulated algebraically, there results

$$\frac{i_1}{C} + R_1 \frac{di_1}{dt} - R_1 \frac{di_2}{dt} = 0 \qquad \text{(general)} \qquad (5\text{-}9)$$

$$\frac{di_1}{dt} = \frac{di_2}{dt} - \frac{i_1}{R_1 C} \qquad \text{(general)} \qquad (5\text{-}10)$$

Both di_2/dt and i_1 are known for $t = 0+$, so that (di_1/dt) may be evaluated as

$$\frac{di_1}{dt}(0+) = \frac{V}{L} - \frac{V}{R_1{}^2 C} \qquad (t = 0+) \qquad (5\text{-}11)$$

Suppose that it is required to evaluate $(d^2 i_2/dt^2)$ at $t = 0+$. From a practical point of view, second- and higher-order derivatives are less frequently required than the first derivative in the solution of differential equations. However, the procedure of continued differentiation and algebraic manipulation can be applied in solving for all derivatives. Differentiation of Eq. 5-7, gives

$$\frac{d^2 i_2}{dt^2} = \frac{1}{L}\left[R_1 \frac{di_1}{dt} - (R_1 + R_2) \frac{di_2}{dt}\right] \qquad \text{(general)} \qquad (5\text{-}12)$$

$$\frac{d^2 i_2}{dt^2}(0+) = -V\left(\frac{1}{R_1 LC} + \frac{R_2}{L^2}\right) \qquad (t = 0+) \qquad (5\text{-}13)$$

In each case, the initial conditions have been given in terms of constants (network and driving force parameters); solutions to problems should not be given in terms of integral or derivative expressions.

Example 1

In the circuit shown in Fig. 5-7, $V = 10$ volts, $R = 10$ ohms, $L = 1$ henry, and $C = 10\ \mu$f. Let it be required to find $i(0+)$, $di/dt(0+)$, and $d^2i/dt^2(0+)$. From the Kirchhoff voltage law,

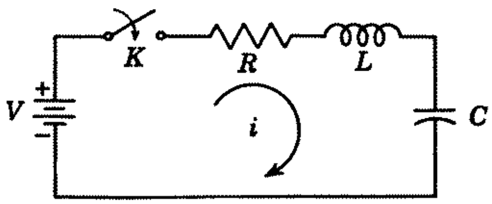


Fig. 5-7. Network of Example 1.

$$V = L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt \qquad \text{(general)} \qquad (5\text{-}14)$$

Analyzing the circuit in terms of equivalent element values for $t = 0$ shows that because of the open circuit,

$$i(0+) = 0 \qquad (t = 0+)$$

The last term in Eq. 5-14, $(1/C)\int i\,dt$, represents the voltage across the capacitor, which is zero at $t = 0$. The general expression in Eq. 5-14 becomes the following for $t = 0+$.

$$V = L\frac{di}{dt}(0+) + R\,0 + 0 \qquad (t = 0+)$$

from which

$$\frac{di}{dt}(0+) = \frac{V}{L} = 10\,\frac{\text{amp}}{\text{sec}} \qquad (t = 0+)$$

To find the second derivative, Eq. 5-14 must be differentiated as

$$L\frac{d^2i}{dt^2} + R\frac{di}{dt} + \frac{1}{C}i = 0 \qquad \text{(general)} \qquad (5\text{-}15)$$

In Eq. 5-15, values for the second and third terms are known at $t = 0+$; thus

$$\frac{d^2i}{dt^2}(0+) = -\frac{R}{L}\frac{di}{dt}(0+) = -100\,\frac{\text{amp}}{\text{sec}^2}$$

Example 2

In the network shown in Figure 5-8, a steady state is reached with the switch K *open*, and at $t = 0$ the switch is *closed*. Let it be required to find the initial value of all three loop currents. We must first find the various currents and voltages in the network at $t = 0-$, before the switch is closed. The current flowing through R_2, R_1, and L will be

$$i_{R_1}(0-) = i_L(0-) = \frac{V}{R_1 + R_2} \qquad (t = 0-)$$

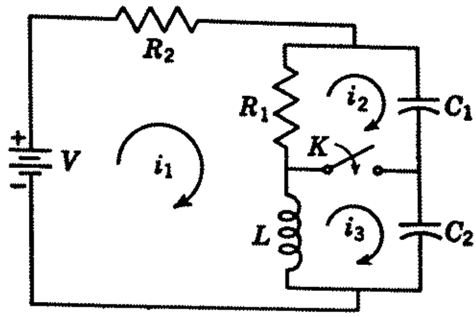


Fig. 5-8. Network of Example 2.

The total voltage across the capacitors will be the same as the drop across R_1; that is,

$$V_{C_1} + V_{C_2} = \frac{R_1}{R_1 + R_2} V \tag{5-16}$$

Since the charge on the capacitors must be equal when connected in series we have $q_1 = q_2$ or $C_1V_{C_1} = C_2V_{C_2}$. Hence the voltage across the capacitors will divide as

$$\frac{V_{C_1}}{V_{C_2}} = \frac{C_2}{C_1} = \frac{S_1}{S_2} \qquad \text{(general)} \tag{5-17}$$

and

$$V_{C_1} = \frac{R_1}{R_1 + R_2}\left[\frac{S_1}{S_1 + S_2}\right] V; \qquad V_{C_2} = \frac{R_1}{R_1 + R_2}\left[\frac{S_2}{S_1 + S_2}\right] V \tag{5-18}$$

To find i_1 at $t = 0+$, apply Kirchhoff's voltage law around the outside loop (not drawn on the diagram). Traversing this loop, we write

$$i_1R_2 = V - V_{C_1} - V_{C_2} = \frac{R_2}{R_1 + R_2} V \tag{5-19}$$

so that

$$i_1(0+) = \frac{V}{R_1 + R_2} \qquad (t = 0+) \tag{5-20}$$

Now, $$i_1(0+) - i_3(0+) = i_L(0+) = \frac{V}{R_1 + R_2} \qquad (t = 0+) \tag{5-21}$$

since the current i_L cannot change instantaneously. Comparing the last two equations shows that

$$i_3(0+) = 0 \tag{5-22}$$

Next, consider the current flowing in the resistor R_1. Since the voltage across the capacitor cannot change instantaneously,

$$i_1(0+) - i_2(0+) = \frac{V_{C_1}}{R_1} \qquad (t = 0+) \tag{5-23}$$

or

$$i_2(0+) = i_1(0+) - \frac{V_{C_1}}{R_1} \qquad (t = 0+) \tag{5-24}$$

so that

$$i_2(0+) = \frac{V}{R_1 + R_2} - \frac{S_1}{S_1 + S_2}\frac{V}{R_1 + R_2} \qquad (t = 0+) \qquad (5\text{-}25)$$

Finally,

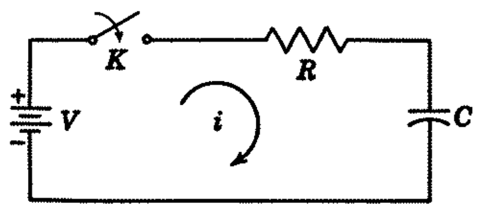
$$i_2(0+) = \frac{V}{R_1 + R_2}\frac{S_2}{S_1 + S_2} \qquad (t = 0+) \qquad (5\text{-}26)$$

FURTHER READING

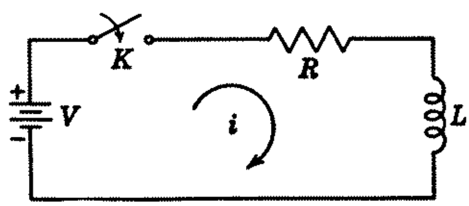
A good discussion of the method of evaluating initial conditions is to be found in Weber's *Linear Transient Analysis* (John Wiley & Sons, Inc., New York, 1954), pp. 42–45, and in Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 26–34.

PROBLEMS

5-1. In the circuit shown, the switch K is closed at $t = 0$. Find the values of i, di/dt, and d^2i/dt^2 at $t = 0+$, when $V = 100$ volts, $R = 1000$ ohms, and $C = 1\ \mu$f. *Answers.* 0.1, -100, 100,000.



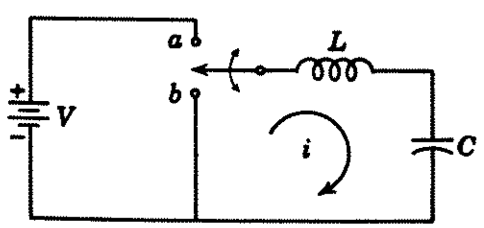
Prob. 5-1.



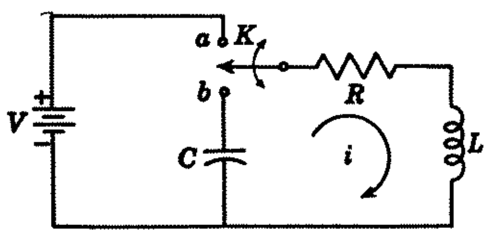
Prob. 5-2.

5-2. In the circuit of the figure, the switch K is closed at $t = 0$. Find the values of i, di/dt, d^2i/dt^2 at $t = 0+$, when $R = 10$ ohms, $L = 1$ henry, and $V = 100$ volts. *Answers.* 0, 100, -1000.

5-3. In the circuit shown, the switch K is changed from position a to position b at $t = 0$, having already established a steady state in position a. Find i, di/dt, d^2i/dt^2, and d^3i/dt^3 at $t = 0+$, when $L = 1$ henry, $C = 10\mu$f, and $V = 100$ volts. *Answers.* 0, -100, 0, 10^7.



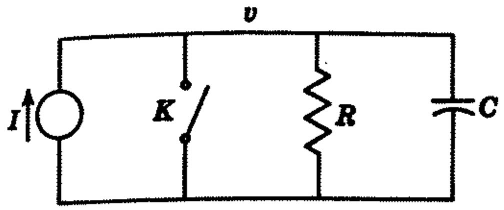
Prob. 5-3.



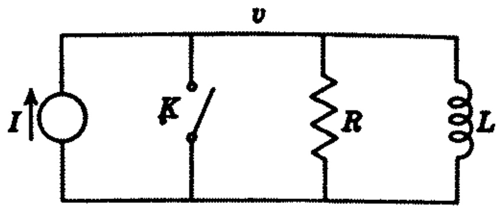
Prob. 5-4.

5-4. In the circuit of the accompanying figure, the switch K is changed from position a to position b at $t = 0$. Solve for i, di/dt, and d^2i/dt^2 at $t = 0+$, when $R = 1000$ ohms, $L = 1$ henry, $C = 1\ \mu$f, and $V = 100$ volts. *Answers.* 0.1, -100, 0.

5-5. In the circuit shown, the switch K is opened at $t = 0$. At $t = 0+$, solve for v, dv/dt, and d^2v/dt^2, when $I = 10$ amp, $R = 1000$ ohms, and $C = 1$ μf. *Answers.* 0, 10^7, -10^{10}.



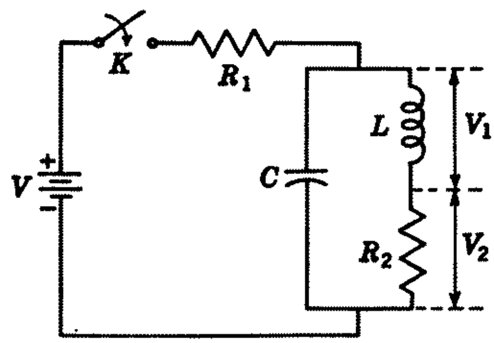
Prob. 5-5.



Prob. 5-6.

5-6. In the circuit of the figure, the switch K is opened at $t = 0$. Solve for v, dv/dt, and d^2v/dt^2 at $t = 0+$, when $I = 1$ amp, $R = 100$ ohms, and $L = 1$ henry. *Answers.* 100, -10^4, 10^6.

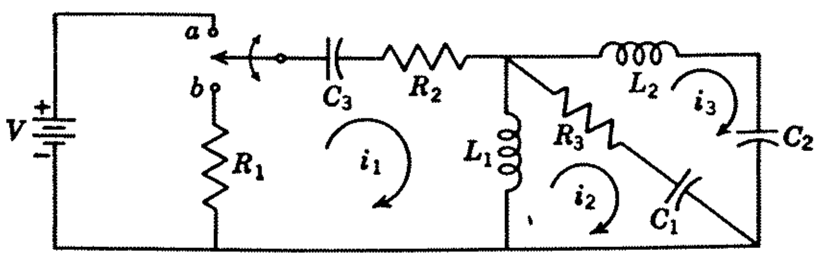
5-7. In the circuit shown, the switch K is closed at $t = 0$. Solve for: (a) v_1 and v_2 at $t = 0+$. (b) v_1 and v_2 at $t = \infty$. (c) dv_1/dt and dv_2/dt at $t = 0+$. (d) d^2v_2/dt^2 at $t = 0+$. *Answers.* (a) 0, 0. (b) 0, $R_2V/(R_1 + R_2)$. (c) $dv_1/dt = V/CR_1$, $dv_2/dt = 0$. (d) $d^2v_2/dt^2 = R_2V/R_1LC$.



Prob. 5-7.

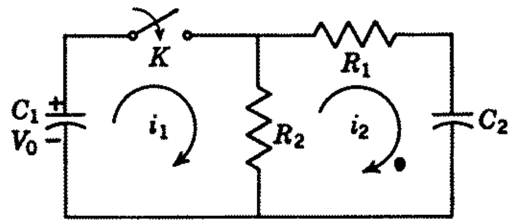
5-8. In the network shown in the accompanying figure, the switch K is changed from a to b at $t = 0$ (a steady state having been established at position a). Show that at $t = 0+$,

$$i_1 = i_2 = -\frac{V}{R_1 + R_2 + R_3}, \qquad i_3 = 0$$

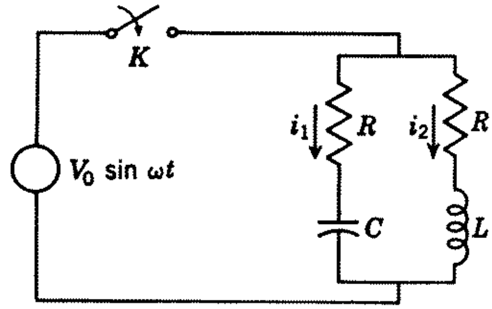


Prob. 5-8.

5-9. In the given network, the capacitor C_1 is charged to voltage V_0 and the switch K is closed at $t = 0$. When $R_1 = 2$ megohms, $V_0 = 1000$ volts, $R_2 = 1$ megohm, $C_1 = 10\ \mu$f, and $C_2 = 20\mu$f, solve for d^2i_2/dt^2 at $t = 0+$. *Answer.* 1.41×10^{-5} amp/sec^2.



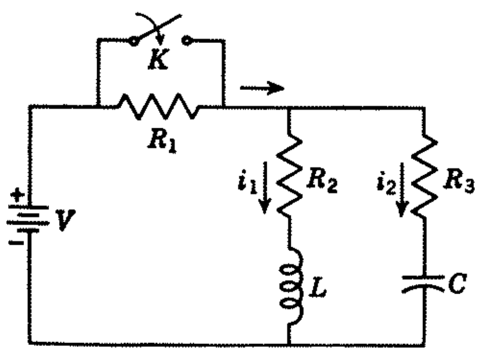
Prob. 5-9.



Prob. 5-10.

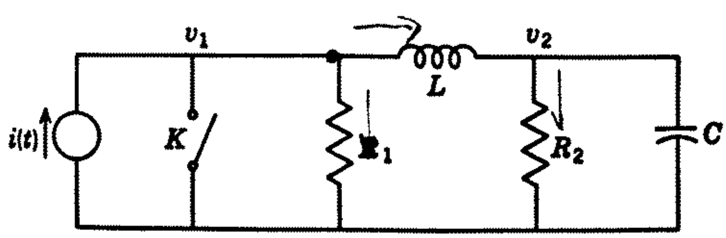
5-10. In the circuit shown in the figure, the switch K is closed at $t = 0$ connecting an alternating voltage, $V_0 \sin \omega t$, to the parallel RL-RC circuit. Find (a) di_1/dt and (b) di_2/dt at $t = 0+$.

5-11. In the network shown, a steady state is reached with the switch K open with $V = 100$ volts, $R_1 = 10$ ohms, $R_2 = 20$ ohms, $R_3 = 20$ ohms, $L = 1$ henry, and $C = 1\ \mu$f. At time $t = 0$, the switch is closed. (a) Write the integrodifferential equations for the network after the switch is closed. (b) What is the voltage V_0 across C before the switch is closed? What is its polarity? *Answer.* 66.7 volts. (c) Solve for the initial value of i_1 and i_2 ($t = 0+$). *Answer.* 3.33 amp, 1.67 amp. (d) Solve for the values of di_1/dt and di_2/dt at $t = 0+$. *Answer.* 33.3, $-83{,}300$. (e) What is the value of di_1/dt at $t = \infty$?



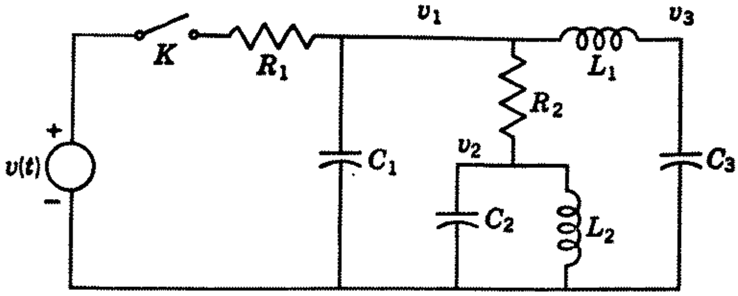
Prob. 5-11.

5-12. The network shown in the figure has two independent node pairs. If the switch K is opened at $t = 0$, find the following quantities at $t = 0+$: (a) v_1, (b) v_2, (c) dv_1/dt, (d) dv_2/dt.



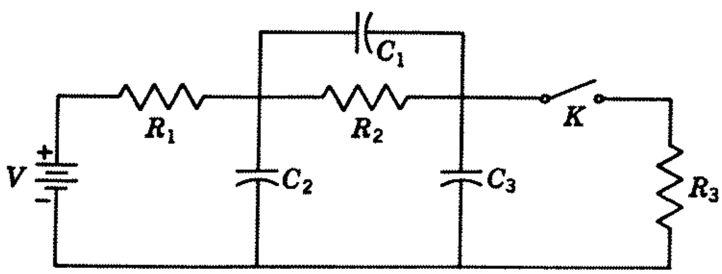
Prob. 5-12.

5-13. In the network shown, the switch K is closed at $t = 0$. Find dv_2/dt at $t = 0+$.

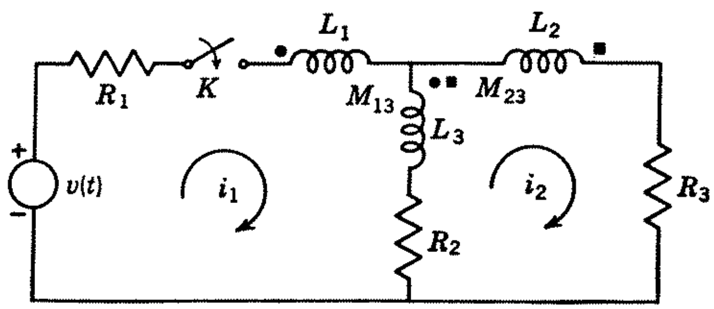


Prob. 5-13.

5-14. In the network shown in the accompanying figure, an equilibrium is reached, and at $t = 0$, switch K is opened. Find the initial voltage across the switch and the initial time derivative of the voltage across the switch.



Prob. 5-14.



Prob. 5-15.

5-15. In the network shown in the figure, the switch K is closed at the instant $t = 0$, connecting an unenergized system to a voltage source. Let $M_{12} = 0$. Determine the values of

$$\frac{di_1}{dt}(0+) \quad \text{and} \quad \frac{di_2}{dt}(0+)$$

Answer.

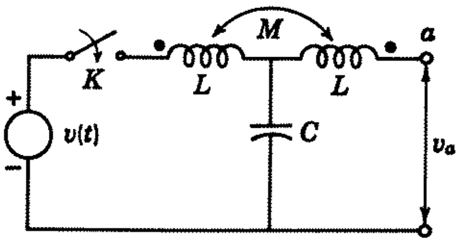
$$\frac{di_1}{dt}(0+) = \frac{V(L_2 + L_3 + 2M_{23})}{(L_1 + L_3 + 2M_{13})(L_2 + L_3 + 2M_{23}) - (L_3 + M_{13} + M_{23})^2}$$

$$\frac{di_2}{dt}(0+) = \frac{V(L_3 + M_{13} + M_{23})}{(L_1 + L_3 + 2M_{13})(L_2 + L_3 + 2M_{23}) - (L_3 + M_{13} + M_{23})^2}$$

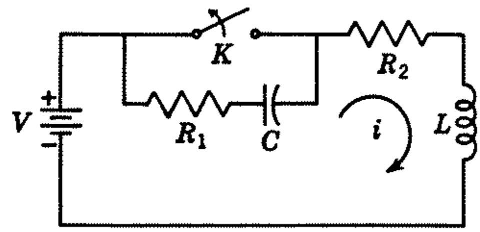
5-16. The given network consists of two coupled coils and a capacitor. At $t = 0$, the switch K is closed connecting a generator of voltage, $v(t) = V \sin (t/\sqrt{MC})$. Determine the values of

$$v_a(0+), \quad \frac{dv_a}{dt}(0+), \quad \text{and} \quad \frac{d^2v_a}{dt^2}(0+)$$

Answer. $v_a(0+) = 0$, $dv_a/dt\,(0+) = (V/L)\sqrt{M/C}$, $d^2v_a/dt^2\,(0+) = 0$.



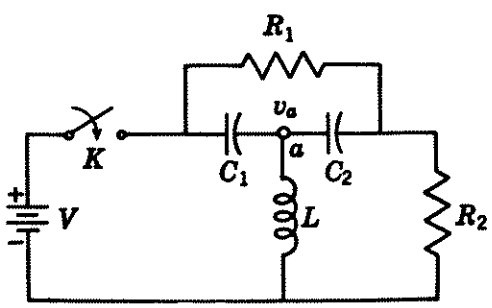
Prob. 5-16.



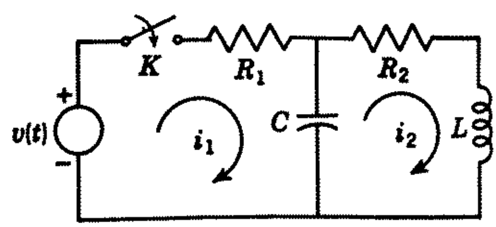
Prob. 5-17.

5-17. In the network of the figure, the switch K is *opened* at $t = 0$ after the network has attained a steady state with the switch closed. (a) Find an expression for the voltage across the switch at $t = 0+$. (b) If the parameters are adjusted such that $i(0+) = 1$ and $di/dt\,(0+) = -1$, what is the value of the derivative of the voltage across the switch, $dv_K/dt\,(0+)$? *Answers.* (a) VR_1/R_2, (b) $1/C - R_1$.

5-18. In the network shown in the figure, the switch K is closed at $t = 0$ connecting the battery with an unenergized system. (a) Find the voltage at point a, V_a at $t = 0+$. (b) Find the voltage across capacitor C_1 at $t = \infty$, $V_{C_1}(\infty)$.



Prob. 5-18.



Prob. 5-19.

5-19. For the network of the figure, show that

$$\frac{d^2i_1}{dt^2}(0+) = -\frac{1}{R_1}\left\{\frac{-1}{R_1C}\left[\frac{v(t)}{R_1C} - \frac{dv(t)}{dt}\right] + \frac{d^2v(t)}{dt^2}\right\}$$

CHAPTER 6

DIFFERENTIAL EQUATIONS, CONTINUED

Differential equations studied in Chapter 4 were limited to linear equations of the first order with constant coefficients. In this chapter, we will continue our study of differential equations with the same restrictions as to linearity and constant coefficients but of higher order. The mathematical procedures given in these two chapters are included under the heading of the *classical* method of solution. As we will see, the classical method affords a better insight into the interpretation of differential equations and the requirements of a solution. Aside from conceptual advantages, the operational method using the Laplace transformation is better suited to our use. For this reason, our treatment will be brief. Topics ordinarily covered using the classical method but more easily developed with the aid of the Laplace transformation will be reserved for the next chapter.

6-1. Solution of a second-order homogeneous differential equation

A second-order differential equation with constant coefficients may be written in the general form

$$a_0 \frac{d^2i}{dt^2} + a_1 \frac{di}{dt} + a_2 i = 0 \tag{6-1}$$

The solution of this differential equation must be of such form that the solution itself, its first derivative, and its second derivative, each multiplied by a constant coefficient, add to zero. To satisfy this requirement, the three terms must be of the same form, differing only in their coefficients. Is there such a function? By whatever method we search, perhaps trying possible functions, the search always leads to the exponential*

$$i(t) = ke^{mt} \tag{6-2}$$

where k and m are constants. Substituting the exponential solution into Eq. 6-1 gives

$$a_0 m^2 ke^{mt} + a_1 mke^{mt} + a_2 ke^{mt} = 0 \tag{6-3}$$

* Taken two at a time, the sine and the cosine or the hyperbolic sine and the hyperbolic cosine satisfy the requirement; however, the exponential solution will be shown to simplify to these forms.

or, since ke^{mt} can never be zero for finite t,

$$a_0m^2 + a_1m + a_2 = 0 \tag{6-4}$$

as the requirement for ke^{mt} to be the solution. This equation is known as the *characteristic* (or *auxiliary*) *equation.* It is satisfied by the two roots given by the quadratic formula

$$m_1, m_2 = -\frac{a_1}{2a_0} \pm \frac{1}{2a_0}\sqrt{a_1^2 - 4a_0a_2} \tag{6-5}$$

We now have discovered that there are two forms of the exponential solution ke^{mt}; they are

$$i_1 = k_1e^{m_1t} \quad \text{and} \quad i_2 = k_2e^{m_2t} \tag{6-6}$$

Now, if i_1 and i_2 are each solutions of the differential equation of Eq. 6-1, the sum of these solutions,

$$i_3 = i_1 + i_2 \tag{6-7}$$

is also a solution. This may be shown by direct substitution of Eq. 6-7 into Eq. 6-1, giving

$$a_0\frac{d^2}{dt^2}(i_1 + i_2) + a_1\frac{d}{dt}(i_1 + i_2) + a_2(i_1 + i_2) = 0 \tag{6-8}$$

$$\left(a_0\frac{d^2i_1}{dt^2} + a_1\frac{di_1}{dt} + a_2i_1\right) + \left(a_0\frac{d^2i_2}{dt^2} + a_1\frac{di_2}{dt} + a_2i_2\right) = 0 \tag{6-9}$$

or $0 + 0 = 0$. The general solution of the differential equation is thus

$$i(t) = k_1e^{m_1t} + k_2e^{m_2t} \tag{6-10}$$

The magnitude of the coefficients in Eq. 6-1 determines the form of the roots of the characteristic equation. In Eq. 6-5, the radical $\pm\sqrt{a_1^2 - 4a_0a_2}$ may be real, zero, or imaginary depending on the value of a_1^2 compared with $4a_0a_2$. The forms of the solutions for these three cases will be given by three simple examples.

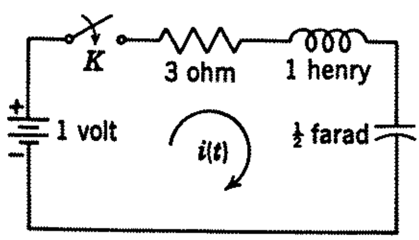


Fig. 6-1. Circuit for Examples 1 and 3.

Example 1

The differential equation for the current in the circuit of Fig. 6-1 is given by Kirchhoff's law as

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = V \tag{6-11}$$

Differentiating and using numerical values for R, L, and C shown in Fig. 6-1 gives

$$\frac{d^2i}{dt^2} + 3\frac{di}{dt} + 2i = 0 \tag{6-12}$$

The characteristic equation can be found by substituting the trial solution $i = e^{mt}$ or by the equivalent of substituting m^2 for (d^2i/dt^2), and m for (di/dt); thus

$$m^2 + 3m + 2 = 0 \tag{6-13}$$

This equation has the roots $m_1 = -1$ and $m_2 = -2$, so that the general solution is

$$i(t) = k_1e^{-t} + k_2e^{-2t} \tag{6-14}$$

The arbitrary constants k_1 and k_2 can be evaluated for a specific problem by a knowledge of the initial conditions. If the switch K is closed at $t = 0$, then $i(0+) = 0$, because current cannot change instantaneously in the inductor. In Eq. 6-11, the second and third voltage terms are zero at the instant of switching, $Ri(0+)$ being zero because $i(0+) = 0$ and $(1/C)\int i\,dt$ being zero because it is the initial voltage across the capacitor. Hence

$$\frac{di}{dt}(0+) = \frac{V}{L} = 1 \qquad \text{amp/sec}$$

The two initial conditions, substituted into the general solution, Eq. 6-14, gives the equations,

$$k_1 + k_2 = 0, \qquad -k_1 - 2k_2 = 1 \tag{6-15}$$

The solution of these equations is $k_1 = +1$ and $k_2 = -1$; hence the particular solution to Eq. 6-12 is

$$i(t) = e^{-t} - e^{-2t} \tag{6-16}$$

A plot of the separate parts and their combination is shown in Fig. 6-2. As discussed in Chapter 4, the total current may be thought of as made

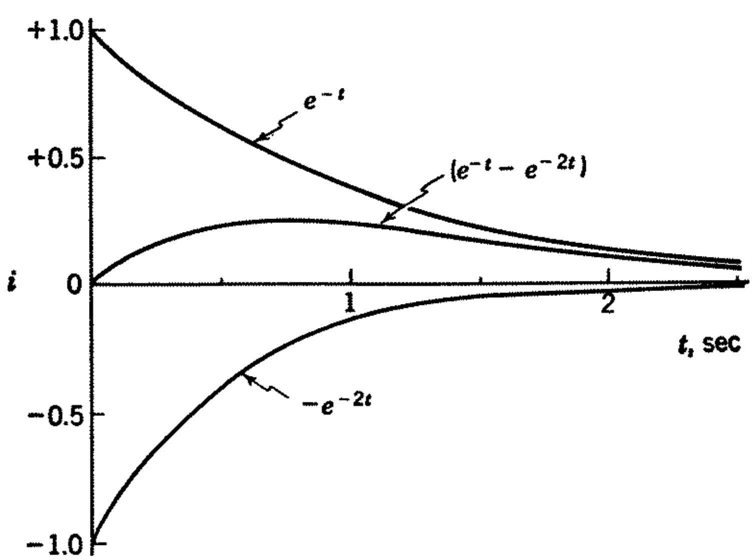


Fig. 6-2. Current as a function of time for the circuit of Example 1.

up of two components which exist from $t = 0$ and combine in such a way as to satisfy the initial conditions.

Example 2

The equilibrium equation for the network shown in Fig. 6-3 formulated on the node basis is

$$C\frac{dv}{dt} + Gv + \frac{1}{L}\int v\,dt = I \tag{6-17}$$

or, by differentiation,

$$C\frac{d^2v}{dt^2} + G\frac{dv}{dt} + \frac{v}{L} = 0 \tag{6-18}$$

Substituting numerical values into this equation as given in Fig. 6-3

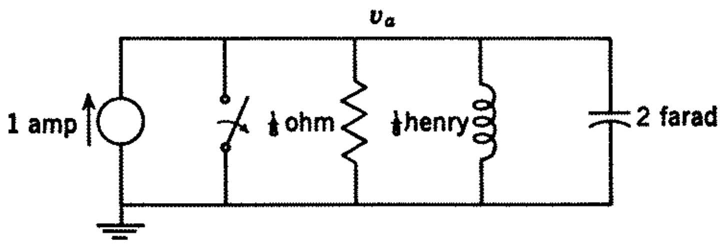


Fig. 6-3. Circuit for Example 2.

gives

$$2\frac{d^2v}{dt^2} + 8\frac{dv}{dt} + 8v = 0 \tag{6-19}$$

The corresponding characteristic equation is

$$2m^2 + 8m + 8 = 0 \tag{6-20}$$

which has as roots $m_1 = -2$ and $m_2 = -2$, or repeated roots. Substituting into the general form of the solution, Eq. 6-10, gives

$$v(t) = k_1e^{-2t} + k_2e^{-2t} = Ke^{-2t} \tag{6-21}$$

where $K = k_1 + k_2$. This is not a complete form of the solution, since the general solution to a second-order differential equation must contain two arbitrary constants. The solution $v = ke^{-2t}$ must be modified in some manner for the condition of repeated roots. If we assume the new solution to be $v = ye^{-2t}$, where y is a factor to be determined, and substitute into Eq. 6-19, we arrive at the requirement that y satisfy the differential equation

$$\frac{d^2y}{dt^2} = 0 \tag{6-22}$$

Two successive integrations of the equation lead to the solution

$$y = k_1 + k_2t \tag{6-23}$$

Thus the solution to our problem with repeated roots becomes

$$v(t) = k_1e^{-2t} + k_2te^{-2t} \tag{6-24}$$

To obtain a particular solution for this problem will require knowledge of two initial conditions. From the network of Fig. 6-3, $v(0+)$ must equal zero, since the capacitor acts as a short circuit at the initial instant. In Eq. 6-17, the second and third terms are equal to zero, the first because $v(0+) = 0$ and the second because there is not current in the inductor at the initial instant. Then, by Eq. 6-17, $dv/dt\ (0+) = I/C = \frac{1}{2}$ volt/sec for this network. Substituting these initial conditions into Eq. 6-24 leads to the result that $k_1 = 0$ and $k_2 = \frac{1}{2}$. The desired particular solution is

$$v(t) = \tfrac{1}{2}te^{-2t} \tag{6-25}$$

A plot of this solution is shown in Fig. 6-4.

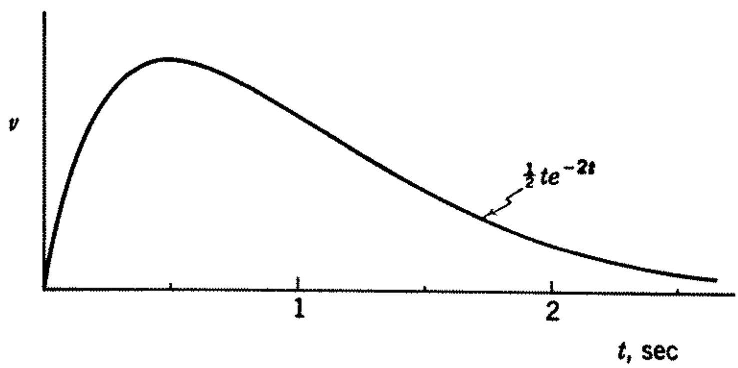


Fig. 6-4. Voltage as a function of time for the circuit of Example 2.

Example 3

For this example, we will use the network of Fig. 6-1 with the following network parameter values: $V = 1$ volt, $L = 1$ henry, $R = 2$ ohms, and $C = \frac{1}{2}$ farad. The characteristic equation becomes

$$m^2 + 2m + 2 = 0 \tag{6-26}$$

with roots*

$$m_1, m_2 = -1 \pm j1$$

The general solution, Eq. 6-10, with these values for m, becomes

$$i(t) = k_1e^{(-1+j1)t} + k_2e^{(-1-j1)t} = e^{-t}(k_1e^{jt} + k_2e^{-jt})$$

This particular form of the solution is not convenient for interpretation. An equivalent form may be found starting with *Euler's equation*,

$$e^{\pm jt} = \cos t \pm j \sin t \tag{6-27}$$

* We will use the letter j for the operator $\sqrt{-1}$ to reserve the letter i for current. The letter j in textbooks of electrical engineering is equivalent to $i = \sqrt{-1}$ in textbooks of mathematics and physics.

which reduces our solution to the form

$$i(t) = e^{-t}(k_3 \cos t + k_4 \sin t)$$

where $k_3 = k_1 + k_2$ and $k_4 = j(k_1 - k_2)$. The initial conditions are the same as in Example 1: $i(0+) = 0$ and $di/dt\,(0+) = 1$ amp/sec. Substituting into the solution, we have

$$i(0+) = 0 = e^{-0}(k_3 \cos 0 + k_4 \sin 0) = k_3$$

With k_3 equal to zero,

$$\frac{di}{dt}(0+) = k_4(e^{-0} \cos 0 - \sin 0\, e^{-0}) = 1$$

whence $k_4 = 1$. The particular solution is

$$i(t) = e^{-t} \sin t \tag{6-28}$$

A plot of the two factors in this solution and their product is shown in Fig. 6-5.

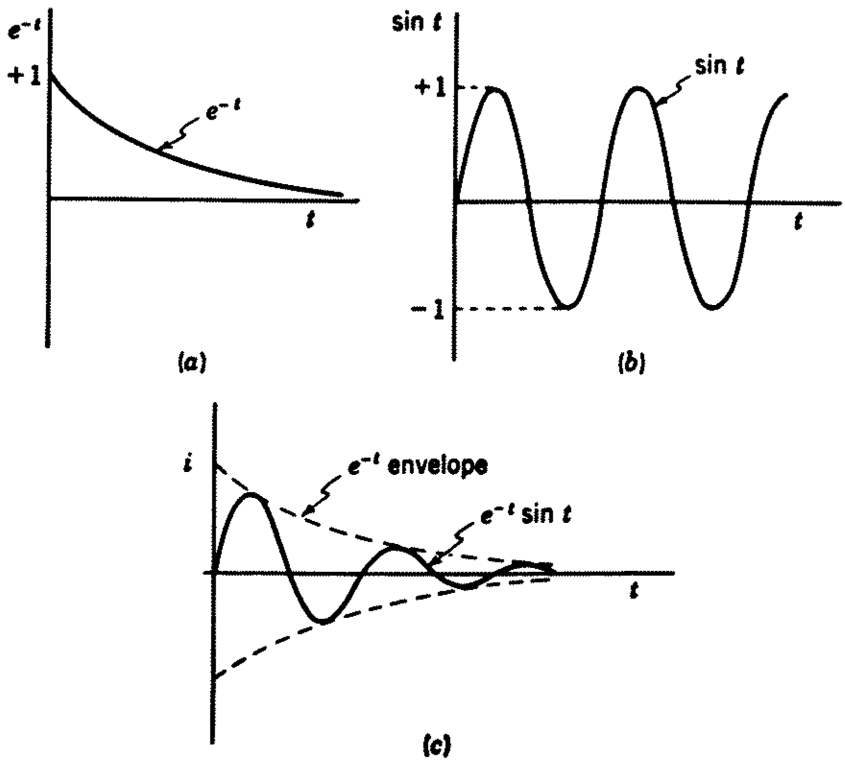


Fig. 6-5. Current as a function of time for the circuit of Example 3: (a) e^{-t}; (b) $\sin t$; (c) $e^{-t} \sin t$.

6-2. The standard form of the solution of second-order differential equations

Consider the Kirchhoff voltage equation that describes a series *RLC* circuit on the loop basis

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = v(t) \tag{6-29}$$

If $v(t)$ is either zero or a constant, differentiation reduces this equation to a homogeneous equation of second order; thus

$$\frac{d^2i}{dt^2} + \frac{R}{L}\frac{di}{dt} + \frac{1}{LC}i = 0 \tag{6-30}$$

The two roots of the corresponding characteristic equation may be found by the quadratic formula to be

$$m_1, m_2 = -\frac{R}{2L} \pm \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}} \tag{6-31}$$

To begin converting Eq. 6-30 to a standard form, we will define the value of resistance that causes the radical term in the above equation to vanish as the *critical resistance*, R_{cr}. This value is found by solving the equation

$$\left(\frac{R_{cr}}{2L}\right)^2 = \frac{1}{LC} \tag{6-32}$$

or

$$R_{cr} = 2\sqrt{L/C} \tag{6-33}$$

We will next introduce two definitions; we define the quantity

$$\zeta = \frac{R}{R_{cr}} = \frac{R}{2}\sqrt{\frac{C}{L}} \tag{6-34}$$

as the dimensionless *damping ratio*. (ζ is the lower-case Greek letter zeta.) The damping ratio is the ratio of the actual resistance to the critical value of resistance. The other definition is

$$\omega_n = \frac{1}{\sqrt{LC}} \tag{6-35}$$

The quantity ω_n is the *undamped natural angular frequency*. The reason for giving ω_n such a name will be discussed under the heading of Case 3 in this section. For the time being, we note that the dimensions of ω_n are $(\text{time})^{-1}$ so that it does not seem unreasonable to define it as a frequency.

Now the product $2\zeta\omega_n$ has the value

$$2\zeta\omega_n = 2\frac{R}{2}\sqrt{\frac{C}{L}}\frac{1}{\sqrt{LC}} = \frac{R}{L} \tag{6-36}$$

and

$$\omega_n^2 = \frac{1}{LC} \tag{6-37}$$

Substituting these relationships into Eq. 6-30 gives

$$\frac{di^2}{dt^2} + 2\zeta\omega_n \frac{di}{dt} + \omega_n{}^2 i = 0 \tag{6-38}$$

This form of the second-order differential equation is called the *standard form.* The corresponding characteristic equation is

$$m^2 + 2\zeta\omega_n m + \omega_n{}^2 = 0 \tag{6-39}$$

and the roots of the characteristic equation are

$$m_1, m_2 = -\zeta\omega_n \pm \omega_n \sqrt{\zeta^2 - 1} \tag{6-40}$$

The general solution may now be written

$$i = K_1 e^{[-\zeta\omega_n + \omega_n\sqrt{(\zeta^2-1)}]t} + K_2 e^{[-\zeta\omega_n - \omega_n\sqrt{(\zeta^2-1)}]t} \tag{6-41}$$

Before simplifying this solution, let us examine the behavior of the roots of the characteristic equation as the dimensionless damping ratio ζ varies from zero (corresponding to $R = 0$) to infinity (corresponding to $R = \infty$). There are evidently three different forms for the roots:

Case 1: $\zeta > 1$, the roots are real.
Case 2: $\zeta = 1$, the roots are real and repeated.
Case 3: $\zeta < 1$, the roots are complex and conjugates.

If we follow the form of the roots for a variation of ζ from 0 to ∞, we will recognize a locus of roots in the complex plane. To start with, for $\zeta = 0$,

$$m_1, m_2 = \pm j\omega_n \tag{6-42}$$

that is, the roots are purely imaginary. For $\zeta < 1$, the roots are complex conjugates as

$$m_1, m_2 = -\zeta\omega_n \pm j\omega_n \sqrt{1 - \zeta^2} \tag{6-43}$$

Since the roots are complex numbers for $\zeta < 1$, let us define the real and imaginary parts of m as

$$m = \sigma + j\omega \tag{6-44}$$

(σ is the lower-case Greek letter sigma). The roots of the equation may be plotted in the complex m plane as shown in Fig. 6-6. The real part is

$$\sigma = -\zeta\omega_n \tag{6-45}$$

and the imaginary part is

$$\omega = \pm\omega_n \sqrt{1 - \zeta^2} \tag{6-46}$$

In other words, the two roots have the same real part, and the imaginary parts differ only by their sign. Since

$$\sigma^2 + \omega^2 = \zeta^2\omega_n^2 + \omega_n^2(1 - \zeta^2) = \omega_n^2 \tag{6-47}$$

it follows that the locus of the roots in the complex m plane is a *circle* of radius ω_n and that this locus is formed by ζ varying from 0 to 1. This locus is shown in Fig. 6-7. It is interesting to note another property of the geometry of the m plane for these second-order roots. The

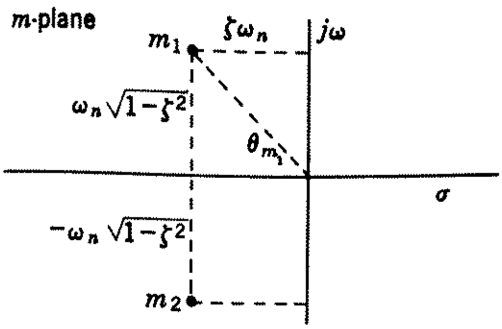


Fig. 6-6. Complex roots.

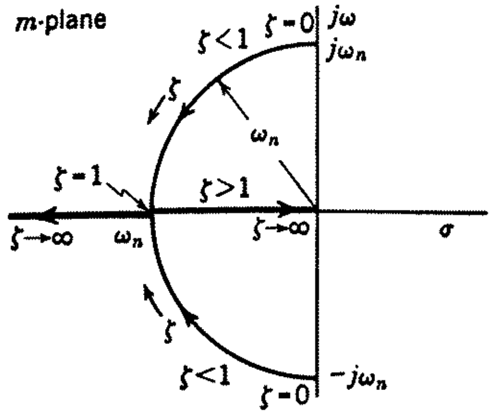


Fig. 6-7. Locus of roots with ζ.

angle to either root measured from the $-\sigma$ axis is given in terms of the inverse tangent as

$$\theta_{m_1} = \tan^{-1} \frac{\omega_n \sqrt{1 - \zeta^2}}{\omega_n\zeta} \tag{6-48}$$

or the inverse tangent of the imaginary part over the real part. Since the term ω_n is common to both the numerator and the denominator, the angle is

$$\theta_{m_1} = \tan^{-1} \frac{\sqrt{1 - \zeta^2}}{\zeta} \tag{6-49}$$

A triangle having the sides ζ and $\sqrt{1 - \zeta^2}$ is shown in Fig. 6-8. Evidently, the hypotenuse has unit value, and

$$\theta_{m_1} = \cos^{-1} \zeta \tag{6-50}$$

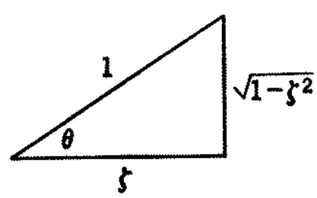


Fig. 6-8. Right triangle relationships.

Thus radial lines from the origin of the m plane are lines of constant ζ as ω_n is varied. When the damping ratio has the value $\zeta = 1$, the imaginary part of the roots vanishes, and the roots have the same value,

$$(m_1, m_2) = -\omega_n \tag{6-51}$$

These two superimposed values are shown in Fig. 6-7. For $\zeta > 1$, the two roots become

$$m_1, m_2 = (-\zeta \pm \sqrt{\zeta^2 - 1})\,\omega_n \tag{6-52}$$

Both of these roots are real. As ζ becomes large such that 1 is small compared with ζ^2, the roots approach $-2\zeta\omega_n$ and 0. The locus of roots shown in Fig. 6-7 illustrates this separation of the real roots as ζ increases.

This discussion has illustrated the three possibilities for the values of ζ. The damping ratio ζ is determined by circuit parameters. With the *RLC* circuit that served as an example in deriving the relationships, a simple adjustment of the resistance R will vary the roots in the complex m plane for the three cases. This general solution, Eq. 6-41, reduces to different forms for each of the three cases. We will next investigate this algebraic reduction for each of the three cases.

Case 1, $\zeta > 1$. For this case, the solution in exponential form is as given in Eq. 6-41. If $e^{-\zeta\omega_n t}$ is factored from this equation, there results

$$i = e^{-\zeta\omega_n t}(K_1 e^{\omega_n\sqrt{\zeta^2-1}\,t} + K_2 e^{-\omega_n\sqrt{\zeta^2-1}\,t}) \tag{6-53}$$

where K_1 and K_2 are arbitrary constants of integration. This equation is sometimes more convenient to evaluate in terms of hyperbolic functions. The *hyperbolic cosine* of x is defined as

$$\cosh x = \tfrac{1}{2}(e^x + e^{-x}) \tag{6-54}$$

and the *hyperbolic sine* of x is defined as

$$\sinh x = \tfrac{1}{2}(e^x - e^{-x}) \tag{6-55}$$

An equivalent relationship can be obtained by successively adding or subtracting these two equations; that is,

$$e^x = \sinh x + \cosh x \tag{6-56}$$

and

$$e^{-x} = \cosh x - \sinh x \tag{6-57}$$

These two identities may be used to convert Eq. 6-53 to terms involving hyperbolic functions; thus

$$\begin{aligned} i = e^{-\zeta\omega_n t}\{&K_1[\cosh(\omega_n\sqrt{\zeta^2-1}\,t) + \sinh(\omega_n\sqrt{\zeta^2-1}\,t)] \\ &+ K_2[\cosh(\omega_n\sqrt{\zeta^2-1}\,t) - \sinh(\omega_n\sqrt{\zeta^2-1}\,t)]\} \end{aligned} \tag{6-58}$$

or

$$i = e^{-\zeta\omega_n t}[K_3\cosh(\omega_n\sqrt{\zeta^2-1}\,t) + K_4\sinh(\omega_n\sqrt{\zeta^2-1}\,t)] \tag{6-59}$$

where

$$K_3 = K_1 + K_2 \tag{6-60}$$

and

$$K_4 = K_1 - K_2 \tag{6-61}$$

This equation is the equivalent of Eq. 6-53. Each has two arbitrary constants, which are usually evaluated to find a particular solution in terms of the initial conditions.

Case 2, $\zeta = 1$. For this case, we have shown that the two roots become identical. With repeated roots, the solution of the equation is given by Eq. 6-24, giving

$$i = (K_1 + K_2t)e^{-\omega_n t} \tag{6-62}$$

The limit of the quantity $te^{-\omega_n t}$ may be investigated by l'Hospital's rule. If this quantity is written as

$$\frac{t}{e^{+\omega_n t}} \tag{6-63}$$

differentiation of numerator and denominator with respect to t shows that

$$\lim_{t \to \infty} te^{-\omega_n t} = 0 \tag{6-64}$$

Case 3, $\zeta < 1$. For Case 3, the roots become complex, and Eq. 6-41 may be written

$$i = e^{-\zeta\omega_n t}(K_1 e^{j\omega_n\sqrt{1-\zeta^2}t} + K_2 e^{-j\omega_n\sqrt{1-\zeta^2}t}) \tag{6-65}$$

This equation may be written in terms of sine and cosine quantities by making use of Euler's equation, Eq. 6-27.

$$e^{\pm jx} = \cos x \pm j \sin x \tag{6-66}$$

Using Euler's equation, the solution for Case 3 reduces to

$$i = e^{-\zeta\omega_n t}[K_5 \cos (\omega_n \sqrt{1 - \zeta^2}\, t) + K_6 \sin (\omega_n \sqrt{1 - \zeta^2}\, t)] \tag{6-67}$$

where $$K_5 = K_1 + K_2 \quad \text{and} \quad K_6 = j(K_1 - K_2) \tag{6-68}$$

which are, again, arbitrary constants of integration. This equation may be written in different form by defining

$$K_5 = K \sin \phi \tag{6-69}$$

$$K_6 = K \cos \phi \tag{6-70}$$

Using the trigonometric identity

$$\sin (x + y) = \sin x \cos y + \sin y \cos x \tag{6-71}$$

Eq. 6-67 becomes

$$i = Ke^{-\zeta\omega_n t} \sin (\omega_n \sqrt{1 - \zeta^2}\, t + \phi) \tag{6-72}$$

These algebraic manipulations have resulted in an equation of one sinusoid equivalent to Eq. 6-67, which contains two sinusoids of the same frequency. In the revised form, the two arbitrary constants are

K and ϕ, which may be related to K_5 and K_6 by means of Eqs. 6-69 and 6-70. Summing the squares of K_5 and K_6 gives the relationship

$$K = \sqrt{K_5^2 + K_6^2} \tag{6-73}$$

Dividing Eq. 6-69 by Eq. 6-70 gives an equation for ϕ in terms of K_5 and K_6:

$$\phi = \tan^{-1} \frac{K_5}{K_6} \tag{6-74}$$

Several terms that have appeared in these solutions are given names. The expression $e^{-\zeta\omega_n t}$ is given the name *damping factor*. The product $\zeta\omega_n$ is called the *decrement factor* or *attenuation factor*. The three cases are given the names,

Case 1. the *overdamped* case
Case 2. the *critically damped* case
Case 3. the *underdamped* (or *oscillatory*) case

In the underdamped case, expressions of the following form appear:

$$\sin (\omega_n \sqrt{1 - \zeta^2}\, t)$$

where $\omega_n \sqrt{1 - \zeta^2}$ is an angular frequency. When $\zeta = 0$, corresponding to $R = 0$, or no damping, the radical term in the angular frequency reduces to unity. On this basis, the following definitions are made:

$$\omega_n \sqrt{1 - \zeta^2} = \text{the actual angular frequency}$$
$$\omega_n = \text{the undamped natural angular frequency}$$

To illustrate these definitions, let us return to the series RLC circuit which is described by the original differential equation of this section. Let the capacitor of Fig. 6-9 be charged to a voltage V_0, and at time $t = 0$ let the switch K be closed. The value of the resistance R with respect to the critical resistance R_{cr} will determine whether the system is overdamped, critically damped, or underdamped. Consider these three possibilities in turn.

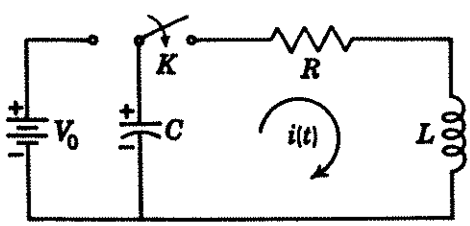


Fig. 6-9. *RLC* circuit.

With $R > R_{cr}$, the system is overdamped. The general solution can be reduced to particular solution for a given set of initial conditions. For the circuit shown in Fig. 6-9, $i(0+) = 0$ because of the inductance. The term $(1/C) \int i\, dt = -V_0$ at $t = 0$ (the initial voltage on the capacitor) such that

$$\frac{di}{dt}(0+) = +\frac{V_0}{L}$$

The requirement that $i(t) = 0$ at $t = 0$, means that K_3 in Eq. 6-59 has zero value; that is,

$$i = K_4 e^{-\zeta\omega_n t} \sinh \omega_n \sqrt{\zeta^2 - 1}\, t \tag{6-75}$$

Constant K_4 can be evaluated from the second initial condition as

$$\frac{di}{dt} = K_4[e^{-\zeta\omega_n t} \cosh (\omega_n \sqrt{\zeta^2 - 1}\, t) \cdot \omega_n \sqrt{\zeta^2 - 1} + \sinh (\omega_n \sqrt{\zeta^2 - 1}\, t) e^{-\zeta\omega_n t}(-\zeta\omega_n)] \tag{6-76}$$

The hyperbolic cosine term approaches unity as $t \to 0$, and the hyperbolic sine term approaches zero as $t \to 0$. Hence

$$\frac{di}{dt}(0+) = K_4\omega_n \sqrt{\zeta^2 - 1} = \frac{V_0}{L} \tag{6-77}$$

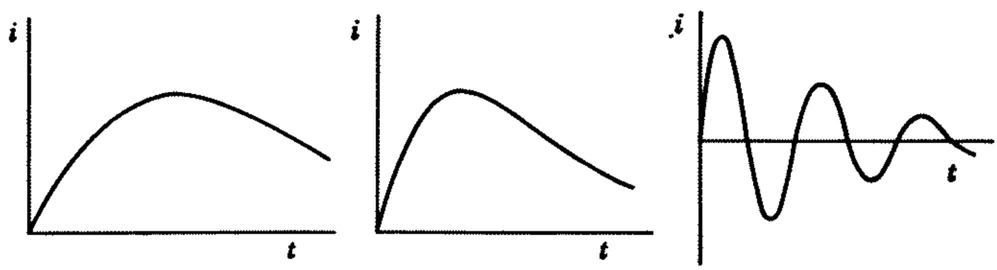
and

$$K_4 = \frac{V_0}{\omega_n L \sqrt{\zeta^2 - 1}} \tag{6-78}$$

The particular solution for the overdamped case thus becomes

$$i = \frac{V_0}{\omega_n L \sqrt{\zeta^2 - 1}} e^{-\zeta\omega_n t} \sinh (\omega_n \sqrt{\zeta^2 - 1}\, t) \tag{6-79}$$

The general shape of the current against time curve for this equation is shown in Fig. 6-10(a).



(a) (b) (c)

Fig. 6-10. Network response corresponding to the three cases: (a) overdamped; (b) critically damped; and (c) underdamped or oscillatory.

For the critically damped case, $R = R_{cr}$ and the solution in general is

$$i = (K_1 + K_2 t)e^{-\omega_n t} \tag{6-80}$$

subject to the same initial conditions as the overdamped case. The initial current condition implies that $K_1 = 0$, since otherwise this equation does not reduce to zero at $t = 0$. To apply the derivative condition, Eq. 6-80 is differentiated as

$$\frac{di}{dt} = K_2[te^{-\omega_n t}(-\omega_n) + e^{-\omega_n t}1] \tag{6-81}$$

Hence

$$\frac{di}{dt}(0+) = K_2 = \frac{V_0}{L} \tag{6-82}$$

and the particular solution for the critically damped case is

$$i = \frac{V_0}{L} te^{-\omega_n t} \tag{6-83}$$

This curve is shown in Fig. 6-10(b) and has much the same appearance as that of Fig. 6-10(a).

For the underdamped, or oscillatory case, $R < R_{cr}$, and the solution is given by Eq. 6-67. The initial condition for the current requires that K_5 be zero, so that the solution can be written

$$i = K_6 e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2} t) \tag{6-84}$$

The constant K_6 is evaluated by using the initial condition of the derivative of the current; thus

$$\frac{di}{dt} = K_6 e^{-\zeta\omega_n t}[\omega_n \sqrt{1 - \zeta^2} \cos(\omega_n \sqrt{1 - \zeta^2}\, t) - \zeta\omega_n \sin(\omega_n \sqrt{1 - \zeta^2}\, t)] \tag{6-85}$$

such that

$$\frac{di}{dt}(0+) = K_6 \omega_n \sqrt{1 - \zeta^2} = \frac{V_0}{L} \tag{6-86}$$

The particular solution for the oscillatory case is

$$i = \frac{V_0}{\omega_n L \sqrt{1 - \zeta^2}} e^{-\zeta\omega_n t} \sin(\omega_n \sqrt{1 - \zeta^2}\, t) \tag{6-87}$$

The variation of current with time for the oscillatory case is shown in Fig. 6-10(c). Since the current is the product of the damping factor and the oscillatory term, the damping factor represents an *envelope* or *boundary curve* for the oscillations. The attenuation factor determines how rapidly the oscillations are damped. As R approaches zero, the oscillations become undamped, and *sustained* oscillations result.

The physical meaning of this mathematical result might be interpreted in terms of an interchange of energy between the electric energy storage element (C) and the magnetic energy storage element (L). After the switch is closed, the energy which is stored in the electric field is transferred to the inductor as magnetic energy. When the current begins to decrease, energy is being returned to the electric field from the magnetic field. This interchange continues as long as any energy remains. If the resistance has zero value, the oscillatory current will be sustained indefinitely. However, if there *is* resistance present (and there always is in any practical circuit) the current flow through the resistor will cause energy to be dissipated, and the total energy will decrease with each cycle. Eventually all the energy will be dissipated and the current will be reduced to zero. If a scheme can be devised to supply the energy that is lost in each cycle, the oscillations can be sustained. This is accomplished in the vacuum-tube oscillator to produce audio frequency or radio frequency power.

6-3. Higher-order homogeneous differential equations

The method of solution discussed for first- and second-order differential equations may be followed in the solution of higher-order equations. For an nth order differential equation, the characteristic equation will be

$$a_0m^n + a_1m^{n-1} + \ldots + a_{n-1}m + a_n = 0 \tag{6-88}$$

A fundamental theorem of algebra states that an equation of order n has n roots. These roots can be found by factoring Eq. (6-88).

$$a_0(m - m_1)(m - m_2)\ldots(m - m_n) = 0 \tag{6-89}$$

Each root gives rise to a factor of the form $k_1e^{m_1t}$ in the solution. The sum of all such factors constitutes the solution of the differential equation. Thus, solution of higher-order homogeneous differential equations is primarily a matter of finding the roots of the characteristic equation.

Fortunately, there is some simplification in finding these roots because the coefficients of Eq. 6-88 are *positive* and *real* coefficients. This follows because these coefficients are made up of the system parameters, R, L, and C. And since R, L, and C must be positive and real (the only way they appear in nature), so must the a-coefficients.

There are three possible forms for the roots: (1) real roots, (2) imaginary roots, and (3) complex roots. For the first-order characteristic equation

$$a_0m + a_1 = 0 \tag{6-90}$$

the root is $m = -a_1/a_0$, which is negative and real because a_0 and a_1 are always positive and real. For a second-order characteristic equation

$$a_0m^2 + a_1m + a_2 = 0 \tag{6-91}$$

the roots are

$$m_1, m_2 = \frac{-a_1}{2a_0} \pm \frac{1}{2a_0} \sqrt{a_1^2 - 4a_0a_2} \tag{6-92}$$

With the positive real restrictions on the a-coefficients, these roots may have any of the three possible forms—real, imaginary, or complex. But if the roots are complex, they occur in conjugate pairs, since this is the only way complex roots can combine to give positive real coefficients. Thus, for characteristic equation roots to be complex, they must occur in conjugate pairs.

Consider next a third-order characteristic equation. In this case, because of the rule just given for complex roots, at least one root must be real. The other two may be both real or a conjugate pair of complex roots.* For a fourth-order characteristic equation, there are more possibilities: four real roots, two real roots and a conjugate complex pair, or two sets of conjugate complex roots. The general pattern is thus established and the following rules may be given:

(1) If the roots are complex, they occur in conjugate pairs.
(2) If the characteristic equation is of odd order, at least one root is real. The remaining roots may be real or occur in conjugate complex pairs.
(3) If the characteristic equation is of even order, the roots may be real or occur in conjugate complex pairs.

Summarizing this discussion, an equation of any order can be factored into its roots, and the roots determine the solution of the homogeneous differential equation as the sum of first-order (or second-order) solutions which have already been considered.

An example will illustrate the method of solution of higher order homogeneous differential equations. The differential equation

$$\frac{d^5i}{dt^5} + 6\frac{d^4i}{dt^4} + 17\frac{d^3i}{dt^3} + 28\frac{d^2i}{dt^2} + 24\frac{di}{dt} + 8i = 0 \tag{6-93}$$

has a characteristic equation which may be factored as

$$(m + 1)(m + 1)(m + 2)(m^2 + 2m + 4) = 0 \tag{6-94}$$

In this equation, there are two repeated real roots, one nonrepeated real root, and one conjugate complex pair with $\omega_n = 2$ and $\zeta = 0.5$. Using the equations already derived for first-order and second-order systems, we see that the solution is

$$i = (K_1 + K_2t)e^{-t} + K_3e^{-2t} + e^{-t}(K_4 \sin \sqrt{3}\, t + K_5 \cos \sqrt{3}\, t) \tag{6-95}$$

* In this discussion, imaginary roots are considered as a special case of complex roots.

6-4. Solution of nonhomogeneous differential equations

In the nonhomogeneous differential equation, the right-hand side of the equation is not zero, but equal to the forcing function or some derivative of the forcing function, $v(t)$. In studying such equations, we first observe that the solution to the corresponding homogeneous differential equation is a part of the solution of the nonhomogeneous equation. To illustrate by a simple example, consider the equation

$$\frac{d^2i}{dt^2} + 5\frac{di}{dt} + 6i = v(t) \tag{6-96}$$

This equation has as roots of its characteristic equation, $m_1 = -2$ and $m_2 = -3$. Thus the complete solution for the case $v(t) = 0$ is

$$i_C = k_1e^{-2t} + k_2e^{-3t} \tag{6-97}$$

Suppose that some function i_P, which we will presently find, satisfies the nonhomogeneous equation, Eq. 6-96. Then i_P plus i_C given above is *also* a solution, since substituting either k_1e^{-2t} or k_2e^{-3t} into Eq. 6-96 would add nothing to the right-hand side of the equation. In other words, part of the solution of a nonhomogeneous differential equation is the solution to the homogeneous differential equation. That part, by analogy to the discussion in Art. 4-3, is termed the *complementary function*. The remaining part of the solution—needed to make the operations of the differential equation add to $v(t)$—is the *particular integral*. Thus we write the total solution as the sum of two parts of the solution

$$i = i_P + i_C \tag{6-98}$$

Since we can find i_C for any equation, as discussed in the last section, there remains to be found only the particular integral i_P.

6-5. The particular integral by the method of undetermined coefficients

In the analysis of electric circuits, the term $v(t)$ in the differential equation is the driving force or a derivative of the driving force. As a practical matter, driving forces are represented by only a few mathematical forms like V (a constant), $\sin \omega t$, kt, $e^{-\alpha t}$, or products of these terms (or linear combinations to give square waves, pulses, etc.). We do not ordinarily encounter physical generators of such functions as the tangent. Several mathematical methods are available for determining the particular integral. If only driving forces of the practical forms mentioned are considered, the method of *undetermined coefficients* is particularly suited to our use.

Ordinarily, the method of undetermined coefficients is applied by selecting trial functions of all possible forms that might satisfy the differential equation. Each trial function is assigned an undetermined coefficient. The sum of the trial functions is substituted into the differential equation, and a set of linear algebraic equations is formed by equating coefficients of like functions in the equation resulting from this substitution. The undetermined coefficients are thus determined by solution of this set of equations. If any trial function is not a solution, its coefficient will be zero.

It is not necessary to study rules for selecting trial functions for the forms of driving force function $v(t)$ we are considering. The required form of the trial functions is given in Table 6-1. In using this table, the following procedure is suggested:

(1) Determine the complementary function i_C. Compare each part of the complementary function with the form of $v(t)$. The rules given in Table 6-1 are modified if these two functions have terms of the same mathematical form.

TABLE 6-1

Factor in $v(t)$*	Necessary choice for the particular integral†
1. V (a constant)	A
2. $a_1 t^n$	$B_0 t^n + B_1 t^{n-1} + \dots + B_{n-1} t + B_n$
3. $a_2 e^{rt}$	Ce^{rt}
4. $a_3 \cos \omega t$	$D \cos \omega t + E \sin \omega t$
5. $a_4 \sin \omega t$	
6. $a_5 t^n e^{rt} \cos \omega t$	$(F_1 t^n + \dots + F_{n-1} t + F_n) e^{rt} \cos \omega t$
7. $a_6 t^n e^{rt} \sin \omega t$	$+ (G_1 t^n + \dots + G_n) e^{rt} \sin \omega t$

* When $v(t)$ consists of a sum of several terms, the appropriate particular integral is the sum of the particular integrals corresponding to these terms individually.

† Whenever a term in any of the trial integrals listed in this column is already a part of the complementary function of the given equation, it is necessary to modify the indicated choice by multiplying it by t before using it. If such a term appears r times in the complementary function, the indicated choice should be multiplied by t^r.

(By permission from *Advanced Engineering Mathematics* by Wylie. Copyright, 1951. McGraw-Hill Book Co., Inc.)

(2) Write the trial form of the particular integral, using Table 6-1. Each different trial solution should be assigned a different letter coefficient, and all similar functions should be combined.

(3) Substitute the trial solution into the differential equation. By equating coefficients of all like terms, form a set of algebraic equations in the undetermined coefficients.

(4) Solve for the undetermined coefficients and so find the particular integral. These coefficients must be in terms of circuit and driving force parameters. There are no arbitrary constants in the particular integral.

Having determined the particular integral, the total solution may be found by adding the complementary function to the particular integral. If a particular solution is required, the arbitrary constants of i_C can be evaluated from a knowledge of the initial conditions. As a precaution, the initial conditions must always be applied to the total solution—never to the complementary function alone unless $i_P = 0$ [when $v(t) = 0$].

Example 4

Consider a series RL circuit with the driving force voltage of the form $v(t) = Ve^{-\alpha t}$, where V and α are constants. By Kirchhoff's voltage law, the differential equation is, after division by L,

$$\frac{di}{dt} + \frac{R}{L}\,i = \frac{V}{L}\,e^{-\alpha t} \tag{6-99}$$

The characteristic equation is $m + (R/L) = 0$, so that the complementary function is

$$i_C = ke^{-Rt/L} \tag{6-100}$$

From Table 6-1, the trial solution should be

$$i_P = Ae^{-\alpha t} \tag{6-101}$$

if $\alpha \neq R/L$, where A is the undetermined coefficient. Substituting this trial solution into the differential equation gives

$$-\alpha Ae^{-\alpha t} + \frac{R}{L}\,Ae^{-\alpha t} = \frac{V}{L}\,e^{-\alpha t} \tag{6-102}$$

or

$$A = \frac{V}{R - \alpha L}, \qquad \alpha \neq \frac{R}{L} \tag{6-103}$$

The solution is the sum of i_P and i_C, or

$$i = \frac{V}{R - \alpha L}\,e^{-\alpha t} + Ke^{-Rt/L}, \qquad \alpha \neq \frac{R}{L} \tag{6-104}$$

The arbitrary constant can be evaluated from knowledge of the initial conditions. If $\alpha = R/L$, the form of the trial solution should be

$$i_P = Ate^{-\alpha t} \tag{6-105}$$

Substituting this solution into the differential equation gives

$$A(-\alpha te^{-\alpha t} + e^{-\alpha t}) + \alpha Ate^{-\alpha t} = \frac{V}{L}e^{-\alpha t} \tag{6-106}$$

or

$$A = \frac{V}{L} \tag{6-107}$$

The solution for this case is thus

$$i = \frac{V}{L}te^{-\alpha t} + Ke^{-Rt/L}, \qquad \alpha = \frac{R}{L} \tag{6-108}$$

Example 5

As a second example, consider a series RC circuit with a sinusoidal driving force voltage $v(t) = V \sin \omega t$. The Kirchhoff voltage equation is

$$Ri + \frac{1}{C}\int i\,dt = V \sin \omega t \tag{6-109}$$

or, differentiating and dividing by R,

$$\frac{di}{dt} + \frac{1}{RC}i = \frac{\omega V}{R}\cos \omega t \tag{6-110}$$

From Table 6-1, the assumed i_P should be the sum of a sine and a cosine term, as

$$i_P = A \cos \omega t + B \sin \omega t \tag{6-111}$$

If this assumed solution is substituted into the differential equation and coefficients of like functions are equated, the following system of linear equations results.

$$\frac{A}{RC} + \omega B = \frac{\omega V}{R}, \qquad \frac{B}{RC} - \omega A = 0 \tag{6-112}$$

Solving for A and B yields

$$A = \frac{\omega CV}{1 + \omega^2R^2C^2}, \qquad B = \frac{\omega^2RC^2V}{1 + \omega^2R^2C^2} \tag{6-113}$$

Substituting these values into the assumed solution, there results, after some simplification,

$$i_P = \frac{V}{1/\omega^2C^2 + R^2}\left(\frac{1}{\omega C}\cos \omega t + R \sin \omega t\right) \tag{6-114}$$

This equation can be reduced to a single sinusoid by defining $1/\omega C = K \cos \phi$ and $R = K \sin \phi$, and making use of the trigonometric identity for the cosine of the difference of two angles. Finally,

$$i_P = \frac{V}{\sqrt{R^2 + 1/\omega^2 C^2}} \cos (\omega t - \phi) \tag{6-115}$$

where

$$\phi = \tan^{-1} \omega RC \tag{6-116}$$

To this value of i_P must be added $i_C = Ke^{-t/RC}$ for the complete solution.

Example 6

Knowledge of the response of systems with sinusoidal driving force voltages is important in studies of power generation and distribution systems. Consider the circuit equivalent of such a system shown in Fig. 6-11. The Kirchhoff voltage equation for this system is

$$L \frac{di}{dt} + Ri = V \sin (\omega t + \theta) \tag{6-117}$$

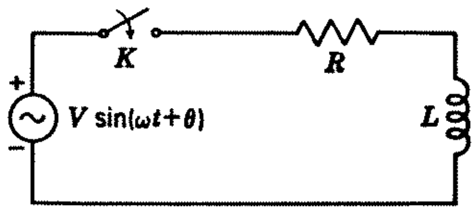


Fig. 6-11. *RL* series circuit.

The method for finding the particular integral is like that illustrated in the last example. The result is

$$i_P = \frac{V}{\sqrt{R^2 + \omega^2 L^2}} \sin \left(\omega t + \theta - \tan^{-1} \frac{\omega L}{R} \right) \tag{6-118}$$

To this result must be added the complementary function, which from Example 4 is

$$i_C = Ke^{-Rt/L} \tag{6-119}$$

The total solution thus becomes

$$i = \frac{V}{\sqrt{R^2 + \omega^2 L^2}} \sin \left(\omega t + \theta - \tan^{-1} \frac{\omega L}{R} \right) + Ke^{-Rt/L} \tag{6-120}$$

Now if the switch is closed at $t = 0$, the initial current has zero value because of the inductor, requiring that

$$\frac{V}{\sqrt{R^2 + \omega^2 L^2}} \sin \left(\theta - \tan^{-1} \frac{\omega L}{R} \right) + Ke^0 = 0 \tag{6-121}$$

or

$$K = - \frac{V}{\sqrt{R^2 + \omega^2 L^2}} \sin \left(\theta - \tan^{-1} \frac{\omega L}{R} \right) \tag{6-122}$$

If the angle θ, which represents the angle of the sinusoid at the time the switch is closed, has the value

$$\theta = \tan^{-1} \frac{\omega L}{R} \tag{6-123}$$

the constant K will have zero value, and the transient term i_C will vanish. In other words, if the switch is closed at the proper instant, there will be no transient. The same conclusion can be reached for the RC series network but not for an RLC network.

6-6. Capacitor charge in an RLC series circuit

The circuit shown in Fig. 6-12 is energized by closing the switch K at $t = 0$. It is desired to find charge on the capacitor as a function of time, $q(t)$. By Kirchhoff's voltage law, the differential equation for the charge is, after division by L,

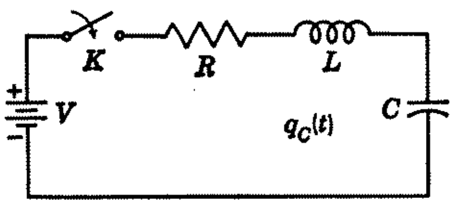


Fig. 6-12. RLC series circuit.

$$\frac{d^2q}{dt^2} + \frac{R}{L}\frac{dq}{dt} + \frac{1}{LC} q = \frac{V}{L} \tag{6-124}$$

This is a nonhomogeneous differential equation, and the solution will be composed of two parts. The particular integral or steady-state solution will be

$$q_{ss} = A \qquad \text{(a constant)} \tag{6-125}$$

Substituting this solution into the differential equation gives

$$0 + 0 + \frac{1}{LC} A = \frac{V}{L} \quad \text{or} \quad A = CV \tag{6-126}$$

For the three cases previously studied, corresponding to overdamped, critically damped or underdamped, the solutions are:

Case 1, $\zeta > 1$.

$$q = q_t + q_{ss} = CV + e^{-\zeta\omega_n t}[K_1 \cosh (\omega_n \sqrt{\zeta^2 - 1}\, t) + K_2(\sinh \omega_n \sqrt{\zeta^2 - 1}\, t)] \tag{6-127}$$

Case 2, $\zeta = 1$.

$$q = CV + (K_1 + K_2 t)e^{-\omega_n t} \tag{6-128}$$

Case 3, $\zeta < 1$.

$$q = CV + e^{-\zeta\omega_n t}[K_1 \cos (\omega_n \sqrt{1 - \zeta^2}\, t) + K_2 \sin (\omega_n \sqrt{1 - \zeta^2}\, t)] \tag{6-129}$$

Each solution has two arbitrary constants, K_1 and K_2. These constants will not be the same for the three cases, but they can be evaluated from the initial conditions. As initial conditions, we will assume that at $t = 0$, $q_C = 0$ (no initial charge on the capacitor) and $i_C = 0$, so $dq_C/dt = 0$ (no initial current because of the inductor). We will carry out the evaluation of the constants in some detail for Case 3. The two other cases follow a similar (and easier) pattern. Using the initial value of q condition gives

$$0 = CV + e^0(K_1 + K_2 \cdot 0)$$

or

$$K_1 = -CV \tag{6-130}$$

The derivative of q with respect to time is

$$\begin{aligned}\frac{dq}{dt} = {} & \\ & K_1 e^{-\zeta\omega_n t}[-\omega_n \sqrt{1-\zeta^2} \sin (\omega_n \sqrt{1-\zeta^2}\, t) - \zeta\omega_n \cos (\omega_n \sqrt{1-\zeta^2}\, t)] \\ & + K_2 e^{-\omega_n t}[\omega_n \sqrt{1-\zeta^2} \cos (\omega_n \sqrt{1-\zeta^2}\, t) - \zeta\omega_n \sin (\omega_n \sqrt{1-\zeta^2}\, t)]\end{aligned} \tag{6-131}$$

At $t = 0$, $dq/dt = 0$, so

$$0 = K_1(-\zeta\omega_n) + K_2(\omega_n \sqrt{1-\zeta^2}) \tag{6-132}$$

Hence

$$K_2 = K_1 \frac{\zeta}{\sqrt{1-\zeta^2}} = -CV \frac{\zeta}{\sqrt{1-\zeta^2}} \tag{6-133}$$

The particular solution is

$$\begin{aligned} q(t) = CV \Big\{ 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} & [\sqrt{1-\zeta^2} \cos (\omega_n \sqrt{1-\zeta^2}\, t) \\ & + \zeta \sin (\omega_n \sqrt{1-\zeta^2}\, t)] \Big\} \end{aligned} \tag{6-134}$$

This equation may also be written in the form

$$q(t) = CV \left[1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin (\omega_n \sqrt{1-\zeta^2}\, t + \theta) \right] \tag{6-135}$$

where $\theta = \tan^{-1} \sqrt{1-\zeta^2}/\zeta$.

For Case 2, the particular solution will be found to be

$$q(t) = CV[1 - (1 + \omega_n t)e^{-\omega_n t}] \tag{6-136}$$

and for Case 1, the particular solution is

$$\begin{aligned} q(t) = CV \Big\{ 1 - e^{-\zeta\omega_n t} \Big[& \cosh (\omega_n \sqrt{\zeta^2 - 1}\, t) \\ & + \frac{\zeta}{\sqrt{\zeta^2-1}} \sinh (\omega_n \sqrt{\zeta^2 - 1}\, t) \Big] \Big\} \end{aligned} \tag{6-137}$$

These three equations give $q(t)$ for the three conditions of damping. Since the damping ratio is determined as the ratio of actual resistance to critical resistance, the three conditions could be obtained merely by an adjustment of the resistor R. Figure 6-13 shows the behavior

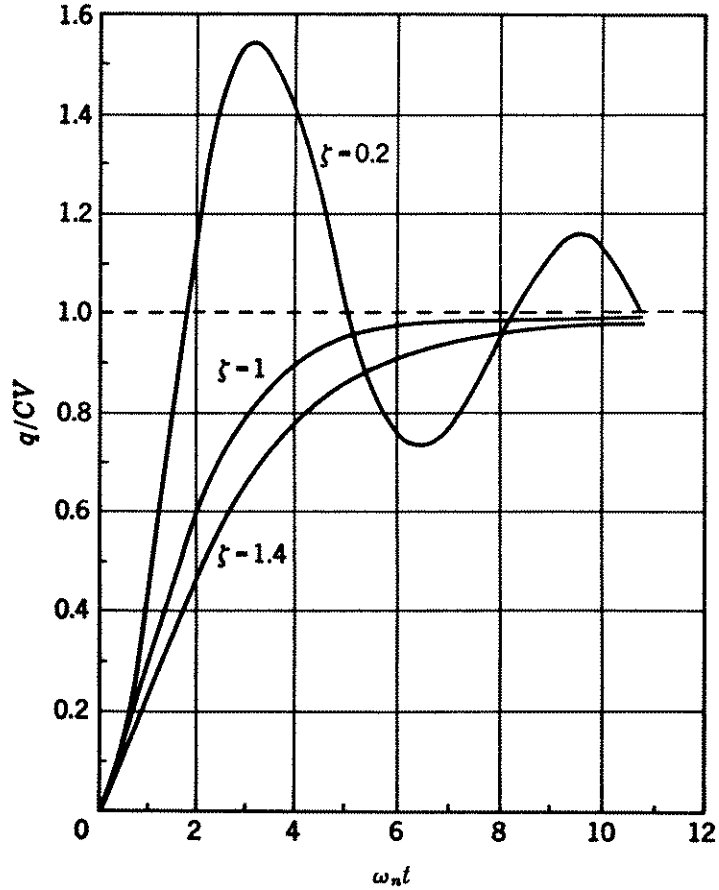


Fig. 6-13. System response for three values of ζ illustrating the underdamped ($\zeta = 0.2$), critically damped ($\zeta = 1.0$), and overdamped ($\zeta = 1.4$) cases.

of the charge in the circuit as the damping ratio changes. These solutions are valid for any second-order system with these particular initial conditions, not only in electric networks but also in mechanical and electromechanical systems (for example, in servomechanisms).

FURTHER READING

Solution of differential equations of the type discussed in this chapter is concisely treated by Wylie in *Advanced Engineering Mathematics* (McGraw-Hill Book Co., Inc., New York, 1951), pp. 1–45. See also Chap. 4, titled "Classical Analysis of Double-Energy Transients" in Fich, *Transient Analysis in Electrical Engineering* (Prentice-Hall, Inc., New York, 1951). Other texts recommended for supplementary reading include: Skilling, *Transient Electric Currents* (McGraw-Hill

Book Co., Inc., New York, 1952), Chaps. 3 and 4; Johnson, *Mathematical and Physical Principles of Engineering Analysis* (McGraw-Hill Book Co., Inc., New York, 1944), Chap. 6; Salvadori and Schwarz, *Differential Equations in Engineering Problems* (Prentice-Hall, Inc., New York, 1954), Chaps. 3 and 4.

PROBLEMS

6-1. Show that $i = ke^{-2t}$ and $i = ke^{-t}$ are solutions of the differential equation

$$\frac{d^2i}{dt^2} + 3\frac{di}{dt} + 2i = 0$$

6-2. Show that $i = ke^{-t}$ and $i = kte^{-t}$ are solutions of the differential equation

$$\frac{d^2i}{dt^2} + 2\frac{di}{dt} + i = 0$$

6-3. Find the general solution each of the following equations:

(a) $\dfrac{d^2i}{dt^2} + 3\dfrac{di}{dt} + 2i = 0$

(b) $\dfrac{d^2i}{dt^2} + 5\dfrac{di}{dt} + 6i = 0$

(c) $\dfrac{d^2i}{dt^2} + 7\dfrac{di}{dt} + 12i = 0$

(d) $\dfrac{d^2v}{dt^2} + 5\dfrac{dv}{dt} + 4v = 0$

(e) $\dfrac{d^2x}{dt^2} + \dfrac{dx}{dt} + 6x = 0$

(*f*) $\dfrac{d^2y}{dx^2} + \dfrac{dy}{dx} + 2y = 0$

(g) $\dfrac{d^2q}{dt^2} + 2\dfrac{dq}{dt} + q = 0$

(h) $\dfrac{d^2i}{dt^2} + 4\dfrac{di}{dt} + 4i = 0$

6-4. Find the particular solution of the differential equation of Prob. 6-3(a) and Prob. 6-3(b) subject to the initial conditions:

$$i(0+) = 1, \qquad \frac{di}{dt}(0+) = 0$$

Answers. $i = 2e^{-t} - e^{-2t}$, $i = 3e^{-2t} - 2e^{-3t}$.

6-5. Repeat Prob. 6-4 for the differential equations of Prob. 6-3, parts (c) through (h) subject to the initial conditions

$$\text{variable } (0+) = 2, \frac{d}{dt}[\text{variable } (0+)] = 1$$

6-6. Write the general solution of the differential equations with the following characteristic equations:

(a) $(m + 1)(m + 2)(m^2 + m + 1) = 0$

(b) $(m + 1)^2(m + 5) = 0$

(c) $(m + a)(m + b)(m - c)(m - d) = 0$

6-7. Plot the roots of the characteristic equation in the m-plane ($m = \sigma + j\omega$) for $L = 1$ henry, $C = 1$ μf, and (a) $R = 500$ ohms, (b) $R = 1000$ ohms, (c) $R = 2000$ ohms, (d) $= 3000$ ohms, (e) $R = 4000$ ohms.

6-8. For the system described in Prob. 6-7 and for (a) $R = 500$ ohms, (b) $R = 2000$ ohms, and (c) $R = 4000$ ohms, find the general solution to the differential equation. Evaluate *all* coefficients but not the arbitrary constants.

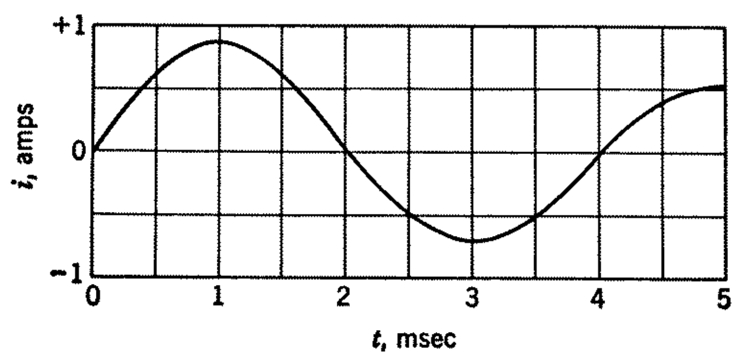
6-9. In a certain network, it is found that the current is given by the expression

$$i = K_1 e^{-\alpha_1 t} - K_2 e^{-\alpha_2 t}, \qquad t > 0, \qquad \alpha_1 > \alpha_2$$

Show that $i(t)$ reaches a maximum value at time

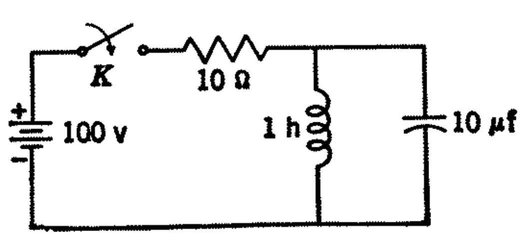
$$t = \frac{1}{\alpha_1 - \alpha_2} \ln \frac{\alpha_1 K_1}{\alpha_2 K_2}$$

6-10. The graph shown is a record of the current as a function of time resulting when a switch is closed at $t = 0$, connecting a battery to a network. Only slightly more than a cycle is shown in the record, but the current eventually reaches zero value. (a) Determine the values of ζ and ω_n for the current waveform. (b) Write the equation of current as a function of time with all coefficients evaluated.



Prob. 6-10.

6-11. In the network shown in the accompanying figure, the switch K is opened at $t = 0$ with equilibrium conditions existing before the the switch is opened. Find the current through the inductor as a function of time. *Answer.* $i_L(t) = 10 \cos 316t$.



Prob. 6-11.

6-12. Show that Eq. 6-67 can be written in the form

$$i = Ke^{-\zeta\omega_n t} \cos\,(\omega_n \sqrt{1 - \zeta^2}\, t + \phi)$$

Give the values for K and ϕ in terms of K_5 and K_6 of Eq. 6-67.

6-13. Solve the following nonhomogeneous differential equations.

(a) $\dfrac{d^2i}{dt^2} + 2\dfrac{di}{dt} + i = 1$

(b) $\dfrac{d^2i}{dt^2} + 3\dfrac{di}{dt} + 2i = 5t$

Answer. $i = \frac{5}{2}t - \frac{15}{4} + k_1e^{-t} + k_2e^{-2t}$.

(c) $\dfrac{d^2i}{dt^2} + 3\dfrac{di}{dt} + 2i = 10 \sin 10t$

(d) $\dfrac{d^2q}{dt^2} + 5\dfrac{dq}{dt} + 6q = te^{-t}$

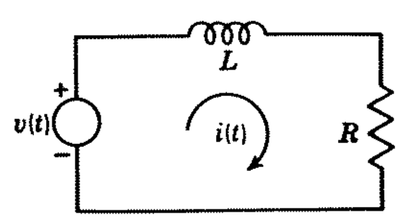
Answer. $q = \dfrac{te^{-t}}{2} - \dfrac{3}{4}e^{-t} + k_1e^{-2t} + k_2e^{-3t}$

(e) $\dfrac{d^2v}{dt^2} + 5\dfrac{dv}{dt} + 6v = e^{-2t} + 5e^{-3t}$

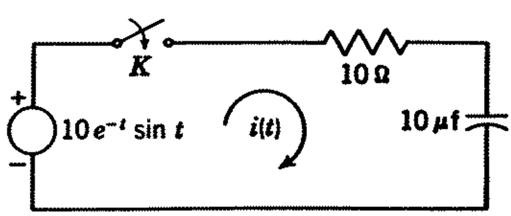
Answer. $v = (K_1 + t)e^{-2t} + (K_2 - 5t)e^{-3t}$.

6-14. A special generator has a voltage variation given by the equation $v(t) = t$ volts, where t is the time in seconds. This generator is connected to an RL series circuit, where $R = 2$ ohms and $L = 1$ henry, at time $t = 0$ by the closing of a switch. Find the equation for the current as a function of time $i(t)$. *Answer.* $i(t) = \frac{1}{2}t - \frac{1}{4} + \frac{1}{4}e^{-2t}$.

6-15. A bolt of lightning having a waveform which is approximated as $v(t) = te^{-t}$ strikes a transmission line having resistance $R = 0.1$ ohm and inductance $L = 0.1$ henry (the line-to-line capacitance is assumed negligible). An equivalent network is shown in the accompanying diagram. What is the form of the current as a function of time? (This current will be in amperes per unit volt of the lightning; likewise the time base is normalized.) *Answer.* $i(t) = 5t^2e^{-t}$.



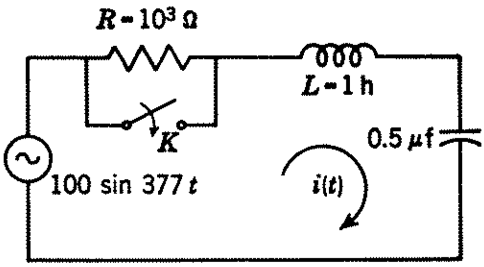
Prob. 6-15.



Prob. 6-16.

6-16. In the circuit shown in the figure, solve for $i(t)$ if K is closed at $t = 0$. *Answer.* $i = 10^{-4}e^{-t}\,(\cos t - \sin t) - 10^{-4}e^{-10^4t}$.

6-17. In the network shown, the switch K is closed at $t = 0$, with a steady state having been established previous to this time. For the parameter values shown in the diagram, find the current as a function of time $i(t)$. *Answer.* $i = (20.3 \cos 377t - 0.8 \cos 1414t + 1485 \sin 1414t) \times 10^{-3}$ amp.



Prob. 6-17.

6-18. On page 118, the statement is made that the transient term in the solution cannot vanish in the *RLC* series circuit with a sinusoidal driving force merely by closing a switch at an appropriate time. Prove that this statement is correct.

6-19. Starting with Eq. 6-127, verify that Eq. 6-137 is the particular solution with the initial conditions given for Case 3.

6-20. A switch is closed at $t = 0$ connecting a battery of voltage V with a series *RL* circuit. (a) Show that the energy in the resistor as a function of time is

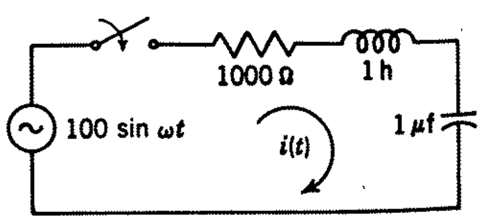
$$w_R = \frac{V^2}{R}\left(t + \frac{2L}{R}e^{-Rt/L} - \frac{L}{2R}e^{-2Rt/L} - \frac{3L}{2R}\right)$$

(b) Find an expression for the energy in the magnetic field as a function of time. (c) Sketch w_R and w_L as a function of time. Show the steady-state asymptotes, that is, the values that w_R and w_L approach as $t \to \infty$. (d) Find the total energy supplied by the voltage source in the steady state.

6-21. In the series *RLC* circuit shown in the accompanying diagram, the frequency of the driving force voltage is

(1) $\omega = \omega_n$ (the undamped natural angular frequency)
(2) $\omega = \omega_n\sqrt{1 - \zeta^2}$ (the natural angular frequency)

These frequencies are applied in two separate experiments. In each experiment we measure (a) the peak value of the transient current when the switch is closed at $t = 0$, and (b) the maximum value of the steady-state current. (a) In which case (that is, which frequency) is the maximum value of the transient greater? (b) In which case (that is, which frequency) is the maximum value of the steady-state current greater?



Prob. 6-21.

CHAPTER 7

THE LAPLACE TRANSFORMATION

7-1. Introduction

The forerunner of the Laplace transformation method of solving differential equations, the *operational calculus*, was invented by the brilliant English engineer Oliver Heaviside (1850–1925). Heaviside was a practical man and his interest was in the practical solution of electric circuit problems rather than careful justification of his methods. He was gifted with an insight into physical problems that enabled him to pick the correct solution from a number of alternatives. This heuristic point of view drew bitter and perpetual criticism from the leading mathematicians of his time. In the years that followed publication of Heaviside's work, the rigor was supplied by such men as Bromwich, Giorgi, Carson, and others. The basis for substantiating the work of Heaviside was found in the writings of Laplace in 1780. As the years have passed, the structural members of the framework of Heaviside's operational calculus have been replaced, piece by piece, by new members derived by the Laplace transformation. This transformation has provided rigorous substantiation of the operational methods; no important errors have been discovered in Heaviside's results.

The Laplace transformation method for solving differential equations offers a number of advantages over the classical methods that were discussed in Chapters 4 and 6. For example:

(1) The solution of differential equations is routine and progresses systematically.
(2) The method gives the total solution—the particular integral and the complementary function—in one operation.
(3) Initial conditions are automatically specified in the transformed equations. Further, the initial conditions are incorporated into the problem as one of the first steps rather than as the last step.

What is a transformation? The *logarithm* is an example of a transformation that we have used in the past. Logarithms greatly simplify such operations as multiplication, division, extracting roots, and raising quantities to powers. Suppose that we have two numbers, given to seven-place accuracy, and we are required to find the product, maintaining the accuracy of the given numbers. Rather than just multiplying the two numbers together, we transform these numbers by taking their logarithm. These logarithms are added (or subtracted in the case of division). The resulting sum itself has little meaning. However, if we perform an *inverse transformation*, if we find the antilogarithm, then we have the desired numerical result. The direct division looks more straightforward, but our experience has been that the use of the logarithm often saves time. If the simple problem of multiplying two numbers is not convincing, consider evaluating $(1437)^{0.1328}$ without logarithms!

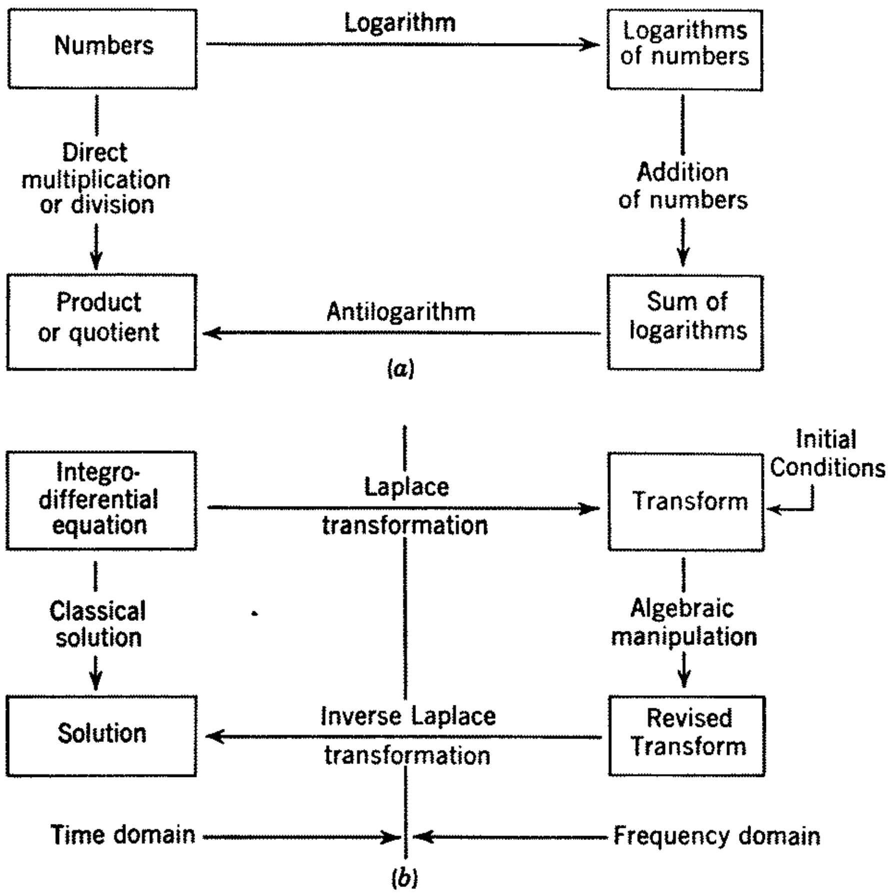


Fig. 7-1. Comparison of logarithms and the Laplace transformation.

A flow sheet of the operation of using logarithms to find a product of a quotient is shown in Fig. 7-1. The individual steps are: (1) find the logarithm of the separate numbers, (2) add or subtract the numbers to obtain the sum of logarithms, and (3) take the antilogarithm to obtain the product or quotient. This is roundabout compared with *direct* multiplication or division, yet we use logarithms to advantage, particularly when a good table of logarithms is available.

The flow sheet idea may be used to illustrate what we will do in using the Laplace transformation to solve a differential equation. The flow sheet for the Laplace transformation is also shown in Fig. 7-1 with a block corresponding to every block of the logarithm flow sheet considered above. The steps will be as follows. (1) Start with an integro-differential equation and find the corresponding Laplace *transform*. This is a mathematical process, but there are tables of transforms just as there are tables of logarithms (and one is included in this chapter). (2) The transform is manipulated algebraically after the initial conditions are inserted. The result is a *revised transform*. As step (3), we perform an inverse Laplace transformation to give us the solution. In this step, we also can use a table of transforms, just as we use the table of logarithms in the corresponding step for logarithms. The flow sheet reminds us that *there is another way:* the classical solution. It looks more direct (and sometimes it is for simple problems). For complicated problems, an advantage will be found for the Laplace transformation, just as an advantage was found for the use of logarithms.

7-2. The Laplace transformation

To construct a Laplace transform for a given function of time $f(t)$, we first multiply $f(t)$ by e^{-st}, where s is a complex number, $s = \sigma + j\omega$. This product is integrated with respect to time from zero to infinity. The result is the Laplace transform of $f(t)$, which is designated $F(s)$. Denoting the Laplace transformation by the script letter $\mathcal{L}$ (in order to reserve L for inductance), the Laplace transformation is given by the mathematical expression

$$\mathcal{L}[f(t)] = F(s) = \int_0^\infty f(t)e^{-st}\,dt \tag{7-1}$$

The letter $\mathcal{L}$ can be replaced by the words "the Laplace transform of" in the above expression.

Although this equation is a rather formidable appearing integral, the actual evaluation of $F(s)$ for a given $f(t)$ is usually not difficult. Furthermore, once the transform of a function is found, it need not be found again for a new problem but can be tabulated. The time function $f(t)$ and the transform $F(s)$ of this function are called *transform pairs*. A table of transform pairs is given on page 146.

The operation which changes a function of s back to a function of time is called the *inverse Laplace transformation*. This operation is symbolized as $\mathcal{L}^{-1}$. Then by definition,

$$\mathcal{L}^{-1}\{\mathcal{L}[f(t)]\} = \mathcal{L}^{-1}[F(s)] = f(t) \tag{7-2}$$

The inverse Laplace transformation is given by the *complex inversion integral*,

$$f(t) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} F(s)e^{st}ds \tag{7-3}$$

where c is a constant. This integral is seldom used because of the uniqueness property of the Laplace transformation that there is a one-to-one correspondence between the direct and inverse transforms. In other words, a table can be used to find the $f(t)$ corresponding to a given $F(s)$ as well as an $F(s)$ for a given $f(t)$.

Not all functions can be integrated to find the Laplace transforms. The usual requirements for an $f(t)$ are that it be (1) piecewise continuous and (2) of exponential order. These requirements are discussed in detail in references cited at the end of this chapter. Suffice to say that all functions of engineering interest have the properties required for the existence of the transform.

As an example of the evaluation of Eq. 7-1, let $f(t) = 1$; then

$$\mathcal{L}[1] = \int_0^\infty e^{-st}\, dt = -\frac{1}{s} e^{-st}\Big]_0^\infty = \frac{1}{s} \tag{7-4}$$

Because of the lower limit of the integral, the value of $f(t)$ for $t < 0$ does not enter into the final solution. Thus the Laplace transform for $f(t) = 1$ is the same as that of a special function having zero value for $t < 0$ and unit value for $t > 0$. Such a function was called a *unit step function* by Heaviside. We will use the symbol $u(t)$ for a function described mathematically as

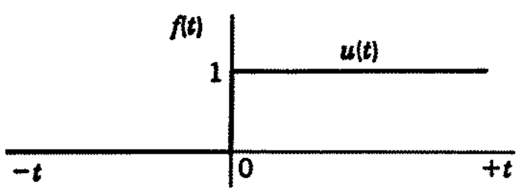


Fig. 7-2. Unit step function.

$$u(t) = \begin{cases} 0 & t < 0 \\ 1 & t > 0 \end{cases} \tag{7-5}$$

and shown as a function of time in Fig. 7-2. Such a function is the mathematical equivalent of physically closing a switch at $t = 0$. If a battery of voltage V is connected to a network at $t = 0$ by closing a switch, that voltage can be described as $Vu(t)$; that is, V times unity for $t > 0$ and V times zero for $t < 0$. The Laplace transform for $Vu(t)$ is

$$\mathcal{L}[Vu(t)] = \frac{V}{s} \tag{7-6}$$

As a second example of the calculation of a transform, let $f(t) = e^{at}$, where a is a constant. Substituting into Eq. 7-1, we have

$$\mathcal{L}[e^{at}] = \int_0^\infty e^{at}e^{-st}\, dt = \int_0^\infty e^{-(s-a)t}\, dt = \frac{1}{s-a}, \qquad \sigma > a \tag{7-7}$$

Thus e^{at} and $1/(s - a)$ constitute a transform pair.

For one further example, let $f(t) = \sin \omega t$. Substituting into the defining equation, we have

$$\mathcal{L}[\sin \omega t] = \int_0^\infty (\sin \omega t)e^{-st}\,dt = \frac{e^{-st}(-s\sin \omega t - \omega \cos \omega t)}{s^2 + \omega^2}\Bigg]_0^\infty$$
$$= \frac{\omega}{s^2 + \omega^2} \tag{7-8}$$

These two computations can form the beginning of a table of transform pairs as shown below.

TABLE OF TRANSFORM PAIRS

$f(t)$	$F(s)$
$u(t)$	$1/s$
e^{at}	$\dfrac{1}{s-a}$
$\sin \omega t$	$\dfrac{\omega}{s^2+\omega^2}$

More pairs can be added as they are computed. Exhaustive tables are to be found in reference books cited at the end of the chapter.

7-3. Basic theorems for the Laplace transformation

(1) Transforms of Linear Combinations. If $f_1(t)$ and $f_2(t)$ are two functions of time and a and b are constants, then

$$\mathcal{L}[af_1(t) + bf_2(t)] = aF_1(s) + bF_2(s) \tag{7-9}$$

This theorem is established with Eq. 7-1. It follows from the fact that the integral of a sum of terms is equal to the sum of the integrals of the terms; that is,

$$\begin{aligned}\mathcal{L}[af_1(t) + bf_2(t)] &= \int_0^\infty [af_1(t) + bf_2(t)]e^{-st}\,dt \\ &= a\int_0^\infty f_1(t)e^{-st}\,dt + b\int_0^\infty f_2(t)e^{-st}\,dt \\ &= aF_1(s) + bF_2(s)\end{aligned} \tag{7-10}$$

We will make use of this theorem in taking the Laplace transformation of the sum of derivative terms appearing in a differential equation.

(2) Transforms of Derivatives. From the defining equation for the Laplace transformation, we write

$$\mathcal{L}\left[\frac{d}{dt}f(t)\right] = \int_0^\infty \frac{d}{dt}f(t)e^{-st}\,dt \tag{7-11}$$

This equation may be integrated by parts by letting

$$u = e^{-st} \quad \text{and} \quad dv = df(t)$$

in the equation

$$\int_a^b u\,dv = uv\Big]_a^b - \int_a^b v\,du \tag{7-12}$$

Then $$du = -se^{-st}\,dt \quad \text{and} \quad v = f(t)$$

so that the transform of a derivative becomes

$$\mathcal{L}\left[\frac{d}{dt}f(t)\right] = e^{-st}f(t)\Big]_0^\infty + s\int_0^\infty f(t)e^{-st}\,dt = sF(s) - f(0+) \tag{7-13}$$

provided $\lim_{t\to\infty} f(t)e^{-st} = 0$, which follows from l'Hospital's rule, provided that $f(t)$ and all its derivatives are not infinite at $t = \infty$.

To find the transform of the second derivative, we follow a similar procedure but make use of the result of Eq. 7-13. Since

$$\frac{d^2}{dt^2}f(t) = \frac{d}{dt}\frac{d}{dt}f(t) \tag{7-14}$$

then

$$\begin{aligned}\mathcal{L}\left[\frac{d^2f(t)}{dt^2}\right] &= s\mathcal{L}\left[\frac{df}{dt}(t)\right] - \frac{df}{dt}(0+) \\ &= s[sF(s) - f(0+)] - \frac{df}{dt}(0+) \\ &= s^2F(s) - sf(0+) - \frac{df}{dt}(0+)\end{aligned} \tag{7-15}$$

In this expression, the quantity $df/dt\,(0+)$ is the derivative of $f(t)$ evaluated at $t = 0+$ (the time immediately after switching action is initiated). The general expression for an nth order derivative is

$$\mathcal{L}\frac{d^nf(t)}{dt^n} = s^nF(s) - s^{n-1}f(0+) - s^{n-2}\frac{df}{dt}(0+) - \cdots - \frac{d^{n-1}}{dt^{n-1}}f(0+) \tag{7-16}$$

(3) *Transforms of Integrals.* The transform for an integral expression, is found by starting with the equation

$$\int_0^\infty f(t)e^{-st}\,dt = F(s) \tag{7-17}$$

and integrating by parts to give

$$F(s) = e^{-st}\int f(t)\,dt\Big]_0^\infty + s\int_0^\infty\left[\int f(t)\,dt\right]e^{-st}\,dt \tag{7-18}$$

The quantity

$$e^{-st}\int f(t)\,dt\Big]_0^\infty$$

becomes zero at the upper limit, and at the lower limit has the value of the integral evaluated at $t = 0$, usually written by the notation

$$\int f(t)\,dt\bigg|_{t=0+} = f^{(-1)}(0+) \tag{7-19}$$

where $f^{(-1)}$ indicates integration. The last term of Eq. 7-18 is recognized as s times the Laplace transform of $\int f(t)\,dt$. Rearranging Eq. 7-18, there results

$$\mathcal{L}\left[\int f(t)\,dt\right] = \frac{F(s)}{s} + \frac{f^{(-1)}(0+)}{s} \tag{7-20}$$

Similarly, it is found that

$$\mathcal{L}\left[\iint f(t)\,dt^2\right] = \frac{F(s)}{s^2} + \frac{f^{(-1)}(0+)}{s^2} + \frac{f^{(-2)}(0+)}{s} \tag{7-21}$$

In the analysis of networks on the loop basis, $f(t)$ is often a current $i(t)$ and the integral of the current is the charge $q(t)$. Equation 20 then has the form

$$\mathcal{L}\left[\int i(t)\,dt\right] = \frac{I(s)}{s} + \frac{q(0+)}{s} \tag{7-22}$$

where $q(0+)$ is the charge (say on the plates of a capacitor) at the time $t = 0+$.

If $f(t)$ is a voltage, then

$$\mathcal{L}\left[\int v(t)\,dt\right] = \frac{V(s)}{s} + \frac{\psi(0+)}{s} \tag{7-23}$$

since the integral of voltage is flux linkage ψ.

7-4. Examples of the solution of problems with the Laplace transformation

With the short table of transforms that has been given on page 129 and the three basic theorems that have been derived in the previous section, we are now equipped to solve a network problem (elementary as yet, to be sure) using the Laplace transformation.

Example 1

For this example, we will write the Kirchhoff voltage law for a series RC network shown in Fig. 7-3. It will be assumed that the switch K is closed at $t = 0$. This information will be included in the formation of the network equations by writing the voltage expression as $Vu(t)$.

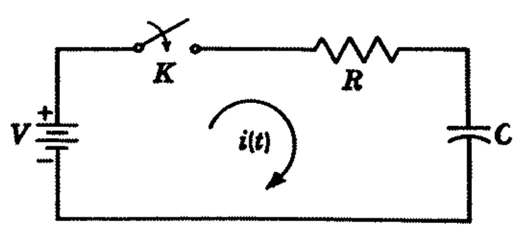


Fig. 7-3. RC series circuit.

Hence

$$\frac{1}{C}\int i\,dt + Ri = Vu(t) \tag{7-24}$$

This is the integral equation we wish to solve. Using Eq. 7-20 for the first term, we take the transforms of the linear combination of terms as

$$\frac{1}{C}\left[\frac{I(s)}{s} + \frac{q(0+)}{s}\right] + RI(s) = V \cdot \frac{1}{s} \tag{7-25}$$

In terms of the flow chart of Fig. 7-1, we have taken the Laplace transformation of the integral equation and there has resulted a transform expression. The required initial conditions are automatically specified and may be inserted as the second step (rather than as the final step as in differential equations solved by classical methods). Now $q(0+)$ is the charge on the capacitor at $t = 0+$. If the capacitor is initially uncharged, $q(0+) = 0$ and the last equation reduces to the form

$$I(s)\left(\frac{1}{Cs} + R\right) = \frac{V}{s} \tag{7-26}$$

The next step, again according to the flow chart, is algebraic manipulation. The objective of this manipulation is to solve for $I(s)$. This is accomplished by multiplying by s and dividing by R to give

$$I(s) = \frac{V/R}{s + 1/RC} \tag{7-27}$$

which is a "revised transform" expression. The next step on our flow chart is to perform the inverse Laplace transformation and obtain the solution. That is,

$$\mathcal{L}^{-1}[I(s)] = \mathcal{L}^{-1}\left(\frac{V/R}{s + 1/RC}\right) = i(t) \tag{7-28}$$

Using the second transform pair of our short table, the solution is

$$i(t) = \frac{V}{R}e^{-t/RC} \tag{7-29}$$

This is the complete solution (the steady state in this case being of zero value). The arbitrary constant emerges evaluated (and has the magnitude V/R).

Example 2

As our second example, consider the *RL* series circuit shown in Fig. 7-4. As in Example 1, the switch is closed at $t = 0$. The differential equation for the circuit is, by Kirchhoff's law,

$$L\frac{di}{dt} + Ri = Vu(t) \tag{7-30}$$

The corresponding transform equation is

$$L[sI(s) - i(0+)] + RI(s) = \frac{V}{s} \tag{7-31}$$

The initial condition specified by the last equation is $i(0+)$, the current after the switch is closed. Because of the inductance, $i(0+) = 0$. Our equation may now be manipulated to solve for $I(s)$; thus

$$I(s) = \frac{V}{L}\frac{1}{s(s + R/L)} \tag{7-32}$$

This transform, however, is not in our short table. We need something new (or a larger table). Notice that this term is made up of the product of the term $(1/s)$ and the term $[1/(s + R/L)]$. We know the inverse Laplace transformation of each of these individual terms. This suggests that the inverse operation could be performed if there were some way to break the transform terms into several parts. As an attempt to perform this operation, let us try the following expansion:

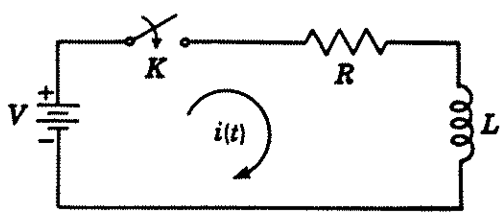


Fig. 7-4. *RL* series circuit.

$$\frac{V/L}{s(s + R/L)} = \frac{K_0}{s} + \frac{K_1}{s + R/L} \tag{7-33}$$

In this equation K_0 and K_1 are unknown coefficients. As the first step, let us simplify the equation by putting all terms over a common denominator. Then

$$\frac{V}{L} = K_0\left(s + \frac{R}{L}\right) + K_1 s$$

By equating coefficients of like functions, we obtain a set of linear algebraic equations:

$$K_0 \cdot \frac{R}{L} = \frac{V}{L}, \qquad K_0 + K_1 = 0$$

From these two equations, we find the required values for K_0 and K_1:

$$K_0 = \frac{V}{R} \quad \text{and} \quad K_1 = -\frac{V}{R}$$

This algebraic manipulation has permitted Eq. 7-32 to be written

$$I(s) = \frac{V}{L} \frac{1}{s(s + R/L)} = \frac{V}{R}\left[\frac{1}{s} - \frac{1}{s + R/L}\right] \tag{7-34}$$

We have transform pairs corresponding to each of these expressions. The current as a function of time is found by taking the inverse Laplace transformation of the individual expressions; thus

$$i(t) = \frac{V}{R}\left(\mathcal{L}^{-1}\frac{1}{s} - \mathcal{L}^{-1}\frac{1}{s + R/L}\right) \tag{7-35}$$

or

$$i(t) = \frac{V}{R}(1 - e^{-Rt/L}) \tag{7-36}$$

This is the final (time-domain) solution. The method we used to expand a transform into the sum of several separate parts is known under the heading of *partial fraction expansion.* It is this subject that we study next.

7-5. Partial fraction expansion

The examples of the last section have suggested the general procedure in applying the Laplace transformation to the solution of integrodifferential equations. A differential equation of the general form

$$a_0 \frac{d^n i}{dt^n} + a_1 \frac{d^{n-1} i}{dt^{n-1}} + \ldots + a_{n-1}\frac{di}{dt} + a_n i = v(t) \tag{7-37}$$

becomes, as a result of the Laplace transformation, an algebraic equation which may be solved for the unknown as

$$I(s) = \frac{\mathcal{L}[v(t)] + \text{initial condition terms}}{a_0 s^n + a_1 s^{n-1} + \ldots + a_{n-1}s + a_n} \tag{7-38}$$

The general form of this equation is a quotient of polynomials in s. Let the numerator and denominator polynomials be designated $P(s)$ and $Q(s)$, respectively, as

$$I(s) = \frac{P(s)}{Q(s)} \tag{7-39}$$

If the transform term $P(s)/Q(s)$ can now be found in a table of transform pairs, the solution $i(t)$ can be written directly. In general, however, the transform expression for $I(s)$ must be broken into simpler terms before any practical transform table can be used. To simplify the expression by partial fraction expansion, we find the roots of the denominator polynomial

$$Q(s) = a_0 s^n + a_1 s^{n-1} + \ldots + a_n = a_0(s + s_1)\ldots(s + s_n) \tag{7-40}$$

Because the coefficients $a_0, a_1, \ldots, a_n$ are positive and real (being functions of network parameters R, L, and C), the roots of $Q(s)$ are restricted to be: (1) simple and real, (2) conjugate complex pairs, or (3) repeated (or nonsimple) as discussed in Art. 6-3.

The rules for expanding $I(s)$ by partial fractions are given in terms of the three possibilities just mentioned:

(1) If all the roots are simple (that is, not repeated), then the partial fraction expansion is

$$\frac{P(s)}{(s+s_1)(s+s_2)\ldots(s+s_n)} = \frac{K_1}{s+s_1} + \frac{K_2}{s+s_2} + \cdots + \frac{K_n}{s+s_n} \tag{7-41}$$

(2) If a root is repeated r times, the partial fraction expansion corresponding to this one (repeated) root is

$$\frac{P(s)}{(s+s_1)^r} = \frac{K_{11}}{s+s_1} + \frac{K_{12}}{(s+s_1)^2} + \cdots + \frac{K_{1r}}{(s+s_1)^r} \tag{7-42}$$

and there will be similar terms for every other repeated root.

(3) An important special rule may be given for two roots which form a complex conjugate pair. For this case, the partial fraction expansion is

$$\frac{P(s)}{Q_1(s)(s+\alpha+j\omega)(s+\alpha-j\omega)} = \frac{K_1}{(s+\alpha+j\omega)} + \frac{K_1^*}{(s+\alpha-j\omega)} + \cdots \tag{7-43}$$

where K_1^* is the complex conjugate of K_1. In other words, when the roots are conjugates, so are the partial fraction expansion coefficients. An expansion of the type shown above is necessary for each pair of complex conjugate roots.

Although the above three rules are sufficient to expand any quotient of polynomials, it is sometimes more convenient to expand second-order denominator terms as

$$\frac{P(s)}{Q_1(s)(s^2+as+b)(s^2+cs+d)} = \frac{As+B}{s^2+as+b} + \frac{Cs+D}{s^2+cs+d} + \cdots$$
$$= \frac{As+B}{(s+\alpha_1)^2+\omega_1^2} + \frac{Cs+D}{(s+\alpha_2)^2+\omega_2^2} + \cdots \tag{7-44}$$

In an expansion of a quotient of polynomials by partial fractions, it may be necessary to use a combination of the three rules given above. Several examples will illustrate the expansion and the determination of the K's.

Example 3

Consider the quotient of polynomials,

$$I(s) = \frac{2s + 3}{s^2 + 3s + 2}$$

The first step is to factor the denominator polynomial and then expand by the appropriate rule. For this example, the expansion is

$$\frac{2s + 3}{(s + 1)(s + 2)} = \frac{K_1}{s + 1} + \frac{K_2}{s + 2}$$

since the roots are real and not repeated. Placing each factor over the common denominator of the equation gives

$$2s + 3 = (s + 2)K_1 + (s + 1)K_2$$

or

$$2s + 3 = (K_1 + K_2)s + (2K_1 + K_2)$$

Equating coefficients of like functions, we obtain the following set of linear equations: $2 = K_1 + K_2$, $3 = 2K_1 + K_2$. From these two equations, we find that $K_1 = 1$ and $K_2 = 1$. The result of the partial fraction expansion is thus

$$\frac{2s + 3}{s^2 + 3s + 2} = \frac{1}{s + 1} + \frac{1}{s + 2}$$

The expansion may be checked by combining the two terms of the partial fraction expansion.

Example 4

For this example, consider a quotient of polynomials with repeated denominator roots:

$$\frac{s + 2}{(s + 1)^2} = \frac{K_{11}}{s + 1} + \frac{K_{12}}{(s + 1)^2}$$

This form is required by rule (2). Multiplying the equation by $(s + 1)^2$ gives

$$s + 2 = (s + 1)K_{11} + K_{12}$$

The resulting set of linear algebraic equations is $K_{11} = 1$, $K_{11} + K_{12} = 2$, or $K_{12} = 1$. The resulting partial fraction expansion is

$$\frac{s + 2}{(s + 1)^2} = \frac{1}{s + 1} + \frac{1}{(s + 1)^2}$$

Again, this expansion can be checked, in this case by multiplying the first term in the expansion by $(s + 1)/(s + 1)$.

Example 5

This example will illustrate the expansion of a quotient of polynomials where the denominator roots are a complex conjugate pair. Consider the quotient

$$\frac{1}{s^2 + 2s + 5} = \frac{K_1}{(s + 1 + j2)} + \frac{K_1^*}{(s + 1 - j2)}$$

If each term is multiplied by a factor to put all terms over a common denominator, the following equation results:

$$1 = K_1(s + 1 - j2) + K_1^*(s + 1 + j2)$$

Equating the coefficients of like terms,

$$(1 - j2)K_1 + (1 + j2)K_1^* = 1 \quad \text{and} \quad K_1 + K_1^* = 0$$

These two equations may be solved simultaneously for K_1 and K_1^*. This gives

$$K_1 = \frac{1}{-j4} = j\tfrac{1}{4} \quad \text{and} \quad K_1^* = \frac{1}{+j4} = -j\tfrac{1}{4}$$

In this development, we did not make use of the conjugate relationship between K_1 and K_1^*. If K_1 is found by any method, it is not necessary to solve for K_1^*. If the values for K_1 and K_1^* are substituted into the first equation of this example, there results

$$\frac{1}{s^2 + 2s + 5} = \frac{j\tfrac{1}{4}}{(s + 1 + j2)} + \frac{-j\tfrac{1}{4}}{(s + 1 - j2)}$$

To use some transform tables, such terms should be revised by completing the square. In this example,

$$(s^2 + 2s + 5) = (s^2 + 2s + 1) + 4 = (s + 1)^2 + 2^2$$

so that

$$\frac{1}{s^2 + 2s + 5} = \frac{1}{(s + 1)^2 + 2^2}$$

In the general form $[(s + a)^2 + b^2]$, a is the real part of the root, and b is the imaginary part.

The three rules just given for partial fraction expansion are restricted to the cases in which the order of the numerator polynomial is less than the denominator polynomial. If this condition is not fulfilled, it is necessary to first divide the denominator into the numerator to obtain an expansion of the form

$$\frac{P(s)}{Q(s)} = B_0 + B_1 s + B_2 s^2 + \ldots + B_{n-d}s^{n-d} + \frac{P_1(s)}{Q(s)} \qquad (7\text{-}45)$$

where n is the order of the numerator, d is the order of the denominator, and $P_1(s)/Q(s)$ is a new quotient of polynomials with the order of the numerator less than that of the denominator such that the usual rules apply.

As an example, let

$$\frac{P(s)}{Q(s)} = \frac{s^2 + 2s + 2}{s + 1}$$

By direct division,

$$\begin{array}{r|l}
s + 1) & s^2 + 2s + 2 \quad (s + 1 \\
 & \underline{s^2 + s \phantom{{}+2}} \\
 & \phantom{s^2 +{}} s + 2 \\
 & \phantom{s^2 +{}} \underline{s + 1} \\
 & \phantom{s^2 + s + {}} 1
\end{array}$$

or

$$\frac{s^2 + 2s + 2}{s + 1} = 1 + s + \frac{1}{s + 1}$$

so that in Eq. 7-45, $B_0 = 1$, $B_1 = 1$, and $P_1(s)/Q(s) = 1/(s + 1)$.

7-6. Heaviside's expansion theorem

Let us return to the problem of Example 3 which was written

$$\frac{2s + 3}{(s + 1)(s + 2)} = \frac{K_1}{(s + 1)} + \frac{K_2}{(s + 2)} \tag{7-46}$$

As the first step, multiply the equation by $(s + 1)$ as

$$\frac{(2s + 3)(s + 1)}{(s + 1)(s + 2)} = K_1 \frac{s + 1}{s + 1} + K_2 \frac{s + 1}{s + 2} \tag{7-47}$$

or, canceling common factors,

$$\frac{2s + 3}{s + 2} = K_1 + K_2 \frac{s + 1}{s + 2} \tag{7-48}$$

In this equation, the coefficient K_1 is not multiplied by any function of s. Now s is merely an algebraic factor that can have any value. If $s = -1$, the coefficient of K_2 reduces to zero and we can solve for K_1 as

$$K_1 = \left.\frac{2s + 3}{s + 2}\right|_{s=-1} = \frac{-2 + 3}{-1 + 2} = 1 \tag{7-49}$$

which is the same result as found previously. To evaluate K_2 and to follow the same pattern, multiply Eq. 7-46 by $(s + 2)$ to obtain

$$\frac{2s + 3}{s + 1} = K_1 \frac{s + 2}{s + 1} + K_2 \tag{7-50}$$

To evaluate K_2, we set $s = -2$ in order to reduce the coefficient of K_1

to zero. Then

$$K_2 = \left.\frac{2s + 3}{s + 1}\right|_{s=-2} = \frac{-4 + 3}{-2 + 1} = 1 \tag{7-51}$$

This method, which the example has shown will eliminate much of the algebraic manipulation of evaluating the coefficients, is known as *Heaviside's partial expansion* method. The method will work as in this example when $Q(s)$ has no repeated roots. In general, if

$$\frac{P(s)}{Q(s)} = \frac{K_1}{s + s_1} + \frac{K_2}{s + s_2} + \frac{K_3}{s + s_3} + \cdots + \frac{K_n}{s + s_n} \tag{7-52}$$

then any of the coefficients $K_1, K_2, K_3, \ldots, K_n$ can be evaluated by multiplying by the denominator of that coefficient and setting s to the value of the root of the denominator. In other words, to find the coefficient K_j,

$$K_j = \left[(s + s_j)\frac{P(s)}{Q(s)}\right]_{s=-s_j} \tag{7-53}$$

In the form given in this equation, Heaviside's expansion theorem applied only to functions with nonrepeated denominator roots. To start our discussion of the case of repeated roots, consider the example,

$$\frac{P(s)}{Q(s)} = \frac{s + 2}{(s + 1)^2} = \frac{K_{11}}{s + 1} + \frac{K_{12}}{(s + 1)^2} \tag{7-54}$$

Multiplying by $(s + 1)^2$ gives

$$s + 2 = (s + 1)K_{11} + K_{12} \tag{7-55}$$

and when $s = -1$, K_{12} is readily evaluated as $K_{12} = 1$. If we attempt to follow the same pattern to evaluate K_{11}, trouble develops. That is,

$$\frac{s + 2}{s + 1} = K_{11} + \frac{K_{12}}{s + 1} \tag{7-56}$$

If, in this equation, $s = -1$, one term becomes infinite and K_{11} cannot be evaluated. However, the problem can be resolved if we return to Eq. 7-55 and differentiate with respect to s. (This is a reasonable thing to do: we have used differentiation before to remove trouble with indeterminant forms.) Differentiating with respect to s,

$$1 + 0 = K_{11} + 0 \quad \text{or} \quad K_{11} = 1$$

The constants are now evaluated and the partial fraction expansion is

$$\frac{s + 2}{(s + 1)^2} = \frac{1}{s + 1} + \frac{1}{(s + 1)^2} \tag{7-57}$$

To consider a general case of r-repeated roots, let

$$\frac{P(s)}{Q(s)} = \frac{R(s)}{(s+s_j)^r} = \frac{K_{j1}}{s+s_j} + \frac{K_{j2}}{(s+s_j)^2} + \cdots \frac{K_{jn}}{(s+s_j)^n} + \cdots + \cdots + \frac{K_{jr}}{(s+s_j)^r} \qquad (7\text{-}58)$$

where n is any term in the partial fraction expansion and $R(s)$ is defined as

$$R(s) = \frac{P(s)}{Q(s)}(s+s_j)^r \qquad (7\text{-}59)$$

Multiplying Eq. 7-58 by $(s+s_j)^r$ gives

$$R(s) = K_{j1}(s+s_j)^{r-1} + K_{j2}(s+s_j)^{r-2} + \ldots + K_{jr} \qquad (7\text{-}60)$$

From this equation, we can visualize the method to be used to evaluate each coefficient. If we let $s = -s_j$, all terms in the equation disappear except K_{jr}, which can be evaluated. Next, differentiate the equation once with respect to s. The term K_{jr} will vanish, but $K_{j,r-1}$ will remain without a multiplying function of s. Again, $K_{j,r-1}$ can be evaluated by letting $s = -s_j$. To find the general term K_{jn}, differentiate Eq. 7-60 $(r - n)$ times and let $s = -s_j$; then

$$K_{jn} = \frac{1}{(r-n)!}\frac{d^{r-n}R(s)}{ds^{r-n}}\bigg|_{s=-s_j} \qquad (7\text{-}61)$$

or

$$K_{jn} = \frac{1}{(r-n)!}\frac{d^{r-n}}{ds^{r-n}}\left\{\frac{P(s)}{Q(s)}(s+s_j)^r\right\}\bigg|_{s=-s_j} \qquad (7\text{-}62)$$

The actual use of this idea is easier than might appear from the complexity of this general equation. For example, consider

$$\frac{2s^2+3s+2}{(s+1)^3} = \frac{K_{11}}{(s+1)} + \frac{K_{12}}{(s+1)^2} + \frac{K_{13}}{(s+1)^3} \qquad (7\text{-}63)$$

Multiplying the equation by $(s+1)^3$, we have

$$2s^2 + 3s + 2 = K_{11}(s+1)^2 + K_{12}(s+1) + K_{13}$$

From this equation,

$$K_{13} = 2s^2 + 3s + 2\bigg|_{s=-1} = 2 - 3 + 2 = 1$$

Next, we differentiate with respect to s to obtain

$$4s + 3 = 2K_{11}(s+1) + K_{12}$$

so that

$$K_{12} = 4s + 3\bigg|_{s=-1} = -1$$

Again, we differentiate the last equation to give

$$4 = 2K_{11} \quad \text{or} \quad K_{11} = 2$$

The partial fraction expansion is

$$\frac{2s^2 + 3s + 2}{(s + 1)^3} = \frac{2}{s + 1} + \frac{-1}{(s + 1)^2} + \frac{1}{(s + 1)^3} \tag{7-64}$$

If $Q(s)$ contains both simple and repeated roots, a combination of both rules may be used. As an example, let

$$\frac{P(s)}{Q(s)} = \frac{s + 2}{(s + 1)^2(s + 3)} \tag{7-65}$$

The form of the partial fraction expansion is

$$\frac{s + 2}{(s + 1)^2(s + 3)} = \frac{K_{11}}{s + 1} + \frac{K_{12}}{(s + 1)^2} + \frac{K_2}{s + 3} \tag{7-66}$$

In this expansion, K_2 may be evaluated by Eq. 7-53 and K_{11} and K_{12} may be found from Eq. 7-62; then

$$K_2 = \left.\frac{s + 2}{(s + 1)^2}\right|_{s=-3} = -\frac{1}{4}$$

Multiplying Eq. 7-66 by $(s + 1)^2$, we have

$$\frac{s + 2}{s + 3} = K_{11}(s + 1) + K_{12} + \frac{(s + 1)^2}{s + 3} K_2 \tag{7-67}$$

Constant K_{12} is evaluated directly by letting $s = -1$; thus

$$K_{12} = \left.\frac{s + 2}{s + 3}\right|_{s=-1} = \frac{1}{2}$$

and K_{11} will be found by differentiating Eq. 7-67 before letting $s = -1$:

$$\frac{(s + 3) \cdot 1 - (s + 2) \cdot 1}{(s + 3)^2} = K_{11} + K_2 \frac{d}{ds}\left[\frac{(s + 1)^2}{s + 3}\right]$$

The coefficient of K_2 vanishes when $s = -1$ because an $(s + 1)$ term remains common to all terms in the numerator. In the example

$$\frac{d}{ds}\left[\frac{(s + 1)^2}{s + 3}\right] = \frac{(s + 3)2(s + 1) - (s + 1)^2 \cdot 1}{(s + 3)^2}$$

and this term vanishes when $s = -1$, because each term in the differentiation contains $(s + 1)$. This is always the case, since the order of the multiplying factor $(s + s_j)^r$ is higher than the number of times differentiation is required.

By using these methods, all the coefficients of the partial fraction expansion can be found and the transform equation can be written

$$F(s) = \sum_{j=1}^{n} \frac{K_j}{s + s_j} \tag{7-68}$$

for simple roots (nonrepeated) and as

$$F(s) = \sum_{k=1}^{r} \frac{K_{jk}}{(s + s_j)^k} \tag{7-69}$$

for a single root, $-s_j$, repeated r times. The corresponding $f(t)$ may now be found, for the general case, by taking the inverse Laplace transformation of $F(s)$ as

$$f(t) = \mathcal{L}^{-1}\left[\frac{P(s)}{Q(s)}\right] = \sum_{j=1}^{n} (s + s_j)\frac{P(s)}{Q(s)} e^{st}\bigg|_{s=-s_j} \tag{7-70}$$

as the time-domain solution for simple roots. Likewise, for repeated roots,

$$f(t) = e^{-s_j t} \sum_{n=1}^{r} \frac{1}{(r-n)!} \frac{d^{r-n}R(-s_j)}{ds^{r-n}} \frac{t^{n-1}}{(n-1)!} \tag{7-71}$$

where s_j, in this equation, is *the root* that is repeated r times. By using both equations for the case of both simple and repeated roots, a general solution is obtained in the form originally given as *Heaviside's expansion theorem.*

The method of the Heaviside partial fraction expansion may be used to give a simplified procedure for finding the inverse transform of the terms for a conjugate complex pair of roots. Suppose that these roots have a real part α and an imaginary part, ω. The first coefficient is evaluated by the procedure,

$$K_1 = \frac{P(s)}{Q(s)}(s - \alpha + j\omega)\bigg|_{s=\alpha-j\omega} = Re^{-j\theta} \tag{7-72}$$

and the second as

$$K_1{}^* = \frac{P(s)}{Q(s)}(s - \alpha - j\omega)\bigg|_{s=\alpha+j\omega} = Re^{+j\theta} \tag{7-73}$$

The inverse transformation of these two terms gives

$$f_1(t) = Re^{j\theta}e^{(\alpha+j\omega)t} + Re^{-j\theta}e^{(\alpha-j\omega)t} \tag{7-74}$$

This equation may be rearranged to the form

$$f_1(t) = 2Re^{\alpha t}\left[\frac{e^{j(\omega t+\theta)} + e^{-j(\omega t+\theta)}}{2}\right]$$
$$= 2Re^{\alpha t}\cos(\omega t + \theta) \tag{7-75}$$

The factors R and θ in this equation are easily found in Eq. 6-72 as the magnitude and phase angle of the coefficient K_1.

7-7. Examples of total solutions by the Laplace transformation

Example 6

As an example of the total solution, now that the methods of partial fraction expansion have been reviewed, consider the differential equation

$$\frac{d^2i}{dt^2} + 4\frac{di}{dt} + 5i = 5u(t) \tag{7-76}$$

The Laplace transformation of this differential equation is

$$\left[s^2I(s) - si(0+) - \frac{di}{dt}(0+)\right] + 4[sI(s) - i(0+)] + 5I(s) = \frac{5}{s}$$

Notice that the required initial conditions are automatically specified in this equation. We must know, from the physical system, $i(0+)$ and $di/dt\,(0+)$. Suppose the following values are found:

$$i(0+) = 1 \quad \text{and} \quad \frac{di}{dt}(0+) = 2$$

Inserting these initial conditions simplifies the transform equation to

$$I(s)(s^2 + 4s + 5) = \frac{5}{s} + s + 6$$

or

$$I(s) = \frac{s^2 + 6s + 5}{s(s^2 + 4s + 5)}$$

This equation may be expanded by partial fractions as

$$I(s) = \frac{s^2 + 6s + 5}{s(s + 2 + j1)(s + 2 - j1)} = \frac{K_1}{s} + \frac{K_2}{s + 2 + j1} + \frac{K_2^*}{s + 2 - j1}$$

To evaluate K_1, multiply the equation by s and let $s = 0$. Then

$$K_1 = \left.\frac{s^2 + 6s + 5}{s^2 + 4s + 5}\right|_{s=0} = 1$$

To evaluate K_2, multiply the equation by $(s + 2 + j1)$ and let

$s = -2 - j1$ as

$$K_2 = \frac{s^2 + 6s + 5}{s(s + 2 - j1)}\bigg|_{s=-2-j1} = \frac{-4 - j2}{(-2 - j1)(-j2)} = \frac{2}{-j2} = j = e^{j90^\circ}$$

The complete partial fraction expansion becomes

$$I(s) = \frac{1}{s} + \frac{j}{s + 2 + j1} + \frac{-j}{s + 2 - j1}$$

To obtain $i(t)$ from this transform equation, we take the inverse Laplace transformation of the first term and use Eq. 7-75 with $R = 1$ and $\theta = 90^\circ$ for the second and third terms to give the solution

$$i(t) = 1 + 2e^{-2t} \sin t \tag{7-77}$$

Example 7

For this example, consider a series RLC circuit with the capacitor initially charged to voltage V_0 as indicated in Fig. 7-5. The differential equation for the current $i(t)$ is

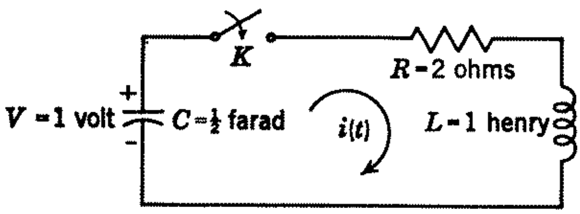


Fig. 7-5. RLC series circuit.

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = 0 \tag{7-78}$$

and the corresponding transform equation is

$$L[sI(s) - i(0+)] + RI(s) + \frac{1}{Cs}[I(s) + q(0+)] = 0$$

The parameters have been specified as $C = \frac{1}{2}$ farad, $R = 2$ ohms, and $L = 1$ henry. The initial current $i(0+) = 0$ because of the inductor, and if C is initially charged to voltage V_0 (with the given polarity),

$$\frac{q(0+)}{Cs} = -\frac{V_0}{s}$$

or $-1/s$ if $V_0 = 1$ volt. The transform equation for $I(s)$ then becomes

$$I(s) = \frac{1}{s^2 + 2s + 2}$$

or, completing the square,

$$I(s) = \frac{1}{(s + 1)^2 + 1}$$

Using transform pair 15 of page 146,

$$i(t) = \mathcal{L}^{-1}I(s) = e^{-t} \sin t \tag{7-79}$$

Example 8

In the network shown in Fig. 7-6, the switch is closed at $t = 0$. With the network parameter values shown, the Kirchhoff voltage equations are

$$\frac{di_1}{dt} + 20i_1 - 10i_2 = 100u(t),$$

$$\frac{di_2}{dt} + 20i_2 - 10i_1 = 0$$

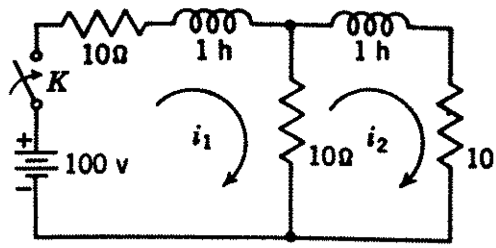


Fig. 7-6. Network of Example 8.

If the network is unenergized before the switch is closed, both i_1 and i_2 are initially zero, and the transform equations may be written

$$(s + 20)I_1(s) - 10I_2(s) = \frac{100}{s}, \qquad -10I_1(s) + (s + 20)I_2(s) = 0$$

Suppose that we are required to find the current i_2 as a function of time. The transform current $I_2(s)$ may be found from the last two algebraic equations by determinants as

$$I_2(s) = \frac{\begin{vmatrix} s + 20 & 100/s \\ -10 & 0 \end{vmatrix}}{\begin{vmatrix} s + 20 & -10 \\ -10 & s + 20 \end{vmatrix}} = \frac{1000}{s(s^2 + 40s + 300)}$$

The partial fraction expansion of this equation is

$$\frac{1000}{s(s + 10)(s + 30)} = \frac{3.33}{s} - \frac{5}{s + 10} + \frac{1.67}{s + 30}$$

The inverse Laplace transformation gives $i_2(t)$ as

$$i_2 = 3.33 - 5e^{-10t} + 1.67e^{-30t}$$

7-8. The initial and final value theorems

The initial value theorem and final value theorem find frequent use in network analysis. To derive the initial value theorem, we allow s to approach infinity in the equation for the transform of a derivative as

$$\lim_{s \to \infty} \int_0^\infty f'(t)e^{-st}\,dt = \lim_{s \to \infty} [sF(s) - f(0+)] \tag{7-80}$$

In writing this equation, we assume that $f(t)$ and its first derivative are transformable and that the limit of $sF(s)$ as s approaches infinity exists. Since the integral has zero value for $s \to \infty$, and $f(0+)$ is

TABLE OF TRANSFORMS

$f(t)$	$F(s)$
1. $u(t)$ or 1	$\dfrac{1}{s}$
2. t	$\dfrac{1}{s^2}$
3. $\dfrac{t^{n-1}}{(n-1)!}$ $\quad n =$ integer	$\dfrac{1}{s^n}$
4. e^{at}	$\dfrac{1}{s-a}$
5. te^{at}	$\dfrac{1}{(s-a)^2}$
6. $\dfrac{1}{(n-1)!}t^{n-1}e^{at}$	$\dfrac{1}{(s-a)^n}$
7. $\dfrac{1}{a-b}(e^{at}-e^{bt})$	$\dfrac{1}{(s-a)(s-b)}$
8. $\dfrac{e^{-at}}{(b-a)(c-a)} + \dfrac{e^{-bt}}{(a-b)(c-b)} + \dfrac{e^{-ct}}{(a-c)(b-c)}$	$\dfrac{1}{(s+a)(s+b)(s+c)}$
9. $1 - e^{+at}$	$\dfrac{-a}{s(s-a)}$
10. $\dfrac{1}{\omega}\sin \omega t$	$\dfrac{1}{s^2+\omega^2}$
11. $\cos \omega t$	$\dfrac{s}{s^2+\omega^2}$
12. $1 - \cos \omega t$	$\dfrac{\omega^2}{s(s^2+\omega^2)}$
13. $\sin(\omega t + \theta)$	$\dfrac{s \sin\theta + \omega\cos\theta}{s^2+\omega^2}$
14. $\cos(\omega t + \theta)$	$\dfrac{s\cos\theta - \omega\sin\theta}{s^2+\omega^2}$
15. $e^{-\alpha t}\sin \omega t$	$\dfrac{\omega}{(s+\alpha)^2+\omega^2}$
16. $e^{-\alpha t}\cos \omega t$	$\dfrac{s+\alpha}{(s+\alpha)^2+\omega^2}$
17. $\sinh \alpha t$	$\dfrac{\alpha}{s^2-\alpha^2}$
18. $\cosh \alpha t$	$\dfrac{s}{s^2-\alpha^2}$

independent of s,

$$\lim_{s \to \infty} [sF(s) - f(0+)] = 0 \tag{7-81}$$

But $f(0+) = \lim_{t \to 0+} f(t)$, from which we conclude that

$$\lim_{s \to \infty} sF(s) = \lim_{t \to 0+} f(t) \tag{7-82}$$

subject to the restrictions mentioned previously. Equation 7-82 shows that the value of $f(t)$ at $t = 0+$ is equal to the limit of the product $sF(s)$ as s approaches infinity.

In deriving the final value theorem we start from the same equation as the initial value theorem, but let s approach zero. Assuming that $f(t)$ and its first derivative are transformable, we write

$$\lim_{s \to 0} \int_0^\infty f'(t)e^{-st}\,dt = \lim_{s \to 0} [sF(s) - f(0+)] \tag{7-83}$$

Since s and t are independent (and $e^{-st} \to 1$, as $s \to 0$) the integral becomes

$$\int_0^\infty f'(t)\,dt = \lim_{t \to \infty} f(t) - f(0+) \tag{7-84}$$

This expression may be equated to Eq. 7-83 to give

$$\lim_{s \to 0} [sF(s)] - \cancel{f(0+)} = \lim_{t \to \infty} [f(t)] - \cancel{f(0+)} \tag{7-85}$$

from which we conclude that

$$\lim_{s \to 0} [sF(s)] = \lim_{t \to \infty} [f(t)] \tag{7-86}$$

which is known as the *final value theorem*. This result holds provided all roots of the denominator of $sF(s)$ have negative real parts. Because of this restriction, the final value theorem does not apply in the case of sinusoidal excitation, because the denominator roots of the transform of the sinusoid are purely imaginary.

FURTHER READING

An interesting historical summary titled "The Work of Oliver Heaviside" by Behrend is contained as an appendix in Berg's *Heaviside's Operational Calculus* (McGraw-Hill Book Co., Inc., New York, 1929), pp. 173–208. Heaviside's original writings have recently been reprinted as *Electromagnetic Theory* (Dover Publications, New York, 1950) and contain an extensive presentation of his method. For a more complete treatment of the Laplace transformation than is offered here, the student is referred to the following: Wylie, *Advanced Engineering Mathematics* (McGraw-Hill Book Co., Inc., New York, 1951), Chap. 6; Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), beginning on p. 93; Churchill, *Modern Operational Mathematics in Engineering* (McGraw-Hill Book Co., Inc., New York, 1951) and Thomson, *Laplace Transformation* (Prentice-Hall, Inc., New York, 1950). Extensive transform tables may be found in Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 334–356 and in Nixon, *Principles of Automatic Controls* (Prentice-Hall, Inc., New York, 1953), pp. 371–399.

PROBLEMS

7-1. Verify the following transform pair by substituting the value of $f(t)$ into Eq. 7-1 and integrating.

$$\mathcal{L}[\cos \omega t] = \frac{s}{s^2 + \omega^2}$$

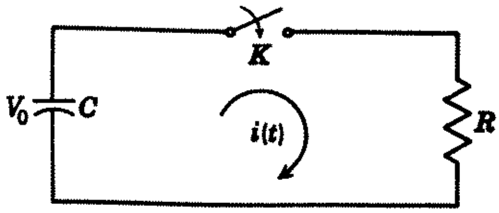
In each of the problems that follow, repeat the procedure of Prob. 7-1 for the various transform pairs of the following table.

	$f(t)$	$F(s)$
7-2.	t^2	$\dfrac{2}{s^3}$
7-3.	$\sinh \alpha t$	$\dfrac{\alpha}{s^2 - \alpha^2}$
7-4.	$\cosh \alpha t$	$\dfrac{s}{s^2 - \alpha^2}$
7-5.	$e^{-\alpha t} \sin \omega t$	$\dfrac{\omega}{(s + \alpha)^2 + \omega^2}$
7-6.	$e^{-\alpha t} \cos \omega t$	$\dfrac{s + \alpha}{(s + \alpha)^2 + \omega^2}$
7-7.	$\sin (\omega t + \theta)$	$\dfrac{s \sin \theta + \omega \cos \theta}{s^2 + \omega^2}$
7-8.	$\cos (\omega t + \theta)$	$\dfrac{s \cos \theta - \omega \sin \theta}{s^2 + \omega^2}$

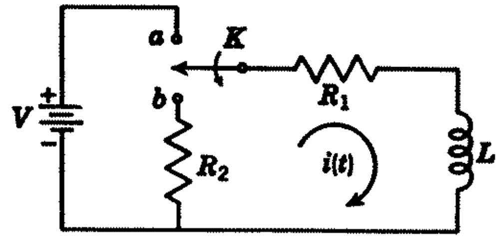
7-9. Letting $u = f(t)$ and $dv = e^{-st}\,dt$, integrate Eq. 7-1 by parts to prove Eq. 7-13.

7-10. Rework Example 1, assuming that the capacitor is originally charged to the voltage $V/2$ with the upper plate positive at $t = 0$. *Answer.* $(V/2R)e^{-t/RC}$.

7-11. In the network shown in the figure, C is charged to V_0, and the switch K is closed at $t = 0$. Solve for the current $i(t)$ using the Laplace transformation. *Answer.* $(V/R)e^{-t/RC}$.



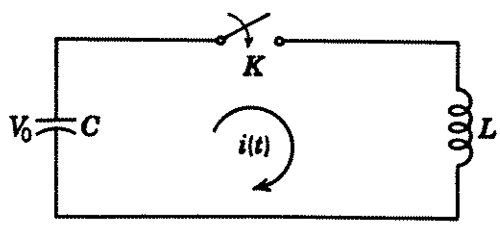
Prob. 7-11.



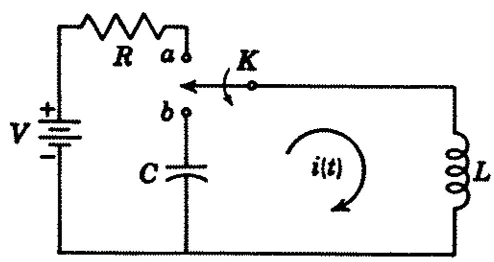
Prob. 7-12.

7-12. In the network shown in the figure, the switch K is moved from position a to position b at $t = 0$, a steady state having previously been established at position a. Solve for the current $i(t)$, using the Laplace transformation. *Answer.* $(V/R_1)e^{-(R_1+R_2)t/L}$.

7-13. In the network shown, C is initially charged to V_0. The switch K is closed at $t = 0$. Solve for the current $i(t)$, using the Laplace transformation. *Answer.* $(V/\sqrt{L/C}) \sin (t/\sqrt{LC})$.



Prob. 7-13.



Prob. 7-14.

7-14. In the network shown, the switch K is moved from position a to position b at $t = 0$ (a steady state existing in position a before $t = 0$). Solve for the current $i(t)$, using the Laplace transformation. *Answer.* $(V/R) \cos (t/\sqrt{LC})$.

7-15. Check the following partial fraction expansions by expanding the given quotient of polynomials in partial fractions. Two of the set are in error.

(a) $\dfrac{2s}{s^2 - 1} = \dfrac{1}{s + 1} + \dfrac{1}{s - 1}$

(b) $\dfrac{7s + 2}{s^3 + 3s^2 + 2s} = \dfrac{1}{s} + \dfrac{2}{s + 2} + \dfrac{-3}{s + 1}$

(c) $\dfrac{5s + 13}{s^2 + 5s + 6} = \dfrac{2}{s + 3} + \dfrac{3}{s + 2}$

(d) $\dfrac{s^2}{s - 1} = \dfrac{1}{s - 1} + s + 1$

(e) $\dfrac{2(s+1)}{s^2+1} = \dfrac{1+j1}{s+j1} + \dfrac{1-j1}{s-j1}$

(f) $\dfrac{s^2+4s+1}{s(s+1)^2} = \dfrac{1}{s+1} + \dfrac{1}{(s+1)^2} + \dfrac{2}{s}$

(g) $\dfrac{3s^3-s^2-3s+2}{s^2(s-1)^2} = \dfrac{1}{s} + \dfrac{2}{s^2} + \dfrac{2}{s-1} + \dfrac{1}{(s-1)^2}$

(h) $\dfrac{s^3-5s^2+9s+9}{s^2(s^2+9)} = \dfrac{1}{s} + \dfrac{1}{s^2} + \dfrac{-j}{s+j3} + \dfrac{+j}{s-j3}$

7-16. Expand the following functions by partial fractions and find the corresponding inverse Laplace transformation, $f(t) = \mathcal{L}^{-1}F(s)$.

(a) $F(s) = \dfrac{3s}{(s^2+1)(s^2+4)}$. *Answer.* $f(t) = \cos t - \cos 2t$.

(b) $F(s) = \dfrac{s+1}{s^2+2s}$. *Answer.* $f(t) = \frac{1}{2}(1+e^{-2t})$.

(c) $F(s) = \dfrac{1}{s(s^2-2s+5)}$. *Answer.* $f(t) = \frac{1}{5}[1 + \frac{1}{2}e^t(-2\cos 2t$ $+ \sin 2t)]$.

(d) $F(s) = \dfrac{1}{(s+1)(s+2)^2}$. *Answer.* $f(t) = e^{-t} - e^{-2t}(1+t)$.

(e) $F(s) = \dfrac{1}{s^3(s^2-1)}$. *Answer.* $f(t) = -1 - t^2/2 + \cosh t$.

(f) $F(s) = \dfrac{s^2+2s+1}{(s+2)(s^2+4)}$. *Answer.* $f(t) = \frac{1}{8}e^{-2t} + \frac{7}{8}\cos 2t + \frac{1}{8}\sin 2t$.

(g) $F(s) = \dfrac{s^2}{(s^2+1)^2}$. *Answer.* $f(t) = \frac{1}{2}t\cos t + \frac{1}{2}\sin t$.

Solve the following differential equations by the Laplace transformation subject to the given initial conditions (where specified).

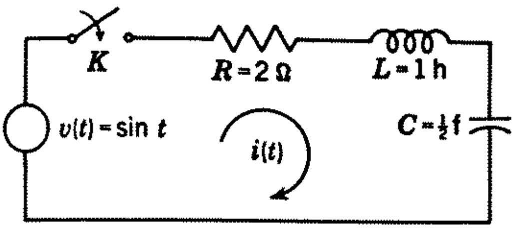
7-17. $\dfrac{d^2i}{dt^2} - i = 25 + e^{2t}$. *Answer.* $i = K_1e^t + K_2e^{-t} - 25 + \frac{1}{3}e^{2t}$.

7-18. $\dfrac{d^2v}{dt^2} + 4v = \sin t - \cos 2t$. *Answer.* $v = K_1 \sin 2t + K_2$ $\cos 2t - \frac{1}{4}t \sin 2t + \frac{1}{3}\sin t$.

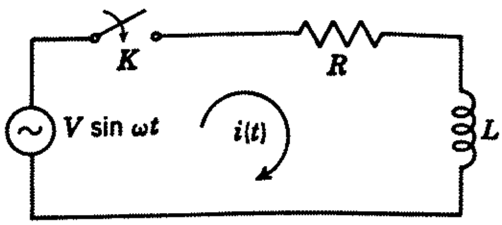
7-19. $\dfrac{d^2q}{dt^2} + \dfrac{dq}{dt} = t^2 + 2t$, $q(0+) = 4$, $\dfrac{dq}{dt}(0+) = -2$. *Answer.* $q = \frac{1}{3}t^3 + 2e^{-t} + 2$.

7-20. Solve Prob. 6-13(d), using the Laplace transformation. Note that in the Laplace transformation method, special conditions of similarity in the form of the driving force $v(t)$ and the roots of the characteristic equation give no concern—the solution of such a problem is as routine as any other problem.

7-21. In the series RLC circuit shown below, the applied voltage is $v(t) = \sin t$. For the parameter values specified, find $i(t)$ if the switch K is closed at $t = 0$. *Answer.* $i(t) = \frac{1}{5}(\cos t + 2 \sin t - e^{-t} \cos t - 3e^{-t} \sin t)$.



Prob. 7-21.



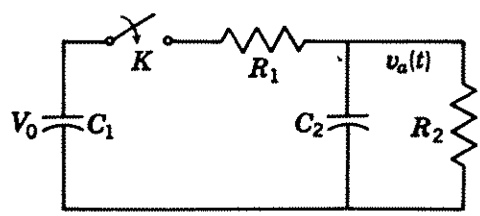
Prob. 7-22.

7-22. At $t = 0$, a switch is closed, connecting a voltage source $v = V \sin \omega t$ to a series RL circuit. By the method of the Laplace transformation, show that the current is given by the equation

$$i = \frac{V}{Z} \sin (\omega t - \phi) + \frac{\omega L V}{Z^2} e^{-Rt/L}$$

where $$Z = \sqrt{R^2 + (\omega L)^2} \quad \text{and} \quad \phi = \tan^{-1} \frac{\omega L}{R}$$

7-23. Dr. L. A. Woodbury of the University of Utah School of Medicine has made use of an electrical analog in studies of convulsions. In the network shown in the figure, the following quantities are duals:



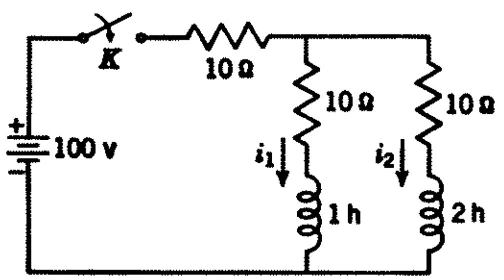
Prob. 7-23.

C_1 represents the volume of drug-containing fluid, R_1 is the "resistance" to the passage of the drug from the pool to the blood stream, C_2 represents the volume of the blood stream, and R_2 is equivalent to the body's excretion mechanism (kidney, etc.). The concentration of the drug dose is represented as V_0 and the voltage $v_a(t)$ at node a is the dual of the amount of drug in the blood stream. The analog network has the advantage that the elements may be readily changed and the effects studied (to say nothing of the saving of cats). Find the transform equation for $V_a(s)$ with the coefficient of the highest-order term normalized to unity.

7-24. This problem is a continuation of Prob. 7-23 concerning Dr. Woodbury's analog. The following constants for the network are selected: $C_1 = 1\ \mu f$, $C_2 = 8\ \mu f$, $R_1 = 9$ megohms, and $R_2 = 5$ megohms. If $V_0 = 100$ volts and the switch is closed at $t = 0$, solve for $v_A(t)$, the equivalent of the concentration of drug in the bloodstream, as a function of time. *Answer.* $v_a(t) = 13e^{-0.0215t} - 13e^{-0.1285t}$.

7-25. Find the time t_m when the concentration of drug in the blood stream for Prob. 7-24, is a maximum. (This information is desired so that a second dose may be given at that time to build up the concentration to the point where a convulsion is induced.) *Answer.* 16.7 sec.

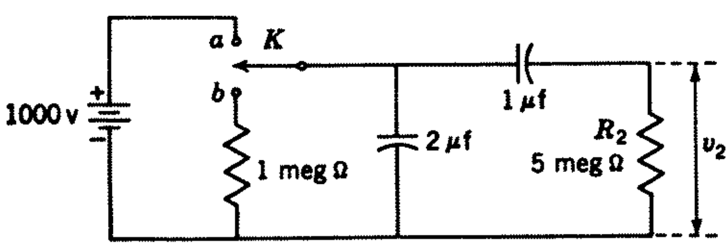
7-26. If a second dose (the voltage equivalent having a magnitude of 100 volts) is injected at $t = t_m$ as found in Prob. 7-25, what will be v_A as a function of time, and what will be the maximum v_A obtained? (*Note:* In giving the second dose we will assume that the total voltage is then 100 volts plus the voltage on the plates at the time the addition is made.) *Answer.* 14 volts in 25 sec.



Prob. 7-27.

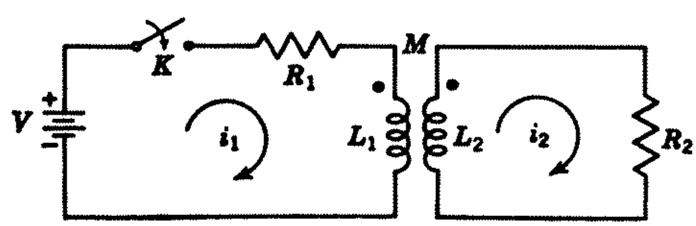
7-27. In the network shown, the switch K is closed at $t = 0$ with the network previously unenergized. For the element values shown on the diagram: (a) find $i_1(t)$, (b) find $i_2(t)$. *Answer* (a). $i_1 = 3.33 + 1.21e^{-6.35t} - 4.54e^{-23.6t}$.

7-28. With switch K in position a, the network shown in the figure attains equilibrium. At time $t = 0$, the switch is moved to position b. Find the voltage across R_2 as a function of time.



Prob. 7-28.

7-29. Find $i_1(t)$ resulting from closing the switch at $t = 0$ with the circuit previously unenergized. The circuit constants are: $L_1 = 1$ henry, $L_2 = 4$ henrys, $M = 2$ henrys, $R_1 = R_2 = 1$ ohm, $V = 1$ volt.



Prob. 7-29.

CHAPTER 8

TOPICS IN THE TIME DOMAIN AND THE FREQUENCY DOMAIN

The time-domain response of networks to various driving forces has been considered in previous chapters. In this chapter, time-domain studies will be extended by specializing the driving forces (current sources and voltage sources) to the following cases: (1) single pulses and related waveforms, (2) time-varying functions which recur a finite number of times, and (3) recurring waveforms which cannot be described by a single equation. Examples of such waveforms are shown in Fig. 8-1. The transient response of networks subjected to these

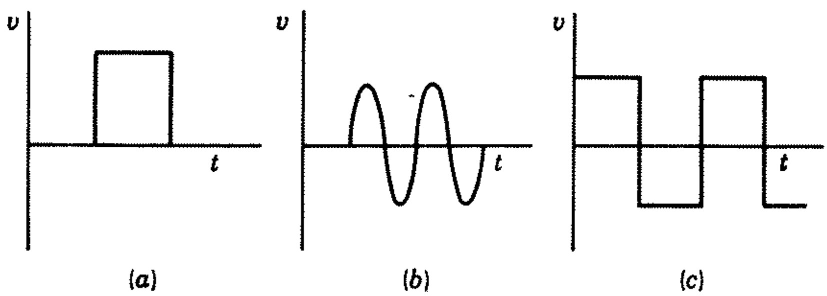


Fig. 8-1. Driving force waveforms: (a) pulse; (b) section of sine wave; (c) square wave.

driving forces will be studied, using the Laplace transformation. The time-domain studies will be followed by related frequency-domain studies, using Fourier series and the Fourier integral.

8-1. The unit step function

The unit step function is defined as

$$u(t) = \begin{cases} 1, & t > 0 \\ 0, & t < 0 \end{cases} \tag{8-1}$$

for a function which changes abruptly from zero to unit value at the time $t = 0$. This expression may be generalized by the definition

$$u(t - a) = \begin{cases} 1, & t > a \\ 0, & t < a \end{cases} \tag{8-2}$$

for a step function which changes abruptly at the time $t = +a$. In general, the step function has unit value when the quantity $(t - a)$

has a positive value, and has zero value when $(t - a)$ is negative. This definition will apply for any form of the variable. Hence the function $u(t + a)$ is one that changes from zero to unit value at $t = -a$. Similarly, the function $u(a - t)$ is one that changes from unit to zero value (with increasing time) at the time $t = a$.* These two functions

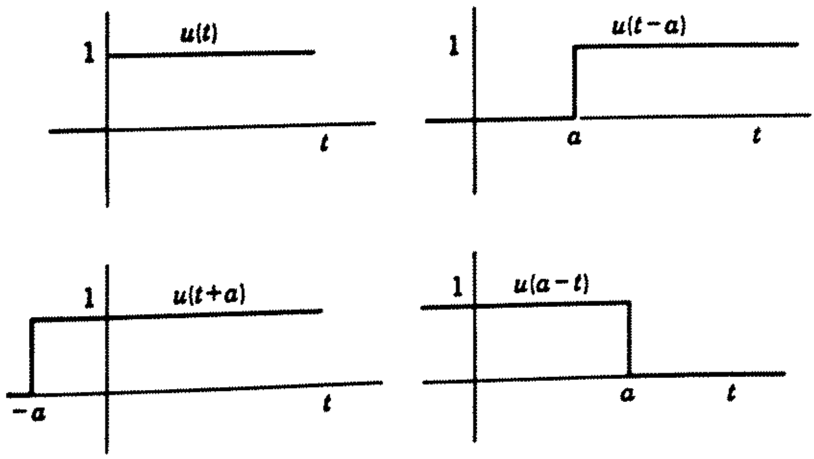


Fig. 8-2. Unit step functions:

$$u(t + a) = \begin{cases} 1, & t > -a; \\ 0, & t < -a \end{cases} \qquad u(a - t) = \begin{cases} 1, & t < a \\ 0, & t > a \end{cases}$$

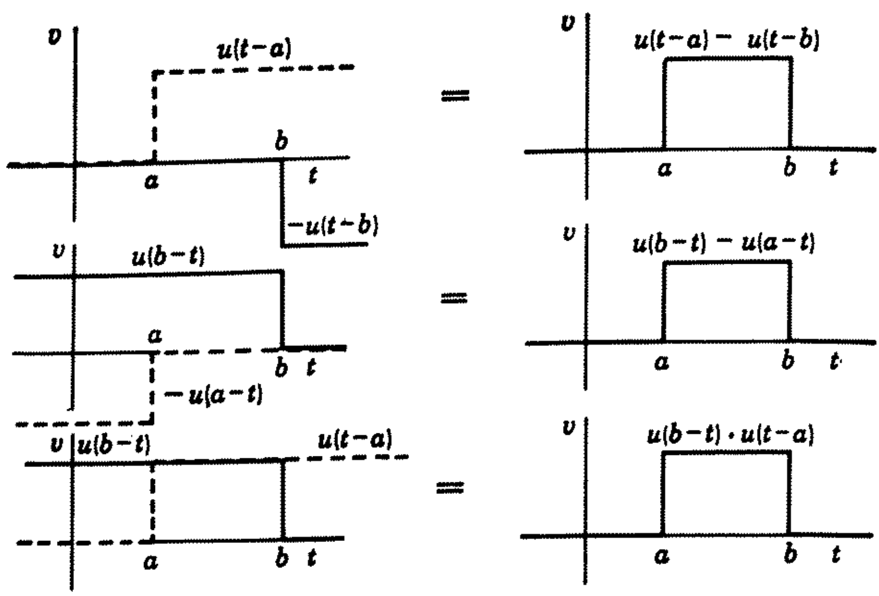


Fig. 8-3. Construction of pulse from unit step functions.

are represented in Fig. 8-2. The use of unit step functions with definitions as illustrated in these examples will make it possible to represent functions with the time axis shifted. Further, the unit step functions can be used as building blocks to represent other time-varying functions such as a pulse. The construction of a pulse from two unit step functions is illustrated in Fig. 8-3. A unit step function $u(t - a)$ and

* Note that $u(a - t) \neq -u(t - a)$.

a unit step function $u(t - b)$ are shown in the figure. By taking the difference between these two step functions,

$$v(t) = u(t - a) - u(t - b) \tag{8-3}$$

a *pulse* is formed of unit amplitude from $t = a$ to $t = b$. The same unit pulse may be formed in terms of the unit step function building blocks as

$$v(t) = u(b - t) - u(a - t) \tag{8-4}$$

or

$$v(t) = u(b - t) \cdot u(t - a) \tag{8-5}$$

These operations are illustrated in Fig. 8-3.

As another example of the use of unit step functions in constructing time-varying waveforms, consider the mathematical representation of a square wave. A pulse of width a starting at $t = 0$ is given by the equation

$$u(t) - u(t - a) \tag{8-6}$$

as illustrated in Fig. 8-4(a). Instead of subtracting $u(t - a)$ from $u(t)$, suppose that $2u(t - a)$ is subtracted. The resulting waveform will

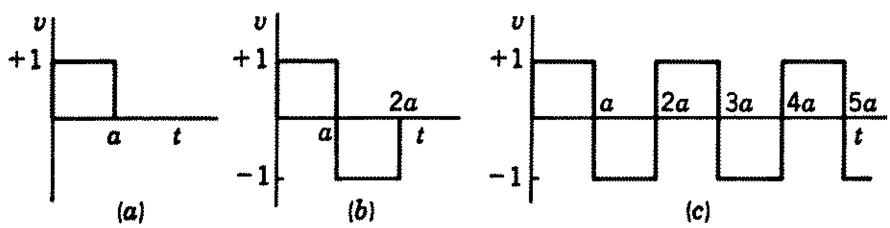


Fig. 8-4. Evolution of a square wave.

then change from the value of $+1$ to -1 at $t = a$. By next adding the unit step function $u(t - 2a)$ to the function, the waveform assumes zero value for all time greater than $t = 2a$. This construction is illustrated in Fig. 8-4(b) for the function

$$v(t) = u(t) - 2u(t - a) + u(t - 2a) \tag{8-7}$$

By following in this pattern, any square wave (square only if $a = 1$, of course, but known as a square wave for any value of a) form of time variation can be synthesized. It is clear that a square waveform which continues infinitely long in duration is given by the infinite series

$$v(t) = u(t) - 2u(t - a) + 2u(t - 2a) - 2u(t - 3a) + \dots \tag{8-8}$$

This waveform is shown in Fig. 8-4(c). This infinite series representation will be found to be more convenient than it appears when it is shown that the Laplace transformation reduces to a closed form. As a third example of the representation of time-varying waveforms, consider the problem of representing the waveform shown in Fig. 8-5. The waveform is sinusoidal from $t = 1$ to $t = 3$ and from $t = 5$ to $t = 7$.

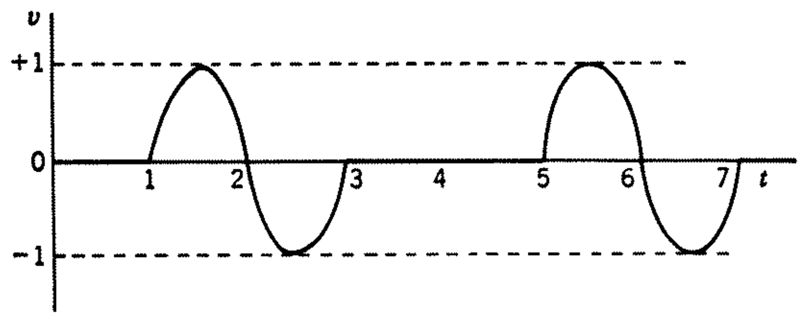


Fig. 8-5. Two sine-wave cycles.

It has zero value for all other times from $t = -\infty$ to $t = +\infty$. A sine wave with a period of T is given as

$$\sin \frac{2\pi}{T} t \tag{8-9}$$

In this particular example, $T = 2$ and the time axis is shifted by 1 unit of time for the first wave and by 5 units of time for the second. We will follow a step-by-step procedure in constructing a function to represent this waveform.

(1) The function $\sin \pi(t - 1)$ has the waveform shown in the interval $t = 1$ to $t = 3$, but the waveform also exists for all other time.
(2) Multiplying $\sin \pi(t - 1)$ by $u(t - 1)$ eliminates all waveform at times *before* $t = 1$. Subtracting from this product a similar product shifted to the time $t = 3$ cancels all times *after* $t = 3$. This product is $u(t - 3) \sin \pi(t - 3)$. Hence the first cycle of sine wave is completely represented by

$$u(t - 1) \sin \pi(t - 1) - u(t - 3) \sin \pi(t - 3) \tag{8-10}$$

(3) By the same reasoning, the second cycle of sine wave is represented as

$$u(t - 5) \sin \pi(t - 5) - u(t - 7) \sin \pi(t - 7) \tag{8-11}$$

(4) The total waveform is the sum of the two expressions. This follows because each function is defined only for its interval (1 to 3 and 5 to 7, respectively) and is zero for all other time. Hence the waveform of Fig. 8-5 may be represented by the equation

$$\begin{aligned} v(t) = u(t - 1) \sin \pi(t - 1) &- u(t - 3) \sin \pi(t - 3) \\ &+ u(t - 5) \sin \pi(t - 5) - u(t - 7) \sin \pi(t - 7) \end{aligned} \tag{8-12}$$

By following similar patterns, any time function can be represented by a time series and unit step functions.

8-2. Other unit functions: the impulse, ramp, and doublet

The waveforms which have been described mathematically in the last section will be used in this chapter to describe driving forces applied to networks consisting of one or more electric elements. The voltage-current relationships for the individual elements are

$$v_R = Ri, \quad v_L = L\frac{di}{dt}, \quad \text{and} \quad v_C = \frac{1}{C}\int i\,dt$$

or

$$i_R = \frac{1}{R}v, \quad i_C = C\frac{dv}{dt}, \quad \text{and} \quad i_L = \frac{1}{L}\int v\,dt$$

Whether the driving force is a voltage source or a current source, voltages and currents in the network are described by integrals and derivatives of waveforms. Consider a step function voltage driving force. If this waveform were applied to an inductor, the current resulting from this voltage would be the integral of the voltage. If a voltage step function were applied to a single capacitor, the current would be the derivative of this voltage. Evidently we shall be concerned with both integrals and the derivatives of driving-force functions. For the unit step function, such waveforms are illustrated by Fig. 8-6. The

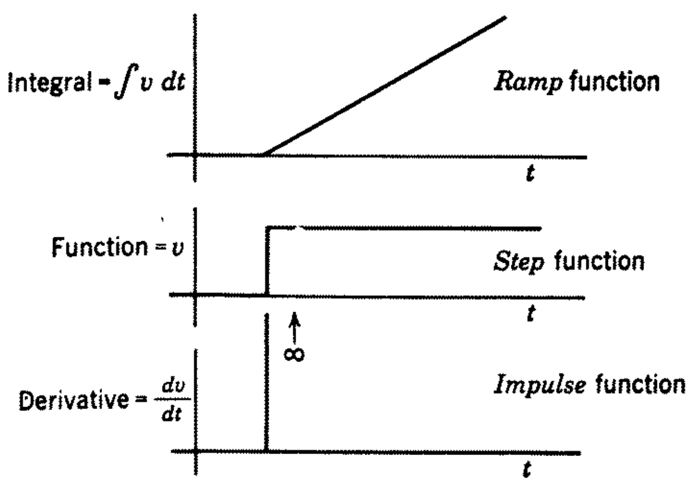


Fig. 8-6. Derivative and integral functions.

integral of the step function varies linearly with time, and is known as a *ramp function* (or a *linear ramp*). The derivative of the step function has a nonzero value only at the beginning of the step function: there the value is infinite, for all other values of time the value is zero. This rather unusual function with only one nonzero value (and that infinite) is known as an *impulse function.*

If the step function has unit magnitude, the slope of the corresponding ramp function is unity since the ramp function is the integral of the step function. A ramp function with unity slope is known as a *unit ramp*. In general, the slope of the ramp function is equal to the magnitude of the step from which it is derived.

The *unit impulse* is not defined so easily as the unit ramp. Consider the modified ramp function shown in Fig. 8-7. This function is linear

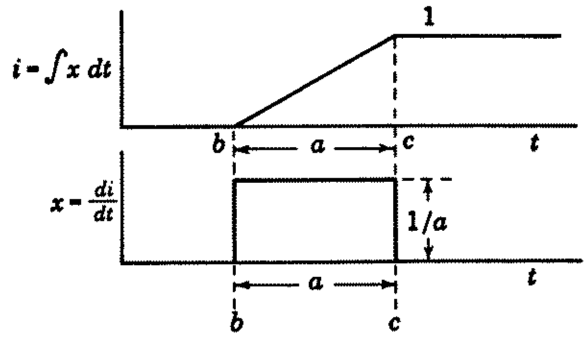


Fig. 8-7. Derivation of unit impulse.

from $t = b$ to $t = c$ and then has a constant value of unity for all time. The time interval $(c - b)$ is defined as a. The derivative of this modified ramp function is a pulse of width a as shown in the figure. (Conversely, the integral of the pulse function is the linear ramp.) If the ramp function is designated as the variable i, the pulse has a magnitude di/dt, the slope of the ramp. The slope of the ramp is the distance 1 divided by the distance a; that is,

$$\frac{di}{dt} = \frac{1}{a} \tag{8-13}$$

Now the area of the pulse is $a \times 1/a = 1$, for any value of a. As a approaches zero, the modified ramp function approaches a unit step function. At the same time, the pulse approaches infinite height and zero width *with the area remaining constant at unity*. In the limit, this function is known as a *unit impulse*, and is designated $\delta(t - b)$.* This symbolism indicates a function which is zero when $t \neq b$ and infinite when $t = b$. Also, since the total area under the curve is unity,

$$\int_{-\infty}^{\infty} \delta(t - b)\, dt = 1 \tag{8-14}$$

(The integral has the same value for any limits which bound the time $t = b$.)

From this discussion, we see that the derivative of a unit step function is a unit impulse. The same process of reasoning might be used

* In mathematical physics, this function is called a Dirac delta function.

to find the derivative of the unit impulse. Consider the waveform shown in Fig. 8-8—a trapezoid made up of a pulse with two ramp functions. In order to visualize the mathematical limit, first let $a \to 0$, then let $d \to 0$. As a approaches zero, i assumes the form of a pulse

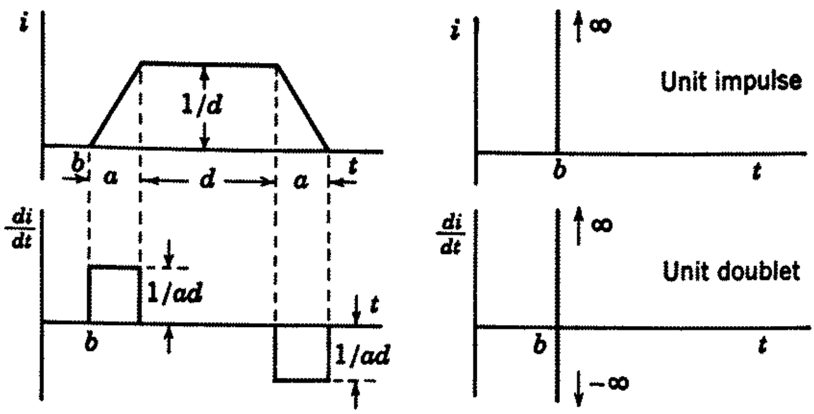


Fig. 8-8. Derivation of unit doublet in the limit, $a \to 0$, $d \to 0$.

and di/dt becomes two impulses, separated by the distance d, one positive going, the other negative going. As d approaches zero, i approaches a unit impulse; di/dt remains in the form of two infinite going impulses, but the two impulses superimpose at $t = b$. This resulting function is called a *unit doublet.* It is the derivative of a unit impulse.

This process might be continued to give a unit triplet. These functions, with discontinuous behavior with time, are known as *singular functions.* Of this family, the unit step function is an old friend. The ramp function, while not too familiar, seems friendly enough. But the unit impulse and unit doublet are rather terrifying! To break the ice, let us see what happens when the impulse is applied to ordinary elements—inductance and capacitance.

The basic equation relating current and voltage in an inductance is

$$i = \frac{1}{L}\int v\,dt \tag{8-15}$$

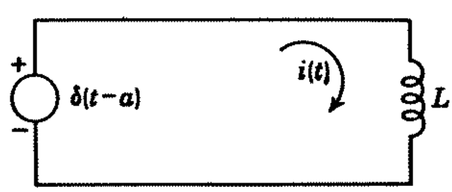


Fig. 8-9.

For this problem,

$$v(t) = \delta(t - a) \tag{8-16}$$

indicating a unit impulse at time $t = a$. Substituting into Eq. 8-15 to find the current, we obtain

$$i(t) = \frac{1}{L}\int_0^b \delta(t - a)\,dt = \begin{cases} \dfrac{1}{L} & \text{for} \quad b > a \\ 0 & \text{for} \quad b < a \end{cases} \tag{8-17}$$

That is, application of a unit impulse at $t = a$ causes a step function of current to start flowing at $t = a$. The current is

$$i(t) = \frac{1}{L} u(t - a) \tag{8-18}$$

This is a rather unusual behavior for the conservative inductance, but after all, it was hit by a rather unusual driving force. Another way of stating this unusual property of the unit impulse is that a voltage impulse of L units will cause a current of 1 amp to be established in an inductance immediately. A similar relationship may be found for a capacitor. Let a current source of value

$$i(t) = \delta(t - a) \tag{8-19}$$

be applied to a capacitance as shown in Fig. 8-10. The voltage across the capacitance is given by the basic relationship

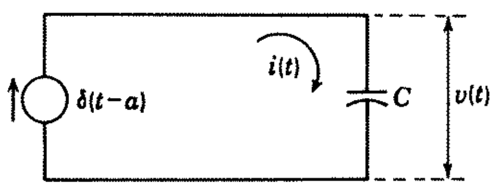


Fig. 8-10.

$$v(t) = \frac{1}{C} \int i(t)\, dt \tag{8-20}$$

This integral is evaluated as before and the result is

$$v(t) = \frac{1}{C} u(t - a) \tag{8-21}$$

In other words, a unit impulse of current applied to a capacitance causes $1/C$ volts to appear instantaneously on the capacitance because a unit impulse of current delivers a unit charge.

To illustrate the application of the concept of a unit impulse in terms of a familiar problem, consider the second-order differential equation

$$L \frac{d^2 i}{dt^2} + R \frac{di}{dt} + \frac{1}{C} i = v(t) \tag{8-22}$$

If the driving-force voltage is taken as a unit step function $v(t) = u(t)$, the solution of the equation is the familiar solution for an RLC circuit studied in Chapter 6. Now suppose that the equation is differentiated once as

$$L \frac{d^2}{dt^2}\left(\frac{di}{dt}\right) + R \frac{d}{dt}\left(\frac{di}{dt}\right) + \frac{1}{C}\frac{di}{dt} = \frac{d}{dt}[v(t)] \tag{8-23}$$

This equation is exactly the same in form as Eq. 8-22 where the variables di/dt and dv/dt have replaced i and v. Now if $v(t)$ is a unit step function, the derivative of $v(t)$ with respect to time is a unit impulse function, $\delta(t)$. With the solution for $i(t)$ of Eq. 8-22 known with a unit step function driving force, the solution for Eq. 8-23 with a unit impulse driving force can *be found by simply differentiating* $i(t)$ found for Eq. 8-22. In other words, the unit impulse response of a network can be found by solving for the unit step function response of the network and differentiating it.

The same process will work in reverse. The step function response may be found by integrating the impulse response. Likewise, the ramp function response may be found by integrating the step function response. Since all the singular functions are related by differentiation and integration, once the solution for one singular function is known, the solution for other singular functions is readily found by simple differentiation or integration. This is an important property of singular functions.

8-3. The Laplace transform for shifted and singular functions

The two previous sections of this chapter have been devoted to mathematical representation of shifted and singular functions. In this section, we will consider the derivation of the transforms of these functions.

The unit step function beginning at $t = a$ (where a is a constant), shown in Fig. 8-2(b), has been represented by the notation $u(t - a)$. The Laplace transform of this function may be computed from the defining equation,

$$F(s) = \int_0^\infty f(t)e^{-st}\,dt$$

For the case, $f(t) = u(t - a)$,

$$\mathcal{L}u(t - a) = \int_a^\infty 1e^{-st}\,dt = \left.\frac{-e^{-st}}{s}\right|_a^\infty$$

$$\mathcal{L}u(t - a) = e^{-as}\left(\frac{1}{s}\right) \tag{8-24}$$

This equation is made up of the product of two factors: the factor $1/s$ is the transform of the unit step function beginning at the time $t = 0$; the term e^{-as} is a function which effectively "*shifts*" the transform from one beginning at $t = 0$ to one beginning at $t = a$.

The example given for a unit step function may be generalized for any time function $f(t)$ which is delayed in its beginning to some other time, $t = a$. Such a time shifted function is represented as

$$f(t - a)u(t - a) \tag{8-25}$$

To find the transform of this equation, we write the defining equation in terms of a new variable, t'; that is,

$$F(s) = \int_0^\infty f(t')e^{-st'}\,dt' \tag{8-26}$$

Let the variable t' be defined as $t' = t - a$ such that the defining equation becomes

$$F(s) = \int_a^\infty f(t-a)e^{-(t-a)s}\,dt \tag{8-27}$$

or

$$= \int_a^\infty f(t-a)e^{-st}(e^{as})\,dt \tag{8-28}$$

The constant factor e^{as} may be removed from within the integral and the lower limit of the integral changed to 0 if $f(t - a)$ is multiplied by $u(t - a)$; thus

$$F(s) = e^{as}\int_0^\infty f(t-a)u(t-a)e^{-st}\,dt \tag{8-29}$$

The integral expression is recognized as the transform of the time function $f(t - a)u(t - a)$, so that

$$\mathcal{L}f(t-a)u(t-a) = e^{-as}\mathcal{L}f(t) \tag{8-30}$$

or, conversely,

$$\mathcal{L}^{-1}e^{-as}\mathcal{L}f(t) = f(t-a)u(t-a) \tag{8-31}$$

These last two equations tell us that the transform of any function delayed to begin at the time $t = a$ is equal to e^{-as} times the transform of the function beginning at the time $t = 0$.

A number of examples will illustrate the use of the last two equations. In the network shown in Fig. 8-11, a pulse of unit amplitude

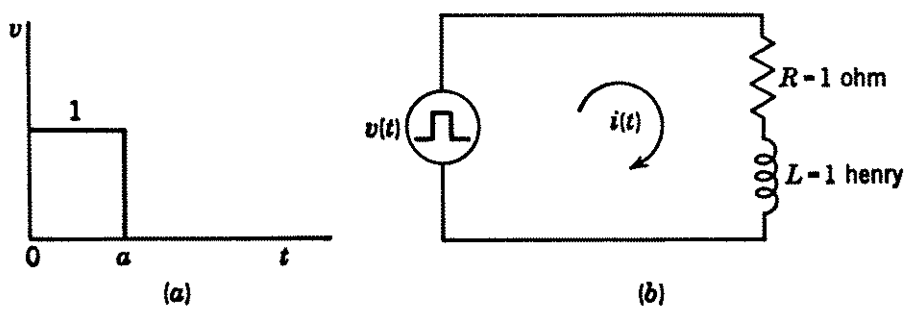


Fig. 8-11. Pulsed *RL* circuit.

and width a is applied to a series *RL* circuit. Let it be required to find the current flowing in the network. The pulse is given by the difference of two unit step functions as

$$v(t) = u(t) - u(t-a) \tag{8-32}$$

From Eq. 8-30 the transform of this voltage is

$$V(s) = \frac{1}{s}(1 - e^{-as}) \tag{8-33}$$

Substituting this value of V into the transform equation,

$$L[sI(s) - i(0+)] + RI(s) = \frac{1}{s}(1 - e^{-as}) \tag{8-34}$$

Substituting the parameter values and the initial condition, $i(0+) = 0$ gives

$$I(s) = \frac{(1 - e^{-as})}{s(s+1)} \tag{8-35}$$

This expression may be written as a sum of terms,

$$I(s) = \frac{1}{s(s+1)} - \frac{e^{-as}}{s(s+1)} \tag{8-36}$$

The first term of this equation is easily expanded by partial fractions to give

$$\frac{1}{s(s+1)} = \frac{1}{s} - \frac{1}{s+1} \tag{8-37}$$

In terms of this expansion, Eq. 8-36 may be written

$$I(s) = \frac{1}{s} - \frac{1}{s+1} - \frac{e^{-as}}{s} + \frac{e^{-as}}{s+1} \tag{8-38}$$

The inverse Laplace transformation may be carried out term by term in this equation to give

$$\mathcal{L}^{-1}I(s) = i(t) = 1 - e^{-t} - u(t-a) + e^{-(t-a)}u(t-a) \tag{8-39}$$

The third and fourth terms of this expression differ from the first and second only in that they are shifted in time and are opposite in sign. The waveform represented by this equation is plotted as Fig. 8-12.

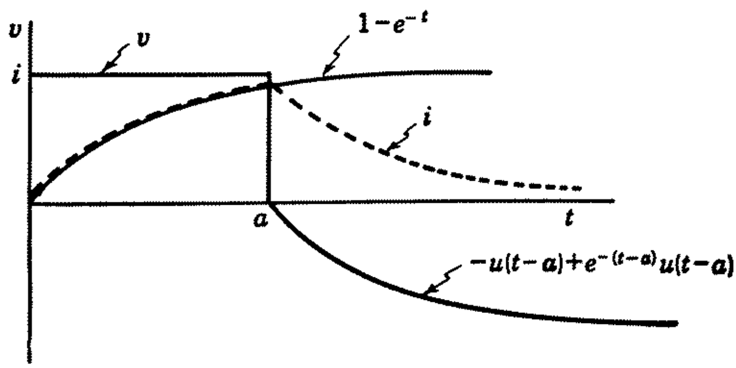


Fig. 8-12. Response of *RL* circuit.

The result we have obtained is the same as would be found by using two voltage sources and the principle of superposition. The resulting current waveform, shown in Fig. 8-12, is the summation of the two responses of the circuit caused by the superimposed voltage sources that make up the pulse.

As a second example, consider the problem of representing the periodic square wave shown in Fig. 8-13 by a transform. The square

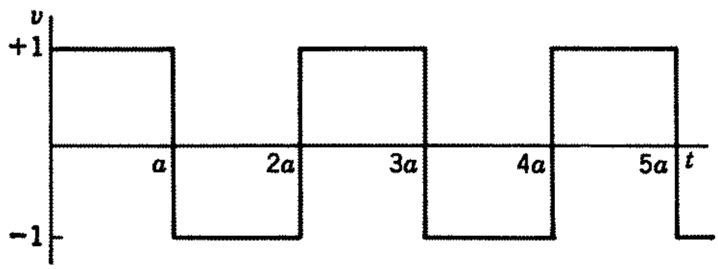


Fig. 8-13. Periodic square wave.

wave has been represented by an infinite sum of step functions of the form

$$v(t) = u(t) - 2u(t-a) + 2u(t-2a) - 2u(t-3a) + \dots \tag{8-40}$$

The Laplace transformation may be applied to this expression term by term to give

$$V(s) = \frac{1}{s} - 2\frac{e^{-as}}{s} + 2\frac{e^{-2as}}{s} - 2\frac{e^{-3as}}{s} + \dots \tag{8-41}$$

By factoring out common terms, the equation becomes

$$V(s) = \frac{1}{s}[1 - 2e^{-as}(1 - e^{-as} + e^{-2as} - e^{-3as} + \dots)] \tag{8-42}$$

The infinite series appearing in this equation may be identified by the following expansion from the binomial theorem,

$$\frac{1}{1+e^{-as}} = 1 - e^{-as} + e^{-2as} - e^{-3as} + \dots \tag{8-43}$$

such that $V(s)$ becomes

$$V(s) = \frac{1}{s}\left(1 - \frac{2e^{-as}}{1+e^{-as}}\right) = \frac{1}{s}\left(\frac{1-e^{-as}}{1+e^{-as}}\right) \tag{8-44}$$

or, finally,

$$V(s) = \frac{1}{s}\tanh\frac{as}{2} \tag{8-45}$$

The procedure outlined in the example may be applied to any periodic function. The transform of any such function of period T is

$$F(s) = \int_0^\infty f(t)e^{-st}\,dt = \int_0^T f(t)e^{-st}\,dt + \int_T^{2T} f(t)e^{-st}\,dt + \ldots \tag{8-46}$$

By successively shifting each transform term by e^{-nsT} where n is the number of shifts necessary to make the limits of the integral expression 0 to T, we have

$$F(s) = (1 + e^{-sT} + e^{-2sT} + \ldots) \int_0^T e^{-st} f(t)\,dt \tag{8-47}$$

Using the binomial theorem to identify the series,

$$F(s) = \frac{1}{1 - e^{-Ts}} \int_0^T e^{-st} f(t)\,dt \tag{8-48}$$

This equation may be used to compute the transform of any periodic waveform, and requires only one integration.

We now turn our attention to the transform of the unit impulse $\delta(t - b)$. The properties of this function were discussed on page 158. In terms of the sketch shown as Fig. 8-14, the unit impulse may be defined as the limit

$$\delta(t - b) = \lim_{a \to 0} \frac{1}{a}\,[u(t - b) - u(t - b - a)] \tag{8-49}$$

The Laplace transform of this limit equation is

$$\mathcal{L}\delta(t - b) = \lim_{a \to 0} \frac{e^{-bs} - e^{-(b+a)s}}{as} \tag{8-50}$$

This limit may be found by the application of l'Hospital's rule. The result is

$$\mathcal{L}\delta(t - b) = e^{-bs} \tag{8-51}$$

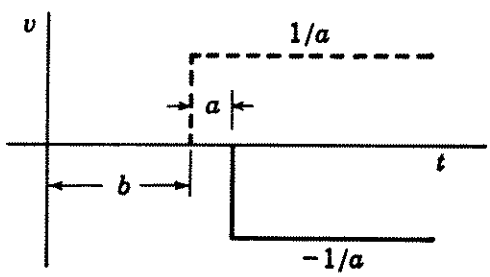


Fig. 8-14. The unit impulse.

where b is the time of appearance of the unit impulse. When $b = 0$ (that is, the impulse occurs at $t = 0$), we have

$$\mathcal{L}\delta(t) = 1 \tag{8-52}$$

This result has significance in terms of a transfer function. The voltage ratio transfer function of a network is

$$\frac{V_{out}(s)}{V_{in}(s)} = G(s) \tag{8-53}$$

If $v_{in}(t) = \delta(t)$, a unit impulse, then $V_{in}(s) = 1$, and

$$V_{out}(s) = G(s) \tag{8-54}$$

that is, the output voltage from a unit impulse input voltage is determined solely by the transfer function of the network. Such a response truly characterizes the network. We will exploit this fact in our study of the convolution integral in the next section.

Consider next the unit doublet and the transform of this function. The procedure parallels that given for the unit impulse. The unit doublet is defined by the limit

$$\lim_{a \to 0} \frac{1}{a^2} [u(t) - 2u(t - a) + u(t - 2a)] \tag{8-55}$$

The Laplace transform, term by term, of this limit is

$$\lim_{a \to 0} \frac{(1 - 2e^{-as} + e^{-2as})}{sa^2} \tag{8-56}$$

Again, l'Hospital's rule may be used. The second differentiation of numerator and denominator yields

$$\lim_{a \to 0} (-se^{-as} + 2se^{-2as}) = s \tag{8-57}$$

Thus the unit doublet has the transform s. This result might have been anticipated from the fact that the doublet is the derivative of the unit impulse. *If initial conditions are ignored*, differentiation corresponds to multiplication by s, while integration corresponds to division by s. The relationship among the family of singular functions (*not* shifted from $t = 0$) is tabulated as follows:

Function	Laplace transform
Unit ramp	$1/s^2$
Unit step	$1/s$
Unit impulse	1
Unit doublet	s
Unit triplet	s^2

The table might be further extended either direction (up or down). As was suggested earlier, if the response to any of the singular functions is known, the response for any other singular function may be found by differentiation or integration. Further, from one known response (frequently the impulse response), the response for any other driving force may be found by the use of the convolution integral to be discussed in the next section.

8-4. The convolution integral

An integral expression that appears frequently in network theory has the form

$$g(t) = \mathcal{L}^{-1}F_1(s)F_2(s) = \int_0^t f_1(t-\lambda)f_2(\lambda)\,d\lambda \tag{8-58}$$

where $\quad \mathcal{L}^{-1}F_1(s) = f_1(t) \quad \text{and} \quad \mathcal{L}^{-1}F_2(s) = f_2(t)$

where λ is a variable of integration. This expression is known as the *convolution integral** in which $f_1(t)$ and $f_2(t)$ are *convolved* to give $g(t)$ by the process of *convolution*. In this section, we will study the applications of the convolution integral: the use of this equation to find new transform pairs and the use of this equation to find the response of networks for complicated inputs.

As an example, suppose that the $f(t)$'s corresponding to $F(s) = 1/s$ and to $F(s) = 1/(s+1)$ are known, and that the inverse transform for $F(s) = 1/s(s+1)$ is to be found. If we designate $F_1(s)$ as $F_1(s) = 1/s$, then $f_1(t) = 1$ or $u(t)$, and similarly, if $F_2(s) = 1/(s+1)$, then $f_2(t) = e^{-t}$. From Eq. 8-58, $f_1(t)$ and $f_2(t)$ may be convolved to give $g(t)$ as follows:

$$\begin{aligned} g(t) &= \mathcal{L}^{-1}\frac{1}{s(s+1)} \\ &= \int_0^t u(t-\lambda)e^{-\lambda}\,d\lambda \end{aligned}$$

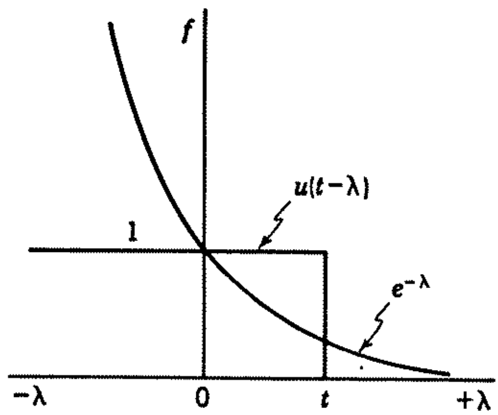


Fig. 8-15. Functions involved in the evaluation of the convolution integral.

The evaluation of this integral expression requires interpretation of the terms in the integral which are shown in Fig. 8-15. The exponential $e^{-\lambda}$ is shown for both positive and negative λ. The unit step function $u(t-\lambda)$ has unit value for $\lambda < t$ and zero value for $\lambda > t$, as was discussed in Art. 8-1. Since the unit step function $u(t-\lambda)$ has unit value over the limits of integration, it may be removed from the integral expression to give

$$g(t) = \int_0^t e^{-\lambda}\,d\lambda = -e^{-\lambda}\Big]_0^t = 1 - e^{-t} \tag{8-59}$$

The same result was given by partial fractions in the last chapter. It

* Proof of this equation can be found in Salvadori and Schwarz, *Differential Equations in Engineering Problems* (Prentice-Hall, Inc., New York, 1954), p. 214; also in Wylie, *Advanced Engineering Mathematics* (McGraw-Hill Book Co., Inc., New York, 1951), p. 188.

should be pointed out that the choice of functions to be designated $f_1(t)$ and $f_2(t)$ is arbitrary and does not affect the result.

Consider next a two-terminal-pair network, shown in Fig. 8-16, with a voltage ratio transfer function $G(s)$.* Assume that the output voltage transform is given in terms of the input voltage transform by the equation

Fig. 8-16. Two-terminal-pair network.

$$V_2(s) = G(s)V_1(s) \qquad (8\text{-}60)$$

where $V_1(s)$ is the input voltage transform and $V_2(s)$ is the output voltage transform. In terms of the convolution integral, let

$$F_1(s) = V_1(s) \quad \text{and} \quad f_1(t) = v_1(t) \qquad (8\text{-}61)$$

and

$$F_2(s) = G(s) \quad \text{and} \quad f_2(t) = h(t) \qquad (8\text{-}62)$$

Function $F_2(s)$ is identified as the transfer function $G(s)$; $f_2(t) = h(t)$ is the related time-domain response. From the discussion of the last section, it will be recognized that $h(t)$ is the *unit impulse response* of the network with a transfer function $G(s)$. For a unit impulse input, the output is determined by the inverse Laplace transform of the transfer function. This function *is* $h(t)$. From the convolution integral,

$$g(t) = \mathcal{L}^{-1}[F_1(s)F_2(s)] = \mathcal{L}^{-1}[V_1(s)G(s)] \qquad (8\text{-}63)$$

By Eq. 8-60,

$$g(t) = \mathcal{L}^{-1}V_2(s) = v_2(t) \qquad (8\text{-}64)$$

Thus $g(t)$ is identified as the output voltage in the time domain. The convolution integral has the form

$$v_2(t) = \int_0^t v_1(t - \lambda)h(\lambda)\, d\lambda \qquad (8\text{-}65)$$

This equation indicates that if $h(t)$, the unit impulse response, is known, only the input voltage $v_1(t)$ need be specified in order to determine the output voltage! In other words, any input convolved with the unit impulse response gives the output.

In order to get a better picture of the meaning of the convolution integral of Eq. 8-65, let us examine each term in the expression. First, consider the term, $v_1(t - \lambda)$. An arbitrary $v_1(t)$ is shown in Fig. 8-17. In terms of this plot, what is λ? When $t = 0$, then $v_1(t - \lambda) = v_1(-\lambda)$ and when $t = \lambda$, then $v_1(t - \lambda) = v_1(0)$. Evidently λ is a quantity measured backward from any t; that is, it is a time interval measured

* Transfer functions will be studied in Chap. 10. For the present, assume that $G(s)$ is an algebraic function relating $V_1(s)$ and $V_2(s)$.

negatively from some specified reference time t. This is illustrated in the figure. The quantity λ can vary from 0 to t, the limits of the integration. As $+\lambda$ varies from 0 to t, then $v_1(t - \lambda)$ ranges through all past values of the input (the input is assumed to start at $t = 0$). What about $h(t)$ and $h(\lambda)$? The quantity $h(t)$ is the transient response to the unit impulse. Its exact form depends on the transfer function, which we have not yet specified. It might have an appearance similar to the waveform shown in Fig. 8-18. A plot of $h(\lambda)$ could be super-

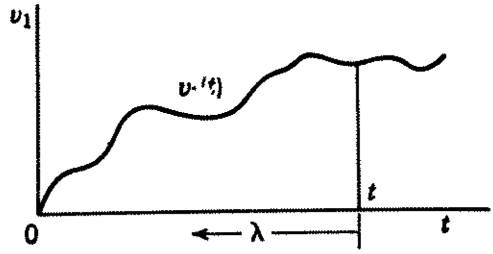


Fig. 8-17. Arbitrary input voltage.

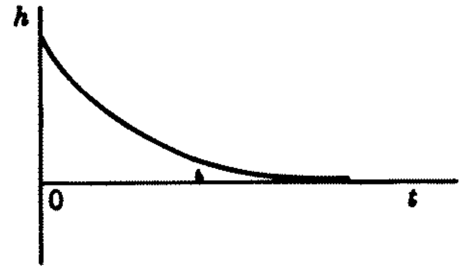


Fig. 8-18. Impulse response.

imposed on Fig. 8-17 with λ increasing from $t = t$ to $t = 0$ (that is, backward from Fig. 8-18). The product of $h(\lambda)$ and $v_1(t - \lambda)$ must next be integrated from 0 to t to give the output response. This process can be thought of as weighting all past values of the input by the unit impulse response. Since $h(\lambda)$ is usually small for large λ, the output at any time—found by integration—is mainly influenced by recent values of input. "Very old" values of input have very little effect on the present output. Strictly speaking, the present output is determined by all past history of the input, weighted by the unit input response. For complicated forms of $v_1(t)$, it may be necessary to use numerical or graphical integration to find $v_2(t)$ from the convolution integral. Further, the integration must be repeated for each value of t of interest.

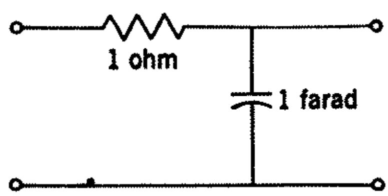


Fig. 8-19. *RC* network.

As a very simple application of this concept, suppose that the response from the driving force $v_1 = e^{-2t}$ is required for the two-terminal-pair network shown in Fig. 8-19. For this network, the transfer function for the voltage ratio is*

$$G(s) = \frac{1}{s + 1} \tag{8-66}$$

From a table of transforms, the corresponding $h(t)$ is found to be

$$h(t) = e^{-t} \tag{8-67}$$

* The computation of this transfer functions is given in Chap. 10.

Then, from the convolution integral,

$$v_2(t) = \int_0^t e^{-2(t-\lambda)}e^{-\lambda}\, d\lambda \tag{8-68}$$

$$= e^{-2t}\int_0^t e^{2\lambda}e^{-\lambda}\, d\lambda = e^{-2t}e^{\lambda}\Big|_0^t \tag{8-69}$$

Finally,
$$v_2(t) = e^{-2t}(e^t - 1) = e^{-t} - e^{-2t} \tag{8-70}$$

For this particular example, expansion by partial fractions is more direct. For more complicated forms of input, the convolution integral can be used to advantage.

Thus far the application of the convolution integral has been in terms of the unit impulse response of a system. If the unit impulse response is known, the unit step function response can be found by integration as discussed in the last section. In some cases, however, the step function response is more conveniently recorded. The convolution integral can be put in another form for this case. Equation 8-65, which is

$$v_2(t) = \int_0^t v_1(t-\lambda)h(\lambda)\, d\lambda \tag{8-71}$$

can be integrated by parts by letting $u = v_1(t-\lambda)$ and $dv = h(\lambda)\, d\lambda$. The resulting equation is

$$v_2(t) = k(t)v_1(0) + \int_0^t k(\lambda)v_1'(t-\lambda)\, d\lambda \tag{8-72}$$

where $k(t)$ is the unit step function response of the system. We note that integration of $h(\lambda)$ to give $k(\lambda)$ is compensated within the integral by differentiation of v_1 with respect to time.

This last equation illustrates a useful property of convolution. If the unit step function response of a system is determined, the response of the system to any input v_1 is fixed and can be determined by convolution. The statements made for the unit impulse and unit step can be extended to any of the family of singular functions.

8-5. Fourier series

This section marks the turning point in our study. Behind are studies in the time domain, the response of a network to a given time-varying driving force. The studies yet to come concern the response of a network to a sinusoidal driving force of variable frequency in addition to time-domain topics.

The modern electrical engineer must be bilingual when speaking of network response. He must speak the language of the time domain and must also be trained in the language of the frequency domain. He may think in either language, but he must be able to translate from one to the other at a moment's notice. The translation process may be purely mechanical, "large bandwidth" equals "desirable step function response." Or it may be based on an understanding of the concepts of the two domains in terms of a common root or origin.

To begin with, we have now extended our time-domain studies to include the response of a network to (1) a nonrecurring pulse and (2) recurring and periodic waveforms such as the square wave. We will next study these two classes of driving force functions in terms of sinusoids.

What do we mean by a periodic function (or waveform)? The familiar sine wave is periodic. If represented as sin ωt, and if θ_1 is some value of ωt after $\omega t = 0$, then for a sine wave,

$$\sin \theta_1 = \sin (2n\pi + \theta_1), \qquad n = \text{any integer}$$

since the function has identical form from $\omega t = 0$ to $\omega t = 2\pi$, from $\omega t = 2\pi$ to $\omega t = 4\pi$, etc. Similarly, any function is periodic in ωt if $f(\theta_1) = f(\theta_1 + 2n\pi)$ and the period is 2π.

Such periodic functions were studied by the French mathematician Fourier (1768–1830) who was the first to show that periodic functions could be expanded in series form in terms of harmonically related sinusoids as

$$f(\omega t) = a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + \dots + a_n \cos n\omega t \\ + \dots + b_1 \sin \omega t + \dots b_n \sin n\omega t + \dots \qquad (8\text{-}73)$$

This series is known as the *Fourier series*, and the process of representing a periodic function by such a series is *Fourier analysis*. The problem of analysis is determination of the values of the coefficients of the Fourier series, $a_0, a_1, a_2, \dots, b_1, b_2, \dots$, for a given time function $f(\omega t)$.

Suppose we select the coefficients of the Fourier series such that $a_1 = 1$, and $b_2 = -1$, and all other coefficients are equal to zero. The plot of the combination of the two functions with t is shown in Fig. 8-20, and the resulting $f(\omega t)$ has small amplitude from 0 to π and large amplitude from π to 2π. This distorted waveform resulted from the combination of merely two harmonically related terms. It seems quite possible that any periodic function could be synthesized with the infinite number of terms that are available in the series.

Equations for the coefficients of the Fourier series for use in analysis are found by the mathematical procedure of (1) multiplication of the series by a suitable factor, (2) integration of the resulting equations

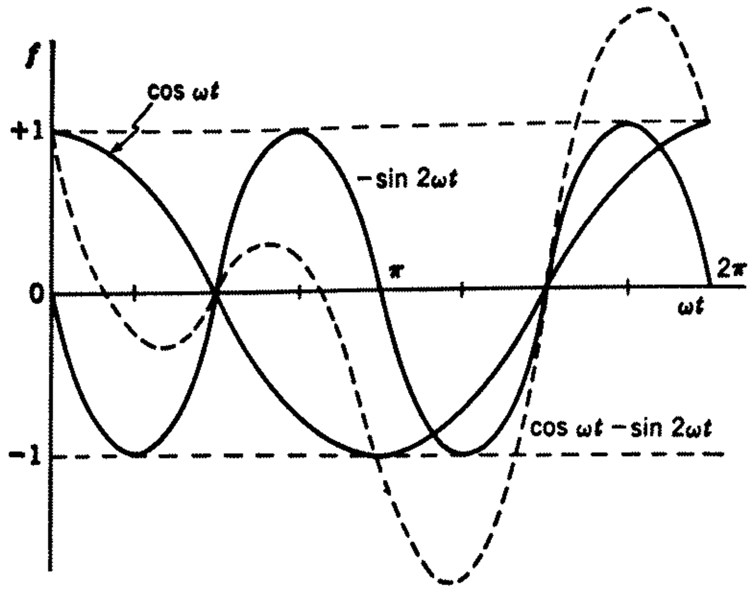


Fig. 8-20. Waveform resulting from the addition of (cos ωt) and ($-$ sin $2\omega t$).

term by term over the period, and (3) simplification by the use of the following definite integrals.

$$\int_0^{2\pi} \cos n\omega t \, d\omega t = 0 \qquad n \neq 0 \tag{8-74}$$

$$\int_0^{2\pi} \sin n\omega t \, d\omega t = 0 \qquad n \neq 0 \tag{8-75}$$

$$\int_0^{2\pi} \sin m\omega t \cos n\omega t \, d\omega t = 0 \qquad m \neq n \tag{8-76}$$

$$\int_0^{2\pi} \sin n\omega t \cos n\omega t \, d\omega t = 0 \qquad n \neq 0 \tag{8-77}$$

$$\int_0^{2\pi} \sin m\omega t \sin n\omega t \, d\omega t = 0 \qquad m \neq n \tag{8-78}$$

$$\int_0^{2\pi} \cos m\omega t \cos n\omega t \, d\omega t = 0 \qquad m \neq n \tag{8-79}$$

$$\int_0^{2\pi} \cos^2 n\omega t \, d\omega t = \pi \qquad n \neq 0 \tag{8-80}$$

$$\int_0^{2\pi} \sin^2 n\omega t \, d\omega t = \pi \qquad n \neq 0 \tag{8-81}$$

These equations also hold for any other period, θ_1 to $\theta_1 + 2\pi$, and the limits of the integrals can be replaced by these more general terms.

In evaluating a_0, no multiplying term specified as step 1 is required. Integration of each term of the Fourier series gives

$$\int_0^{2\pi} f(\omega t)\,d\omega t = a_0 \int_0^{2\pi} d\omega t + a_1 \int_0^{2\pi} \cos \omega t\,d\omega t + a_2 \int_0^{2\pi} \cos 2\omega t\,d\omega t$$
$$+ \ldots + a_n \int_0^{2\pi} \cos n\omega t + \ldots + b_1 \int_0^{2\pi} \sin \omega t\,d\omega t + \ldots$$
$$+ b_n \int_0^{2\pi} \sin n\omega t\,d\omega t + \ldots \quad (8\text{-}82)$$

assuming that such term-by-term integration is permitted. In Eq. 8-82, all terms on the right except the first have zero value by Eqs. 8-74 and 8-75. Hence the total equation reduces to

$$\int_0^{2\pi} f(\omega t)\,d\omega t = a_0(2\pi) \quad (8\text{-}83)$$

or

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\omega t)\,d\omega t \quad (8\text{-}84)$$

To find a_n for n other than zero, each term in the Fourier series is multiplied by $\cos n\omega t$ and integrated from 0 to 2π. In the resulting expression, all integrals will vanish except the one of the form of Eq. 8-80, which is the integral with a coefficient a_n. The equation thus simplifies to give

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(\omega t) \cos n\omega t\,d\omega t \quad (8\text{-}85)$$

Similarly, the b_n coefficient is evaluated by multiplying by $\sin n\omega t$ and integrating over the period 0 to 2π, giving

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(\omega t) \sin n\omega t\,d\omega t \quad (8\text{-}86)$$

These three equations determine all coefficients of the Fourier series. These integrals hold when $f(\omega t)$ represents a finite periodic function with at most a finite number of maxima and minima and a finite number of discontinuities in every finite interval. These are the *Dirichlet conditions*, which must be satisfied for the Fourier series representation of $f(\omega t)$ to be valid. The practical consequence in terms of engineering application is that the Fourier series can be written for engineering functions without concern.

The amount of labor involved in the evaluation of the coefficients can be reduced when there is symmetry with respect to the axis in the plot of $f(\omega t)$. Figure 8-21(a) shows a plot of sine and cosine functions for positive and negative values of ωt. The cosine function is seen to have symmetry about the f axis, the same value for $+\omega t$ and $-\omega t$. Such a function is said to be an *even function*. In the case of the sine function, the value of the function for $-\omega t$ is the negative of that for

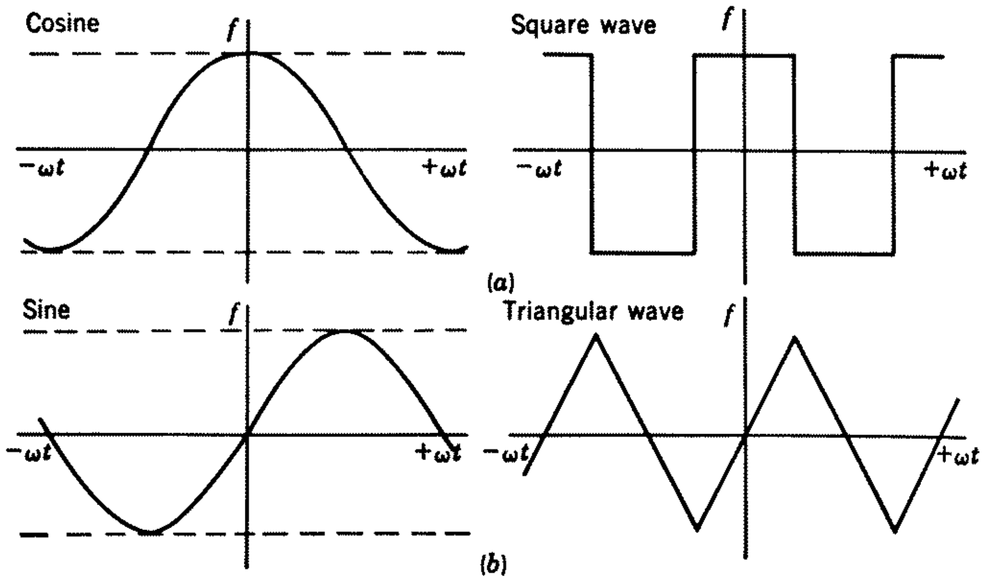


Fig. 8-21. (a) Even functions—cosine and square wave: (b) odd functions—sine and triangular waves.

$+\omega t$ and vice versa. Such a function is an *odd function*. Any general function may be described as odd or even when it meets the following conditions.

$$\text{Even function:} \quad f(\omega t) = f(-\omega t)$$
$$\text{Odd function:} \quad f(\omega t) = -f(-\omega t)$$

An even and an odd function are shown in Fig. 8-21(b), a square wave and triangular wave, respectively. The square wave is an even function (although it might be made odd by shifting the ωt axis). Being an even function, every term in its Fourier series representation must also be even; a single odd term would destroy the even symmetry. The same argument can be applied to the triangular wave in that its Fourier series must contain only odd terms. These conclusions of this discussion can be verified mathematically.*

The equation for the a_0 coefficient of the Fourier series is

$$a_0 = \frac{1}{2\pi} \int_0^{2\pi} f(\omega t)\, d\omega t$$

This integral represents the area under the $f(\omega t)$ curve from 0 to 2π.

* For example, see Wylie, *op. cit.*, p. 122, or Salvadori and Schwarz, *op. cit.*, p. 359.

If, as in the case of the square wave and the triangular wave of Fig. 8-21(b), there is as much positive area as negative area, the value of a_0 is zero. These three conclusions are summarized in the following table.

Condition	Simplification of Fourier series
$f(\omega t) = f(-\omega t)$	$b_n = 0$, all n
$f(\omega t) = -f(-\omega t)$	$a_n = 0$, all n including a_0
Equal positive and negative areas under the waveform over one cycle.	$a_0 = 0$

Example 1

Figure 8-22 shows a square wave function which we wish to represent by a Fourier series. From the figure it is seen that the symmetry is

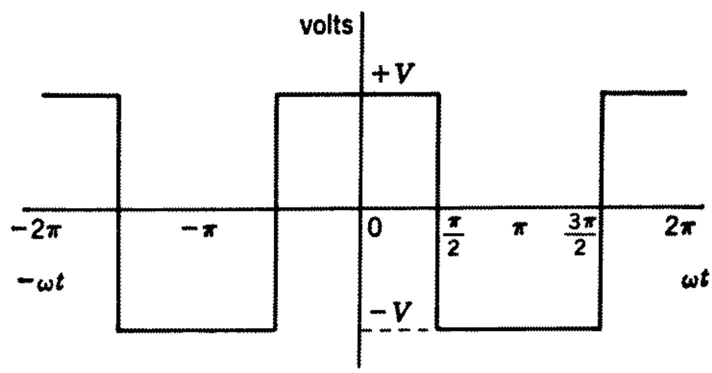


Fig. 8-22. Square wave function.

such that $v(\omega t) = v(-\omega t)$, and so $b_n = 0$. Since the total area under any cycle adds to zero, the coefficient a_0 is zero. The coefficients a_n are determined by evaluating the integral

$$a_n = \frac{1}{\pi} \int_0^{2\pi} v(\omega t) \cos n\omega t \, d\omega t$$

The voltage $v(\omega t)$ has the following set of values over one cycle:

Interval	$v(\omega t)$
0 to $\pi/2$	V
$\pi/2$ to $3\pi/2$	$-V$
$3\pi/2$ to 2π	V

These values may be substituted into the integral equation to give

a_n as

$$a_n = \frac{1}{\pi}\left(V\int_0^{\pi/2}\cos n\omega t\, d\omega t - V\int_{\pi/2}^{3\pi/2}\cos n\omega t\, d\omega t + V\int_{3\pi/2}^{2\pi}\cos n\omega t\, d\omega t\right)$$
$$= \frac{V}{n\pi}\left(\sin n\omega t\Big|_0^{\pi/2} - \sin n\omega t\Big|_{\pi/2}^{3\pi/2} + \sin n\omega t\Big|_{3\pi/2}^{2\pi}\right)$$

The term in parenthesis has the value of ± 4 for odd values of n, and zero for even values; hence

$$a_n = \begin{cases} \dfrac{+4V}{n\pi}, & n = 1, 5, 9, \ldots \\ \dfrac{-4V}{n\pi}, & n = 3, 7, 11, \ldots \\ 0, & n = \text{even integers} \end{cases}$$

Thus the Fourier series is

$$v(\omega t) = \frac{4V}{\pi}\left(\cos \omega t - \tfrac{1}{3}\cos 3\omega t + \tfrac{1}{5}\cos 5\omega t - \tfrac{1}{7}\cos 7\omega t + \ldots\right)$$

By a Fourier expansion, we have shown that a sum of voltage terms each varying sinusoidally is equivalent to a square wave, as illustrated in Fig. 8-23. By the principle of superposition, the response of each

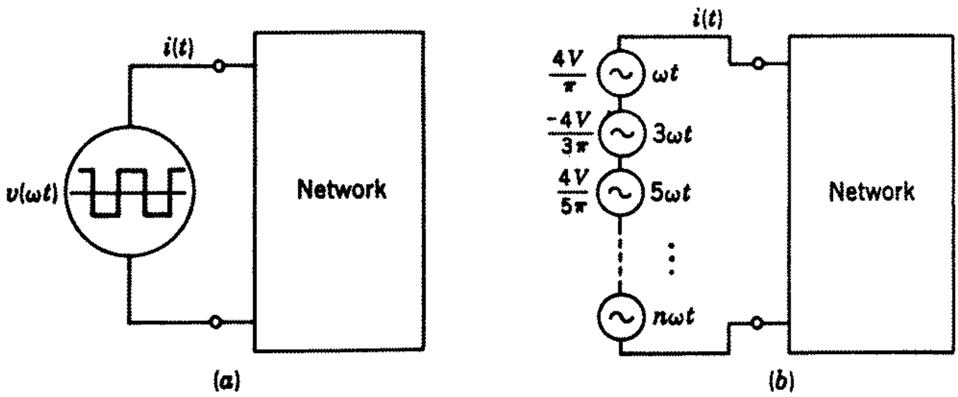


Fig. 8-23. Two equivalent systems: the Fourier series expansion of a time-varying driving-force function.

generator can be determined with all other generators short-circuited, and the total response will be the sum of the individual responses so found (as discussed in Chapter 4). This may appear to be complicating the problem rather than simplifying it, many solutions instead of just one being required. We have yet to show that it is easy to compute the response as a function of frequency by complex algebra and that in many cases we are not so interested in the actual response as we are in the frequency response required for a given waveform. In solving a problem, we have a choice between the solution of the transient problem (the time domain) and the solution of response in terms of a sinusoidal generator of variable frequency (frequency domain).

Example 2

A triangular (or saw-tooth) waveform is shown in Fig. 8-24. This voltage function is an odd function with a_n equal to zero for all n

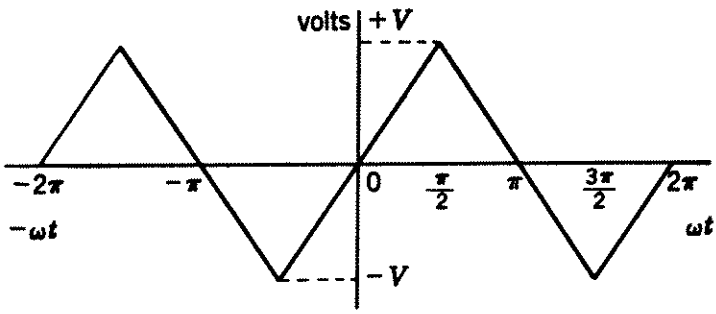


Fig. 8-24. Triangular (or sawtooth) function.

including $n = 0$. To find the b-coefficients, we represent the waveform by the following equations, each derived in terms of the equation for a straight line, $y = mx + b$, where m is the slope, b is the y intercept and for this problem, $y = v$ and $x = t$.

Interval	$v(\omega t)$
0 to $\dfrac{\pi}{2}$	$\dfrac{2V}{\pi}\,\omega t$
$\dfrac{\pi}{2}$ to $\dfrac{3\pi}{2}$	$\dfrac{-2V}{\pi}\,(\omega t) + 2V$
$\dfrac{3\pi}{2}$ to 2π	$\dfrac{2V}{\pi}\,(\omega t) - 4V$

Carrying out the integration, as in Example 1, gives

$$v(\omega t) = \frac{8V}{\pi^2}\left(\sin \omega t - \frac{1}{3^2}\sin 3\omega t + \frac{1}{5^2}\sin 5\omega t - \cdots\right)$$

Example 3

In some practical problems, the waveforms are not known in the form of mathematical equations but rather as recorded graphs. In such cases, the coefficients may be evaluated by approximate graphical integration. As shown in Fig. 8-25, the waveform is divided into m rectangles of width $\Delta\omega t$ and height $f(\omega t)$. In terms of a summation

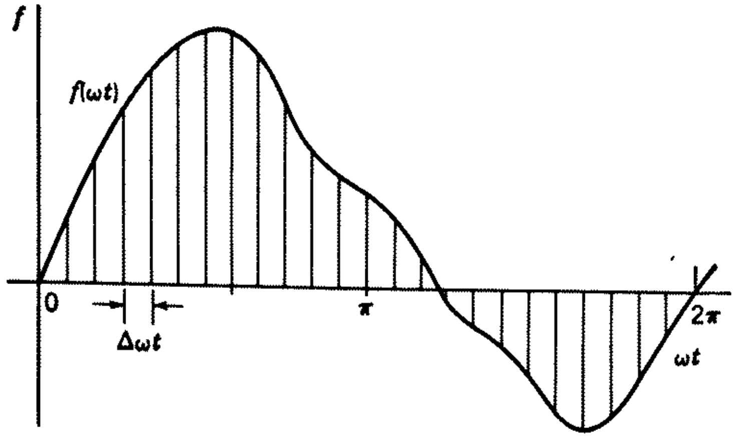


Fig. 8-25. A recorded waveform. The Fourier series equivalent of this waveform may be found by graphical integration.

rather than integration, the approximate equations for the Fourier coefficients become

$$a_0 = \frac{1}{2\pi} \sum_{j=1}^{m} f\left(\frac{j}{m} 2\pi\right) \Delta\omega t \tag{8-87}$$

$$a_n = \frac{1}{\pi} \sum_{j=1}^{m} f\left(\frac{j}{m} 2\pi\right) \cos n \left(\frac{j}{m} 2\pi\right) \Delta\omega t \tag{8-88}$$

$$b_n = \frac{1}{\pi} \sum_{j=1}^{m} f\left(\frac{j}{m} 2\pi\right) \sin n \left(\frac{j}{m} 2\pi\right) \Delta\omega t \tag{8-89}$$

where

$$\Delta\omega t = \frac{2\pi}{m} \tag{8-90}$$

The summations required are most conveniently carried out in tabular form, for example, as follows.

TABLE FOR CALCULATION OF a_n

$n = 3 \qquad \Delta\omega t = 15° \qquad m = 24$

j	$\frac{j}{m}(360°) = \theta$	$\cos n\left(\frac{j}{m} 360°\right)$	$f\left(\frac{j}{m} 360°\right)$	$f(\theta) \cos n\theta$
1	15°	cos 45° = 0.707	1.52	1.07
2	30°	cos 90° = 0	1.77	0
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
m				$\overline{\Sigma}$

The coefficient has the value

$$a_3 = 2\,\frac{\Sigma}{m} \tag{8-91}$$

8-6. Complex exponential form of the Fourier series

The Fourier series studied in the last section can be expressed in equivalent form in terms of exponential quantities. Suppose that the terms in the series are grouped together by harmonic number as

$$f(\omega t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t) \tag{8-92}$$

Now the cosine and sine may be expressed in exponential form, as we learned in Art. 6-2. Starting with Euler's equation,

$$e^{\pm j\omega t} = \cos \omega t \pm j \sin \omega t \tag{8-93}$$

the cosine is found in terms of exponentials by adding positive and negative exponential forms as

$$\cos \omega t = \tfrac{1}{2}(e^{j\omega t} + e^{-j\omega t}) \tag{8-94}$$

Similarly, the sine is found by subtracting these quantities as

$$\sin \omega t = \frac{1}{2j}\,(e^{j\omega t} - e^{-j\omega t}) \tag{8-95}$$

Substituting these equations into Eq. 8-92, there results

$$f(\omega t) = a_0 + \sum_{n=1}^{\infty} \left(a_n\,\frac{e^{jn\omega t} + e^{-jn\omega t}}{2} + b_n\,\frac{e^{jn\omega t} - e^{-jn\omega t}}{2j} \right) \tag{8-96}$$

In order to simplify this equation, like exponential terms are grouped. Noting that $1/j = -j$, our equation becomes

$$f(\omega t) = a_0 + \sum_{n=1}^{\infty} \left[\left(\frac{a_n - jb_n}{2} \right) e^{jn\omega t} + \left(\frac{a_n + jb_n}{2} \right) e^{-jn\omega t} \right] \tag{8-97}$$

To simplify this expression, we next introduce a new coefficient to replace the a and b coefficients. By definition,

$$c_n = \frac{a_n - jb_n}{2}, \quad c_{-n} = \frac{a_n + jb_n}{2}, \quad \text{and} \quad c_0 = a_0 \tag{8-98}$$

The new form for Eq. 8-97 is

$$f(\omega t) = c_0 + \sum_{n=1}^{\infty} (c_n e^{j\omega nt} + c_{-n} e^{-j\omega nt}) \tag{8-99}$$

We are now in a position to understand better all the maneuvering we have just been through. Letting n range through values from 1 to ∞ in this equation is equivalent to letting n range from $-\infty$ to $+\infty$ (including zero) in a compact equation,

$$f(\omega t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega t} \tag{8-100}$$

Here we have the exponential form of the Fourier series. The coefficients c_n can easily be evaluated in terms of a_n and b_n, which we already knew. Then

$$\begin{aligned} c_n &= \frac{1}{2\pi}\int_0^{2\pi} f(\omega t)\cos n\omega t\,d\omega t - \frac{j}{2\pi}\int_0^{2\pi} f(\omega t)\sin n\omega t\,d\omega t \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\omega t)(\cos n\omega t - j\sin n\omega t)\,d\omega t \\ &= \frac{1}{2\pi}\int_0^{2\pi} f(\omega t)e^{-jn\omega t}\,d\omega t \end{aligned} \tag{8-101}$$

This equation for c_n holds whether n is positive as we have assumed, negative, or zero, as can be shown by exactly the same procedure. Has this form any advantage over the other form of the Fourier series? In computing coefficients, the sine and cosine form usually may be used to advantage. But in discussing the concepts of frequency spectra and introducing the Fourier integral, which we will study next, we need this exponential form.

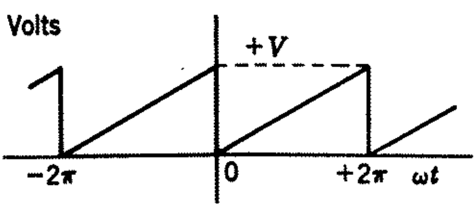


Fig. 8-26. Sweep voltage of the form used in a cathode ray oscillograph.

Example 4

The sweep voltage waveform shown in Fig. 8-26 may be represented over one cycle by the equation of a straight line, $v = (V/2\pi)\omega t$. The c_n-coefficients, defined by Eq. 8-101, are

$$\begin{aligned} c_n &= \frac{1}{2\pi}\int_0^{2\pi} \frac{V}{2\pi}\,\omega t e^{-jn\omega t}\,d\omega t \\ &= \frac{jV}{2n\pi}, \qquad n \neq 0 \\ &= \frac{V}{2}, \qquad n = 0 \end{aligned}$$

Hence the exponential form of the Fourier series for this waveform is

$$v(\omega t) = \ldots - \frac{jV}{6\pi} e^{-j3\omega t} - \frac{jV}{4\pi} e^{-j2\omega t} - \frac{jV}{2\pi} e^{-j\omega t} + \frac{V}{2} + \frac{jV}{2\pi} e^{j\omega t} + \frac{jV}{4\pi} e^{j2\omega t} + \ldots \qquad (8\text{-}102)$$

If we wish to reduce this result to the alternate form of Fourier series, the a and b coefficients may be found from the equations which follow from the definitions of Eq. 8-98.

$$a_n = c_n + c_{-n}, \qquad b_n = j(c_n - c_{-n}), \qquad a_0 = c_0$$

From these equations, $a_n = 0$, $a_0 = V/2$, and $b_n = -V/n\pi$, and the Fourier series becomes

$$v(\omega t) = V\left[\frac{1}{2} - \frac{1}{\pi}\left(\sin \omega t + \frac{1}{2}\sin 2\omega t + \frac{1}{3}\sin 3\omega t + \ldots\right)\right] \qquad (8\text{-}103)$$

8-7. The frequency spectra of periodic waveforms

The second form of the Fourier series of our example of the last section is the most easily interpreted. We can visualize a large number of sinusoidal generators of voltage as specified by the appropriate Fourier coefficient, all connected in series to produce a sweep voltage. We have some difficulty picturing generators of exponential voltage terms of the form appearing in Eq. 8-102, but the coefficients contain the same information as those in Eq. 8-103. This information is conveniently displayed in a plot of the magnitude of c_n, and sometimes of the phase,* as a function of frequency. Such a plot shows the *frequency spectrum* corresponding to a particular waveform.

The plot of the magnitude and phase of c_n as a function of frequency requires special interpretation. Actually, c_n *has values only for discrete values of frequency*, the harmonics of ω, the fundamental frequency. Such a plot is actually for different values of n, or if we identify ω_0 as the fundamental frequency (merely ω in the equations derived thus far), the plot is for discrete values of the ratio ω/ω_0, which is equivalent to n. Such a plot is shown in Fig. 8-27 for both positive and negative values of n or ω/ω_0.

We do not attach any particular significance to negative frequencies as plotted in the frequency spectrum. We do note that these frequencies are related to exponential factors of the form $e^{jn\omega t}$; such a term *and* a term of the form $e^{-jn\omega t}$ may be added to be equivalent to a sine or cosine function. Thus a positive frequency and a negative frequency

* The phase angle of c_n is abbreviated as Ang c_n.

combine to form the frequency associated with actual sinusoidal generators.

Figure 8-28 shows two other spectra, for the examples of Art. 8-5 for the square wave and the triangular wave. Comparing these spectra

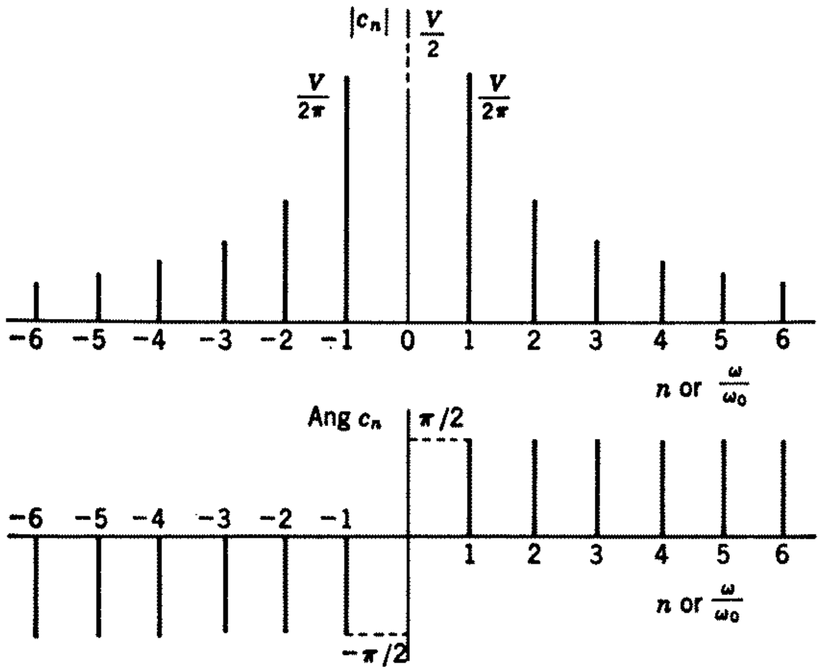


Fig. 8-27. The line spectrum corresponding to the sweep voltage of Fig. 8-26. Separate plots are made for magnitude and phase angle (Ang) of c_n.

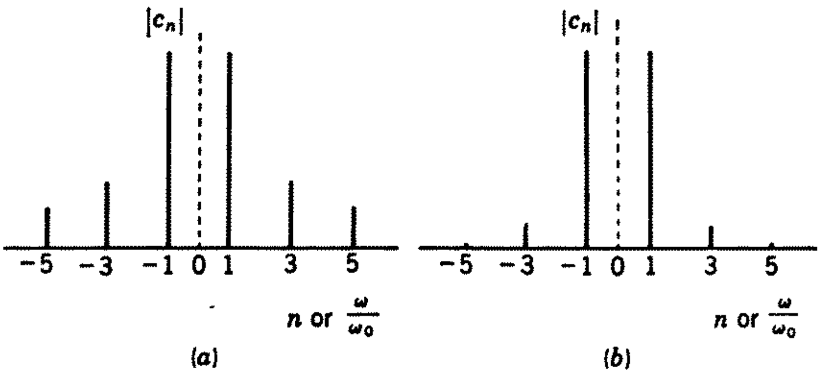


Fig. 8-28. The line spectrum for (a) a square wave, and (b) a triangular wave.

to that shown in Fig. 8-27, we see that the amplitude distribution in terms of the harmonics of the Fourier series must be quite different for the different waveforms. The triangular wave contains little in addition to the fundamental, while the sweep voltage waveform contains many harmonic terms of larger magnitude than for the triangular waveform.

Because there are components of frequency for discrete values of frequency only, such plots of the magnitude of c_n are known as *line spectra*.

8-8. The Fourier integral and continuous-frequency spectra

Figure 8-29 shows the waveform of a periodic pulse of magnitude V and duration a. The period, marked on the figure as T, extends from $\omega_0 t = -\pi$ to $\omega_0 t = +\pi$, where, following the practice established in

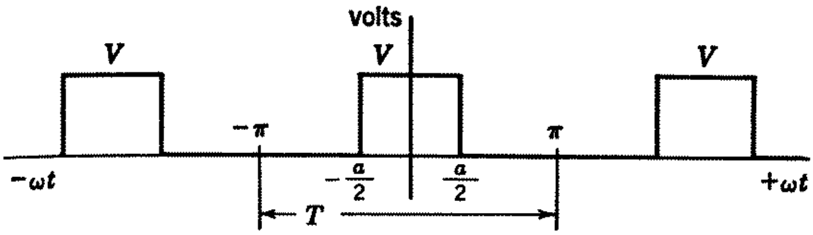


Fig. 8-29. A recurrent pulse of duration a and period T.

the last section, ω_0 is used as the frequency of the fundamental of the Fourier series rather than ω. The Fourier coefficients for the exponential form of the series may be computed for this problem from Eq. 8-101 written

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} f(\omega_0 t) e^{-jn\omega_0 t}\, d\omega_0 t \tag{8-104}$$

Since the voltage waveform has zero value except between the limits $(a/2)$ and $(-a/2)$, the integral becomes

$$c_n = \frac{1}{2\pi} \int_{-a/2}^{+a/2} V e^{-jn\omega_0 t}\, d\omega_0 t \tag{8-105}$$

$$c_n = \frac{-V}{2\pi} \frac{1}{jn} e^{-jn\omega_0 t} \Big]_{-a/2}^{+a/2} = \frac{V}{n\pi} \left(\frac{e^{jn\omega_0 a/2} - e^{-jn\omega_0 a/2}}{2j} \right)$$

$$= V \frac{\omega_0 a}{2\pi} \left(\frac{\sin (n\omega_0 a/2)}{n\omega_0 a/2} \right) \tag{8-106}$$

Now since $T = 2\pi/\omega_0$, the equation finally may be written

$$c_n = V \frac{a}{T} \frac{\sin (n\omega_0 a/2)}{n\omega_0 a/2} \tag{8-107}$$

For any particular problem, the ratio a/T will be fixed, and c_n will vary as the mathematical function $(\sin x)/x$.

This analysis brings up a number of questions of interest: (1) how does c_n change as the width of the pulse or the ratio a/T changes, and (2) what happens if the period becomes infinite, leaving us with one nonrecurring pulse?

To help answer the first question, two plots are made in Fig. 8-30, one for $(T/a) = 2$ and one for $(T/a) = 5$. The two plots differ in two respects: (1) there are more lines in the second plot and (2) the amplitude is smaller by a factor of $\frac{2}{5}$ in the second plot. From another point of view, there are more lines because ω_0 is smaller for the second pulse. More frequency components are required to make up the shorter pulse, but the amplitude of the frequency components is smaller.

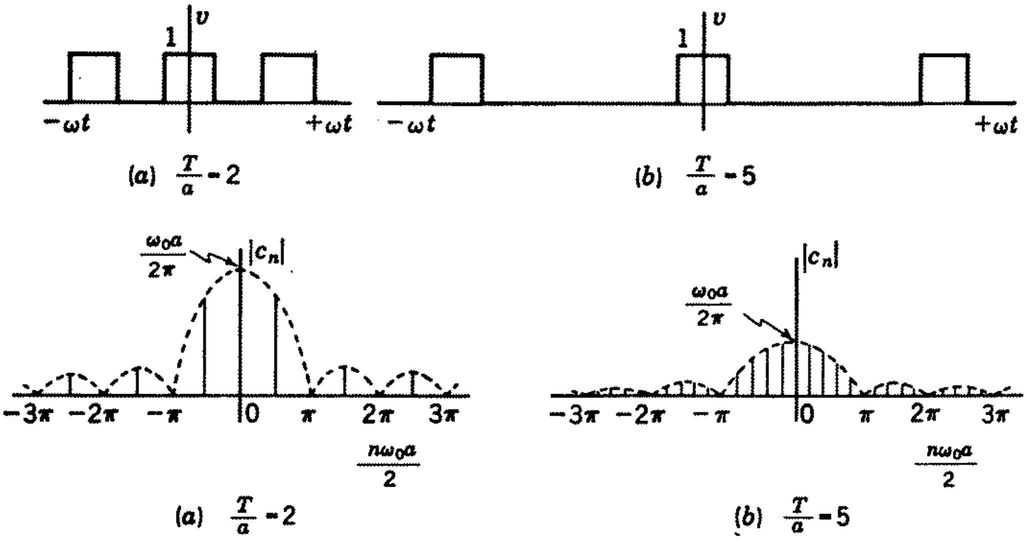


Fig. 8-30. Two recurrent pulses with different values of (T/a) and the corresponding line spectra holding a constant. The envelope of the line spectra is of the general form $(\sin x/x)$. Because $|c_n|$ is plotted, the envelope is always positive whereas $(\sin x/x)$ is negative from π to 2π, etc.

We begin to see a trend that should help answer our second question as to what happens in the case of an aperiodic pulse. As the period approaches infinity, more frequency components of smaller amplitude will be added. To accomplish this limit in terms of the expressions for the Fourier series,* we begin with Eq. 8-100, which is

$$f(\omega_0 t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_0 t} \tag{8-108}$$

and substitute the equation for c_n, giving

$$f(\omega_0 t) = \sum_{n=-\infty}^{\infty} \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega_0 t)e^{-jn\omega_0 t}\, d\omega_0 t\right] e^{jn\omega_0 t} \tag{8-109}$$

* The following discussion is intended to provide a heuristic proof or motivation for the Fourier integral theorem. It is not a rigorous derivation.

We next let $\omega_0 = \Delta\omega$ as $T \to \infty$ and introduce a new variable by letting $n\,\Delta\omega = \omega$. If we write Eq. 8-109 in terms of $f(t)$ rather than $f(\omega_0 t)$, the limits of integration change to give the following expression.

$$f(t) = \sum_{n=-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\pi/\Delta\omega}^{\pi/\Delta\omega} f(t) e^{-j\omega t}\, dt \right] e^{j\omega t}\, \Delta\omega \tag{8-110}$$

In the limit, as $\Delta\omega \to 0$, the summation becomes a process of integration, and

$$f(t) = \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \right] e^{j\omega t}\, d\omega \tag{8-111}$$

This equation is one form of what is known as a *Fourier integral*. It may be written in a slightly different form by calling the bracket term in the equation $g(\omega)$, as

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \tag{8-112}$$

so that

$$f(t) = \int_{-\infty}^{\infty} g(\omega) e^{j\omega t}\, d\omega \tag{8-113}$$

and these two equations constitute a *Fourier transform pair*. These equations can be used to represent $f(t)$ provided $f(t)$ satisfies the Dirichlet conditions, mentioned in Art. 8-5, and if the integral

$$\int_{-\infty}^{\infty} |f(t)|\, dt \tag{8-114}$$

is finite.

Now $|c_n|$ determined the frequency spectrum in the case of the Fourier series. The term corresponding to c_n in the Fourier integral expression is $g(\omega)\, d\omega$. The amplitude of $g(\omega)\, d\omega$ is vanishingly small, of course, since $d\omega$ is an infinitesimal quantity. However, the function $g(\omega)$ is finite and is plotted in magnitude and phase as the *frequency spectrum* corresponding to an aperiodic $f(t)$. No longer is this spectrum given for only discrete values of frequency. The function $g(\omega)$ is a continuous function for all ω. Because of this difference in spectra, $|g(\omega)|$ is sometimes called a *continuous spectrum*, while $|c_n|$ is a *line spectrum*. In terms of the synthesis of a pulse by addition of frequency components, the continuous spectrum requires *all* frequencies combined as required by Eq. 8-113.

For the single pulse shown in Fig. 8-29, the other two having moved to infinity in opposite directions, the frequency spectrum may be found from Eq. 8-112 as

$$g(\omega) = \frac{1}{2\pi} \int_{-a/2}^{a/2} Ve^{j\omega t}\, dt = \frac{Va}{2\pi} \frac{\sin(\omega a/2)}{(\omega a/2)} \tag{8-115}$$

This equation is similar in form to Eq. 8-107 given for recurring pulses, but is a continuous rather than a line spectrum. The plots of the mathematical expression $|(\sin x)/x|$ shown in Fig. 8-30 constitute the *envelope* of the spectrum $|g(\omega)|$ for the single pulse.

The Fourier transform pair of equations, Eqs. 8-112 and 8-113, serves to illustrate the relationship of the time-domain function $f(t)$ and the frequency-domain quantity, the frequency spectrum, $g(\omega)$. For a given $f(t)$, we can find the corresponding $g(\omega)$. And for a given $g(\omega)$, we can similarly find the corresponding $f(t)$. The Fourier transform equations provide us with a two-way street with which we can go from time domain to frequency domain or vice versa. The same two-way street exists for the Fourier series and the associated concept of the line spectrum and the time domain.

We know that the waveform of a single pulse can be synthesized from frequency components specified by Eq. 8-115. Suppose that we apply a single pulse to the input of a system which does not transmit the higher-frequency components. The output of this system will no longer be a square pulse but will be some other time-domain function that could be computed from Eq. 8-113 using the $g(\omega)$ of the output of the frequency-selective system. As we change frequency response, we

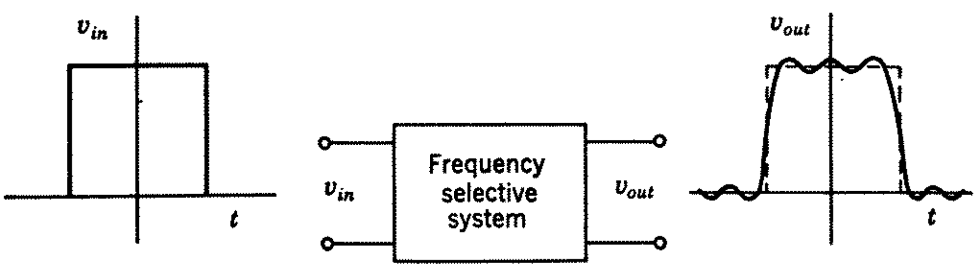


Fig. 8-31. Input and output waveforms for a two-terminal-pair frequency-selective network.

also change time response. An input and the corresponding output for a frequency sensitive system are illustrated in Fig. 8-31.

8-9. Fourier transforms and their relationship to Laplace transforms

The Fourier transforms, defined by Eqs. 8-112 and 8-113, have been used in the last section to illustrate the relationship of frequency domain and time domain concepts. We have not yet discussed the transform property of these equations which can be used in solving circuit equations in much the same manner as the Laplace transformation has been used. For reference, Eqs. 8-112 and 8-113 are repeated here.

$$g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t) e^{-j\omega t}\, dt \tag{8-116}$$

$$f(t) = \int_{-\infty}^{\infty} g(\omega) e^{j\omega t}\, d\omega \tag{8-117}$$

The transform character of these equations is emphasized by the use of the following notation.

$$\mathfrak{F} f(t) = g(\omega) \tag{8-118}$$

$$\mathfrak{F}^{-1} g(\omega) = f(t) \tag{8-119}$$

Equations 8-116 and 8-118 define the *direct Fourier transformation,* while Eqs. 8-117 and 8-119 define the *inverse Fourier transformation.* These four equations are similar in appearance to the corresponding equations for the Laplace transformation, given in Chapter 7 as Eqs. 7-1 and 7-3, which are

$$F(s) = \int_{0}^{\infty} f(t) e^{-st}\, dt \tag{8-120}$$

$$f(t) = \frac{1}{2\pi j} \int_{\sigma - j\infty}^{\sigma + j\infty} F(s) e^{st}\, ds \tag{8-121}$$

and

$$\mathcal{L} f(t) = F(s) \tag{8-122}$$

$$\mathcal{L}^{-1} F(s) = f(t) \tag{8-123}$$

Comparison of these two sets of equations reveals several differences: (1) The $j\omega$ in the Fourier transform occupies the same position as s in the Laplace transform. (2) The letters F and g signify functions with similar roles. (3) The limits of integration in Eqs. 8-116 and 8-120 are different, $-\infty$ in the Fourier transform corresponding to 0 in the Laplace transform. (4) The multiplying constants $(1/2\pi)$ and $(1/2\pi j)$ occupy different positions although this is a matter of convention since $1/2\pi$ may be associated with either $f(t)$ or $g(\omega)$ in Eq. 8-111. Since we are stressing the similarities of the two systems of transforms, the factor $1/2\pi$ will be shifted from Eq. 8-116 for $g(\omega)$ to Eq. 8-117 for $f(t)$ for the remainder of this discussion.

To illustrate the consequences of these differences, we will study an example of the computation of a Fourier transform. In most circuits studied in past chapters, interest has centered on what happens in a circuit after an instant of time corresponding to the opening or closing of a switch. This instant of time is conveniently taken as the reference time, $t = 0$. A function that has been used to denote the closing of a switch to connect a driving force to a circuit is the unit step function $u(t)$. To determine the Fourier transform of $u(t)$, we make use of Eq. 8-116. In this case, the lower limit of the integral may be changed to zero, since $u(t)$ has zero value for all negative t. In this usual circuit situation in which $f(t)$ has zero value before $t = 0$, Eq. 8-116 may be written with the lower limit of zero, and is then known as the *unilateral Fourier transformation*. Carrying out the operations we have just described, we have, with $1/2\pi$ removed from Eq. 8-116

$$g(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt = \int_0^{\infty} e^{-j\omega t}\,dt \tag{8-124}$$

$$g(\omega) = \frac{-1}{j\omega} e^{-j\omega t}\Big|_0^{\infty} = \frac{-1}{j\omega}(\cos \omega t - j \sin \omega t)\Big|_0^{\infty} \tag{8-125}$$

This equation has no meaning, since neither the sine nor the cosine is defined for infinite ωt. This difficulty can be avoided by introducing a convergence factor defined in the equation

$$f_1(t) = e^{-\sigma t}f(t) \tag{8-126}$$

where $f_1(t)$ is a modified function and σ is real and positive. This procedure provides the convergence necessary to avoid the difficulty in evaluating Eq. 8-125, and permits computation of the Fourier transform as the limit as $\sigma \to 0$. Substituting $f_1(t) = e^{-\sigma t}u(t)$ into Eq. 8-116 without the factor $1/2\pi$ gives

$$g(\omega) = \lim_{\sigma \to 0} \int_{-\infty}^{\infty} f_1(t)e^{-j\omega t}\,dt = \lim_{\sigma \to 0} \int_0^{\infty} e^{-\sigma t}e^{-j\omega t}\,dt \tag{8-127}$$

and

$$g(\omega) = \lim_{\sigma \to 0} \int_0^{\infty} e^{-(\sigma + j\omega)t}\,dt = \lim_{\sigma \to 0} \frac{1}{\sigma + j\omega} = \frac{1}{j\omega} \tag{8-128}$$

From this equation, we have the Fourier transform of $u(t)$ with $\sigma = 0$ as

$$\mathfrak{F}u(t) = \frac{1}{j\omega} \tag{8-129}$$

Thus the convergence difficulty has been avoided effectively by introducing a real part to be added to $j\omega$. Since s of the Laplace transformation is defined as a complex number, $s = \sigma + j\omega$, we see that the

Laplace transformation automatically has the advantage of stronger convergence by incorporating a "built-in" convergence factor. However, the Laplace transform of $f(t)$ is identical with the Fourier transform of $f(t)$ multiplied by the convergence factor $e^{-\sigma t}$; that is,

$$\mathcal{L}f(t) = \mathfrak{F}[f(t)e^{-\sigma t}] \tag{8-130}$$

Recognition of this relationship of transforms unifies two important topics in the study of electric circuits.

The preceding discussion might be regarded as a heuristic derivation of the Laplace transformation from the Fourier transformation. Since the Fourier transform conveys more physical meaning than the Laplace transform, arising as it does out of the Fourier integral and Fourier series, this tie-in is conceptually important. Aside from this advantage, the Laplace transformation is a more powerful mathematical tool than the Fourier transformation and is more extensively used.

Tables of Fourier transforms are available, an example being the extensive compilation of Campbell and Foster.*

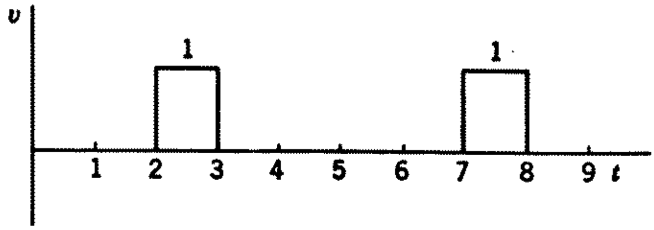
FURTHER READING

For further reading on the subject of the response of a system to such excitations as the step function, impulse, ramp function, square wave, etc., see Thomson, *Laplace Transformation* (Prentice-Hall, Inc., New York, 1950), pp. 23–26. The convolution integral is also discussed by Thomson, pp. 37–38, and in such references as Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 228–241; Salvadori and Schwarz, *Differential Equations in Engineering Problems* (Prentice-Hall, Inc., New York, 1954), pp. 214–219; and Wylie, *Advanced Engineering Mathematics* (McGraw-Hill Book Co., Inc., New York, 1951), pp. 188–197. For further reading on Fourier series, see Kerchner and Corcoran, *Alternating-Current Circuits* (John Wiley & Sons, Inc., New York, 1951), Chap. 6. A very complete discussion of the Fourier series and integral is given by Guillemin, *The Mathematics of Circuit Analysis* (John Wiley & Sons, Inc., New York, 1949), Chap. 7. Chapter 5 of Wylie (*op. cit.*) is very concise on these subjects and is especially recommended. Two additional references containing valuable information on the Fourier integral and frequency spectra are Fich, *Transient Analysis in Electrical Engineering* (Prentice-Hall, Inc., New York, 1951), pp. 199–214; and LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 444–463.

* Campbell and Foster, *Fourier Integrals for Practical Applications* (D. Van Nostrand Company, Inc., New York, 1950).

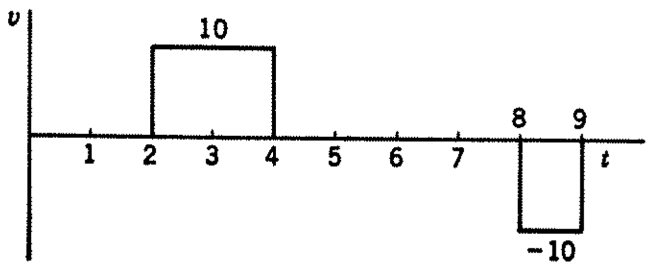
PROBLEMS

8-1. Write an equation for the nonrecurring waveform shown in the figure in terms of unit step functions.



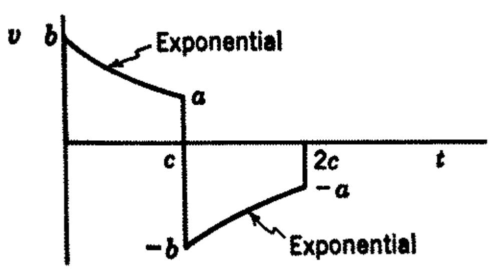
Prob. 8-1.

8-2. Write an equation for the nonrecurring waveform shown in terms of unit step functions.

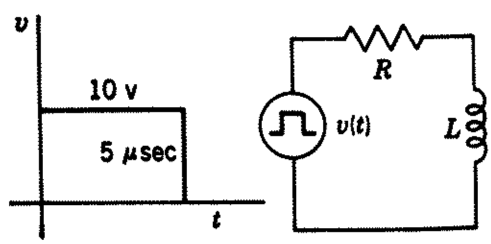


Prob. 8-2.

8-3. In the nonrecurring waveform shown, the function suddenly increases to a value b at the time $t = 0$ and then decreases exponentially to a value a at $t = c$ before decreasing suddenly to zero. The waveform then goes through the same cycle with negative magnitudes. Write an expression for this waveform, using unit step functions. *Partial answer.* $v = be^{-[\ln(a/b)]t/c}u(t) - \ldots.$



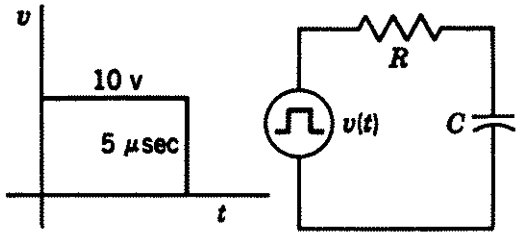
Prob. 8-3.



Prob. 8-4.

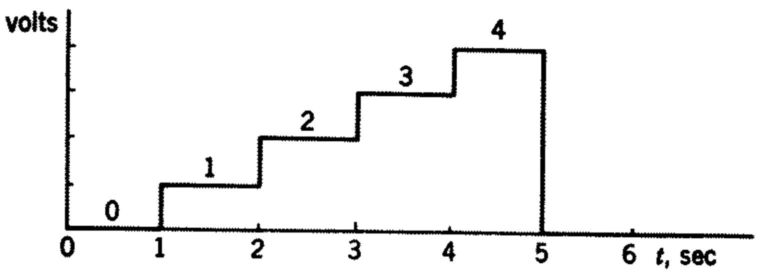
8-4. A voltage pulse of 10 volts magnitude and of 5 μsec duration is applied to an RL series circuit where $R = 2$ ohms and $L = 10$ μhenry. Plot the waveform of the current as a function of time.

8-5. A voltage pulse of 10 volts magnitude and of 5 μsec duration is applied to an RC series circuit where $R = 100$ ohms and $C = 0.05$ μf. Find the equation for the current and plot the current waveform as a function of time.



Prob. 8-5.

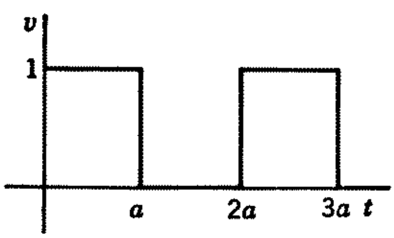
8-6. A voltage waveform known as a "staircase" is used to shift the frequency of a radio transmitter. One cycle of staircase is shown in the figure. (a) Write the equation for this voltage waveform $v(t)$, assuming it is not repeated. (b) Suppose that this voltage is applied to a series RL circuit with $R = 1$ ohm and $L = 1$ henry. Sketch the current waveform approximately to scale on the same coordinates as the "staircase" voltage.



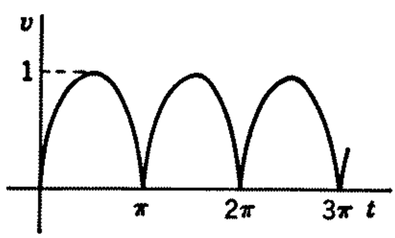
Prob. 8-6.

8-7. Show that the transform of the square wave is

$$F(s) = \frac{1}{s(1 + e^{-as})}$$



Prob. 8-7.



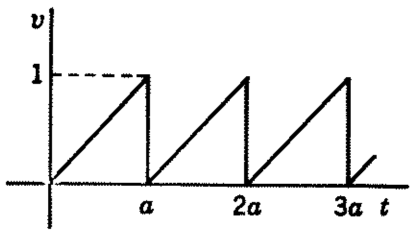
Prob. 8-8.

8-8. The waveform shown in the figure is that of a full-wave rectified voltage. The equation for the waveform is sin t from 0 to π, $-\sin t$ from π to 2π, etc. Show that the transform of this function is

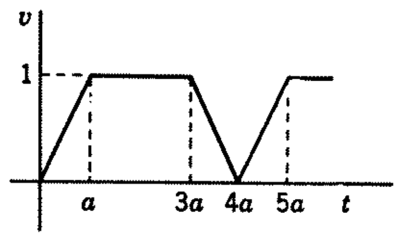
$$F(s) = \frac{1}{s^2 + 1} \coth \frac{\pi s}{2}$$

8-9. The waveform shown is a sweep voltage used to deflect the beam in a cathode ray oscilloscope. Show that the transform of this function is

$$F(s) = \frac{1}{as^2} - \frac{e^{-as}}{s(1 - e^{-as})}$$



Prob. 8-9.



Prob. 8-10.

8-10. Find the transform of the voltage waveform shown in the figure.

8-11. By convolution, find the time functions corresponding to the following transform functions starting with the transform pair $f(t) = e^{at}$, $F(s) = 1/(s - a)$.

(a) $\dfrac{1}{(s-a)^2}$ (c) $\dfrac{1}{(s+a)(s+b)}$

(b) $\dfrac{1}{(s-a)(s-b)}$ (d) $\dfrac{1}{(s-a)(s-b)(s-c)}$

8-12. By convolution, find the inverse Laplace transformation of the following functions.

(a) $\dfrac{1}{(s^2+1)^2}$ (b) $\dfrac{s}{(s+1)(s^2+1)}$

8-13. Tests on a certain network showed that the current output was $i(t) = -2e^{-t} + 4e^{-3t}$ when a unit voltage was suddenly applied to the input terminals at $t = 0$. What voltage must be applied to give an output current of $i(t) = 2e^{-t}$ if the network remains in the same form as for the previous test? *Answer.* $v = 4e^{t} - 3$.

8-14. The output of a half-wave rectifier is given by the equation

$$v(\omega t) = \begin{cases} \cos \omega t, & 0 \leqq \omega t < \dfrac{\pi}{2} \\ 0, & \dfrac{\pi}{2} \leqq \omega t \leqq \dfrac{3\pi}{2} \\ \cos \omega t, & \dfrac{3\pi}{2} \leqq \omega t \leqq 2\pi \end{cases}$$

Show that this periodic waveform can be represented by the Fourier series

$$v(\omega t) = \frac{1}{\pi}\left(1 + \frac{\pi}{2}\cos\omega t + \frac{2}{3}\cos 2\omega t - \frac{2}{15}\cos 4\omega t + \ldots\right)$$

8-15. Find the Fourier series representation of the trapazoidal waveform shown for Prob. 8-10. Draw the line spectrum for this waveform.

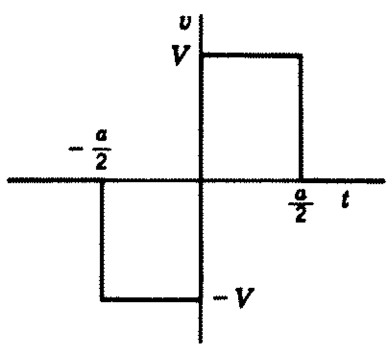
8-16. Draw the line spectrum for the waveform of Prob. 8-14.

8-17. The following table gives the ordinates of a waveform as a function of ωt. The values for π to 2π are defined by the relationship $f(\pi + \omega t) = -f(\omega t)$; in other words, the negative loop from π to 2π is similar to the positive loop from 0 to π.

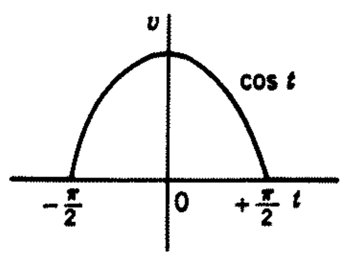
ωt	$f(\omega t)$	ωt	$f(\omega t)$
0	0	105°	85.0
15°	49.7	120	77.9
30°	75	135	77.8
45°	77.8	150	75
60°	77.9	165	49.7
75°	85.0	180	0
90°	90		

(a) Determine the Fourier coefficients for the first five harmonics. (b) Draw the line spectra for this waveform.

8-18. For the waveform shown in the figure, determine the continuous spectrum and sketch $|g(\omega)|$ and Ang $g(\omega)$.



Prob. 8-18.



Prob. 8-19.

8-19. The aperiodic function shown in the figure is part of a cosine wave defined only from $-\pi/2$ to $+\pi/2$. Determine the continuous spectrum and sketch $|g(\omega)|$ and Ang $g(\omega)$.

8-20. An aperiodic function is defined by the equation

$$v(t) = Ve^{-\alpha t}\sin\omega_0 t, \qquad t \geqq 0$$

and represents a damped oscillation. Determine the continuous spectrum for this function and sketch both $|g(\omega)|$ and Ang $g(\omega)$.

CHAPTER 9

IMPEDANCE AND ADMITTANCE FUNCTIONS

In this chapter, the operational method studied by the Laplace transformation will be used to introduce the concepts of impedance and admittance.

9-1. The concept of complex frequency

The solution of the differential equations for networks has given rise to time-domain functions of the form

$$K_n e^{s_n t} \tag{9-1}$$

where s_n is a complex number, a root of the characteristic equation, expressed as

$$s_n = \sigma_n + j\omega_n \tag{9-2}$$

Here ω_n, the imaginary part of s_n, has been interpreted as *radian frequency* (or angular frequency) and it appears in time-domain equations in the forms

$$\sin \omega_n t \quad \text{or} \quad \cos \omega_n t \tag{9-3}$$

Radian frequency has the dimensions of radians per second and may be expressed in terms of frequency, f_n, in cycles per second, or in terms of the period T, in seconds, by the equation

$$\omega_n = 2\pi f_n = \frac{2\pi}{T} \tag{9-4}$$

By Eq. 9-2 we see that σ_n and ω_n must be identical in dimensions. The dimension of ω_n is $(\text{time})^{-1}$, since the radian is a dimensionless quantity (being *length* of arc per *length* of radius). The dimension of σ_n must be "something" per unit time. Since σ_n appears as an exponential factor,

$$I = I_0 e^{\sigma_n t} \tag{9-5}$$

such that

$$\sigma_n = \frac{1}{t} \ln \frac{I}{I_0} \tag{9-6}$$

it is evident that the "something" per unit time should be a nondimensional logarithmic unit. The usual unit for the natural (or Naperian) logarithm is the *neper*. This unit is commonly used making the dimension for σ the *neper per second.*

The complex sum

$$s_n = \sigma_n + j\omega_n \tag{9-7}$$

is defined as the *complex frequency*. The imaginary part of the complex frequency is the *radian frequency* (or real frequency), and the real part of complex frequency is *neper frequency** (rather than the misleading term "imaginary frequency"). The physical interpretation of complex frequency appearing in the function $e^{s_n t}$ will be studied by considering a number of special cases for the value of s_n.

(1) Let $s_n = \sigma_n + j0$ and let σ_n have positive, zero, and negative values. The exponential function of Eq. 9-1 becomes $K_n e^{\sigma_n t}$, an exponential function which increases exponentially for $\sigma_n > 0$ and decreases (or decays) exponentially for $\sigma_n < 0$. When $\sigma_n = 0$, so that $s_n = 0 + j0$, the term becomes

$$K_n e^{s_n t} = K_n e^{0t} = K_n \tag{9-8}$$

a time-invariant quantity which in terms of current and voltage is described as "direct current." The time variation for the three possibilities for real s_n are shown in Fig. 9-1.

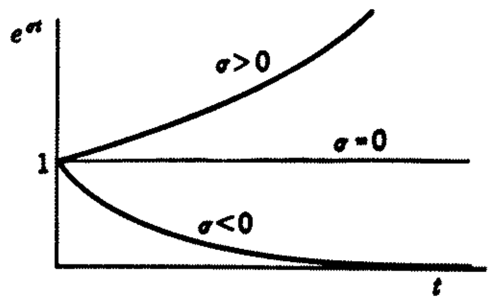


Fig. 9-1. Variation of $e^{\sigma t}$ with time (σ = neper frequency).

(2) Let $s_n = 0 \pm j\omega_n$ (radian frequency only). In this case, the exponential factor becomes

$$K_n e^{\pm j\omega_n t} = K_n(\cos \omega_n t \pm j \sin \omega_n t) \tag{9-9}$$

by Euler's equation. The exponential $e^{\pm j\omega_n t}$ is usually interpreted in terms of the physical model (with no actual physical significance) of a unit rotating *phasor*,† the direction of rotation being determined by the sign. A positive sign, $e^{+j\omega_n t}$ implies counterclockwise (or positive) rotation, while a negative sign, $e^{-j\omega_n t}$ implies clockwise (or negative) rotation. For positive rotation, the real part of $e^{j\omega_n t}$ (or the projection on the real axis) varies as the cosine of $\omega_n t$, while the imaginary part (or projection on the imaginary axis) varies as the sine of $\omega_n t$. This concept is illustrated by Fig. 9-2. The variation of the exponential function with time is sinusoidal and corresponds to the case of the sinusoidal steady state.

* The terms *radian frequency* and *neper frequency* were used by W. H. Huggins, "The potential analog in network synthesis and analysis," Air Force Cambridge Research Laboratories, Report No. E5066, March 1951.

† Many texts used the word *vector* in place of phasor. A phasor is characterized by magnitude and phase with respect to a reference.

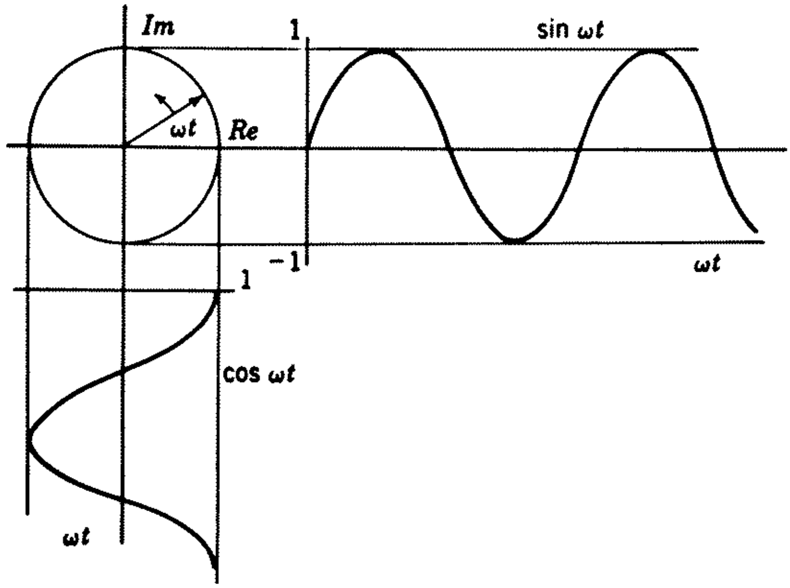


Fig. 9-2. Rotating phasor and imaginary and real axis projections (sine and cosine).

(3) Let $s_n = \sigma_n + j\omega_n$ (this is the general case and the frequency is complex). For this case,

$$K_n e^{s_n t} = K_n e^{(\sigma_n + j\omega_n)t} = K_n e^{\sigma_n t} e^{j\omega_n t} \tag{9-10}$$

This expression shows that such a term has a time variation which is the product of the result for $s_n = \sigma_n$ and for $s_n = \pm j\omega_n$. One term is represented by the rotating phasor model, the other term by an exponentially increasing or decreasing function. This result can be thought of as a rotating phasor with a magnitude which changes with time. Such a phasor is illustrated in Fig. 9-3. The real and imaginary projections of this phasor are

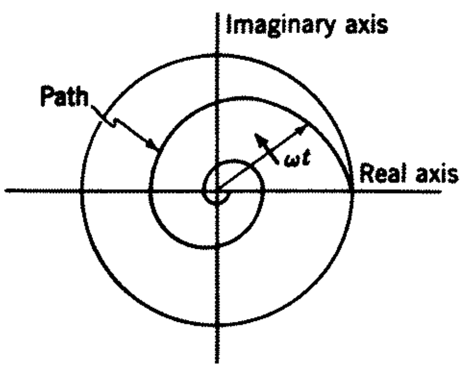


Fig. 9-3. Rotating phasor decreasing in magnitude with time.

$$\mathrm{Re}(e^{s_n t}) = e^{-\sigma_n t} \cos \omega_n t \tag{9-11}$$

and

$$\mathrm{Im}(e^{s_n t}) = e^{-\sigma_n t} \sin \omega_n t \tag{9-12}$$

for a phasor rotating in the positive direction and negative σ. These projections are shown in Fig. 9-4. Such waveforms have been classified as *damped sinusoids*.

From this discussion, we see that there is nothing really new in the concept of complex frequency. The imaginary part of complex frequency, the radian (or real) frequency corresponds to oscillations. The real part of complex frequency, neper frequency, corresponds to exponential decay or exponential increase (depending on sign) or to no variation for zero neper frequency. We have talked about such exponential functions before in terms of the time constant. Since the role of the two "kinds" of frequency is the same, even though the consequences are different, we unify the two concepts under one name—complex frequency.

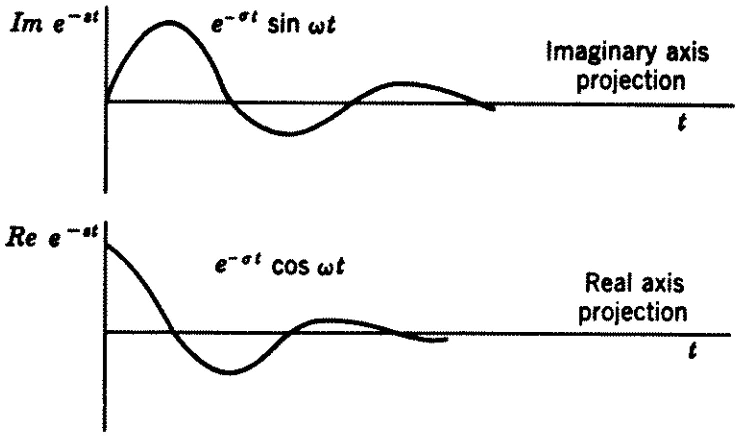


Fig. 9-4. Time variation of e^{-st} where s is complex frequency.

We should constantly guard against semantic difficulties in the use of the work "imaginary" as one part of a complex quantity. The imaginary part of a quantity is *not* physically imaginary (that is invisible or ghostlike) in the sense that it is not physically real. We have borrowed the words "real" and "imaginary" from the mathematicians as designations of two distinct parts of a quantity or function (which we often reinterpret in terms of magnitude and phase). The mathematician's "imaginary" carries no connotation about the physical universe about us!

9-2. Transform impedance and admittance

The ratio of the transform of a voltage to the transform of a current is defined as the *transform* (or generalized) *impedance*. The reciprocal ratio is defined as the *transform* (or generalized) *admittance*. We will next determine an expression for the impedance and admittance for each of the network parameters.

Resistance. The time-domain expression for the voltage across a resistor is given by Ohm's law in the forms

$$v(t) = Ri(t) \quad \text{or} \quad i(t) = Gv(t) \tag{9-13}$$

The corresponding transform equations are

$$V(s) = RI(s) \quad \text{or} \quad I(s) = GV(s) \tag{9-14}$$

Following the definitions given above for the transform impedance and transform admittance, we have

$$\frac{V(s)}{I(s)} = Z(s) = R \tag{9-15}$$

where $Z(s)$ is the transform impedance, and

$$\frac{I(s)}{V(s)} = Y(s) = G \tag{9-16}$$

where $Y(s)$ is the transform admittance.

The schematic which shows the actual resistor and the time-domain voltage and current can be replaced by a diagram to represent equivalent transform quantities. Two such diagrams are shown in Fig. 9-5.

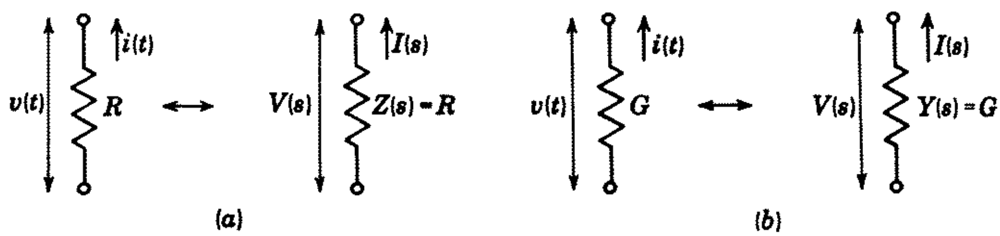


Fig. 9-5. Resistor impedance and admittance.

The time-domain schematic is a representation of the actual physical system. The transform diagram is composed of time-domain element representations, but the letter symbol for the actual element is replaced by an impedance or admittance symbol.

Inductance. The time-domain relationship between voltage and current in an inductor is expressed by the following equations.

$$v(t) = L\frac{di(t)}{dt} \quad \text{and} \quad i(t) = \frac{1}{L}\int v(t)\,dt \tag{9-17}$$

The equivalent transform equation for the voltage expression is

$$V(s) = L[sI(s) - i(0+)] \tag{9-18}$$

Regrouping the terms, we have

$$LsI(s) = V(s) + Li(0+) \tag{9-19}$$

In this expression, $V(s)$ is the transform of the applied voltage, and $Li(0+)$ is a transform voltage resulting from the initial current in the inductor. Designating the transform voltage across $Z(s)$ as $V_1(s)$, where $V_1(s) = V(s) + Li(0+)$, the transform impedance becomes

$$\frac{V_1(s)}{I(s)} = Z(s) = Ls$$

The equivalent transform diagram thus contains a transform impedance and a voltage source due to the initial current. This equivalence to the time-domain schematic is shown in Fig. 9-6.

The transform equation for the current is

$$I(s) = \frac{1}{L}\left[\frac{V(s)}{s} + \frac{v^{-1}(0+)}{s}\right] \tag{9-20}$$

The initial-value integral $v^{-1}(0+)$ can be evaluated in terms of flux linkages Li as

$$v^{-1}(0+) = \int v(t)\,dt\Big|_{t=0+} = Li(0+) \tag{9-21}$$

The equation for $I(s)$ may be rewritten

$$I(s) = \frac{1}{L}\frac{V(s)}{s} + \frac{i(0+)}{s} \tag{9-22}$$

or

$$\frac{1}{Ls}V(s) = I(s) - \frac{i(0+)}{s} \tag{9-23}$$

In this equation, $i(0+)/s$ is an equivalent transform current source resulting from the initial current in the inductor. Designating the transform current in $Y(s)$ as $I_1(s) = I(s) - i(0+)/s$, the transform admittance becomes

$$\frac{I_1(s)}{V(s)} = Y(s) = \frac{1}{Ls} \tag{9-24}$$

The equivalent transform diagram thus contains an admittance of value $1/Ls$ and an equivalent current source defined in Eq. 9-23. This equivalent schematic for the time domain diagram is shown in Fig. 9-6. We note that,

$$Z(s) = \frac{1}{Y(s)} = Ls \tag{9-25}$$

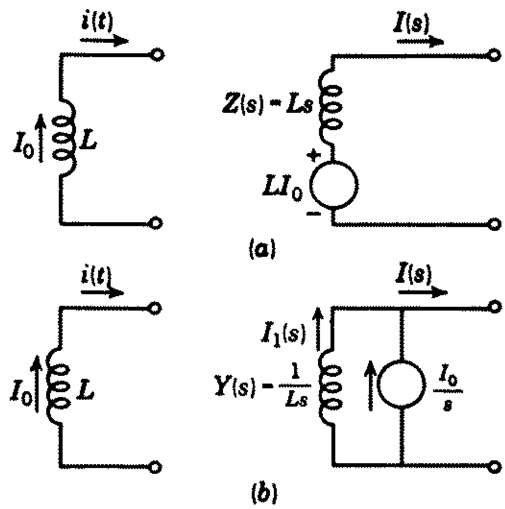


Fig. 9-6. (a) Impedance diagram for L, (b) admittance diagram for L.

Capacitance. The time-domain relationship between voltage and current for a capacitor is given as

$$i(t) = C\frac{dv(t)}{dt} \quad \text{and} \quad v(t) = \frac{1}{C}\int i(t)\,dt \tag{9-26}$$

The equivalent transform equation for the voltage expression is

$$V(s) = \frac{1}{C}\left[\frac{I(s)}{s} + \frac{q(0+)}{s}\right] \tag{9-27}$$

where $q(0+)/C$ is the initial voltage of the capacitor which, due to the charge polarity, is $-V_0$. This equation may be written

$$\frac{1}{Cs} I(s) = V(s) + \frac{V_0}{s} \tag{9-28}$$

Designating the transform voltage of $Z(s)$ as $V_1(s) = V(s) + V_0/s$, the ratio of the transform voltage to the transform current is

$$\frac{V_1(s)}{I(s)} = Z(s) = \frac{1}{Cs} \tag{9-29}$$

The capacitor with an initial charge thus has an equivalent transform diagram with an impedance $1/Cs$ in series with a voltage source having a transform $-v(0+)/s$. The schematic of this combination is shown in Fig. 9-7.

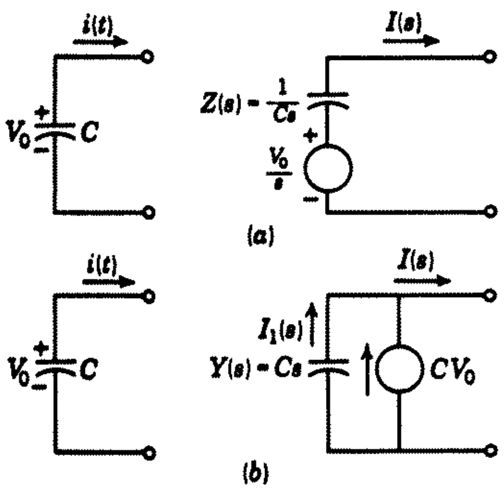


Fig. 9-7. (a) Impedance diagram for C, (b) admittance diagram for C.

The transform equation for the current expression of Eq. 9-26 is

$$I(s) = C[sV(s) - v(0+)] \tag{9-30}$$

or
$$CsV(s) = I(s) - CV_0 \tag{9-31}$$

Designating the transform current in $Y(s)$ as $I_1(s) = I(s) - CV_0$, the ratio of transform current to transform voltage becomes

$$\frac{I_1(s)}{V(s)} = Y(s) = Cs \tag{9-32}$$

The capacitor with an initial charge has an equivalent transform schematic representation of an admittance of value Cs in parallel with a current source of value CV_0. This schematic is shown in Fig. 9-7(b). For the capacitor,

$$Z(s) = \frac{1}{Y(s)} = \frac{1}{Cs} \tag{9-33}$$

9-3. Series and parallel combinations

In this section, we will consider the impedance and admittance of series, parallel and series-parallel combinations of different elements. To simplify schematic diagrams, we will use the symbol normally reserved for the resistor *together* with the letters $Z(s)$ or $Y(s)$ to designate a *transform impedance or admittance.* We will not consistently use the broken zigzag line for a transform impedance symbol, but either this symbol or the actual element symbol depending on which is most convenient and descriptive. Consider the series combination of elements shown in Fig. 9-8(a) and the equivalent transform impedance diagram shown in Fig. 9-8(b). In Fig. 9-8(a), the same current $i(t)$

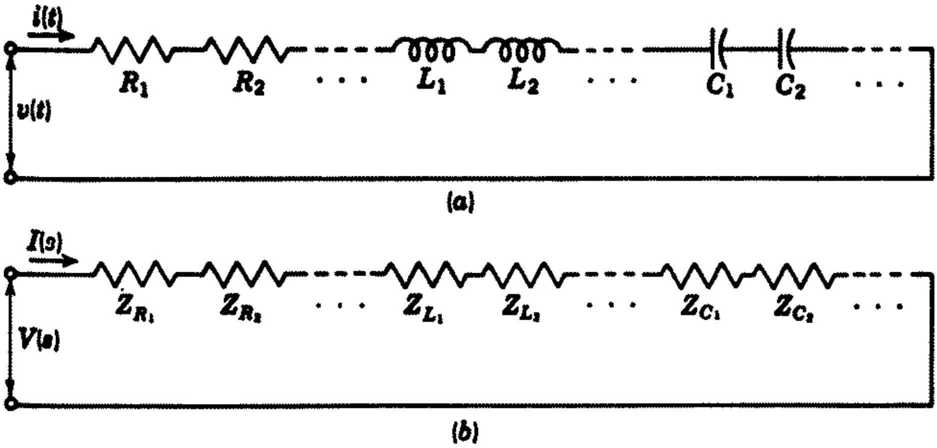


Fig. 9-8. Impedance of series networks.

flows in all elements, and so in Fig. 9-8(b), $I(s)$ is common to all elements. By Kirchhoff's voltage law, the sum of the drops of voltage for all elements is equal to $v(t)$. Hence the transform of all voltages of the elements sum to $V(s)$; that is,

$$V(s) = V_{R_1}(s) + \ldots + V_{L_1}(s) + \ldots + V_{C_1}(s) + \ldots \tag{9-34}$$

Dividing this equation by $I(s)$ and recognizing that the ratio of the voltage of each element divided by the current for that element is impedance, we have

$$Z(s) = Z_{R_1}(s) + \ldots + Z_{L_1}(s) + \ldots + Z_{C_1}(s) + \ldots \tag{9-35}$$

or

$$Z(s) = \sum_{k=1}^{n} Z_k(s) \tag{9-36}$$

for a series combination of elements, where n is the total number of elements in series.

It should be recognized that in performing such a summation, the elements are *not* being combined. Rather, only a characteristic feature of the element (its impedance) is being summed and added to a characteristic of another element.

Consider next the parallel combination of elements shown in Fig. 9-9(a) and in equivalent transform form in Fig. 9-9(b). In this network, the voltage drop $v(t)$ is the same across all elements, and so $V(s)$ is the same for all elements. From Kirchhoff's current law, the sum of

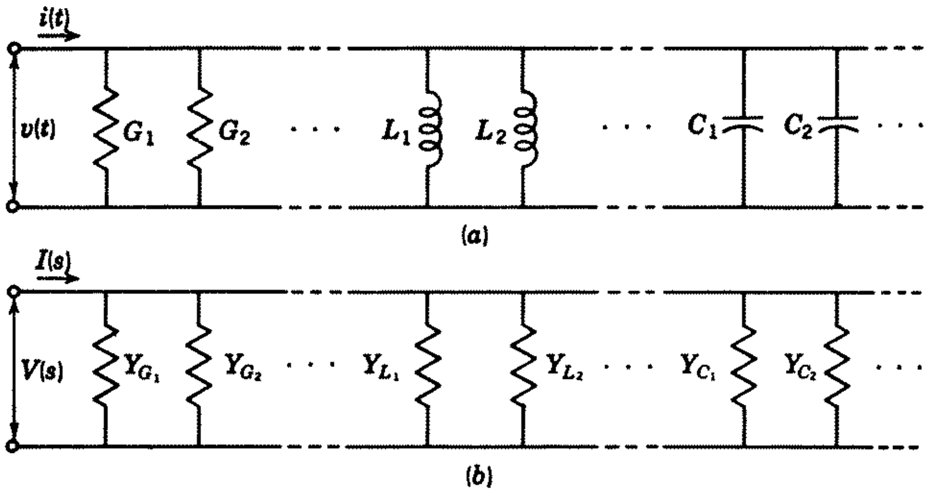


Fig. 9-9. Admittance of parallel networks.

the currents in the elements is equal to the total current supplied to the network; that is,

$$i(t) = i_{G_1}(t) + \ldots + i_{L_1}(t) + \ldots + i_{C_1}(t) + \ldots \tag{9-37}$$

and the corresponding transform equation is

$$I(s) = I_{G_1}(s) + \ldots + I_{L_1}(s) + \ldots + I_{C_1}(s) + \ldots \tag{9-38}$$

If this equation is divided by $V(s)$ and it is recognized that the ratio of the current transform to the voltage transform is transform admittance, there results

$$Y(s) = Y_{G_1}(s) + \ldots + Y_{L_1}(s) + \ldots + Y_{C_1}(s) + \ldots \tag{9-39}$$

or

$$Y(s) = \sum_{k=1}^{n} Y_k(s) \tag{9-40}$$

for a parallel combination of elements, where n is the total number of all kinds of elements in parallel.

For a series-parallel network, rules for the combination of impedance and of admittance can be used successively to reduce a network to a single equivalent impedance or admittance. This procedure will be illustrated with a number of examples.

Example 1

In the series circuit shown in Fig. 9-10, the switch K is held in position a until such a time that a current I_0 flows in the inductor and the capacitor is charged to voltage V_0. At that instant, the switch is thrown to position b, connecting the circuit to a voltage source $v(t)$. The problem is to find $I(s)$ and so $i(t)$. An equivalent circuit diagram marked with transform impedances is shown in Fig. 9-11. The imped-

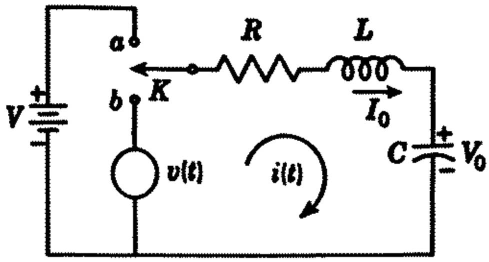


Fig. 9-10. *RLC* circuit.

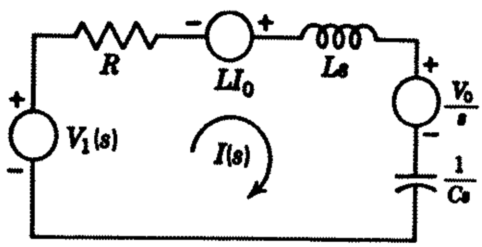


Fig. 9-11. Equivalent diagram for impedance of Fig. 9-10.

ance values and the equivalent voltage source values are taken from the derivations of this section given on pages 199 and 200. In this revised form, the current $I(s)$, a transform current, may be found by Ohm's law. The current $I(s)$ is given as the total transform voltage in the network divided by the total transform impedance. Then

$$I(s) = \frac{V(s)}{Z(s)} = \frac{V_1(s) + LI_0 - V_0/s}{R + Ls + 1/Cs} = \frac{sV_1(s) + LI_0s - V_0}{Ls^2 + Rs + 1/C} \quad (9\text{-}41)$$

This transform equation can be expanded by partial fractions to find the corresponding $i(t)$ by the inverse Laplace transformation. This solution has been found *without* writing the differential equation of the system, and automatically incorporates the required initial conditions.

Example 2

The dual of the network of Example 1 is shown in Fig. 9-12. In this network, the switch K_1 is opened at an instant when the inductor cur-

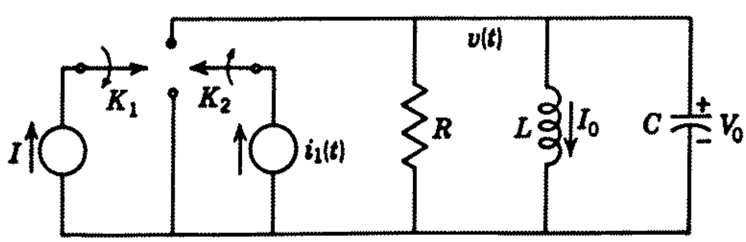


Fig. 9-12. Parallel *RLC* network.

rent is I_0 and the capacitor is charged to V_0. At the same instant, $t = 0$, the switch K_2 is closed. It is required to find the transform of the node voltage $V(s)$ so that $v(t)$ can be determined. From the equivalent admittance diagram shown in Fig. 9-13, the transform voltage

$V(s)$ is found as

$$V(s) = \frac{I(s)}{Y(s)} = \frac{I_1(s) + CV_0 - I_0/s}{Cs + G + 1/Ls} = \frac{sI_1(s) + CV_0 s - I_0}{Cs^2 + Gs + 1/L} \quad (9\text{-}42)$$

This transform is the dual of the transform of Eq. 9-41 (and could

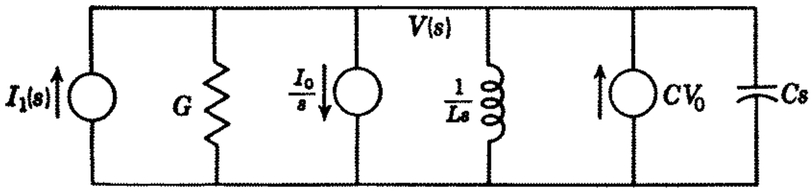


Fig. 9-13. Equivalent diagram for admittance of Fig. 9-12.

therefore have been written by inspection). The corresponding time-domain voltage, $v(t)$ can be found by taking the inverse Laplace transformation after the above transform has been expanded by partial fractions.

Example 3

In this example, we will make use of the laws for the series combination of impedance and the parallel combination of admittance to determine current. In the network shown in Fig. 9-14, it will be assumed

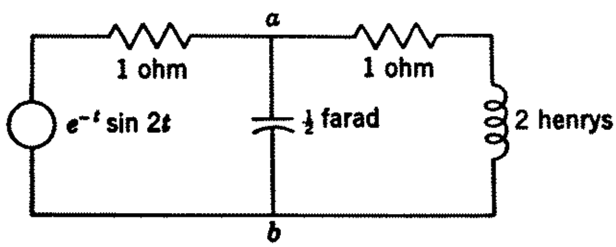


Fig. 9-14. Two-loop network.

that the network is initially relaxed (no current, no charge) and that the switch was closed at $t = 0$. It is required to find the current in the generator $i(t)$ by finding the transform of this current $I(s)$. The impedance of the branch containing the 1-ohm resistor and 2-henry inductor is

$$Z(s) = 1 + 2s \quad (9\text{-}43)$$

This impedance is in parallel with the impedance $2/s$ of the capacitor. The admittances may be added directly. Thus

$$Y_{ab}(s) = Y_C + Y_{RL} = \frac{s}{2} + \frac{1}{2s+1} = \frac{2s^2 + s + 2}{2(2s+1)} \quad (9\text{-}44)$$

The impedance from a to b is the reciprocal of the admittance; thus

$$Z_{ab}(s) = \frac{1}{Y_{ab}(s)} = \frac{2(2s+1)}{2s^2 + s + 2} \quad (9\text{-}45)$$

The total impedance is now found by adding to $Z_{ab}(s)$ the impedance of the 1-ohm resistor. Then the total impedance is

$$Z_{\text{total}}(s) = 1 + \frac{2(2s+1)}{2s^2+s+2} = \frac{2s^2+5s+4}{2s^2+s+2} \tag{9-46}$$

This total impedance in series with the transform of the voltage source

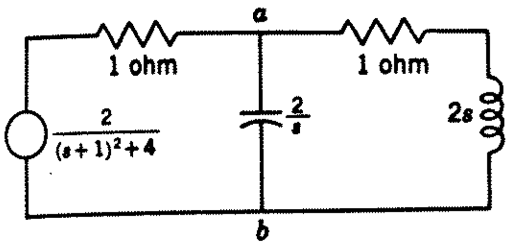


Fig. 9-15. Equivalent diagram for impedance of Fig. 9-14.

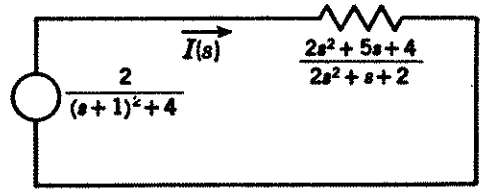


Fig. 9-16. Equivalent diagram of Fig. 9-15.

is shown in Fig. 9-16. The current may now be found by Ohm's law for transform quantities; that is,

$$I(s) = \frac{V(s)}{Z(s)} = \frac{2(2s^2+s+2)}{[(s+1)^2+4](2s^2+5s+4)} \tag{9-47}$$

If the inductor has a current flowing through it at $t = 0$ or if the capacitor is charged at $t = 0$, the problem is somewhat more complicated, since several voltage sources are involved.

9-4. Thévenin's theorem and Norton's theorem

When several voltage or current sources are present in a network, the net effect of all sources, as far as the current in one branch or the voltage at one node are concerned, may be taken into account by a theorem due to Thévenin* (and the dual of this theorem due to Norton).

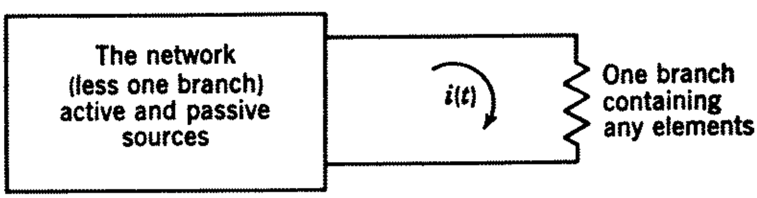


Fig. 9-17. Arbitrary network.

Suppose that we are interested in the current in one branch of a network. The single branch and the remainder of the network as a box are shown in Fig. 9-17. We will assume that the remainder of the

* This theorem was first proposed by M. L. Thévenin in the French scientific journal, *Comptes rendus*, in 1883. The dual of Thévenin's theorem is due to E. L. Norton of the Bell Telephone Laboratories.

network is arbitrarily complicated and that it contains an arbitrary number of voltage sources of arbitrary waveform as a function of time. If a generator is now inserted in the branch under consideration and is adjusted until the current in this branch *is equal to zero* at all time, that voltage source may be said to be an *equivalent voltage source* in the sense that it is identical to the net effect of all generators in the entire network as far as this *one branch* is concerned. With this equivalent generator connected in the circuit, no current flows in the branch being considered. If no current flows, the branch could be broken without affecting the network. The voltage across the network terminals with the branch and generator removed is the voltage of the equivalent generator except for polarity. Thus the equivalent generator voltage is the same as the voltage measured by removing the branch and considering the open circuit voltage but of opposite polarity.

Now if the equivalent voltage generator is placed in the loop being considered with polarity reversed, all active sources within the network could be removed by replacing them with short circuits, and the current in the branch will be the same as in the original network. The concept of an equivalent voltage source for a single branch, which is the basis of Thévenin's theorem, is illustrated in Fig. 9-18. This network is equivalent to that of Fig. 9-17 as far as the current in the one branch is concerned.

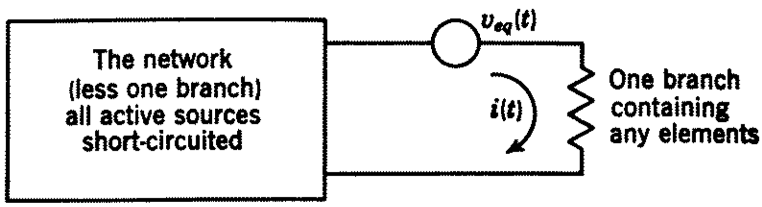


Fig. 9-18. Thévenin's equivalent network.

The statements made thus far apply to the time domain. The currents considered are time-domain currents, and the resulting voltages are time-domain. To convert to the frequency domain it is necessary only to find the Laplace transforms for all time-domain quantities involved.

To summarize our discussion, we can say that by Thévenin's theorem we have the equivalent network shown in Fig. 9-18 with one voltage source, one passive branch (although it is not necessary that it be passive) and with a network containing passive elements *only*. We may consider this network in terms of an equivalent transform voltage and by impedances of network elements. These impedances may be combined by the rules for series and parallel combination previously discussed. Finally, the entire passive network may be made equivalent to a single transform impedance. Let this impedance be $Z_{eq}(s)$ and let the impedance of the branch being considered be $Z_{br}(s)$. Then the current in this branch is given as

$$I(s) = \frac{V_{eq}(s)}{Z_{eq}(s) + Z_{br}(s)} \tag{9-48}$$

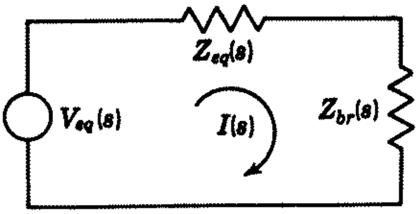


Fig. 9-19. Thévenin's equivalent transform network.

It must be recognized that this analysis applies only to a given branch. The equivalent circuit, shown in Fig. 9-19, does not hold for any branch other than the one under consideration. If another branch current is needed, it is necessary to start over and reapply the theorem. Thèvenin's theorem can be stated as follows:

> As far as the current in one branch is concerned, the remainder of the network may be replaced by an equivalent network having: (1) as a transform voltage source, the transform of that voltage appearing at the open-circuit terminals resulting from the removal of the branch, and (2) as series transform impedance, an equivalent impedance equal to that of the network from the terminals of the branch with all energy sources replaced by their internal impedances—zero impedance for voltage sources and infinite impedance for current sources.

The dual of Thévenin's theorem is Norton's theorem. Once more consider the network shown in Fig. 9-17. The current in the single branch being considered may be reduced to zero by placing a current source in parallel with the branch and adjusting the current until the voltage across the branch is zero. The voltage across the branch is zero because the current from the network is just balanced by an opposite current from the parallel current source. Since the voltage across

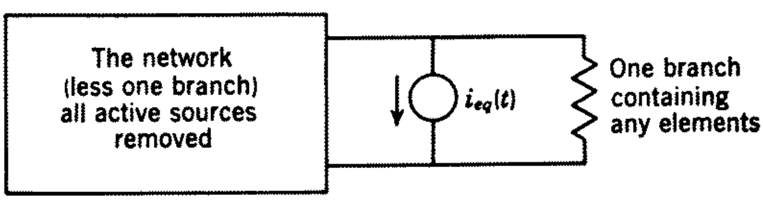


Fig. 9-20. Norton's equivalent network.

the branch is zero, a short circuit may be placed in parallel with the branch without affecting the network. The current in the short circuit will actually be zero, because there will be a current from the equivalent source which exactly cancels the current from the network. If all sources within the network are replaced by their internal impedance and the equivalent current source which caused the voltage across the branch to reduce to zero (but with opposite direction) is placed in parallel with the branch, then this network, shown in Fig. 9-20, is equivalent to the original network. This equivalent current source has the value and direction that is found by short-circuiting the branch in consideration and measuring the *current in the short circuit.* In summary, Norton's theorem can be stated as follows:

> As far as the voltage across any branch is concerned, the remainder of the network may be replaced by an equivalent network having: (1) as a transform current source, the transform of that current in a short circuit across the branch, and (2) as parallel transform admittance, an equivalent admittance equal to that of the network from the terminals of the branch with all energy sources replaced by their internal impedances—zero impedance for voltage sources and infinite impedance for current sources.

The equivalent network is shown in Fig. 9-21 as a node basis network with a current source and two parallel transform admittances. The unknown voltage is given as

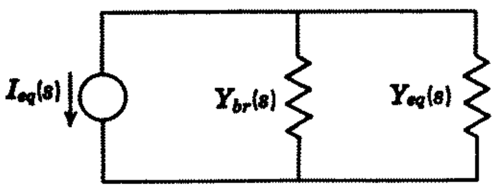


Fig. 9-21. Norton's equivalent transform network.

$$V(s) = \frac{I(s)}{Y(s)} = \frac{I_{eq}(s)}{Y_{br}(s) + Y_{eq}(s)} \tag{9-49}$$

These two theorems will allow us to reduce the form of any network to an equivalent simple series circuit, and from this circuit the transform of the current can be found. These operations will be illustrated by two examples.

Example 4

The network shown is unenergized until the instant $t = 0$, when the switch K is closed. It is required to find the current $i_1(t)$ flowing in

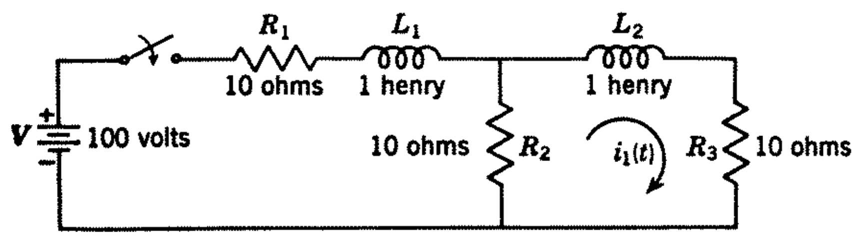


Fig. 9-22. Two-loop network.

the resistor R_3. The values given on the schematic have the units of the ohm, the henry, and the farad. An equivalent schematic showing element impedances is shown in Fig. 9-23. Thèvenin's theorem will be applied by disconnecting the branch containing L_2 and R_2. The network that remains is a simple series network, and the voltage across the 10-ohm impedance will be the open-circuit voltage for the Thévenin equivalent network. This voltage is found by finding the current in

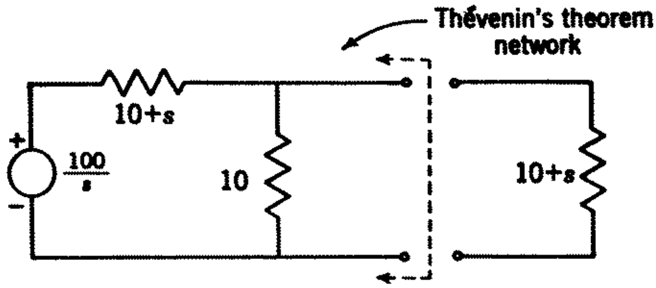


Fig. 9-23. Impedance schematic for Fig. 9-22.

the series network and multiplying this current by 10 ohms, the impedance of the resistor; thus

$$V_{oc}(s) = \frac{10(100/s)}{10 + s + 10} = \frac{1000}{s(s + 20)} \tag{9-50}$$

The impedance of the network with the voltage source $100/s$ short-circuited is

$$Z_{eq}(s) = \frac{10(s + 10)}{s + 20} \tag{9-51}$$

The current transform, by Eq. 9-48, is

$$I(s) = \frac{V_{oc}(s)}{Z_{eq}(s) + Z_{br}(s)} = \frac{1000/s(s + 20)}{10(s + 10)/(s + 20) + (s + 10)} \tag{9-52}$$

which simplifies to

$$I(s) = \frac{1000}{s(s^2 + 40s + 300)} \tag{9-53}$$

This equation may be expanded by partial fractions as

$$\frac{1000}{s(s^2 + 40s + 300)} = \frac{K_1}{s} + \frac{K_2}{(s + 10)} + \frac{K_3}{(s + 30)} \tag{9-54}$$

With K_1, K_2, and K_3 evaluated, the current transform becomes

$$I(s) = \frac{3.33}{s} + \frac{-5}{s + 10} + \frac{1.67}{s + 30} \tag{9-55}$$

The time-domain current $i(t)$ is found by the inverse Laplace transformation as

$$i(t) = 3.33 - 5e^{-10t} + 1.67e^{-30t} \tag{9-56}$$

As a check, this equation reduces to the correct values for initial and final conditions.

Example 5

In the network shown in Fig. 9-24, it is required to find the current in the resistor R_2. The equivalent impedance schematic is shown in Fig. 9-25. It is assumed that the capacitor C_2 is initially uncharged and that the switch K is closed at $t = 0$. Thévenin's theorem is applied at terminals a-a', and the equivalent impedance and equivalent voltage at

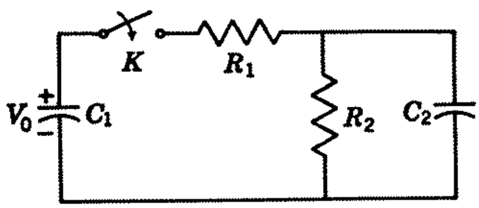


Fig. 9-24. Two-loop network.

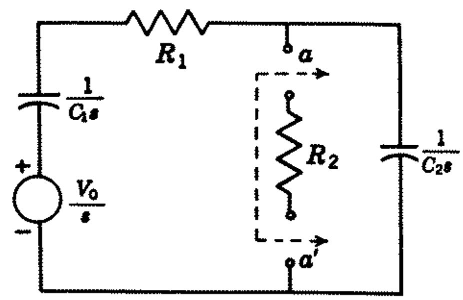


Fig. 9-25. Thévenin's equivalent of Fig. 9-24.

these terminals will be found. The equivalent impedance is a parallel combination of the impedance of two branches; thus

$$Z_{eq}(s) = \frac{(R_1 + 1/C_1s)1/C_2s}{R_1 + 1/C_1s + 1/C_2s} \tag{9-57}$$

and

$$V_{oc}(s) = \frac{(V_0/s)(1/C_2s)}{R_1 + 1/C_1s + 1/C_2s} \tag{9-58}$$

The current through R_2 is

$$\begin{aligned} I_2(s) &= \frac{V_{oc}(s)}{Z_{eq}(s) + R_2} \\ &= \frac{V_0/C_2}{R_1R_2s^2 + (R_1/C_2 + R_2/C_1 + R_2/C_2)s + 1/C_1C_2} \end{aligned} \tag{9-59}$$

Suppose that the following values are given for the network: $C_1 = 8\ \mu$f, $C_2 = 8\ \mu$f, $R_1 = 9$ megohms, $R_2 = 5$ megohms, and $V_0 = 75$ volts. With these parameter values, Eq. 9-59 reduces to

$$I_2(s) = \frac{0.208 \times 10^{-6}}{(s + 0.045)(s + 0.0077)} \tag{9-60}$$

This equation can be expanded by partial fractions to give

$$I_2(s) = 5.55 \times 10^{-6}\left[\frac{1}{s + 0.0077} - \frac{1}{s + 0.045}\right] \tag{9-61}$$

The inverse Laplace transformation gives $i_2(t)$ as

$$i_2(t) = 5.55 \times 10^{-6}(e^{-0.0077t} - e^{-0.045t}) \tag{9-62}$$

which is the required current. If the current in any *other* branch is required it is necessary to start over, applying Thévenin's theorem.

FURTHER READING

For further discussion of the concept of complex frequency, the student is referred to LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 189–193 and to Bode, *Network Analysis and Feedback Amplifier Design* (D. Van Nostrand Co., Inc., New York, 1945), pp. 18–30. For a discussion of the direct use of transforms in solving equations for a network, read Gardner and Barnes, *Transients in Linear Systems* (John Wiley & Sons, Inc., New York, 1942), pp. 176–214.

PROBLEMS

For systems described by the differential equations that follow, determine the complex frequencies that will appear in the solution, and designate whether these frequencies are natural frequencies determined by the passive parameters of the system or frequencies determined by the nature of the driving force. Call these two kinds of frequencies "free" and "forced," respectively.

9-1. (a) $\dfrac{d^2i}{dt^2} + \dfrac{di}{dt} + i = Ae^{-t}$

(b) $(p^2 + 4p + 5)(p^2 + 2p + 5)v = Be^{-t}\sin t, \qquad p = d/dt$

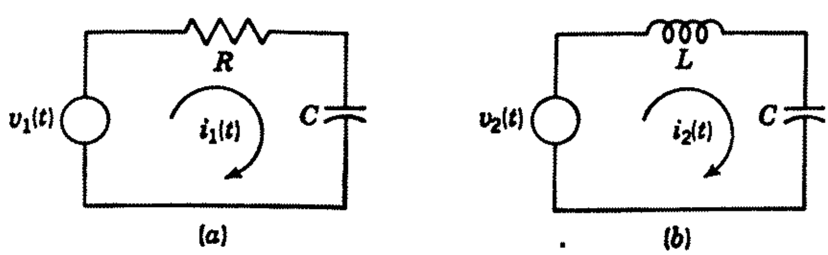
9-2. (a) $\dfrac{d^2i}{dt^2} + 5\dfrac{di}{dt} + 6i = \cos t$

(b) $\dfrac{di}{dt} + 5i = De^{-2t} + E\sin 7t$

9-3. $\dfrac{d^3v}{dt^3} + 2\dfrac{d^2v}{dt^2} + 2\dfrac{dv}{dt} + v = 1 + e^{-3t}\sin 3t$

9-4. $\dfrac{di_1}{dt} + i_1 - 3i_2 = \sin t, \qquad 2\dfrac{di_2}{dt} + i_2 - 3i_1 = 1$

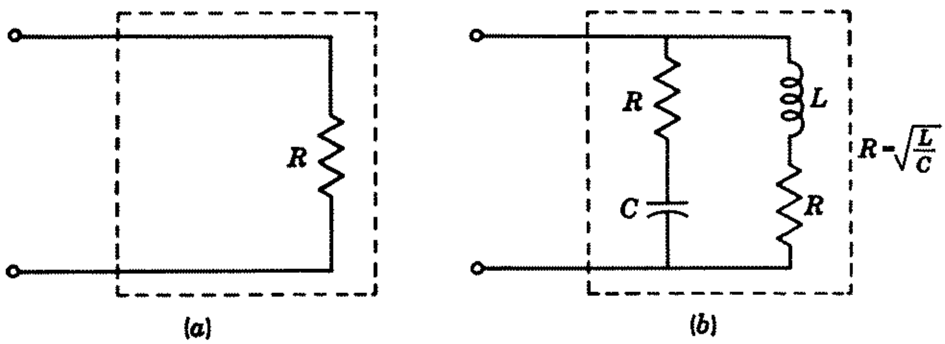
9-5. Consider the two series circuits shown in the accompanying figure. Given that $v_1(t) = \sin 10^3t$, $v_2(t) = e^{-1000t}$ for $t > 0$, and $C =$



Prob. 9-5.

1 μf. (a) Show that it is possible to have $i_1(t) = i_2(t)$ for all $t > 0$. (b) Determine the required values of R and L for (a) to hold. (c) Discuss the physical meaning of this problem in terms of the complex frequencies of the two series circuits.

9-6. Two black boxes with two terminals each are externally identical. It is known that one box contains the network shown as (a) and the other contains the network shown as (b) with $R = \sqrt{L/C}$. (a) Show that the input impedance, $Z_{in}(s) = V_{in}(s)/I_{in}(s) = R$ for both

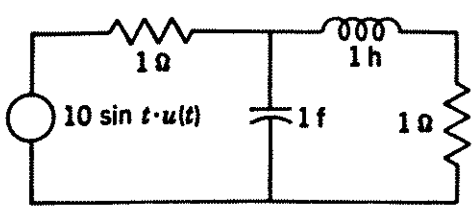


Prob. 9-6.

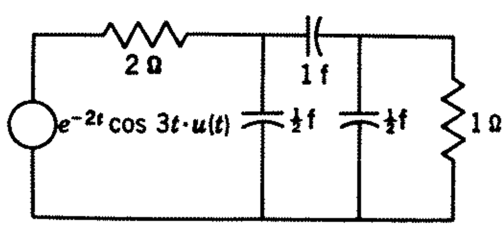
networks. (b) Investigate the possibility of distinguishing the purely resistive network. Any external measurements may be made, initial and final conditions may be examined, etc.

9-7. If the capacitors are initially uncharged and no current flows in the inductors at $t = 0$, determine the transform of the generator current $I(s)$ for the network shown in the accompanying figure. *Answer.*

$$I(s) = \frac{10(s^2 + s + 1)}{(s^2 + 1)(s^2 + 2s + 2)}$$



Prob. 9-7.

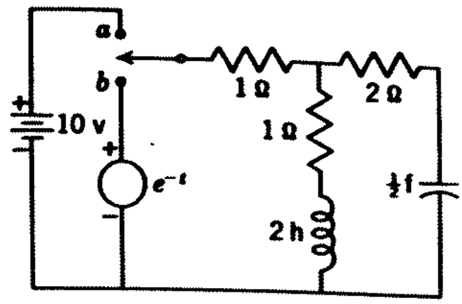


Prob. 9-8.

9-8. Repeat Prob. 9-7 for the network shown in the figure. *Answer.*

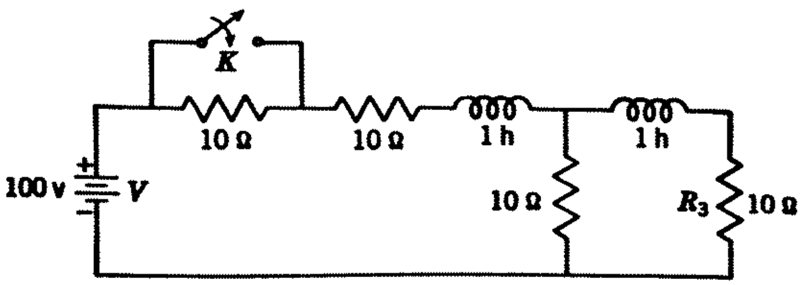
$$I(s) = \frac{s(s + 2)(5s + 6)}{(s^2 + 4s + 13)(10s^2 + 18s + 4)}$$

9-9. In the given network, the switch K is in position a until the network reaches a steady state. Then at $t = 0$, the switch K is moved to position b. Find the transform of the voltage across the 0.5-farad capacitor, using Thévenin's theorem.



Prob. 9-9.

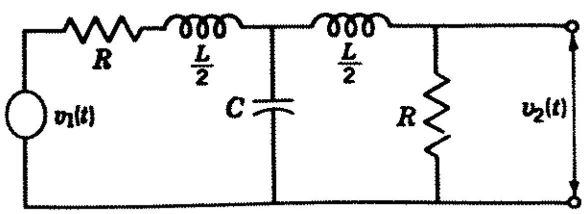
9-10. The network of Example 4, Fig. 9-22, has been modified as shown in the accompanying figure. If the switch K is closed at $t = 0$, a steady state having previously existed, find the current in R_3, using Thèvenin's theorem. Compare this result with Eq. 9-56.



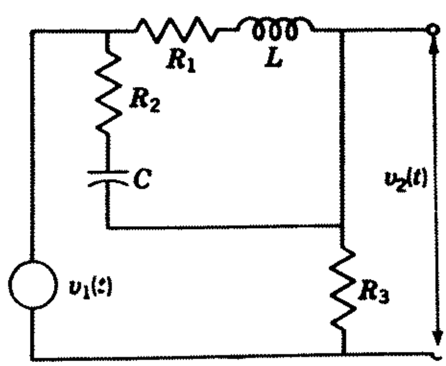
Prob. 9-10.

9-11. The network shown in the figure is a low-pass filter (to be studied in Chap. 14). The input voltage $V_1(t)$ is a unit step function, and the input and load resistors have the value $R = \sqrt{L/C}$. By using Thèvenin's theorem, show that the transform of the output voltage is

$$V_2(s) = \frac{4}{(LC)^{3/2}}\left[\frac{1}{s(s^3 + 4\sqrt{1/LC}\,s^2 + 8s/LC + 8/(LC)^{3/2}}\right]$$



Prob. 9-11.



Prob. 9-12.

9-12. In the network shown in the accompanying sketch, the elements are chosen such that $L = CR_1^2$ and $R_1 = R_2$. If $V_1(t)$ is a voltage pulse of 1-volt amplitude and T sec duration, show that $V_2(t)$ is also a pulse, and find its amplitude and time duration.

CHAPTER 10

NETWORK FUNCTIONS

In this chapter, the concept of transform impedance and transform admittance which was introduced in the last chapter will be studied and extended. Further, a function relating currents or voltages at different parts of the network, called a *transfer function*, will be found to be mathematically similar to the transform impedance function. These two functions are called *network functions*.

10-1. Terminals and terminal pairs

Consider an arbitrary network made up entirely of passive elements. To indicate the general nature of the network, let it be represented by the symbol of a rectangle (or a box). If a conductor is fastened to any node in the network and brought out of the box for access, the end of this conductor is designated as a *terminal*. Terminals are required for connecting driving forces to the network, for connecting some other network (say a load), or for making measurements. The minimum number of terminals that are useful is *two*. Further, the terminals are associated in pairs, one pair for a driving force, another pair for the load, etc. Two associated terminals are given the name *terminal pair*.

In Fig. 10-1(a) is shown a symbolic representation of a one-terminal-pair (or two-terminal) network. The terminal pair is customarily connected to a driving force and so is sometimes given the name *driving*

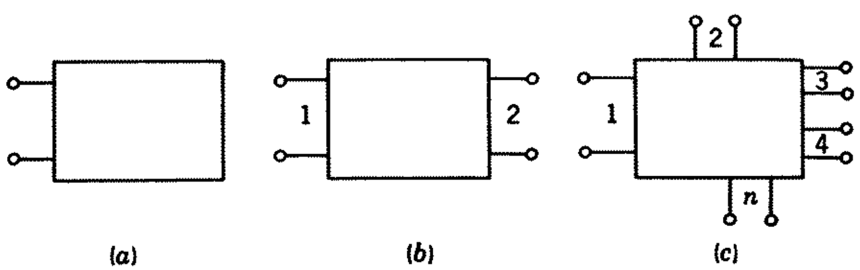


Fig. 10-1. Network representations.

point. Figure 10-1(b) shows a two-terminal-pair network. The terminal pair designated 1 is usually connected to a driving force (or input) while the terminal pair marked 2 is usually connected to a load (as an output). The number of terminal pairs in a network can increase without limit: Fig. 10-1(c) shows a representation of an n-terminal-pair network. All the discussion in this chapter, however, will be concerned with one- and two-terminal-pair networks.

10-2. Driving-point immittances

The transform impedance has been defined as the ratio of the voltage transform to the current transform; that is,

$$Z(s) = \frac{V(s)}{I(s)} \tag{10-1}$$

Similarly, the transform admittance is defined as the ratio

$$Y(s) = \frac{I(s)}{V(s)} \tag{10-2}$$

The voltage transform and current transform that define transform impedance and transform admittance must relate to the same pair of terminals. The impedance or admittance found at a given terminal pair is called a *driving-point impedance* (*or admittance*).

Because of the similarity of impedance and admittance (and to avoid writing "impedance and admittance"), the two quantities are assigned one name, *immittance* (a combination of *im*pedance and ad*mittance*). An immittance is thus an impedance *or* an admittance.

The driving-point immittance of a network is found by combining impedance terms (Ls, R, and $1/Cs$) or admittance terms (Cs, G, and $1/Ls$) by adding, multiplying, or dividing. This algebraic combination of terms results in an immittance function in the form of a quotient of polynomials as

$$\frac{a_0s^n + a_1s^{n-1} + \ldots + a_{n-1}s + a_n}{b_0s^m + b_1s^{m-1} + \ldots + b_{m-1}s + b_m} \tag{10-3}$$

which is a *rational* function of s (n and m are integers).

In this equation, n is the order of the numerator polynomial and m is the order of the denominator polynomial. The polynomials may be of any order including zero, although we will later show that there is a restriction in the difference in order of the two polynomials.

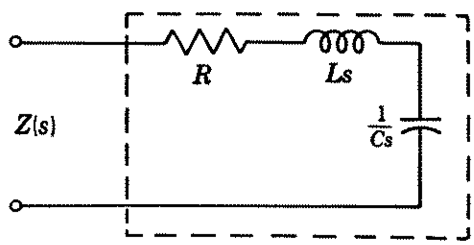


Fig. 10-2. Series network.

Example 1

Figure 10-2 shows an *RLC* series one-terminal-pair network with transform impedances marked for each element. The driving-point impedance $Z(s)$ is

$$Z(s) = R + Ls + \frac{1}{Cs} = \frac{LCs^2 + RCs + 1}{Cs} \tag{10-4}$$

or

$$Z(s) = L\,\frac{s^2 + Rs/L + 1/LC}{s} \tag{10-5}$$

The numerator polynomial for this driving-point impedance is of second order, while the denominator polynomial is of first order.

Example 2

Figure 10-3 shows a more complicated network consisting of a series RL network shunted by a capacitor. The driving-point impedance is

$$Z(s) = \frac{1}{Cs + 1/(R + Ls)} = \frac{1}{C}\frac{s + R/L}{s^2 + Rs/L + 1/LC} \tag{10-6}$$

In this driving-point impedance function, the numerator is of first order and the denominator is of second order. The driving-point admittance function $Y(s)$ for this network is the reciprocal of Eq. 10-6.

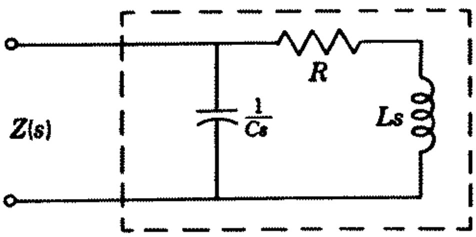


Fig. 10-3. Network of Example 2.

Fig. 10-4. Two-terminal-pair network.

10-3. Transfer functions

The concept of a transfer function is identified with networks having at least two terminal pairs. Such a network is shown in Fig. 10-4. Although the driving-point immittances at terminal pair 1 and terminal pair 2 are of interest, we are also interested in the ratio of excitation and response for the two terminal pairs. The function relating the transform of a quantity at one terminal pair to the transform of another quantity at another terminal pair is given the name *transfer function.** There are several forms for transfer functions in electric networks:

(1) The ratio of one voltage to another voltage, or the voltage transfer ratio.
(2) The ratio of one current to another current, or the current transfer ratio.
(3) The ratio of one current to another voltage or one voltage to another current.

The transfer function for a voltage or current ratio is assigned the symbol $G(s)$. If terminal pair 2 of Fig. 10-4 is designated the output terminal-pair, and terminal pair 1 is designated the input, then the

* In computing the transfer function, all initial conditions are assumed to be zero.

voltage transfer ratio of the output to the input is

$$G(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{V_2(s)}{V_1(s)} \tag{10-7}$$

The ratio of voltage to current or current to voltage is dimensionally immittance, but since the two quantities are not measured at the same terminals, such a ratio is designated a *transfer immittance* in ohms or mhos. The transfer immittance is given the same symbol as the driving-point immittance with subscripts to identify the terminals. For example,

$$Z_{12}(s) = \frac{V_1(s)}{I_2(s)} \quad \text{and} \quad Y_{21}(s) = \frac{I_2(s)}{V_1(s)} \tag{10-8}$$

where the first subscript identifies the numerator quantity and the second identifies the denominator quantity. The transfer function is determined by the network immittances and can always be reduced to a quotient of polynomials,

$$G(s) = \frac{P(s)}{Q(s)} = \frac{a_0s^n + a_1s^{n-1} + \ldots + a_n}{b_0s^m + b_1s^{m-1} + \ldots + b_m} \tag{10-9}$$

The transfer function thus has the same general form as the driving-point immittance function.

Example 3

The two-terminal-pair network shown in Fig. 10-5 has marked $V_1(s)$ as the input voltage and $V_2(s)$ as the output voltage transform. This network acts as a voltage divider. With no current in the output terminals, the voltage equations are

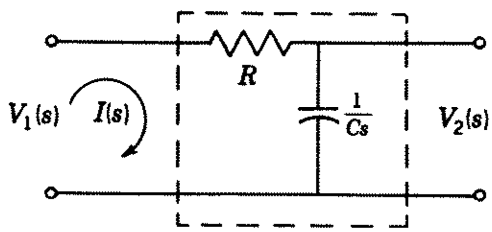


Fig. 10-5. Two-terminal-pair network.

$$RI(s) + \frac{1}{Cs} I(s) = V_1(s) \tag{10-10}$$

$$\frac{1}{Cs} I(s) = V_2(s) \tag{10-11}$$

The ratio of these equations is

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{(1/Cs)I(s)}{(R + 1/Cs)I(s)} = \frac{1}{RCs + 1} \tag{10-12}$$

or

$$G(s) = \frac{1/RC}{s + 1/RC} \tag{10-13}$$

for this network. This transfer function has a numerator polynomial of zero order and a denominator polynomial of first order.

Example 4

The two-terminal-pair network shown in Fig. 10-6 is similar to that of Example 1 except that the resistor has been replaced by an inductor. It is not necessary to write Kirchhoff's equations as above to find the transfer function, since this network is essentially a voltage divider. The transfer function for the voltage ratio becomes

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{1/Cs}{Ls + 1/Cs} = \frac{1}{LCs^2 + 1} \tag{10-14}$$

or

$$G(s) = \frac{1/LC}{s^2 + 1/LC} \tag{10-15}$$

The numerator polynomial is of zero order, and the denominator polynomial is of second order.

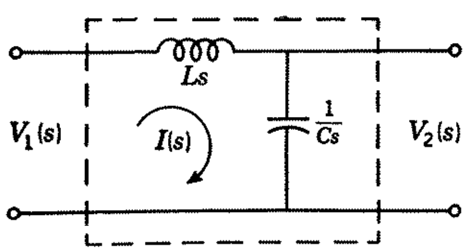


Fig. 10-6. Two-terminal-pair network.

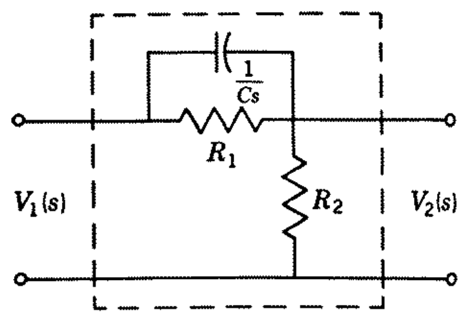


Fig. 10-7. Network of Example 5.

Example 5

The same voltage-divider network concept can be used with more than one current loop in the network by using network reduction. Figure 10-7 shows such a network. The transform impedances R_1 and $1/Cs$ can be combined into an equivalent impedance having the value

$$Z_{eq}(s) = \frac{1}{Cs + 1/R_1} = \frac{R_1}{R_1Cs + 1} \tag{10-16}$$

Then the transfer function becomes

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{R_2}{R_2 + Z_{eq}(s)} \tag{10-17}$$

or

$$G(s) = \frac{R_2R_1Cs + R_2}{R_2R_1Cs + R_1 + R_2} \tag{10-18}$$

which may be reduced to

$$G(s) = \frac{s + 1/R_1C}{s + (R_1 + R_2)/R_1R_2C} \tag{10-19}$$

In this transfer function, the order of the numerator and order of the denominator are the same. This particular network finds application in servomechanisms where it is known as a "lead" network.

10-4. Poles and zeros

All network functions have the form of a quotient of polynomials as

$$\frac{a_0 s^n + a_1 s^{n-1} + \ldots + a_{n-1}s + a_n}{b_0 s^m + b_1 s^{m-1} + \ldots + b_{m-1}s + b_m} \tag{10-20}$$

If the numerator polynomial is factored into its n roots, and the denominator polynomial is factored into its m roots, the equation can be written in the form

$$H\,\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \tag{10-21}$$

where $H = a_0/b_0$ is a constant known as the *scale factor*, and the roots $s_1, s_2, \ldots, s_a, s_b, \ldots$ are complex frequencies. When the variable s has the values $s_1, s_2, \ldots, s_n$, the network function vanishes. Such complex frequencies are called *zeros* of the network function. When s has the values $s_a, s_b, \ldots, s_m$, the network function becomes infinite. These complex frequencies are called *poles* of the network function. Poles and zeros are important in network theory; a comparison of the last two equations shows that a network function is completely specified by its poles, zeros, and the scale factor.

There is the possibility that roots of Eq. 10-21 may coincide. Such multiple roots, corresponding to a factor of the form $(s - s_q)^r$, are described as poles or zeros (depending on location in the numerator or denominator) *of order* r. For a nonrepeated root, such that $r = 1$, the pole or zero is said to be *simple.*

Both zero and infinite values of s are possible pole or zero locations. From Eq. 10-21 it is seen that:

(1) When $n > m$, $s = \infty$ is a pole of order $n - m$.
(2) When $n < m$, $s = \infty$ is a zero of order $m - n$.
(3) When $n = m$, $s = \infty$ is neither a zero nor a pole but an ordinary point.

If, for any rational network function, poles and zeros at zero and infinity are taken into account in addition to finite poles and zeros, *the total number of zeros is equal to the total number of poles.* For example, the network function

$$\frac{(s + 1)(s + 2 + j1)(s + 2 - j1)}{s^3(s + 3)(s + 5)} \tag{10-22}$$

has five zeros and five poles. The zeros are at $s_1 = -1$, $s_2 = -2 - j1$, $s_3 = -2 + j1$, and $s_4 = s_5 = \infty$. The poles occur at the complex frequencies $s_a = s_b = s_c = 0$, $s_d = -3$, $s_e = -5$. These poles and zeros are plotted on the complex s plane ($s = \sigma + j\omega$) in Fig. 10-8. The real

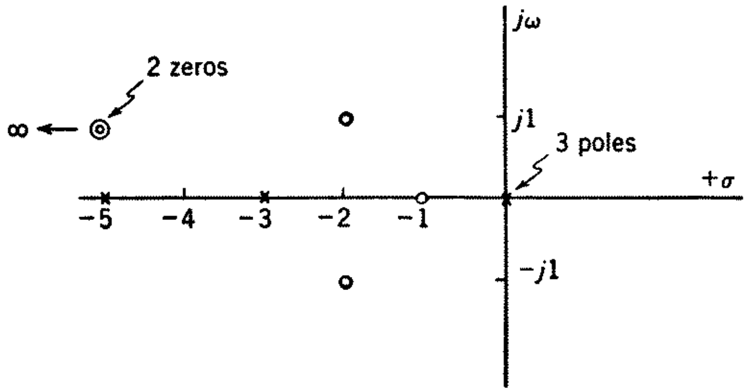


Fig. 10-8. Poles and zeros in the s plane.

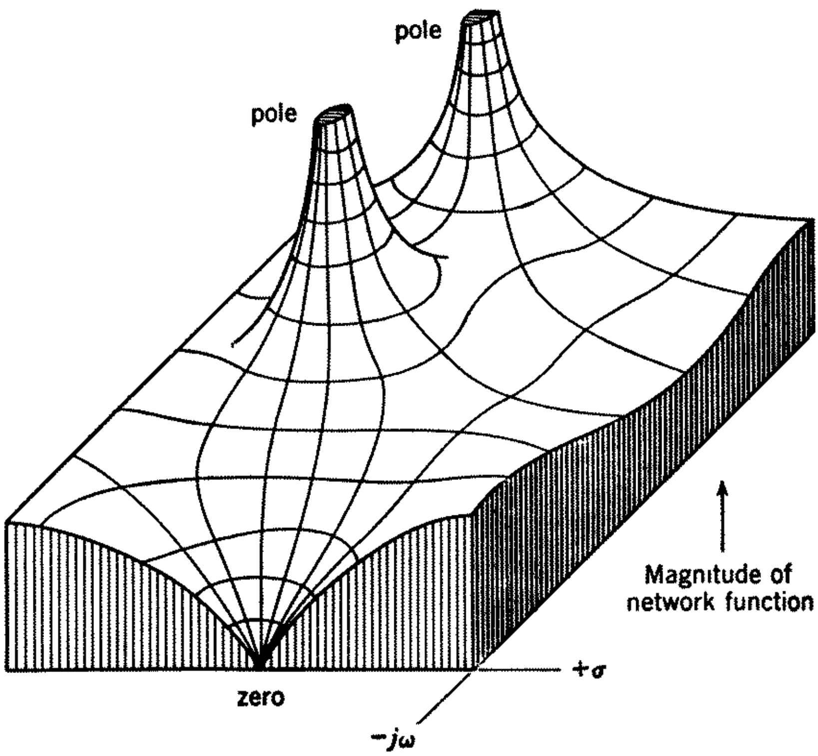


Fig. 10-9. The magnitude of a network function plotted in the complex frequency plane, showing two poles and one zero.

part is plotted along the σ axis, and the imaginary part along the $j\omega$ axis. The symbol O is used to designate the location of a zero and the symbol × for the location of a pole.

Poles and zeros designate *critical frequencies.* At poles the network function becomes infinite, while at zeros the network function becomes zero. At other complex frequencies, the network function has a finite, nonzero value. A three-dimensional representation of the magnitude of the transfer function as a function of complex frequency is shown in Fig. 10-9, for one quadrant of the s plane. The portion of the complex plane represented in Fig. 10-9 is shown in Fig. 10-10. This particular network function has four finite poles, one finite zero, and a third-order zero at infinity.

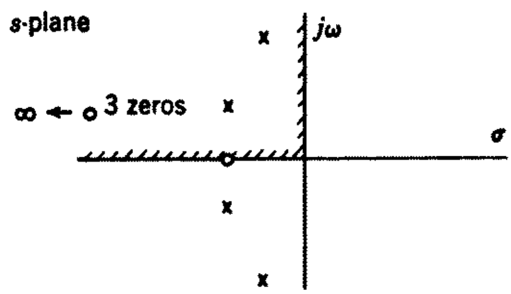


Fig. 10-10. s plane for Fig. 10-9.

The pole represents a frequency at which the network function "blows up." The zero represents a frequency at which the opposite behavior takes place: the network function becomes nothing at all. Either "blowing up" or "becoming nothing" sounds like rather drastic behavior for the network function. We might wonder if it would not be wise to completely avoid poles and zeros, to select network functions without poles or zeros. Such is not the case at all. Poles and zeros are the lifeblood of a function; without poles and zeros the function reduces to a dull, drab, grubby constant—a function which does not change under any conditions. Without poles and zeros, the three-dimensional representation of the network function becomes a tedious expanse of mathematical desert—absolutely flat. But add a few poles and a few zeros and we have a land of spectacular peaks (elevation: ∞) and beautiful springs (elevation: 0). This picture will become clearer as we study concepts of network behavior with the aid of poles and zeros.

Consider the transfer function for a voltage ratio

$$\frac{V_{out}(s)}{V_{in}(s)} = G(s) \tag{10-23}$$

which may be written

$$V_{out}(s) = G(s)V_{in}(s) \tag{10-24}$$

In the usual problem, $v_{in}(t)$ is specified, and $G(s)$ can be computed from the network. The problem is to find the response, $v_{out}(t)$. When the last equation is expanded by partial fractions, the denominator of each partial fraction term gives a pole of either $G(s)$ or $V_{in}(s)$; that is, with no repeated roots in the denominator of $V_{out}(s)$

$$G(s)V_{in}(s) = \sum_{j=1}^{p} \frac{K_j}{s - s_j} + \sum_{k=1}^{v} \frac{K_k}{s - s_k} \tag{10-25}$$

where p is the number of poles of $G(s)$, and v is the number of poles of $V_{in}(s)$. Performing the inverse Laplace transformation of this equation gives

$$V_{out}(t) = \mathcal{L}^{-1}G(s)V_{in}(s) = \sum_{j=1}^{p} K_j e^{s_j t} + \sum_{k=1}^{v} K_k e^{s_k t} \tag{10-26}$$

Thus the frequencies s_j are the natural complex frequencies corresponding to *free oscillations*. The frequencies s_k are the driving-force complex frequencies corresponding to *forced oscillations*. The poles therefore determine the waveform of the time variation of the response, the output voltage. The zeros determine the magnitude of each part of the response, since they determine the magnitude of K_j and K_k in the partial fraction expansion, as we shall see.

In terms of driving-point immittances, poles and zeros have easily visualized meanings. Since $Z(s) = V(s)/I(s)$, a pole of $Z(s)$ implies zero current for a finite voltage, which means an *open circuit*. A zero of $Z(s)$, on the other hand, means no voltage for a finite current, or a *short circuit*. Thus a one-terminal-pair network is an open circuit for pole frequencies and a short circuit for zero frequencies. This can be visualized easily in terms of single element networks. For a capacitor, the driving-point impedance is $Z(s) = 1/Cs$. This network function has a pole at $s = 0$ and a zero at $s = \infty$. It behaves as an open circuit at the pole frequency ($\omega = 0$) and as a short circuit at infinite frequency. Likewise, for an inductor, the driving-point impedance $Z(s) = Ls$ (zero at $s = 0$, pole at $s = \infty$) and this element behaves as a short circuit at zero frequency and as an open circuit at infinite frequency.

10-5. Restrictions on pole and zero locations in s-plane

The poles and zeros of network functions have limitations as to their location in the s plane. These restrictions follow from two facts: (1) the terms in the polynomials of the form

$$a_0 s^n + a_1 s^{n-1} + \ldots + a_{n-1}s + a_n \tag{10-27}$$

have coefficients (a_0, a_1, $\ldots$, a_n) which are positive and real. This follows because each of these coefficients is determined by some combination of R, L, and C, and these parameters must be positive and real (the only way they appear in nature). (2) The networks being considered are made up of passive elements *only*. The rules for location of poles and zeros are different for driving-point functions and for transfer functions, and so will be considered separately.

Driving-Point Immittance Functions. (1) Since the coefficients of both numerator and denominator polynomials of driving-point immittance functions are positive and real, poles and zeros are either real or occur in conjugate pairs. This was discussed in more detail in Art. 6-3.

(2) All poles and zeros of driving-point immittance functions have *negative* real parts. Consider a denominator factor $(s - s_a)$, where s_a is a pole having a real and imaginary part, $s_a = \sigma_a + j\omega_a$. If σ_a is positive, this pole will give rise to a time-domain factor (by finding the inverse Laplace transformation) of the form

$$K_a e^{s_a t} = K_a e^{\sigma_a t} e^{j\omega_a t} \tag{10-28}$$

The exponential term $(e^{\sigma_a t})$ increases exponentially as t increases. For such a pole in $Z(s)$, the voltage would increase without limit for any current input, and for such a pole in $Y(s)$ the current would increase without limit for any voltage input. Since this cannot happen physically with only passive elements in the network, the poles and zeros of a driving-point immittance function *have negative real parts*. In terms of pole and zero location in the s plane, all poles and zeros must be in the left half plane (LHP) and can never occur in the right half plane (RHP). Poles and zeros can be on the boundary (the $j\omega$ axis) subject to the limitations we discuss next.

(3) Poles and zeros on the $j\omega$ axis of the s plane (corresponding to real radian frequency) will always be *simple*. The reason for this restriction is the same as that listed in (2). Multiple poles give rise to time domain functions of the type $(t \cos \omega t)$, $(t \sin \omega t)$, etc., and such terms increase without limit as t increases. Such an increase is not possible for a network made up of passive elements only. For example, consider the following transform pair.

$$\mathcal{L}^{-1} \frac{s}{(s^2 + \omega^2)^2} = \frac{t}{2\omega} \sin \omega t \tag{10-29}$$

The transform expression corresponds to two poles at $-j\omega$ and two at $+j\omega$. The time domain factor of the transform pair is a linearly increasing sinusoid.

Multiple poles and zeros are permitted at other locations in the left half of the s plane, since such poles give rise to terms of the form $t^n e^{-\sigma t}$, having the required zero limit since

$$\lim_{t \to \infty} t^n e^{-\sigma t} = 0$$

for finite n by l'Hospital's rule.

(4) The order of the numerator polynomial and denominator polynomial for a driving-point immittance function can differ at most by unity. If the driving-point immittance function is found without algebraic error, this restriction will always be observed. It can, how-

ever, be shown to be a requirement which follows from the restriction listed under (2) above. First, we must further discuss the meaning of infinite frequency. We usually visualize high frequency in terms of an increasing sinusoidal frequency. Starting with conventional 60-cycle generators, we next visualize an audio oscillator, a radio frequency oscillator, a microwave frequency oscillator, and then, somewhere far beyond, lies infinite frequency. This is a rather nebulous and hazy concept, but it is the best we have. In terms of the s plane, we have followed the $j\omega$ axis from a value near the origin on out to infinity. It is not necessary to follow the $j\omega$ axis. Following any other path in the s plane will eventually lead us to infinity, and once we get there we are at the same place as if we had followed the $j\omega$ axis. In other words, infinite frequency is just *one frequency*, and is reached by traveling *any* direction from the origin of the s plane. Infinity is a unique point (or it would not be infinity). We can say that the s plane is really not a plane at all—it is a sphere, similar to the earth. Let the north pole represent the origin of the s plane. Standing at the north pole, the s plane looks flat, which is really not too unreasonable since it is part of a sphere of infinite radius. But if you go far enough in any direction from the north pole of the s plane, you end up at the south pole, which is one point infinitely far removed from the north pole of the s plane.

Now if infinity is only one point in the s plane, it includes the $j\omega$ axis (sort of an international date line in the s world). But by item (3), poles and zeros on the $j\omega$ axis must be simple. Hence poles and zeros at infinity for a driving-point immittance function must be simple.* The only way we can get a simple pole or a simple zero at infinity is to have the order of the numerator exceed that of the denominator by unity for a pole at infinity, and have the order of the denominator exceed that of the numerator by unity for a zero at infinity. This rule is satisfied automatically when you compute the driving-point immittance function. If it is not, you have made an algebraic mistake.

Transfer Functions for Output/Input. The restrictions for transfer functions are not so rigid as those for driving-point immittances, because the transfer function is determined as the ratio of two different quantities at different points in one network. The restrictions will be given by analogy to those for driving-point immittance functions and in the same order.

(1) This restriction also holds for transfer functions. Poles and zeros are either real or occur in conjugate pairs.

(2) The poles of a transfer function must have negative real parts, but this restriction does not hold for the zeros. A network with zeros in

* This is intended to be only a suggestive or heuristic proof.

the left half plane only is classified as *minimum phase;* those with zeros in the right half plane are *nonminimum phase.*

(3) This restriction holds for transfer function poles. Poles and zeros on the $j\omega$ axis will always be simple.

(4) For (output/input) transfer functions, the order of the numerator may exceed the order of the denominator by one. However, the numerator order may be any value less than that of the denominator.

10-6. Time-domain behavior from the pole and zero plot

In this section, we will show that the time-domain behavior of a system can be determined from the s plane plot of the poles and zeros of its transfer function and those of the transform of the active-source driving-forces. Suppose that the transform of some variable, say a current $I(s)$, is found, and the poles and zeros are determined as

$$I(s) = Y(s)V(s) = \frac{P(s)}{Q(s)} \tag{10-30}$$

where

$$\frac{P(s)}{Q(s)} = H\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \tag{10-31}$$

It was shown in Art. 10-4 that the poles of this function determine the time-domain behavior of $i(t)$. It was suggested that the zeros determine the magnitude of each of the terms of $i(t)$. In this section, we will amplify these concepts by showing how $i(t)$ can be determined from a knowledge of the poles, the zeros, and the scale factor H.

In terms of the damping ratio ζ and the undamped natural frequency, ω_n as discussed on page 104, the poles and zeros of the last equation will have the following forms.

$$s_1, s_2 = -\zeta\omega_n \pm j\omega_n\sqrt{1 - \zeta^2} \qquad \zeta < 1 \tag{10-32}$$

$$s_1, s_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1} \qquad \zeta > 1 \tag{10-33}$$

$$s_1, s_2 = -\omega_n \qquad \zeta = 1 \tag{10-34}$$

$$s_1, s_2 = \pm j\omega_n \qquad \zeta = 0 \tag{10-35}$$

It was also shown on page 105 that contours of constant ω_n are circles in the s plane, that contours of constant damping ratio are straight lines through the origin, and that contours of constant damping ($\zeta\omega_n$) are straight lines parallel to the $j\omega$ axis of the s plane. Further, lines parallel to the σ axis of the s plane are lines of constant actual frequency of oscillation, $\omega_n\sqrt{1 - \zeta^2}$. These facts are summarized in Fig. 10-11.

The location of the poles in the s plane can be interpreted in terms of the general time-domain response in terms of ζ and ω_n.

$$i(t) = K_1 e^{(-\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})t} + K_2 e^{(-\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})t} \tag{10-36}$$

To illustrate the use of the contours of Fig. 10-11, consider the array of

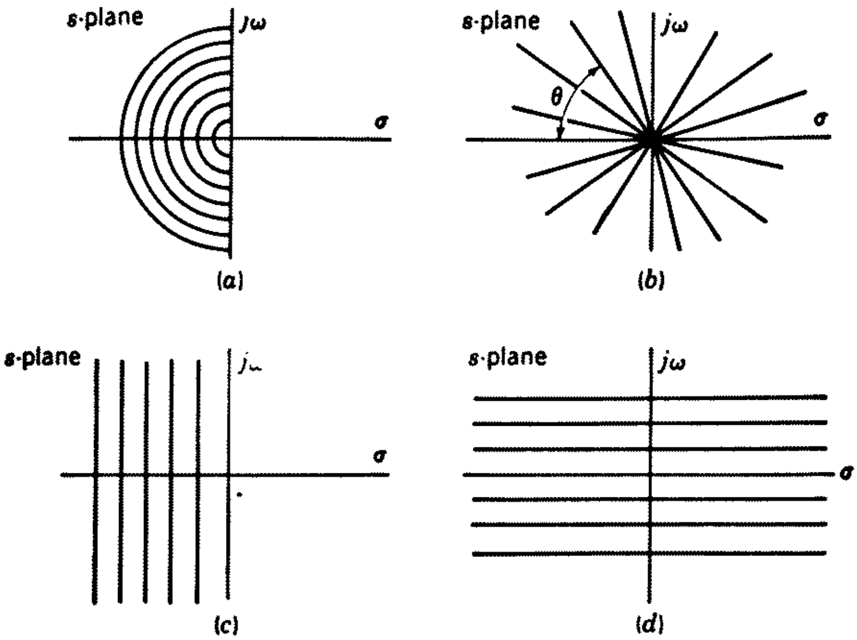


Fig. 10-11. Constant contours in the s plane: (a) constant radius $= \omega_n$; (b) constant damping ratio line $\theta = \tan^{-1} (\sqrt{1 - \zeta^2}/\zeta)$; (c) constant negative damping line $\sigma = -\zeta\omega_n$ (or any real part of s); (d) constant actual frequency of oscillation lines $\omega = \pm\omega_n \sqrt{1 - \zeta^2}$. ζ and ω_n are defined by the second-order characteristic equation $s^2 + 2\zeta\omega_n s + \omega_n^2 = 0$.

poles shown in Fig. 10-12 (zeros have been omitted for clarity). The pair of poles s_a and s_a^* and the pair s_c and s_c^* correspond to oscillatory expressions in the time domain. The actual frequency of oscillation corresponding to s_a and s_a^* is higher than that of s_c and s_c^*, just as the damping (or rate of decreasing amplitude) is less for s_a and s_a^* than for s_c and s_c^*. The natural frequency of the two pole pairs is approximately the same, since they are on about the same radius from the origin. The difference in actual frequency of oscillation is due to a lower damping ratio for s_a and s_a^*.

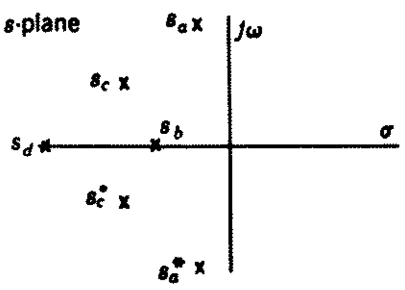


Fig. 10-12. Typical poles in the s plane.

The poles s_b and s_d are quite different from the conjugate pairs just considered. They correspond to the overdamped case, and have an exponential decay form in the time domain. The damping is greater for s_d than for s_b. From another point of view, the time constant for the pole s_b is greater than that for s_d. Typical time-domain response corresponding to each pole is shown in Fig. 10-13 for an arbitrary

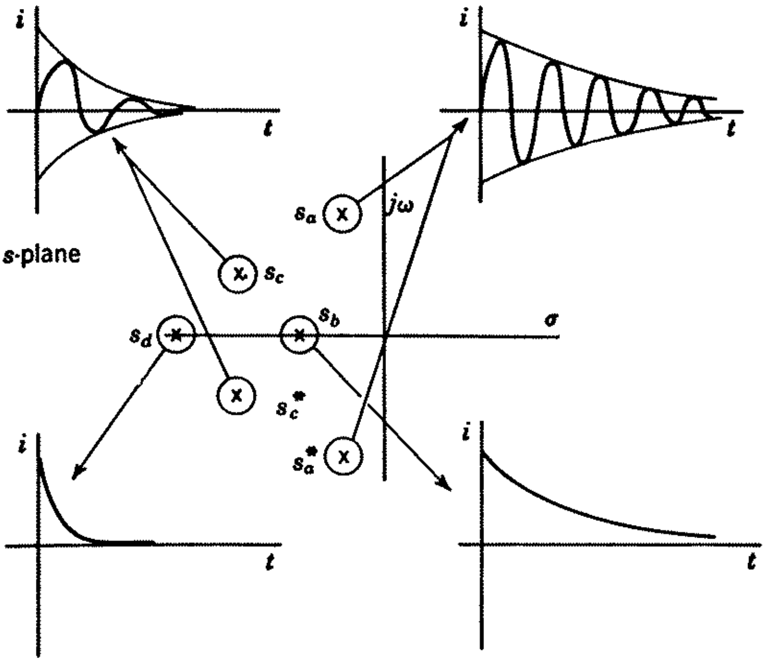


Fig. 10-13. Response comparison for various poles in the s plane (arbitrary amplitudes).

amplitude for each factor. The total response corresponding to these poles is found by *adding* each of the individual factors as

$$i(t) = K_a e^{s_a t} + K_a^* e^{s_a^* t} + K_b e^{s_b t} + K_c e^{s_c t} + K_c^* e^{s_c^* t} + K_d e^{s_d t} \quad (10\text{-}37)$$

As usual, the terms corresponding to conjugate pairs will combine to give damped sinusoidal expressions.

There remains the problem of determining the multiplying constant (or magnitude) for each of the terms (or modes). The starting point is Eq. 10-31. To find the time-domain response corresponding to this transform equation, we expand by partial fractions. Hence

$$I(s) = \frac{K_a}{s - s_a} + \frac{K_b}{s - s_b} + \cdots + \frac{K_r}{s - s_r} + \cdots + \frac{K_m}{s - s_m} \quad (10\text{-}38)$$

Any of the K-coefficients, say K_r, can be found by the Heaviside method as

$$K_r = H\,\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)\ldots\cancel{(s - s_r)}\ldots(s - s_m)}\,\cancel{(s - s_r)}\bigg|_{s = s_r} \quad (10\text{-}39)$$

Substituting s_r for s in Eq. 10-38 gives the following value for K_r.

$$K_r = H \frac{(s_r - s_1)(s_r - s_2)\ldots(s_r - s_n)}{(s_r - s_a)(s_r - s_b)\ldots(s_r - s_m)} \tag{10-40}$$

This equation is composed of factors of the general form $(s_r - s_n)$, where both s_r and s_n are known complex numbers. The difference of two complex numbers is another complex number which may be written in polar form as

$$(s_r - s_n) = M_{nr}e^{j\phi_{nr}} \tag{10-41}$$

where M_{nr} is the magnitude of the phasor $(s_r - s_n)$, and ϕ_{nr} is the phase angle of the same phasor. The difference of the two complex quantities s_r and s_n is illustrated in Fig. 10-14 (other poles and zeros are omitted

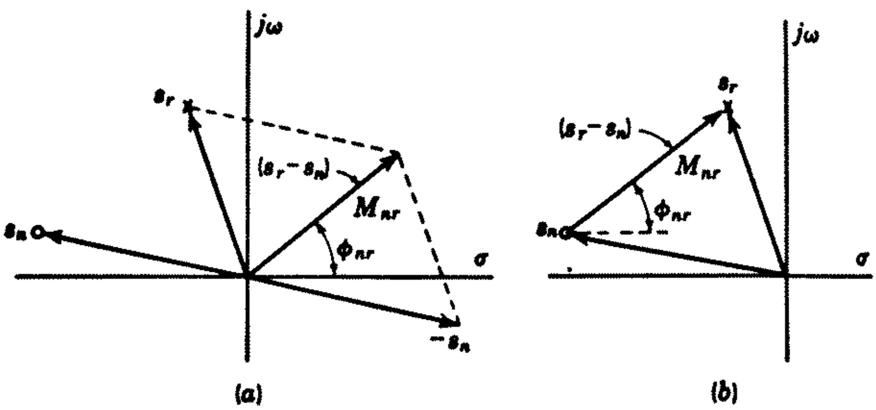


Fig. 10-14. Magnitude and phase of $(s_r - s_n)$ (other poles and zeros omitted): (a) polar diagram; (b) string diagram.

again for clarity). The term $(s_r - s_n)$ is interpreted as a phasor directed *from* s_n *to* s_r. The magnitude M_{nr} is the *distance* from s_n to s_r; the phase angle ϕ_{nr} is the angle of the line from s_n to s_r, measured with respect to the $\phi = 0$ line. The magnitude and phase of the factor $(s_r - s_n)$ are thus easily measured, and so all terms of this general type in Eq. 10-40 are easily found. In terms of M and ϕ for each factor in Eq. 10-40, the value of K_r is seen to become

$$K_r = H \frac{M_{1r}M_{2r}M_{3r}\ldots M_{nr}}{M_{ar}M_{br}M_{cr}\ldots M_{mr}} e^{j(\phi_{1r}+\phi_{2r}+\ldots-\phi_{ar}-\phi_{br}-\ldots)} \tag{10-42}$$

This equation gives K_r as a magnitude and phase. By performing the operations indicated by this equation, the constant K_r can be evaluated. Determining the quantities in Eq. 10-42 is readily accomplished by a graphical procedure which may be outlined as:

(1) Plot the poles and zeros of $I(s) = P(s)/Q(s)$ *to scale* on the complex s plane.

(2) Measure (or compute) the distance *from* each of the other finite poles and zeros *to* a given pole s_r.
(3) Measure (or compute) the angle *from* each of the other finite poles and zeros *to* a given pole s_r.
(4) Substitute these quantities into Eq. 10-42 and so evaluate K_r.

An example will illustrate this procedure. Suppose $I(s)$ has poles $s = -1$ and -3 and a zero at the origin, and H is given as 5. The current transform has the form

$$I(s) = \frac{5s}{(s+1)(s+3)} \qquad (10\text{-}43)$$

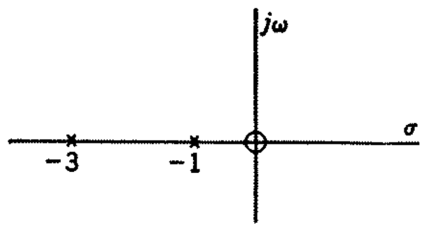


Fig. 10-15. Pole-zero configuration.

This function is easily expanded by partial fractions, but can also be evaluated as outlined above. Referring to Fig. 10-15, it is seen that

$$M_{01}e^{j\phi_{01}} = 1e^{j180^\circ} \qquad (10\text{-}44)$$

$$M_{31}e^{j\phi_{31}} = 2e^{j0^\circ} \qquad (10\text{-}45)$$

Hence
$$K_1 = H\frac{M_{01}e^{j\phi_{01}}}{M_{31}e^{j\phi_{31}}} = 5 \times \tfrac{1}{2}e^{+j180^\circ} = -2.5 \qquad (10\text{-}46)$$

Similarly,
$$K_3 = H\frac{M_{03}e^{j\phi_{03}}}{M_{13}e^{j\phi_{13}}} = 5 \times \frac{3e^{j180^\circ}}{2e^{j180^\circ}} = 7.5 \qquad (10\text{-}47)$$

Since the poles determine the frequency (in this case neper frequency), we write for the general solution,

$$i(t) = K_1e^{-t} + K_3e^{-3t} \qquad (10\text{-}48)$$

and since K_1 and K_3 have been evaluated from a knowledge of the pole and zero locations, we have as a particular solution,

$$i(t) = -2.5e^{-t} + 7.5e^{-3t} \qquad (10\text{-}49)$$

From this discussion and with the aid of Eq. 10-40, the influence of a zero on the time-domain response can be visualized. Consider one pole, say s_r, in Fig. 10-14. If all other poles and zeros in the s plane remain fixed in position and the zero s_n is moved, the proximity of a zero to a pole is seen by Eq. 10-42 to *reduce* the magnitude of the K-coefficient associated with the complex frequency of the pole s_r. Again, from Eq. 10-42, proximity of a pole to s_r is seen to have the opposite effect—since pole magnitudes appear in the denominator—and proximity of another pole to s_r *increases* the magnitude of the coefficient K_r. When the zero s_n is moved so close to s_r that they *coincide*, the pole and zero *cancel* and reduce the value of the particular K_r to zero.

The magnitude of the K-coefficient corresponding to a particular pole is thus determined by the proximity of both poles and zeros. If, in the design of a network, the position of the poles and zeros can be selected, they should be selected according to the following pattern:

(1) Select pole locations to give the required time behavior. Do this in terms of complex frequencies.
(2) Fix the position of the zeros in the complex plane to adjust the magnitudes of the various K coefficients.

It should be noted that the graphical interpretation of the position of poles and zeros was discussed for the case of nonrepeated (or simple) poles. In the case of multiple poles, it is suggested that expansion by partial fractions be followed rather than seeking a modification of the procedures that have been discussed to fit the new case.

10-7. Procedure for finding network functions for general two-terminal-pair networks

For complicated two-terminal-pair networks, the computation of transfer functions and driving-point immittances may become quite involved. In this section, we will discuss systematic procedures for finding such network functions.

Any network can be thought of as made up of the combination of a number of one-terminal-pair networks. There is no unique rule for

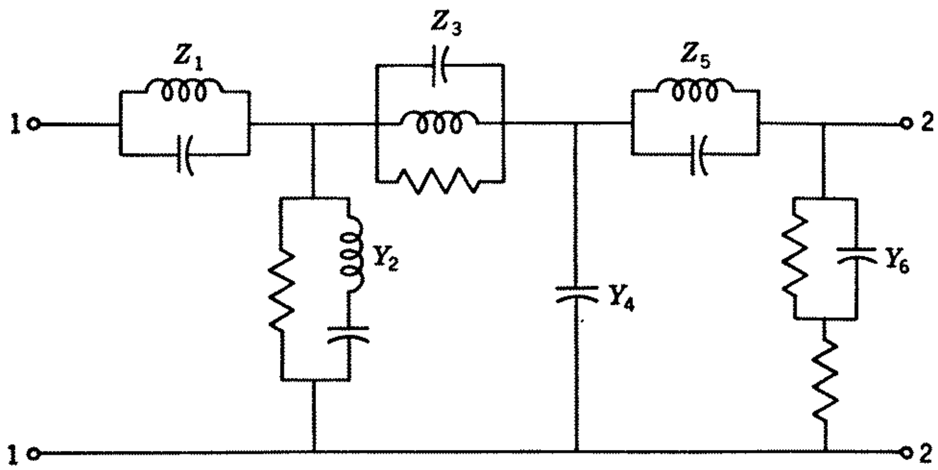


Fig. 10-16. Grouping of elements in a network to form a system of alternate impedances in series and admittances in parallel.

dividing the arbitrary network into elementary one-terminal-pair components. However, such a division is made on the basis of interest in the voltage of certain nodes or the current in certain branches in many cases. The network of Fig. 10-16, for example, is grouped into a number of one-terminal-pair networks. For each of the one-terminal-pair networks, the transform impedance $Z(s)$ or the transform admittance $Y(s)$ can be computed. This is illustrated for a number of examples in Fig. 10-17. Such combination is accomplished by the usual rules for series and parallel combination of immittances discussed earlier.

Several two-terminal-pair networks occur so often in useful networks that they are given special names. The general network sometimes reduces to a series impedance, a parallel impedance, and another series

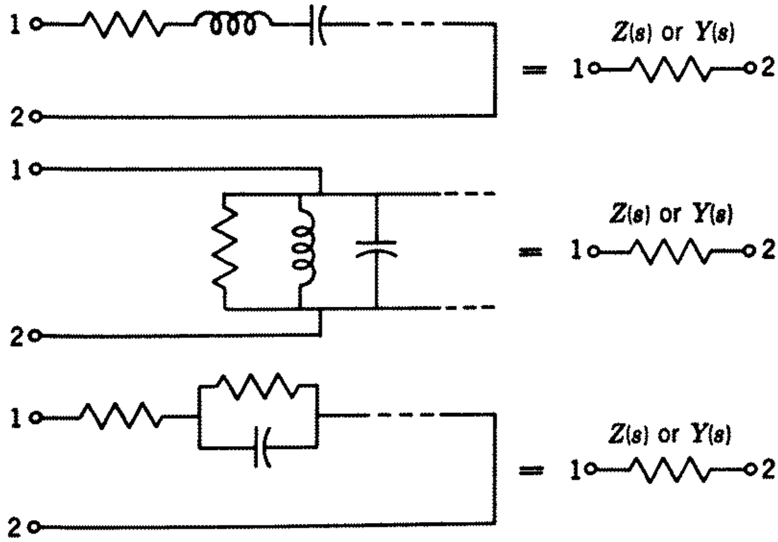


Fig. 10-17. Immittance of one-terminal-pair networks.

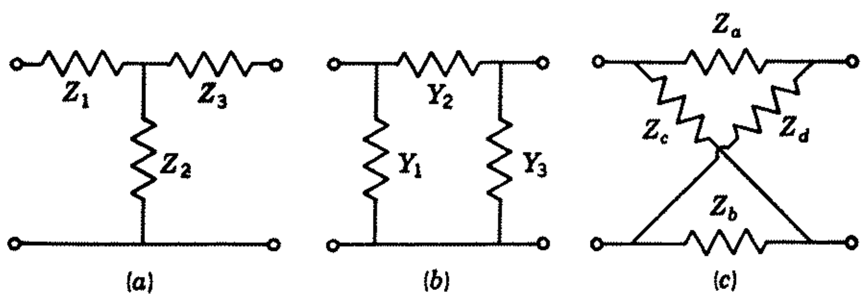


Fig. 10-18. Network configurations.

impedance as shown in Fig. 10-18(a). Such a network is known as a *T network*. Further, if the series impedances are equal, that is if $Z_1 = Z_3$, the network is designated as a *symmetrical T*. Figure 10-18 shows two other network configurations. The network of Fig. 10-18(b) is a *π network*, and with $Y_1 = Y_3$, the network is a *symmetrical π*. The network of Fig. 10-18(c) is a *lattice network* or a *symmetrical lattice* when $Z_c = Z_d$ and $Z_a = Z_b$.

If the two terminals of a terminal pair of a two-terminal-pair network are connected to a terminal pair of another two-terminal-pair network, the networks are said to be connected in *tandem* or *cascade*. If T networks or π networks are connected in cascade, the resulting network is an important network structure. It may contain any number of sections and may begin as shown in Fig. 10-19, or may begin with $Z_1(s) = 0$. The same thing may be said about the manner in which the network ends (at terminal pair 2). For convenience in computation, the series immittances are computed as impedances, and the parallel (or shunt) immittances are computed as admittances.

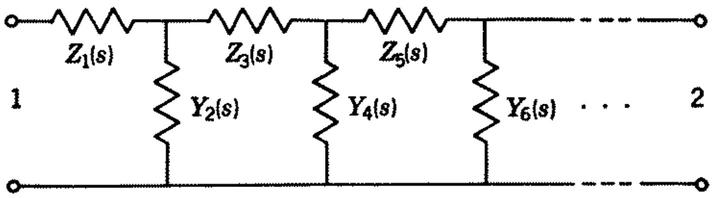


Fig. 10-19. Ladder network.

The driving-point immittance of any network can be found by writing loop or node equations. If the network can be made into a "ladder structure," it is possible to find the driving-point immittance by series and parallel combination of immittances without writing loop or node equations directly. Assume that the ladder network of Fig. 10-19 is made up of the six immittances shown. Combining immittances at the terminals *opposite* those for which the driving-point immittance is required, $Y_6(s)$ is first inverted and combined with $Z_5(s)$. Next this sum is inverted and combined with $Y_4(s)$. This pattern may be continued until the network reduces to a single immittance. In summary,

$$Z_{dp}(s) = Z_1(s) + \cfrac{1}{Y_2(s) + \cfrac{1}{Z_3(s) + \cfrac{1}{Y_4(s) + \cfrac{1}{Z_5(s) + \cfrac{1}{Y_6(s)}}}}} \qquad (10\text{-}50)$$

which is read from the bottom to the top to give the pattern of combining immittances. Such an algebraic configuration is called a *continued fraction* or a *Stieltjes continued fraction*. Forming a continued fraction for a ladder network provides a systematic procedure for finding the driving-point immittance.

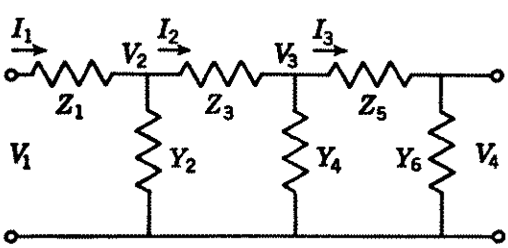


Fig. 10-20. Two-terminal-pair network.

The use of a continued fraction for network reduction applies only when the network function of interest is a driving-point immittance. When the transfer function is desired, a different procedure must be followed. In Fig. 10-20, the ladder network of Fig. 10-19 is redrawn with several pertinent voltages and currents designated. As in finding driving-point immittance, the following procedure requires that computations begin at the terminals opposite the driving-point terminals. For the network, we may write the following equations for Kirchhoff's voltage and current laws:

$$I_3 = Y_6 V_4 \tag{10-51}$$

$$V_3 = V_4 + Z_5 I_3 \tag{10-52}$$

$$I_2 = I_3 + Y_4 V_3 \tag{10-53}$$

$$V_2 = V_3 + Z_3 I_2 \tag{10-54}$$

$$I_1 = I_2 + Y_2 V_2 \tag{10-55}$$

$$V_1 = V_2 + Z_1 I_1 \tag{10-56}$$

These equations describe the network and they contain the usual transfer quantities, $G(s) = V_4(s)/V_1(s)$, $Z_{41}(s) = V_4(s)/I_1(s)$, etc. Starting with the first equation and substituting it into the second equation gives

$$V_3 = (1 + Y_6 Z_5) V_4 \tag{10-57}$$

In turn, this equation may be substituted with Eq. 10-51 into the next equation to give

$$I_2 = [Y_6 + Y_4(1 + Y_6 Z_5)] V_4 \tag{10-58}$$

Continuing according to this pattern,

$$V_2 = \{(1 + Y_6 Z_5) + Z_3[Y_6 + Y_4(1 + Y_6 Z_5)]\} V_4 \tag{10-59}$$

$$I_1 = ([Y_6 + Y_4(1 + Y_6 Z_5)] + Y_2\{(1 + Y_6 Z_5) + Z_3[Y_6 + Y_4(1 + Y_6 Z_5)]\}) V_4 \tag{10-60}$$

This equation gives the transfer impedance $Z_{41}(s)$ as the inverse of the admittance $Y_{41}(s)$, which is

$$Y_{41}(s) = \frac{I_1(s)}{V_4(s)} = Y_6 + Y_4\delta + Y_2[\delta + Z_3(Y_6 + Y_4\delta)] \tag{10-61}$$

where

$$\delta = 1 + Y_6 Z_5 \tag{10-62}$$

The transfer function for the voltage ratio may be found by carrying this procedure one step further by substituting into Eq. 10-56; that is,

$$\begin{aligned} G(s) &= \frac{V_1(s)}{V_4(s)} \\ &= \delta + Z_3(Y_6 + Y_4\delta) + Z_1\{Y_6 + Y_4\delta + Y_2[\delta + Z_3(Y_6 + Y_4\delta)]\} \end{aligned} \tag{10-63}$$

The transfer function $V_4(s)/V_1(s)$ is the reciprocal of $G(s)$ as given. The method illustrated above holds for any number of sections in the ladder network. The general pattern has been established by this example.

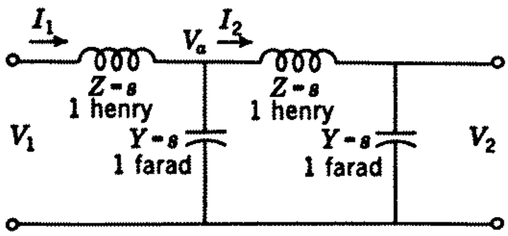


Fig. 10-21. Two-terminal-pair network.

To illustrate the method with a specific example, consider the network of Fig. 10-21, which is made up of series inductances (1 henry) and shunt capacitors (1 farad). The impedance of the series elements is s, and likewise, the admittance of the shunt elements is s. For this network, the driving-point impedance and the transfer function of the voltage ratio will be found. Other currents and voltages that will aid in the computation (but not appear in the solution) are shown on the figure. The driving-point impedance, written in continued-fraction form, is

$$Z_{dp}(s) = s + \cfrac{1}{s + \cfrac{1}{s + \cfrac{1}{s}}} \tag{10-64}$$

This equation can be reduced, starting at the bottom and working up, to the form

$$Z_{dp}(s) = \frac{s^4 + 3s^2 + 1}{s^3 + 2s} \tag{10-65}$$

To find the voltage ratio transfer function start at the V_2 terminals and proceed as

$$I_2(s) = YV_2 = sV_2 \tag{10-66}$$

$$V_a(s) = V_2 + I_2Z = (s^2 + 1)V_2 \tag{10-67}$$

$$I_1(s) = I_2 + YV_a = [s + s(s^2 + 1)]V_2 \tag{10-68}$$

$$V_1(s) = V_a + ZI_1 = (s^2 + 1)V_2 + s(s^3 + 2s)V_2 \tag{10-69}$$

The voltage ratio transfer function thus becomes

$$\frac{V_2(s)}{V_1(s)} = \frac{1}{s^4 + 3s^2 + 1} \tag{10-70}$$

for this particular network. We will show in another chapter that this network behaves as a low-pass filter.

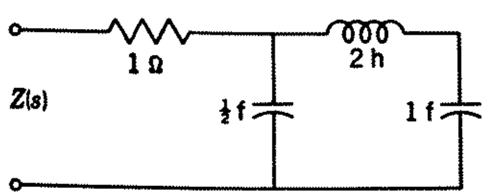
FURTHER READING

For additional discussion relating to network functions, the reader is referred to Gardner and Barnes, *Transients in Linear Systems* (John

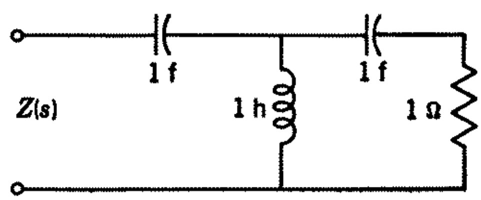
Wiley & Sons, Inc., New York, 1942), Chap. 5; to Valley and Wallman, *Vacuum Tube Amplifiers* (Vol. 18 of the Radiation Laboratory series, McGraw-Hill Book Co., Inc., New York, 1948), pp. 42–53; to Tuttle, *Network Synthesis*, 2 vols. (John Wiley & Sons, Inc., New York, in preparation); and to Guillemin, *Communications Networks, Vol. I* (John Wiley & Sons, Inc., New York, 1932). It should be noted that the quantity s is equivalent to p and λ as used by some authors.

PROBLEMS

10-1. Find the driving-point impedance for the network shown in the figure. Arrange the polynomials of this function with the highest ordered term normalized to unity coefficient. *Answer.* $Z(s) = \dfrac{s^3 + 2s^2 + \frac{3}{2}s + 1}{s^3 + \frac{3}{2}s}$.



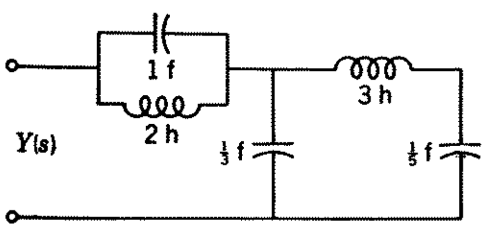
Prob. 10-1.



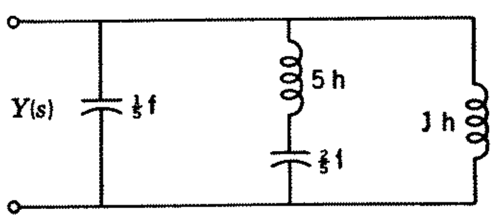
Prob. 10-2.

10-2. Repeat Prob. 10-1 for the network shown in the figure. *Answer.* $Z(s) = \dfrac{s^3 + 2s^2 + s + 1}{s^3 + s^2 + s}$.

10-3. Find the driving-point admittance for the network shown in the accompanying figure. Arrange the polynomials with the highest-ordered term normalized to unity coefficient. *Answer.* $Z(s) = 4\,\dfrac{s^4 + \frac{55}{24}s^2 + \frac{5}{8}}{s(s^4 + \frac{19}{6}s^2 + \frac{4}{3})}$.



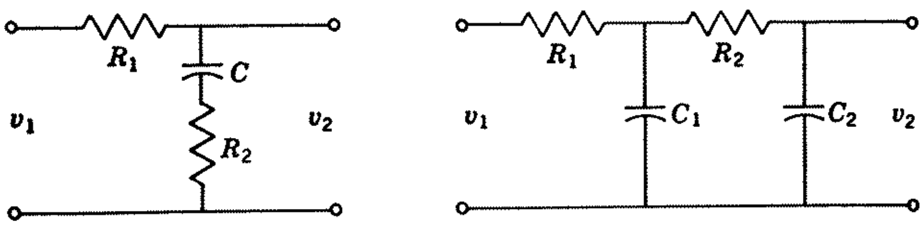
Prob. 10-3.



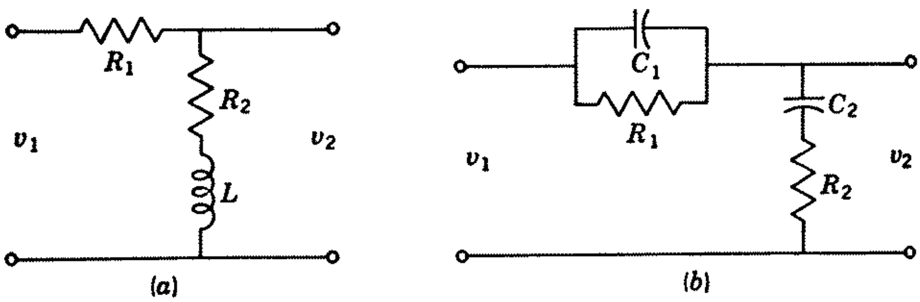
Prob. 10-4.

10-4. Repeat Prob. 10-3 for the network shown above. *Answer.* $Y(s) = \dfrac{s^4 + \frac{13}{2}s^2 + \frac{5}{2}}{\frac{5}{2}s(2s^2 + 1)}$.

10-5. Find the transfer function, the output voltage to input voltage ratio, for the networks shown in the figure. Arrange the polynomials with the highest-ordered term normalized to unity coefficient. *Answer to* (*b*). $G(s) = \dfrac{1/R_1R_2C_1C_2}{s^2 + (R_1C_1 + R_1C_2 + R_2C_2)s/R_1R_2C_1C_2 + 1/R_1R_2C_1C_2}$.



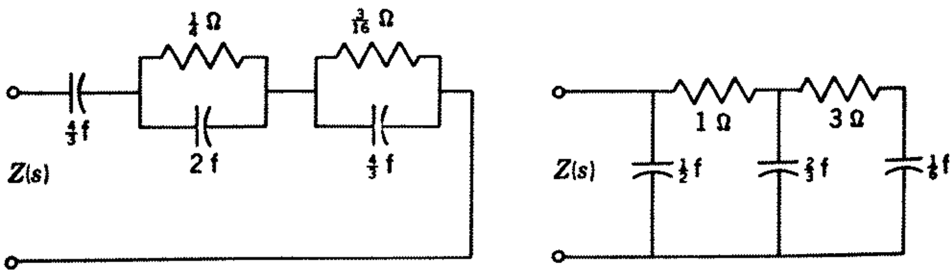
Prob. 10-5.



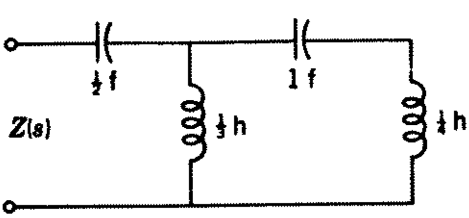
Prob. 10-6.

10-6. Repeat Prob. 10-5 for the networks shown. *Answer to* (*b*). $\left[\dfrac{s^2 + (R_1C_1 + R_2C_2)s/R_1R_2C_1C_2 + 1/R_1R_2C_1C_2}{s^2 + (R_1C_1 + R_1C_2 + R_2C_2)s/R_1R_2C_1C_2 + 1/R_1R_2C_1C_2}\right]$.

10-7. Show that both of the networks of the figure have the same driving-point impedance $Z(s) = 2\left[\dfrac{(s+1)(s+3)}{s(s+2)(s+4)}\right]$.



Prob. 10-7.

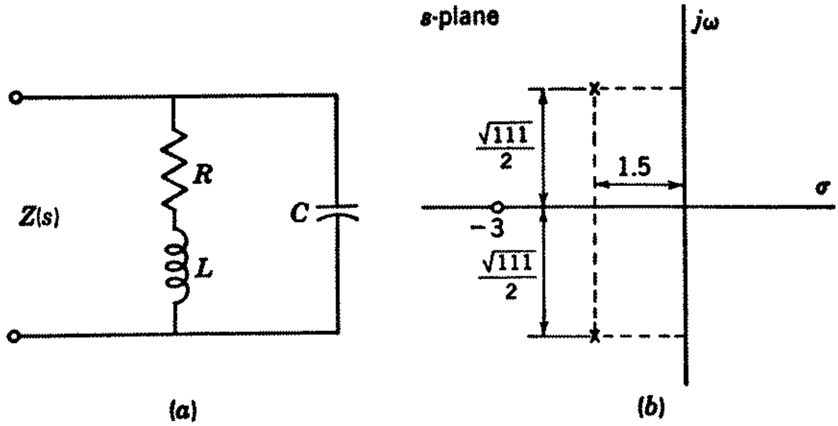


Prob. 10-8.

10-8. Show that the network of the accompanying figure has the driving-point impedance $Z(s) = \dfrac{s^4 + 18s^2 + 24}{7s^3 + 12s}$.

10-9. Prove that the total number of poles is equal to the total number of zeros for any network function having the form of quotient of rational polynomials.

10-10. The network shown below is known to have the pole-zero configuration shown in the figure. In addition, it is known that the



Prob. 10-10.

impedance at zero frequency ($s = 0$ or direct current) is 1 ohm, that is $Z(0) = 1$. Determine the values of R, L, and C in the network. *Answer.* $R = 1$ ohm, $L = \frac{1}{3}$ henry, $C = 0.1$ farad.

10-11. It is known that the response in the time domain of a system is the summation of terms having the following characteristics: (a) $\omega_n = 2$, $\zeta = 0.5$ (second order); (b) T (the time constant) $= 3$ sec; (c) $\omega_n = 1$, $\zeta = 0$. Plot the poles of this system in the s plane.

10-12. A transient is found to be of the form

$$i(t) = 2e^{-t} - 1e^{-5t}$$

Find a pole-zero configuration for $I(s)$ that gives this time-domain response.

10-13. A transient is found to be of the form

$$i(t) = \tfrac{7}{4}e^{-t} - 2e^{-3t} + \tfrac{1}{4}e^{-5t}$$

Find a pole-zero configuration for $I(s)$ that gives this time domain response.

10-14. Given that the scale factor of a current transform $I(s)$ has the value 10 and that the following finite poles and zeros describe the system:

Poles	Zeros
$-2 \pm j1$	none
-5	

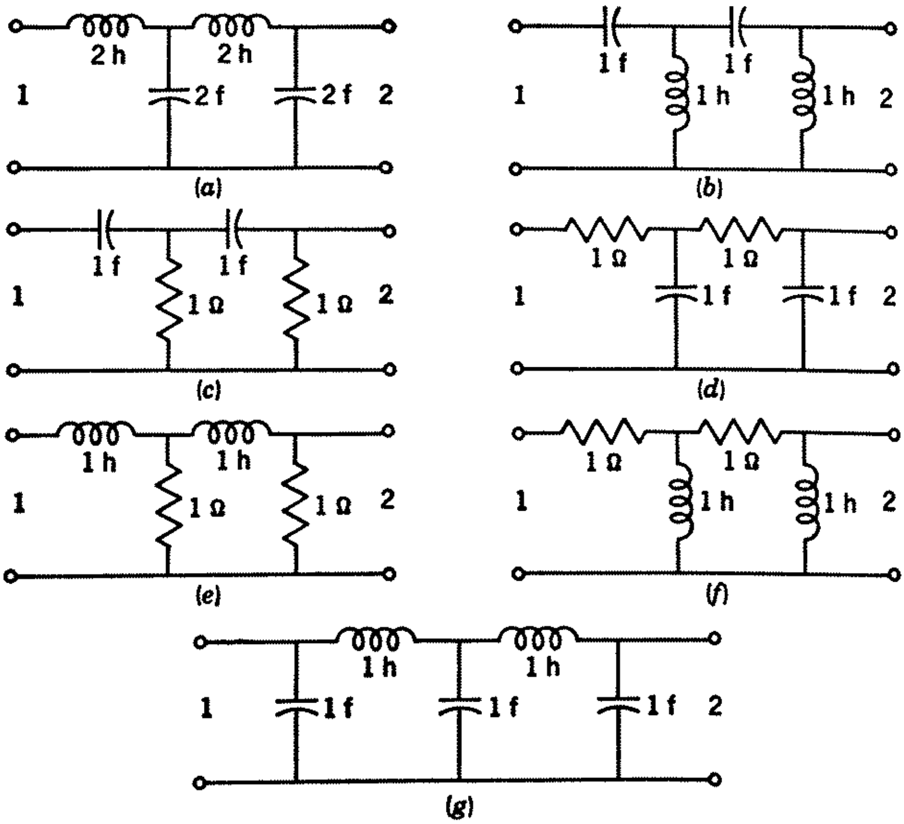
Find the time-domain response corresponding to this $I(s)$. *Answer.* $i(t) = e^{-5t} + \sqrt{10}\, e^{-2t} \cos\,(t - 108.4°)$.

10-15. Given that the scale factor of a current transform $I(s)$ has the value 5.0 and that the following finite poles and zeros describe the system.

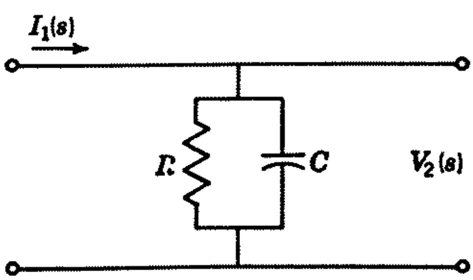
Poles	Zeros
$-1 \pm j2$	-6
-3	

Find the time-domain response $i(t)$ corresponding to this $I(s)$.

10-16. For each of the networks shown in the figure, find the transfer impedance $Z_{tr}(s) = V_{out}(s)/I_{in}(s)$ and the voltage ratio transfer function $G(s) = V_{out}(s)/V_{in}(s)$.



Prob. 10-16.



Prob. 10-17.

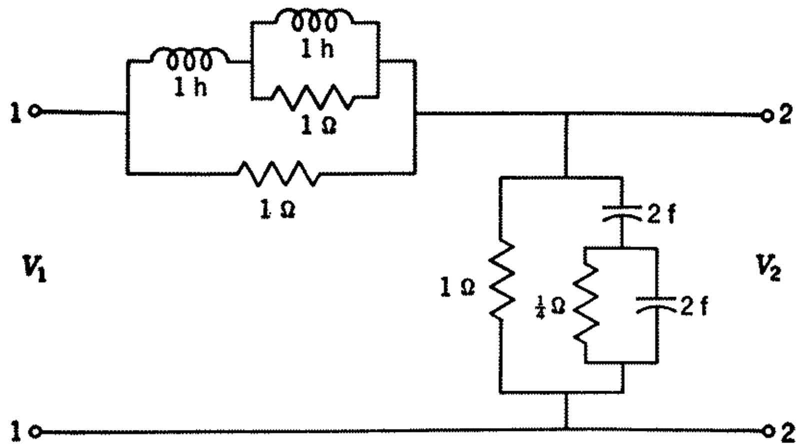
10-17. The network shown in the figure is driven by a current source I_1. The output voltage is V_2. (a) Find the transfer impedance, $Z_{21}(s) = V_2(s)/I_1(s)$. (b) Show the pole-zero configuration for $Z_{21}(s)$.

10-18. For a given network, it is known that

$$Z_{21} = V_2/I_1 = \frac{(s - s_1)(s - s_1)}{s}$$

where s_1, $s_1 = -1 \pm j10$. If $i_1(t) = e^{-0.5t}$, find $v_2(t)$.

10-19. (a) For the network shown, show that the input impedance at terminal-pair 1 is $Z_{11}(s) = 1$ ohm. (b) Find the transfer function $G(s) = V_2(s)/V_1(s)$ for this network.



Prob. 10-19.

10-20. Two conjugate complex poles are required to meet the various specifications given below. For each specification, sketch the region in the s plane (using crosshatching for identification) that the poles may be located. (a) $\zeta \geqq 0.707$, $\omega_n \geqq 1$, $\zeta\omega_n \geqq -4$. (b) $0 \leqq \zeta \leqq 0.5$, actual frequency $\leqq 2$, $\zeta\omega_n$ negative. (c) $1 \leqq \omega_n \leqq 4$, $\zeta\omega_n \leqq 0.5$. (d) $0.5 \leqq \zeta \leqq 0.866$, $\omega_n \leqq 2.5$.

CHAPTER 11

SINUSOIDAL STEADY-STATE ANALYSIS FROM POLE-ZERO CONFIGURATIONS

There is something distinctive about the sinusoidal waveform. If a sinusoidal driving force is applied to a network of linear passive elements, every voltage and every current in that network will be sinusoidal in the steady state, differing from the driving-force sinusoid only in amplitude and phase angle. This property follows from two facts:

(1) The sinusoid may be repeatedly differentiated or integrated and still be a sinusoid of the same frequency.
(2) The sum of a number of sinusoids of one frequency with arbitrary amplitudes and phase angles is a sinusoid of the same frequency.

In addition to this mathematical distinction, the sinusoid is generated rather commonly in nature: a bottle bobbing in the water, a pendulum, the shadow of a crank-handle on a wheel—all these devices describe sinusoidal motion. A sinusoidal voltage is generated by a conductor constrained to move in a circular path at right angles to a magnetic field.

Analysis under the assumption of a sinusoidal driving force and a steady state is used in such fields as electronics, network theory, and servomechanisms. In these fields, however, the driving forces are seldom sinusoidal. We might rightfully question how valid such analysis is. Part of the justification of this method stems from Fourier analysis: periodic waveforms can be approximated by a finite sum of sinusoids. Further, nonperiodic (and nonrecurring) waveforms can be expressed in terms of sinusoids by use of the Fourier integral. This concept of analysis in terms of harmonic frequency components allows the response of a network to a nonsinusoidal waveform to be predicted from a known response as a function of frequency.

In this section, we will develop the relationship between the general solution of a network problem and the solution for the sinusoidal steady state. This will be accomplished in terms of the pole-zero configuration of network functions.

11-1. Radian frequency and the sinusoid

The term *sinusoid* includes the sine wave, cosine wave, or either the sine or the cosine with a phase angle. The transforms of the sine and the cosine are

$$\mathcal{L} \sin \omega t = \frac{\omega}{s^2 + \omega^2}, \qquad \mathcal{L} \cos \omega t = \frac{s}{s^2 + \omega^2} \tag{11-1}$$

The poles and zeros for these transform equations, as shown in Fig. 11-1, appear on the $j\omega$ axis. Such frequencies have been defined as

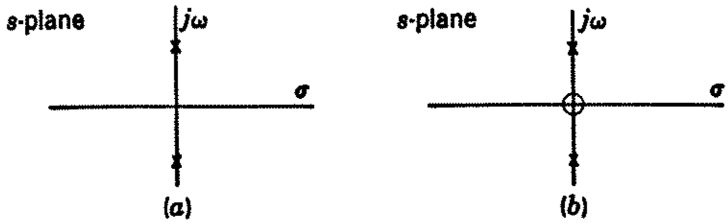


Fig. 11-1. Pole-zero configurations for sinusoids: (a) sine wave; (b) cosine wave.

radian frequencies. Frequencies described by positions on the $j\omega$ axis of the s plane represent pure radian frequencies such as occur in the sinusoidal steady state and correspond to the time-domain factors $e^{j\omega t}$ and $e^{-j\omega t}$.

Sine and cosine functions are related to exponential factors by the equations

$$\sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j} \tag{11-2}$$

$$\cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \tag{11-3}$$

The term $e^{j\omega t}$ is commonly interpreted in terms of a unit rotating phasor* rotating in the positive (or counterclockwise) direction; $e^{-j\omega t}$

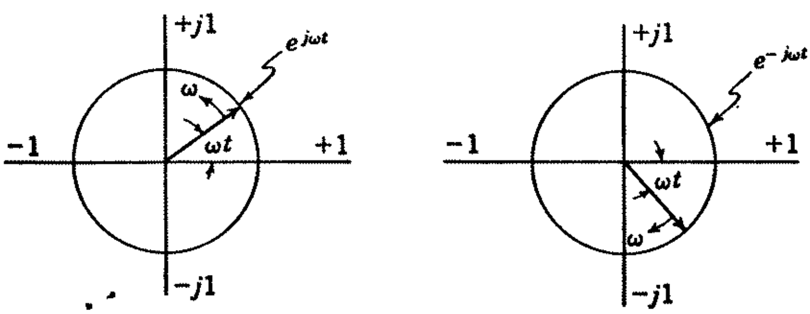


Fig. 11-2. Unit rotating phasors.

likewise is interpreted as a unit rotating phasor rotating in the negative (clockwise) direction. The unit phasors are illustrated in Fig. 11-2.

* For those who prefer, the term *phasor* may be read as *vector*.

Now the sinusoid, according to Eq. 11-2, is made up of the difference of two rotating unit phasors, rotating in opposite directions, divided by the factor $(2j)$. The construction of a sine wave in terms of these unit exponentials is illustrated in Fig. 11-3. The combination of the

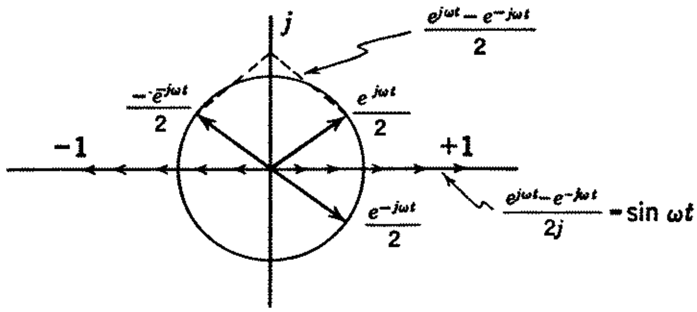


Fig. 11-3. The sine wave from rotating phasors.

phasor $(e^{j\omega t}/2)$ and $(-e^{-j\omega t}/2)$ gives a phasor on the $j\omega$ axis. The factor $(1/j) = -j$ corresponds to a negative rotation of 90° ($-\pi/2$ radians). The sine of ωt is a real number (on the axis of reals); it has a value of zero when $\omega t = 0$, and a value of unity when $\omega t = \pi/2$. As ωt increases from 0 to 2π, the sine function is seen to have values between the limits of 1 and -1.

The cosine function may be similarly constructed in terms of exponential factors as is illustrated in Fig. 11-4. The cosine is also a real

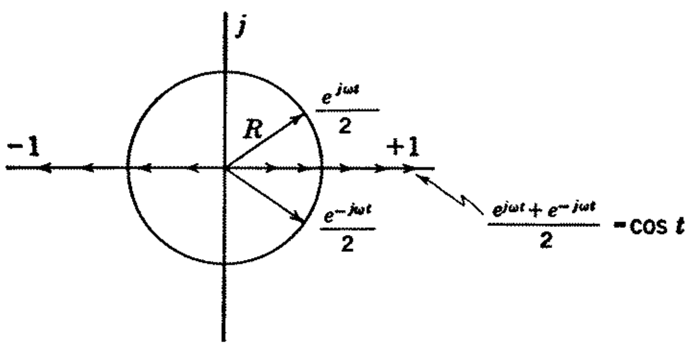


Fig. 11-4. The cosine wave from rotating phasors.

number having a total variation from $+1$ to -1. When $\omega t = 0$, the cosine has a value of unity; when $\omega t = \pi/2$, the cosine has zero value. Both the cosine and the sine are generated by two "frequencies": $+j\omega$ and $-j\omega$. This is also shown from the pole locations of Fig. 11-1.

The exponential factors corresponding to the cosine or sine terms can be used in computing impedance for the sinusoidal steady state. Consider a series RL circuit with a cosine driving force given as

$$V \cos \omega t = V \left(\frac{e^{j\omega t}}{2} + \frac{e^{-j\omega t}}{2} \right) \tag{11-4}$$

The single cosine generator generating $v(t) = V \cos \omega t$ is thus seen to be equivalent to *two* generators, one generating $(V/2)e^{j\omega t}$, the other generating $(V/2)e^{-j\omega t}$. Using the principle of superposition, we may consider the driving forces separately and then combine the resulting currents to obtain the final solution. For the first generator, the differential equation becomes

$$L\frac{di}{dt} + Ri = \frac{V}{2}e^{j\omega t} \tag{11-5}$$

The steady-state part of the solution (the particular integral) will be of the form $i_{ss}(t) = Ae^{j\omega t}$. Substituting this solution into the equation gives

$$j\omega LA + RA = V/2 \tag{11-6}$$

or

$$A = \frac{V/2}{R + j\omega L} = \frac{V/2}{Z}$$

where Z is the impedance for the sinusoidal steady state. Similarly, we may let $Be^{-j\omega t}$ be the steady-state solution of Eq. 11-5 with $e^{-j\omega t}$ replacing $e^{j\omega t}$ to give

$$B = \frac{V/2}{R - j\omega L} = \frac{V/2}{Z^*} \tag{11-7}$$

The total solution for the steady state becomes

$$i_{ss}(t) = \frac{V}{2}\left(\frac{e^{j\omega t}}{Z} + \frac{e^{-j\omega t}}{Z^*}\right) \tag{11-8}$$

If V/Z is defined as I, this equation may be written in the form

$$i_{ss}(t) = \tfrac{1}{2}(Ie^{j\omega t} + I^*e^{-j\omega t}) \tag{11-9}$$

In this equation I is a complex number, I^* is the conjugate of this complex number, and the exponential factors $e^{j\omega t}$ and $e^{-j\omega t}$ are complex numbers relating to the cosine and sine functions according to Euler's equation,

$$e^{\pm j\omega t} = \cos \omega t \pm j \sin \omega t \tag{11-10}$$

If we let $I = a + jb$, Eq. 11-9 reduces to the form

$$i_{ss}(t) = a \cos \omega t - b \sin \omega t \tag{11-11}$$

$$= \text{Re}\,[(a + jb)(\cos \omega t + j \sin \omega t)] \tag{11-12}$$

where the letters Re mean "*the real part of*" (similarly, Im means "the imaginary part of"). But $(a + jb) = I$ and $(\cos \omega t + j \sin \omega t)$ is defined by Euler's equation as $e^{j\omega t}$. Hence

$$i_{ss}(t) = \text{Re}\,(Ie^{j\omega t}) \tag{11-13}$$

or

$$i_{ss}(t) = \text{Re}\left(\frac{V}{Z}\,e^{j\omega t}\right) \tag{11-14}$$

This last equation tells us that we may use the exponential in place of the sinusoid for the steady-state solution providing we *take only the real part of the solution.* Since the exponential is easily differentiated and integrated, this method of the solution is convenient. Provided it is always understood that only the real part has meaning, currents and voltages may be written in the forms

$$i(t) = Ie^{j\omega t} \tag{11-15}$$

$$v(t) = Ve^{j\omega t} \tag{11-16}$$

To illustrate the use of the exponential equivalent of the sinusoid, suppose that we consider the differential equation for an RLC series circuit given as

$$L\frac{di}{dt} + Ri + \frac{1}{C}\int i\,dt = Ve^{j\omega t} \tag{11-17}$$

The form of the solution must be $Ie^{j\omega t}$. Performing the required differentiation and integration gives

$$\left(j\omega L + R + \frac{1}{j\omega C}\right) I = V \tag{11-18}$$

The sinusoidal impedance is defined as

$$Z(j\omega) = \frac{V}{I} = R + j\left(\omega L - \frac{1}{\omega C}\right) \tag{11-19}$$

The current is given as

$$i(t) = \frac{V}{Z}\,e^{j\omega t} \tag{11-20}$$

provided only the real part of this expression is taken; that is,

$$i(t) = \text{Re}\left(\frac{V}{Z}\,e^{j\omega t}\right) \tag{11-21}$$

For this example,

$$i(t) = \text{Re}\left[\frac{V}{R + j(\omega L - 1/\omega C)}\,e^{j\omega t}\right] \tag{11-22}$$

$$i(t) = \text{Re}\left\{\frac{V}{R^2 + (\omega L - 1/\omega C)^2}\left[R - j\left(\omega L - \frac{1}{\omega C}\right)\right](\cos\omega t + j\sin\omega t)\right\} \tag{11-23}$$

Since $Z = R + j(\omega L - 1/\omega C)$, the real part of $i(t)$ in this equation is

$$i(t) = \frac{V}{|Z|^2}\left[R \cos \omega t + \left(\omega L - \frac{1}{\omega C}\right) \sin \omega t\right] \tag{11-24}$$

or, finally,

$$i(t) = \frac{V}{|Z|} \cos \left(\omega t - \tan^{-1} \frac{\omega L - 1/\omega C}{R}\right) \tag{11-25}$$

This same exponential factor may be used to represent a sinusoid at a phase angle; for example, let

$$v(t) = V \cos (\omega t + \phi) \tag{11-26}$$

In exponential form, this equation becomes

$$v(t) = V\left[\frac{e^{j(\omega t+\phi)} + e^{-j(\omega t+\phi)}}{2}\right] \tag{11-27}$$

or

$$v(t) = \tfrac{1}{2}[(Ve^{j\phi})e^{j\omega t} + (Ve^{-j\phi})e^{-j\omega t}] \tag{11-28}$$

The quantity $(Ve^{j\phi})$ is a phasor of magnitude V and phase angle ϕ, which will be represented by the notation $\mathbf{V}$. Then

$$v(t) = \tfrac{1}{2}(\mathbf{V}e^{j\omega t} + \mathbf{V}^*e^{-j\omega t}) \tag{11-29}$$

This equation is of the same form as Eq. 11-9, and by the same reasoning as previously given is equivalent to

$$v(t) = \text{Re } (\mathbf{V}e^{j\omega t}) \tag{11-30}$$

This exponential factor may be used in place of Eq. 11-26 to give the same result with less mathematical manipulation. Again, it is not necessary to carry the "real part of" notation so long as the requirement that only the real part of the result has meaning is kept in mind.

In this section, we have seen that the sine and the cosine correspond to exponential factors and have complex frequencies located on the $j\omega$ axis of the s plane.

11-2. Magnitude and phase of network functions

All network functions may be written as a quotient of polynomials in s; in general form,

$$\left.\begin{matrix} G(s) \\ Z(s) \end{matrix}\right\} = \frac{P(s)}{Q(s)} = \frac{a_0s^r + a_1s^{n-1} + \dots + a_n}{b_0s^m + b_1s^{m-1} + \dots + b_m} \tag{11-31}$$

For the sinusoidal steady state, $s = j\omega$, and the network function becomes

$$\left.\begin{matrix} G(j\omega) \\ Z(j\omega) \end{matrix}\right\} = \frac{P(j\omega)}{Q(j\omega)} = \frac{a_0(j\omega)^n + a_1(j\omega)^{n-1} + \dots}{b_0(j\omega)^m + b_1(j\omega)^{m-1} + \dots} \tag{11-32}$$

In this equation, terms will alternately be real and imaginary. Which terms of the general expression are real and which imaginary depends on whether n and m are even or odd. However, in every case it will be possible to write the quotient of polynomials in the form

$$\left.\begin{matrix} G(j\omega) \\ Z(j\omega) \end{matrix}\right\} = \frac{A(\omega) + jB(\omega)}{C(\omega) + jD(\omega)} = R(\omega) + jX(\omega) \tag{11-33}$$

where

$$\begin{aligned} A(\omega) &= \operatorname{Re} P(j\omega) \quad \text{and} \quad B(\omega) = \operatorname{Im} P(j\omega) \\ C(\omega) &= \operatorname{Re} Q(j\omega) \quad \text{and} \quad D(\omega) = \operatorname{Im} Q(j\omega) \end{aligned} \tag{11-34}$$

The quantities $R(\omega)$ and $X(\omega)$ are found by rationalizing the expression involving $A(\omega)$, $B(\omega)$, $C(\omega)$, and $D(\omega)$. Phase and magnitude of the general network function are defined in terms of R and X. The defining equation for the phase is

$$\phi(\omega) = \tan^{-1}\frac{X(\omega)}{R(\omega)} \tag{11-35}$$

and the defining equation for magnitude is

$$M(\omega) = \sqrt{[X(\omega)]^2 + [R(\omega)]^2} \tag{11-36}$$

The complex variable, $R(\omega) + jX(\omega)$, is thus defined in polar coordinates as

$$M(\omega)e^{j\phi(\omega)} \tag{11-37}$$

Alternately, the magnitude and phase of the network function may be computed directly from the quotient form of Eq. 11-33 as

$$[M(\omega)]^2 = \frac{[A(\omega)]^2 + [B(\omega)]^2}{[C(\omega)]^2 + [D(\omega)]^2} \tag{11-38}$$

$$\phi(\omega) = \tan^{-1}\frac{B(\omega)}{A(\omega)} - \tan^{-1}\frac{D(\omega)}{C(\omega)} \tag{11-39}$$

By either of the methods that have been described the magnitude and phase of a network function may be found as a function of frequency. The magnitude and phase characteristics of networks are important in network theory, partly because measurements of these quantities are easily made.

The problem before us is the computation and plotting of magnitude and phase as a function of frequency. The amount of computations can frequently be reduced by first considering the asymptotic values of these functions in terms of the original quotient of polynomial form given as Eq. 11-32.

High-Frequency Asymptotes. Assume that the network function being considered is a transfer function $G(j\omega)$. For large values of ω, only the highest-ordered terms in the numerator and denominator of $G(j\omega)$ are significant; that is,

$$\lim_{\omega \to \infty} G(j\omega) = \lim_{\omega \to \infty} H \frac{(j\omega)^n + \cdots}{(j\omega)^m + \cdots} \tag{11-40}$$

$$= \lim_{\omega \to \infty} H(j\omega)^{n-m} \tag{11-41}$$

The limit of this function depends on which number, n or m, is larger; that is,

$$\lim_{\omega \to \infty} G(j\omega) = \begin{cases} \infty\, e^{j(n-m)\pi/2} & n > m \\ 0\, e^{j(n-m)\pi/2} & n < m \\ H & n = m \end{cases} \tag{11-42}$$

The limiting value of the magnitude is zero, infinity, or a constant H. The angles in each case are some multiple of $\pi/2$ radians.

Low-Frequency Asymptotes. The low-frequency behavior of the network function is determined by the lowest-ordered terms in the quotient of polynomials. The important part of the network function for this case is

$$\frac{\cdots + a_{n-1}(j\omega) + a_n}{\cdots + b_{m-1}(j\omega) + b_m} \tag{11-43}$$

If neither a_n nor b_m is zero, the low-frequency asymptote is

$$\lim_{\omega \to 0} G(j\omega) = H \frac{a_n}{b_m} \tag{11-44}$$

However, if one or more terms are zero such as b_m, b_{m-1}, a_n, a_{n-1}, etc., then the network function may be written

$$\frac{H}{(j\omega)^p} \left[\frac{a_0'(j\omega)^n + \cdots + a_n'}{b_0'(j\omega)^m + \cdots + b_m'} \right] \tag{11-45}$$

where p may be positive or negative. The limit of this function as ω becomes small is

$$\lim_{\omega \to 0} G(j\omega) = \frac{H(a_n'/b_m')}{(j\omega)^p}$$

The limit of this function depends on whether p is positive, zero, or negative; that is,

$$\lim_{\omega \to 0} G(j\omega) = \begin{cases} \infty\, e^{-jp\pi/2}, & p > 0 \\ 0\, e^{-jp\pi/2}, & p < 0 \\ H(a_n'/b_m'), & p = 0 \end{cases} \tag{11-46}$$

Again, the limiting value of the magnitude is zero, infinity, or a constant, while the angle is some multiple of $(\pi/2)$ radians.

In practice, the phase and magnitude information is plotted in two ways: a polar coordinate plot, and separate plots of M and ϕ against frequency. These two types of plots are illustrated in Fig. 11-5. The polar plot can be made in terms of either $M(\omega)$ and $\phi(\omega)$ or the imaginary part and real part of $G(j\omega)$.

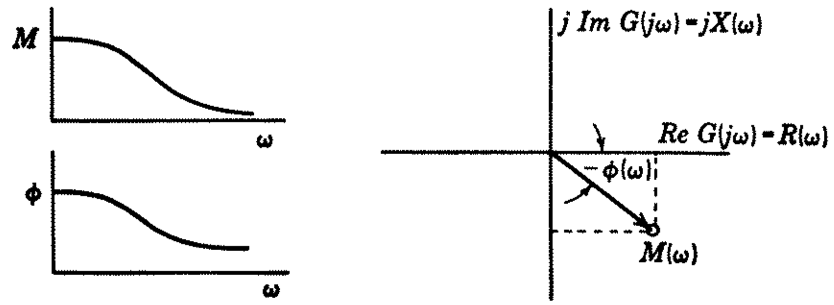


Fig. 11-5. Plotting of phase and magnitude.

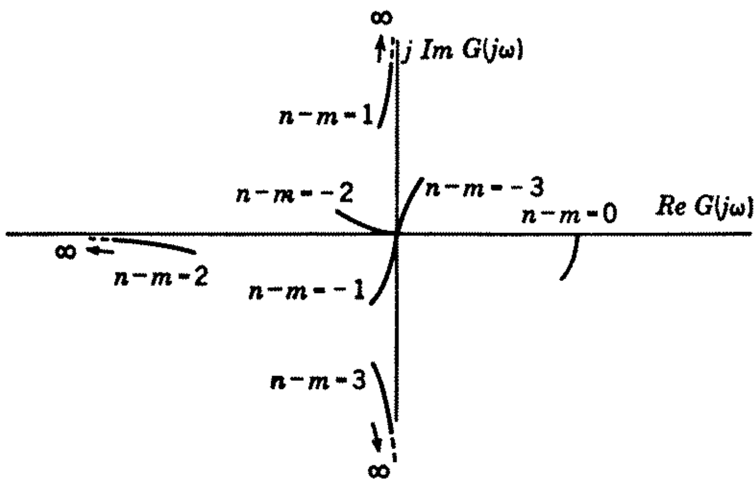


Fig. 11-6. High frequency asymptotes of $G(j\omega)$, where m = order of denominator and n = order of numerator.

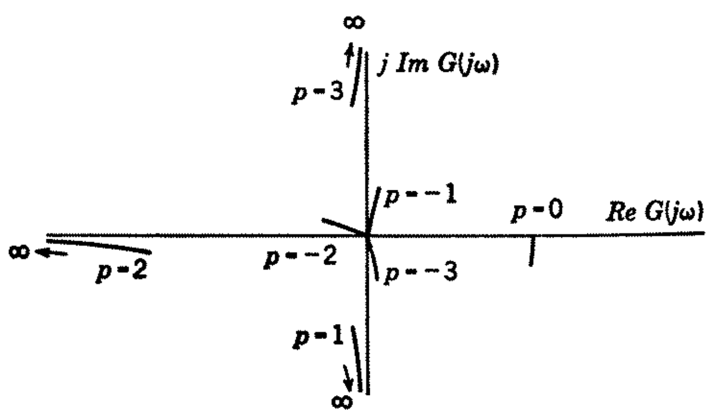


Fig. 11-7. Low-frequency asymptotes of $G(j\omega)$.

Plots on the $M(\omega)$ and $\phi(\omega)$ coordinates are made as continuous curves. The quantity $M(\omega)e^{j\phi(\omega)}$, however, is usually thought of as a phasor represented by an *arrow* as shown in Fig. 11-5. To avoid confusion, only the "tip" of the phasor is plotted. The locus of the "tip" of the phasor is known as the *phasor locus of the network function*. The asymptotic values of the network functions for low and high frequency are shown in Fig. 11-6 and Fig. 11-7.

Thus the low-frequency and high-frequency behavior of a network function can be determined by inspection. There remains the tedious task of computing intermediate points. A number of examples will illustrate the general procedure.

Example 1

A two-terminal-pair network made up of one resistor and one capacitor is shown in Fig. 11-8. The voltage ratio transfer function is

$$G(s) = \frac{V_2(s)}{V_1(s)} = \frac{1/Cs}{R + 1/Cs} = \frac{1}{RCs + 1} \tag{11-47}$$

The transfer function for the sinusoidal steady state is found by letting $s = j\omega$; thus

$$G(j\omega) = \frac{1}{j\omega RC + 1} = \frac{1/RC}{j\omega + 1/RC} \tag{11-48}$$

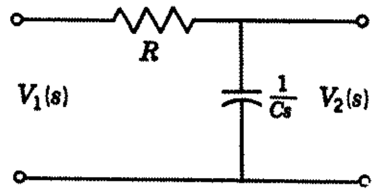


Fig. 11-8. Two-terminal-pair RC network.

For low frequencies $G(j\omega) \to 1$, while for high frequencies, the asymptotic value becomes $0 \times e^{-j\pi/2}$; that is, zero magnitude and $-90°$ phase angle. One other frequency is especially convenient for computation: $\omega = 1/RC$. At this frequency, the magnitude is 0.707 and the phase is $-45°$. This information is summarized in Table 11-1.

TABLE 11-1

ω	$G(j\omega)$
0	1 at 0°
$1/RC$	0.707 at $-45°$
∞	0 at $-90°$

The complete phasor locus is that of a circle as shown in Fig. 11-9. The equivalent $M(\omega)$ and $\phi(\omega)$ plots are also shown in the figure.

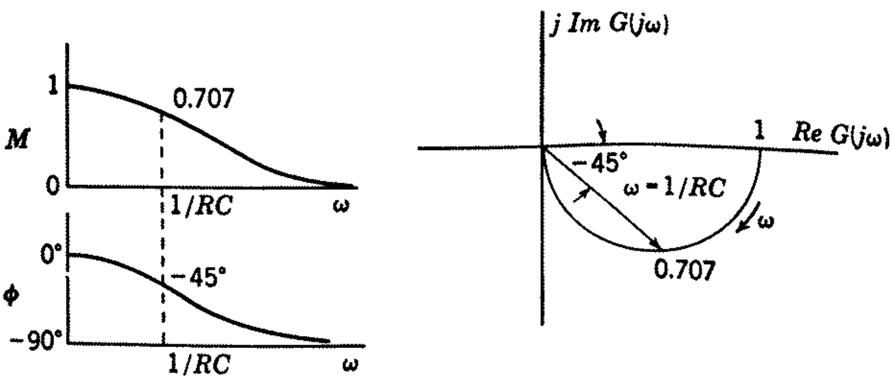


Fig. 11-9. RC network characteristics.

Example 2

If the resistor and the capacitor of the RC network used for Example 1 are interchanged, there results the two-terminal-pair network shown in Fig. 11-10. Again, the voltage ratio transfer function is to be studied. This function has the value

$$G(s) = \frac{R}{R + 1/Cs} = \frac{RCs}{RCs + 1} = \frac{s}{s + 1/RC} \tag{11-49}$$

corresponding to a zero at $s = 0$ and a pole at $s = -1/RC$. Letting $s = j\omega$ gives the transfer function for the sinusoidal steady state as

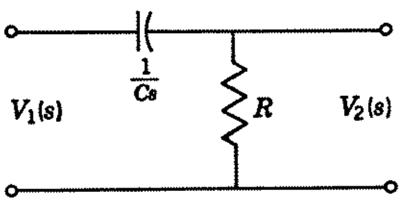


Fig. 11-10. Network of Example 2.

$$G(j\omega) = \frac{j\omega}{j\omega + 1/RC} \tag{11-50}$$

This function will be examined for low-frequency and high-frequency asymptotes. As ω becomes large, the term $1/RC$ can be neglected, and $G(j\omega)$ approaches unity (alternately, l'Hospital's rule can be applied). For small values of ω, $G(j\omega)$ approaches zero magnitude and 90° phase angle. Again, one frequency causes the function to reduce to an especially simple form: $\omega = 1/RC$. For that frequency, $G(j\omega) = 0.707\ e^{j\pi/4}$. In tabular form, these computations may be summarized as follows:

TABLE 11-2

ω	$G(j\omega)$
0	0 at +90°
$1/RC$	0.707 at +45°
∞	1 at 0°

These values serve as a guide to the computation. In order to make the complete plot, several other values will have to be found. The complete phasor locus is that of a circle, as shown in Fig. 11-11. The equivalent $M(\omega)$ and $\phi(\omega)$ plots for this same function are also shown in the figure. The two plotting systems display the same information. With practice, it will be possible to visualize one form from an inspection of the other.

Comparing the two networks of Example 1 and Example 2, it is seen that the first provides positive phase shift (or phase lead) for all frequencies, while the latter provides negative phase shift (or phase lag) for all frequencies. For the first, the output per unit input is high at low frequencies and low at high frequencies. The opposite behavior takes place in the second network. This result can be correlated with the behavior of the individual elements of the two networks. In the network of Fig. 11-8, the capacitor acts as an open circuit at low frequencies and hence the same voltage appears on the output terminals

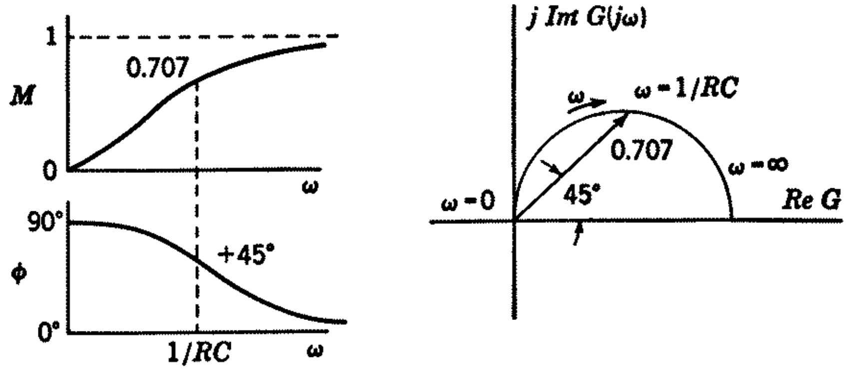


Fig. 11-11. *RC* network characteristics.

as appears on the input terminals. At high frequencies, however, the capacitor acts as a short circuit and the output voltage approaches zero magnitude (being the drop across the capacitor). For the network of Fig. 11-10, the capacitor acts as an open circuit for low frequencies, so that there is *no* output voltage; however, at high frequencies the capacitor behaves as a short circuit, causing approximately the same voltage to appear on the output terminals as appears on the input terminals.

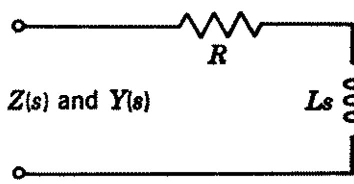


Fig. 11-12. *RL* network.

Example 3

As the third example, consider the driving-point immittance of a series RL network shown in Fig. 11-12. The immittance functions have the forms

$$Z(s) = R + Ls = L\left(s + \frac{R}{L}\right) \tag{11-51}$$

$$Y(s) = \frac{1}{Z(s)} = \frac{1/L}{s + R/L} \tag{11-52}$$

Thus the impedance function has a zero at $s = -R/L$ and a pole at infinity, while the admittance function has the opposite pole-zero configuration (poles become zeros; zeros become poles). In the sinusoidal steady state, $s = j\omega$, and the immittance functions become

$$Z(j\omega) = L\left(j\omega + \frac{R}{L}\right) = R\left(1 + j\omega\frac{L}{R}\right) \tag{11-53}$$

$$Y(j\omega) = \frac{1/R}{j\omega L/R + 1} \tag{11-54}$$

By Eq. 11-53 the real part of $Z(j\omega)$ is constant and the imaginary part increases with frequency. Equation 11-54 has the same form as Eq. 11-48, which has a circular locus. The phasor loci for impedance and

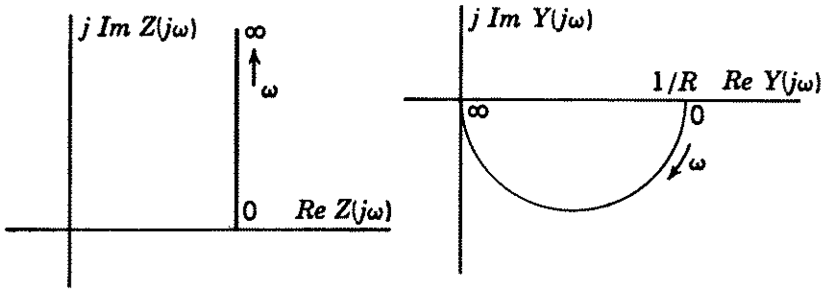


Fig. 11-13. Immittance characteristics for RL network.

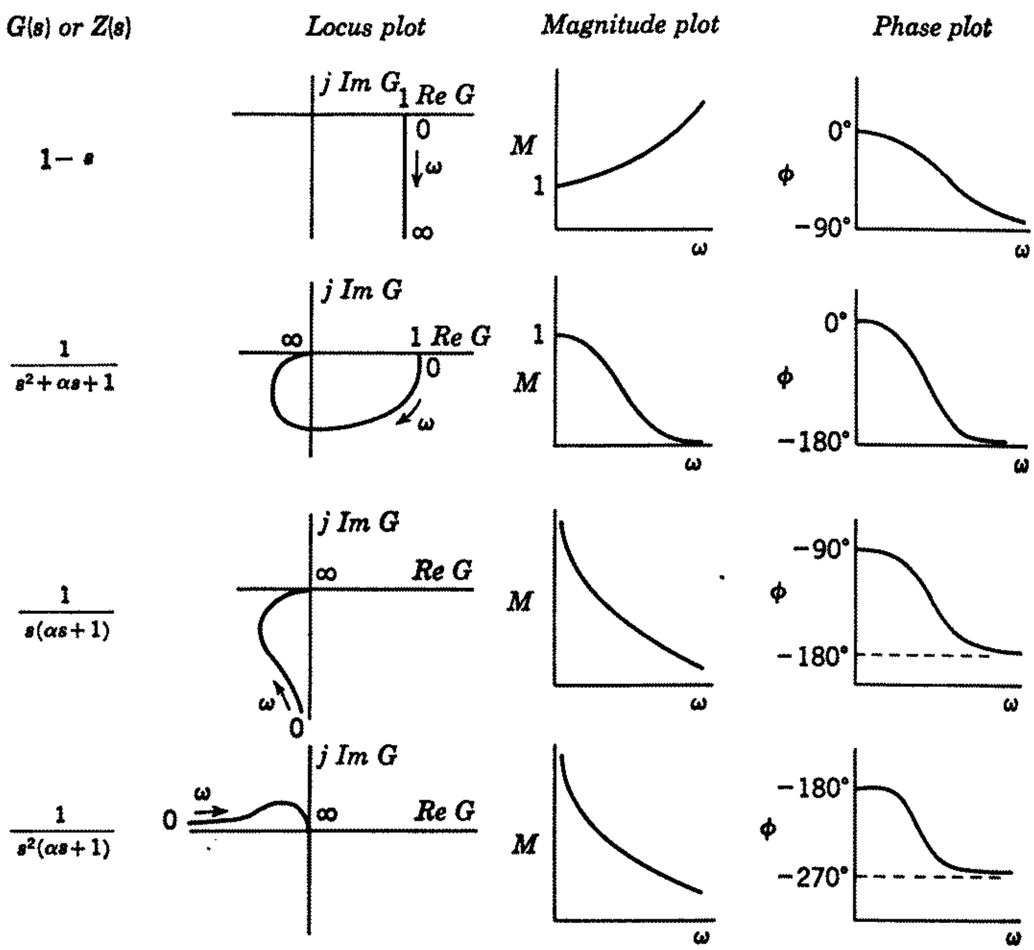


Fig. 11-14.

admittance are shown in Fig. 11-13. Several other plots are shown in Fig. 11-14 for given transfer functions or immittance functions.

11-3. Sinusoidal network functions in terms of poles and zeros

As was pointed out earlier in this chapter, all voltage and current waveforms in any linear network are sinusoidal in the steady state if the network is driven by sinusoidal waveforms. For this reason, if we are given $V(j\omega)$ in this equation,

$$I(j\omega) = Y(j\omega)V(j\omega) \tag{11-55}$$

we are not interested in solving for the waveform of $I(j\omega)$. We know the waveform in advance: *it is a sinusoid.* The information we do need is: (1) for a given magnitude of the voltage $V(j\omega)$, what is the magnitude of $I(j\omega)$, and (2) what is the phase relationship of $I(j\omega)$ in terms of $V(j\omega)$? In other words, we are interested only in magnitude and phase relating $V(j\omega)$ and $I(j\omega)$. To find this information, it is not necessary to know the magnitude of $V(j\omega)$. Since the networks under consideration are linear, the magnitude of $I(j\omega)$ is linearly dependent on the magnitude of $V(j\omega)$: if V is 1 volt to give I of 1 amp, V of 10 volts will give an I of 10 amp. The quantity that relates the phase of $V(j\omega)$ to that of $I(j\omega)$ is $Y(j\omega)$; likewise, $Y(j\omega)$ relates the magnitude of $V(j\omega)$ to that of $I(j\omega)$. In the case of the last equation, the relationship between $V(j\omega)$ and $I(j\omega)$ is given *completely* (for all values of frequency) by the magnitude and phase of $Y(j\omega)$,

$$Y(j\omega) = |Y(j\omega)|e^{j\phi(\omega)} \tag{11-56}$$

If Eq. 11-55 is written in the form

$$\frac{I(j\omega)}{V(j\omega)} = Y(j\omega) \tag{11-57}$$

this ratio is often described as the *complex ratio* of current to voltage. The term complex ratio thus implies not only the magnitude of the ratio of one quantity to another but also the phase of the one quantity with respect to the other. Network behavior as a function of frequency (and, of course, we are now specializing in *radian* frequency) is determined entirely by complex ratios; that is, by immittance functions and transfer functions.

The same arguments given in the previous paragraph apply to the following typical equations because of their similarity to Eq. 11-55.

$$V_2(j\omega) = G(j\omega)V_1(j\omega) \tag{11-58}$$

$$V(j\omega) = Z(j\omega)I(j\omega) \tag{11-59}$$

$$V_2(j\omega) = Z_{tr}(j\omega)I(j\omega) \tag{11-60}$$

and so on (Z_{tr} is the transfer impedance). Thus it is seen that the discussions for $Y(j\omega)$ apply in general to any network function in the sinusoidal steady state.

The admittance function $Y(s)$ has the form of a quotient of polynomials which may be factored into roots of the form

$$(s - s_r) \tag{11-61}$$

where s_r is either a pole or a zero. In the sinusoidal steady state, $s = j\omega$, and the typical term becomes

$$(j\omega - s_r) \tag{11-62}$$

In the complex plane, $j\omega$ and s_r are phasors. We are interested in the *difference* of these two phasors—which is also a phasor. The phasor s_r is, in general, complex; the phasor $j\omega$ is purely imaginary and is on the $j\omega$ axis. These two phasors and their difference are shown in Fig. 11-15.

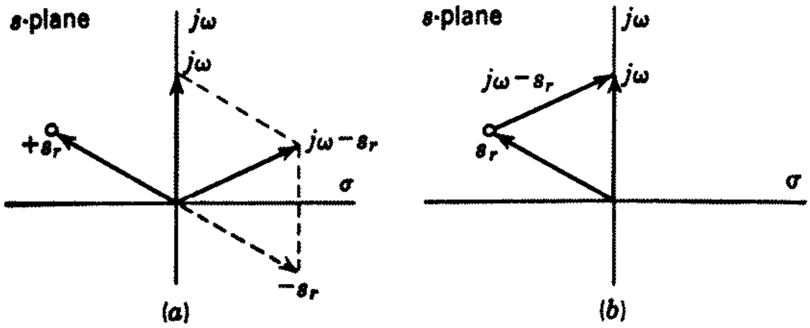


Fig. 11-15. Direction of the phasor $(j\omega - s_r)$: (a) polar diagram; (b) string diagram.

Figure 11-15(a) shows the phasors with respect to the s plane origin. Figure 11-15(b) shows the equivalent "string" phasor diagram. The phasor difference $(j\omega - s_r)$ is seen to be a phasor directed *from* s_r *to* $j\omega$. As ω changes from 0 to ∞, the position of $j\omega$ changes—always remaining on the $j\omega$ axis. The combination of several of these phasors can be used to determine sinusoidal network functions. This will be illustrated by a number of examples.

Consider first the admittance of a series RL circuit. Such a circuit is shown in Fig. 11-12. The impedance of this network is

$$Z(s) = L\left(s + \frac{R}{L}\right) \tag{11-63}$$

and so the admittance has the form

$$Y(s) = \frac{1}{L}\,\frac{1}{(s + R/L)} \tag{11-64}$$

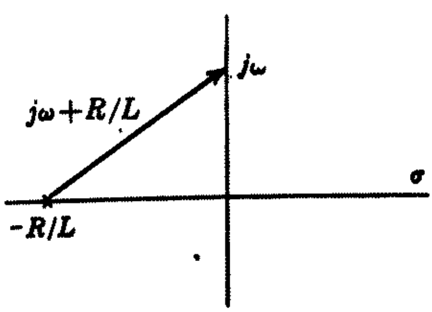


Fig. 11-16. Phasor diagram.

Thus $Y(s)$ has a finite pole at $s = -R/L$ and a zero at infinity. This pole-zero configuration is shown in Fig. 11-16. As ω increases from zero to infinity, the phasor changes position

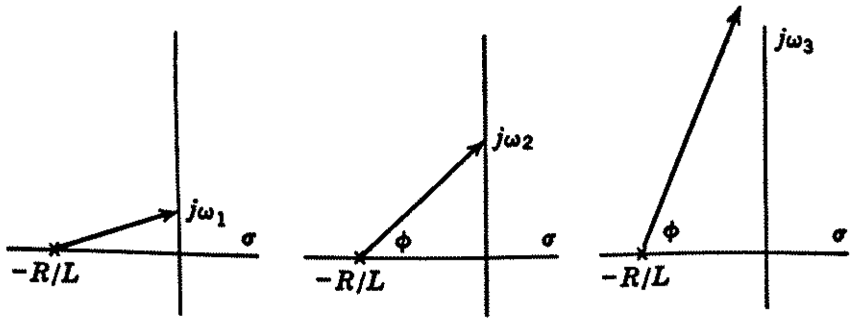


Fig. 11-17. Phasor diagram changing with frequency.

as shown in Fig. 11-17. The frequency variation of the impedance is found by writing

$$L\left(j\omega + \frac{R}{L}\right) = M(\omega)e^{j\phi(\omega)} \tag{11-65}$$

Then the admittance may be found from the equation

$$Y(j\omega) = \frac{1}{M(\omega)}\,e^{-j\phi(\omega)} \tag{11-66}$$

It can be seen that the magnitude changes from $1/R$ to 0 as ω changes from zero to infinity; similarly, the phase changes from 0° to $-90°$ as ω varies from zero to infinity. The polar coordinate representation of this variation is shown in Fig. 11-18. The variation of admittance with frequency is exactly the same as the variation of current with frequency for a constant magnitude of voltage of $1/R$ volts. If the voltage has a different magnitude, the current will increase or decrease linearly for all values of frequency.

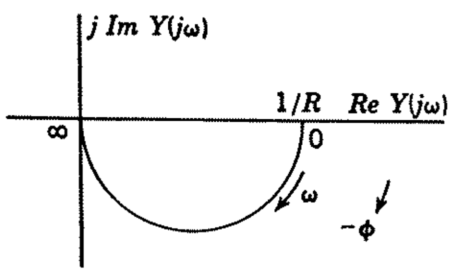


Fig. 11-18. Variation of phasor with frequency.

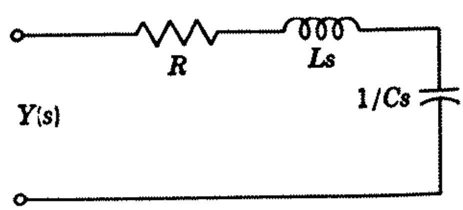


Fig. 11-19. Series *RLC* circuit.

11-4. Resonance, circuit Q, and bandwidth

The method that has been illustrated by means of the study of the *RL* series circuit applies to other networks. Consider a series *RLC* circuit as shown in Fig. 11-19. The driving-point impedance for this network is

$$Z(s) = Ls + R + \frac{1}{Cs} \tag{11-67}$$

and the admittance is the reciprocal of $Z(s)$; that is

$$Y(s) = \frac{1}{L}\left(\frac{s}{s^2 + Rs/L + 1/LC}\right) \tag{11-68}$$

or

$$Y(s) = \frac{1}{L}\left(\frac{s}{(s - s_a)(s - s_a^*)}\right) \tag{11-69}$$

where

$$s_a,\, s_a^* = -\frac{R}{2L} \pm j\sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} \tag{11-70}$$

In this expression, ω_n is the natural undamped frequency of the system, and ζ is the dimensionless damping ratio. If we consider only the underdamped (or oscillatory) case where $\zeta < 1$, the variation of the position of the poles of $Y(s)$ for constant ω_n and variable ζ is shown in Fig. 11-20, where the locus is a circle. In addition, $Y(s)$ has a zero at the origin of the s plane.

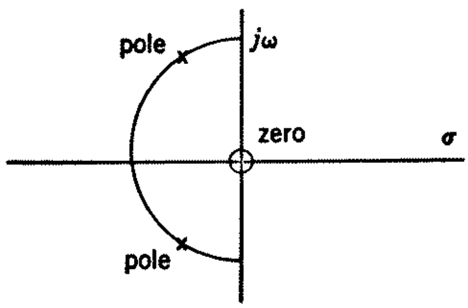


Fig. 11-20. Pole-zero configuration for $Y(s)$.

In the sinusoidal steady state ($s = j\omega$), the *frequency response* of the system, such quantities as $|I(j\omega)|$, $|Y(j\omega)|$ etc., may be found by allowing ω to vary over a range of frequencies. Several steps in such a frequency variation are shown in Fig. 11-21, together with the com-

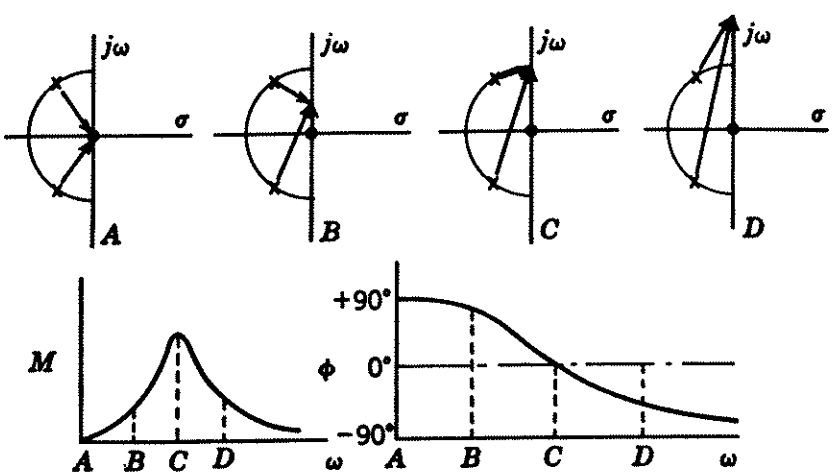


Fig. 11-21. Frequency response of an *RLC* network.

plete magnitude and phase characteristic. Over the range of frequencies, the phase changes from $+90°$, through $0°$ to $-90°$, while the magnitude starts from zero value, attains a maximum value, and then asymptotically approaches zero for high frequencies. The function $Y(j\omega)$ is found from Eq. 11-69, letting $s = j\omega$ in terms of magnitude and phase factors in the form

$$Y(j\omega) = \frac{M_0}{M_a M_a^*} e^{j(\phi_0 - \phi_a - \phi_a^*)} \tag{11-71}$$

The maximum value of $Y(j\omega)$ evidently takes place near the frequency at which M_a has a minimum value. The frequency to cause M_a to have a minimum value is a frequency very near to the point of closest approach to one of the conjugate poles. In that frequency range, M_a changes rapidly, and at the same time M_a^* and M_0 are changing very slowly. The frequency corresponding to a maximum $Y(j\omega)$ is defined as the *frequency of resonance*. Since $I(j\omega)$ varies just as $Y(j\omega)$, the frequency of resonance is also the frequency of maximum $I(j\omega)$.

The magnitude of $Y(j\omega)$ may be written in the form

$$|Y(j\omega)| = \frac{1}{\sqrt{R^2 + (\omega L - 1/\omega C)^2}} \tag{11-72}$$

and from this equation it is seen that $Y(j\omega)$ has a maximum value of $(1/R)$ when

$$\omega L - \frac{1}{\omega C} = 0 \quad \text{or} \quad \omega = \frac{1}{\sqrt{LC}} = \omega_n \tag{11-73}$$

that is to say that resonance occurs at $\omega = \omega_n$ (and not at the point opposite the pole on the $j\omega$ axis). An enlarged view of the various phasors for the condition of resonance is shown in Fig. 11-22. The phase angles from the poles to $j\omega_n$ are marked ϕ_a and ϕ_a^*. The phasor from the zero to $j\omega_n$ is along the $j\omega$ axis and thus has a constant phase angle of $+90°$. The sum of ϕ_a and ϕ_a^* is equal to 90° because the triangle ABC is a right triangle (being inscribed in a semicircle). The total phase angle, which is

$$\phi_0 - \phi_a - \phi_a^* = +90° - (+90°) = 0° \tag{11-74}$$

thus has zero value. The phase angle of $Y(j\omega)$ is *zero degrees at resonance*. The magnitude of $Y(j\omega)$ at resonance is $(1/R)$, and the magnitude of $I(j\omega)$ at resonance is (V/R).

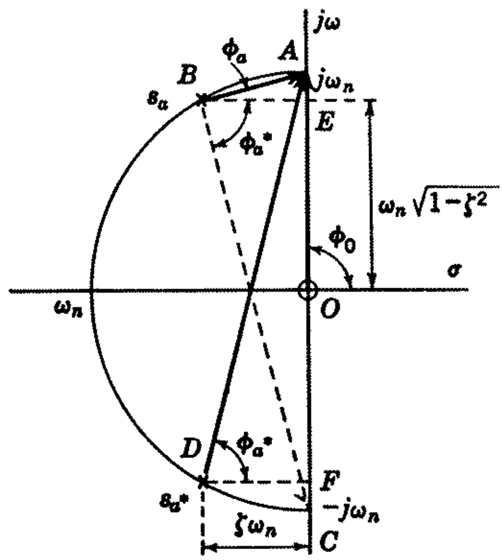


Fig. 11-22. Phasors drawn for resonance.

The *circuit Q* or simply Q for an RLC series circuit is defined as

$$Q = \frac{\omega_n L}{R} = \frac{1}{2} \frac{\omega_n}{R/2L} \tag{11-75}$$

Now the quantity $(R/2L)$ is the same as $\zeta\omega_n$ by Eq. 11-70. Thus the circuit Q is defined as

$$Q = \frac{1}{2} \frac{\omega_n}{\zeta\omega_n} = \frac{1}{2} \times \frac{\text{the length } OA}{\text{the length } EB} \tag{11-76}$$

in terms of quantities shown in Fig. 11-22. The circuit Q can thus be taken directly from a scale plot of the poles and zeros of the immittance function for an RLC circuit. The circuit Q can alternately be written in the form

$$Q = \frac{1}{2\zeta} \tag{11-77}$$

$$= \frac{1}{2 \cos \theta} \tag{11-78}$$

where θ is the angle from the $-\sigma$ axis to the line OB (or OD) of Fig. 11-22. From these last three equations, several conclusions can be written:

(1) The closer poles s_a and s_a^* are to the $j\omega$ axis, the higher the Q. (This follows since Q varies inversely with the distance EB.)
(2) The value of Q varies inversely with damping ratio, ζ. A high value of Q infers a low value of damping ratio. A circuit with low R thus has high Q.

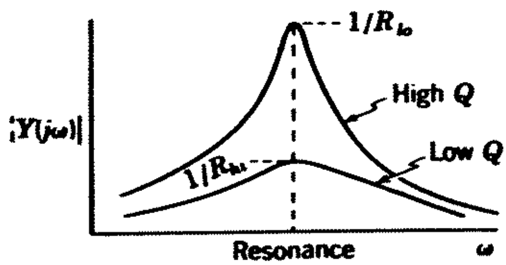


Fig. 11-23. Variation of $|Y(j\omega)|$ with Q.

Plots of the magnitude of $Y(j\omega)$ for various values of Q are shown in Fig. 11-23. The circuit Q is an important factor in circuits (of the type being considered) used for *selectors* (filters).

Another means of specifying the circuit Q is specification in terms of *half-power points*. As has been shown, the current at resonance has the magnitude V/R. When the current has the magnitude

$$I = \frac{V}{\sqrt{2}\,R} \tag{11-79}$$

the power will be *half* of that at resonance (being equal to I^2R). At the half-power points, the magnitude of the admittance $Y(j\omega)$ is $(1/\sqrt{2}\,R)$;

this requires that

$$\sqrt{R^2 + (\omega L - 1/\omega C)^2} = \sqrt{2}\, R \tag{11-80}$$

or

$$\left(\omega L - \frac{1}{\omega C}\right) = \pm R \tag{11-81}$$

This equation reduces to the form

$$\omega^2 \pm \frac{R}{L}\omega - \frac{1}{LC} = 0 \tag{11-82}$$

The values of ω that satisfy this equation are

$$\omega = \pm \frac{R}{2L} \pm \sqrt{(R/2L)^2 + 1/LC} \tag{11-83}$$

or, in terms of damping ratio and undamped natural frequency,

$$\omega = \omega_n(\pm\zeta \pm \sqrt{\zeta^2 + 1}) \tag{11-84}$$

In most practical networks used as selectors, the damping ratio ζ is very small, so that ζ^2 is negligible compared with unity. Under this condition, the last equation reduces to an especially simple form,

$$\omega = \omega_n \pm \zeta\omega_n \tag{11-85}$$

(considering only positive frequencies). The frequencies defined by this approximation are the *half-power frequencies*. Let the highest half-power frequency be designated ω_1, which is defined as

$$\omega_1 = \omega_n + \zeta\omega_n \tag{11-86}$$

and let the smaller half-power frequency be ω_2 be given as

$$\omega_2 = \omega_n - \zeta\omega_n \tag{11-87}$$

The quantity $(\zeta\omega_n)$ is the distance EB of Fig. 11-22, or the distance

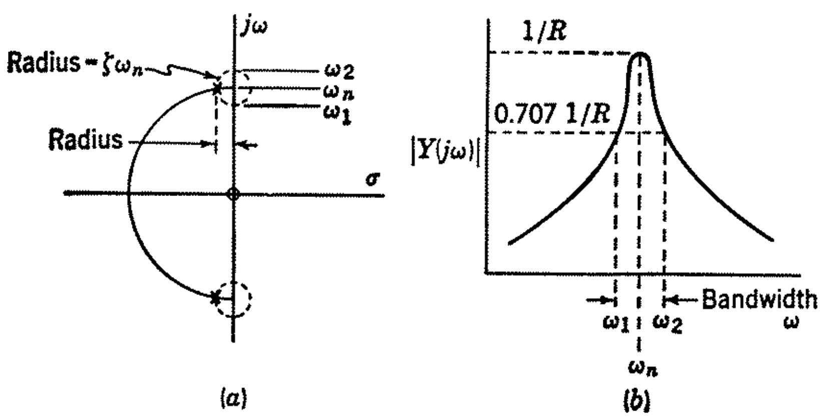


Fig. 11-24. Bandwidth on the s plane.

from the $j\omega$ axis to the pole s_a (or s_a^*). The location of the half-power points in the s plane is shown in Fig. 11-24. A circle of radius $(\zeta\omega_n)$

and centered at $j\omega_n$ crosses the $j\omega$ axis at the half-power points. At these half-power frequencies, $Y(j\omega)$ has the magnitude $0.707(1/R)$ by Eq. 11-79. The range of frequencies given as $(\omega_1 - \omega_2)$ is defined as the *bandwidth*. Bandwidth varies inversely with Q. A small bandwidth corresponds to a selective network.

The concepts of resonance, circuit Q, and bandwidth can thus be visualized in terms of the pole-zero configuration of $Y(s)$ for the RLC circuit. These concepts are easily visualized and do not depend upon algebraic manipulation of complex numbers. The specific definitions of resonance, circuit Q, and bandwidth given in this section do not apply to all possible network configurations. For example, resonance in the sense of a maximum impedance or admittance does not coincide with the frequency of unity power factor for most networks. However, all these quantities can be visualized in terms of phasor magnitude and phase, and design can be accomplished by means of simple graphical constructions. A number of additional examples of network analysis by pole-zero configuration will further illustrate these concepts in Chapter 14.

In some applications, parallel RLC networks are used as selectors. Since the parallel RLC network is the *dual* of the series RLC network, the analysis given in this section applies in terms of impedance and voltage instead of admittance and current.

11-5. Asymptotic change of magnitude with frequency in terms of poles and zeros

Both transfer functions and driving-point immittances are made up of frequency factors of the form

$$(s - s_a)^{\pm 1} \tag{11-88}$$

which in the sinusoidal steady state become

$$(j\omega - s_a)^{\pm 1} \tag{11-89}$$

For very small values of ω, this factor can be approximated as $(-s_a)^{\pm 1}$, a phasor from the point s_a to the origin of the s plane. As ω becomes larger, no such an approximation is valid. But as ω becomes very large compared with the phasor s_a, the frequency factor can be represented as

$$(j\omega)^{\pm 1} = (\omega)^{\pm 1} e^{\pm j\pi/2} \tag{11-90}$$

For large ω, the magnitude of this factor changes either linearly or inversely with ω. The asymptotic phase angle is either $+90°$ or $-90°$ depending on whether the factor is a zero or a pole. This behavior of magnitude and phase with frequency is illustrated in Fig. 11-25.

To illustrate, suppose that we are interested in a voltage ratio transfer function,

$$G(j\omega) = \frac{V_2(j\omega)}{V_1(j\omega)} \tag{11-91}$$

Assume that the sinusoidal input voltage has a magnitude of unity and that for frequencies in excess of $\omega = |s_a|$ the magnitude of the voltage $V_2(j\omega)$ is of the form

$$|V_2(j\omega)| = \frac{1}{\omega} \tag{11-92}$$

Such variation of V_2 would result if $G(s)$ had the form

$$G(s) = \frac{1}{s + \alpha} \tag{11-93}$$

provided α is very small compared to unity. The magnitude of the transfer function would then be

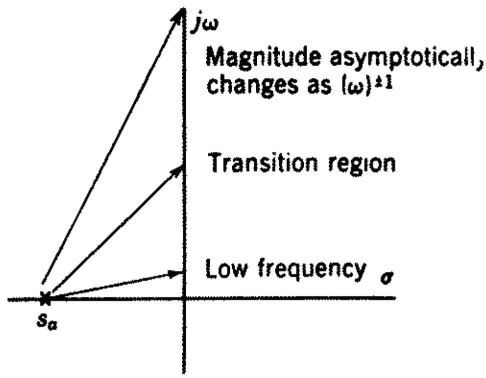


Fig. 11-25. Phasor variation with frequency.

$$|G(j\omega)| = \frac{1}{\omega} \tag{11-94}$$

The logarithmic unit, the decibel, was originally defined for a ratio of powers but is now often used for voltage and current ratios. The voltage amplitude ratio in decibels is defined by the equation

$$|G(j\omega)| = 20 \log_{10} \frac{V_2(j\omega)}{V_1(j\omega)} \tag{11-95}$$

For the example being considered, the voltage amplitude ratio in *decibels* (abbreviated db) has the form

$$20 \log_{10} \frac{1}{\omega} = -20 \log_{10} \omega \tag{11-96}$$

When $\omega = 1$, $G(j\omega)$ has a value of 0 db, and when $\omega = 2$, $G(j\omega)$ has the value

$$-20 \log_{10} 2 = -20 \times 0.301 \approx -6 \quad \text{db} \tag{11-97}$$

Two frequencies having a ratio of 2:1 are said to be separated by an *octave*. In one octave, the magnitude of this example has decreased 6 decibels. In another octave (to $\omega = 4$), the magnitude would decrease an additional 6 decibels. The magnitude is thus changing at the rate of

$$-6 \, \frac{\text{decibels}}{\text{octave}} \tag{11-98}$$

due to one pole in $G(s)$ given in Eq. 11-93.* Had $G(s)$ been of the form

$$G(s) = s + \alpha \tag{11-99}$$

corresponding to one zero in $G(s)$, the asymptotic change in the magnitude of the voltage ratio transfer function would have been an increase of 6 decibels per octave. This case would correspond to the asymptotic form resulting from the choice of the positive sign in Eq. 11-90.

A transfer function in general will be made up of a number of factors of the form considered; that is,

$$G(s) = H\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \tag{11-100}$$

Each zero in this expression will cause an asymptotic increase in the magnitude of $G(s)$ of $+6$ db per octave, while each pole will cause an asymptotic decrease in magnitude of $G(s)$ of -6 db per octave. In the frequency limit, these increases and decreases will cancel in pole-zero pairs (one pole cancels the effect of one zero). The net asymptotic change of magnitude with frequency will thus be the number of finite zeros less the number of finite poles times 6 db per octave.

As an example, consider a two-terminal-pair network having a voltage ratio transfer function of the form

$$G(s) = H\frac{s}{(s - s_a)(s - s_b)(s - s_c)} \tag{11-101}$$

This transfer function has three finite poles and one finite zero. For $s = j\omega$ and for large values of ω, the magnitude of $G(j\omega)$ has the form

$$G(j\omega) = H\left.\frac{s}{s^3}\right|_{s=j\omega} \tag{11-102}$$

The output of this two-terminal-pair network will fall off at a rate determined by the excess of poles over zeros. The rate for this particular network is -12 db per octave. The asymptotic phase of the output compared to the input will be 180° as given by Eq. 11-90.

The *RLC* selector network studied in the previous section has the admittance function

$$Y(s) = \frac{1}{L}\frac{1}{(s - s_a)(s - s_a^*)} \tag{11-103}$$

corresponding to two poles and one zero. The current passing through this selective network for constant voltage will vary with frequency

* Quantities in the ratio of 10:1 are said to be separated by a decade. Six decibels per octave is equivalent to 20 decibels per decade.

as the $M(\omega)$ curve of Fig. 11-21. For large frequency, the current will decrease with frequency at the rate of -6 db per octave.

11-6. An application: the symmetrical lattice

The *symmetrical lattice network* shown in Fig. 11-26 is a two-terminal-pair network that finds frequent application for phase correction. The

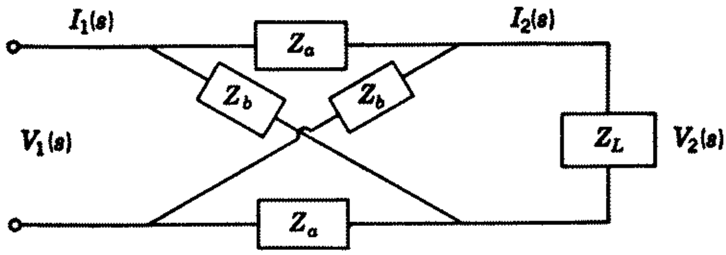


Fig. 11-26. The symmetrical lattice.

properties of this network are easily visualized in terms of the pole-zero configuration, as we shall show in this section.

The lattice structure can be put in a more familiar form by "unwrapping" it as shown in Fig. 11-27 as a *bridge network*. Assume that the network is terminated in a load impedance $Z_L(s)$ and that we are interested in the voltage ratio transfer function, $G(s) = V_2(s)/V_1(s)$. Several currents are identified in Fig. 11-27. The two currents marked I and the two currents marked I' are equal because of the symmetry of the network configuration. The load current is marked as I_2. From Kirchhoff's voltage law, we write

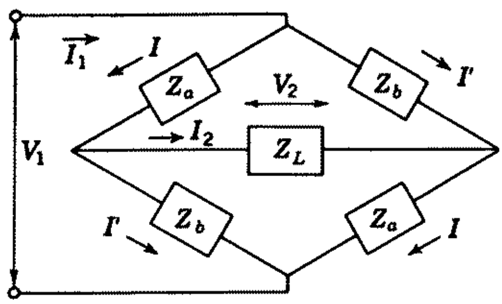


Fig. 11-27. Bridge form of the symmetrical lattice.

$$V_1 = Z_aI + Z_L(I - I') + Z_aI \tag{11-104}$$

$$V_1 = Z_aI + Z_bI' \tag{11-105}$$

In these equations, the functional notation has been omitted for simplicity—each of the quantities shown is a function of the complex frequency s. If these equations are arranged in the forms

$$V_1 = (2Z_a + Z_L)I - Z_LI' \tag{11-106}$$

$$V_1 = Z_aI + Z_bI' \tag{11-107}$$

the unknown currents I and I' may be found conveniently by the use of determinants; thus

$$I = \frac{\begin{vmatrix} V_1 & -Z_L \\ V_1 & Z_b \end{vmatrix}}{\begin{vmatrix} (2Z_a + Z_L) & -Z_L \\ Z_a & Z_b \end{vmatrix}} = \frac{V_1(Z_b + Z_L)}{Z_b(2Z_a + Z_L) + Z_aZ_L} \tag{11-108}$$

$$I' = \frac{\begin{vmatrix} (2Z_a + Z_L) & V_1 \\ Z_a & V_1 \end{vmatrix}}{\Delta} = \frac{V_1(Z_a + Z_L)}{Z_b(2Z_a + Z_L) + Z_aZ_L} \tag{11-109}$$

The load current I_2 may be found in terms of I and I' as

$$I_2 = I - I' = \frac{V_1(Z_b - Z_a)}{Z_b(2Z_a + Z_L) + Z_aZ_L} \tag{11-110}$$

The voltage across the load impedance is I_2Z_L; hence the voltage ratio transfer function becomes

$$\frac{V_2}{V_1} = \frac{I_2Z_L}{V_1} = \frac{Z_L(Z_b - Z_a)}{2Z_aZ_b + Z_L(Z_a + Z_b)} \tag{11-111}$$

The lattice network has very useful properties when the network elements are selected such that

$$Z_L = R \quad \text{and} \quad Z_aZ_b = R^2 \tag{11-112}$$

With these restrictions, the equation becomes

$$\begin{aligned} \frac{V_2}{V_1} &= \frac{(Z_b - Z_a)}{2R + (Z_a + Z_b)} = \frac{R^2 - Z_a^2}{R^2 + Z_a^2 + 2RZ_a} \\ &= \frac{(R - Z_a)(R + Z_a)}{(R + Z_a)(R + Z_a)} = \frac{R - Z_a}{R + Z_a} \end{aligned} \tag{11-113}$$

To apply this derivation to a specific network, let the impedance Z_a be represented by the parallel LC network shown in Fig. 11-28. The net-

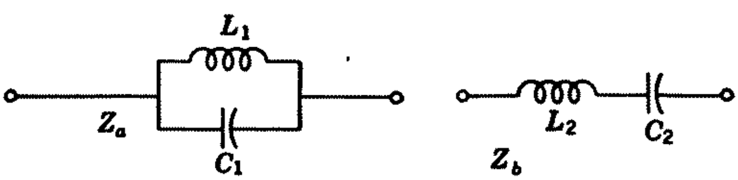


Fig. 11-28. Networks for Z_a and Z_b.

work to represent Z_b must be the dual of that representing Z_a. The series LC network for Z_b is also shown in Fig. 11-28. The impedance for the parallel LC network is

$$Z_a(s) = \frac{1}{C_1s + 1/L_1s} = \frac{L_1s}{L_1C_1s^2 + 1} \tag{11-114}$$

while the series LC network has impedance given by

$$Z_b(s) = L_2 s + \frac{1}{C_2 s} = \frac{L_2 C_2 s^2 + 1}{C_2 s} \tag{11-115}$$

Now the parameters L_1, L_2, C_1, C_2, and R must be selected to satisfy Eq. 11-112. One set of parameters that satisfies the requirements is $L_1 = L_2 = 1$ henry, $C_1 = C_2 = 1$ farad, and $R = 1$ ohm. There are other such values; these are selected for their simplicity in illustrating the frequency behavior of the lattice network. With this choice of parameters, the impedance functions become

$$Z_a = \frac{s}{s^2+1}; \qquad Z_b = \frac{s^2+1}{s}; \qquad Z_a Z_b = 1 = R^2 \tag{11-116}$$

The voltage ratio transfer function is given by Eq. 11-113. With the assigned parameters, this transfer function becomes

$$\frac{V_2(s)}{V_1(s)} = \frac{1 - s/(s^2+1)}{1 + s/(s^2+1)} \tag{11-117}$$

or

$$\frac{V_2(s)}{V_1(s)} = \frac{s^2 - s + 1}{s^2 + s + 1} = \frac{(s - s_1)(s - s_1^*)}{(s - s_a)(s - s_a^*)} \tag{11-118}$$

The two zeros of the voltage transfer function have the values

$$s_1,\ s_1^* = +\frac{1}{2} \pm j\frac{\sqrt{3}}{2} \tag{11-119}$$

and the poles have the values

$$s_a,\ s_a^* = -\frac{1}{2} \pm j\frac{\sqrt{3}}{2} \tag{11-120}$$

The pole-zero configuration for the network of Fig. 11-29 is shown in Fig. 11-30. The poles and zeros are located on a unit circle about the

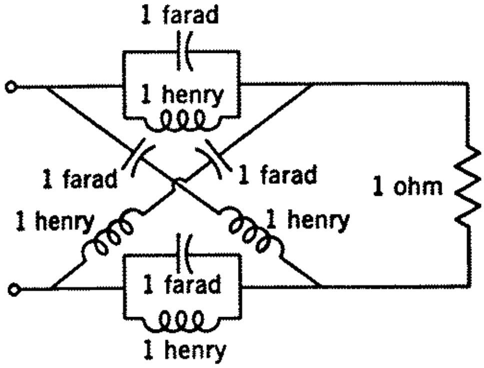


Fig. 11-29. Symmetrical lattice with element values assigned.

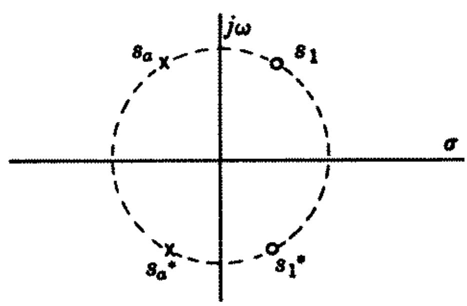


Fig. 11-30. Pole-zero configuration.

origin of the s plane. They are symmetrically located with respect to the axis of reals and axis of imaginaries. The two poles and two zeros so arranged in the s plane are known as a *quad*. If other parameter values are selected still satisfying the requirements of Eq. 11-112, the poles and zeros will still be located symmetrically with respect to the axis of imaginaries. The zeros will always be located in the right half plane and the poles in the left half plane, and both poles and zeros will occur either in conjugate pairs or on the real axis.

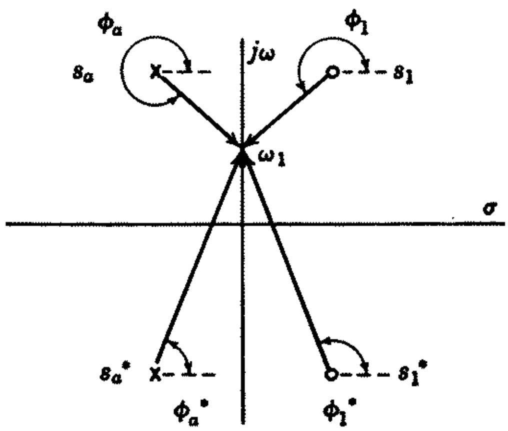


Fig. 11-31. Frequency response computation.

We now come to the problem of finding the frequency response—the variation of the magnitude and phase angle of the transfer function with frequency. As outlined in previous sections, the frequency response is found by drawing phasors to different points on the $j\omega$ axis. A typical graph is shown in Fig. 11-31 for $s - j\omega_1$. Each phase angle is found with respect to the positive σ axis. First let us examine the frequency behavior of the transfer function relating V_2 and V_1. For $s = j\omega$, Eq. 11-118 becomes

$$\frac{V_2(j\omega)}{V_1(j\omega)} = \frac{(j\omega - s_1)(j\omega - s_1^*)}{(j\omega - s_a)(j\omega - s_a^*)} \tag{11-121}$$

But from the figure, we see that the magnitude of $(j\omega - s_1)$ is always equal to the magnitude of $(j\omega - s_a)$; likewise, the magnitude of $(j\omega - s_1^*)$ is always equal to the magnitude of $(j\omega - s_a^*)$. In terms of the last equation we have discovered that

$$\left|\frac{V_2(j\omega)}{V_1(j\omega)}\right| = 1 \tag{11-122}$$

In other words, we have arrived at the remarkable conclusion that for this network, the magnitude of the output is *always* equal to the magnitude of the input—*for any frequency*. Our network is made up of four inductors, four capacitors, and one resistor, and yet it has the same frequency invariant characteristic associated with purely resistive networks. There must be something else of interest in this network after we have come this far. Let us examine the phase of the transfer function as a function of frequency.

In computing the phase, we regard the phase from zero terms as positive and from pole terms as negative. When $\omega = 0$ the phase is given as

$$\phi = \phi_1 + \phi_1^* - \phi_a - \phi_a^* = 240° + 120° - 300° - 60° = 0° \qquad (11\text{-}123)$$

There is no phase shift at zero frequency. (It is seen from the network of Fig. 11-29 that the input and output are identical at zero frequency, with the inductor acting as a short circuit, the capacitor acting as an open circuit, and thus the two-terminal-pairs directly connected together.) As the frequency increases, the phase of $V_2(j\omega)/V_1(j\omega)$ becomes negative, approaching $-360°$ as the frequency becomes infinite. The phase and magnitude characteristics are shown in Fig. 11-32. This network finds application as a phase-shifting network in

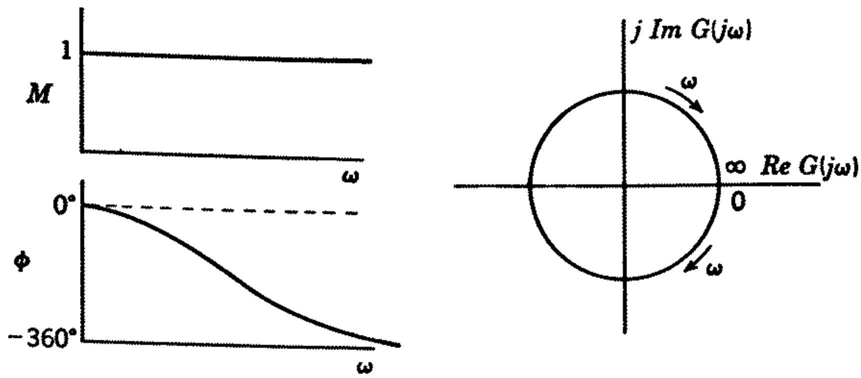


Fig. 11-32. Phase and magnitude characteristics of a symmetrical lattice network.

telephone circuits. Note, incidentally, that for the first time we have found zeros located in the right half plane. As discussed in the last chapter, zeros are permitted in the right half plane for (output/input) transfer functions, but poles are not. This is true only for the transfer function; neither poles nor zeros are permitted in the right half plane for driving-point immittances.

FURTHER READING

For further reading on analysis in the sinusoidal steady state in terms of poles and zeros, see LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 8–12, 193–196; and Guillemin, *Introductory Circuit Theory* (John Wiley & Sons, Inc., New York, 1953), Chap. 6. See also D. F. Tuttle, Jr., *Network Synthesis*, 2 vols. (John Wiley & Sons, Inc., New York, in preparation).

PROBLEMS

11-1. By manipulating unit phasors as in Art. 11-1, show that

$$\sin^2 \omega t + \cos^2 \omega t = 1$$

(That is, start with the phasors $e^{j\omega t}$ and $e^{-j\omega t}$ and manipulate these phasors to prove the identity given above by a graphical construction.)

11-2. Find the steady-state solution of the differential equation

$$Ri(t) + \frac{1}{C}\int i(t)\,dt = v(t)$$

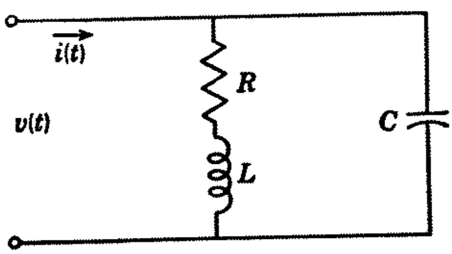
for a sinusoidal $v(t)$, by letting $v(t)$ have the form $Ve^{\pm j\omega t}$.

11-3. Find the steady-state solution of the differential equation

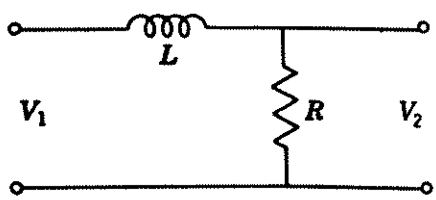
$$L\frac{di(t)}{dt} + \frac{1}{C}\int i(t)\,dt = v(t)$$

for a sinusoidal driving force $v(t)$, by letting $v(t)$ have the exponential form $Ve^{\pm j\omega t}$.

11-4. For the network shown in the figure, find the steady-state component of $i(t)$ when $v(t) = V \sin \omega t$, by using Eq. 11-14.



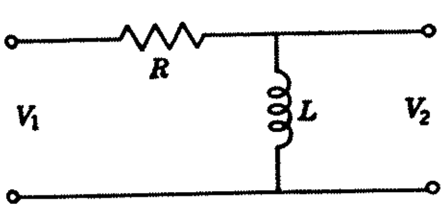
Prob. 11-4.



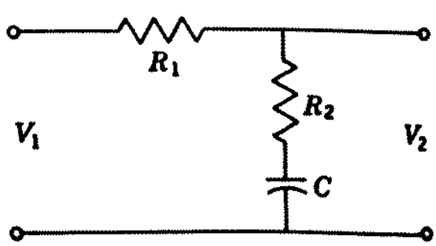
Prob. 11-5.

11-5. For the network shown in the figure, sketch $G(j\omega) = V_2(j\omega)/V_1(j\omega)$ as a function of ω for (a) polar coordinates $M(\omega)$ and $\phi(\omega)$, and (b) rectangular coordinates M vs ω and ϕ vs ω. On the plots, clearly indicate the low- and high-frequency asymptotes.

11-6. Repeat Prob. 11-5 for the network shown.



Prob. 11-6.

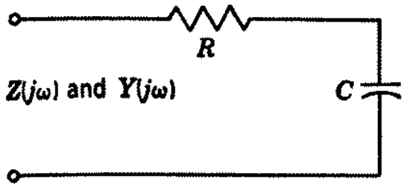


Prob. 11-7.

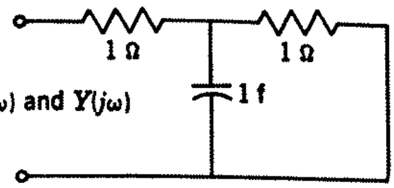
11-7. Repeat Prob. 11-6 for the network shown.

11-8. Show that the phasor locus representation of Eq. 11-48 given as Fig. 11-9 is a semicircle centered at Re $G(j\omega) = 0.5$.

11-9. For the one-terminal-pair network shown in the figure, sketch: (a) the driving-point impedance $Z(j\omega)$ as a function of ω, and (b) the driving-point admittance $Y(j\omega)$ as a function of ω, using polar coordinates as in Fig. 11-13. The sketches should have one point located accurately and the low- and high-frequency asymptotes clearly indicated.



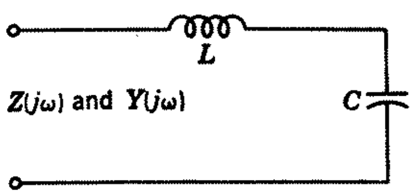
Prob. 11-9.



Prob. 11-10.

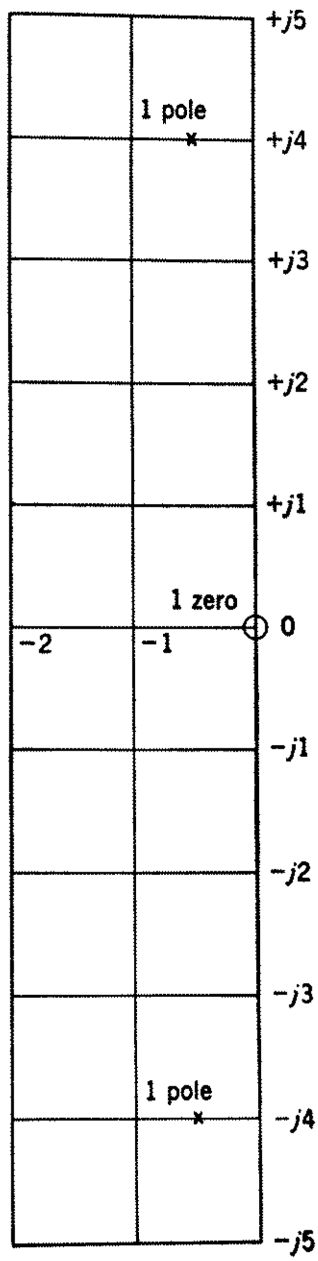
11-10. Repeat Prob. 11-9 for the one-terminal-pair network shown.

11-11. Repeat Prob. 11-9 for the one-terminal-pair network shown.



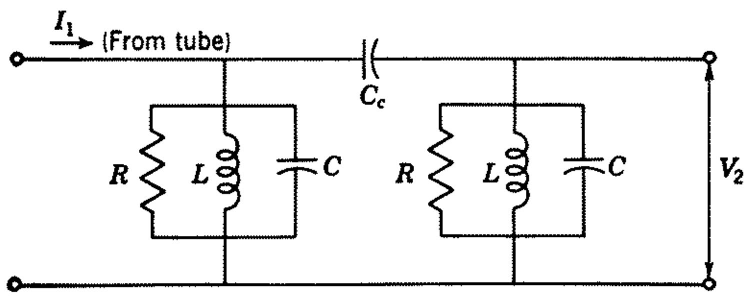
Prob. 11-11.

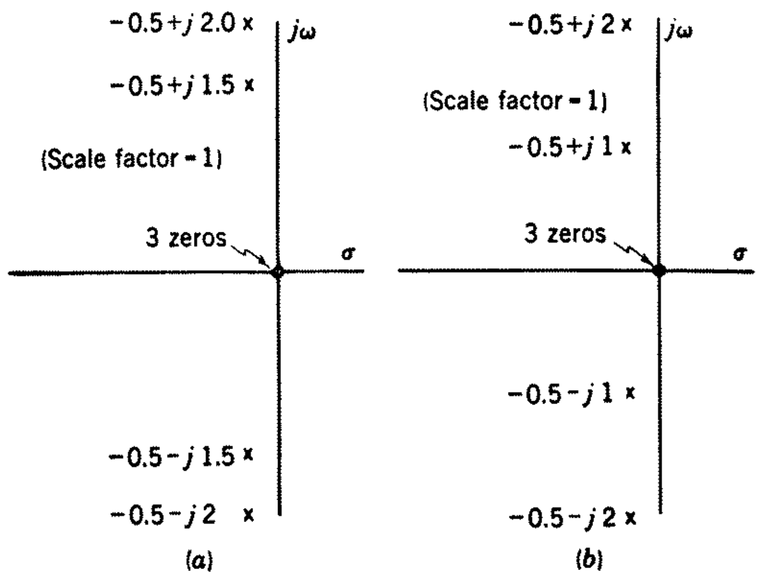
11-12. The pole-zero configuration shown in the figure represents the admittance function for the series *RLC* circuit shown in Fig. 11-19. From the pole-zero configuration, determine: (a) the undamped natural frequency ω_n, (b) the damping ratio ζ, (c) the circuit Q, (d) the bandwidth (to the half-power points), (e) the actual frequency of oscillation of the transient response, (f) the damping factor of the transient response, (g) the frequency of resonance, (h) the parameter values (in terms of L if the values cannot otherwise be uniquely determined). (i) Sketch the magnitude of the admittance $|Y(j\omega)|$ as a function of frequency. (j) If the frequency scale is magnified by a factor of 1000, how do the values of the parameters, R, L, and C change? *Answers.* (a) 4.04, (b) 0.124, (c) 4.04, (d) 1.0, (e) 4.0, (f) 0.5, (g) 4.04, (h) $R = L$, $C = 0.061/L$.



Prob. 11-12.

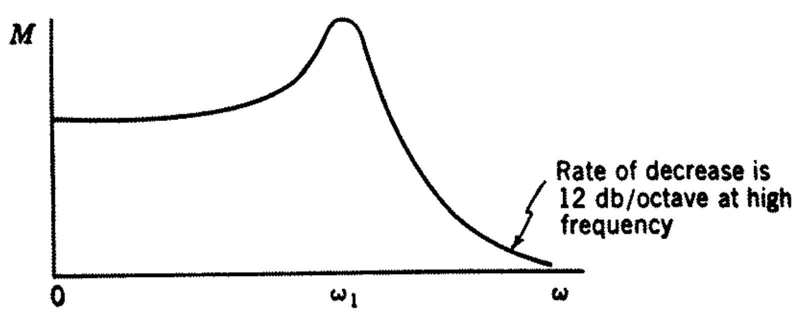
11-13. The passive network shown in the accompanying figure is known as a double-tuned circuit. It consists of two parallel RLC networks coupled with a capacitor C_c. For a certain combination of





Prob. 11-13.

parameters, the pole-zero configuration is as shown in the figure as (a) and (b) for the transfer impedance, $Z_{21}(s) = V_2(s)/I_1(s)$. From the pole-zero configuration (accurately plotted), *plot* the magnitude of the transfer impedance as a function of frequency ω.



Prob. 11-14.

11-14. The frequency response shown in the figure is observed for a given network. Draw a pole-zero configuration that can give this response. (*Note:* there is no unique solution to the problem, but every solution must be such as to meet the requirements at low frequencies, high frequencies, and resonance.)

11-15. A black box is marked "*RLC* Series Circuit" but no component values are indicated. You are seeking a circuit that will oscillate if a battery is connected to the box by the closing of a switch. In your laboratory, you have standard test equipment such as vacuum tube voltmeters, ammeters, sine wave generators—any frequency range. However, you have no adequate cathode ray oscillograph. You are not certain that you could detect oscillation with the instruments you have, since the frequency of oscillation may be very high. The problem you face is this: with measurements made in the sinusoidal steady state, how can you determine whether the current through the black box will oscillate when the switch is closed *and* what will be the frequency of oscillation. Describe the experiment you would perform.

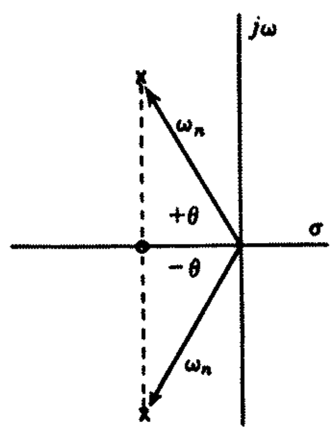
11-16. Show that the bandwidth B varies inversely with the circuit Q for a *series* RLC circuit.

11-17. Show that for an RLC series network the product of $|Y|_{max}$ and the bandwidth B equals $1/L$, where L is the inductance.

11-18. Draw the phasor locus corresponding to the transfer function, $G(s) = 1/(s^2 + \alpha)$. Carefully identify the high- and low-frequency asymptotes.

11-19. Draw the phasor locus for the function $G(s) = 1/s(s^2 + \alpha s + 1)$. Carefully identify the high-frequency and low-frequency asymptotes.

11-20. The two poles and zero shown in the s plane of the accompanying sketch are for the transfer function of a two-terminal-pair network, $G(s) = V_2(s)/V_1(s)$. The zero is on the real axis at a position to correspond with the same real part of the poles. The poles have positions corresponding to $\zeta = 0.707(\theta = 45°)$; ω_n is the distance from the origin to the pole as shown. In this problem, we will investigate the effect of the finite zero by computations with and without the zero. (a) The bandwidth of the system is modified from the definition given in the chapter as the range of frequencies from $\omega = 0$ to the halfpower point. Compute the bandwidth of the system with the pole-zero configuration shown above; compute the bandwidth with the zero removed. In which case is the bandwidth greater and by what factor? A graphical construction is suggested. (b) We

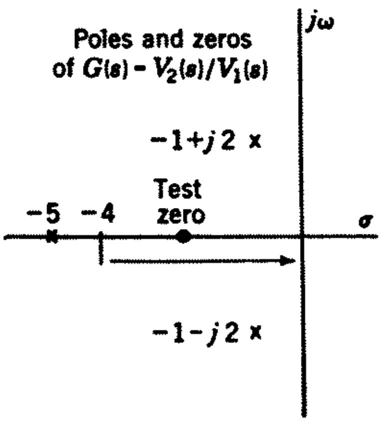


Prob. 11-20.

define the *per cent overshoot* in response to a step function input as

$$\frac{\text{maximum value} - \text{final value}}{\text{final value}} 100\%$$

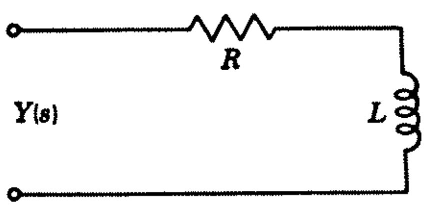
Compute the per cent overshoot for the two cases described in (a). In which case is the overshoot greater? By what factor? Check point: 4.3% without the zero. (c) Discuss qualitatively the effect of another pole on the real σ axis but with a position ten times further from the origin than the zero with respect to (1) the transient response and (2) the bandwidth.

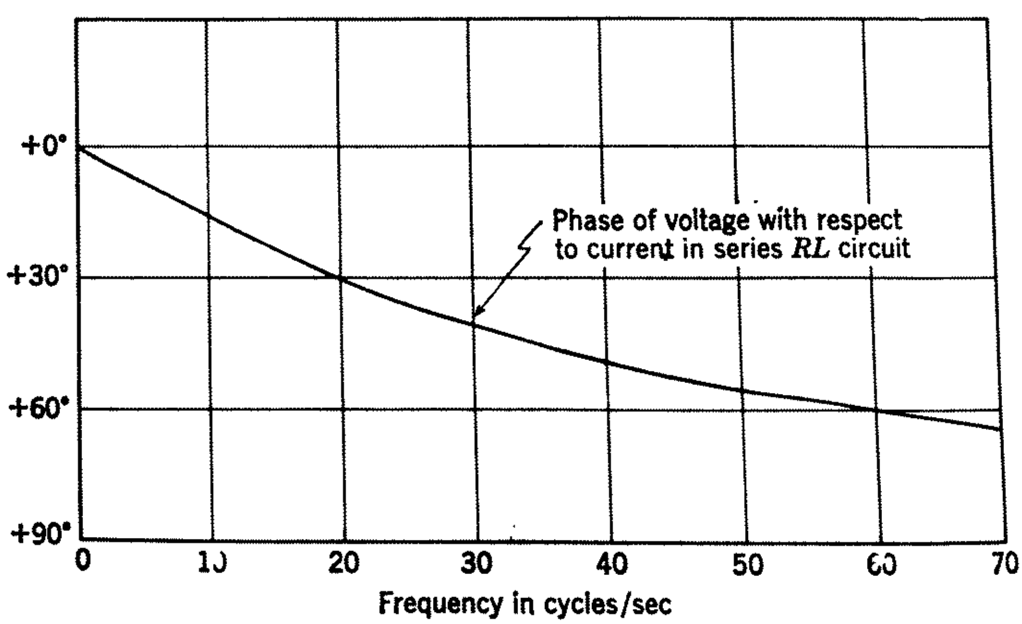


Prob. 11-21.

11-21. For the pole-zero configuration shown in the figure, compute a curve of bandwidth (as defined in Prob. 11-20) as a function of zero position from $\sigma = -4$ to $\sigma = 0$. Show any changes in curvature carefully.

11-22. Consider the pole-zero configuration of Prob. 11-21 without the "test zero." To the configuration is added a so-called "dipole" of a zero at $\sigma = -0.1$ and a pole at $\sigma = -0.105$. Show that this dipole

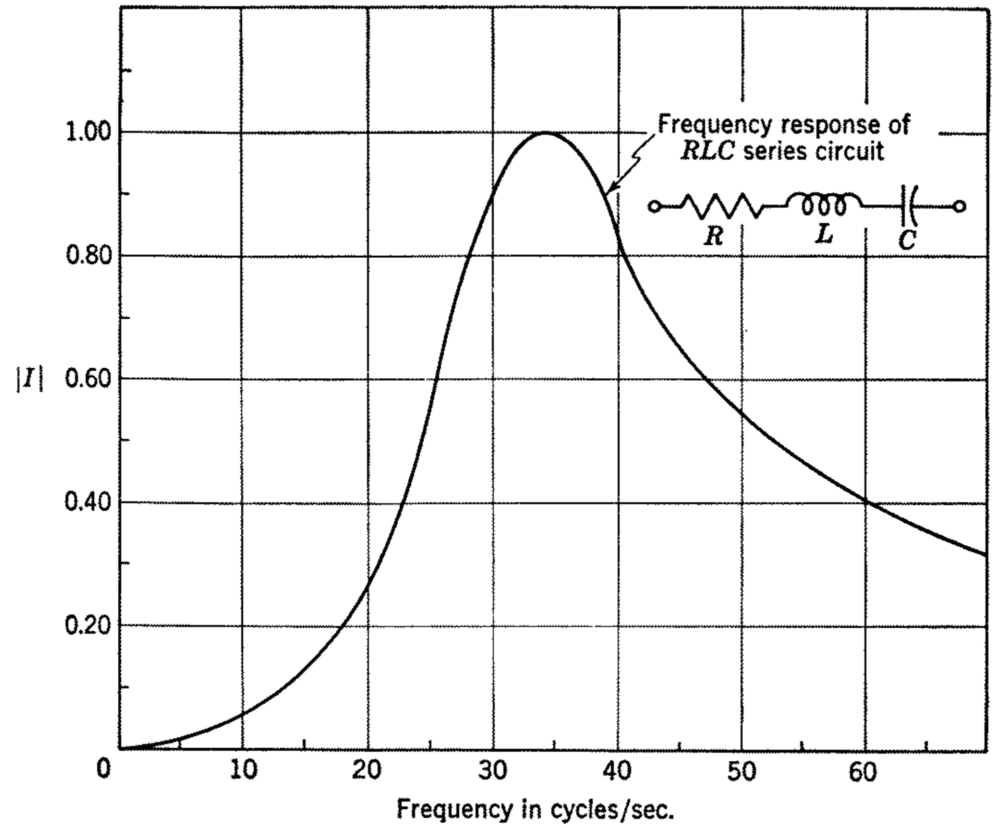




Prob. 11-23.

does not appreciably affect the bandwidth or the transient response to a step input. (The use of a dipole to change certain characteristics of network response without changing bandwidth is described in the literature relating to servomechanisms as lag or integral compensation.)

11-23. For a series RL circuit, the phase of the voltage waveform with respect to the current waveform is measured and plotted in the figure. Plot (with coordinate values) the pole-zero configuration for the driving-point admittance of the RL circuit.



Prob. 11-24.

11-24. The curve of the accompanying figure represents the current magnitude as a function of frequency with constant input voltage for an RLC series circuit. From this plot, determine the locations of the poles and zeros in the s plane for the network under study.

CHAPTER 12

ONE-TERMINAL-PAIR REACTIVE NETWORKS

12-1. Reactive networks

The networks to be studied in this chapter will be restricted in two ways. (1) The networks will be assumed to be made up of inductances and capacitances only. Since these networks contain no resistive elements, they are said to be *dissipationless*. (2) Only one-terminal-pair networks will be considered. The appropriate network function for the one-terminal-pair network is the driving-point immittance (either impedance or admittance). The driving-point impedance and admittance are

$$Z(s) = \frac{V(s)}{I(s)}, \qquad Y(s) = \frac{I(s)}{V(s)} \tag{12-1}$$

respectively, where $V(s)$ is the voltage and $I(s)$ is the current at the driving-point terminals.

Driving-point immittances are found by combining impedance or admittance expressions for elements in the network. These expressions for inductance and capacitance are summarized in the following table.

	Impedance	Admittance
Inductance	Ls	$1/Ls$
Capacitance	$1/Cs$	Cs

Any arbitrarily complicated network can be broken into parts consisting of series and parallel combinations of elements. For a series combination of any number of inductances and capacitances, the total impedance is

$$Z_t(s) = Z_1(s) + Z_2(s) + \dots + Z_n(s) \tag{12-2}$$

or

$$Z_t(s) = (L_1 + L_2 + L_3 + \dots)s + \left(\frac{1}{C_1} + \frac{1}{C_2} + \dots\right)\frac{1}{s} \tag{12-3}$$

$$= L_{eq}s + \frac{1}{C_{eq}s} \tag{12-4}$$

In this expression L_{eq} is the equivalent inductance and C_{eq} is the equivalent capacitance of the series system. Equation 12-4 may be manipulated algebraically to the form

$$Z_t(s) = L_{eq}\,\frac{s^2 + 1/L_{eq}C_{eq}}{s} \tag{12-5}$$

Similarly, the total admittance of a parallel combination of any number of inductances and capacitances is

$$Y_t(s) = Y_1(s) + Y_2(s) + \ldots + Y_n(s) \tag{12-6}$$

or
$$Y_t(s) = (C_1 + C_2 + C_3 + \ldots)s + \left(\frac{1}{L_1} + \frac{1}{L_2} + \ldots\right)\frac{1}{s} \tag{12-7}$$

$$= C_{eq}s + \frac{1}{L_{eq}s} \tag{12-8}$$

where C_{eq} and L_{eq} are the equivalent capacitance and inductance of the parallel combination of elements (hence this set of equivalent values is different from those symbolized identically in Eq. 12-4). The last equation may be rearranged in a form similar to Eq. 12-5; thus

$$Y_t(s) = C_{eq}\frac{s^2 + 1/L_{eq}C_{eq}}{s} \tag{12-9}$$

for a parallel system. The only difference in the form of the two equations for Y_t and Z_t is the multiplying constant.

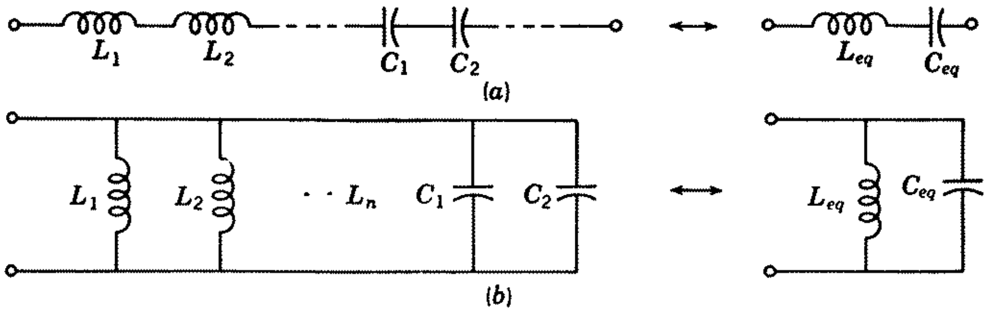


Fig. 12-1. Equivalent immittance function representations.

The driving-point immittance for a complex network is found by adding impedances and the reciprocals of admittances, or admittances and the reciprocals of impedances. Since all networks can be arbitrarily divided into a number of series and parallel networks, the expression for driving-point immittance will be a combination of terms of the form of Eqs. 12-5 and 12-9, and their reciprocals. Several examples will illustrate such combinations.

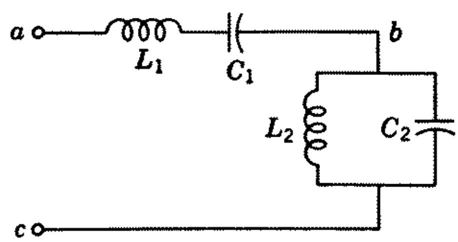


Fig. 12-2. *LC* network.

Example 1

The network shown in Fig. 12-2 is seen to be made up of a series *LC* network in series with a parallel *LC* network. The driving-point impedance is given as

$$Z_{ac} = Z_{ab} + \frac{1}{Y_{bc}} \tag{12-10}$$

or, in terms of individual element immittances,

$$Z_{ac} = L_1 s + \frac{1}{C_1 s} + \frac{1}{C_2 s + 1/L_2 s} \tag{12-11}$$

This equation may be rearranged in the form of Eqs. 12-5 and 12-9 as

$$Z_{ac} = L_1 \frac{s^2 + 1/L_1C_1}{s} + \frac{s}{C_2(s^2 + 1/L_2C_2)} \tag{12-12}$$

or $$Z_{ac} = L_1 \frac{s^4 + (1/L_1C_1 + 1/L_2C_2 + 1/L_1C_2)s^2 + 1/L_1L_2C_1C_2}{s^3 + s/L_2C_2} \tag{12-13}$$

Example 2

The network shown in Fig. 12-3 is a ladder network with specific element values designated. The driving-point admittance for this network will be found by grouping the network elements into several series and parallel combinations. The impedance from node a to node b is

$$Z_{ab}(s) = 2s + \frac{2}{s} = \frac{2(s^2 + 1)}{s} \tag{12-14}$$

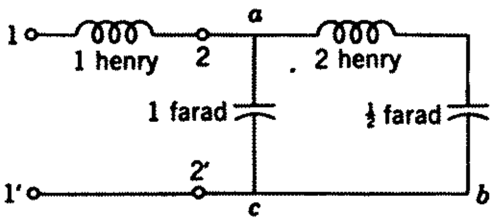


Fig. 12-3. LC ladder network.

The reciprocal of this impedance Y_{ab} can be combined with the admittance of the 1-farad capacitor; thus

$$Y_{22'}(s) = Y_{ac} + \frac{1}{Z_{ab}} = s + \frac{s}{2(s^2 + 1)}, \tag{12-15}$$

or $$Y_{22'}(s) = \frac{2s^3 + 3s}{2(s^2 + 1)} \tag{12-16}$$

The driving-point impedance at terminals 1-1′ is found by combining the impedance of the 1-henry inductance with the reciprocal of $Y_{22'}$; thus

$$Z_{11'}(s) = Z_{1a} + \frac{1}{Y_{22'}} = s + \frac{2(s^2 + 1)}{2s^3 + 3s} \tag{12-17}$$

Simplifying, $$Z_{11'}(s) = \frac{2s^4 + 5s^2 + 2}{2s^3 + 3s} \tag{12-18}$$

The driving-point admittance is the reciprocal of $Z_{11'}$ and so is given by the expression

$$Y_{11'}(s) = \frac{2s^3 + 3s}{2s^4 + 5s^2 + 2} \tag{12-19}$$

or $$Y_{11'}(s) = \frac{s^3 + 1.5s}{s^4 + 2.5s^2 + 1} \tag{12-20}$$

A comparison of the various derived expressions for driving-point immittance will show a number of common features:

(1) All the equations—Eqs. 12-5, 12-9, 12-13, and 12-20, as well as intermediate steps—are quotients of polynomials in s with a constant multiplier.
(2) The order of the numerator and denominator polynomials never differs by more than unity.
(3) The polynomials have only even powers of s or odd powers of s in any one polynomial. Further, if the numerator polynomial has only odd powers of s, the denominator polynomial *always* has only even powers of s, and vice versa.
(4) In a polynomial with even powers of s, no even term of degree less than the term of maximum degree can be missing. The same condition holds for odd terms in odd polynomials.

Now a polynomial with only even powers of s (or whatever the variable may be) is by definition an *even polynomial.* Similarly, a polynomial with only odd powers of s is by definition an *odd polynomial.* Hence statement (3) may be expressed in different words as: the driving-point immittance functions are all odd to even or even to odd quotients of polynomials.

To illustrate further the concept of even and odd polynomials, the equation

$$P_1(s) = a_8s^8 + a_6s^6 + a_4s^4 + a_2s^2 + a_0s^0 \tag{12-21}$$

is an *even* polynomial, since it contains only even powers of s. Similarly, the equation

$$P_2(s) = a_7s^7 + a_5s^5 + a_3s^3 + a_1s \tag{12-22}$$

is an *odd* polynomial containing only odd powers of s. The equation

$$P_3(s) = a_4s^4 + a_3s^3 + a_2s^2 + a_1s + a_0 \tag{12-23}$$

contains both even and odd powers of s and so is neither an even nor an odd polynomial, but has both an even and an odd part.

The four statements just given are made on the basis of only a limited number of examples. However, these statements are shown to be true in general for LC networks in advanced textbooks.* Further, statements (1) and (2) are true for the driving-point immittance of any network—RL, RC, or the general RLC. Only statements (3) and (4) apply only in the case of the dissipationless LC network.

* See Guillemin, *Communications Network, Vol. II* (John Wiley & Sons, Inc., New York, 1935), pp. 184 ff., or Tuttle, *Network Synthesis*, 2 vols. (John Wiley & Sons, Inc., New York, in preparation).

On the basis of the above discussion, we will assume that driving-point immittances for dissipationless (LC) networks are quotients of even to odd polynomials or odd to even polynomials. Further, the order of the numerator and denominator polynomials will never differ by more than unity. Such a general impedance (and the same will hold for admittance) can be written in the form

$$Z(s) = \frac{a_{2n}s^{2n} + a_{2n-2}s^{2n-2} + \dots + a_2 s^2 + a_0}{b_{2n-1}s^{2n-1} + b_{2n-3}s^{2n-3} + \dots + b_1 s} \tag{12-24}$$

as a quotient of an even to odd polynomial, and with $a_0 = 0$ such that an s may be factored out of both numerator and denominator, an odd to even polynomial. The numerator polynomial may be considered to be an equation in s^2 which may be factored into its n roots. After the common s is factored from the denominator of the last equation, the equation can be factored into $n - 1$ roots in s^2. In factored form, the equation becomes

$$Z(s) = H\frac{(s^2 + s_1^2)(s^2 + s_3^2)\dots(s^2 + s_{2n-1}^2)}{s(s^2 + s_2^2)(s^2 + s_4^2)\dots(s^2 + s_{2n-2}^2)} \tag{12-25}$$

where

$$H = \frac{a_{2n}}{b_{2n-1}}, \qquad \text{a constant} \tag{12-26}$$

Typical form of the factors in Eq. 12-25 is $(s^2 + s_1^2)$, which factors into two roots as

$$s^2 = -s_1^2; \qquad s = \pm js_1 \tag{12-27}$$

The roots are thus purely imaginary. Such imaginary values of roots have previously been associated with radian frequency. To emphasize this identification, we will change notation at this point by letting terms of the form s_n become ω_n. Hence typical roots of the impedance equation will have roots occurring in pairs as purely imaginary numbers of the form

$$s = \pm j\omega_n \tag{12-28}$$

Then the driving-point impedance expression becomes

$$Z(s) = \frac{H}{s}\frac{(s^2 + \omega_1^2)(s^2 + \omega_3^2)\dots}{(s^2 + \omega_2^2)(s^2 + \omega_4^2)\dots} \tag{12-29}$$

or, if $a_0 = 0$ in Eq. 12-24,

$$Z(s) = Hs\frac{(s^2 + \omega_2^2)(s^2 + \omega_4^2)\dots}{(s^2 + \omega_1^2)(s^2 + \omega_3^2)\dots} \tag{12-30}$$

The reason for the particular choice of subscripts for ω will be discussed later.

From this point on, we will restrict our considerations to the special case of the sinusoidal steady state. The mathematical consequence of this restriction is that $s = j\omega$ and that the s^2 terms in the previous equations become $s^2 = -\omega^2$. For the two cases of the last two equations, the substitution $s = j\omega$ gives

$$Z(j\omega) = \pm \frac{H}{j\omega} \frac{(\omega^2 - \omega_1{}^2)(\omega^2 - \omega_3{}^2)\ldots}{(\omega^2 - \omega_2{}^2)(\omega^2 - \omega_4{}^2)\ldots} \tag{12-31}$$

or

$$Z(j\omega) = \pm j\omega H \frac{(\omega^2 - \omega_2{}^2)(\omega^2 - \omega_4{}^2)\ldots}{(\omega^2 - \omega_1{}^2)(\omega^2 - \omega_3{}^2)\ldots} \tag{12-32}$$

The $\pm$ sign in these equations is introduced to account for there being more (-1) factors removed from the numerator than the denominator or vice versa.

In the sinusoidal steady state, the general driving-point immittance functions are complex and of the form

$$Z(j\omega) = R(\omega) + jX(\omega) \tag{12-33}$$

or

$$Y(j\omega) = G(\omega) + jB(\omega) \tag{12-34}$$

where $R(\omega)$ = resistance, $X(\omega)$ = reactance, $G(\omega)$ = conductance, $B(\omega)$ = susceptance. According to Eqs. 12-31 and 12-32 written for the driving-point impedance (and of the same form for the driving-point admittance), $Z(j\omega)$ is purely imaginary. This follows because terms of the form $(\omega^2 - \omega_n{}^2)$ are always real, and likewise the multiplying constant H is always real as well as positive. Thus for LC networks,

$$Z(j\omega) = jX(\omega) \quad \text{and} \quad Y(j\omega) = jB(\omega) \tag{12-35}$$

Since the impedance function is purely reactive, LC networks are spoken of as *reactive networks*.

12-2. Separation property for reactive networks

When the reactance function $X(\omega)$ discussed in the previous section is differentiated with respect to radian frequency ω, the resultant func-

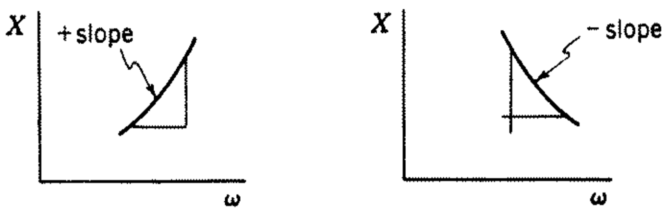


Fig. 12-4. Geometry of positive and negative slope.

tion is always positive; that is,

$$\frac{dX}{d\omega} > 0 \tag{12-36}$$

We will postpone the proof for this statement until the partial fraction expansion of $X(\omega)$ is studied. In terms of a plot of X as a function of frequency, the slope of the curve must always be positive; that is, must be increasing with increasing ω. If we start with a given value of reactance, X_1 at some frequency ω_1, then as frequency increases, X must increase, finally to an infinite value. This is illustrated in Fig. 12-5. At the frequency of infinite reactance, the sign of X changes.

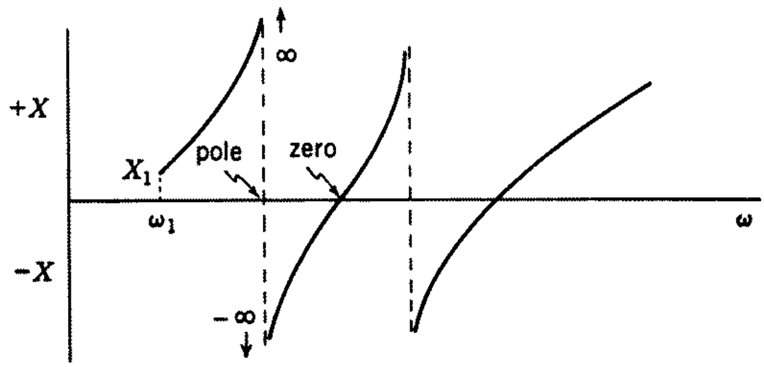


Fig. 12-5. Reactance as a function of frequency.

Starting again at $X(\omega) = -\infty$, the slope must again be positive. The curve of X is shown to increase to zero value, thence increase to infinite value before the cycle of change is repeated. The reactance function has a magnitude which varies from zero value to infinite value as frequency changes. Those values of frequency which result in a zero value for the reactance are *zeros* (*of frequency*). Frequencies resulting in infinite magnitude of reactance are *poles* (*of frequency*). The zero frequencies are also sometimes spoken of as *resonant frequencies* (frequencies of zero reactance), and the pole frequencies are called *antiresonant frequencies* (infinite reactance).

Because of the property of reactive networks that the derivative $dX/d\omega$ always be positive, the poles and zeros of the reactive network function must *alternate*. The poles must be separated by zeros and the zeros by poles. This is referred to as the *separation property* for reactive networks.

The poles and zeros of the reactance function illustrated in Fig. 12-5 can be located by inspection of the reactance function. A term of the form $(\omega^2 - \omega_n^2)$ in the numerator of $X(\omega)$ locates two zeros at $\omega = \pm\omega_n$. Similarly, a term $(\omega^2 - \omega_m^2)$ in the denominator of $X(\omega)$ locates two poles at $\omega = \pm\omega_m$. There remain only to consider the poles and zeros at zero frequency and at infinite frequency. We will start our study of these lower and upper limit frequencies by considering the behavior of elements and combinations of elements at $\omega = 0$ and $\omega = \infty$.

The reactance of an inductance varies with frequency as given by the equation

$$X_L = \omega L \tag{12-37}$$

Thus X_L has a zero at $\omega = 0$ and a pole at $\omega = \infty$, since X_L varies directly with ω. The expression for the reactance of a capacitance is

$$X_C = \frac{-1}{\omega C} \tag{12-38}$$

showing that since X_C varies inversely with ω, X_C is infinite at zero frequency so that zero frequency is a pole and is zero at infinite frequency so that infinite frequency defines a zero. These relationships for inductance and capacitance are summarized below.

ω	X_C	X_L
0	pole	zero
∞	zero	pole

In the case of LC networks, we can attach a physical significance to the terms pole and zero. A pole of reactance means an infinite value of reactance which we interpret physically as an *open circuit*. The capacitance does appear to be an open circuit at zero frequency (direct current) since the capacitor plates are not in physical contact. Similarly, an inductance appears to be an open circuit at very high (approaching infinite) frequencies. The word "choke" is applied to the inductor because of this high reactance at high frequency. By dual reasoning, a zero of reactance means zero value of reactance. Zero reactance (and thus zero impedance, since there is no resistance present in the networks being considered) is interpreted as a *short circuit*. An inductance appears to be a short circuit at zero frequency (direct current) because $(d/dt)(Li) = 0$ and no voltage appears across the inductance. Likewise, the capacitance appears to be a short circuit at infinite frequency. This physical interpretation of poles and zeros at $\omega = 0$ and $\omega = \infty$ allows the zero and infinite poles and zeros of a network to be found by inspection. Several examples will illustrate this feature.

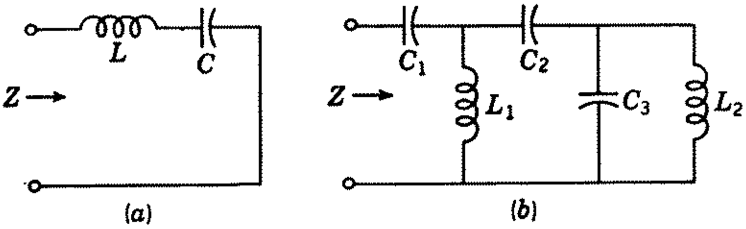


Fig. 12-6. *LC* networks for examples.

Figure 12-6(a) shows a simple series LC network. At zero frequency (direct current) the capacitor acts as an open circuit. Hence the driving-point impedance for this network has a pole at zero. At infinite frequency, the inductance acts as an open circuit. Thus the impedance function also has a pole at infinity. A more complicated network is shown in Fig. 12-6(b). At zero frequency, C_1 acts as an open circuit, making the behavior of any other elements in the network at zero frequency irrelevant. The driving-point impedance thus has a pole at zero frequency. At $\omega = \infty$, C_1, C_2, and C_3 behave as short circuits, and L_1 and L_2 as open circuits. There is a zero impedance path from terminal to terminal through C_1-C_2-C_3. Then the driving-point impedance has a zero at infinite frequency. We conclude that the network of Fig. 12-6(b) has *a pole at zero and a zero at infinity.*

The behavior of the driving-point impedance at zero and infinite frequencies can be determined from the mathematical form of the impedance function. Using s in place of $j\omega$ to simplify notation, the driving-point impedance given as Eq. 12-24 may be rewritten as

$$Z(s) = \frac{a_n s^n + a_{n-2}s^{n-2} + \ldots + a_0}{b_m s^m + b_{m-2}s^{m-2} + \ldots + b_1 s} \tag{12-39}$$

where n is even and m odd. As $s = j\omega$ approaches a very large value, only the highest-ordered term of the numerator and denominator polynomials need be considered. This is to say that

$$\lim_{s\to\infty} Z(s) = \lim_{s\to\infty} \frac{a_n s^n}{b_m s^m} \tag{12-40}$$

Now n and m can differ at most by unity and are never equal, as discussed in the last section. Hence as s approaches infinity, $Z(s)$ approaches either zero or infinity, depending on whether m is larger than n or n larger than m. In either case, because n and m can differ by unity at most, the pole or zero at infinity will be *simple* (not multiple). In summary: If the order of the numerator polynomial is greater than the order of the denominator polynomial, there will be a simple pole at infinity. If the converse is true, there will be a simple zero at infinity.

For the low-frequency case, only the lowest-ordered terms in the polynomials of the impedance function need be considered. In Eq. 12-39, the higher-order terms may be ignored as

$$Z(s) = \frac{\ldots + a_2 s^2 + a_0}{\ldots + b_3 s^3 + b_1 s} \tag{12-41}$$

The two possible cases of an even-to-odd or odd-to-even quotient of polynomials can be taken into account by considering two possibilities in this equation: (1) $a_0 \neq 0$, and (2) $a_0 = 0$. For case (1),

$$\lim_{s \to 0} Z(s) = \lim_{s \to 0} \frac{H}{s} = +\infty \tag{12-42}$$

and there is a pole at zero frequency. For case (2),

$$\lim_{s \to 0} Z(s) = \lim_{s \to 0} Hs = 0 \tag{12-43}$$

and there is a zero at zero frequency. From this discussion, we see that when the lowest-ordered term of the numerator is of higher order than the lowest-ordered term of the denominator, there will be a simple zero at zero. If the converse is true, there will be a simple pole at zero.

In all cases, zero and infinity (frequency) are either poles or zeros for LC networks. Further, these poles and zeros are always simple. Two examples will illustrate these conclusions.

Example 3

For this example, suppose that the driving-point impedance is given by the expression

$$Z(s) = 2\,\frac{(s^2 + 1)(s^2 + 9)}{s(s^2 + 4)} \tag{12-44}$$

The order of the numerator polynomial is 4, and that of the denominator is 3. Applying the rule for infinite frequency stated on page 282, infinity is a pole, since the order of the numerator is greater than that of the denominator. Since for small values of s, $Z(s)$ approaches the form H/s, zero frequency is also a pole. By inspection, there are finite

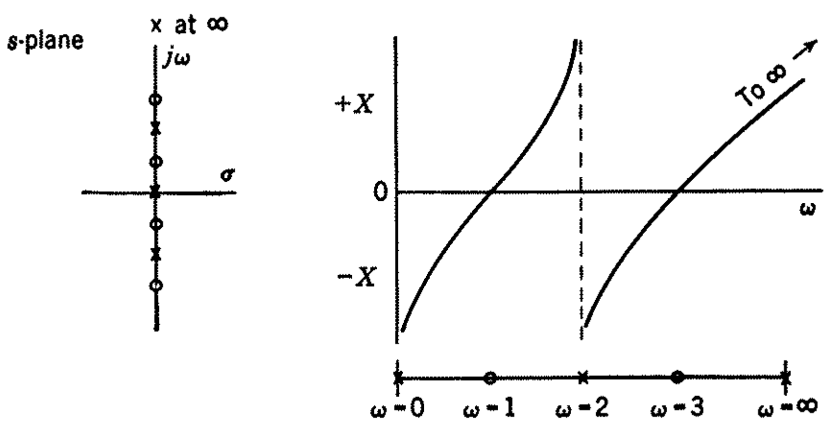


Fig. 12-7. *s* plane and reactance plot of Example 3.

zeros at $\omega = \pm 1$ and $\omega = \pm 3$, and a finite pole at $\omega = \pm 2$. The pole-zero configuration and a plot of this reactance function are shown in Fig. 12-7. The reactance function $X(\omega)$ may be found from Eq. 12-44 and different values of ω substituted into this equation to make the plot.

Example 4

For this example, consider an impedance function with the order of the denominator greater than that of the numerator. Let

$$Z(s) = 3\,\frac{s(s^2 + 4)}{(s^2 + 1)(s^2 + 9)} \tag{12-45}$$

In this equation, the order of the numerator is 3 and that of the denominator is 4. Analysis of this equation shows that both zero and infinity are zeros and that there are finite poles at $\omega = \pm 1$ and $\omega = \pm 3$

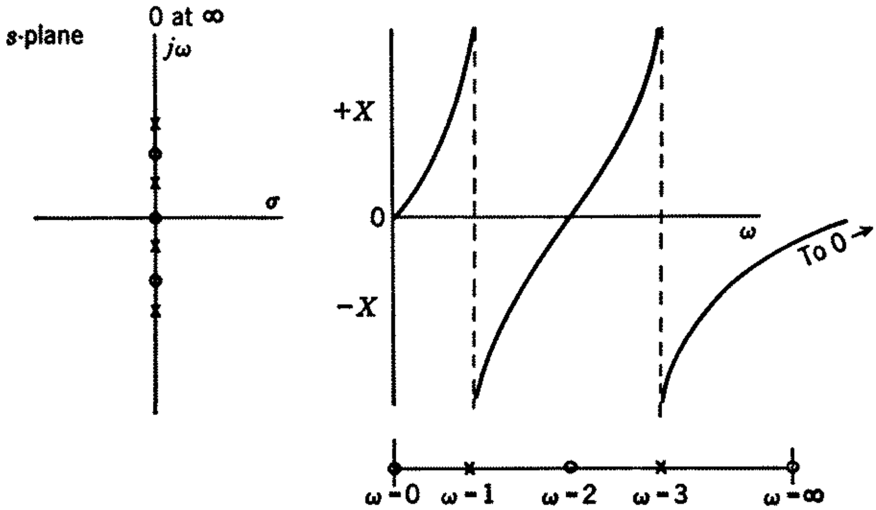


Fig. 12-8. Plots of Example 4.

and one finite zero at $\omega = \pm 2$. The s-plane representation and the reactance function plot are shown in Fig. 12-8.

12-3. The four reactance function forms

It has been shown in the previous section that zero and infinite frequencies are always either poles or zeros. The four possibilities for the two possible conditions at the two frequencies are tabulated below.

Case	$\omega = 0$	$\omega = \infty$
1	pole	pole
2	zero	zero
3	pole	zero
4	zero	pole

There remains the task of finding the form of the driving-point impedance (or admittance) corresponding to each of these four cases. There are but two forms of factors for the numerator and the denominator: $(s^2 + s_n^2)$ and s.

Case 1. With a pole at both zero and infinity, there must be an s in the denominator and one more $(s^2 + s_n^2)$ type factor in the numerator than in the denominator. The general form for this driving-point impedance is

$$Z(s) = \frac{H}{s}\frac{(s^2+\omega_1^2)\ldots(s^2+\omega_n^2)}{(s^2+\omega_2^2)\ldots(s^2+\omega_{n-1}^2)} \tag{12-46}$$

where
$$\omega_1 < \omega_2 < \omega_3 < \ldots < \omega_{n-1} < \omega_n \tag{12-47}$$

The first and last finite critical frequencies (poles or zeros) are in the numerator of this equation. The general form of the corresponding reaction function is shown in Fig. 12-9.

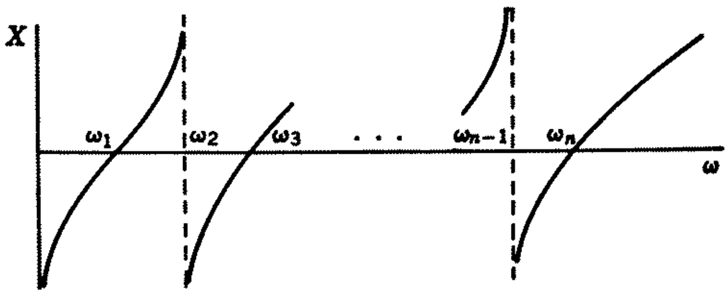


Fig. 12-9. Reactance plot for Case 1.

Case 2. This is the inverse of Case 1. With a zero both at zero and at infinity, it is necessary that there be an s term in the numerator and an additional $(s^2 + s_n^2)$ type term in the denominator. For Case 2, the driving-point impedance has the form

$$Z(s) = Hs\frac{(s^2+\omega_2^2)\ldots(s^2+\omega_{n-1}^2)}{(s^2+\omega_1^2)\ldots(s^2+\omega_n^2)} \tag{12-48}$$

where the same relationship exists for the ω's of this equation as given in Eq. 12-47. The reactance function plot for Case 2 is shown in Fig. 12-10.

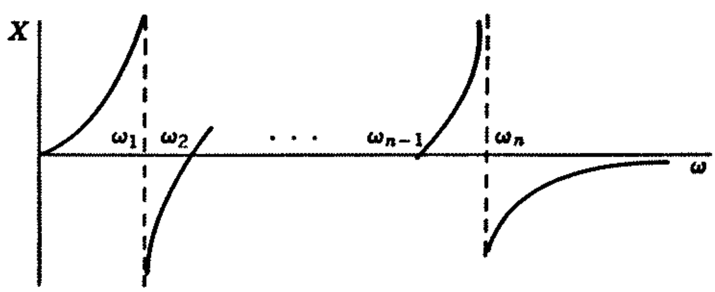


Fig. 12-10. Reactance plot for Case 2.

Case 3. To have a pole at zero frequency requires an H/s multiplier for the impedance function. For there to be a zero at infinity, the total order of the denominator must be greater than that of the numerator. Since there is already an s in the denominator, there must be just as many $(s^2 + s_n^2)$ type terms in the numerator as in the denominator.

The driving-point impedance becomes

$$Z(s) = \frac{H}{s} \frac{(s^2 + \omega_1^2)\ldots(s^2 + \omega_{n-1}^2)}{(s^2 + \omega_2^2)\ldots(s^2 + \omega_n^2)} \tag{12-49}$$

The Case 3 reactance function is plotted in Fig. 12-11.

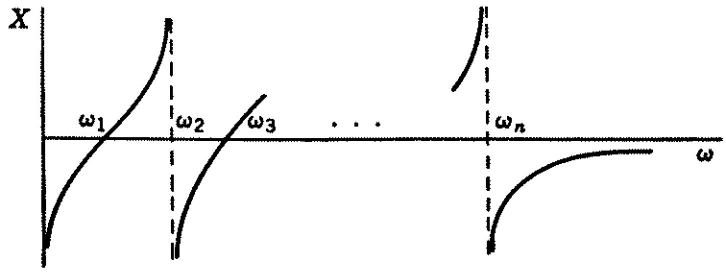


Fig. 12-11. Reactance plot for Case 3.

Case 4. For this case, there must also be an equal number of $(s^2 + s_n^2)$ factors in the numerator and denominator with, in addition, an (Hs) multiplying factor in the numerator. The driving-point impedance for Case 4 is

$$Z(s) = Hs \frac{(s^2 + \omega_2^2)\ldots(s^2 + \omega_n^2)}{(s^2 + \omega_1^2)\ldots(s^2 + \omega_{n-1}^2)} \tag{12-50}$$

The Case 4 reactance function is plotted in Fig. 12-12.

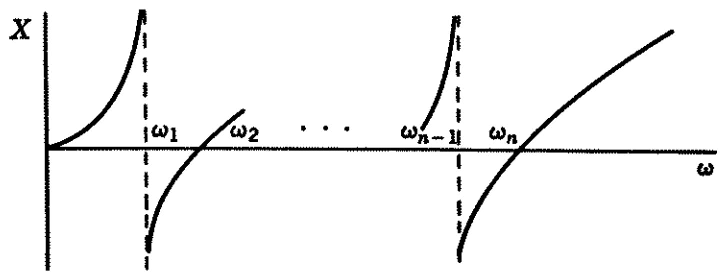


Fig. 12-12. Reactance plot for Case 4.

From the four cases just discussed, it is seen that in the sinusoidal steady state ($s = j\omega$) the form of the reactance function for Case 1 and Case 3 is

$$X(\omega) = \pm \frac{H}{\omega} \frac{(\omega^2 - \omega_1^2)(\omega^2 - \omega_3^2)\ldots}{(\omega^2 - \omega_2^2)(\omega^2 - \omega_4^2)\ldots} \tag{12-51}$$

and that for Case 2 and Case 4, the form is

$$X(\omega) = \pm H\omega \frac{(\omega^2 - \omega_2^2)(\omega^2 - \omega_4^2)\ldots}{(\omega^2 - \omega_1^2)(\omega^2 - \omega_3^2)\ldots} \tag{12-52}$$

In each case the sign of the reactance function must be selected to give a positive slope to $dX/d\omega$ (see Prob. 12-7). In order to have this positive slope, the value of X at $\omega = 0$ must either be zero or negative infinity. The sign of X changes each time the frequency passes a pole or a zero. Consider the factor

$$(\omega^2 - \omega_n^2) \tag{12-53}$$

For ω less than ω_n, the sign of the factor is negative. When ω exceeds ω_n, the sign of the factor becomes positive. Since the poles and zeros must alternate (the separation property), the sign of the reactance function alternates from positive to negative, changing successively at the pole and zero frequencies. In the two forms of the reactance function given above, the pole and zero frequencies must satisfy the condition

$$0 < \omega_1 < \omega_2 < \omega_3 < \omega_4 < \dots \tag{12-54}$$

The statements made in this section for the reactance function $X(\omega)$ apply directly to the admittance case, where $Y(j\omega) = jB(\omega)$, and $B(\omega)$ is the *susceptance function.*

The factor H which appears in all the reactance equations is a positive real constant known as the *multiplying or scale factor.* The function of H in terms of the reactance is to fix the scale of the reactance. Doubling the value of H, for example, doubles the values of the reactance function for all values of frequency. Thus H fixes the level of impedance.

12-4. Specifications for reactance functions

In this section, we will discuss the nature and number of quantities which must be known to completely specify the impedance function for an LC network. The term *critical frequency* will be defined to mean *either* a pole or a zero frequency. From the last section, we know that zero frequency and infinite frequency are always critical frequencies. These poles and zeros at zero and infinity (frequency) are defined as *external* critical frequencies. Poles and zeros at finite, nonzero frequencies are defined as *internal* critical frequencies. The internal and external critical frequencies for a particular pole-zero configuration are identified in Fig. 12-13. We know that the poles and zeros must alternate as frequency increases because of the separation property for LC networks. If the internal critical frequencies are specified as poles or zeros, there remains no choice for the external critical frequencies. They must be the opposite of the nearest finite (or internal) critical frequency in order that the poles and zeros alternate

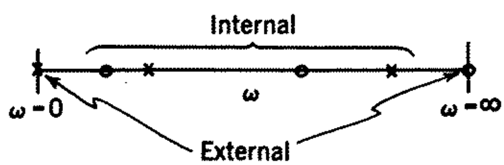


Fig. 12-13. Designations for critical frequencies.

as required. In summary: If the nature of the internal critical frequencies is specified, the nature of the external critical frequencies is fixed. The specification of internal critical frequencies specifies all critical frequencies.

We next observe, from Eqs. 12-51 and 12-52, the only possible forms of the reactance function, that if all critical frequencies are known, there remains only the scale factor H (or the equivalent) to be specified in order to determine completely the driving-point impedance function for reactive networks. In place of the value of H, an equivalent specification would be either (a) a value of the reactance at some noncritical frequency, or (b) a value of the slope of the reactance curve $dX/d\omega$ at some frequency other than a pole frequency. This information is summarized below.

Specifications for Reactive Networks

A. The internal critical frequencies.

B. One additional bit of information to give H or to allow H to be computed. The three most common forms of this second specification are:

(a) the value of H,
(b) the value of X at a noncritical frequency, or
(c) the value of $dX/d\omega$ at some nonpole frequency.

When these two types of specification are made, $Z(s)$ can be found by making the substitution $\omega = s/j$ in the equations of the form of Eq. 12-51 or Eq. 12-52. We will next study the problem of *designing* networks to meet the $Z(s)$ specification.

Reactance functions of the type studied in the chapter thus far were first investigated in 1924 by R. M. Foster, then of the Bell Telephone Laboratories but now at Brooklyn Polytechnic Institute. Features of this study are classified under the heading of *Foster's reactance theorem.*

12-5. Foster form of reactive networks

The partial fraction expansion of reactance functions may be studied by considering Case 1 and then specializing to the other three cases. The driving-point impedance function for Case 1 is given in Eq. 12-46, which is

$$Z(s) = \frac{H}{s}\frac{(s^2+\omega_1^2)(s^2+\omega_3^2)\dots}{(s^2+\omega_2^2)(s^2+\omega_4^2)\dots} \tag{12-55}$$

where there is one more factor of the form $(s^2 + \omega_1^2)$ in the numerator than in the denominator. In the partial fraction expression of this equation, terms of the type $(s^2 + \omega_2^2)$ will expand as

$$\frac{N(s)}{(s^2 + \omega_2^2)} = \frac{N(s)}{(s + j\omega_2)(s - j\omega_2)} = \frac{K_2}{(s + j\omega_2)} + \frac{K_2^*}{(s - j\omega_2)} \qquad (12\text{-}56)$$

where K_2^* is the conjugate of K_2. It is shown in advanced texts that the K-coefficients in the partial fraction expansion of terms with imaginary roots are always positive and real. This being the case, K_2 and K_2^* are equal, so that

$$\frac{K_2}{(s + j\omega_2)} + \frac{K_2^*}{(s - j\omega_2)} = \frac{2K_2 s}{s^2 + \omega_2^2} \qquad (12\text{-}57)$$

Using this form of expansion for the $(s^2 + \omega_2^2)$ terms, Eq. 12-55 for Case 1 expands as

$$Z(s) = \frac{K_0}{s} + \frac{2K_2 s}{s^2 + \omega_2^2} + \frac{2K_4 s}{s^2 + \omega_4^2} + \ldots + Hs \qquad (12\text{-}58)$$

The last term in the expansion, Hs, is necessary to give the pole at infinity in Case 1. The H value of the coefficient may be verified either by application of l'Hospital's rule or by direct division.

For a series combination of impedances, the total impedance will be the sum of the series impedances; that is,

$$Z(s) = Z_1(s) + Z_2(s) + Z_3(s) + \ldots + Z_n(s) \qquad (12\text{-}59)$$

The philosophy of the design procedure to arrive at a Foster network is to identify each term in the last equation as the impedance of a simple network configuration. These configurations will then be combined in series to give a composite network having the required driving-point impedance $Z(s)$.

Following this philosophy, we will associate the impedance Z_1 in the last equation with the term K_0/s in Eq. 12-58. Thus $Z_1(s) = K_0/s$ represents a *capacitor* of value $C = 1/K_0$. Similarly associating $Z_n(s)$ and Hs leads to the conclusion that $Z_n(s)$ represents an *inductor* of value $L = H$ henrys. All other terms in the impedance expression are of the same form and represent a parallel combination of inductor and capacitor. The impedance of such a network is

$$Z(s) = \frac{1}{Cs + 1/Ls} = \frac{(1/C)s}{s^2 + 1/LC} \qquad (12\text{-}60)$$

Comparing this equation with, for example, Eq. 12-57, gives values for the capacitors and inductors in the network as

$$C_n = \frac{1}{2K_n} \tag{12-61}$$

$$L_n = \frac{1}{\omega_n{}^2 C_n} = \frac{2K_n}{\omega_n{}^2} \tag{12-62}$$

for the nth term in the partial fraction expansion. These conclusions are summarized in Fig. 12-14.

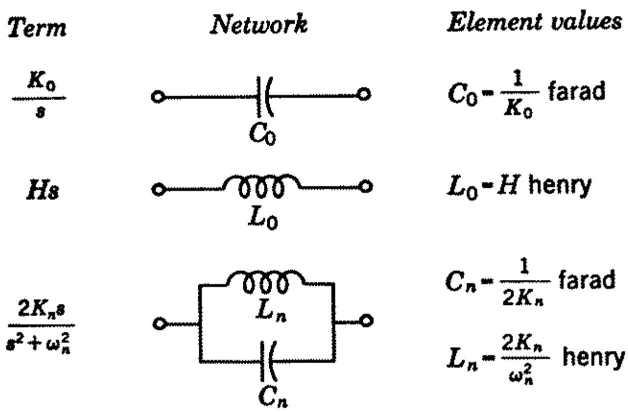


Fig. 12-14. Impedance expressions and equivalent networks.

The realization of the network corresponding to the expanded expression for $Z(s)$ is shown in Fig. 12-15. This realization is known as the *first Foster form* (series impedances). The network of the figure is for Case 1. There remains the problem of specializing to the other

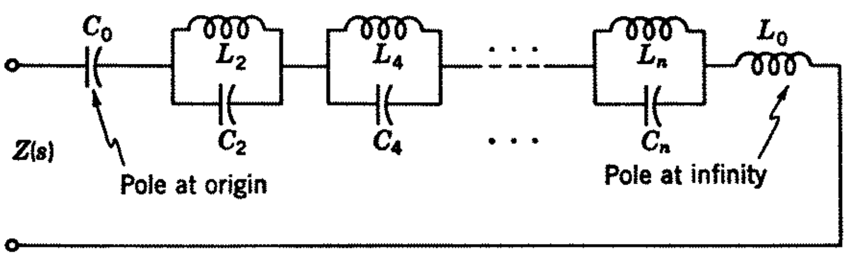


Fig. 12-15. Network of the first Foster type.

three cases. Before undertaking this study, let us consider the role of each element in the network in terms of the known pole-zero configuration.

(1) *Capacitor* C_0. Capacitor C_0 appears in the network because of the term K_0/s in the partial fraction expansion. This term corresponds to a *pole at the origin*. Hence the presence of C_0 depends on there being a pole at the origin.

(2) *Inductor* L_0. Inductor L_0 appears in the network from the impedance expression Hs in the partial fraction expansion. This term is pres-

ent because of the *pole at infinity*. Hence the presence of L_0 depends on there being a pole at infinity.

(3) *Parallel L_nC_n network*. Each term of the type $(s^2 + \omega_n^2)$ in the denominator of $Z(s)$ gives rise to a parallel LC network. The frequency ω_n of each term is a pole frequency. Thus antiresonance in the individual LC parallel networks gives rise to the poles of $Z(s)$. The zeros of $Z(s)$ cannot be associated with any specific elements. In terms of the reactance function $X(\omega)$, each of the parallel circuits changes sign as the frequency increases through antiresonance. At some frequency, the reactance of a group of parallel circuits is equal to and opposite in sign to the reactance of all remaining parallel circuits. Under this condition there is a zero of $Z(s)$. There will be as many zeros as there are poles because of the separation property. These zeros can be thought of as being caused by the first parallel network being in "series" resonance (the resonance of zero impedance) with the rest of the network, then the first and second parallel networks in resonance with the remaining network and so on, until finally the last parallel network resonates with the combined preceding network.

Since the distinguishing features of the four cases of reactive networks considered are the poles and zeros at zero and infinity, it follows that Cases 2, 3, and 4 can be specialized from Case 1 simply by leaving out either or both of C_0 and L_0. For example, if there is no s term in the denominator of $Z(s)$, there will be no K_0/s term in the partial fraction expansion. Similarly, if the order of the denominator polynomial is greater than the numerator polynomial, there will be no Hs factor in the partial fraction expansion, and consequently, in terms of the physical elements, no series inductor in the circuit. The nature of the "end elements" for the four cases is summarized below.

End Element Values in First Foster Networks

Case	$\omega = 0$	$\omega = \infty$	C_0	L_0
1	pole	pole	present	present
2	zero	zero	short-circuited	short-circuited
3	pole	zero	present	short-circuited
4	zero	pole	short-circuited	present

In his 1924 paper, Foster pointed out that if the admittance corresponding to a given impedance function $Y(s) = 1/Z(s)$ is determined, it is possible to realize a physical network for the admittance function. Networks found by an admittance expansion are known as networks of the *second Foster form*.

Consider an admittance function of Case 1 with a pole at both zero and infinity. This function has the same form as Eq. 12-55 with $Y(s)$

replacing $Z(s)$. The partial fraction expansion for $Y(s)$ is the same as Eq. 12-58. There remains the problem of identifying individual terms in the partial fraction expansion as admittance expressions for LC network configurations. The driving-point admittance of parallel networks is the sum of the admittances of the networks in parallel; that is,

$$Y(s) = Y_1(s) + Y_2(s) + Y_3(s) + \dots + Y_n(s) \qquad (12\text{-}63)$$

Comparing this equation with Eq. 12-58, we see that the term K_0/s corresponds to the first admittance, so that $Y_1(s) = K_0/s$. Then $Y_1(s)$ represents the admittance of an inductor of value $L = 1/K_0$ henry. Similarly $Y_n(s) = Hs$ is the admittance of a capacitor of H farad value. All other terms in the admittance expression are the same and represent the admittance of a series inductor and capacitor (the dual of the network found for the first Foster form). The admittance of the series LC network is

$$Y(s) = \frac{1}{Ls + 1/Cs} = \frac{s/L}{s^2 + 1/LC} \qquad (12\text{-}64)$$

Comparing this equation with Eq. 12-57 permits identification of the required values of L and C for the network corresponding to the nth pole of admittance as

$$L_n = \frac{1}{2K_n} \qquad (12\text{-}65)$$

$$C_n = \frac{1}{L_n\omega_n{}^2} = \frac{2K_n}{\omega_n{}^2} \qquad (12\text{-}66)$$

These conclusions are summarized in Fig. 12-16.

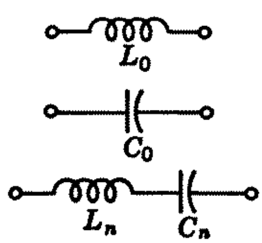
Term	*Network*	*Element values*
$\frac{K_0}{s}$		$L_0 = \frac{1}{K_0}$ henry
Hs		$C_0 = H$ farad
$\frac{2K_n s}{s^2+\omega_n^2}$		$L_n = \frac{1}{2K_n}$ henry $C_n = \frac{2K_n}{\omega_n^2}$ farad

Fig. 12-16. Admittance expressions and equivalent networks.

The combination of networks of the types shown in Fig. 12-16 to conform with Eq. 12-63 for parallel admittances is shown in Fig. 12-17. The network shown is for Case 1 but specializes to the other three cases just as in the first Foster type of network. The role of each type of network configuration in terms of the poles and zeros of the driving-point admittance function is summarized as follows:

(1) *Inductor* L_0. Inductor L_0 appears because of the pole (of Y) at the origin which gives rise to the K_0/s term in the partial fraction expansion.

(2) *Capacitor* C_0. Capacitor C_0 appears because of the pole (of Y) at infinity which causes the Hs term to appear in the partial fraction expansion.

(3) *Series* L_nC_n *network.* Each series LC network corresponds to a factor of the type $(s^2 + \omega_n^2)$ in the denominator of $Y(s)$. The frequency ω_n of each such term is a pole frequency which is associated

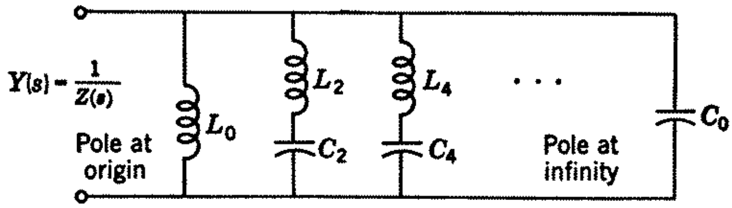


Fig. 12-17. Network of the second Foster type.

with *resonance* (in contrast to antiresonance in the series impedance realization) of the LC networks. Thus all poles can be associated directly with specific elements in the sense that resonance within individual networks cause the entire network to have a pole of admittance. As in the case of the first Foster form, the zeros of $Y(s)$ cannot be so identified. Every element contributes some part to the conditions associated with a zero of admittance, in much the same way as discussed for the first Foster form.

The "end elements" (that is, L_0 and C_0) are associated with the poles of $Y(s)$ at zero and infinity (frequency). The presence of a zero in place of a pole thus infers the absence of an end element. End conditions for the four cases of the second Foster type of networks are summarized below for impedance functions.

End Element Values in Second Foster Networks

Case	$\omega = 0$	$\omega = \infty$	L_0	C_0
1	pole	pole	present	present
2	zero	zero	absent	absent
3	pole	zero	present	absent
4	zero	pole	absent	present

From this discussion, we see that from any specifications two completely equivalent networks can be designed. The first Foster network form is a series impedance realization of $Z(s)$; the second Foster network form is found by forming $Y(s)$ by inverting $Z(s)$, as $Y(s) = 1/Z(s)$, and then expanding as parallel admittance functions which are identified with specific network forms. It is important to observe that $Z(s)$ and $Y(s)$ will never be of the same case classifications for one set of specifications. Inverting a function interchanges poles and zeros, and since $Z(s)$ will never have the same number of internal poles as internal zeros, the case designation will differ for $Z(s)$ and $Y(s) = 1/Z(s)$. The following conclusions can easily be verified.

If a given $Z(s)$ is:	The corresponding $Y(s)$ is:
Case 1	Case 2
Case 2	Case 1
Case 3	Case 4
Case 4	Case 3

An example will illustrate the procedure for finding the first and second Foster networks for a given $Z(s)$. Consider the impedance

$$Z(s) = 2\,\frac{(s^2+1)(s^2+9)}{s(s^2+4)} \tag{12-67}$$

This impedance function is of Case 1 and the partial fraction expansion is

$$Z(s) = \frac{K_0}{s} + \frac{K_2}{s+j2} + \frac{K_2{}^*}{s-j2} + Hs \tag{12-68}$$

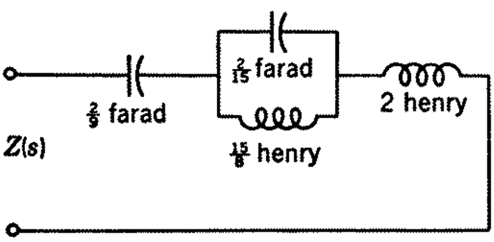


Fig. 12-18. First Foster network of example.

The constants of this equation may be evaluated by the Heaviside rule as follows.

$$K_0 = \frac{2\times 9}{4} = \frac{9}{2}$$

$$K_2 = \left.\frac{2(s^2+1)(s^2+9)}{s(s-j2)}\right|_{s=-j2} = \frac{15}{4}$$

and since $H = 2$ by inspection,

$$Z(s) = \frac{9/2}{s} + \frac{2(15s/4)}{s^2+4} + 2s \tag{12-69}$$

Using the equations displayed in Fig. 12-14, the element values and first Foster form network are found as shown in Fig. 12-18.

To determine the second Foster form of network, the admittance function is found as

$$Y(s) = \frac{1}{Z(s)} = \frac{1}{2}\,\frac{s(s^2+4)}{(s^2+1)(s^2+9)} \tag{12-70}$$

$$= \frac{2K_1 s}{s^2+1} + \frac{2K_3 s}{s^2+3^2} \tag{12-71}$$

The two constants K_1 and K_3 are found by the Heaviside method as

$$K_1 = \frac{s(s^2+4)}{2(s-j1)(s^2+9)}\bigg|_{s=-j1} = \frac{3}{32}$$

$$K_3 = \frac{s(s^2+4)}{2(s^2+1)(s-j3)}\bigg|_{s=-j3} = \frac{5}{32}$$

Making use of the chart of Fig. 12-16, the network configuration and element values are determined as shown in Fig. 12-19. The two equivalent networks have the same number of elements.

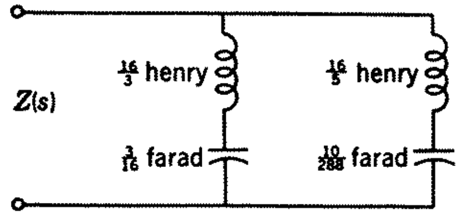


Fig. 12-19. Second Foster network of example.

Before we consider two other forms of equivalent networks, let us digress to consider the unfinished business of a proof for the separation property. Now that we have completed the partial fraction expansion for immittance functions, we know that any immittance function is composed of, at most, three types of terms:

$$\frac{K_0}{s}, \qquad \frac{2K_n s}{s^2+\omega_n^2}, \qquad Hs \tag{12-72}$$

The reactance expressions for these terms are

$$X_1 = -\frac{K_0}{\omega} \tag{12-73}$$

$$X_n = \frac{2K_n\omega}{-\omega^2+\omega_n^2} \tag{12-74}$$

$$X_3 = \omega H \tag{12-75}$$

(Susceptance expressions have exactly the same form.) The derivatives of these three expressions are

$$\frac{dX_1}{d\omega} = +\frac{K_0}{\omega^2} \tag{12-76}$$

$$\frac{dX_n}{d\omega} = \frac{+2K_n\omega^2 + 2K_n\omega_n^2}{(\omega_n^2-\omega^2)^2} \tag{12-77}$$

$$\frac{dX_3}{d\omega} = +H \tag{12-78}$$

Each of the three typical terms is positive for both positive and negative values of frequency ω. Since the reactance function for the driving-point terminals (and the same thing might be said for the susceptance function) is given as a sum of reactances,

$$X = \sum_{j=1}^{n} X_j \tag{12-79}$$

which may be differentiated term by term to give

$$\frac{dX}{d\omega} = \sum_{j=1}^{n} \frac{dX_j}{d\omega} \tag{12-80}$$

Then $dX/d\omega$ is positive,

$$\frac{dX}{d\omega} > 0 \tag{12-81}$$

for both positive and negative values of frequency ω. From this conclusion, the separation property follows.

12-6. Cauer form of reactive networks

An important extension to the Foster reactance theorem was made in Germany by W. Cauer* in 1927. He first pointed out that the reactance function could be represented by two different network configurations by a *continued fraction expansion* of the driving-point impedance. The basic form of the network for the Cauer realization is the ladder

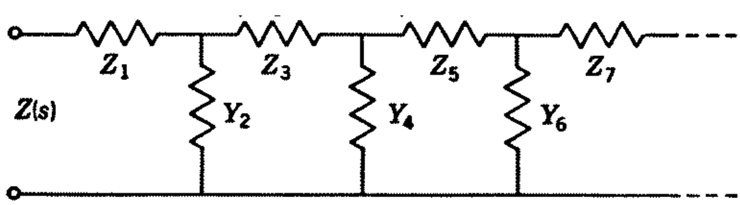


Fig. 12-20. Ladder network.

shown with impedance and admittance designations in Fig. 12-20. The driving-point impedance of such a network may be written in the form of a continued fraction as

$$Z(s) = Z_1 + \cfrac{1}{Y_2 + \cfrac{1}{Z_3 + \cfrac{1}{Y_4 + \cfrac{1}{Z_5 + \cfrac{1}{Y_6 + \cdots}}}}} \tag{12-82}$$

* W. Cauer, *Arch. Elektrotech.*, **17**, 355 (1927).

Let us restrict our discussion to an impedance function with a pole at infinity. In the general expression for the driving-point impedance

$$Z(s) = \frac{a_n s^n + a_{n-2} s^{n-2} + \dots}{b_m s^m + b_{m-2} s^{m-2} + \dots} \tag{12-83}$$

this means that n is greater than m. The procedure for forming the continued fraction is to divide, then invert and divide, invert and divide, and continue this process until the expansion terminates (as it must). This procedure can best be illustrated with a numerical example. Consider the impedance function

$$Z(s) = \frac{12s^4 + 12s^2 + 1}{6s^3 + 3s} \tag{12-84}$$

Direct division proceeds as follows.

$$\begin{array}{rll} 6s^3 + 3s) & 12s^4 + 12s^2 + 1 & (2s \\ & \underline{12s^4 + 6s^2} & \\ & 6s^2 + 1 & \end{array}$$

so that

$$Z(s) = 2s + \frac{6s^2 + 1}{6s^3 + 3s} \tag{12-85}$$

Inverting the remainder term and dividing gives

$$\begin{array}{rll} 6s^2 + 1) & 6s^3 + 3s & (s \\ & \underline{6s^3 + s} & \\ & 2s & \end{array}$$

or

$$Z(s) = 2s + \frac{1}{s + 2s/(6s^2 + 1)} \tag{12-86}$$

Continuing the invert and divide procedure,

$$\begin{array}{rll} 2s) & 6s^2 + 1 & (3s \\ & \underline{6s^2} & \\ & 1 & \end{array}$$

such that, finally,

$$Z(s) = 2s + \cfrac{1}{s + \cfrac{1}{3s + \cfrac{1}{2s}}} \tag{12-87}$$

Comparing this expression with Eq. 12-82, we see that: $Z_1 = 2s$ represents an inductor of 2 henrys, $Y_2 = s$ represents a capacitor of 1 farad, $Z_3 = 3s$ represents an inductor of 3 henrys, $Y_4 = 2s$ represents a capacitor of 2 farads. The network configuration for this specific example is shown in Fig. 12-22. For this case, a pole at infinity, the first Cauer network will always be of the form of Fig. 12-21 with "series" inductors and "parallel" capacitors extending as far as required by the continued fraction expansion. There is only one other

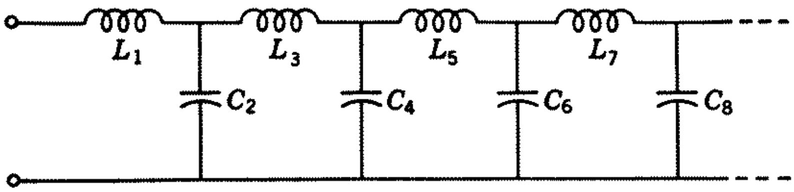


Fig. 12-21. General form of first Cauer network (pole at ∞).

case to be considered: a zero at infinity. For this case, m exceeds n by unity in Eq. 12-83, and before the continued fraction expansion can be made, it is necessary to invert the polynomial. The form of the expansion will be

$$Z(s) = \cfrac{1}{Y_2 + \cfrac{1}{Z_3 + \cfrac{1}{Y_4 + \cfrac{1}{Z_5 + \dots}}}} \tag{12-88}$$

Comparing this equation with Eq. 12-82, we see that the only difference is that $Z_1 = 0$ in the second case. If $Z_1 = 0$, then L_1, the first series inductor, is absent from Fig. 12-22. The form of the first Cauer network with a zero at infinity is shown in Fig. 12-23. In summary, the *first* or *"end" element* in the first Cauer network form is determined by the behavior of the network function at infinite frequency. A pole at infinity requires that the first element be a series inductor. A zero at infinity requires that the first element be a parallel (or shunt) capacitor.

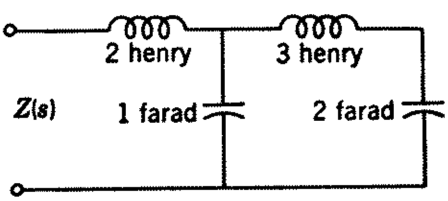


Fig. 12-22. First Cauer network for Eq. 12-84.

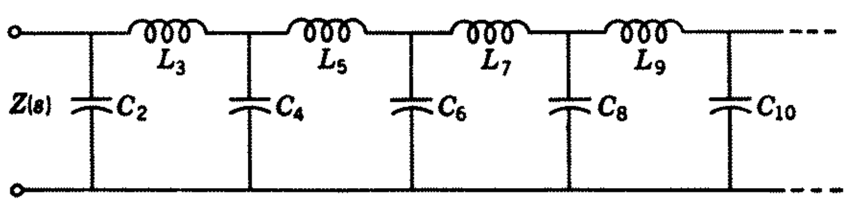


Fig. 12-23. General form of first Cauer network (zero at ∞).

Let us focus our attention on the way the Cauer network *ends* now that we have studied the factor controlling the way it begins. There are but two possible ways the ladder network can end. One is with an inductor as shown in Fig. 12-24(a); the other is with a capacitor as in Fig. 12-24(b). Imagine these networks attached to the general form of the network (with any number of elements in a ladder arrangement) shown in Figs. 12-22 and 12-23. With the inductor as the last element, there are series inductors in a path from one terminal to the other. At

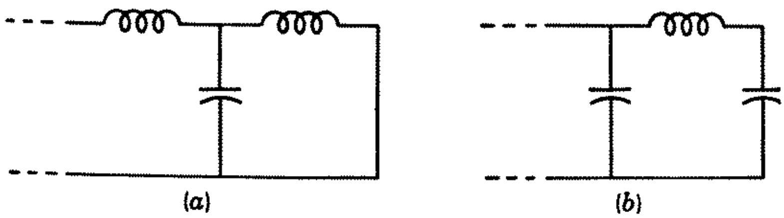


Fig. 12-24. Last element forms for first Cauer network.

zero frequency (direct current), these inductors offer zero impedance such that zero frequency is a zero of impedance. Alternately, if the last element is a capacitor, there is a "break" in the path from terminal to terminal and at zero frequency there is a pole of impedance. In summary: The *last* (or far end) element in a first Cauer network is determined by the nature of the impedance function at zero frequency. A zero at zero requires that the last element be an inductor. A pole at zero requires a capacitor as the last element. These conclusions are summarized below.

First Cauer Network End Elements

Case	$\omega = 0$	$\omega = \infty$	First element	Last element
1	pole	pole	L	C
2	zero	zero	C	L
3	pole	zero	C	C
4	zero	pole	L	L

The basic form for the *second Cauer network* is again the ladder network, but with the position of capacitor and inductor interchanged.

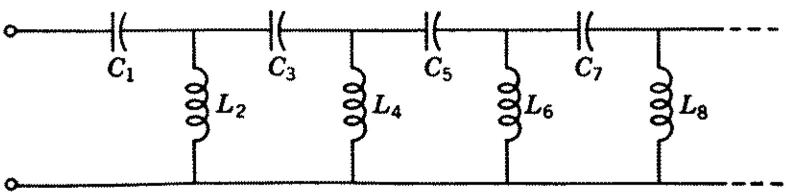


Fig. 12-25. Second Cauer network form (pole at 0).

Such a ladder network is illustrated in Fig. 12-25. The impedance of the first element is $Z_1(s) = 1/C_1s$; similarly, the admittance of the second (parallel or shunt) element is $Y_2 = 1/L_2s$. The continued fraction expansion for the realization of this network form must be

$$Z(s) = \frac{1}{C_1 s} + \cfrac{1}{\cfrac{1}{L_2 s} + \cfrac{1}{\cfrac{1}{C_3 s} + \cdots}} \tag{12-89}$$

To obtain this form of continued fraction expansion will require a different procedure than that used for the first Cauer network. To illustrate, consider the case of an impedance function with a pole at zero. An example of such a function is

$$Z(s) = \frac{s^4 + 7s^2 + 6}{2s^3 + 3s} \tag{12-90}$$

To expand this function in the form of Eq. 12-89, we first turn it end for end as

$$Z(s) = \frac{6 + 7s^2 + s^4}{3s + 2s^3} \tag{12-91}$$

The "invert-and-divide" procedure may now be started. We divide the denominator into the numerator to obtain one term; the remainder is then inverted and the division repeated. In our example,

$$\begin{array}{r|l}
3s + 2s^3) & 6 + 7s^2 + s^4 \quad (2/s \\
 & \underline{6 + 4s^2} \\
 & \phantom{6 +{}} 3s^2 + s^4
\end{array}$$

or

$$Z(s) = \frac{2}{s} + \frac{3s^2 + s^4}{3s + 2s^3} \tag{12-92}$$

Inverting the remainder term and dividing gives

$$\begin{array}{r|l}
3s^2 + s^4) & 3s + 2s^3 \quad (1/s \\
 & \underline{3s + s^3} \\
 & s^3
\end{array}$$

such that $Z(s)$ is

$$Z(s) = \frac{2}{s} + \cfrac{1}{\cfrac{1}{s} + \cfrac{s^3}{3s^2 + s^4}} \tag{12-93}$$

The final "invert-and-divide" step gives

$$\begin{array}{r|l}
s^3) & 3s^2 + s^4 \quad (3/s \\
 & \underline{3s^2} \\
 & \phantom{3s^2 +{}} s^4
\end{array}$$

such that the final form of the continued fraction is

$$Z(s) = \frac{2}{s} + \cfrac{1}{\cfrac{1}{s} + \cfrac{1}{\cfrac{3}{s} + \cfrac{1}{\cfrac{1}{s}}}} \tag{12-94}$$

Comparing this expression with Eqs. 12-82 and 12-89, the following identifications are made:

$Z_1(s) = \dfrac{2}{s}$ represents a capacitor of $\frac{1}{2}$ farad.

$Y_2(s) = \dfrac{1}{s}$ represents an inductor of 1 henry.

$Z_3(s) = \dfrac{3}{s}$ represents a capacitor of $\frac{1}{3}$ farad.

$Y_4(s) = \dfrac{1}{s}$ represents an inductor of 1 henry.

The network configuration of the second Cauer form, which is equivalent to the last equation, is shown in Fig. 12-26. The impedance function for LC networks must be an even-to-odd or odd-to-even quotient of polynomials. For the procedure just shown by example to work, the quotient must be an even-to-odd polynomial. This is equivalent to saying that the function expanded must have a pole at zero. If the impedance function has a zero at zero, it is necessary to invert first before dividing such that the continued fraction will be of the form

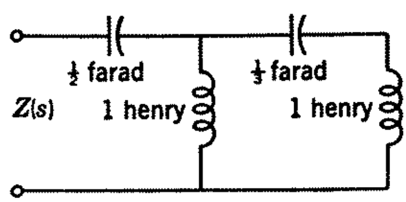


Fig. 12-26. Second Cauer network for example.

$$Z(s) = \cfrac{1}{\cfrac{1}{L_2 s} + \cfrac{1}{\cfrac{1}{C_2 s} + \cfrac{1}{\cfrac{1}{L_4 s} + \cdots}}} \tag{12-95}$$

Comparison of Eq. 12-95 and Eq. 12-89 shows that with a zero at zero the first element is an inductor and that C_1 is not present. In summary: The first element in the second Cauer network is a capacitor if the impedance function has a pole at zero; it is an inductor if the impedance function has a zero at zero.

By analogy to the other Cauer case, the infinite frequency behavior of the network is determined by the way the network ends; that is, whether the last element is an inductor or a capacitor. The two possible networks are shown in Fig. 12-27. Attaching these terminating

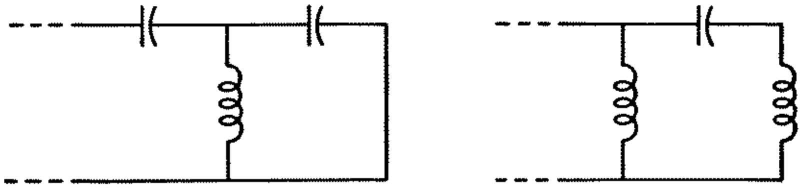


Fig. 12-27. Last element forms for second Cauer network.

networks to the general network of Fig. 12-25, we can see that if the last element is a capacitor, the network has zero impedance at infinite frequency, there being a short-circuited path from terminal to terminal. On the other hand, if the last element is an inductor, the network has a pole of impedance at infinite frequency. In summary: The last element in the second Cauer network is an inductor if the impedance function has a pole at infinity; it is a capacitor if the impedance function has a zero at infinity. These conclusions are summarized below.

SECOND CAUER NETWORK END ELEMENTS

Case	$\omega = 0$	$\omega = \infty$	First element	Last element
1	pole	pole	C	L
2	zero	zero	L	C
3	pole	zero	C	C
4	zero	pole	L	L

12-7. Choice of network realizations

In discussing the specifications for a reactive network leading to the summary on page 288 it was found that the specification of (1) the *internal* critical frequencies and (2) the scale factor H, or some equivalent specification, was sufficient to fix the driving-point impedance $Z(s)$. In the last two sections, we have shown that a network can be realized in four basic forms from the driving-point impedance only. In any network there are as many unknowns as there are elements, and these unknowns must be specified by $Z(s)$. In the solution of any system of equations, there must be as many specifications as there are unknowns. The total number of specifications that we have found sufficient is one more than the number of internal critical frequencies. The total number of unknowns equals the total number of elements. From the equality of specifications and unknowns, we conclude that the minimum number of elements in a network realization is one more than the number of internal critical frequencies. It should be noted that in this statement each *pair* of conjugate critical frequencies count as *one* critical frequency. We do not distinguish between ω_n and $-\omega_n$.

A knowledge of the number of elements required to realize a network specification along with our association of end elements with network behavior at zero and infinite frequency can be used in drawing the four possible network configurations directly by inspection of the pole-zero configuration. As example will illustrate the procedure.

Example 5

Consider the pole-zero specifications given in Fig. 12-28. The internal critical frequencies consist of two poles and one zero for $Z(s)$. The

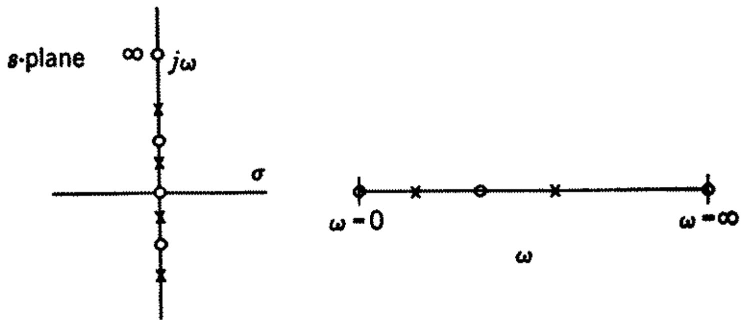


Fig. 12-28. Pole-zero configuration of Example 5.

external critical frequencies are constrained by the separation property to be zeros. Since there are three internal critical frequencies, there must be *four* elements in each network realization. Consider the realizations one at a time.

(1) *First Foster network.* Because zero and infinity are both zeros, the end elements are missing in the basic Foster form of network Fig. 12-15. The network must have two parallel LC networks to give the two poles (antiresonant frequencies). The network is shown in Fig. 12-29(a).

(2) *Second Foster network.* In finding the critical frequencies for $Y(s) = 1/Z(s)$, the poles of $Z(s)$ become zeros of $Y(s)$ and vice versa. Since both zero and infinity are poles, both end elements are present. The one internal pole is caused by a single series LC network in the basic network shown in Fig. 12-17. The four-element network is shown in Fig. 12-29(b).

(3) *First Cauer network.* The end elements will first be found for the first Cauer network. Referring to the table of page 299, we see that a zero of $Z(s)$ at infinity means the first element is a capacitor. Also since there is a zero of $Z(s)$ at zero, the last element is an inductor.

There being only four elements in the network, specification of the first and last element determines the schematic shown in Fig. 12-29(c).

(4) *Second Cauer network.* The table of page 302 may be used to advantage in investigating the second Cauer network form. From the

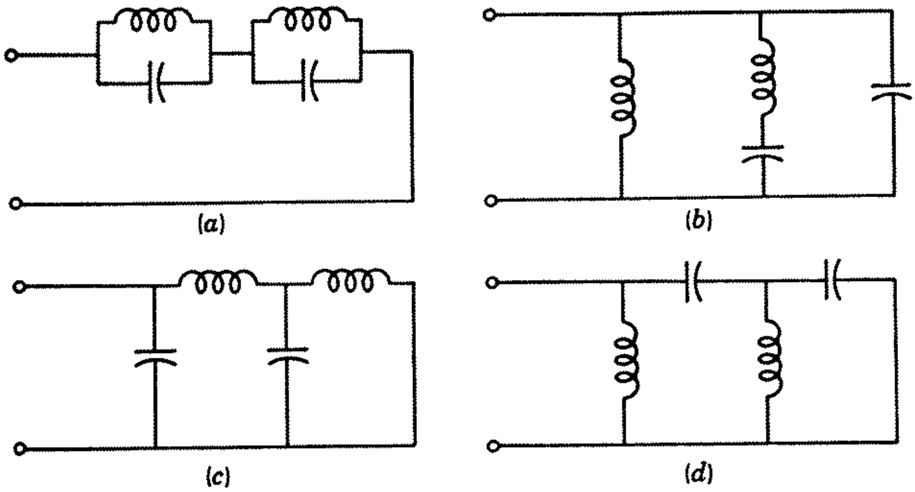


Fig. 12-29. The four network realizations of Example 5.

table, a zero at zero implies that the first element is an inductor. The zero at infinity identifies the last element as a capacitor. The four-element realization is shown in Fig. 12-29(d).

To find element values, the value of H must be given, and the partial or continued fraction expansions must be completed.

There are a number of practical matters involved in selecting a network from the four possible forms for a specific application. These include:

(1) *Element component values.* There are but a limited range of component sizes available in commercial quantities. For example, 1-farad capacitors with any reasonable voltage rating are hard to come by. Economic factors may thus give one network form the advantage over the other three.

(2) *Stray capacitance.* It is impossible to construct inductors without stray capacitance. This capacitance can be taken into account by reducing the size of the parallel capacitor in the first Foster form of realization. This is especially important when specifications are rigid or when the operating frequency is high.

(3) *Use in vacuum tube circuits.* Vacuum tube circuits frequently require blocking capacitors in interstage coupling networks. This requirement may specify which network must be used.

12-8. Use of normalized frequency

The examples given in this chapter have made use of small integer critical frequency values to advantage in simplifying numerical operations. In many practical problems, the critical frequencies are in thousands or millions of radians per second. In such cases, the arithmetic can be simplified by normalizing frequency to some value that makes the critical frequency values small integers. To normalize frequency and observe the effect on element values, let

$$X_L = \omega L_{act} = \omega_0 L_{act}\left(\frac{\omega}{\omega_0}\right) = L_{norm}\omega' \tag{12-96}$$

and

$$B_C = \omega C_{act} = \omega_0 C_{act}\left(\frac{\omega}{\omega_0}\right) = C_{norm}\omega' \tag{12-97}$$

where L_{act} = the actual inductance, C_{act} = the actual capacitance

$$L_{norm} = \text{the normalized inductance} = \omega_0 L_{act} \tag{12-98}$$

$$C_{norm} = \text{the normalized capacitance} = \omega_0 C_{act} \tag{12-99}$$

$$\omega' = \text{the normalized frequency} = \omega/\omega_0 \tag{12-100}$$

The actual element values can be found from the normalized values from the equations

$$L_{act} = \frac{L_{norm}}{\omega_0} \tag{12-101}$$

$$C_{act} = \frac{C_{norm}}{\omega_0} \tag{12-102}$$

With frequency normalized, the actual values of capacitance and inductance can be found from the normalized values found in the partial fraction or continued fraction expansion. An example will illustrate the procedure in normalizing an equation in frequency.

Example 6

A driving-point impedance is known to have zeros at 1000 and 4000 cycles per second and a pole at 3000 cycles per second. From the data, the reactance function is

$$X(\omega) = -H\frac{[-\omega^2 + (2\pi \times 10^3)^2][-\omega^2 + (8\pi \times 10^3)^2]}{\omega[-\omega^2 + (6\pi \times 10^3)^2]} \tag{12-103}$$

Let the normalizing frequency be

$$\omega_0 = 2\pi \times 10^3 \quad \text{radians/sec} \tag{12-104}$$

Several other choices of normalizing frequency might have been made. Substituting $\omega = \omega_0\omega'$ into Eq. 12-103 gives

$$X(\omega') = \frac{-H\omega_0}{\omega'}\frac{(-\omega'^2 + 1)(-\omega'^2 + 16)}{(-\omega'^2 + 9)} \tag{12-105}$$

Suppose that it is required that the reactance be 100 ohms when $\omega = 4\pi \times 10^3$ radians/sec corresponding to $\omega' = 2$. Substituting these values into Eq. 12-105 fixes the value of $H\omega_0$ at 1000/36 so that the normalized impedance function becomes

$$Z(s') = \frac{1000(s'^2 + 1)(s'^2 + 16)}{36s'(s'^2 + 9)} \tag{12-106}$$

After normalized values of the elements are found by the expansion of this equation, the actual values can be found by division by ω_0 as

$$L_{act} = \frac{L_{norm}}{2\pi \times 1000} \tag{12-107}$$

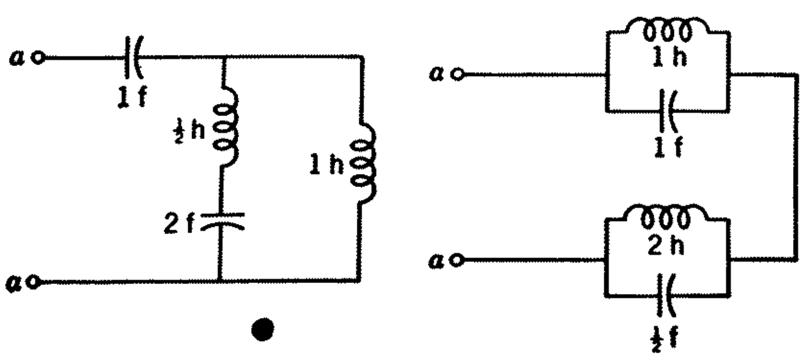
$$C_{act} = \frac{C_{norm}}{2\pi \times 1000} \tag{12-108}$$

FURTHER READING

For further study on the topic of one-terminal-pair networks, E. A. Guillemin's *Communications Networks, Vol. II* (John Wiley & Sons, Inc., New York, 1935), Chap. 5, is recommended, as well as D. F. Tuttle, Jr., *Network Synthesis*, 2 vols. (John Wiley & Sons, New York, in preparation). Source material may be found in articles by R. M. Foster, "A reactance theorem," *Bell System Tech. J.*, **3**, 259 (1924), and W. Cauer, "Die Verwirklichung von Wechselstromwiderständen vorgeschriebener Frequenzabhängigkeit," *Arch. Elektrotech.*, **17**, 355 (1927).

PROBLEMS

12-1. For the networks shown in the figure, find the driving-point impedance as a quotient of polynomials. Identify the even and odd

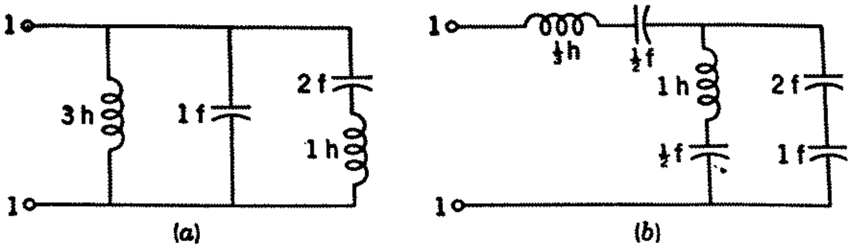


(a) (b)

Prob. 12-1.

polynomials. Compare the order of the numerator and denominator polynomials. *Answer.* (a) $(s^4 + 4s^2 + 1)/(3s^3 + s)$; (b) $3s/(s^2 + 1)$.

12-2. Repeat Prob. 12-1 but find the driving-point admittance of the networks shown below. *Answer.* (a) $(s^4 + 11s^2/6 + 1/6)/(s^3 + s/2)$; (b) $3(s^3 + 7s/2)/(s^4 + 14s^2 + 30)$.



Prob. 12-2.

12-3. A certain LC network is known to have a driving-point impedance with poles at the frequencies of 0 and 2 radians/sec and zeros at 1 and 3 radians/sec. Determine the driving-point impedance as a quotient of polynomials (expanded) if H, the multiplying factor, is unity. *Answer.* $Z(s) = (s^4 + 10s^2 + 9)/(s^3 + 4s)$.

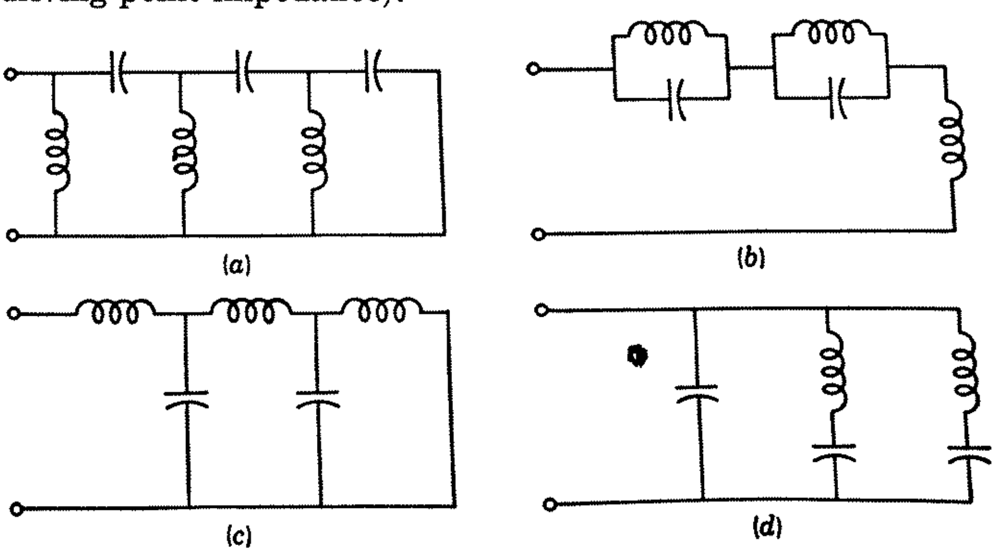
12-4. Which of the following functions may represent driving-point impedances for LC networks? In each case, why?

(a) $\pm 73 \times \dfrac{(\omega^2 - 4)(\omega^2 - 25)}{(\omega^2 - 16)(\omega^2 - 64)}$ (b) $\pm \dfrac{10}{j\omega} \times \dfrac{(\omega^2 - 25)(\omega^2 - 36)}{(\omega^2 - 49)}$

(c) $\pm j\omega \times \dfrac{(\omega^2 - 3)}{(\omega^2 - 1)(\omega^2 - 5)}$ (d) $\pm j5\omega \times \dfrac{(\omega^2 - 16)(\omega^2 - 25)}{(\omega^2 - 4)(\omega^2 - 9)}$

Answer. (a) No—no pole or zero at ∞; (b) No—separation property.

12-5. For the LC networks shown in the figure, determine: (a) whether zero frequency represents a pole or a zero, and (b) whether infinity (frequency) represents a pole or a zero of the impedance function. Do this by inspection of the networks (and not by determining the driving-point impedance).



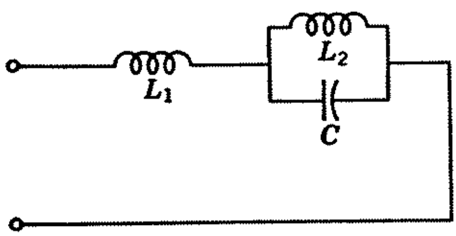
Prob. 12-5.

12-6. For the following network functions (representing LC networks), (a) sketch the pole-zero configuration in the s plane, and (b) sketch the reactance function X as a function of frequency ω. Examples of the type of sketches desired are given in Fig. 12-7 and in Fig. 12-8.

(a) $5s\dfrac{(s^2+36)(s^2+100)}{(s^2+16)(s^2+81)}$ (b) $\dfrac{10}{s}\dfrac{(s^2+1)(s^2+36)}{(s^2+9)(s^2+64)}$

(c) $\dfrac{s(s^2+5)(s^2+49)}{(s^2+4)(s^2+25)(s^2+81)}$ (d) $64\dfrac{(s^2+25)(s^2+81)(s^2+144)}{s(s^2+64)(s^2+100)}$

12-7. In this problem, we will consider which sign should be used in Eqs. 12-51 and 12-52. Show that for Case 2 and Case 3 the sign of the equation should be negative and that for case 1 and Case 4 the sign should be positive.



Prob. 12-8.

12-8. A network function has a pole at $\omega = 4$ and a zero at $\omega = 10$. These two are the only internal critical frequencies. It is required that the magnitude of reactance be 100 ohms at $\omega = 6$ radians/sec. Determine (a) the schematic diagrams of the *two* Foster networks corresponding to these specifications, and (b) the element values for the two networks.

12-9. A reactive network is to be designed to serve as the load for a vacuum tube amplifier. The following specifications are given for the LC impedance function. (1) The internal critical frequencies are: 1000 cycles/sec (a zero), 3000 cycles/sec, 4000 cycles/sec. (2) The slope of the reactance vs. frequency curve must be 100 ohms per kilocycle/sec at a frequency of 1000 cycles per second. From these specifications: (a) Sketch the pole-zero configuration and the $X(\omega)$ curve. (b) Determine the schematic diagrams for the two Foster networks. (c) Determine each element value in the two networks of part (b). (d) Which of the two networks would you select for a pràctical application? Consider such factors as estimated cost of elements, taking into account the stray capacitance of coils, etc.

12-10. A driving-point impedance is given by the equation,

$$Z(s) = \frac{15s^5 + 29s^3 + 6s}{7.5s^4 + 7s^2 + 1}$$

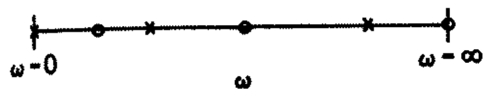
For this impedance function: (a) Determine the first Cauer network configuration. (b) Determine the value of each element in the network. (c) Find the nature of the external critical frequencies (that is, does zero frequency represent a pole or a zero, etc.). (d) Sketch the pole-zero configuration for $Z(s)$.

12-11. A driving-point impedance is given as

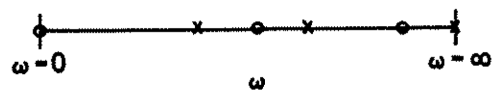
$$Z(s) = \frac{2s^4 + 13s^2 + 5}{12s^3 + 5s}$$

For this impedance function: (a) Determine the second Cauer network configuration. (b) Find the value for each element in the network. (c) Investigate the nature of the external critical frequencies (are they poles or zeros?). (d) Sketch the pole-zero configuration for $Z(s)$.

12-12. For the pole-zero configuration shown, draw the schematic diagrams of the two Foster and two Cauer networks. (Do not determine element values.) Plots are of impedance unless otherwise noted.



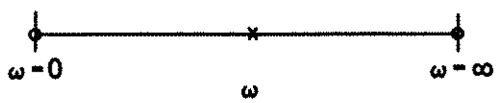
Prob. 12-12.



Prob. 12-13.

12-13. For the pole-zero configuration shown, draw the schematic diagrams of the two Foster and two Cauer networks. (Do not determine element values.)

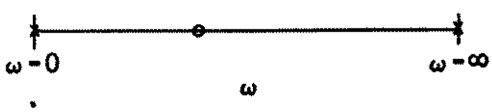
12-14. Repeat Prob. 12-13 for the pole-zero configuration of the figure.



Prob. 12-14.

12-15. Starting with Eq. 12-103, verify Eq. 12-106.

12-16. Draw the two Foster and two Cauer networks for the pole-zero configuration shown in the figure.



Prob. 12-16.

12-17. The following specifications are made for an LC network: (a) the first element must be a capacitor in series (to avoid a d-c path), (b) the network must have zero impedance at $\omega = 2$ radians/sec, (c) at 1 radian/sec, the impedance must have a magnitude of 10 ohms; that is, $|Z(j1)| = 10$, (d) the network must have the smallest possible number of elements. Draw the network schematic and indicate element values. *Answer.* $L = \frac{10}{3}$ henrys, $C = \frac{3}{40}$ farad.

CHAPTER 13

TWO-TERMINAL-PAIR REACTIVE NETWORKS (FILTERS)

The discussions in Chapter 12 were confined to networks with *one* terminal pair (the driving-point terminals). In this chapter, we will study *two*-terminal-pair networks. One of the terminal pairs will be identified as the *input*, the other as the *output*. Our ultimate objective is to design networks to give a specified relationship between voltages or currents at one terminal pair and voltages or currents at the other. The concepts in this chapter are thus *transfer* as well as driving-point in nature in contrast to exclusively driving-point in Chapter 12. A representation of a two-terminal-pair network is shown in Fig. 13-1. In the work to follow, the two terminals marked 1-1 will be identified with the input and the two marked 2-2 with the output unless otherwise specifically noted.

Fig. 13-1. Two-terminal-pair network.

13-1. The ladder network

Most of our studies will concern the *ladder* network structure. The ladder structure is important historically; it was the first structure used in constructing a design procedure for filters. In addition, the concepts developed for the ladder structure can be applied to other structures

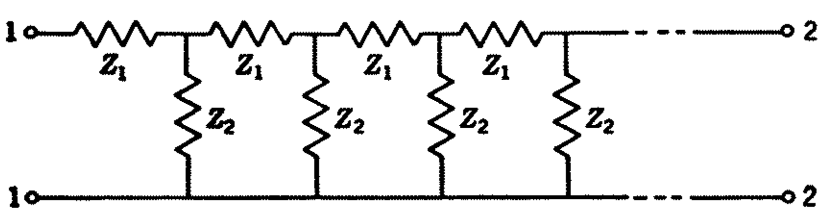


Fig. 13-2. Standard ladder network designations.

such as the lattice. *A standard* ladder network is shown in Fig. 13-2. By convention, the impedance of all series elements is $Z_1(s)$ and of all shunt elements is $Z_2(s)$. For our studies, it will be convenient to separate the standard ladder network into two other network structures: the T section and the π section. This separation and the resulting values for the series and shunt impedances is illustrated in Fig. 13-3.

The network shown in Fig. 13-3(c), together with the specific impedance designations, will be adopted as a *standard* T *section*. The network of Fig. 13-3(d) will likewise be adopted as a *standard* π *section*. All equations we shall develop for T or π sections will refer to these specific standard networks. The T and π section building blocks can

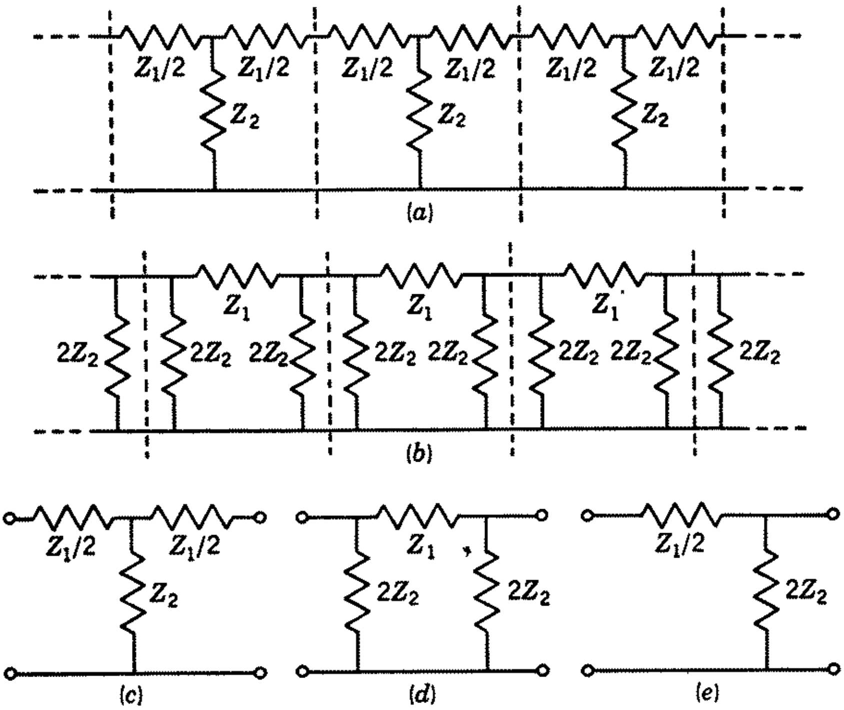


Fig. 13-3. Evolution of the ladder network into tandem networks of (a) T sections; (b) π sections; (c) a T section; (d) a π section; and (e) an L section.

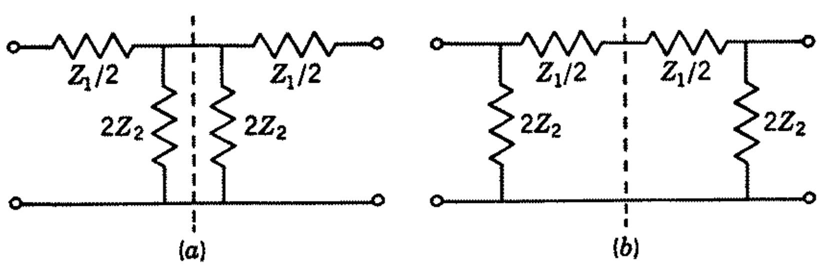


Fig. 13-4. (a) T section from two L sections; (b) π section from two L sections.

be divided one step further into a more primitive network shown in Fig. 13-3(e). This primitive network is designated as a *standard* L *section* (although the term "inverted L," or gamma network when turned end-for-end, might seem more appropriate from the point of view of geometrical similarity). The construction of the standard T section and the standard π section from the primitive L section is shown in Fig. 13-4. These three network structures will form the basis of the studies to follow.

13-2. Image impedance

The T section and π section just discussed are *symmetrical* in the sense that terminals 1-1 and 2-2 could be interchanged. The L section, however, is *unsymmetrical*. In order to derive an equation that will apply to either a symmetrical or an unsymmetrical network, consider the unsymmetrical T section shown in Fig. 13-5. Let the generator impedance be Z_g and the load impedance be Z_L. We will also

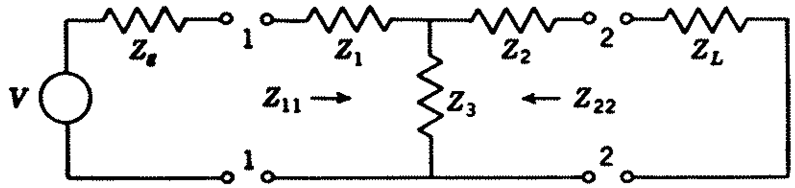


Fig. 13-5. T network.

define Z_{11} as the impedance at terminals 1-1 with Z_L connected (and Z_g disconnected), and similarly Z_{22} as the impedance at terminals 2-2 with Z_g connected (but Z_L disconnected). When the impedances Z_g and Z_L are adjusted such that

$$Z_g = Z_{11} \quad \text{and} \quad Z_L = Z_{22} \tag{13-1}$$

an *image match* is said to exist at terminals 1-1 and 2-2. To add emphasis to the special impedance defined by Eq. 13-1, Z_{11} will be written Z_{1i}, the image impedance at terminal pair 1, and similarly Z_{22} will be written Z_{2i}, the image impedance at terminal pair 2.

The reason for using the word *image* is suggested by Eq. 13-1. Under the specified conditions, the impedance seen "looking in" at terminals 1-1 is the same as that "seen" in a mirror (constructed to see impedance) which views the generator impedance. An image match exists when the driving-point impedance is the same as the image impedance of the generator if terminals 2-2 are also terminated in their image impedance.

For the network of Fig. 13-5, expressions for Z_{1i} and Z_{2i} can be written in terms of Z_1, Z_2, and Z_3. These equations are

$$Z_{1i} = Z_1 + \frac{Z_3(Z_2 + Z_{2i})}{Z_2 + Z_3 + Z_{2i}} \tag{13-2}$$

$$Z_{2i} = Z_2 + \frac{Z_3(Z_1 + Z_{1i})}{Z_1 + Z_3 + Z_{1i}} \tag{13-3}$$

We have here two equations in two unknowns, Z_{1i} and Z_{2i}. By routine algebraic operation, it is possible to solve for the unknowns. However, a more useful result is found by solving for Z_{1i} and Z_{2i} in terms of open-circuit and short-circuit impedances. Let

Z_{1o} = the impedance at terminal-pair 1 with terminal-pair 2 open,
Z_{1s} = the impedance at terminal-pair 1 with terminal-pair 2 short-circuited,
Z_{2o} = the impedance at terminal-pair 2 with terminal-pair 1 open,
Z_{2s} = the impedance at terminal-pair 2 with terminal-pair 1 short-circuited.

For the network of Fig. 13-5, the open-circuit and short-circuit impedances for terminal pair 1 have the values

$$Z_{1o} = Z_1 + Z_3 \tag{13-4}$$

$$Z_{1s} = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} \tag{13-5}$$

By algebraic manipulation of the last four equations, it is found that

$$Z_{1i} = \sqrt{Z_{1o} Z_{1s}} \tag{13-6}$$

Similarly,

$$Z_{2i} = \sqrt{Z_{2o} Z_{2s}} \tag{13-7}$$

These two equations are the foundation of much of the analysis to follow. For the symmetrical network, the image impedances are equal. For this case, the notation will be simplified by letting

$$Z_{1i} = Z_{2i} = Z_i \tag{13-8}$$

Several examples, important in terms of the discussion to follow, will be given next.

Image Impedance of the T *Section.* For the T section shown in Fig. 13-3(c) the image impedance is

$$Z_{iT} = \sqrt{Z_{1o} Z_{1s}} = \sqrt{\left(\frac{Z_1}{2} + \frac{Z_1 Z_2/2}{Z_1/2 + Z_2}\right)\left(\frac{Z_1}{2} + Z_2\right)} \tag{13-9}$$

$$Z_{iT} = \sqrt{Z_1^2/4 + Z_1 Z_2} \tag{13-10}$$

Image Impedance of the π Section. The π section is shown in Fig. 13-3(d). The image impedance is found as follows:

$$Z_{i\pi} = \sqrt{Z_{1o} Z_{1s}} = \sqrt{\frac{2Z_2(Z_1 + 2Z_2)}{2Z_2 + 2Z_2 + Z_1} \frac{2Z_1 Z_2}{(Z_1 + 2Z_2)}} \tag{13-11}$$

$$Z_{i\pi} = \frac{Z_1 Z_2}{\sqrt{Z_1^2/4 + Z_1 Z_2}} \tag{13-12}$$

Image Impedance of the L *Section.* The L section shown in Fig. 13-3(e) is reproduced below as Fig. 13-6. This network is unsymmetrical, and the image impedance must be computed for each terminal pair. At terminals 1-1, the image impedance is

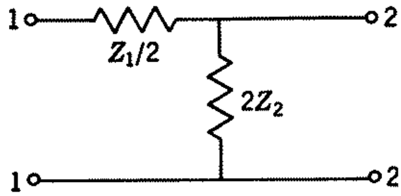


Fig. 13-6. L section.

$$Z_{1iL} = \sqrt{(Z_1/2 + 2Z_2)Z_1/2} = \sqrt{Z_1^2/4 + Z_1Z_2} \qquad (13\text{-}13)$$

Comparison with Eq. 13-10 shows that

$$Z_{1iL} = Z_{iT} \qquad (13\text{-}14)$$

or that the image impedance of the L section at terminals 1-1 is the image impedance of the T section. At the other terminal pair

$$Z_{2iL} = \sqrt{(2Z_2)\frac{2Z_2(Z_1/2)}{2Z_2 + Z_1/2}} = \frac{Z_1Z_2}{\sqrt{Z_1^2/4 + Z_1Z_2}} \qquad (13\text{-}15)$$

Comparison of this equation with Eq. 13-12 shows that

$$Z_{2iL} = Z_{i\pi} \qquad (13\text{-}16)$$

Thus at terminal pair 2 the image impedance is that of the symmetrical π section. The image impedance of the L section appears as a T section looking in one direction, and as a π section looking in the other. From another point of view this conclusion seems reasonable. Two L sections can be combined *with an image match* to form a T section, with the image impedance Z_{iT} on each end as shown in Fig. 13-7. Similarly,

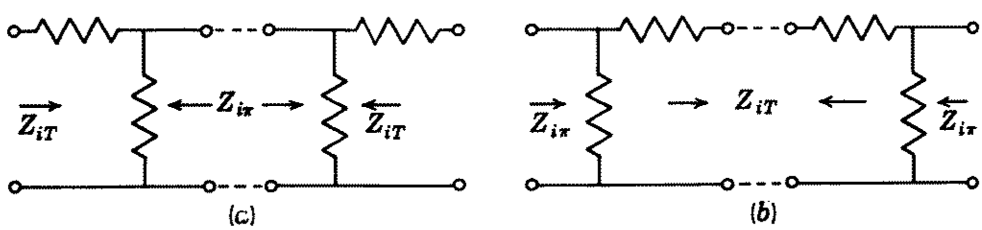


Fig. 13-7. Combination of image matched L sections to form (a) the T section, and (b) the π section.

L sections combine with an image match to form a π section with $Z_{i\pi}$ at both terminal pairs as required.

13-3. Image transfer function

It is evident that we need something in addition to the image impedance concept in the two-terminal-pair problem. The image impedance is a driving-point concept. We need a function to relate variables at one terminal pair to the other terminal pair. For the time being, we will restrict the transfer function to the ratio of currents; thus

$$\frac{I_1(s)}{I_2(s)} = G(s) \tag{13-17}$$

where I_1 is the input current, I_2 is the output (or load) current, and $G(s)$ is the transfer function. In the sinusoidal steady state, the transfer function becomes a complex number which may be expressed as a magnitude and phase angle as

$$G(j\omega) = |G(j\omega)|e^{j\,\mathrm{Ang}\,G(j\omega)} \tag{13-18}$$

In practice, the magnitude $|G(j\omega)|$ is measured in a logarithmic unit (the neper) so that

$$|G(j\omega)| = e^{\alpha} \tag{13-19}$$

The angle of $G(j\omega)$ is designated β, so that

$$G(j\omega) = e^{\alpha}e^{j\beta} = e^{\gamma} \tag{13-20}$$

where α = the attenuation (nepers), β = the phase shift (radians), γ = the image transfer function. As a practical matter, the most common unit for attenuation is the decibel, abbreviated db, even though the definition for the decibel involves the ratio of powers—not voltages or currents—as follows.

$$\alpha_{db} = 10 \log_{10} (P_1/P_2) \qquad \text{db} \tag{13-21}$$

If the power ratio is related to the voltage or current ratio as

$$\frac{P_1}{P_2} = \left|\frac{E_1}{E_2}\right|^2 \quad \text{or} \quad \frac{P_1}{P_2} = \left|\frac{I_1}{I_2}\right|^2 \tag{13-22}$$

Then

$$\alpha_{db} = 20 \log_{10} \left|\frac{E_1}{E_2}\right| \quad \text{or} \quad 20 \log_{10} \left|\frac{I_1}{I_2}\right| \tag{13-23}$$

To find the number of decibels corresponding to a neper, under the restriction that Eq. 13-22 applies, we substitute for the current ratio in the last equation as

$$\alpha_{db} = 20 \log_{10} e^{\alpha} = 8.686\,\alpha \qquad \text{db} \tag{13-24}$$

Thus α_{db} in decibels is found by multiplying α in nepers by the factor 8.686 (the neper being the larger unit).

It should be emphasized that the quantities α and β are *transfer* in nature. The attenuation is a measure of the ratio of the magnitude of the input current (which must be sinusoidal for α to have meaning) to the magnitude of the output current. The phase angle β is the phase of the input sinusoid measured with respect to the output sinusoid. Knowing the input current and the image transfer function, the output current is determined when the network is terminated in the image impedance.

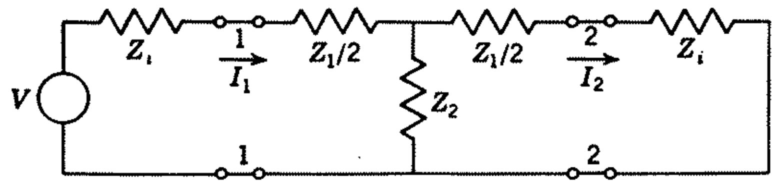


Fig. 13-8. Network for computing image transfer function.

To find the image transfer function for a symmetrical network, consider the T section shown in Fig. 13-8. The currents are related by the equation

$$\frac{I_1}{I_2} = e^\gamma = \frac{Z_2 + Z_1/2 + Z_i}{Z_2} = 1 + \frac{Z_1}{2Z_2} + \frac{Z_i}{Z_2} \tag{13-25}$$

Solving this equation for Z_i, there results

$$Z_i = Z_2\left[(e^\gamma - 1) - \frac{Z_1}{2Z_2}\right] \tag{13-26}$$

This image impedance is the image impedance of the T section which is given by Eq. 13-10, which is

$$Z_{iT}^2 = \frac{Z_1^2}{4} + Z_1Z_2 \tag{13-27}$$

Squaring Eq. 13-26 and equating this squared equation to Eq. 13-27, there results, after common terms are canceled,

$$Z_2^2(e^{2\gamma} - 2e^\gamma + 1) - Z_1Z_2e^\gamma = 0 \tag{13-28}$$

or

$$\frac{e^{2\gamma} - 2e^\gamma + 1}{e^\gamma} = \frac{Z_1}{Z_2} \tag{13-29}$$

This equation can be put in hyperbolic form by recognizing that $\frac{1}{2}(e^\gamma + e^{-\gamma}) = \cosh\gamma$, so that finally

$$\cosh\gamma = 1 + \frac{Z_1}{2Z_2} \tag{13-30}$$

A similar expression for the hyperbolic sine is found by making use of the identity

$$\cosh\gamma + \sinh\gamma = e^\gamma \tag{13-31}$$

Comparing this equation with Eq. 13-25, it is seen that

$$\sinh \gamma = \frac{Z_i}{Z_2} \tag{13-32}$$

Dividing Eq. 13-32 by 13-30 and canceling the common Z_2, an equation is found for the hyperbolic tangent of γ, which is

$$\tanh \gamma = \frac{Z_i}{Z_1/2 + Z_2} \tag{13-33}$$

This equation can be expressed in terms of the open- and short-circuit parameters by making use of Eq. 13-6.

$$Z_i = \sqrt{Z_{1o}Z_{1s}} \tag{13-34}$$

(which could be written for side 2 by replacing the 1's by 2's, since the network is symmetrical) and the expression for the open-circuit impedance for the T network.

$$Z_1/2 + Z_2 = Z_{1o} \tag{13-35}$$

Using these two identities, Eq. 13-33 may be written

$$\tanh \gamma = \sqrt{Z_{1s}/Z_{1o}} \tag{13-36}$$

This is a most useful form of the equation for the image-transfer function, which with Eq. 13-34 forms the basis of much of the discussion to follow. This equation is more general than is implied by our derivation. It holds for any passive reciprocal network. (See Prob. 13-2, for example.)

13-4. Application to LC networks

In a very important class of two-terminal-pair networks, all the elements within the network are inductors and capacitors. Since all practical elements have resistance, any result based on the assumption

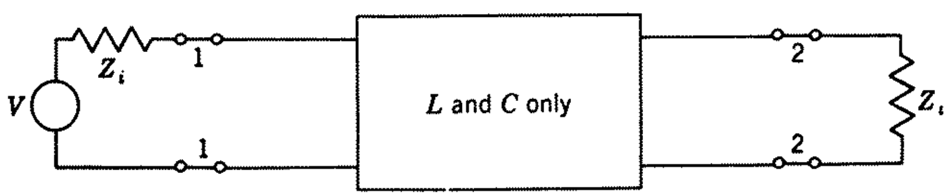


Fig. 13-9. Two-terminal-pair LC network.

of purely LC networks is approximate. The results conform to measurements sufficiently well, however, to be of engineering value. From Chapter 12, we know that the driving-point impedance of an LC network is purely reactive (and the admittance purely susceptive) such that

$$Z(j\omega) = \pm jX(\omega) \quad \text{and} \quad Y(j\omega) = \pm jB(\omega) \tag{13-37}$$

The foundation equations for our analysis are

$$Z_i = \sqrt{Z_{1o}Z_{1s}} = \sqrt{Z_{2o}Z_{2s}} \tag{13-38}$$

$$\tanh \gamma = \sqrt{Z_{1s}/Z_{1o}} = \sqrt{Z_{2s}/Z_{2o}} \tag{13-39}$$

for symmetrical networks. In the equations to follow, the subscript 1 will be used with the understanding that it can be replaced by a 2 as in these equations given above, as long as the network is symmetrical. For reactive networks in the sinusoidal steady state, equations for the image impedance and image transfer function become

$$Z_i = \sqrt{(\pm jX_{1o})(\pm jX_{1s})} \tag{13-40}$$

$$\tanh \gamma = \sqrt{\frac{\pm jX_{1s}}{\pm jX_{1o}}} \tag{13-41}$$

From our knowledge of the properties of LC networks discussed in Chapter 12, we know that the sign of the reactance function changes with frequency for driving-point reactances. Here X_{1s} and X_{1o} are driving-point reactance functions, although their quotient relates to a transfer function. As frequency changes, the sign as well as the magnitude of X_{1s} and X_{1o} changes. There are four possible sign conditions summarized below.

Case	X_{1s}	X_{1o}	Z_i	$\tanh \gamma$
1	+	+	jX_i	real
2	−	−	jX_i	real
3	+	−	R_i	imaginary
4	−	+	R_i	imaginary

If the signs are the same for X_{1s} and X_{1o}, the image impedance is imaginary and $\tanh \gamma$ is real. For opposite signs, Z_i is real and $\tanh \gamma$ is imaginary. These are the only choices. Both Z_i and $\tanh \gamma$ must be real *or* imaginary, but can *never be complex*. We next turn our attention to an investigation of the conditions under which $\tanh \gamma$ can be real or imaginary. Since $\gamma = \alpha + j\beta$, $\tanh \gamma$ can be expanded as

$$\tanh \gamma = \frac{\sinh \gamma}{\cosh \gamma} = \frac{\sinh \alpha \cos \beta + j \cosh \alpha \sin \beta}{\cosh \alpha \cos \beta + j \sinh \alpha \sin \beta} \tag{13-42}$$

Dividing both numerator and denominator of the equation by the factor $\cosh \alpha \cos \beta$ gives

$$\tanh \gamma = \frac{\tanh \alpha + j \tan \beta}{1 + j \tanh \alpha \tan \beta} \tag{13-43}$$

Our problem is to make this expression either purely real or purely imaginary. There are several possibilities.

(1) Let $\alpha = 0$ such that $\tanh \alpha = 0$. For this condition,

$$\tanh \gamma = j \tan \beta \qquad (\text{for } \alpha = 0) \tag{13-44}$$

and $\tanh \gamma$ is purely imaginary.

(2) Let $\beta = 0$ such that $\tan \beta = 0$; then

$$\tanh \gamma = \tanh \alpha \qquad (\text{for } \beta = 0, \pm\pi, \pm 2\pi, \ldots) \tag{13-45}$$

and $\tanh \gamma$ is purely real.

(3) Let $\beta = \pi/2$ (or odd multiples of this angle) such that $\tan \beta$ approaches infinity. In the limit,

$$\tanh \gamma = \frac{1}{\tanh \alpha} \qquad \left(\text{for } \beta = \pm\frac{\pi}{2}, \pm\frac{3\pi}{2}, \pm\frac{5\pi}{2}, \ldots\right) \tag{13-46}$$

and again $\tanh \gamma$ is purely real.

Any other values of α and β will make $\tanh \gamma$ complex, and this is not permitted. Hence there are *only three* possibilities for values for α and for β, as summarized below.

Value of $\tanh \gamma$	Conditions of α and β	
Real	$\alpha = \tanh^{-1}\sqrt{\dfrac{X_{1s}}{X_{1o}}}$, $\beta = 0, \pm\pi, \pm 2\pi, \ldots$	(13-47)
	or	
	$\alpha = \tanh^{-1}\sqrt{\dfrac{X_{1o}}{X_{1s}}}$, $\beta = \pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2}, \ldots$	(13-48)
Imaginary	$\alpha = 0$, $\beta = \tan^{-1}\sqrt{\dfrac{X_{1s}}{X_{1o}}}$	(13-49)

It is now possible to extend the table given on page 318 to include the values of α and β for the four cases.

Case	jX_{1s}	jX_{1o}	Z_i	$\tanh \gamma$	α	β
1	+	+	jX_i	real	$\alpha \neq 0$	0 or $\pi/2$
2	−	−	jX_i	real	$\alpha \neq 0$	0 or $\pi/2$
3	+	−	R_i	imaginary	0	$\beta \neq 0$
4	−	+	R_i	imaginary	0	$\beta \neq 0$

We now have sufficient information to examine the different cases in terms of both a transfer quantity, the image transfer function, and a driving-point quantity, the image impedance. At some value of frequency, assume that the signs of X_{1s} and X_{1o} are such that the conditions of Case 1 or Case 2 apply. Then the image impedance at terminals 1-1 is imaginary. This is the impedance of the network presented to the generator when terminated in Z_i. We associate this reactive load with the condition of no power transfer. From the transfer point of view, the attenuation α is positive and real, which means that the load current I_2 is *smaller* than the generator current I_1 (see Eq. 13-17). The current is attenuated or, so to speak "stopped." Under the conditions of Case 1 and Case 2, the frequencies are designated *stop frequencies*, and the band of stop frequencies is designated the *stop band.*

When the signs of X_{1s} and X_{1o} are opposite, we have Case 3 and Case 4, where the image impedance is real and the attenuation α is zero. From the driving-point impedance point of view, the load is now resistive and there is power transfer. With no attenuation, the magnitude of I_2 is equal to the magnitude of I_1 (although there will be a difference in the phase of the two currents). For these cases, the current is "passed." Such frequencies as give the conditions of Case 3 and Case 4 are designated *pass frequencies.* A band of pass frequencies is identified as a *pass band.* The frequency of transition from pass band to stop band or vice versa is assigned the name *cutoff frequency.*

The reactance function X_{1s} and X_{1o} vary with frequency according to several rules discussed in Chapter 12:

(1) The slope of the reactance curve $dX/d\omega$ is always positive.
(2) As a consequence of the slope property, the poles and zeros (or points of resonance and antiresonance) *alternate* as a function of frequency.
(3) The external frequencies (that is, $\omega = 0$ and $\omega = \infty$) are always either poles or zeros.

Since the critical frequencies of the reactance function determine the nature of the reactance versus frequency plot, we suspect that the critical frequencies somehow relate to the pass band, the stop band, and the cutoff frequencies. We can study the relationships by comparing the reactance curves made for X_{1o} and X_{1s}. Both plots will have the general appearance of the plot shown in Fig. 13-10. In comparing the critical frequencies of X_{1o} and X_{1s}, we recognize that there are three possibilities: (1) the critical frequencies will coincide but be opposite in nature, (2) the critical frequencies will coincide but be the same type, and (3) the critical frequencies will not coincide. These three possibilities are illustrated in Fig. 13-11.

The reactance plots for X_{1s} and X_{1o} for the first possibility are shown in Fig. 13-12. At noncritical frequencies, the signs for X_{1s} and X_{1o} are *always opposite*. At critical frequencies, the two functions change signs at the same time to preserve this opposite sign nature. From the table on page 319, we see that opposite signs for X_{1s} and X_{1o}

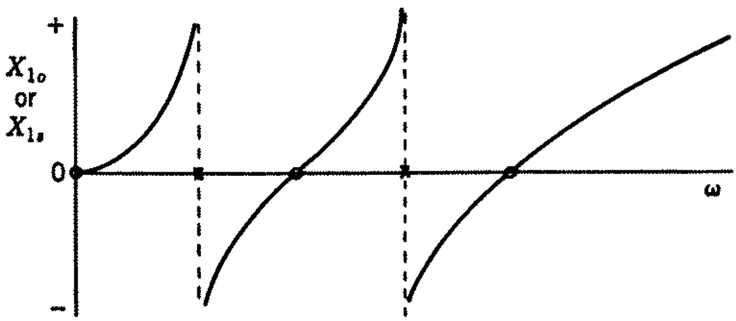


Fig. 13-10. Reactance plot for LC network.

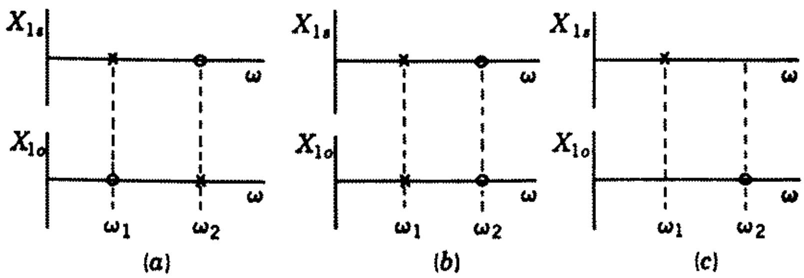


Fig. 13-11. Comparison of critical frequencies of X_{1o} and X_{1s}.

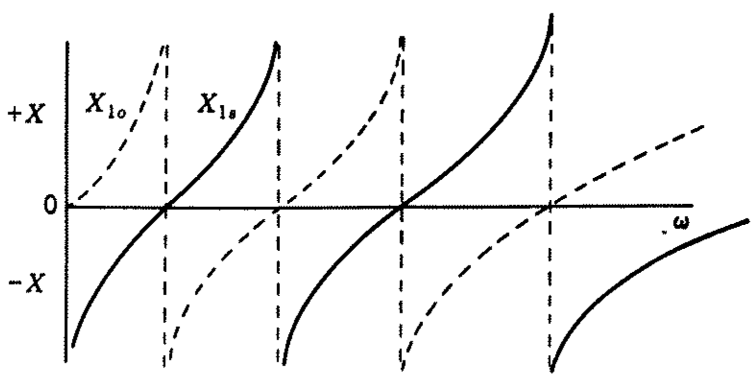


Fig. 13-12. Reactance plot for opposite signs of X_{1s} and X_{1o}.

correspond to Case 3 and Case 4 for which $\alpha = 0$. At all frequencies where X_{1s} and X_{1o} are opposite in sign, there is no attenuation and the frequencies are pass frequencies. A band of frequencies for which this condition holds is thus a *pass band.*

The second possibility shown in Fig. 13-11(b) results in reactance plots having the same sign for all values of frequency. Such a reactance plot is shown in Fig. 13-13. With the same sign for X_{1s} and X_{1o}, we have Case 1 or Case 2 on the table of page 319, corresponding to attenuation. A band of frequencies with the same sign for X_{1s} and X_{1o} is therefore a *stop band.*

The last possibility is illustrated in Fig. 13-14: a critical frequency exists in either X_{1s} or X_{1o} without being in the other one. At the crit-

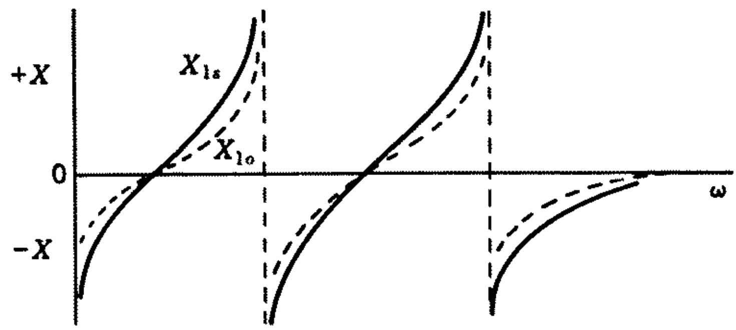


Fig. 13-13. Reactance plot for same sign of X_{1s} and X_{1o}.

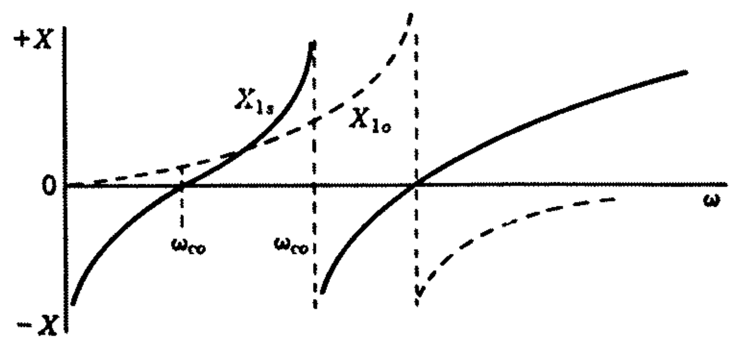


Fig. 13-14. Condition for cutoff frequency from X_{1s} and X_{1o}.

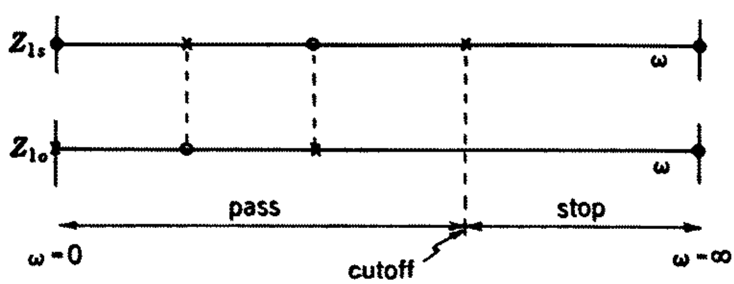


Fig. 13-15. Stop and pass bands and cutoff frequency.

ical frequency, the sign of one reactance function changes but the other does not. This condition corresponds to changing from a pass band to a stop band or vice versa. Such critical frequencies are thus *cutoff frequencies.* An example of a plot with each of the three conditions existing and with the corresponding designation of pass band, stop band, and cutoff frequency is shown in Fig. 13-15.

We are now in a position to study further the nature of the image impedance Z_i in terms of the critical frequencies of the open- and short-circuit impedance functions, Z_{1o} and Z_{1s}. The image impedance is given by Eq. 13-6 as

$$Z_i = \sqrt{Z_{1o}Z_{1s}} \tag{13-50}$$

In terms of the poles and zeros of Z_{1o} and Z_{1s}, the image impedance may be written

$$Z_i = \sqrt{H_1 s \frac{(s^2 + \omega_2^2)(s^2 + \omega_4^2)\ldots}{(s^2 + \omega_1^2)(s^2 + \omega_3^2)\ldots} \frac{H_2}{s} \frac{(s^2 + \omega_a^2)\ldots}{(s^2 + \omega_b^2)\ldots}} \tag{13-51}$$

where two of the possible impedance forms have been assumed for Z_{1o} and Z_{1s}. In the *pass band*, the poles of Z_{1o} are zeros of Z_{1s} or vice versa. Such factors, for example $(s^2 + \omega_2^2)$ and $(s^2 + \omega_b^2)$ if $\omega_2 = \omega_b$, *cancel* term by term and hence are not critical frequencies of Z_i. In the *stop band*, the poles and zeros of Z_{1o} and Z_{1s} *coincide* and so may be removed from the radical. Stop-band critical frequencies are thus critical frequencies of Z_i. A *cutoff frequency* appears as a critical frequency in either Z_{1o} or Z_{1s} (never both). In a typical frequency term (in the sinusoidal steady state where $s = j\omega$),

$$(-\omega^2 + \omega_1^2) \tag{13-52}$$

the sign of the term changes as ω exceeds ω_1, causing a change of sign within the radical. This change of sign changes Z_i from a real number to an imaginary number or vice versa. One plot of Z_i for a filter is

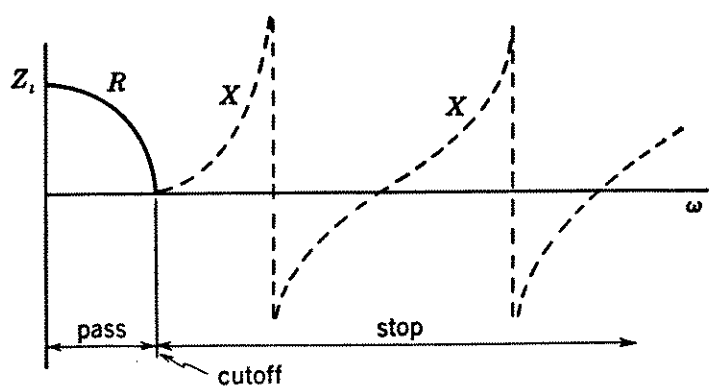


Fig. 13-16. Image impedance variation in the pass band and stop band.

shown in Fig. 13-16. Note that Z_i changes from a resistance to a reactance at the cutoff frequency, and again, that the critical frequencies of Z_{1s} and Z_{1o} in the stop band are the critical frequencies of Z_i in the stop band.

Example 1

To illustrate the application of the concepts of this section to specific network configurations, consider first the T section shown in Fig. 13-17.

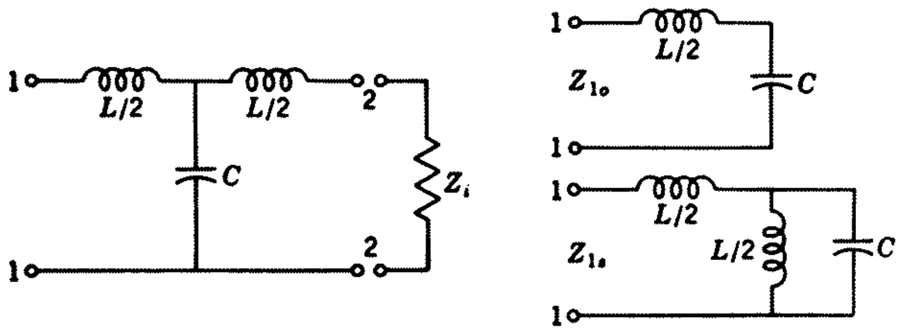


Fig. 13-17. T network of Example 1 with open-circuit and short-circuit networks.

We will determine the poles and zeros of the impedance functions Z_{1s} and Z_{1o}, and from this determine the pass band, the stop band, and the cutoff frequency. The open-circuit impedance is

$$Z_{1o}(s) = \frac{Ls}{2} + \frac{1}{Cs} = \frac{LCs^2 + 2}{2Cs} \tag{13-53}$$

and the short-circuit impedance function is

$$Z_{1s}(s) = \frac{Ls}{2} + \frac{1}{Cs + 2/Ls} = \frac{L(LCs^3 + 4s)}{2(LCs^2 + 2)} \tag{13-54}$$

From these two impedance functions, the poles and zeros are found to have the values tabulated below.

	Z_{1o}	Z_{1s}
Poles	zero infinity	infinity $\omega = \sqrt{2/LC}$
Zeros	$\omega = \sqrt{2/LC}$	zero $\omega = \dfrac{2}{\sqrt{LC}}$

These are shown in Fig. 13-18, together with the designations of the frequencies in the pass band and in the stop band. The value of the cutoff frequency for this network is seen to be

$$\omega_0 = \frac{2}{\sqrt{LC}} \tag{13-55}$$

which is a zero of Z_{1s}. As a filter, this network passes the low frequencies and rejects the high frequencies. Filters of this type are given the name *low-pass filters.*

Example 2

Another T section is shown in Fig. 13-19, together with the corresponding open-circuit and short-circuit networks. For this network we

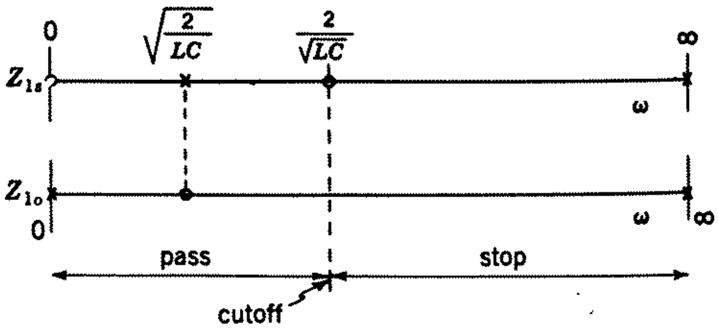


Fig. 13-18. Low-pass filter poles and zeros.

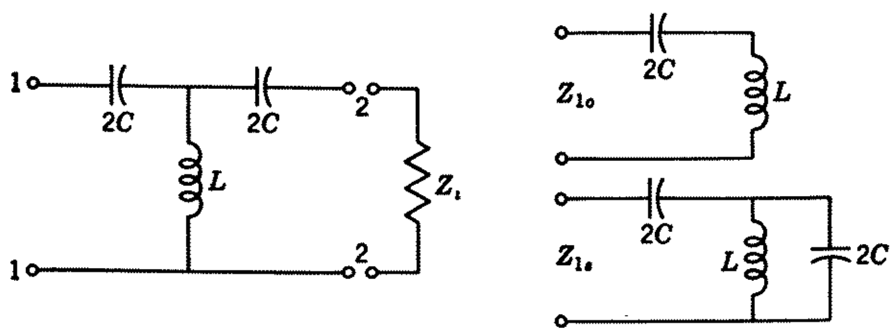


Fig. 13-19. T network of Example 2 with the open-circuit and short-circuit networks.

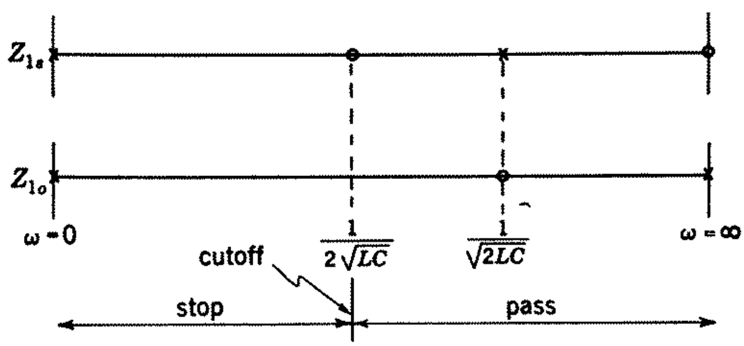


Fig. 13-20. High-pass filter poles and zeros.

see that

$$Z_{1o} = \frac{1}{2Cs} + Ls = \frac{2LCs^2 + 1}{2Cs} \tag{13-56}$$

$$Z_{1s} = \frac{1}{2Cs} + \frac{1}{2Cs + 1/Ls} = \frac{4LCs^2 + 1}{2Cs(2LCs^2 + 1)} \tag{13-57}$$

The poles and zeros for these two functions are shown in Fig. 13-20, together with the pass band, stop band, and cutoff frequency designations. As a filter, this network rejects low frequencies but passes high frequencies. This type of filter is designated a *high-pass filter*.

Example 3

The network for this example has element values given in order to simplify the algebra in computing impedances. The network is a T

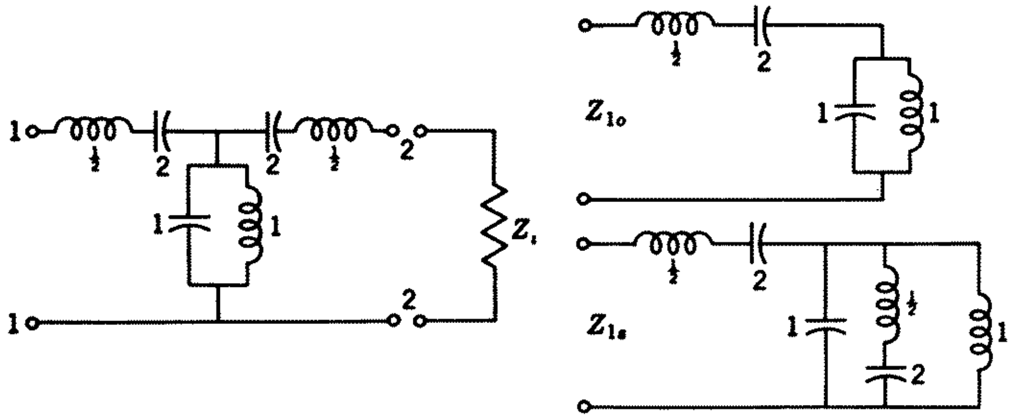


Fig. 13-21. T network of Example 3 with the open-circuit and short-circuit networks.

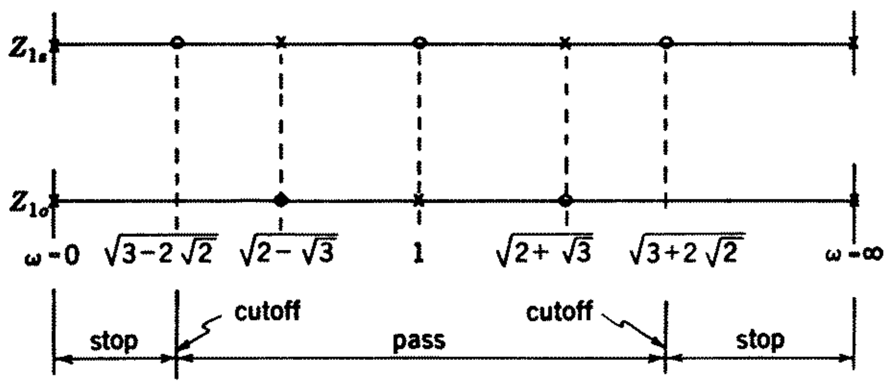


Fig. 13-22. Band-pass filter poles and zeros.

with two elements in each branch. From the schematic diagrams the impedance expressions are found to be

$$Z_{1o} = \frac{s}{2} + \frac{1}{2s} + \frac{1}{s + 1/s} = \frac{s^4 + 4s^2 + 1}{2s(s^2 + 1)} \tag{13-58}$$

and similarly,

$$Z_{1s} = \frac{s}{2} + \frac{1}{2s} + \frac{1}{s + 1/s + 1/(s/2 + 1/2s)} = \frac{s^6 + 7s^4 + 7s^2 + 1}{2s(s^4 + 4s^2 + 1)} \tag{13-59}$$

This last equation looks rather formidable, being of sixth order (or rather third order in s^2). However, the fact that an $s^2 + 1$ term appears in the equation for Z_{1o} as a pole leads one to suspect that $(s^2 + 1)$ might also be a zero of Z_{1s}. Inspection of the pole zero plot of Fig. 13-22 with all poles and zeros plotted except the zeros of Z_{1s} adds to the suspicion. Factoring, there results

$$s^6 + 7s^4 + 7s^2 + 1 = (s^2 + 1)(s^4 + 6s^2 + 1) \tag{13-60}$$

from which all zeros can be found using the quadratic formula. A summary of values for the poles and zeros follows.

	Z_{1o}	Z_{1s}
Poles	zero $\omega = 1$ infinity	zero $\omega = \sqrt{2 \pm \sqrt{3}}$ infinity
Zeros	$\omega = \sqrt{2 \pm \sqrt{3}}$	$\omega = 1$ $\omega = \sqrt{3 \pm 2\sqrt{2}}$

A plot of these poles and zeros is shown in Fig. 13-22. It is seen that there are two cutoff frequencies and that both low-frequency and high-frequency bands are stop bands. Frequencies in a center band are pass-band frequencies. Filters of this type are designated *band-pass filters*. An opposite type of filter with pass bands at low and high frequencies and a stop band at a center band of frequencies is a *band-elimination filter*.

Short-circuit and open-circuit measurements can be used as a practical means of analyzing an unknown two-terminal-pair network in the laboratory. Suppose that the two-terminal-pair network is connected

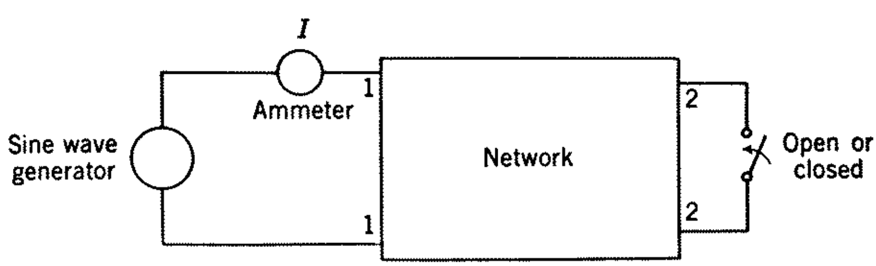


Fig. 13-23. Experimental apparatus to study filters.

to a sine wave generator as shown in Fig. 13-23. If the current is measured by an ammeter marked I (or a dropping resistor together with a cathode ray oscillograph) as frequency is changing with output voltage maintained constant, a plot shown in Fig. 13-24 will be obtained. The current will have a maximum value at a zero of impedance and a minimum value at a pole of impedance. The incidental resistance of the elements prevents the current from becoming zero or infinity. The same rules that have been used in our network analysis can be applied to the experimental results. If I_{sc} is at a minimum value when I_{oc} is at a maximum and vice versa, measurements are being

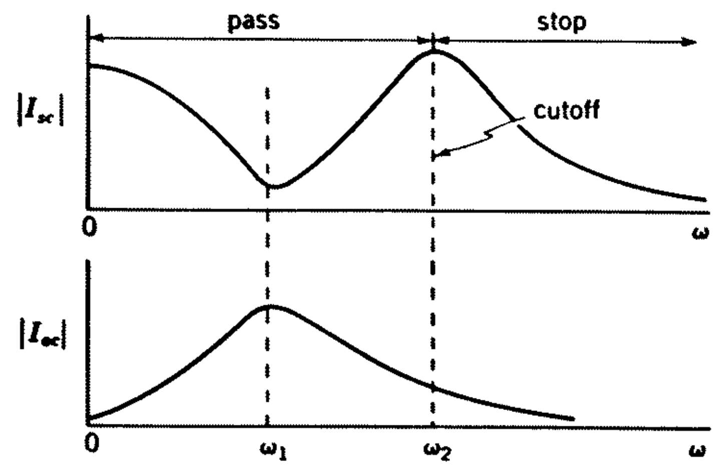


Fig. 13-24. Results for low-pass filter.

made in the pass band. When a maximum or minimum is recorded for I_{sc} or I_{oc} but not for the other, that frequency is a cutoff frequency. When I_{sc} and I_{oc} are both maximum or minimum at a given frequency, that frequency is in the stop band. The case illustrated in Fig. 13-24 evidently corresponds to a low-pass filter.

We now have practice in locating pass bands, stop bands, and cutoff frequencies. We next turn our attention to the problem of computing the attenuation α in the stop band, and the phase shift β in the pass band, for a number of important filter networks. We will also be concerned with the variation of the image impedance with frequency for these networks.

13-5. Attenuation and phase shift in symmetrical T and π networks

Expressions for the image impedance of symmetrical T and π networks were derived as Eqs. 13-10 and 13-12. These equations may be rearranged to show the significance of the factor $(Z_1/4Z_2)$ as

$$Z_{iT} = \sqrt{Z_1^2/4 + Z_1Z_2} = \sqrt{Z_1Z_2(1 + Z_1/4Z_2)} \tag{13-61}$$

$$Z_{i\pi} = \frac{Z_1Z_2}{\sqrt{Z_1^2/4 + Z_1Z_2}} = \sqrt{\frac{Z_1Z_2}{1 + Z_1/4Z_2}} \tag{13-62}$$

We have shown that in the pass band the image impedance is real and that in the stop band the image impedance is imaginary. These equations permit determination of pass bands and stop bands, using this image impedance criterion. When the factor

$$1 + Z_1/4Z_2 \tag{13-63}$$

changes sign, the nature of the image impedance must change, real to imaginary or imaginary to real. Hence cutoff frequencies occur when

$$\frac{Z_1}{4Z_2} = -1 \tag{13-64}$$

This equation offers an alternate method for analyzing a network directly in terms of the series and shunt impedances rather than in terms of open- and short-circuit impedances used in the last section. The factor $(Z_1/4Z_2)$ is evidently of importance in the analysis of filters, since once cutoff frequencies are known, the nature of the network is readily established by knowing whether the network passes or stops at one additional frequency.

An equation may be derived relating the image transfer function to the factor $(Z_1/4Z_2)$. The derivation begins with Eq. 13-30, which is

$$\cosh \gamma = 1 + \frac{Z_1}{2Z_2} \tag{13-65}$$

This equation was derived for a T section, but also applies to a π section (see Prob. 13-2). Dividing both sides of the equation by 2 and rearranging, we have

$$\frac{\cosh \gamma - 1}{2} = \frac{Z_1}{4Z_2} \tag{13-66}$$

By an identity for hyperbolic functions,

$$\frac{\cosh \gamma - 1}{2} = \sinh^2 \frac{\gamma}{2} = \frac{Z_1}{4Z_2} \tag{13-67}$$

Expanding this equation in terms of the real and imaginary part of γ gives

$$\sinh \frac{\alpha + j\beta}{2} = \sinh \frac{\alpha}{2} \cos \frac{\beta}{2} + j \cosh \frac{\alpha}{2} \sin \frac{\beta}{2} = \sqrt{\frac{Z_1}{4Z_2}} \tag{13-68}$$

Now for reactive networks, $Z_1 = \pm jX_1$ and $Z_2 = \pm jX_2$, so that there are two possibilities in the radical expression depending on whether Z_1 and Z_2 have the same or opposite signs. We will consider these possibilities separately.

(1) When Z_1 and Z_2 have opposite signs, that is, $Z_1 = +jX_1$ and $Z_2 = -jX_2$ or $Z_1 = -jX_1$ and $Z_2 = +jX_2$, then $(Z_1/4Z_2)$ is *negative*.

The last equation may be interpreted in terms of a positive factor $(-Z_1/4Z_2)$. Then

$$\sinh\frac{\alpha}{2}\cos\frac{\beta}{2} = 0 \tag{13-69}$$

$$\cosh\frac{\alpha}{2}\sin\frac{\beta}{2} = \sqrt{\frac{-Z_1}{4Z_2}} \tag{13-70}$$

This equation may be satisfied in two ways: $\alpha = 0$ or $\beta = \pm\pi, \pm 3\pi$, etc. Thus either

$$\alpha = 0 \tag{13-71}$$

and

$$\beta = 2\sin^{-1}\sqrt{-Z_1/4Z_2} \tag{13-72}$$

or

$$\beta = \pm\pi,\ \pm 3\pi,\ \pm 5\pi,\ \ldots \tag{13-73}$$

and

$$\alpha = 2\cosh^{-1}\sqrt{-Z_1/4Z_2} \tag{13-74}$$

(2) When $(Z_1/4Z_2)$ is *positive*, the radical of Eq. 13-68 is real, and

$$\cosh\frac{\alpha}{2}\sin\frac{\beta}{2} = 0 \tag{13-75}$$

$$\sinh\frac{\alpha}{2}\cos\frac{\beta}{2} = \sqrt{\frac{Z_1}{4Z_2}} \tag{13-76}$$

In this case $\cosh(\alpha/2)$ can never equal zero, so that there is only one way in which these equations can be satisfied: β must be $0, \pm 2\pi$, etc. Thus the solution is

$$\beta = 0,\ \pm 2\pi,\ \pm 4\pi,\ \ldots \tag{13-77}$$

$$\alpha = 2\sinh^{-1}\sqrt{Z_1/4Z_2} \tag{13-78}$$

These two equations apply when $(Z_1/4Z_2)$ is positive. However, when $(Z_1/4Z_2)$ is negative there are the two possibilities corresponding to Eqs. 13-71 and 13-72 for a pass band, and to Eqs. 13-73 and 13-74 for a stop band.

Since we arrive at different conclusions for positive and for negative values of $(Z_1/4Z_2)$, zero value for $(Z_1/4Z_2)$ is evidently a point of division for the various forms of equations for α and β. When $(Z_1/4Z_2)$ is negative there are two possibilities. The point of division for these two equations is given by Eq. 13-64 as $Z_1/4Z_2 = -1$. The different possibilities are summarized in the following table. The equations for negative values of $(Z_1/4Z_2)$ find the most frequent application, because most of our studies will concern networks with opposite signs for X_1 and X_2.

$Z_1/4Z_2$ Lower limit	$Z_1/4Z_2$ Upper limit	Type of band	Attenuation α	Phase shift β
$-\infty$	-1	stop	$2\cosh^{-1}\sqrt{-Z_1/4Z_2}$	$\pm\pi, \pm 3\pi, \ldots$
-1	0	pass	0	$2\sin^{-1}\sqrt{-Z_1/4Z_2}$
0	∞	stop	$2\sinh^{-1}\sqrt{Z_1/4Z_2}$	$0, \pm 2\pi, \pm 4\pi, \ldots$

13-6. Constant-K filters

An important class of filters is designed under the condition that Z_1 and Z_2 (as defined for the standard T and standard π section) are related by the equation below, where R is a constant both positive and real.

$$Z_1 Z_2 = R^2 \tag{13-79}$$

This equality requires that the impedances Z_1 and Z_2 be purely reactive and of opposite sign. In the first discussion of filters of this type, Zobel* used the letter K in place of the R of our equation. Actually R is a preferred symbol because the quantity is dimensionally ohms, and R turns out to be the value of the terminating resistance. Even though R has replaced K in the defining equation, filters designed on the assumption of Eq. 13-79 are universally designated as *constant-K filters*. The advantages, if not the justification of the assumed relationship between Z_1 and Z_2, will become evident by algebraic simplification, and later by simple network structures.

To simplify the equations derived in the last section, we will define a new variable for the quantity $(Z_1/4Z_2)$ as

$$x^2 = \frac{-Z_1}{4Z_2} \tag{13-80}$$

This particular choice of sign is made in order to make x^2 a positive quantity, since Z_1 and Z_2 have opposite signs for constant-K filters. The impedance expressions in Eq. 13-79 are, for reactive networks,

$$Z_1 Z_2 = (\pm jX_1)(\mp jX_2) = +X_1 X_2 = R^2 \tag{13-81}$$

It is now possible to write the expressions for the image impedance in very simple form in terms of R and x. Equations 13-61 and 13-62

* See reference at end of chapter.

become

$$Z_{iT} = R\sqrt{1 - x^2} \tag{13-82}$$

$$Z_{i\pi} = \frac{R}{\sqrt{1 - x^2}} \tag{13-83}$$

Similarly, the expressions for the attenuation and phase of the last section become simple in form. Since we are considering only positive values for x^2, only negative values for $(Z_1/4Z_2)$ in the table on page 331 need be considered for the time being. In the *stop band*, by Eq. 13-74, we have

$$\alpha = 2 \cosh^{-1} x \tag{13-84}$$

$$\beta = \pm\pi, \ \pm 3\pi, \text{ etc.} \tag{13-85}$$

In the *pass band*, by Eq. 13-72,

$$\alpha = 0 \tag{13-86}$$

$$\beta = 2 \sin^{-1} x \tag{13-87}$$

Plots of Z_{iT}, $Z_{i\pi}$, α, and β against x are shown in Fig. 13-25. These are generalized plots. In order for the plots to be specialized to specific

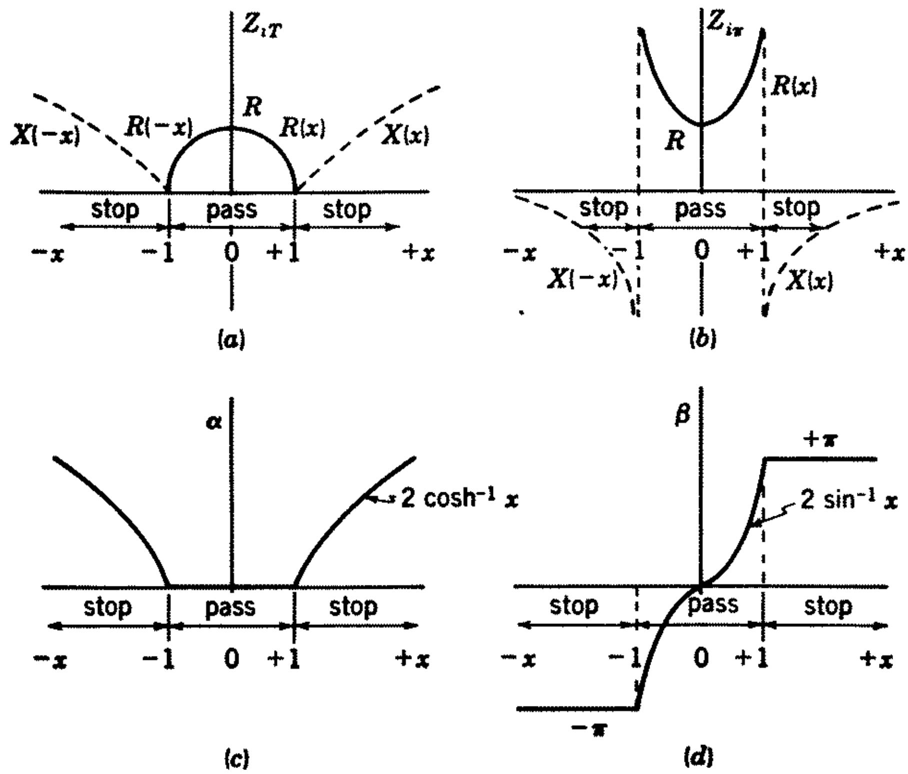


Fig. 13-25. Normalized plots in terms of $x = \sqrt{-Z_1/4Z_2}$.

networks, only x as a function of frequency ω need be determined. Once $x(\omega)$ is known, the coordinates may be adjusted for the special cases. This procedure will be illustrated for several examples considered earlier in the chapter.

Example 4

Consider the standard T section shown in Fig. 13-17, (page 324). For this network, $Z_1 = j\omega L$ and $Z_2 = -j/\omega C$ (note that Z_1 and Z_2 are opposite in sign as required). The normalized variable x then becomes

$$x = \sqrt{\frac{-Z_1}{4Z_2}} = \sqrt{\frac{\omega L}{4/\omega C}} = \sqrt{\frac{\omega^2}{4/LC}} \tag{13-88}$$

Now by Eq. 13-55, the cutoff frequency for this T section is

$$\omega_0 = \frac{2}{\sqrt{LC}} \tag{13-89}$$

so that x becomes

$$x = \frac{\omega}{\omega_0} \tag{13-90}$$

Then for this T section, we have the following information:

$$\alpha = 0 \quad \text{and} \quad \beta = 2 \sin^{-1} \frac{\omega}{\omega_0} \qquad 0 \leqq \omega \leqq \omega_0 \tag{13-91}$$

$$\alpha = 2 \cosh^{-1} \frac{\omega}{\omega_0}, \qquad \beta = \pi, \qquad \omega \geqq \omega_0 \tag{13-92}$$

$$\omega_0 = \frac{2}{\sqrt{LC}} \tag{13-93}$$

$$R = \sqrt{L/C} \tag{13-94}$$

The plots for Z_{iT}, α, and β given in Fig. 13-25 apply directly to this network, with x replaced by ω/ω_0. As discussed previously, the attenuation α is usually computed in decibels using the relationship, $\alpha_{db} = 8.686\alpha_{\text{nepers}}$.

Example 5

For this example, consider the T network of Fig. 13-19 (page 325). The variation of x with ω is found as

$$x = \sqrt{\frac{1/\omega C}{4\omega L}} = \sqrt{\frac{1}{4\omega^2 LC}} = \frac{\omega_0}{\omega} \tag{13-95}$$

where the cutoff frequency ω_0 is identified from Fig. 13-20. This T section has an *inverse* relationship to that of Example 1. The equations become

$$\alpha = 0 \quad \text{and} \quad \beta = -2\sin^{-1}\frac{\omega_0}{\omega}, \qquad \omega \geqq \omega_0 \tag{13-96}$$

$$\alpha = 2\cosh^{-1}\frac{\omega_0}{\omega} \quad \text{and} \quad \beta = -\pi, \qquad 0 \leqq \omega \leqq \omega_0 \tag{13-97}$$

$$\omega_0 = \frac{1}{2\sqrt{LC}} \tag{13-98}$$

$$R = \sqrt{L/C} \tag{13-99}$$

These equations evidently identify a high-pass filter. In order to make use of the normalized plots of Fig. 13-25, only the x axis need be inverted, the origin becoming infinity and infinity becoming the origin. The resulting plots for Z_{iT}, α, and β are shown in Fig. 13-26.

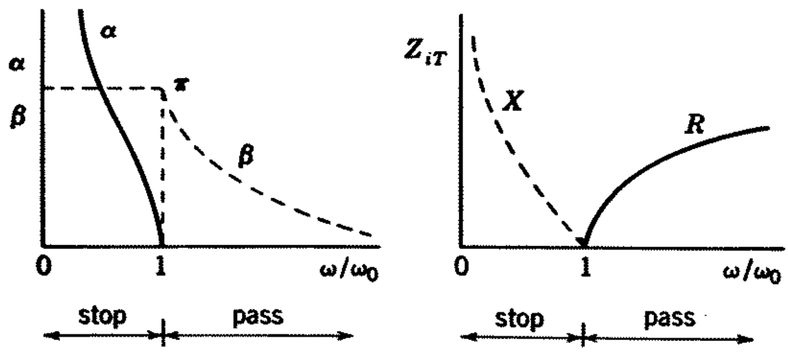


Fig. 13-26. Characteristics of high-pass filter.

Example 6

For the π section of Fig. 13-27, $Z_1 = j\omega L$ and $Z_2 = -j/\omega C$. The variable x for this network becomes

$$x = \sqrt{\frac{\omega L}{4/\omega C}} = \sqrt{\frac{\omega^2}{4/LC}} = \frac{\omega}{\omega_0} \tag{13-100}$$

since the cutoff condition, $Z_1/4Z_2 = -1$ defines ω_0. This equation is identical with Eq. 13-90, indicating that the attenuation and phase shift of this π network are *identical* with those of the T network of Example 1, and are given in Fig. 13-25. There is one important difference, however. The image impedance variation with ω/ω_0 is *different*, being that given in Fig. 13-25(b). For these two different networks, the impedance characteristics are quite different even though α and β are identical.

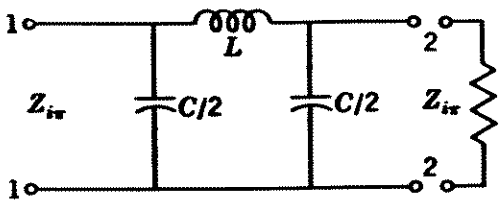


Fig. 13-27. π section network of Example 3.

Example 7

For this example consider the network of Example 3 shown in Fig. 13-21. This network was shown to be a band-pass filter. The reactance functions for this network are

$$Z_1 = j\,\frac{\omega^2 - 1}{\omega} \quad \text{and} \quad Z_2 = -j\,\frac{\omega}{\omega^2 - 1} \tag{13-101}$$

so that

$$x = \sqrt{\frac{-Z_1}{4Z_2}} = \frac{\omega^2 - 1}{2\omega} \tag{13-102}$$

This last equation relates the frequency ω for the filter to the variable x of the standard attenuation, phase shift, and image impedance characteristics. By this equation, we perform a frequency transformation. The same technique can be used for any filter—band-pass, band-elimination, or any combination of such specifications. Plots of α and β for the band-pass filter of this example as shown in Fig. 13-28.

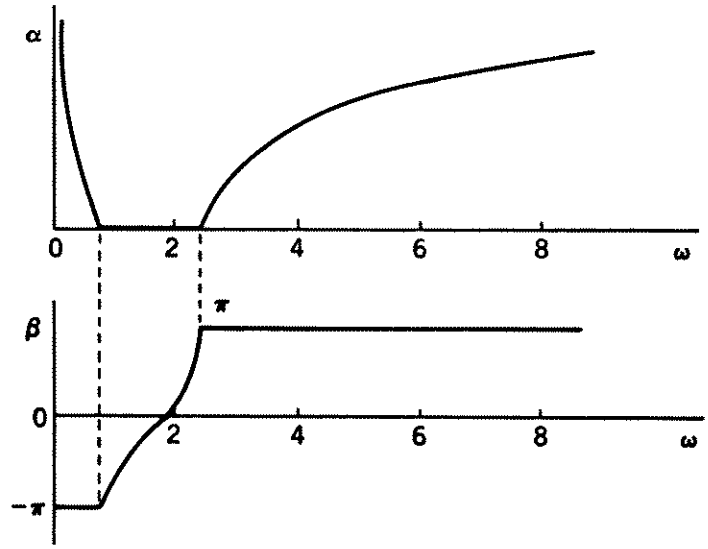


Fig. 13-28. Attenuation and phase characteristics for a band-pass filter.

The networks of our four examples have been very simple, but the same concepts apply to more complicated networks necessary to accomplish multiple-pass or multiple-stop bands. All networks of the constant-K type must have Z_1 and Z_2 obeying the inverse relationship

$$Z_1 = \frac{R^2}{Z_2} \tag{13-103}$$

and this restriction limits the possible forms for Z_1 and Z_2. We are familiar with two forms of networks that obey this relationship, the *Foster* forms of networks. The relationship

$$\frac{Z_1}{Y_2} = R^2 \tag{13-104}$$

is satisfied if, for example,

$$Z_1 = R^2 \frac{(s^2 + s_1^2)\ldots}{s(s^2 + s_2^2)\ldots} \tag{13-105}$$

and

$$Y_2 = \frac{(s^2 + s_1^2)\ldots}{s(s^2 + s_2^2)\ldots} \tag{13-106}$$

(Other multiplying factors will also satisfy Eq. 104.) Hence if Z_1 is realized as a Foster No. 1 type of network and Y_2 is realized as a Foster

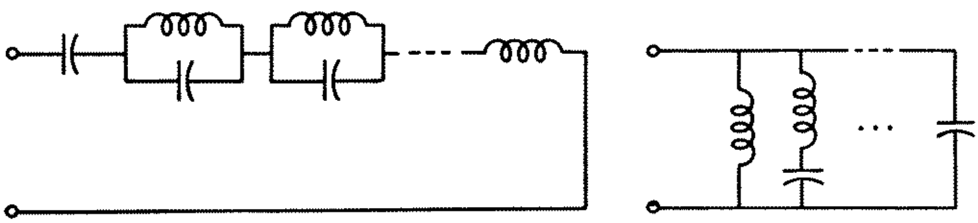


Fig. 13-29. Foster network forms.

No. 2 admittance network, the resulting filter will be constant-K. A typical term in the admittance expansion of Y_2 is given by Eq. 12-64; it is

$$Y(s) = \frac{s/L_n}{(s^2 + 1/L_nC_n)} \tag{13-107}$$

for a series LC network. Similarly, typical terms of the impedance expansion of $Z(s)$ will have a form given by Eq. 12-60, as

$$Z(s) = \frac{s/C_m}{(s^2 + 1/L_mC_m)} \tag{13-108}$$

Since by Eq. 13-104 these equations for $Y(s)$ and $Z(s)$ containing the above typical terms must be equal, we have

$$Z_1(s) = R^2Y_2(s) \tag{13-109}$$

or, in terms of the product forms for $Y(s)$ and $Z(s)$,

$$\frac{N_1(s)}{(s^2 + 1/L_mC_m)(s^2 + 1/L_{m+2}C_{m+2})\ldots} = R^2 \frac{N_2(s)}{(s^2 + 1/L_nC_n)\ldots} \tag{13-110}$$

This equality is possible only if, term by term,

$$L_mC_m = L_nC_n = \frac{1}{\omega_n^2} \tag{13-111}$$

or, for similar network configurations, in the two Foster forms,

$$\frac{L_m}{L_n} = \frac{C_n}{C_m} \tag{13-112}$$

for the constant-K network. As an illustration of these conclusions, consider the network shown in Fig. 13-30. In order for this network to be constant-K, it is necessary that

$$L_1C_1 = L_2C_2 = \frac{1}{\omega_0^2} \tag{13-113}$$

This concept is useful in the design of constant-K filters.

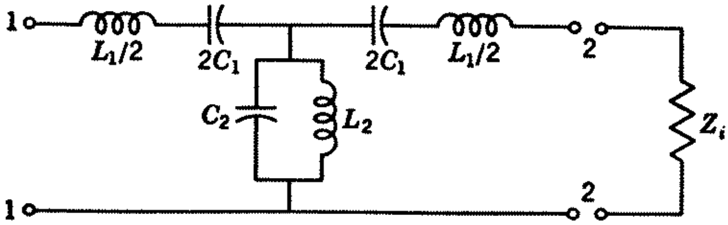


Fig. 13-30. Constant-K filter when $L_1C_1 = L_2C_2$.

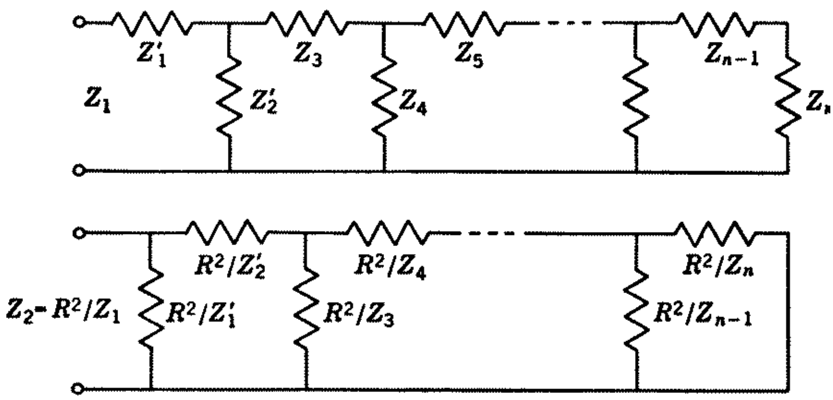


Fig. 13-31. Ladder networks satisfying the requirement of reciprocal impedances, $Z_1Z_2 = R^2$. Note that Z_1 is the driving-point impedance and that Z_1' is the impedance of a ladder element.

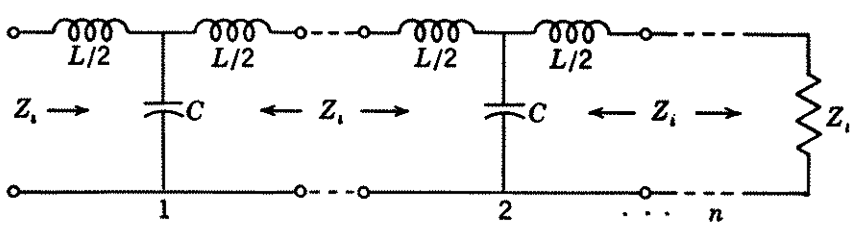


Fig. 13-32. Composite filter of T sections.

The networks used for Z_1 and Z_2 need not be of the Foster forms. The two ladder structures shown in Fig. 13-31 satisfy the requirement that $Z_1Z_2 = R^2$.

We will define a *composite filter* as a filter made up of the cascade connection of a number of standard T or standard π sections. Constant-K networks can be connected in tandem to form a composite filter provided an image impedance match exists at each terminal pair. To illustrate, consider a standard T section. Figure 13-32 shows a representation of n such sections connected in tandem. Note that at each terminal pair in the composite structure there exists an *image match* (such that the equations we have derived hold). The attenuation of n sections is

$$\alpha = 2n \cosh^{-1} \frac{\omega}{\omega_0} \tag{13-114}$$

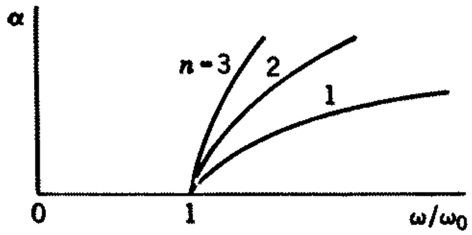


Fig. 13-33. Attenuation in composite filters.

from Eq. 13-92. Attenuation curves for several values of n are shown in Fig. 13-33. Evidently the effectiveness of the composite filter in providing attenuation in the stop band is increased by increasing the number of T sections in tandem. However, if a "sharp" cutoff is required by specifications, a large number of T sections must be used in the composite filter. The limitations of constant-K composite filters are:

(1) A large number of T sections is required to attain high attenuation in the stop band near the cutoff frequency. This large number of elements may make the cost of the composite filter prohibitive.

(2) The composite network cannot be terminated in the required image impedance shown in Fig. 13-25(a), because no such terminating impedance exists. Terminating the filter with a constant resistance R introduces mismatch at all but one frequency.

13-7. The m-derived filter

The need for a filter section with high attenuation in the stop band near cutoff frequency led to development of the m-derived filter by O. J. Zobel in 1923. The filters considered in the last section were a

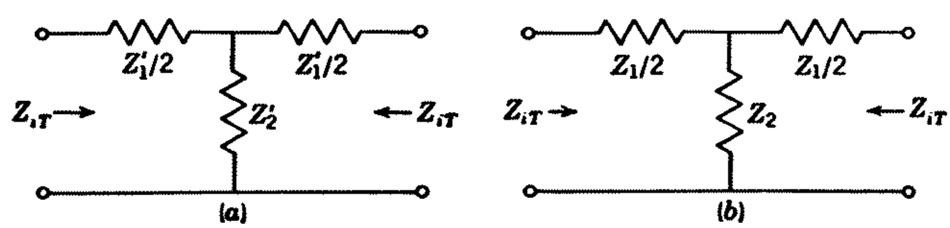


Fig. 13-34. Networks with the same image impedances: (a) new; (b) old.

very restricted class of filters satisfying the requirement that the product Z_1Z_2 be a constant. If other combinations of elements are permitted, it seems intuitively possible that some arrangement of elements will give the required high attenuation near cutoff frequency. Zobel approached this problem with one specification for the new network.

Anticipating that the resulting section might be used in tandem with constant-K filter sections, he specified that the image impedance of the new network be the same as the standard T or standard π section. This requirement is illustrated in Fig. 13-34. The new impedance terms are designated Z_1' and Z_2'; the impedance terms for the constant-K sections are designated Z_1 and Z_2. We have the requirement that the two networks have the same image impedances; we need some further premise in order to relate Z_1, Z_2, Z_1', and Z_2'. Zobel assumed that Z_1' and Z_1 were related by the equation

$$Z_1' = mZ_1 \tag{13-115}$$

where m is a constant. This may seem to be an unusual assumption to make. We should expect that our assumption might instead relate to the attenuation properties of the new filters, or perhaps the desired form of image impedance. It is difficult to anticipate the surprising results that follow from this simple beginning.

With Z_1' fixed, let us see what happens to Z_2' in terms of Z_1 and Z_2. Equating image impedances,

$$\frac{Z_1'^2}{4} + Z_1'Z_2' = \frac{Z_1^2}{4} + Z_1Z_2 \tag{13-116}$$

Substituting the condition $Z_1' = mZ_1$ into this equation and solving for Z_2', there results

$$Z_2' = \left(\frac{1 - m^2}{4m}\right) Z_1 + \frac{1}{m} Z_2 \tag{13-117}$$

The schematic representation of the new m-derived network is shown in Fig. 13-35.

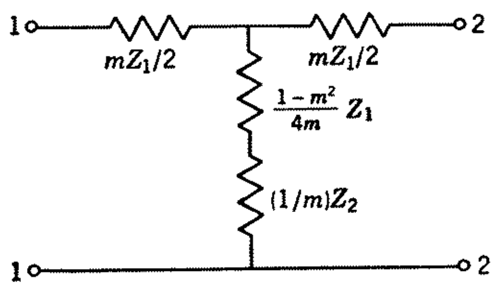


Fig. 13-35. m-derived T section.

A similar derivation may be given for the π section by assuming that any new network structure must have the same image impedance as the standard π network and further making the assumption that

$$Z_2' = \frac{Z_2}{m} \tag{13-118}$$

The image impedance of the π section is given by Eq. 13-12. Using this equation and 13-118, the impedance Z_1' is found to have the value

$$Z_1' = \cfrac{1}{\cfrac{1}{mZ_1} + \cfrac{1}{\cfrac{4m}{1 - m^2} Z_2}} \tag{13-119}$$

which is the impedance of a parallel combination of the impedances

$$mZ_1 \quad \text{and} \quad \frac{4m}{1 - m^2} Z_2 \tag{13-120}$$

The resulting m-derived π section is shown in Fig. 13-36, together with the schematic for the standard π section.

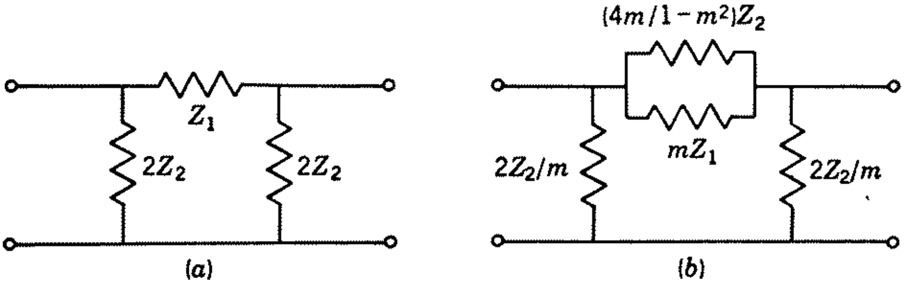


Fig. 13-36. Comparison of (a) standard section, and (b) m-derived section.

Now that we have the m-derived structures, our next task is to compute the attenuation to see that we have attained our objective. First, a physical interpretation of the results thus far can be seen in a specific example. Suppose that we select a low-pass filter of the type shown in Fig. 13-17 and in Fig. 13-27 for the T and π sections, respec-

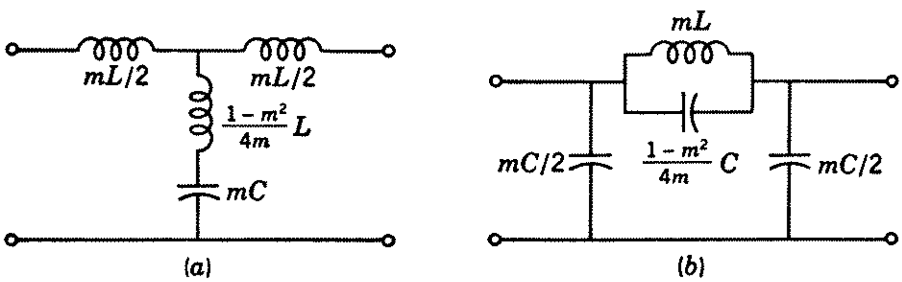


Fig. 13-37. m-derived low-pass filter sections: (a) T section; (b) π section.

tively. The equivalent m-derived sections are shown in Fig. 13-37. For these networks, let us ask, what can possibly give infinite attenuation at a particular frequency? The m-derived T section shown in Fig. 13-37 will have infinite attenuation (or no transmission) when the series LC circuit is in resonance. Under this resonance condition, the series LC circuit is the equivalent of a short circuit, such that all current by-passes the load. This resonant frequency is thus a frequency of infinite attenuation designated as ω_∞. It has the value

$$\omega_\infty = \frac{1}{\sqrt{[(1 - m^2)/4m]L(mC)}} \tag{13-121}$$

or

$$\omega_\infty = \frac{\omega_0}{\sqrt{1 - m^2}} \tag{13-122}$$

since the cutoff frequency ω_0 has the value $2/\sqrt{LC}$ by Eq. 13-55.

At this same frequency, the parallel circuit of Fig. 13-37(b) will be in parallel resonance (antiresonance) having infinite impedance. Under this open-circuit condition, no current can pass through the series elements, and hence there will be no current in the output. This corresponds to infinite attenuation for the π section. We now begin to see that altering Z_1 in the T, and Z_2 in the π has introduced a new element either in series or in parallel in such a way as to prevent transmission at one particular frequency.

The equations for attenuation and phase shift in the m-derived filter sections can be found by computing the factor x defined by Eq. 13-80 in terms of the new reactance functions Z_1' and Z_2'. Thus

$$x'^2 = \frac{-Z_1'}{4Z_2'} = \frac{-mZ_1}{4[Z_1(1 - m^2)/4m + Z_2/m]} \tag{13-123}$$

$$= \frac{-m^2}{1 - m^2 + 4Z_2/Z_1} \tag{13-124}$$

In this expression, we recognize that $-4Z_2/Z_1 = 1/x^2$, where x is the factor used in the study of the constant-K filter sections. When this factor is substituted into Eq. 13-124, there results

$$x'^2 = \frac{m^2}{-(1 - m^2) + 1/x^2} \tag{13-125}$$

From this equation, we see that as x approaches the value given by

$$x^2 = \frac{1}{1 - m^2} \tag{13-126}$$

then x'^2 approaches infinity such that $\cosh^{-1} x'$, the attenuation, also approaches an infinite value. The value of x causing infinite attenuation will be designated x_∞ such that

$$x_\infty^2 = \frac{1}{1 - m^2} \tag{13-127}$$

With this definition, Eq. 13-125 may be written

$$x'^2 = \frac{m^2}{-1/x_\infty^2 + 1/x^2} \tag{13-128}$$

In this equation, x = a normalized frequency relating to constant-K filters, x' = a normalized frequency derived for m-derived filters as a function of x, x_∞ = the value of x causing x' to be infinite, and m = a constant for the m-derived filter section. Inspection of this equation shows that when x is less than x_∞, then x'^2 is negative, while for x larger than x_∞, x'^2 is positive. In order to associate these conditions with pass band or stop band and the equation for attenuation, we will refer to the table on page 331. Comparing this table with the different possibilities for the sign of x^2, we reach the conclusions summarized below.

Sign of x'^2	Limits of x^2	Type of band	Attenuation	Phase shift β	Derived as equations
positive	$0 \leqq x^2 \leqq 1$	pass	$\alpha = 0$	$\beta = 2 \sin^{-1} x'$	13-71, 13-72
positive	$1 \leqq x^2 \leqq x_\infty^2$	stop	$\alpha = 2 \cosh^{-1} x'$	$\pm\pi$	13-73, 13-74
negative	$x_\infty^2 \leqq x^2 \leqq \infty$	stop	$\alpha = 2 \sinh^{-1} x'$	0	13-77, 13-78

$$x' = \frac{m}{\sqrt{(-1/x_\infty^2 + 1/x^2)}}, \qquad x_\infty = \frac{1}{\sqrt{1 - m^2}}$$

A plot of these equations (for positive values of x) is shown in Fig. 13-38, together with the same characteristics for the constant-K filter

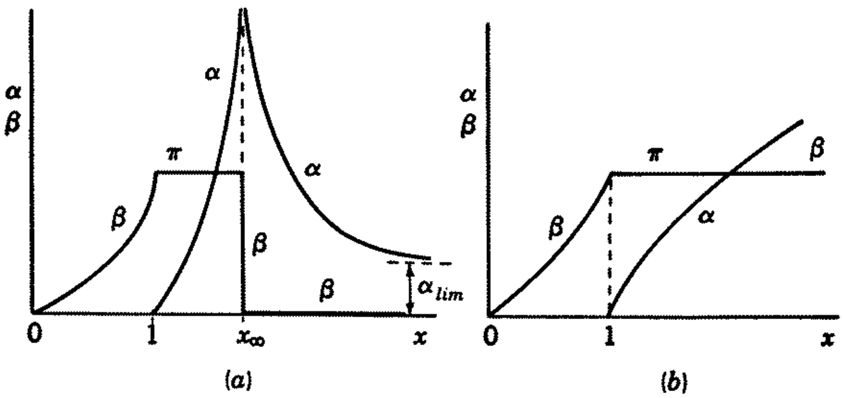


Fig. 13-38. Characteristics of the m-derived filter: (a) m-derived; (b) constant-K.

section. The plot illustrates the high attenuation near the edge of the stop-band feature of the m-derived filter. Comparing the plot with that for the constant-K filter, it is seen that the attenuation for the m-derived filter approaches a minimum value for large values of x, whereas the constant-K filter attenuation approaches a large value for large x. Each type of filter sections has advantages and disadvantages. Since both sections have the same image impedance, there is a possibility of use of a *combination* of both filter types to get *both* the high attenuation near the edge of the stop band and at large values of x.

Figure 13-38 shows that when $x = x_\infty$, the phase shift in the m-derived filter section changes abruptly from π to 0 degrees. In terms of the equations, this is caused by the sign of x'^2 changing from positive to negative. A physical reason for this phenomena can be seen in the low-pass filter illustrated in Fig. 13-37(a) in terms of the series LC circuit. As frequency increases through resonance, the LC circuit reactance changes from negative to positive (or from capacitive to inductive). Since no phase shift is possible in a network with the same kind of reactance in all arms, the "effective" inductive network above resonance gives no phase shift. This resonant frequency, incidently, corresponds to x_∞.

In designing m-derived filters, the question will arise, why not make $x_\infty = 1$ by making $m = 0$ to make the filter have an extremely sharp cutoff in the stop band? More generally, what happens to the attenuation characteristics as m is varied? To answer this question, we will investigate the attenuation of the m-derived filter at large values of x. From the equation for attenuation given in the table on page 342, the attenuation for large x approaches a value α_{lim} given as

$$\alpha_{lim} = 2 \sinh^{-1} \frac{m}{\sqrt{1 - m^2}} \tag{13-129}$$

or, from the identity $\cosh^2 \alpha_{lim} - \sinh^2 \alpha_{lim} = 1$,

$$\cosh^2 \frac{\alpha_{lim}}{2} = 1 + \sinh^2 \frac{\alpha_{lim}}{2} = 1 + \frac{m^2}{1 - m^2} = \frac{1}{1 - m^2} \tag{13-130}$$

Now
$$\frac{\sinh (\alpha_{lim}/2)}{\cosh (\alpha_{lim}/2)} = \tanh \frac{\alpha_{lim}}{2} = \frac{m/\sqrt{1 - m^2}}{1/\sqrt{1 - m^2}} = m \tag{13-131}$$

Hence
$$\alpha_{lim} = 2 \tanh^{-1} m \tag{13-132}$$

From the last equation, it is seen that as m becomes small, approaching zero, the attenuation for large x, α_{lim} also becomes small. Also, from Eq. 13-125, x' is seen to become small as m becomes small (for any value of x), and this in turn reduces the magnitude of the attenuation for all values of x. Thus the price paid for sharp cutoff is reduced attenuation for all frequencies, x being some function of frequency. These conclusions are illustrated in Fig. 13-39 for two values of m.*

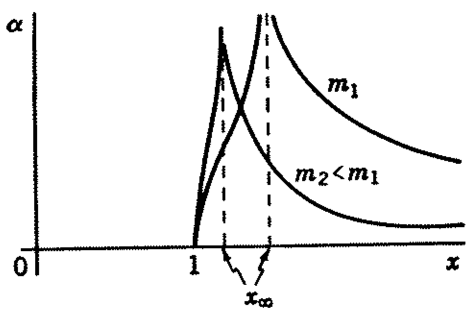


Fig. 13-39. Variation of attenuation characteristics of m-derived filters as x_∞ moves closer to cutoff ($x = 1$).

* Another reason for avoiding small values of m is that finite dissipation in the filter elements results in finite attenuation at x_∞, this finite attenuation being smaller as $m \to 0$.

From this discussion, the value of the constant m of the m-derived filter is seen to determine the nature of the variation of attenuation with x. The constant m also determines element values in the m-derived T and π sections. Since in these structures, element values are determined by a multiplying factor $(1 - m^2)$, it follows that m cannot exceed unity in value; that is,

$$0 < m < 1 \tag{13-133}$$

(Note: This limitation applies only to the ladder structures. Values of m larger than 1 are used in lattice structures to give linear phase characteristics.) There is another significance attached to the value of m in terms of image impedance of the m-derived sections. This will be our next subject for study.

13-8. Image impedance of m-derived half (or L) sections

The m-derived T section is shown in Fig. 13-35, and the m-derived π section in Fig. 13-36. These sections were found under the assumption that the image impedances are the same as the constant-K filter sections. If the T and π sections are divided into *half* (or *L*) *sections*, an unexpected image impedance characteristic is found. This result we must regard as a bonus; certainly it is not a consequence of any requirements made of the m-derived filter. A different image impedance

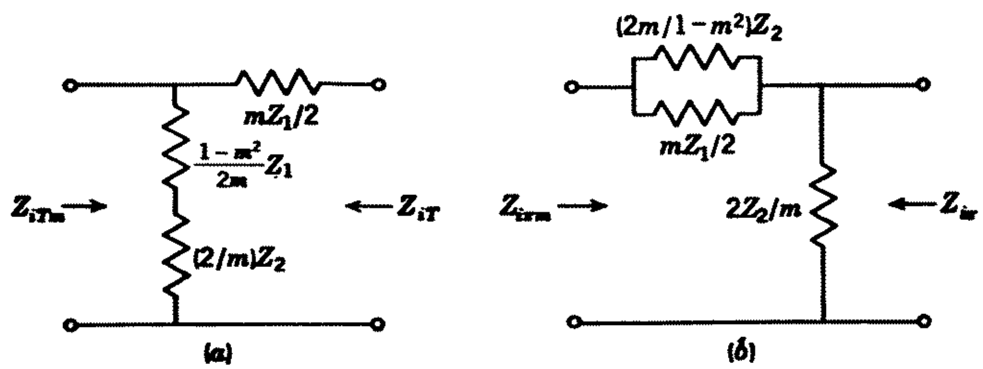


Fig. 13-40. m-derived filter half-sections: (a) m-derived half T; (b) m-derived half π.

behavior at the other terminals of a divided T or π section (the half section) might be expected from the discussion of the image impedance of the L section on page 314. There it was found that the image impedance at one terminal pair was Z_{iT}, and at the other terminal pair was $Z_{i\pi}$. The half-sections for the m-derived filter are shown in Fig. 13-40. The image impedances at the "back door" terminals are designated Z_{iTm} and $Z_{i\pi m}$. They may be determined by using Eq. 13-6.

For the T section, we have

$$Z_{iTm} = \sqrt{Z_{1o}Z_{1s}}$$

$$= \sqrt{\left[\frac{1-m^2}{2m} Z_1 + \frac{2}{m} Z_2\right]\left[\frac{\frac{mZ_1}{2}\left(\frac{1-m^2}{2m} Z_1 + \frac{2}{m} Z_2\right)}{\left(\frac{m}{2} + \frac{1-m^2}{2m} Z_1 + \frac{2}{m} Z_2\right)}\right]} \tag{13-134}$$

$$= \left[1 + (1 - m^2)\frac{Z_1}{4Z_2}\right]\sqrt{\frac{Z_1Z_2}{1 + Z_1/4Z_2}} \tag{13-135}$$

Since $R^2 = Z_1Z_2$, the image impedance becomes

$$Z_{iTm} = R\,\frac{[1 - (1 - m^2)x^2]}{\sqrt{1 - x^2}} \tag{13-136}$$

where $x^2 = -Z_1/4Z_2$ is the factor defined for the constant-K filter.

The same procedure may be used to find $Z_{i\pi m}$ for the filter of Fig. 13-40(b); thus

$$Z_{i\pi m} = \sqrt{Z_{1o}Z_{1s}}$$

$$= \sqrt{\left[\frac{1}{2/mZ_1 + (1 - m^2)/2mZ_2} + \frac{2}{m} Z_2\right]\left[\frac{1}{2/mZ_1 + (1 - m^2)/2mZ_2}\right]} \tag{13-137}$$

$$= \frac{\sqrt{Z_1Z_2 + Z_1^2/4}}{[1 + (1 - m^2)Z_1/4Z_2]} \tag{13-138}$$

so that

$$Z_{i\pi m} = R\,\frac{\sqrt{1 - x^2}}{[1 - (1 - m^2)x^2]} \tag{13-139}$$

In Fig. 13-41, Z_{iTm} and $Z_{i\pi m}$ are shown plotted as a function of x for several values of m. The plot for $m = 0.6$ gives an image impedance constant within 4% over 90% of the pass band. Other values of m give variations greater than this. This image impedance variation is much more constant than the constant-K image impedance functions We note that for $m = 0$, $Z_{i\pi m} = Z_{i\pi}$ by Eq. 13-83, and that for $m = 1$, $Z_{i\pi m} = Z_{iT}$ by Eq. 13-82 and vice versa for Z_{iTm}. In other words, the m-derived filter sections reduce to constant-K filter sections with $m = 1$. The new image impedance function $Z_{i\pi m}$ or Z_{iTm} for m $= 0.6$ more nearly approximates a constant, and so an image impedance match with a constant terminating resistor is a reasonable approximation. This is a very important advantage for the m-derived half section. With these remarkable properties of the m-derived filter established, we next turn to the use of such filters in combination with constant-K filters.

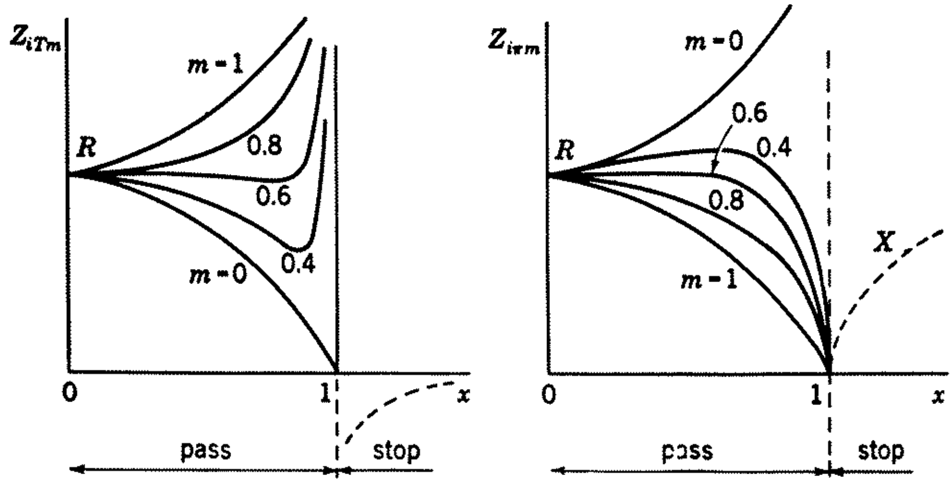


Fig. 13-41. Z_{iTm} and $Z_{i\pi m}$ characteristics.

13-9. Composite filters

The following table summarizes the advantages and disadvantages of constant-K and m-derived filters.

	Constant-K	**m-derived**
Attenuation near cutoff ($x = 1$):	small	large
Attenuation at large x:	large	small
Image impedance in pass band:	not constant	more nearly constant; depends on m; best when $m = 0.6$.

The table illustrates the inverse attenuation characteristics of the two types and suggests that a combination of the two types would have advantages over either type alone. Such a filter is called a *composite filter*. The constant-K filter which forms the nucleus about which the composite filter is designed is known as the *prototype*. The m-derived filter sections will have the same image impedance as the prototype and will have element values found in terms of the constant-K section element values.

In designing a composite filter, two factors must be kept in mind: (1) there must be an image impedance match at the terminals of each filter section, and (2) the attenuation properties of each section must be so selected that the composite attenuation characteristic is that desired. The impedance and attenuation properties of the various network configurations are summarized in Fig. 13-42 and Fig. 13-43. These building blocks are the basis of filter design on the image basis. Several examples will illustrate.

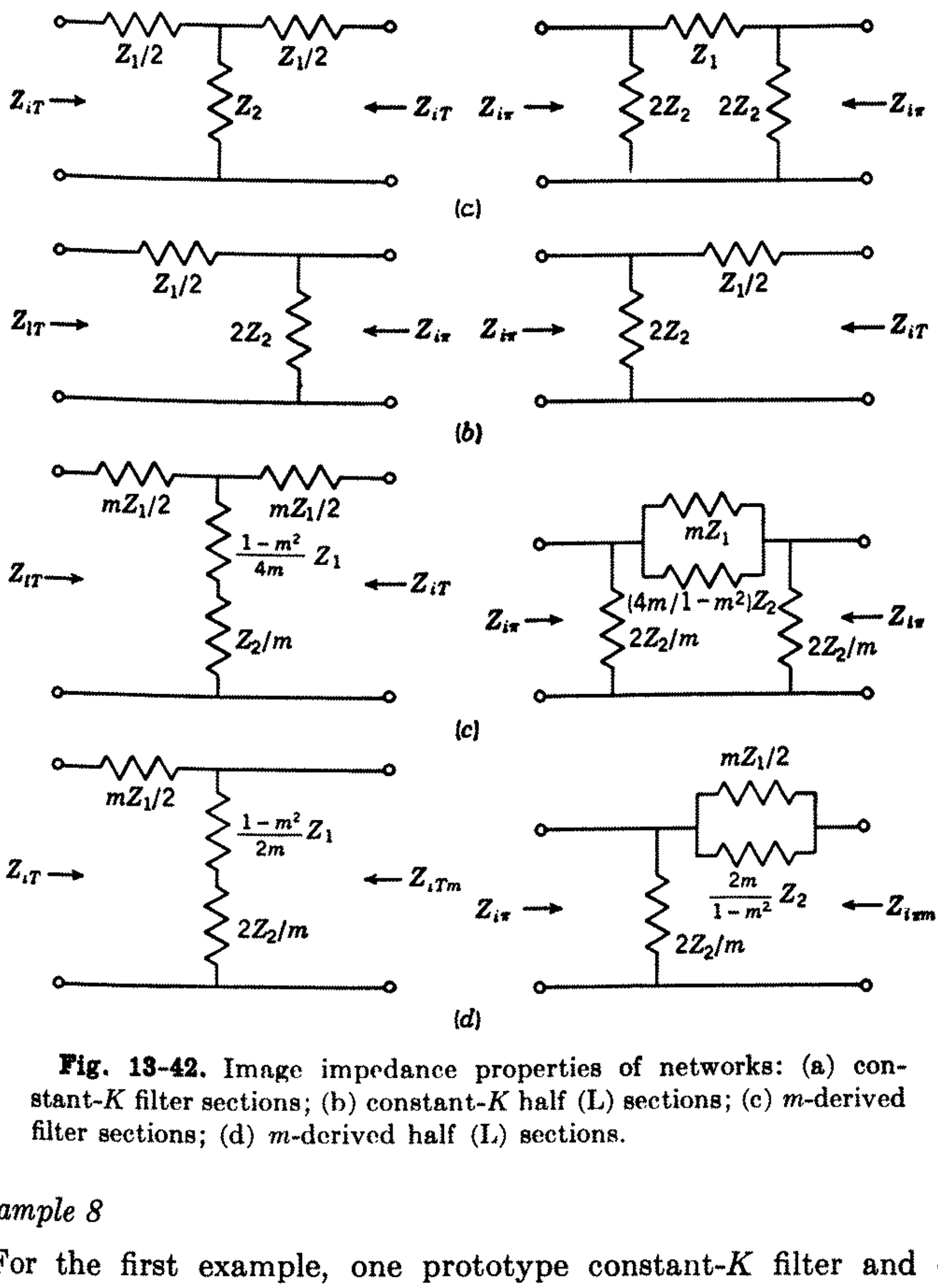


Fig. 13-42. Image impedance properties of networks: (a) constant-K filter sections; (b) constant-K half (L) sections; (c) m-derived filter sections; (d) m-derived half (L) sections.

Example 8

For the first example, one prototype constant-K filter and one m-derived filter will be used in a cascade connection. This is to be a low-pass filter. For this case, we have shown in Eq. 13-90 that $x = \omega/\omega_0$. The prototype (a T section in this case) and the m-derived section are shown in Fig. 13-44. These two sections will make up the composite section of filter. The m-derived section is first split into half sections in order to realize the best impedance properties. An arrangement of the three sections such that there is an image match at each input is shown in Fig. 13-44. The match, however, is only approximate at the load, so that the computed properties are only approximately

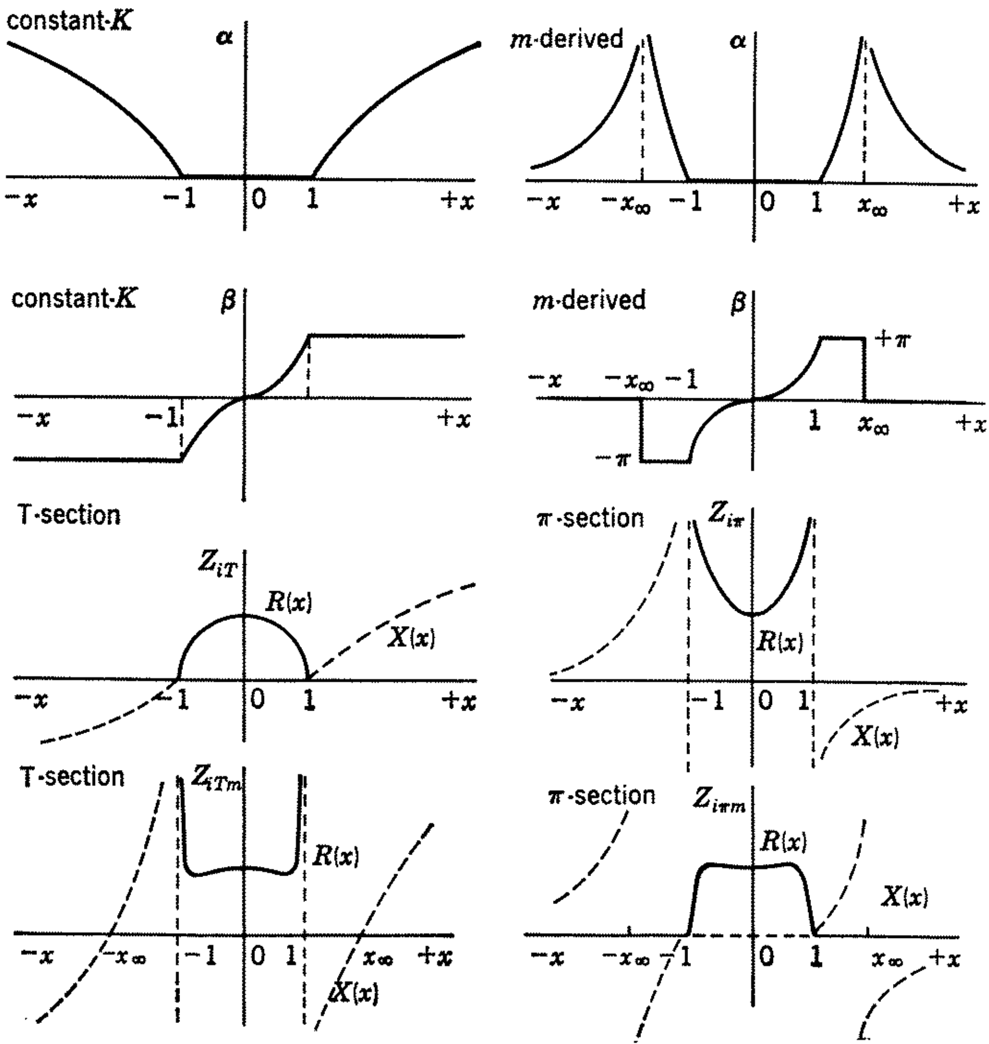


Fig. 13-43. Attenuation, phase and image impedance properties as a function of x.

correct. The input impedance is Z_{iTm} for this arrangement of sections. The attenuation is found by adding the separate attenuations,

$$\alpha_t = \alpha_k + \alpha_m \tag{13-140}$$

as shown in the figure. Similarly,

$$\beta_t = \beta_k + \beta_m \tag{13-141}$$

The input impedance and attenuation of the composite filter are superior to those of either the constant-K filter or the m-derived filter separately. The series inductors of the prototype and the m-derived half section are lumped together when actually constructing the filter.

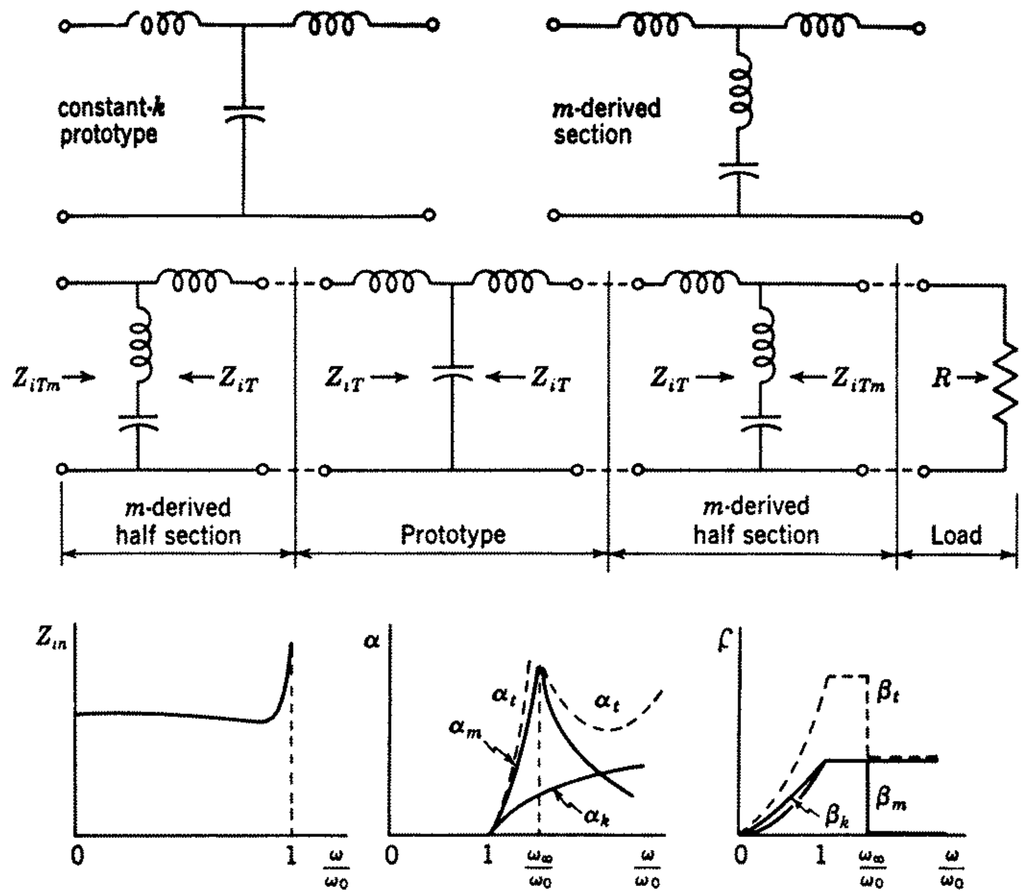


Fig. 13-44. Composite low-pass filter characteristics.

Example 9

In some cases, the attenuation property of the low-pass filter of Example 8 would not be satisfactory, either because the attenuation near cutoff frequency was not sufficiently sharp or because the α_t curve dropped to too low a value before beginning to rise again. In this case,

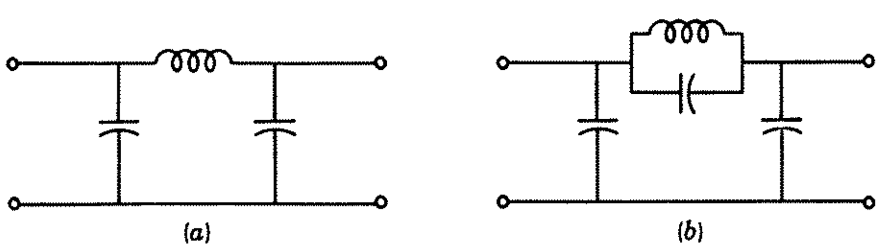


Fig. 13-45. (a) Constant-K prototype and (b) m-derived low-pass sections.

two or more m-derived sections can be used as long as an image match is realized at each input terminal of the cascade connection of sections. To illustrate, suppose that a π section is selected for the prototype of the low-pass filter and a decision is made to use two m-derived sections, one with $m = 0.6$ and one with $m = 0.3$. The prototype and m-derived section are shown in Fig. 13-45. Since the $m = 0.6$ section has superior image impedance characteristics, it will be divided into two half sections for use at the ends of the filter. An arrangement of sections and half sections giving an image impedance match at each terminal pair is shown in Fig. 13-46. The input impedance in this case is $Z_{i\pi m}$ and the total attenuation is

$$\alpha_t = \alpha_k + \alpha_m(0.3) + \alpha_m(0.6) \tag{13-142}$$

The impedance and attenuation properties of the resulting filter section are also shown in Fig. 13-46. Note the improvement in the attenuation

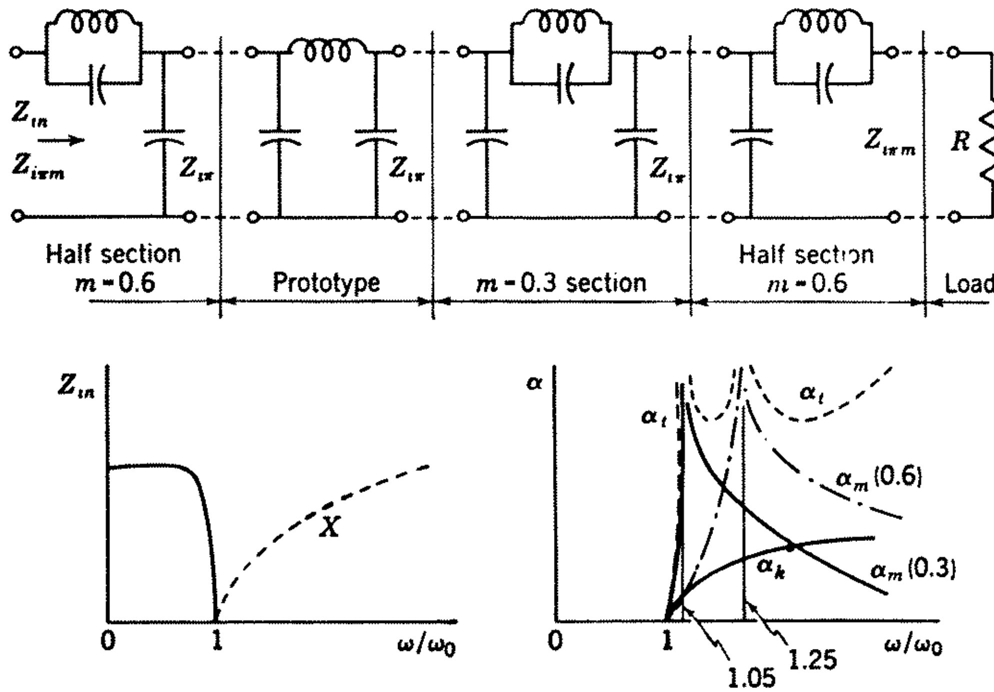


Fig. 13-46. Composite low-pass filter characteristics.

with higher attenuation at all frequencies and sharper cutoff in the stop band.

In practice, the various parallel capacitors of the schematic of Fig. 13-46 would be combined into equivalent capacitors.

Example 10

From the conclusions of the first two examples, let us now formulate a design procedure to use on any filter sections.

(1) First, we should decide on the specifications to be required for the attenuation as a function of frequency. The attenuation requirement may be met with some combination of constant-K filters and m-derived filters according to the equation

$$\alpha_t = A\alpha_k + \sum_{j=1}^{n} B_j \alpha_{mj} \tag{13-143}$$

where A is the number of constant-K sections, B_j is the number of sections for a specific m, and n is the number of different m used. There is usually no unique solution to the type and number of sections required, and cut-and-try must often be used to find a solution to meet the specifications.

(2) Select a constant-K prototype and find the element values. From these element values, find all element values for the m-derived sections.

(3) Include at least one m-derived filter section with $m = 0.6$ for the beginning and ending section of the filter. This gives the optimum image impedance properties.

(4) The type of prototype selected and the number of sections used are usually limited by cost considerations.

13-10. The problem of termination

The constant-K and m-derived sections and half sections can easily be arranged such that there is an image impedance match at each point

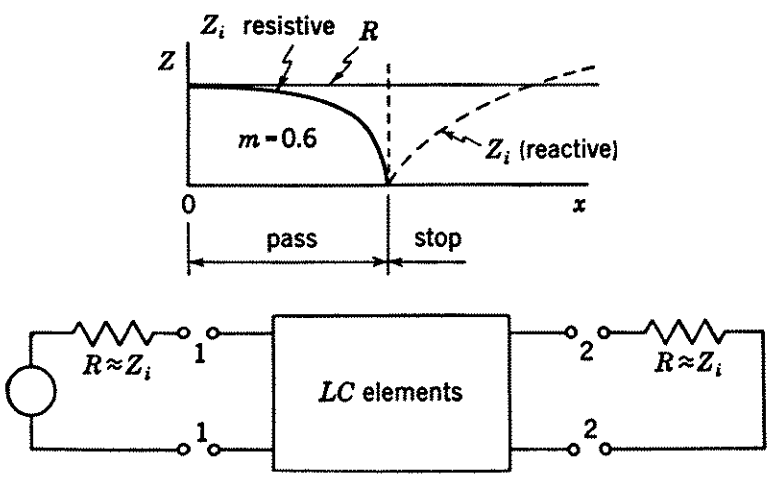


Fig. 13-47. Actual termination of image designed filters.

of connection. But when we come to the beginning or to the termination of the filter, we have a problem in approximating an image match. There are no resistors with the properties of our image impedances, especially the ability to change from resistance to reactance at the cutoff frequency. The best that can be done is to terminate (and make the generator impedance) a constant. What value should this constant resistance be? Let us review the expressions for image impedances as a function of x. By Eqs. 13-82 and 13-83,

$$Z_{iT} = R\sqrt{1 - x^2} \tag{13-144}$$

$$Z_{i\pi} = \frac{R}{\sqrt{1 - x^2}} \tag{13-145}$$

From Eqs. 13-136 and 13-139,

$$Z_{iTm} = R\frac{1 - (1 - m^2)x^2}{\sqrt{1 - x^2}} \tag{13-146}$$

$$Z_{i\pi m} = R\frac{\sqrt{1 - x^2}}{1 - (1 - m^2)x^2} \tag{13-147}$$

Note that all these impedance expressions reduce to $Z_i = R$ when $x = 0$. This constant value is a good approximation to either Z_{iTm} or $Z_{i\pi m}$ with $m = 0.6$ in the pass band. This is the termination commonly used.* The terminating resistor has a value determined from the constant-K prototype as

$$R^2 = +Z_1Z_2 \tag{13-148}$$

All results given thus far have been found on the basis of (1) dissipationless elements in the network and (2) an image match throughout the filter network including the termination. What are the consequences of using elements with finite dissipation (primarily the resistance of inductors)? What are the consequences of terminating the filter in a nonimage impedance, R?

In answer to the first question, the computed values of attenuation are only approximately correct because of finite dissipation. This dissipation causes attenuation in the pass band and finite attenuation at the so-called frequencies of infinite attenuation. A rule of thumb states that the results of image basis design will be acceptable in most engineering applications if the Q of the inductors is 15 or higher.

An answer to the second question requires that we first define the quantity *insertion loss* as the loss resulting when a network is introduced between a generator and a load. The insertion loss is defined by the equation

$$e^N = \left|\frac{I_2'}{I_2}\right| \tag{13-149}$$

where N is the insertion loss in nepers, I_2' is the current in the load connected directly to the generator, and I_2 is the load current with the network in place. For numerical computation of insertion loss, we let $I_2' = I_1$. We thus assume that the generator current is the same with and without the network and so have a basis for comparison. This computation is best made by assuming a unit value for I_2 and then tracing through the network to find the corresponding value for I_1. A

* A slightly better approximation results if the terminating resistor is smaller than R for the m-derived half π section or larger than R for the m-derived half T section as the termination.

comparison of insertion loss N and attenuation α for the network of Fig. 13-44 with $m = 0.6$ is shown in Fig. 13-48.

Even though the image basis attenuation is only approximately equal to the actual insertion loss, the results are usually sufficiently

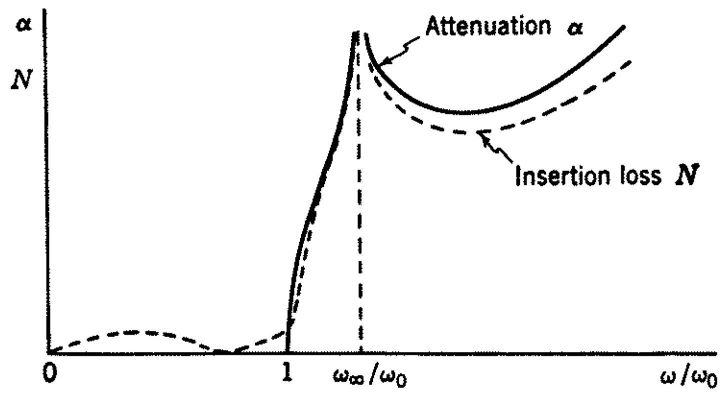


Fig. 13-48. The effect of non-image termination on filter attenuation.

close to be useful. Design on the image basis has the advantage of being simple and routine. Tables showing various networks with their corresponding attenuation characteristics are found in handbooks.*

13-11. Lattice filters

The discussion to this point has related to the ladder structure of networks. Another common structure used in filter design is the *lattice*. A *symmetrical lattice* is shown in Fig. 13-49(a), and the "bridge circuit" equivalent is shown in Fig. 13-49(b). The advantage of the lattice representation over that of the bridge is that sections connected in cascade are more easily drawn as lattice structures.

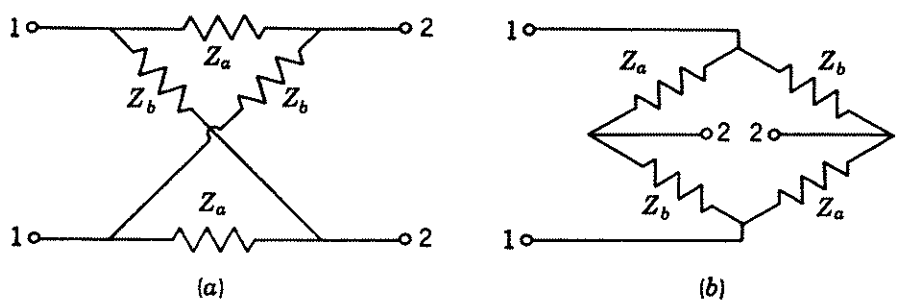


Fig. 13-49. Lattice network structure for filters.

For the symmetrical lattice, the image impedance may be computed from open-circuit and short-circuit impedances. Since

$$Z_{1o} = \frac{Z_a + Z_b}{2} \tag{13-150}$$

* For example, see Terman, *Radio Engineers' Handbook* (McGraw-Hill Book Co., Inc., New York, 1943), pp. 228–236.

and
$$Z_{1s} = \frac{2Z_aZ_b}{Z_a + Z_b} \tag{13-151}$$

it follows that
$$Z_i = \sqrt{Z_{1o}Z_{1s}} = \sqrt{Z_aZ_b} \tag{13-152}$$

The equation for the image transfer function is also a function of Z_a and Z_b as

$$\gamma = 2 \tanh^{-1} \sqrt{Z_a/Z_b} \tag{13-153}$$

Comparing these equations with Eq. 13-6, which is

$$Z_i = \sqrt{Z_{1o}Z_{1s}} \tag{13-154}$$

and Eq. 13-39,
$$\gamma = \tanh^{-1} \sqrt{Z_{1s}/Z_{1o}} \tag{13-155}$$

it is seen that the analysis made previously for pass-band, stop-band, and cutoff frequency in terms of the poles and zeros of Z_{1s} and Z_{1o} holds for the lattice filter, with Z_a replacing Z_{1s} and Z_b replacing Z_{1o}! In summary, when poles of Z_a coincide with zeros of Z_b, or vice versa, there is defined a *pass band*. When poles or zeros of Z_a coincide with poles or zeros, respectively, of Z_b, there is defined a *stop band*. A critical frequency in Z_a but not in Z_b, or vice versa, defines a cutoff frequency. These rules assume that Z_a and Z_b are reactance functions (that is, LC elements only).

We also have the results we need to compute the attenuation and phase shift. By Eq. 13-43,

$$\tanh \frac{\gamma}{2} = \frac{\tanh (\alpha/2) + j \tan (\beta/2)}{1 + j \tanh (\alpha/2) \tan (\beta/2)} = \sqrt{\frac{Z_a}{Z_b}} \tag{13-156}$$

If the *sign* of Z_a is *opposite* to that of Z_b, then $\tanh (\gamma/2)$ is imaginary, and

$$\tan \frac{\beta}{2} = \sqrt{\frac{-Z_a}{Z_b}} \tag{13-157}$$

$$\alpha = 0 \tag{13-158}$$

If Z_a and Z_b have the *same sign*, then either

$$\tanh \frac{\gamma}{2} = \tanh \frac{\alpha}{2} \quad \text{and} \quad \beta = 0 \tag{13-159}$$

by Eq. 13-45, or

$$\tanh \frac{\gamma}{2} = \frac{1}{\tanh (\alpha/2)} \quad \text{and} \quad \beta = \pi \tag{13-160}$$

by Eq. 13-46. Now $\tanh (\alpha/2)$ cannot exceed unit value corresponding to infinite α. It follows that the choice of these two possibilities depends on the magnitude of Z_a and Z_b. Then

$$\alpha = 2 \tanh^{-1} \sqrt{Z_a/Z_b}, \qquad \beta = 0 \quad \text{when } Z_b > Z_a \tag{13-161}$$
$$\alpha = 2 \tanh^{-1} \sqrt{Z_b/Z_a}, \qquad \beta = \pi \quad \text{when } Z_b < Z_a \tag{13-162}$$

These equations permit computation of α and β in the pass and stop bands.

From Eq. 13-161, we see that the attenuation becomes infinite as Z_a/Z_b approaches unit value. Similarly, the attenuation is small when Z_a/Z_b is small. In selecting positions for the poles and zeros for Z_a and Z_b, the pass-band poles and zeros determine the phase variation, and their position is determined by the desired form of phase variation (for example, linear variation is often required). In the stop band, however, the poles and zeros of Z_a and Z_b are selected so that the quotient Z_a/Z_b remains as nearly unity as possible as frequency varies. A procedure for locating these poles and zeros has been given by Bode and Dietzold.*

From Eqs. 13-161 and 13-162, it is seen that as Z_b exceeds Z_a in magnitude, or vice versa, the phase of the output changes by 180°. Under this condition, the output voltage of the lattice effectively reverses polarity. The filtering action in the case of the lattice takes place by there being a balance of the bridge circuit shown in Fig. 13-49(b). For a perfect balance, corresponding to infinite attenuation, the components must be of high quality and carefully matched. This is one disadvantage of lattice filters. However it is possible to get infinite attenuation at the frequency of balance even with finite dissipation, if the effective resistances also balance.

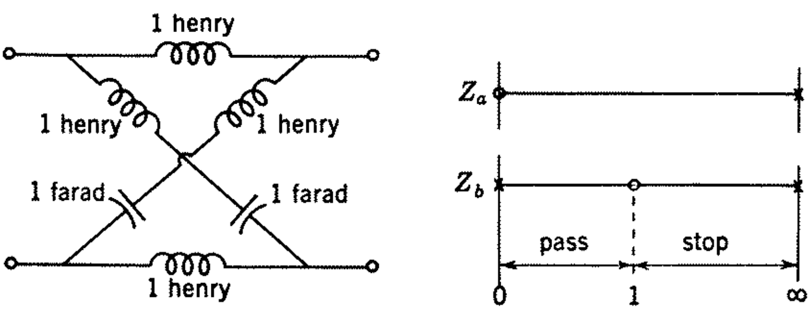


Fig. 13-50. Lattice and poles and zeros of Z_a and Z_b.

Example 11

The lattice shown in Fig. 13-50 has element values such that

$$Z_a = s \quad \text{and} \quad Z_b = \frac{s^2 + 1}{s} \tag{13-163}$$

* H. W. Bode and R. L. Dietzold, "Ideal Wave Filters," *Bell System Tech. J.*, **14**, 215 (1935).

The pole-zero plot shown in Fig. 13-50 indicates that this is a low-pass filter with a cutoff frequency of $\omega_0 = 1$. Since $Z_a = j\omega$ and $Z_b = j(\omega^2 - 1/\omega)$,

$$\alpha = 0, \qquad \beta = 2 \tan^{-1} \sqrt{\omega^2/(\omega^2 - 1)}, \qquad 0 \leqq \omega \leqq 1 \quad (13\text{-}164)$$

by Eqs. 13-157 and 13-158 and

$$\alpha = 2 \tanh^{-1} \sqrt{(\omega^2 - 1)/\omega^2}, \qquad \beta = \pi, \qquad 1 \leqq \omega \leqq \infty \quad (13\text{-}165)$$

by Eq. 13-162, since $Z_a > Z_b$ for $\omega \geqq 1$. For this particular lattice, the image impedance is

$$Z_i = \sqrt{1 - \omega^2} \quad (13\text{-}166)$$

Another example illustrating properties of the symmetrical lattice was given in Art. 11-6, page 263. For that particular case, the entire frequency range was pass band, and the phase characteristic was given by an equation of the form of Eq. 13-157.

13-12. Bartlett's bisection theorem

A relationship between the lattice impedances Z_a and Z_b and the open-circuit and short-circuit impedances was suggested on page 354. The equivalence of these quantities is given in a theorem originally due to *Bartlett.* This theorem applies only for *symmetrical* two-terminal-pair networks. Bartlett's bisection theorem provides a means for finding the lattice impedances for a lattice network *equivalent* to a symmetrical ladder network.

The first step in the application of this theorem is *bisection* of the symmetrical ladder network. By the term *bisection,* we mean that we divide the network into identical parts such that the two networks, when reversed end for end, have identical geometrical as well as electrical properties. Such a bisected network with only connecting wires

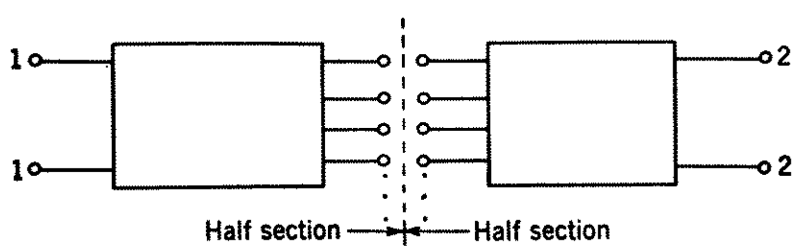


Fig. 13-51. Bisected symmetrical network.

showing appears in Fig. 13-51. Bartlett's bisection theorem,* given here without proof, states: The lattice equivalent of a symmetrical ladder network has a series arm Z_a equal to the impedance of a half

* Bartlett, A. C., *Theory of Electrical Artificial Lines and Filters,* (John Wiley & Sons, Inc., New York, 1931), pp. 53–58; Brune, Otto, "Note on Bartlett's bisection theorem," *Phil. Mag.*, **14**, 806 (1932).

of the bisected network measured at terminals 1-1 or 2-2, with the other terminals *short-circuited;* the shunt arm Z_b is equal to the impedance of the half network with the bisected terminals *open.* Two examples will illustrate the application of this theorem.

Example 12

The standard T section is shown in original form and also bisected in Fig. 13-52. Following Bartlett's bisection theorem, the open-circuit

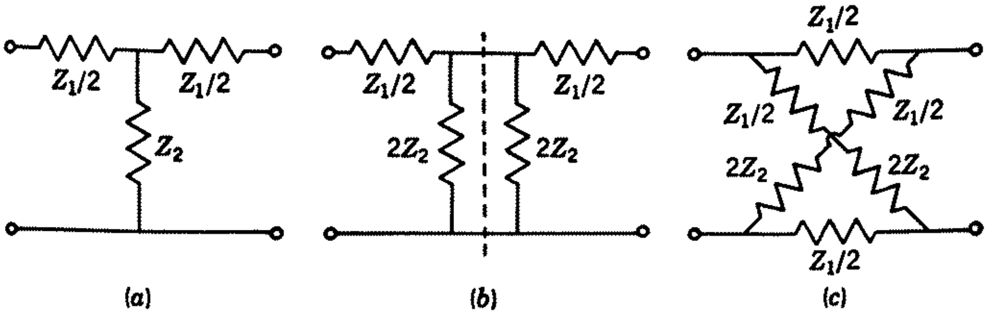


Fig. 13-52. Application of Bartlett's theorem: (a) original network; (b) bisected network; (c) equivalent lattice.

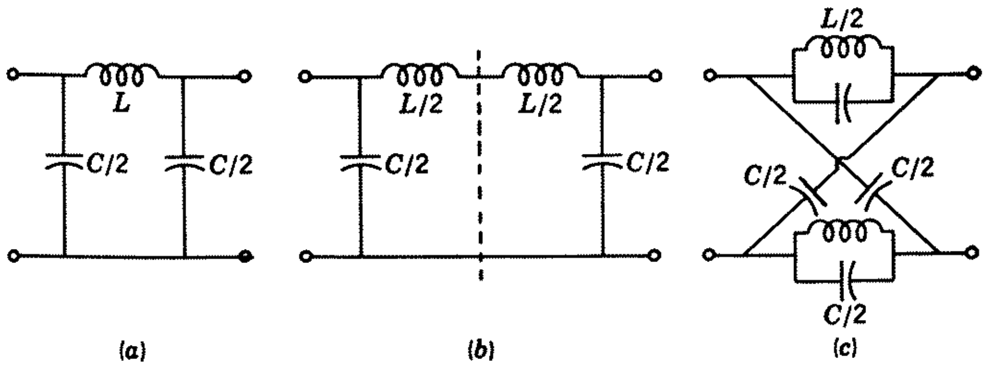


Fig. 13-53. Networks of Example 13: (a) original network; (b) bisected network; (c) equivalent lattice.

and short-circuit impedances are found for the half network. The resulting equivalent lattice is shown in the figure.

Example 13

For this example, the standard π section is shown for the low-pass filter case. The bisected network and resulting equivalent lattice are also shown in the figure.

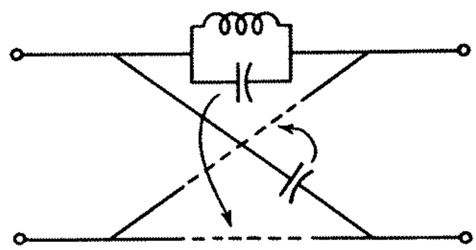


Fig. 13-54. Conventional representation of the lattice network.

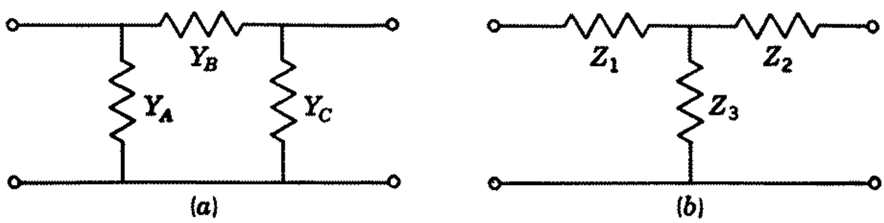
To avoid the complicated structure of the lattice in drawings, it is usual practice to replace one of the series arms and one of the shunt arms by a dashed line. Thus the lattice of Fig. 13-54 is defined to be identical with the lattice of Fig. 13-53.

FURTHER READING

The original articles on the constant-K and m-derived filters by O. J. Zobel are found in the *Bell System Technical Journal* under the titles, "Theory and design of uniform and composite electric wave filters" in the volume for 1923 and "Extensions to the theory and design of electric wave filters" in 1931. For additional discussion of these topics see: Guillemin's *Communication Networks, Vol. II* (John Wiley & Sons, Inc., New York, 1935), Chaps. 5, 8, 9; D. F. Tuttle, Jr., *Network Synthesis*, 2 vols. (John Wiley, & Sons, Inc., New York, in preparation); W. L. Everitt, *Communication Engineering* (McGraw-Hill Book Co., Inc., New York, 1937), pp. 179–240; J. D. Ryder, *Networks, Lines, and Fields* (Prentice-Hall, Inc., New York, 1949), pp. 114–163; LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 218–236; and Reed, *A-C Circuit Theory* (Harper & Brothers, New York, 1948), pp. 553–597.

PROBLEMS

13-1. The T and π networks shown in the accompanying figure are also known in electrical engineering literature as Y *and delta networks*,



Prob. 13-1. (a) π or delta network; (b) T or wye network.

respectively. Networks can sometimes be simplified by converting from a Y to an equivalent delta or from a delta to an equivalent Y.* (a) Show that, if a delta equivalent of a Y network exists, the following relationships hold.

$$Y_A = \frac{Z_2}{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}$$

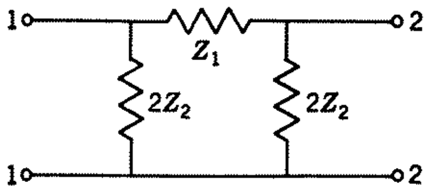
$$Y_B = \frac{Z_3}{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}$$

$$Y_C = \frac{Z_1}{Z_1Z_2 + Z_2Z_3 + Z_3Z_1}$$

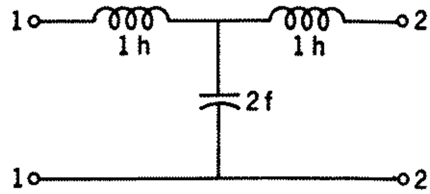
* The notion of delta-Y equivalence is originally due to A. E. Kennelly in 1899. His article, "The equivalence of triangles and three-point stars in conducting networks," appeared in *Electric World and Engineering*.

(b) Find the corresponding transformation for the Y network equivalent of a delta network. Give the values for Z_1, Z_2, and Z_3 in terms of Y_A, Y_B, and Y_C.

13-2. A two-terminal-pair π network is shown in the figure. Show that for this network, $\coth \gamma = \sqrt{Z_{1o}/Z_{1s}}$, where $\gamma = \alpha + j\beta$ and $I_1/I_2 = e^\gamma$, and Z_{1o} = the impedance at terminal pair 1 with terminal pair 2 open, Z_{1s} = the impedance at terminal pair 1 with terminal-pair 2 short-circuited.



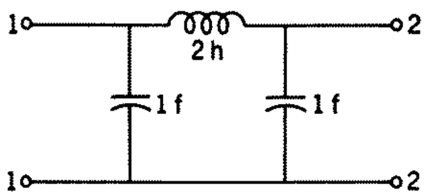
Prob. 13-2.



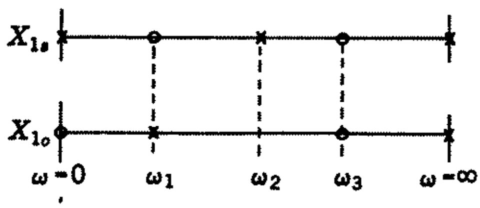
Prob. 13-3.

13-3. For the T network shown above in the figure, determine and, plot the image-impedance Z_i for the frequency range $\omega = 0$ to $\omega = 1$. *Answer.* $Z_i = \sqrt{1 - \omega^2}$.

13-4. Repeat Prob. 13-3 for the π network shown in the figure. *Answer.* $Z_i = 1/\sqrt{1 - \omega^2}$.



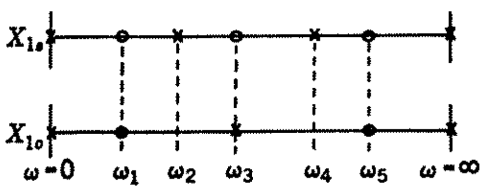
Prob. 13-4.



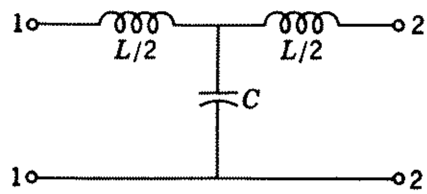
Prob. 13-5.

13-5. A pole-zero plot for Z_{1s} and Z_{1o} is shown in the accompanying figure. From the plots, Determine: (a) the pass bands, (b) the stop bands, (c) the cutoff frequencies.

13-6. Repeat Prob. 13-5 for the pole-zero plot shown in the figure.



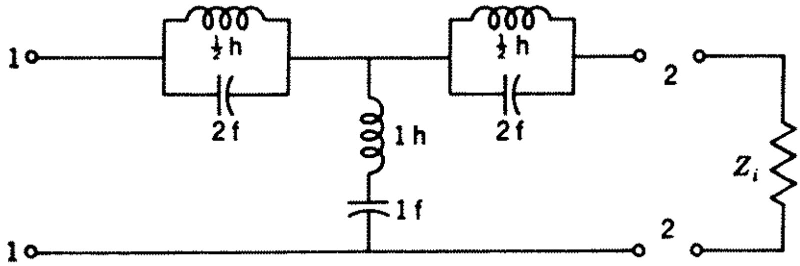
Prob. 13-6.



Prob. 13-7.

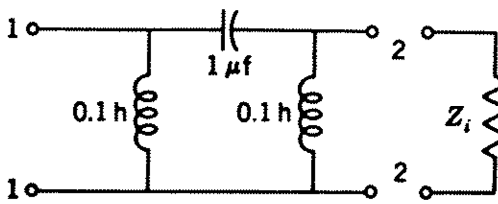
13-7. The symmetrical T network shown in the figure has element values as indicated. Starting with Eqs. 13-30 and/or 13-32, which apply to the T network, derive an equation in terms of L, C, and ω for the attenuation α in the stop band and for the phase shift β in the pass band. *Answer.* $\alpha = \cosh^{-1} (2 - \omega^2 LC)/2$, $\beta = \cos^{-1} (2 - \omega^2 LC)/2$.

13-8. For the schematic shown in the figure, (a) Analyze the network to determine the pass bands and the stop bands. (b) Determine all cutoff frequencies. *Answer.* $\omega_{C1} = 0.781$, $\omega_{C2} = 1.282$.



Prob. 13-8.

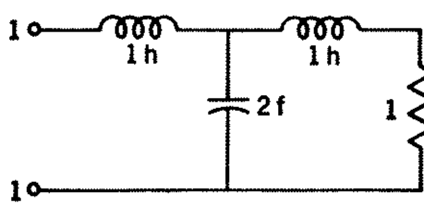
13-9. Repeat Prob. 13-8 for the π section shown below.



Prob. 13-9.

13-10. Design a low-pass filter having a cutoff frequency of 1000 cycles per second and a purely resistive image impedance of 100 ohms at 0 cycles per second. Give element values. *Answer.* $C = 3.18\ \mu$f, $L = 31.8$ mh.

13-11. In Prob. 13-3, it was found that $Z_i = \sqrt{1 - \omega^2}$ in the pass band. The attenuation found, for example, in Prob. 13-7 applies only



Prob. 13-11.

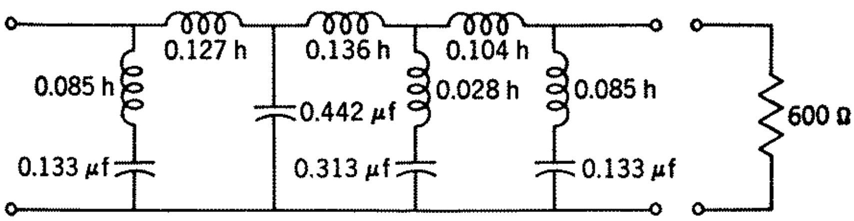
when the T section is terminated in this Z_i. As a practical approximation, let $Z_i = R = 1$ ohm (constant). To investigate how good such an approximation is: (a) Plot the insertion loss of the circuit shown above by solving for $|I_2'/I_2| = e^N$ defined by Eq. 13-149 for $0 \leq \omega \leq 2$. (b) On the same coordinates, plot the function, $\alpha = 2 \cosh^{-1} \omega$ for $1 \leq \omega \leq 2$. This is the approximate result.

13-12. For the filter given in Prob. 13-8, determine the value of x as a function of ω. Make a careful sketch of: (a) the attenuation α as a function of ω, (b) the phase shift β as a function of ω, (c) the image impedance as a function of ω. Make use of the normalized plots of Fig. 13-25 in working this problem. (d) What is the value of R in Eq. 13-81?

13-13. Repeat Prob. 13-12 for the network given in Prob. 13-9.

13-14. Show that if F is the number of cutoff frequencies, a single section of a constant-K filter requires $3F$ elements.

13-15. Design a composite m-derived low-pass filter to the following specifications: (a) the termination is a 600-ohm resistor, (b) the cutoff



Prob. 13-15. Solution.

frequency is 1200 cycles per second, (c) the frequencies of infinite attenuation, ω_∞, are 1500 and 1700 cycles per second. Draw the schematic diagram for the filter with all possible series and parallel elements combined. Indicate all element values.

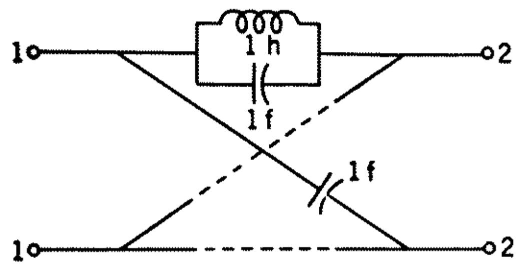
13-16. Design a filter to the following specifications:

Pass band:	0 to 2000 cycles per second.
Cutoff:	Output must be no more than 5% of the input at 2100 cycles per second and all higher frequencies.
Termination:	Load resistor will have a value of 600 ohms.
End section:	End sections should be m-derived half sections with $m = 0.6$.

(a) Draw the schematic diagram of the filter and indicate all element values. (Note: As in most design problems, there is no unique solution to this problem.) (b) How many sections of constant-K filter are needed to meet specifications?

13-17. A network is to be composed of the cascade connection of *four* constant-K half sections (or L sections) and *two* m-derived half sections. Draw schematic diagrams of all the possible ways these half sections can be combined such that there is an *image impedance match* at each of the cascaded terminal pairs. Consider *both* the π and T m-derived half sections.

13-18. For the lattice filter shown below, determine (a) the pass band, (b) the stop band, (c) the cutoff frequency, (d) the phase shift

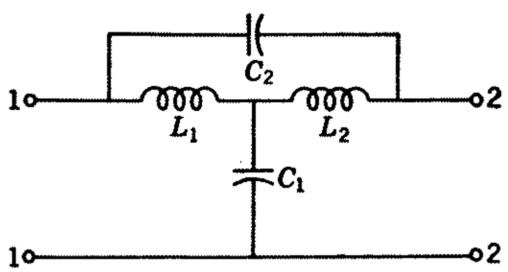


Prob. 13-18.

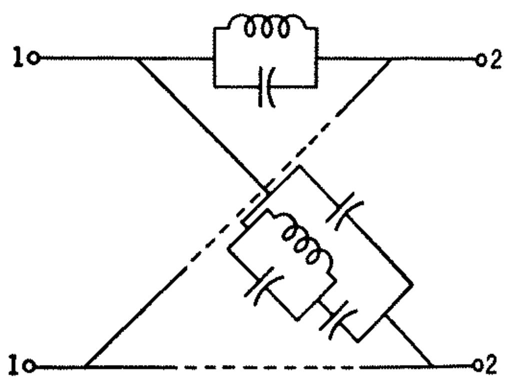
in the pass band, (e) the attenuation in the stop band, (f) the phase shift in the stop band and (g) the image impedance. *Answer.* (a) $0 - 1$ radians/sec, (b) $1 - \infty$ radians/sec, (c) $\omega_\infty = 1$ radians/sec, (d) $\beta = 2\tan^{-1}\sqrt{\omega^2/(\omega^2 - 1)}$, (e) $\alpha = 2\tanh^{-1}\sqrt{(\omega^2 - 1)/\omega^2}$, (f) π, (g) $Z_i = 1/\sqrt{1 - \omega^2}$.

13-19. Determine the lattice equivalent of the network of Prob. 13-8. Show all element values.

13-20. The network shown in the figure is known as a bridged T. Determine the lattice equivalent of the network if $L_1 = L_2$.



Prob. 13-20.



Prob. 13-21.

13-21. A lattice structure is shown without element values. Determine a possible *ladder* equivalent of this lattice by studying Z_a and Z_b. Mark element values (L_1, L_2, C_1, etc.) noting which elements in Z_a and Z_b must be equal.

CHAPTER 14

AMPLIFIER NETWORKS

14-1. Shunt peaked amplifier network

Frequency-sensitive networks are often used in conjunction with vacuum tube amplifiers to give a combination filter-amplifier. Such a network is shown in Fig. 14-1. In some practical networks, the inductor may be the only element connected to the plate of the vacuum tube, usually a pentode. The resistor R represents the resistance of the inductor, and the capacitor C may represent the wiring capacitance and interelectrode capacitance of the tube. If the plate resistance of the vacuum tube is very high (as in the case of pentodes), the output voltage is given by the equation*

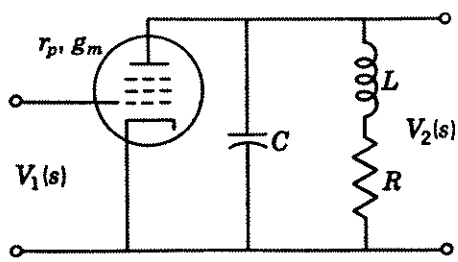


Fig. 14-1. Shunt peaked amplifier network.

$$V_2 = -g_m Z V_1 \tag{14-1}$$

where g_m is the tube transconductance and Z is the impedance of the network connected to the plate. The impedance for the network of Fig. 14-1 is

$$Z(s) = \frac{(1/Cs)(Ls + R)}{1/Cs + Ls + R} = \frac{1}{C}\left[\frac{s + R/L}{s^2 + Rs/L + 1/LC}\right] \tag{14-2}$$

The denominator polynomial has been encountered many times before and may also be written in terms of the dimensionless damping ratio ζ and the natural undamped frequency ω_n. From Eq. 14-1, the voltage ratio transfer function may be written in terms of the impedance expression. The resulting equation is

$$\frac{V_2(s)}{V_1(s)} = -\frac{g_m}{C}\,\frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{14-3}$$

The pole-zero configuration of the transfer function is evident from this equation: there are a pair of conjugate poles and one real zero. However, analysis is frequently made in terms of the circuit Q discussed

* Equation 14-1 is derived in Ryder, *Electronic Engineering Principles*, 2d ed. (Prentice-Hall, Inc., New York, 1952), p. 220.

in Art. 11-4, and so this equation will be rearranged in terms of this quantity. Circuit Q is defined as

$$Q = \frac{\omega_n L}{R} = \frac{1}{2}\frac{\omega_n}{R/2L} = \frac{1}{2\zeta} \tag{14-4}$$

The complex frequency s will be normalized by division by ω_n. This new frequency variable will be designated by the letter p and defined as

$$p = s/\omega_n = \sigma/\omega_n + j\omega/\omega_n \equiv \sigma_p + j\omega_p \tag{14-5}$$

With these substitutions, the equation for the transfer function becomes

$$\frac{V_2(p)}{V_1(p)} = -\frac{g_m}{\omega_n C}\left[\frac{p + 1/Q}{p^2 + p/Q + 1}\right] \tag{14-6}$$

The poles of this function are evidently

$$p_a, p_a^* = -\frac{1}{2Q} \pm \sqrt{\left(\frac{1}{2Q}\right)^2 - 1} \qquad Q < \tfrac{1}{2} \tag{14-7}$$

$$= -\frac{1}{2Q} \pm j\sqrt{1 - \left(\frac{1}{2Q}\right)^2} \qquad Q > \tfrac{1}{2} \tag{14-8}$$

In practical networks, Q is much larger than $\frac{1}{2}$, and so the second equation will be considered the typical case. The pole-zero configuration in terms of circuit Q is shown in Fig. 14-2. The locus of the poles and

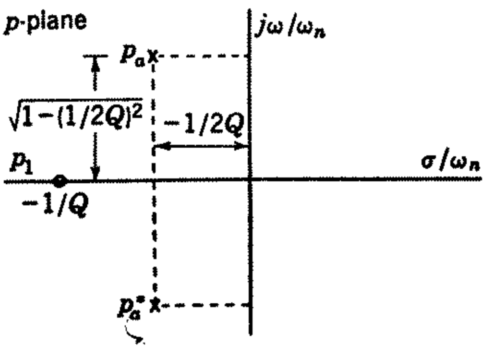


Fig. 14-2. Pole-zero location in terms of Q.

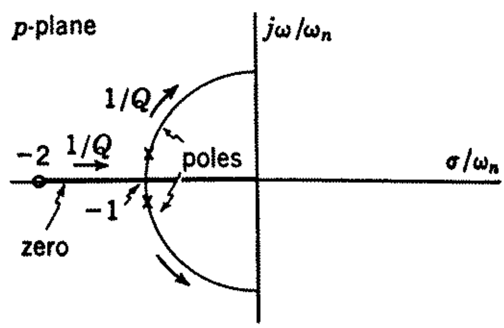


Fig. 14-3 Pole-zero locus in terms of Q.

zero is shown in Fig. 14-3 for various values of Q greater than the critical value of $\frac{1}{2}$. As Q decreases toward $\frac{1}{2}$, the zero approaches the point -2, and the poles approach the point -1.

Frequency response of the shunt peaked amplifier is found by letting $p = j\omega_p$ and computing the gain and phase for a number of frequencies. The magnitude of the voltage ratio or the amplifier *gain* is found from the relationship

$$\left|\frac{V_2(j\omega_p)}{V_1(j\omega_p)}\right| = \frac{g_m}{\omega_n C} \frac{|j\omega_p - p_1|}{|j\omega_p - p_a||j\omega_p - p_a^*|} \tag{14-9}$$

As the frequency varies from 0 to ∞, the length $|j\omega_p - p_a|$ changes rapidly, going through a minimum value when ω_p is nearest p_a. The response magnitude reaches a maximum value at a frequency near (but not at) this minimum. As frequency increases, the magnitude function falls at the rate of 6 db per octave for high frequencies. A typical response curve is shown in Fig. 14-4. In contrast to the series *RLC* circuit considered in Art. 11-4, the phase angle is *not* zero at resonance for this network. This fact can be readily established by inspection of the pole-zero configuration. This method of visualizing the results in terms of the pole-zero configuration is simpler than an algebraic investigation which involves manipu'ation of complex numbers.

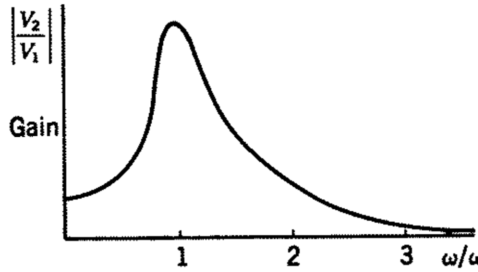


Fig. 14-4. Typical response curve for a shunt peaked amplifier network.

14-2. Stagger-tuned amplifier networks

An important property of the shunt peaked amplifier network, shown in Fig. 14-1, was not discussed in the previous section. This property follows from the fact that the input to the grid of the vacuum tube draws negligible current and thus has high internal impedance. Specifically, the property is that this amplifier network may be connected to another network *without loading;* that is, without causing any significant current to flow such that the output voltage of the other network would be altered by connecting the amplifier network. For this reason, successive stages of amplifier networks may be connected together in *cascade* (or *tandem*), and each network will be independent of all others.

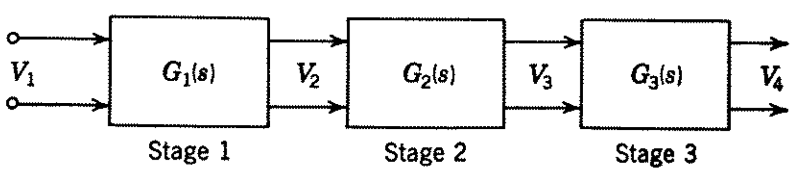


Fig. 14-5. Cascade connection of amplifier networks.

A cascade connection of amplifier stages is shown in Fig. 14-5, where each stage is represented by a block. The output of the first amplifier is connected to the input of the second, the output of the second amplifier is connected to the input of the third, and so on. Any number of stages may be so connected. If the amplifier input has a very high impedance and the amplifier output has a low impedance, the stages of amplification and filtering will be *independent* in the sense that the second amplifier will not affect the first, the third will not affect the second, and so on. Each stage is isolated from all other stages; each stage lives in a world of its own, accepting the voltage it receives without influencing the "giver," and in turn without being influenced by the stage that receives its output.

The practical reason for cascade connection of stages of amplifiers is that one stage does not provide enough voltage gain. The ordinary superheterodyne receiver uses at least two stages of amplification (the so-called intermediate frequency, or IF amplifiers); it is common for six to eight stages of amplification to be used in radar receivers. Each stage is ordinarily terminated in a network made up of resistors, inductors, and capacitors. For such networks, the output voltage is a function of frequency. In this section, we will investigate desirable forms of the variation of the output voltage to the input voltage with radian frequency.

Under the assumption of no loading, the voltage ratio transfer functions may be written

$$G_1(s) = \frac{V_2(s)}{V_1(s)}, \qquad G_2(s) = \frac{V_3(s)}{V_2(s)}, \qquad G_3(s) = \frac{V_4(s)}{V_3(s)} \tag{14-10}$$

where G_1, G_2, and G_3 are successively the voltage ratio transfer functions for the first, second, and third stages. The output voltage $V_4(s)$ may be found in terms of the input voltage $V_1(s)$ by the following manipulation.

$$\frac{V_2(s)}{V_1(s)}\frac{V_3(s)}{V_2(s)}\frac{V_4(s)}{V_3(s)} = \frac{V_4(s)}{V_1(s)} = G_t(s) = G_1(s)G_2(s)G_3(s) \tag{14-11}$$

where $G_t(s)$ is the voltage transfer function for the three stages connected in cascade. This mathematical operation would not have been possible had not the output voltage been a function of the input voltage only for each stage (i.e., each stage isolated from the others).

In the sinusoidal steady state, two properties of the total transfer function $G(j\omega)$ are important in design. The first is the maximum magnitude of the transfer function, or the maximum gain. The other is the variation of magnitude with radian frequency. It has been found that desired combinations of gain and gain variation with frequency cannot be attained by simply cascading *identical* amplifier network stages. Better performance can be realized if each stage is made slightly different. The composite amplifier network is then said to be *stagger tuned.* The design of stagger-tuned amplifier networks is easily accomplished in terms of the pole-zero configuration. Our approach to this design problem will be to consider first the desirable pole-zero configurations for stagger-tuned amplifiers. We will then show how such pole-zero configurations can be realized by the design of the amplifier networks of each stage.

In the discussion to follow, we will restrict ourselves to the case of a low-pass filter; that is, a filter which passes low frequencies with high gain and attenuates high frequencies by means of a low gain. The techniques will apply directly to the case of band-pass filters and high-pass filters, as will later be illustrated by several examples. For the time being, we will also restrict our discussion to voltage ratio transfer functions having the form

$$\frac{V_{n+1}(s)}{V_1(s)} = K \frac{1}{b_0 s^n + b_1 s^{n-1} + \ldots + b_n} \tag{14-12}$$

where n is the number of stages of amplification. In the sinusoidal steady-state, the magnitude of this transfer function has the form

$$\left| \frac{V_{n+1}(j\omega)}{V_1(j\omega)} \right| = K' \frac{1}{\sqrt{\omega^{2n} + A_1 \omega^{2n-2} + \ldots + A_n}} \tag{14-13}$$

This magnitude is a function of ω^2, designated $M(\omega^2)$, as may be seen by reviewing the way the magnitude of a complex function is formed; that is, the magnitude is equal to the square root of the real part of the function squared, plus the imaginary part squared. The frequency ω raised to either an even or an odd power has an even exponent when squared. For example,

$$\left| \frac{1}{a + j\omega} \right| = \frac{1}{\sqrt{\omega^2 + a^2}} \tag{14-14}$$

and the magnitude function contains ω to even exponents only.

The function represented by Eq. 14-13 has the value $K'/\sqrt{A_n}$ for $\omega = 0$. As frequency becomes larger, the magnitude decreases at the rate of $-n \times 6$ db per octave. An ideal form for the magnitude as a function of frequency curve to have for intermediate frequencies is shown in Fig. 14-6. The curve is *flat* from $\omega = 0$ up to a frequency called the *cutoff frequency* and then asymptotically approaches the $-n \times 6$ db per octave rate of decrease. We now face a number of problems: Just how flat can the curve be made? How do we go about accomplishing this flat characteristic in terms of Eq. 14-13?

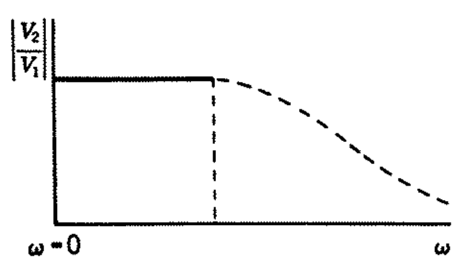


Fig. 14-6. Ideal flat characteristic.

The slope of the voltage ratio function with frequency can be made equal to zero by requiring that the derivative of M with respect to ω^2 (rather than ω since M is a function of ω^2) be equal to zero. If the rate of change of the slope of the $M(\omega^2)$ curve is also made equal to zero, the flatness of the curve will be improved. The pattern becomes clear: make just as many derivatives of $M(\omega^2)$ zero as we can. The more derivatives that are zero at $\omega = 0$, the flatter the curve will be at a band of frequencies above $\omega = 0$.

If Eq. 14-13 is differentiated with respect to ω^2, there results

$$\frac{d}{d\omega^2} \cdot \left| \frac{V_{n+1}(j\omega)}{V_1(j\omega)} \right| = K' \frac{A_1'\omega^{2n-2} + A_2'\omega^{2n-4} + \ldots + A_{n-1}}{(\omega^{2n} + A_1\omega^{2n-2} + \ldots + A_n)^{3/2}} \tag{14-15}$$

This expression can be made equal to zero at $\omega = 0$ by setting A_{n-1} equal to zero. Setting successively higher derivatives equal to zero will make all the A-coefficients zero up to A_1. Under these conditions, Eq. 14-13 has the form

$$\left| \frac{V_{n+1}(j\omega)}{V_1(j\omega)} \right| = \frac{K'}{\sqrt{\omega^{2n} + A_n}} \tag{14-16}$$

To simplify the form of this equation, let $A_n = \beta^{2n}$ and $\omega/\beta = \omega_p$. Then

$$\left| \frac{V_{n+1}(j\omega_p)}{V_1(j\omega_p)} \right| \frac{\beta^n}{K'} = \frac{1}{\sqrt{\omega_p^{2n} + 1}} \tag{14-17}$$

When staggered amplifiers are designed to satisfy the relationship given by this equation, they are said to be *maximally flat*. Such staggered amplifiers are also called *Butterworth* amplifiers after an author who first described such amplifier design in 1930. (Networks with no amplifiers and no isolation may also have the maximally flat characteristic and are called *Butterworth filters*.) The actual curve realized by the last equation does not exactly fit the ideal shown in Fig. 14-6. The larger the value of n can be made, the better the approximation to the ideal curve. A plot of the function given by Eq. 14-17 for several values of n is shown in Fig. 14-7. All the curves pass through the point 0.707 (the half-power point) at $\omega_p = 1$. This will be shown later in the chapter.

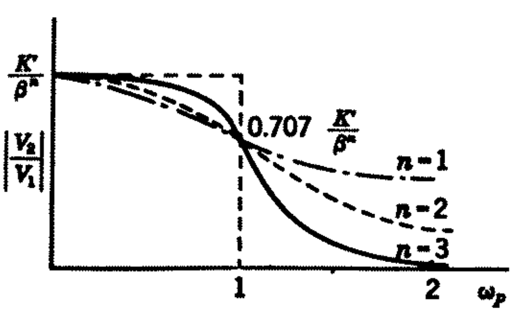


Fig. 14-7. Maximally flat characteristic.

The next problem is to find the positions of the poles of the voltage ratio transfer function that will give an absolute magnitude function of the form of Eq. 14-17. To simplify notation, let $p = j\omega_p$ or $\omega_p = p/j$. The square root factor of Eq. 14-17 then becomes

$$-p^{2n} + 1 \qquad \text{if } n \text{ is odd} \tag{14-18}$$

$$p^{2n} + 1 \qquad \text{if } n \text{ is even} \tag{14-19}$$

We proceed in our investigation by taking what appears to be an indirect route. Consider the functions

$$F(p) = \frac{1}{p^{2n} + 1}, \qquad n \text{ is even} \tag{14-20}$$

$$H(p) = \frac{1}{-p^{2n} + 1}, \qquad n \text{ is odd} \tag{14-21}$$

The poles of $F(p)$ occur when the denominator of the first of these equations is equal to zero; that is,

$$p^{2n} + 1 = 0 \quad \text{or} \quad p^{2n} = -1 \tag{14-22}$$

Similarly, the poles of $H(p)$ occur under the condition

$$-p^{2n} + 1 = 0 \quad \text{or} \quad p^{2n} = +1 \tag{14-23}$$

To find the $2n$th roots of ± 1, we write this number in polar form as

$$-1 = e^{\pm j\pi} = e^{\pm j3\pi} = e^{\pm j5\pi} = \dots \tag{14-24}$$

$$+1 = e^{\pm 0\pi} = e^{\pm j2\pi} = e^{\pm j4\pi} = \dots \tag{14-25}$$

These equations may be written in generalized form as

$$-1 = e^{\pm j(2m-1)\pi}, \qquad m = 1, 2, 3, \dots \tag{14-26}$$

$$+1 = e^{\pm j2k\pi}, \qquad k = 0, 1, 2, \dots \tag{14-27}$$

Setting these equations equal to p^{2n} and taking the $2n$th root of both sides of the equation gives

$$p_m = e^{\pm j(2m-1)\pi/2n}, \qquad m = 1, 2, 3, \dots, n \tag{14-28}$$

$$p_k = e^{\pm jk\pi/n}, \qquad k = 0, 1, 2, \dots, n \tag{14-29}$$

These expressions locate the poles of $F(p)$ and $H(p)$ given above. The magnitude of each root of the last two equations is unity, and the roots are separated by π/n radians. The location of the roots for an odd n ($n = 3$) and an even n ($n = 4$) are shown in Fig. 14-8. Similar plots for other values of n are readily made by following these rules:

(1) For odd values of n, a root is always located on the $+\sigma_p$ and on the $-\sigma_p$ axis. Other roots are displaced from these real roots by π/n radians.

(2) For even values of n, no roots are located on the real axis. Roots are located $\pi/2n$ radians from the positive and negative real axis. All roots are displaced by π/n radians.

(3) There is a symmetry with respect to both the real axis and the imaginary axis. No roots occur on the $j\omega_p$ axis.

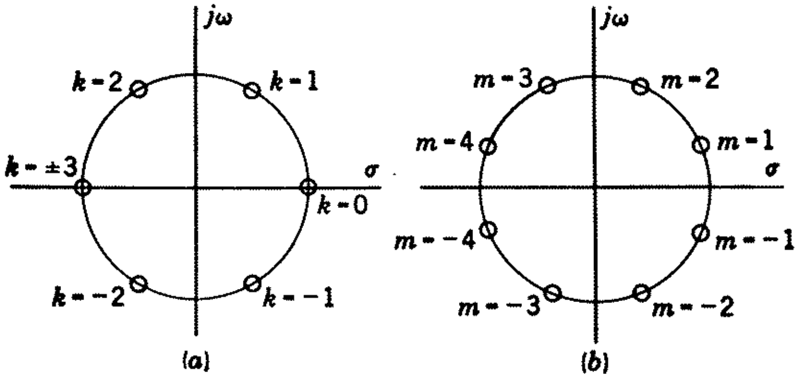


Fig. 14-8. Location of roots of ± 1: (a) $n = 3$ (odd); (b) $n = 4$ (even).

The roots plotted in Fig. 14-8 are really *poles* of the functions $F(p)$ and $H(p)$ defined by Eqs. 14-20 and 14-21. We know that impedance functions and transfer functions cannot have poles in the right half plane as $F(p)$ and $H(p)$ have. These functions, however, are not necessarily network functions. They are only arbitrary functions that have been invented in the expectation that they might somehow relate to the transfer functions having magnitudes of the maximally flat form.

Because of rule (3) stated above, there are always as many poles of the functions $F(p)$ and $H(p)$ in the right half plane as in the left half plane. If the right half-plane poles are grouped together and designated $f_r(p)$, and the left half-plane poles are similarly grouped as $f_l(p)$, we can write

$$F(p) = f_r(p)f_l(p) \tag{14-30}$$

When $p = j\omega_p$ in the sinusoidal steady state, the magnitude of $f_r(p)$ is always equal to the magnitude of $f_l(p)$. This can be seen from the pole configuration: phasors drawn from each of the poles to a point on the $j\omega_p$ axis can be matched in identical pairs as far as magnitude is concerned. Because of this equality,

$$|f_r(j\omega_p)| = |f_l(j\omega_p)| \tag{14-31}$$

the magnitude of $F(j\omega)$ may be written

$$|F(j\omega_p)| = |f_r(j\omega_p)||f_l(j\omega_p)| = |f_l(j\omega_p)|^2 \tag{14-32}$$

Extracting the square root of this equation gives

$$|f_l(j\omega_p)| = \sqrt{|F(j\omega_p)|} \tag{14-33}$$

Now $F(p)$ is defined by Eq. 14-20 for *even* n only. For these even values of n, $F(j\omega)$ is a *real number* having the value

$$F(j\omega) = \frac{1}{1 + \omega_p^{2n}} \tag{14-34}$$

Substituting this magnitude into Eq. 14-33 yields

$$f_l(j\omega) = \sqrt{1/(1 + \omega_p^{2n})} \tag{14-35}$$

This equation is precisely of the form of Eq. 14-17 which defined the maximally flat function. Thus the pole configuration described by $f_l(p)$ is the one required to give a maximally flat magnitude characteristic. These poles are readily found by Eqs. 14-28 and 14-29—alternately by the rules of page 370—provided *only poles in the left half plane are retained.* Similarly, it follows that for *odd* values of n, that

$$|h_l(j\omega_p)| = \sqrt{H(j\omega_p)} = \sqrt{1/(1 + \omega_p^{2n})} \tag{14-36}$$

under the same requirement that only left half-plane poles be con-

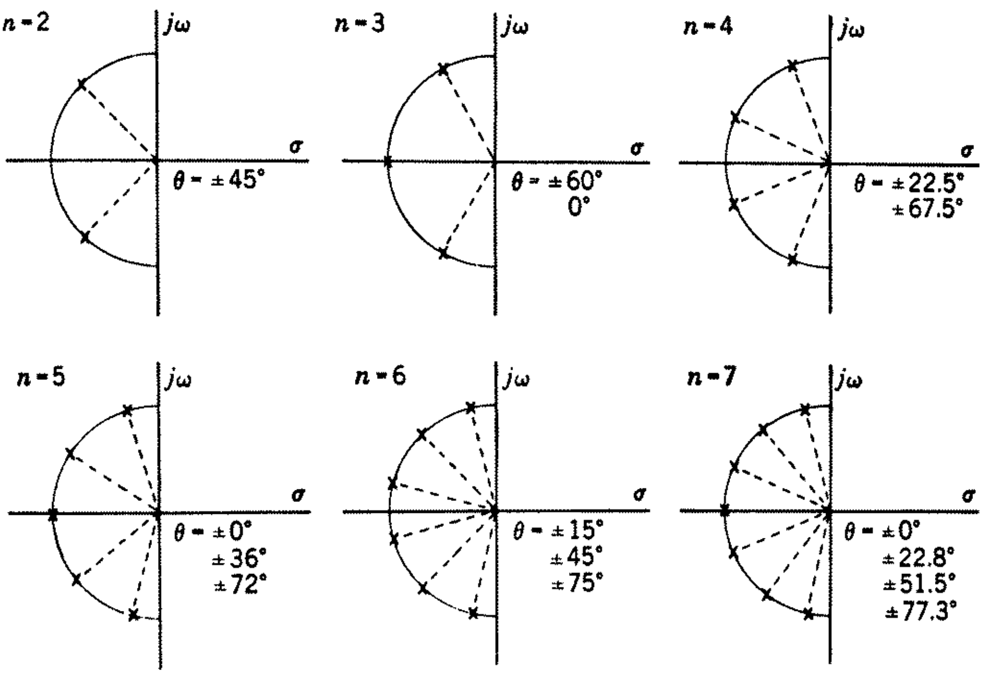


Fig. 14-9. Pole locations for maximally flat frequency response (θ measured from the negative real axis).

sidered. The pole configurations to give maximally flat magnitude characteristics for several values of n are shown in Fig. 14-9.

Now the ω_p of the two network functions we have found, Eq. 14-35 and Eq. 14-36, is normalized radian frequency. If this term is rewritten as ω/ω_n, then in each case we are considering terms of the form

$$\frac{1}{\sqrt{1 + (\omega/\omega_n)^{2n}}} \tag{14-37}$$

This function has the value 1 at $\omega/\omega_n = 0$. And when $\omega/\omega_n = 1$ or $\omega = \omega_n$, the function has the value $(1/\sqrt{2})$ independent of the value of n! This frequency is designated the *half-power frequency* as was done in Eq. 11-79, and there is another half-power frequency at negative ω_n. The response curve more nearly approximates a constant value for larger values of n.

14-3. Overstaggered amplifiers (Chebyshev polynomials)

This last discussion points to a disadvantage of stagger-tuning an amplifer for the maximally flat (or Butterworth) condition. The magnitude function approximates a constant for a range of frequencies, but as frequency becomes larger the approximation is poor. This is illustrated in Fig. 14-10. The ideal characteristic is a constant shown

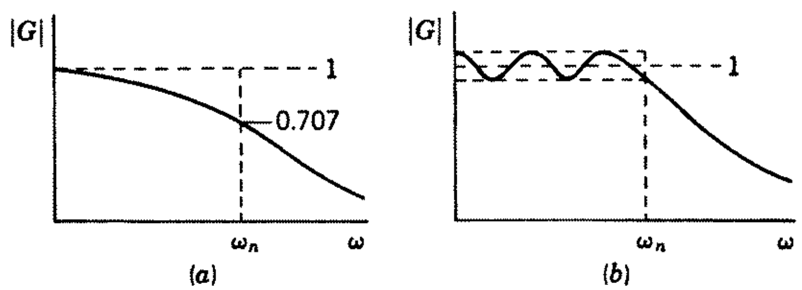


Fig. 14-10. Comparison of responses.

as a dashed line. The Butterworth response closely approximates the ideal characteristic for low frequencies, but the difference between the ideal and actual characteristic becomes large as frequency increases. All the error (the difference between the ideal and the actual) is lumped at high frequency. The total characteristic would seem to be better if this error could be spread out over the entire band of low frequencies (for this low-pass filter case). Such a frequency response is shown in Fig. 14-10(b). The error is spread out from $\omega = 0$ to $\omega = \omega_n$ as an "equal ripple," a series of hills and dales. The maximum error is the same for several points. Such a frequency response appears to be better than the maximally flat response. We face two problems: Can we write an expression in mathematical form for this response? Can we find the pole configuration that gives this response?

The equal ripple type of functions illustrated in Fig. 14-10(b) were originally studied by the Russian mathematician P. L. Chebyshev some 100 years ago. Chebyshev found a certain type of polynomial useful in his studies of the action of linkages used in steam engines. These polynomials, which we will call Chebyshev* polynomials, approximate a constant in the characteristic equal ripple form we have illustrated. The general Chebyshev polynomial of order n is defined by the equation

$$C_n(\omega) = \cos\,(n \cos^{-1}\omega) \tag{14-38}$$

Chebyshev polynomials for several values of n are as follows.

$$C_1 = \cos\,(\cos^{-1}\omega) = \omega \tag{14-39}$$

$$C_2 = 2\omega^2 - 1 \tag{14-40}$$

$$C_3 = 4\omega^3 - 3\omega \tag{14-41}$$

$$C_4 = 8\omega^4 - 8\omega^2 + 1 \tag{14-42}$$

$$C_5 = 16\omega^5 - 20\omega^3 + 5\omega \tag{14-43}$$

$$C_6 = 32\omega^6 - 48\omega^4 + 18\omega^2 - 1 \tag{14-44}$$

$$C_{n+1} = 2\omega C_n - C_{n-1} \tag{14-45}$$

The last equation may be used to calculate higher-ordered Chebyshev polynomials.

The equation for the magnitude of the voltage ratio transfer function (for the sinusoidal steady state) corresponding to the maximally flat case defined by Eq. 14-17 has the form, for n stages,

$$\frac{V_{n+1}(j\omega_p)}{V_1(j\omega_p)} = G(j\omega_p) = \frac{1}{\sqrt{1 + \epsilon C_n^2(\omega_p)}} \tag{14-46}$$

where ϵ is a constant (to be defined). Just as in the maximally flat case, n will be related to the number and location of the poles (and hence to the number of amplifier stages).

To construct this equal ripple function, we must start with a Chebyshev polynomial of order n and square it. The squared function is multiplied by the constant ϵ and added to unity. The reciprocal of the square root of the resulting function is the transfer function magnitude. This process can be duplicated by performing each step graphically. The Chebyshev polynomial

$$C_n(\omega) = \cos\,(n \cos^{-1}\omega) \tag{14-47}$$

is defined for a range of ω from $+1$ to -1 (that is $-1 \leq \omega \leq +1$).

* Chebyshev is also variously spelled as Tchebycheff, Tchebichef, etc. These forms apparently resulted from repeated translation: Russian to German, German to French, etc. The spelling used is considered the best Russian to English translation.

When $|\omega| > 1$, $\cos^{-1}\omega = j\cosh^{-1}\omega$ and

$$C_n(\omega) = \cosh\,(n\cosh^{-1}\omega) \tag{14-48}$$

This function does not have a "rippling" nature but increases with ω at a rate determined by the value of n. Plots of $C_n(\omega)$ for several values of n are shown in Fig. 14-11. The functions vary from $+1$ to -1 over the frequency range $0 \leqq \omega \leqq 1$. Several features of the Chebyshev

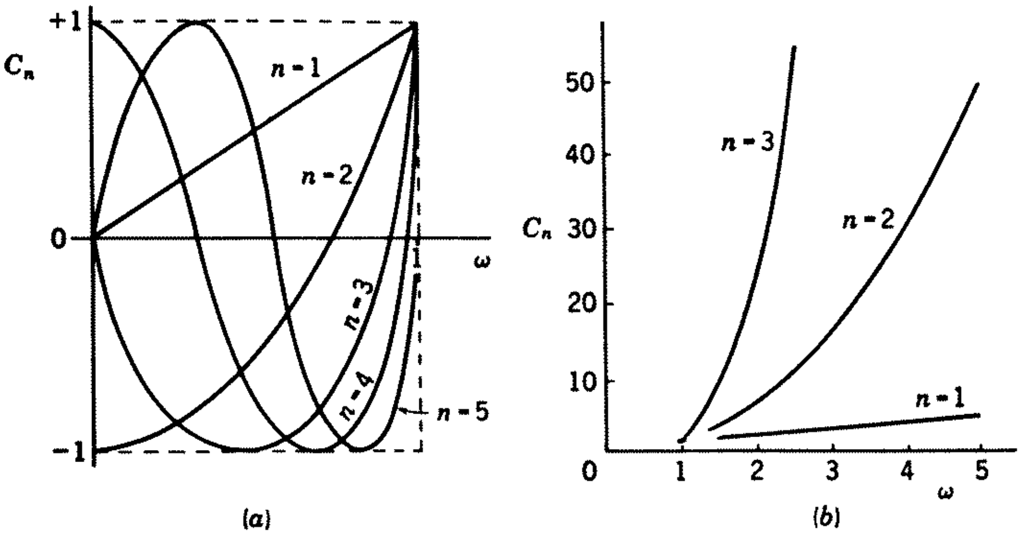


Fig. 14-11. Chebyshev polynomials: (a) plot for $0 \leq \omega \leq 1$; (b) plot for $\omega \geq 1$.

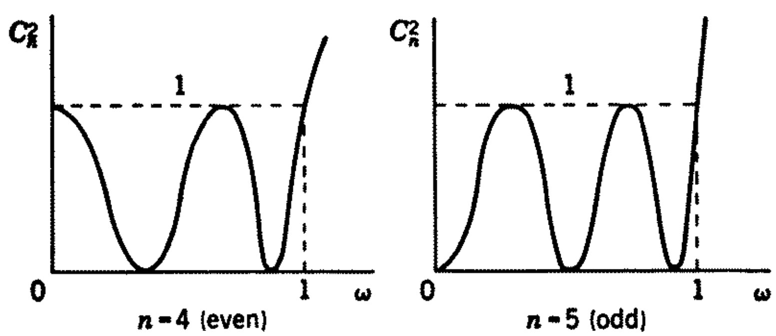


Fig. 14-12. Squared Chebyshev polynomials.

polynomials can be seen from the plots of Fig. 14-11. For all *odd* values of n, $C_n(\omega)$ has zero value at $\omega = 0$ and the initial slope is alternately positive and negative. For *even* values of n, $C_n(\omega)$ alternately has the value $+1$ and -1 at $\omega = 0$. When the Chebyshev polynomials are squared, all the negative values of the $C_n(\omega)$ plot will be "reflected" as positive values. For *odd* n all plots of $C_n^2(\omega)$ start from zero and have initially increasing values; for *even* n all plots of $C_n^2(\omega)$ start from $+1$ and have an initially decreasing values. Typical plots of the squared function are shown in Fig. 14-12.

To construct our equal-ripple frequency response, the curves of Fig. 14-12 above are multiplied by ϵ (which will merely reduce the scale) and substituted into the equation

$$|G(j\omega_p)| = \frac{1}{\sqrt{1 + \epsilon C_n^2(\omega_p)}} \tag{14-49}$$

When $C_n(\omega_p)$ has zero value, $G(j\omega_p)$ will have unit value; when $C_n(\omega_p)$ has the value of unity (the maximum value it can have), the magnitude function will be

$$\frac{1}{\sqrt{1 + \epsilon}} \tag{14-50}$$

The ripples will vary between these two limits as shown in Fig. 14-13 for two typical values of n (corresponding to Fig. 14-12). The ripple

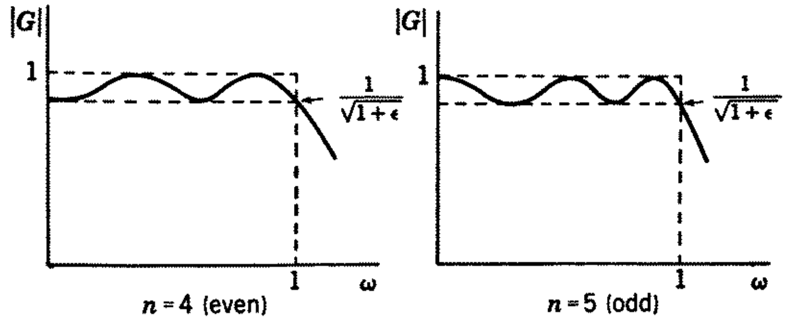


Fig. 14-13. Equal-ripple frequency response.

width is often specified in decibels. This width and ϵ are related by the equation

$$\text{ripple width} = \delta = +20 \log_{10} 1 - 20 \log_{10} \frac{1}{\sqrt{1 + \epsilon}} \tag{14-51}$$

or

$$\delta = 10 \log_{10} (1 + \epsilon) \tag{14-52}$$

where δ is in decibels. From this equation ϵ can be found if δ is specified (in a design problem).

At this point, the quantity ϵ will be defined in terms of a new factor, a in order to simplify the computation. This relationship will be justified later in this chapter and will appear as Eq. 14-71. Then, by definition,

$$\epsilon = \frac{1}{\sinh^2 (na)} \tag{14-53}$$

or

$$a = \frac{1}{n} \sinh^{-1} \frac{1}{\sqrt{\epsilon}} \tag{14-54}$$

where n is the order of the Chebyshev polynomial.

In the maximally flat case, the frequency response curve had dropped to 0.707 at $\omega_p = 1$. For the Chebyshev equal-ripple frequency response, the curve has a value $1/\sqrt{1+\epsilon}$ at the same frequency. In the Chebyshev case, another frequency has significance. Let $\omega_{p_1} = \cosh a$, corresponding to a particular frequency larger than $\omega_p = 1$. By the equation for the Chebyshev polynomial,

$$C_n(\cosh a) = \cosh [n \cosh^{-1} (\cosh a)] \tag{14-55}$$

$$= \cosh na \tag{14-56}$$

Now the factor a is defined by Eq. 14-54. Substituting this value for a in Eq. 14-56 gives

$$C_n(\cosh a) = \cosh \left(\sinh^{-1} \frac{1}{\sqrt{\epsilon}}\right) \tag{14-57}$$

The hyperbolic sine and hyperbolic cosine are related by the identity

$$\cosh^2 x - \sinh^2 x = 1 \tag{14-58}$$

If, from Eq. 14-57,

$$\sinh x = \frac{1}{\sqrt{\epsilon}} \tag{14-59}$$

then, by the above identity,

$$\cosh x = \sqrt{1 + 1/\epsilon} \tag{14-60}$$

or

$$x = \cosh^{-1} \sqrt{1 + 1/\epsilon} \tag{14-61}$$

Then Eq. 14-57 may be written

$$C_n(\cosh a) = \cosh (\cosh^{-1} \sqrt{1 + 1/\epsilon}) = \sqrt{1 + 1/\epsilon} \tag{14-62}$$

If this equation is squared and unity is added to both sides of the equation, there results

$$1 + \epsilon C_n^2(\cosh a) = 2 + \epsilon \tag{14-63}$$

This equation is in the form of the square root factor of the frequency response given by Eq. 14-49. Making this substitution gives

$$|G(j \cosh a)| = \frac{1}{\sqrt{2+\epsilon}} \approx 0.707 \tag{14-64}$$

if ϵ is much smaller than 1. Thus for the approximation that ϵ is small (the usual case), the frequency $\omega_{p_1} = \cosh a$ corresponds to the "half-power" frequency in the maximally flat case. This is illustrated in Fig. 14-14 (along with other information that we have deduced thus far). The approximate half-power frequency, cosh a, is determined by the value of a, which is determined by both n and ϵ. For a large value of n, a is small and cosh a has a value only slightly larger than unity. In other words, the larger the value of n, the faster the frequency response falls off with frequency above $\omega_p = 1$ (giving better filtering action). This steepness of the response characteristic is also dependent upon ϵ, which is in turn dependent upon the size of the ripples in the frequency range $\omega_p = 0$ to $\omega_p = 1$. The summary of our knowledge of equal-ripple frequency response characteristics, shown in Fig. 14-14,

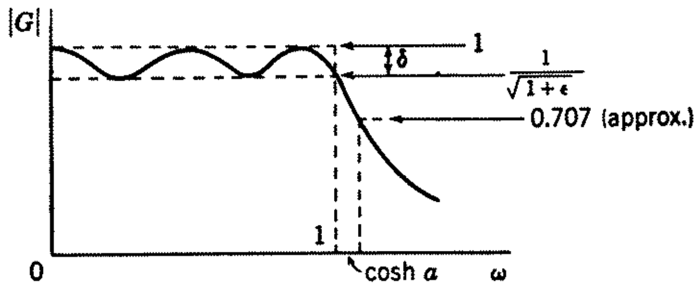


Fig. 14-14. Half-power point of frequency response (n = number of half cycles of ripple).

indicates that specifications are complete and given in terms of ϵ, a, δ, and n. We next turn our attention to the pole configuration that will give the equal-ripple characteristics.

The procedure for determining the locations for the poles parallels that for the maximally flat (or Butterworth) case. In this procedure, a function of the form of Eq. 14-49 with ω_p replaced by some other variable is written, and the poles of this function are determined. As in the maximally flat case, the poles in the right half plane are rejected to give the magnitude function. Paralleling the discussion leading to Eqs. 14-18 and 14-19, we let $p = j\omega_p$ or $\omega_p = p/j$ and examine the function appearing under the radical in Eq. 14-49 which is

$$1 + \epsilon C_n^2(p/j) = 0 \quad \text{or} \quad C_n^2(p/j) = -1/\epsilon \tag{14-65}$$

Now $C_n(x)$, where x is any variable, is defined as $C_n(x) = \cos n \cos^{-1} x$, $|x| \leqq 1$, so that the last equation becomes

$$\cos n \cos^{-1} (p/j) = \pm j/\sqrt{\epsilon} \tag{14-66}$$

Since the inverse cosine of a complex number is complex in general, we define $\cos^{-1} (p/j) = \alpha - ja$ such that

$$\cos (n\alpha - jna) = \pm j/\sqrt{\epsilon} \tag{14-67}$$

Expansion of the cosine of the difference of two angles gives

$$\cos n\alpha \cosh na + j \sin n\alpha \sinh na = \pm j/\sqrt{\epsilon} \tag{14-68}$$

The real and imaginary parts of this equation may be equated to give

$$\cos n\alpha \cosh na = 0 \qquad \text{and} \qquad \sin n\alpha \sinh na = \pm 1/\sqrt{\epsilon} \tag{14-69}$$

Since $\cosh na$ cannot equal zero for any value of a, α must have values given by the equation

$$\alpha = \frac{2N+1}{n}\frac{\pi}{2} \text{ radians}, \qquad N = 0, 1, 2, \ldots, n \tag{14-70}$$

For these values of α, $\sin n\alpha = \pm 1$ and

$$a = \pm \frac{1}{n} \sinh^{-1} (1/\sqrt{\epsilon}) \tag{14-71}$$

This equation was introduced without proof as Eq. 14-54 in order to simplify the discussion at that point.

Since we have now determined the required values of α and a in the equation, $p/j = \cos(\alpha - ja)$, we write

$$p = j \cos\left(\frac{2N+1}{n}\frac{\pi}{2} - ja\right), \qquad N = 0, 1, 2, \ldots, n \tag{14-72}$$

or $$p = \cosh a \left(-\tanh a \sin \frac{2N+1}{n}\frac{\pi}{2} + j \cos \frac{2N+1}{n}\frac{\pi}{2}\right),$$
$$N = 0, 1, 2, \ldots, n \tag{14-73}$$

This equation defines the roots of Eq. 14-65 as required. Our next step will be to modify the form of this equation for comparison with the results of the maximally flat case. From the identities, $\cos x = \sin\left(\frac{\pi}{2} - x\right)$ and $\sin x = \cos\left(\frac{\pi}{2} - x\right)$, Eq. 14-73 may be written

$$p = \cosh a\,(-\tanh a \cos b + j \sin b) \tag{14-74}$$

where $$b = \frac{n - 2N - 1}{2n}\pi, \qquad N = 0, 1, 2, \ldots, n \tag{14-75}$$

We have already found that the poles for the maximally flat case are located on a circle with locations given by the equation

$$p = e^{jb'} = \cos b' + j \sin b' \tag{14-76}$$

where b' values are given by Eq. 14-28 for even values of n and by Eq. 14-29 for odd values of n as

$$b' = \frac{2m-1}{2n}\pi, \qquad n \text{ even} \tag{14-77}$$

$$b' = k\pi/n, \qquad n \text{ odd} \tag{14-78}$$

We next compare these angles with those of Eq. 14-73. These two angles are equal for the following integer values for m and k, found by equating Eq. 14-75 to Eqs. 14-77 and 14-78

$$m = (n - 2N)/2 \qquad \text{for even } n$$
and
$$k = (n - 1 - 2N)/2 \qquad \text{for odd } n \tag{14-79}$$

where $N = 0, 1, 2, \ldots, n$. For a given value of n, the angles b and b' are equal although they are specified in different orders according to Eq. 14-79. The equality of these angles for the equal ripple and maximally flat cases is the basis for a simplified method for locating the roots in the equal ripple case.

Comparing the equations, $p_1 = \cos b' + j \sin b'$ and $p_2 = \cosh a\,(-\tanh a \cos b + j \sin b)$, we see that the roots for the equal ripple case can be found from the roots for the maximally flat case by the following steps: (1) Change the radius of the circle of the maximally flat case from 1 to $\cosh a$. (2) Multiply the real part of the poles located for the maximally flat case by $\tanh a$. This construction is illustrated in Fig. 14-15. The angles tabulated in Fig. 14-9 will be found useful in constructing the new pole configurations for the equal ripple case.

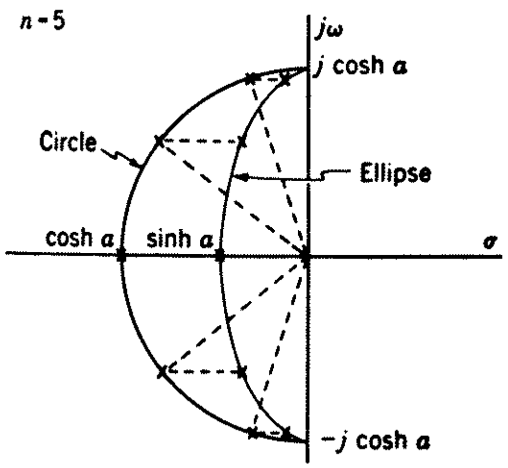


Fig. 14-15. Location of poles for Chebyshev case. Real parts of poles on circle are multiplied by ($\tanh a$).

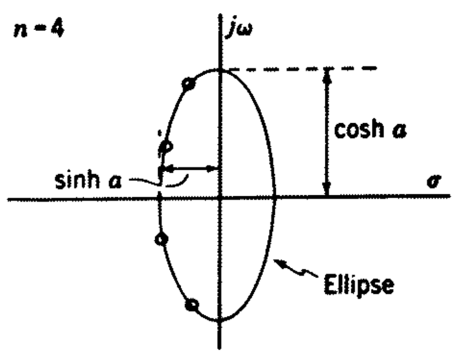


Fig. 14-16. Pole locations for Chebyshev case.

The roots for the equal ripple case are located on the periphery of an ellipse. This can be demonstrated by noting from Eq. 14-74 that if $p = \sigma_p + j\omega_p$, then $\sigma_p = -\sinh a \cos b$ and $\omega_p = \cosh a \sin b$. From these two equations it follows that

$$\frac{\sigma_p^2}{\sinh^2 a} + \frac{\omega_p^2}{\cosh^2 a} = 1 \tag{14-80}$$

This is the equation of an ellipse with its major axis along the $j\omega_p$ axis having a major semiaxis of length $\cosh a$ and a minor semiaxis of length $\sinh a$. These features are illustrated in Fig. 14-16.

All computations thus far have been for unit transfer function magnitude and unit cutoff frequency. In a practical case where the cutoff frequency may be kilocycles/sec and the magnitude is large (as determined by the gain of the tube), the magnitude and frequency can be scaled by multiplying all normalized frequencies by ω_n (usually taken as the half-power frequency) and the magnitude which is the maximum gain of the system under study. In many cases, there will be advantage in leaving the scaling to the last step in the design, because of the relative ease of working with small numbers. We turn next to the actual amplifier networks used to realize these stagger-tuned characteristics.

Low-Pass Filter Amplifier. An examination of the complex plane for the Chebyshev case shows that only poles are present; no zeros have been required. There are networks with this transfer characteristic, or a network can be constructed by using results that have already been found. Consider the shunt peaked amplifier network studied in Art. 14-1. The voltage ratio transfer function for this network has, for the high Q case, two poles (conjugate pair) and one zero. The location of the poles in the complex plane can be controlled by controlling the Q of the network. This is accomplished in practice by adjusting the inductance by means of a tuning slug (alternately, the resistance might be varied). But we still have a zero, and that zero cannot be ignored. Keep this problem in mind, and let us turn our attention to another network shown in Fig. 14-17. If the tube has a high plate resistance (and acts as a current source), the transfer function for the voltage ratio is

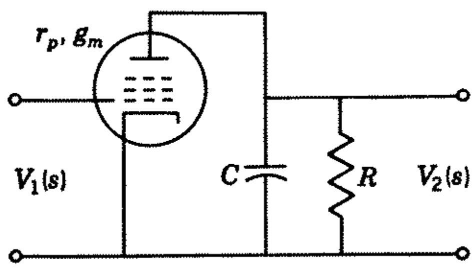


Fig. 14-17. *RC* amplifier network.

$$\frac{V_2}{V_1} = -g_m Z(s) \tag{14-81}$$

where $Z(s)$ is the impedance of the plate circuit (termination). For the network shown, the impedance has the value

$$Z(s) = \frac{1}{C}\frac{1}{s + 1/RC} \tag{14-82}$$

and the voltage transfer function is, in terms of $p = s/\omega_n$

$$\frac{V_2(p)}{V_1(p)} = -\frac{g_m}{\omega_n C}\left(\frac{1}{p + 1/\omega_n RC}\right) \tag{14-83}$$

The network has a pole on the negative real axis of the s plane; its position can be adjusted by adjusting either R or C. Return now to our problem of the zero: Here we have a network with a single pole. Might this pole be used to cancel the unwanted zero? If the two stages can be connected together in cascade such that the two networks are isolated (and this *is* the case because of the isolating action of the amplifiers), the answer is yes. The cascade connection has a transfer function

$$G(p) = G_1(p)G_2(p) \tag{14-84}$$

The two appropriate transfer functions are Eqs. 14-6 and 14-83; the over-all transfer function is

$$\frac{V_{out}(p)}{V_{in}(p)} = \frac{g_{m_1}g_{m_2}}{\omega_n{}^2C_1C_2}\left[\frac{\left(p + \frac{1}{Q}\right)}{(p^2 + p/Q + 1)(p + 1/\omega_nRC)}\right] \tag{14-85}$$

where $p = s/\omega_n$ is normalized frequency. If we set

$$\frac{1}{Q} = \frac{1}{\omega_nRC} \tag{14-86}$$

the pole and zero cancel, and the transfer function has only two poles defined in terms of Q. This cancellation is illustrated in Fig. 14-18.*

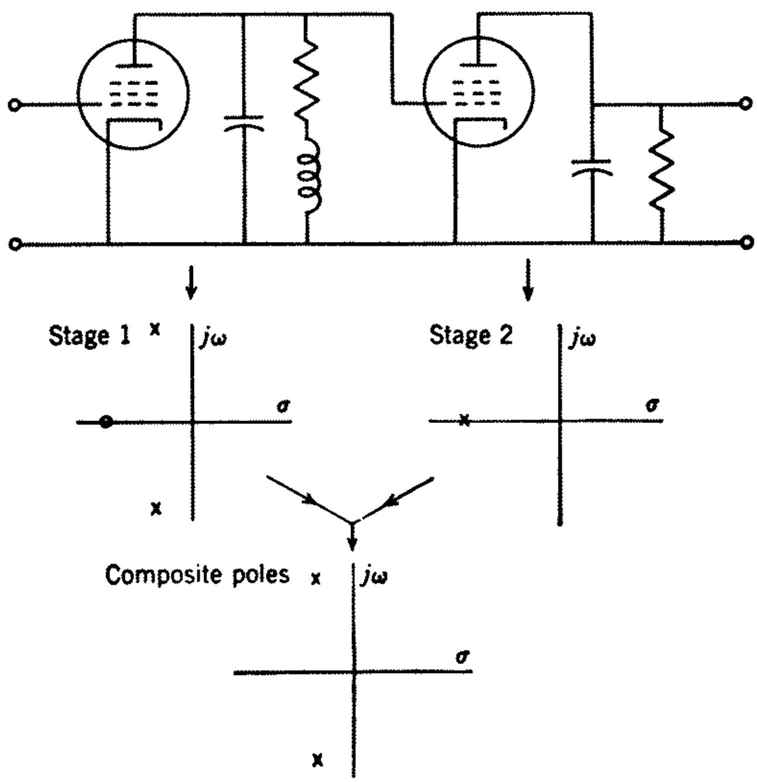


Fig. 14-18. Use of two stages of amplification to give conjugate poles (no zeros).

* Another amplifier network with the same characteristics is shown in Prob. 14-6. For a discussion of stagger-tuned amplifier design, see McWhorter and Pettit, *Proc. IRE*, **43**, 923 (1955).

This basic network unit shown in Fig. 14-18 may now be used to build a pole configuration to give either maximally flat or Chebyshev frequency response. This is illustrated in Fig. 14-19, where the networks are adjusted by varying parameters to give maximally flat response. These poles might be adjusted to give a Chebyshev frequency response, using the same basic building blocks in the form of the networks of Fig. 14-18.

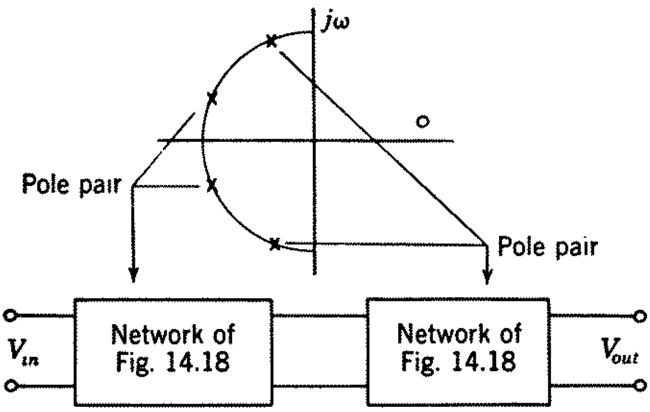


Fig. 14-19. Cascade connection of networks to give maximally flat response ($n = 4$).

Band-Pass Filter Amplifier. Since the poles and zeros can be moved to any position in the s plane by merely adjusting parameters (of course, within the range of practical adjustment: the network Q can be made only so large with practical elements), the poles can be adjusted to give band-pass filter characteristics, using the same basic ideas as in the low-pass filter case. As the building block in this example, we will make use of the shunt peaked network which is the first half of the network of Fig. 14-18. The transfer function of this network amplifier was derived as Eq. 14-6 and has two poles and one zero as

$$\frac{V_2(p)}{V_1(p)} = -\frac{g_m}{\omega_n C}\left[\frac{(p - p_1)}{(p - p_a)(p - p_a{}^*)}\right] \tag{14-87}$$

where, as in the low-pass filter case, $p = s/\omega_n$. Suppose that four such networks are cascaded and the poles are adjusted for the configuration shown in Fig. 14-20. This time we *have* zeros. The effect of the zeros must be taken into account. The over-all transfer function for the four stages is (using subscript numbers to designate the stage)

$$\frac{V_{out}(p)}{V_{in}(p)} = \frac{g_{m_1}g_{m_2}g_{m_3}g_{m_4}}{\omega_{n_1}\omega_{n_2}\omega_{n_3}\omega_{n_4}C_1C_2C_3C_4} \times \frac{(p - p_{11})(p - p_{12})}{(p - p_{a_1})(p - p_{a_1}{}^*)(p - p_{a_2})(p - p_{a_2}{}^*)} \times \frac{(p - p_{13})(p - p_{14})}{(p - p_{a_3})(p - p_{a_3}{}^*)(p - p_{a_4})(p - p_{a_4}{}^*)} \tag{14-88}$$

The cascaded system thus has eight poles and four zeros. The position of the poles and zeros can be fixed by adjusting the circuit Q for each stage (fixing the poles also fixes the zeros, however). For a maximally flat response, suppose that the poles are assigned positions as shown in Fig. 14-20(b). One stage contributes one pole in the upper half plane, one pole in the lower half plane, and one zero. The poles have positions

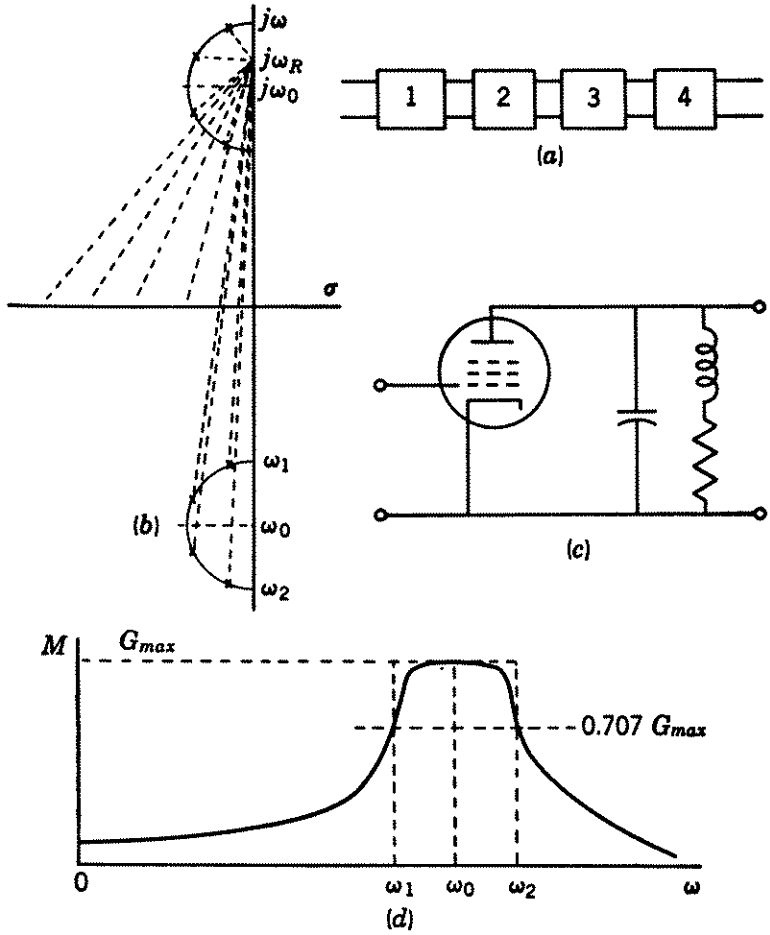


Fig. 14-20. Maximally flat band-pass filter characteristics (four stages of amplifier networks): (a) four cascaded stages; (b) pole-zero configuration for maximally flat response; (c) basic unit of each stage (shunt peaking network); (d) maximally flat response curve.

on the periphery of a circle having a diameter $(\omega_{p1} - \omega_{p2})$, where ω_{p2} and ω_{p1} are half-power frequencies for the maximally flat case. For the sinusoidal steady state, when $p = j\omega_p$, each frequency term in Eq. 14-88 can be represented by a phasor as shown in Fig. 14-20(b). For frequencies from ω_{p1} to ω_{p2}, the poles in the upper half plane have the greatest effect on the transfer function magnitude. The phasors from these poles are changing rapidly in magnitude, while the phasors from the poles in the lower half plane and the zeros are changing slowly. In many practical designs, the *mid-band* frequency ω_0 is high and the band of pass frequencies (ω_{p2} to ω_{p1}) is small. Under these conditions the last equation can be written in *approximate* form as

$$\frac{V_{out}(p)}{V_{in}(p)} = K \frac{1}{(p - p_{a_1})(p - p_{a_2})(p - p_{a_3})(p - p_{a_4})} \tag{14-89}$$

where K is only *approximately* constant and is given as

$$K = \frac{g_{m_1}g_{m_2}g_{m_3}g_{m_4}}{\omega_{n_1}\omega_{n_2}\omega_{n_3}\omega_{n_4}C_1C_2C_3C_4} \frac{(p - p_{11})(p - p_{12})(p - p_{13})(p - p_{14})}{(p - p_{a_1}{}^*)(p - p_{a_2}{}^*)(p - p_{a_3}{}^*)(p - p_{a_4}{}^*)} \tag{14-90}$$

This equation for the voltage ratio is of the form required for the two types of responses that have been studied. The poles of this equation can be adjusted to give either maximally flat (Butterworth) frequency response as illustrated *or* equal ripple (Chebyshev) frequency response. The response for the maximally flat case is shown in Fig. 14-20(d). The band-pass features of this response are evident. Such response characteristics are required in such applications as intermediate frequency amplifiers in superheterodyne receivers.

The usual specifications for design are: (a) the mid-band frequency ω_0, (b) the over-all bandwidth, (c) the rate of decrease of the frequency response outside of the pass band, and (d) the over-all gain. From these specifications, n (the number of stages) is determined.

For the *maximally flat case* (Butterworth), the pole configuration is then selected from the chart of Fig. 14-9 or from corresponding equations. The parameters of the actual network are used and then adjusted to give the required real and imaginary part for each pole.

For the *equal ripple* (Chebyshev) case, the ripple width is usually specified in addition to the list given above. From these specifications,

(1) Determine δ from Eq. 14-52 as

$$\delta = 10 \log_{10} (1 + \epsilon) \qquad \text{decibels} \tag{14-91}$$

(2) Calculate the factor a from Eq. 14-54.

$$a = \frac{1}{n} \sinh^{-1} \frac{1}{\sqrt{\epsilon}} \tag{14-92}$$

(3) Compute cosh a and tanh a. Draw a circle of radius (cosh a), equal to the *bandwidth*, with a center at the *mid-band frequency*. Locate the poles on this circle as in the maximally flat case (given above). Multiply the *real part* of each pole by tanh a. This gives the pole locations for the Chebyshev case.

(4) Adjust the parameters of the network being used to give these pole locations.

An example will illustrate the design procedure just outlined. Three stages of amplification with shunt peaking are to be stagger-tuned with an equal-ripple characteristic. The mid-band frequency is to be 5.0 megacycles/sec, and the bandwidth to the half-power frequencies is to be 500 kilocycles/sec. The ripple width is specified as 1.0 db. We are required to find the mid-band frequency and the Q for each of the three stages. The parameters R, L, and C of the shunt peaking network can in turn be found if one of the three is fixed (as it often is in practice—for example, the interelectrode and wiring capacitance).

Following the steps just given we first find ϵ from the equation

$$\delta = 10 \log_{10} (1 + \epsilon) \qquad \text{decibels} \tag{14-93}$$

Since $\delta = 1$ db, we have

$$1 = 10 \log_{10} (1 + \epsilon) \quad \text{or} \quad \epsilon = 10^{0.1} - 1 \tag{14-94}$$

and

$$\epsilon = 0.259 \tag{14-95}$$

We next compute the factor a from Eq. 14-92 as

$$a = \frac{1}{n} \sinh^{-1} \frac{1}{\sqrt{\epsilon}} = \frac{1}{3} \sinh^{-1} 1.963 = 0.475 \tag{14-96}$$

For this value of the factor a, the hyperbolic tangent has the value

$$\tanh a = 0.442 \tag{14-97}$$

We next turn our attention on the pole-zero configuration—in particular to the location of the poles. Figure 14-9 shows the pole locations for $n = 3$ as occurring at $\theta = 0°$ and $\pm 60°$ with respect to the negative real axis. These pole locations for a bandwidth of 0.500 megacycle are shown in Fig. 14-21. With respect to the midband frequency, the poles

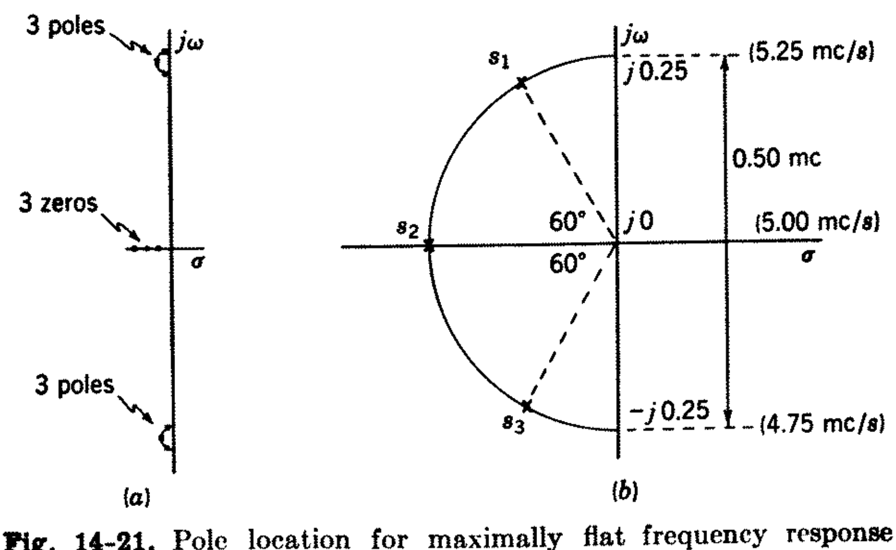


Fig. 14-21. Pole location for maximally flat frequency response: (a) full scale (approximately); (b) region of interest.

have the locations:

$$\begin{aligned} s_1 &= (-0.5 + j0.866)0.25, \\ s_2 &= (-1.0 + j0)0.25, \\ s_3 &= (-0.5 - j0.866)0.25 \end{aligned} \tag{14-98}$$

The actual locations in the s plane are shown in Fig. 14-21 (hardly to scale even so). For the equal ripple response, the real part of the pole location is multiplied by ($\tanh a = 0.442$). The new locations for the poles then become

$$\begin{aligned} s_1' &= (-0.221 + j0.866)0.25, \\ s_2' &= (-0.442 + j0)0.25, \\ s_3' &= (-0.221 - j0.866)0.25 \end{aligned} \tag{14-99}$$

where all measurements indicated by these equations are made with respect to 5.0 megacycles/sec.

For the high-Q case (corresponding to $\zeta \ll 1$), the bandwidth B is found from Eqs. 11-86 and 11-87 as

$$B = 2\zeta\omega_n \tag{14-100}$$

and Q is defined by Eq. 11-77 as

$$Q = \frac{1}{2\zeta} \tag{14-101}$$

Combining these two equations, we have

$$Q = \frac{\omega_n}{B} = \frac{f_n}{B_f} \tag{14-102}$$

where f_n is the natural undamped frequency in cycles per second, and B_f is the bandwidth in cycles per second (the common 2π term cancels). In Fig. 14-22(b) the distance from the $j\omega$ axis to a pole location is

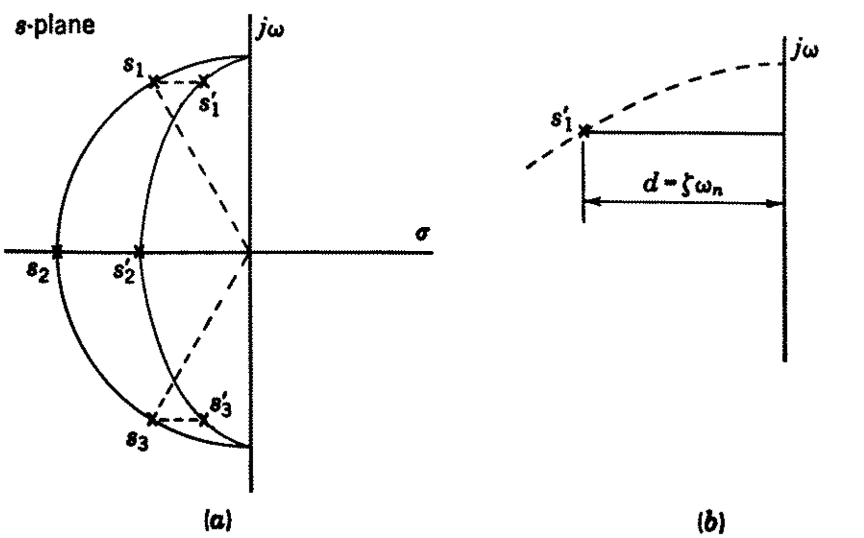


Fig. 14-22. (a) New pole location for the equal ripple case; (b) s_1' pole enlarged to show distance to $j\omega$ axis.

marked d. By Eq. 11-70 this real part of the complex pole has the value $(\zeta\omega_n)$. But by Eq. 14-100, the bandwidth is given as $B = 2\zeta\omega_n$. From this it is seen that

$$d = \frac{B}{2} \tag{14-103}$$

or that the distance d is the "half bandwidth."

We next make use of these last two identities to design the stagger-tuned amplifier corresponding to the computed pole configuration, under the assumption that the zeros have negligible effect. Assume that stage 1 will be made to correspond to the pole s_1' and its conjugate, stage 2 to s_2' and its conjugate, and stage 3 to pole s_3' and its conjugate.

For stage 1 (s_1'),

$$f_n = 5.0 + (0.866 \times 0.25) = 5.217 \quad \text{megacycles/sec}$$

$$B_f = 2 \times 0.221 \times 0.25 = 0.110 \quad \text{megacycle/sec}$$

$$Q = f_n/B_f = 47.2$$

For stage 2 (s_2'),

$$f_n = 5.0 \quad \text{megacycles/sec}$$

$$B_f = 2 \times 0.442 \times 0.25 = 0.221 \quad \text{megacycle/sec}$$

$$Q = 22.6$$

For stage 3 (s_3'),

$$f_n = 5.0 - 0.866 \times 0.25 = 4.783 \quad \text{megacycles/sec}$$

$$B_f = 2 \times 0.221 \times 0.25 = 0.110 \quad \text{megacycle/sec}$$

$$Q = 43.2$$

The amplifier-network realization of the required equal-ripple characteristic is shown in Fig. 14-23. If required, the circuit parameters

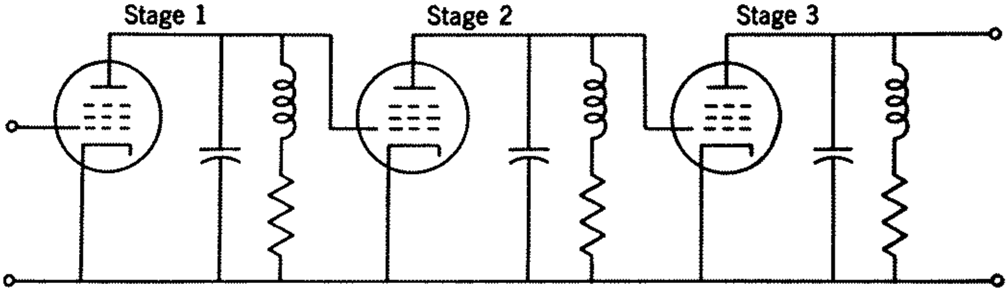


Fig. 14-23. Equal-ripple realization.

	Stage 1	Stage 2	Stage 3
f_n, mc/sec	5.217	5.00	4.783
B_f, mc/sec	0.110	0.221	0.110
Q	47.2	22.6	43.2

R, L, and C can be found, if one is assumed fixed. Figure 14-24 shows how the characteristics of the individual stages are combined by the tandem connection to give the equal-ripple frequency response.

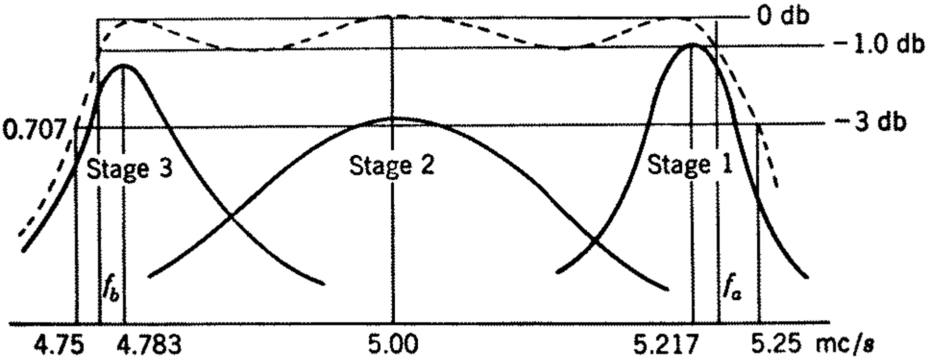


Fig. 14-24. Combined frequency characteristic.

The frequencies marked f_a and f_b are frequencies corresponding to the end of the ripple band. In most of the preceding discussion this frequency has been normalized to unity. By the specifications of this problem, it was more convenient to work with the 3-db point frequencies (4.75 and 5.25 megacycles/sec). The frequencies f_a and f_b are given as

$$f_b, f_a = 500 \pm \frac{0.25}{\cosh a} \quad \text{megacycles/sec} \tag{14-104}$$

$$= 4.776 \quad \text{megacycles/sec}, \qquad 5.224 \quad \text{megacycles/sec} \tag{14-105}$$

One advantage of the equal-ripple case over the maximally flat case is that the gain is higher for the equal-ripple configuration. This follows because the poles are closer to the $j\omega$ axis in the equal-ripple case. Referring to Fig. 14-22(a), the gain at the mid-band frequency (5.00 megacycles/sec) may be found in terms of phasor lengths. The ratio of gains of the equal-ripple case to the maximally flat case is given as

$$\text{ratio of gains} = \frac{|s_1| \cdot |s_2| \cdot |s_3| \cdot \text{(other pole and zero distances)}}{|s_1'| \cdot |s_2'| \cdot |s_3'| \cdot \text{(the same pole and zero distances)}} \tag{14-106}$$

The distances to each of the other poles and zeros is approximately the same for both cases. The three significant distances may be found by converting the complex numbers of Eqs. 14-98 and 14-99 to polar form. For this particular problem

$$\frac{\text{equal ripple gain}}{\text{maximally flat gain}} = \frac{(0.25)^3}{(0.223)^2 \times 0.1105} = 2.8 \tag{14-107}$$

at 5.00 megacycles/sec.

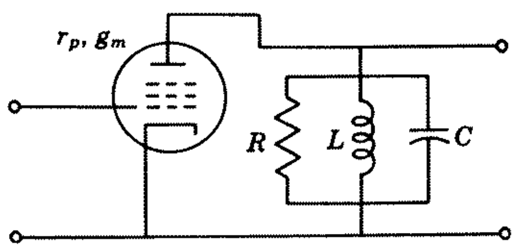
Offsetting this gain advantage is the disadvantage that the phase angle of the output with respect to the input as a function of frequency is not so linear in the equal-ripple case as in the maximally flat case. This nonlinear characteristic restricts the application of equal-ripple designed amplifiers in such applications as television.

FURTHER READING

The best known treatment of stagger tuning is that in Valley and Wallman, *Vacuum Tube Amplifiers*, Vol. 18 of the Radiation Laboratory Series (McGraw-Hill Book Co., Inc., New York, 1948) Chap. 4. This account is based on an MIT Radiation Laboratory report by Henry Wallman issued in 1944. Two earlier papers on stagger tuning are Butterworth, "On the theory of filter amplifiers," *Wireless Engineer*, **7**, 536 (1930) and V. D. Landon, "Cascade amplifiers with maximal flatness," *RCA Rev.*, **5**, 347 (1941). Further discussions of stagger-tuned amplifiers are to be found in T. L. Martin, Jr., *Electronic Circuits* (Prentice-Hall, Inc., New York, 1955), LePage and Seely, *General Network Analysis* (McGraw-Hill Book Co., Inc., New York, 1952), pp. 238–256, and R. F. Baum, "Design of broad-band IF amplifiers," *Jour. Appl. Phys.*, **17**, 519 and 721 (1946). An extensive bibliography on the subject is given by H. A. Wheeler, "The potential analog applied to the synthesis of stagger-tuned filters," *Proc. IRE*, **CT-2**, 86 (1955).

PROBLEMS

14-1. For the amplifier network shown in the accompanying figure, find the voltage ratio transfer function. The vacuum tube has high

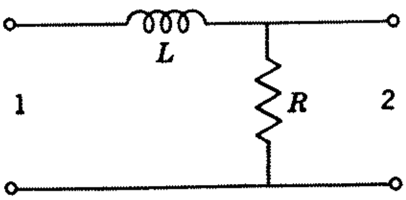


Prob. 14-1.

plate resistance and can be considered a current source. Show a typical pole-zero configuration if the components have values to correspond to the oscillatory case. Networks of this type find application in cascade-connected amplifiers.

14-2. The amplifier network shown in Fig. 14-3 is to be designed according to the following specifications. The tube used in a 6AK5, for which a g_m of 4500 μmho may be assumed. The capacitance of the amplifier stage is the interstage capacitance *only* (including the tube, socket, and wiring) which has a value of 12 $\mu\mu$f. If the circuit Q has a value of 75 and the frequency at resonance is 12.00 megacycles/sec, determine: (a) the values of R and L, (b) the maximum gain of the amplifier stage, and (c) the bandwidth. *Answer.* (a) 14.7 ohms, 14.7 μh, (b) 375, (c) 160 kc.

14-3. The network is to be used as a voltage coupling network. (a) For the voltage ratio transfer function to have a maximally flat



Prob. 14-3.

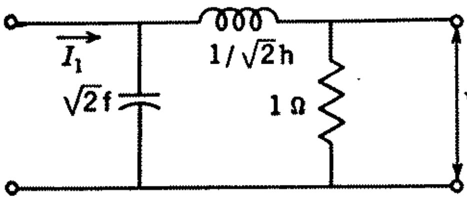
frequency characteristic of the form, $1/\sqrt{1 + \omega^{2n}}$, where ω is in radians/sec, what must be the relationship between R and L? (b) Determine a value for R and for L such that the half-power frequency (of the maximally flat characteristic) is 10 radians/sec.

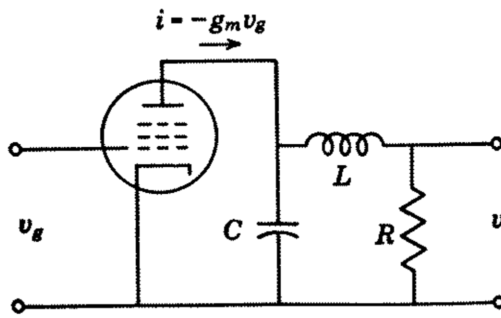
14-4. Plot the function $1/\sqrt{1 + \omega^{2n}}$ for $0 \leqq \omega \leqq 4$ for $n = 1, 2,$ and 3.

14-5. The network shown in the figure is known as a second-order *Butterworth filter.* Find the magnitude of the transfer impedance defined as

$$|Z_{12}(j\omega)| = \left| \frac{V_2(j\omega)}{I_1(j\omega)} \right|$$

and show that it has maximally flat frequency characteristics. Sketch the frequency response curve and identify significant points such as the half-power frequencies, etc.





Prob. 14-5. **Prob. 14-6.**

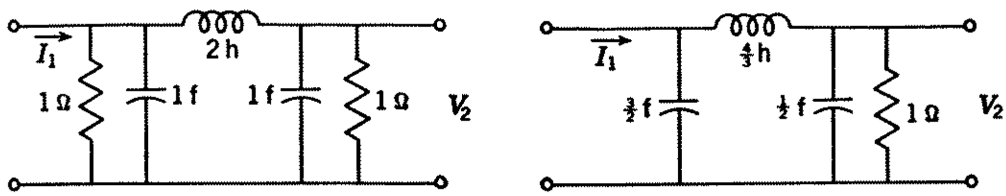
14-6. In this problem, the second-order Butterworth network of Prob. 14-5 is to be used in the amplifier network shown. It is desired that the frequency characteristic be maximally flat, but for this network two changes must be made: (1) the load resistance R must be 50 ohms (purely resistive) and (2) the half-power frequency must be

250 kilocycles/sec. Determine the values of L and C to meet these requirements. *Answer.* $L = 1.414 \times 10^{-4}$ henry, $C = 0.113\ \mu$f.

14-7. The network shown in the figure is a third-order *Butterworth filter* (that might be used in a pentode circuit as shown in the last problem). For this network, show that the magnitude of the transfer impedance, $|Z_{12}(j\omega)|$ has a maximally flat frequency characteristic and that the transfer impedance has the form

$$Z_{12}(s) = \frac{\frac{1}{2}}{(s - s_1)(s - s_2)(s - s_3)}$$

where $s_1 = -1$, $s_2 = -\frac{1}{2} + j\sqrt{3}/2$, and $s_3 = s_2^*$.



Prob. 14-7. **Prob. 14-8.**

14-8. Show that the transfer impedance, $Z_{12}(s) = V_2(s)/I_1(s)$, of the network shown in the figure differs from that given in Prob. 14-7 only by a constant multiplier. Determine the constant.

14-9. Plot the function $1/\sqrt{1 + \epsilon C_n^2(\omega)}$ where $C_n(\omega)$ is the nth order Chebyshev polynomial defined by Eq. 14-38 if $\epsilon = 0.25$ for $0 \leqq \omega \leqq 4$ for $n = 1$, 2, and 3.

14-10. Write the 8th order Chebyshev polynomial, $C_8(\omega)$, in the form of a polynomial.

14-11. Write the 9th order Chebyshev polynomial, $C_9(\omega)$, in the form of a polynomial.

14-12. An equal-ripple frequency response of the form given by Eq. 14-49 has the following fixed parameters: $n = 5$ and $\epsilon = 0.1$. For this response, determine: (a) the maximum value of $|G(j\omega)|$, (b) the minimum value of $|G(j\omega)|$ in the pass band, (c) the ripple width δ in decibels, (d) the half-power frequency. *Answers.* (a) 1, (b) 0.953, (c) 0.414 db, (d) 1.07.

14-13. Repeat Prob. 14-12 for $n = 4$ and $\epsilon = 0.2$. *Answers.* (a) 1, (b) 0.913, (c) 0.79 db, (d) 1.076.

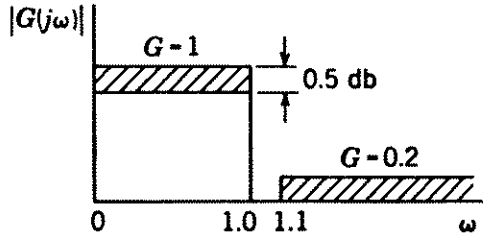
14-14. It is required that a system having the frequency-response equation given by Eq. 14-49 have a half-power frequency at $\omega = 1.10$. If the ripple width is limited to 0.5 db, what is the minimum value n may have?

14-15. The constant factor a in Eq. 14-54 is assumed fixed in value. Under this condition, derive a relationship between n and δ. Sketch δ and a function of n.

14-16. Refer to Fig. 14-16 plotted for $n = 5$. For this case, give the locations of the poles for the Chebyshev case if $a = \frac{1}{2}$.

14-17. A maximally flat low-pass filter-amplifier with $n = 3$ has a half-power frequency of $\omega = 1.00$. An equal ripple (Chebyshev) filter is to be designed with the same half-power frequency, the same value of n, and a ripple width of 0.5 db. Determine the upper and lower frequency limits of the pass band (the frequency at which the equal ripple ends).

14-18. A filter-amplifier is to be designed on the Chebyshev basis. It is specified that the magnitude of the voltage transfer ratio must be



Prob. 14-18.

inside the crosshatch sections. (No specifications are given for the range $\omega = 1.0$ to 1.1.) What is the required value of n?

14-19. The network of Fig. 14-18 is to be used to give a maximally flat frequency characteristic for a low-pass filter amplifier. The amplifier is to have a bandwidth of 100 kilocycles/sec (measured from $\omega = 0$ to the half-power frequency). (a) Design an amplifier to meet these requirements. Select a vacuum tube for each stage. Specify all component values. (b) Compute the maximum gain for the composite system. (c) Plot the frequency response (gain vs. ω) for the first stage. Repeat for the second stage. Plot the frequency response for the composite system.

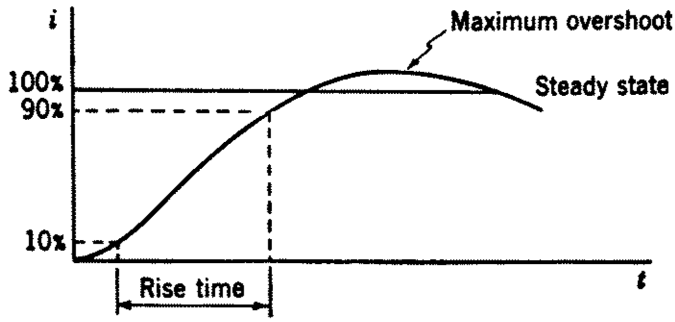
14-20. A band-pass amplifier network is to be designed to the following specifications: the mid-band frequency is 10.0 megacycles/sec, the bandwidth to the half-power frequencies is to be 500 kilocycles/sec the ripple is limited to 0.5 db. Based on gain requirements, a decision is made to use 3 stages. Stagger tune these three stages of amplification to give a Chebyshev equal ripple characteristic, using the basic network shown in Fig. 14-23. Determine f_n, B_f, and Q for each stage. Determine the ratio of the maximally flat gain to the equal ripple gain at the mid-band frequency.

14-21. Design a band-pass amplifier network having the mid-band frequency and bandwidth given in Prob. 14-20 but for the maximally flat case.

14-22. In this problem, we will investigate the transient response of amplifier networks designed on the equal ripple and maximally flat basis. For this problem, let $n = 2$ corresponding to two poles in the s plane of the form,

$$Z_{12}(s) = \frac{V_2(s)}{I_1(s)} = \frac{1}{(s - s_1)(s - s_2)}$$

where s_1 and s_2 have different values for the two cases (equal ripple and maximally flat). Two time-domain quantities of interest are the



Prob. 14-22.

rise time and *overshoot* resulting from a step-function input. One definition of rise time is the time interval, measured in seconds between the times the response, that has 10% and 90% of the final (steady-state) value. The overshoot is defined as

$$\frac{(\text{maximum transient value}) - (\text{final value})}{(\text{final value})} \times 100\%$$

If the driving-force current is a step function; that is, $i(t) = u(t)$, determine the rise time and overshoot for (a) a maximally flat designed amplifier network, and (b) an equal-ripple design amplifier network. Compare these quantities for the two amplifier networks. Which amplifier seems to have superior transient response? (Check point: the overshoot for the maximally flat amplifier is about 4%.)

CHAPTER 15

BLOCK DIAGRAMS

Block diagrams are widely used by engineers as a shorthand symbolism in describing a system. The block represents components: entirely, in part, or in combination. Lines given direction by arrows indicate the sequence of operations that take place in the system. Block diagrams, as we shall use them, are not pictorial representations of components. Several blocks may be used to represent a single component *or* one block may represent a complex mechanism such as a digital computer. Indeed, the blocks may represent the solution of a mathematical equation with no direct physical significance. The block diagram provides a method for representing a system in such a way as to express a cause-and-effect relationship between the input and the output. Block diagrams are also referred to by the descriptive name *signal flow diagrams*, where we define the signal as any causal quantity intentionally introduced into a system (in contrast to noise).

As we study block diagrams, we should keep in mind the underlying objectives in their use:

(1) Block diagrams are easier to draw than detailed schematic diagrams. The block diagram is shorthand notation.
(2) Block diagrams, together with the transfer function, indicate the dynamic behavior of the system. They tell the engineer not only "what it is" but "what it does."
(3) The construction of a block diagram is one step in system analysis. Once constructed, the blocks may be manipulated by a system of algebra to find a simplified equivalent block diagram. This reduction in complexity is equivalent to algebraic reduction of system equations.

15-1. Basic operations for block diagrams

Let the input to a system be designated as v_1 and the output as v_2 (where v is any variable such as voltage, current, etc.). The relationship between the input and output may be expressed in terms of an operator G, such that

Fig. 15-1. Block diagram.

$$v_2 = Gv_1 \qquad (15\text{-}1)$$

We define the block diagram shown in Fig. 15-1 to be equivalent to this equation. The function G may be any operator or any linear combination of operators. For example, suppose that G represents differentiation with respect to time. This means that the input $v_1(t)$ is differentiated as $dv_1(t)/dt$ to be equal to the output, $v_2(t)$. The block diagram equivalent to this statement is shown in Fig. 15-2(a). If the operation to be performed is multiplication of a transform function $V_1(s)$ by a transfer function K_1s, then the block diagram and algebraic equivalent are as shown in Fig. 15-2(b). A linear combination of operations,

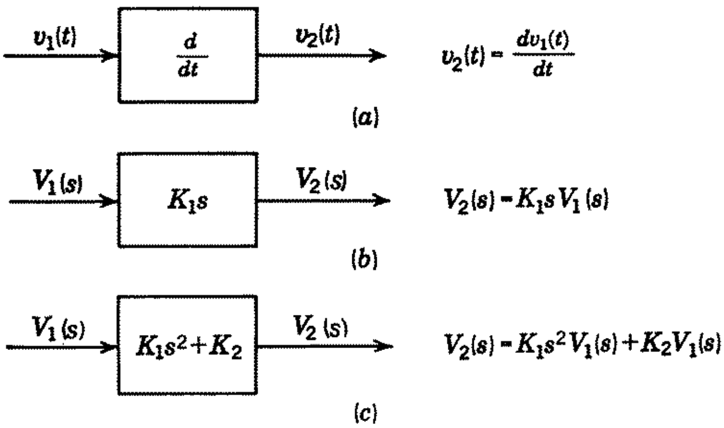


Fig. 15-2. Block diagram and equation equivalents.

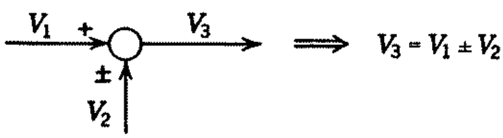


Fig. 15-3. Summing point symbol.

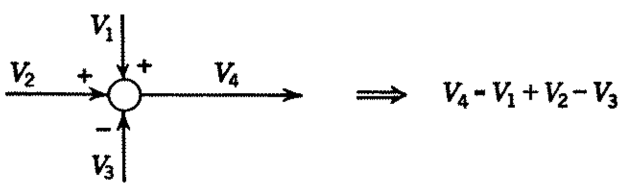


Fig. 15-4. Summing point.

expressed as $G(s) = K_1s^2 + K_2$, is shown in Fig. 15-2(c). Each term in the expression operates on $V_1(s)$ independently as illustrated.

Another basic symbol is shown in Fig. 15-3. The circle with arrows pointing into it is a *summing point.* Quantities entering the circle or summing point are either added or subtracted according to directions in the form of a + sign or a − sign on the arrow. If the sign on the arrow marked V_1 is omitted in a diagram, it is presumed to be positive. In constructing a block diagram, however, it is wise to use a sign for each input whenever there might be doubt. This is very necessary in the case when more than two inputs enter a summing point as illustrated in Fig. 15-4.

When a single line on a block diagram separates into two or more lines leaving the same point, all lines carry the same unaltered variable. Such a point is called a *pickoff point.* A pickoff point in a system is illustrated in Fig. 15-5.

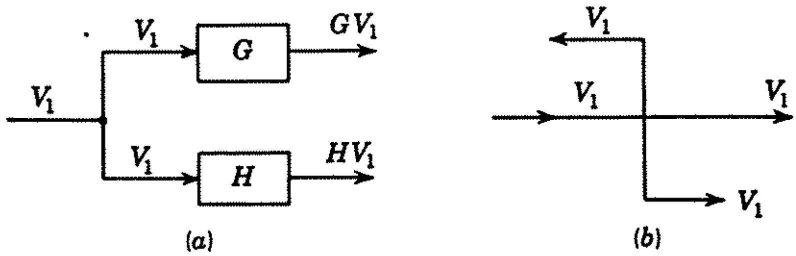


Fig. 15-5. Pickoff points.

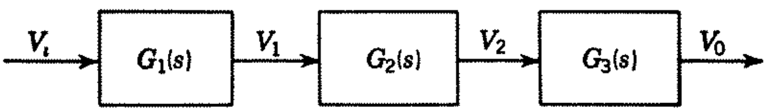


Fig. 15-6. Tandem system.

Suppose that several blocks are connected in *cascade* (or *tandem*) as shown in Fig. 15-6. The transfer function for each block is given on the figure, and is the ratio of the output to the input for each of the three cases. Since

$$\frac{V_2}{V_1} \times \frac{V_3}{V_2} \times \frac{V_4}{V_3} = \frac{V_4}{V_1} \tag{15-2}$$

the total (or over-all) transfer function is

$$G_t = G_1G_2G_3 \tag{15-3}$$

This equality depends on there being "no loading" between blocks, as will be discussed in a later section. Thus blocks connected in tandem

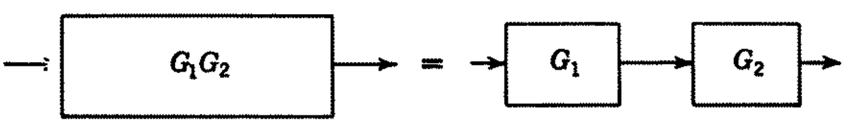


Fig. 15-7. Expansion of blocks.

may be combined into a single equivalent block by *multiplying* the transfer functions together for the equivalent transfer function. Likewise a system may be expanded into several parts as illustrated in Fig. 15-7.

15-2. Block diagrams for electrical elements

Instantaneous value of voltage and current for the electrical elements are related by the following equations.

$$\begin{aligned} &\text{resistance:} \quad v_R(t) = Ri_R(t) \quad \text{or} \quad i_R(t) = Gv_R(t), \quad G = 1/R \\ &\text{inductance:} \quad v_L(t) = L\, di_L(t)/dt \quad \text{or} \quad i_L(t) = 1/L \int v_L(t)\, dt \\ &\text{capacitance:} \quad v_C(t) = 1/C \int i_C(t)\, dt \quad \text{or} \quad i_C(t) = C\, dv_C(t)/dt \end{aligned} \tag{15-4}$$

These equations are equivalent to the block diagrams shown in Fig. 15-8.

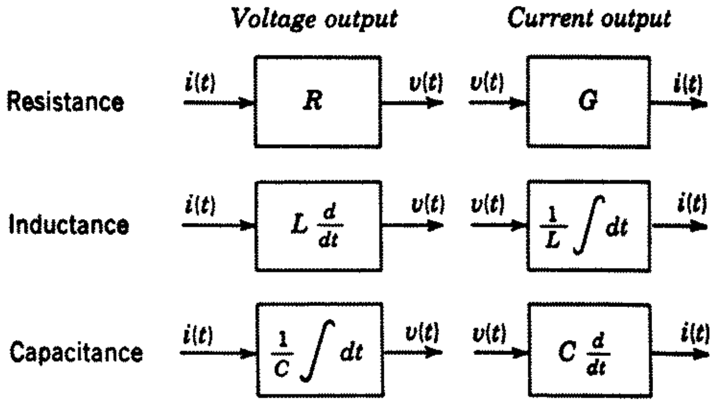


Fig. 15-8. Block diagrams for electrical elements.

To construct a similar chart for the transform voltages and currents, the Laplace transformation of each expression in Eq. 15-4 will be taken with initial conditions ignored. The resulting equations are:

$$\begin{aligned} &\text{resistance:} \quad V_R(s) = RI_R(s) \quad \text{or} \quad I_R(s) = GV_R(s), \quad G = 1/R \\ &\text{inductance:} \quad V_L(s) = LsI_L(s) \quad \text{or} \quad I_L(s) = (1/Ls)V_L(s) \\ &\text{capacitance:} \quad V_C(s) = (1/Cs)I_C(s) \quad \text{or} \quad I_C(s) = CsI_C(s) \end{aligned} \tag{15-5}$$

The same equations apply for the sinusoidal steady state, with s replaced by $(j\omega)$. Since the *transform impedance* is the ratio of the voltage to the current transform (and, similarly, the transform admittance is the ratio of the current to the voltage transform), a chart of block diagrams for electrical elements may be constructed as shown in Fig. 15-9 for transform impedance or admittance. The block diagrams and the expressions for impedance and admittance shown in Fig. 15-9

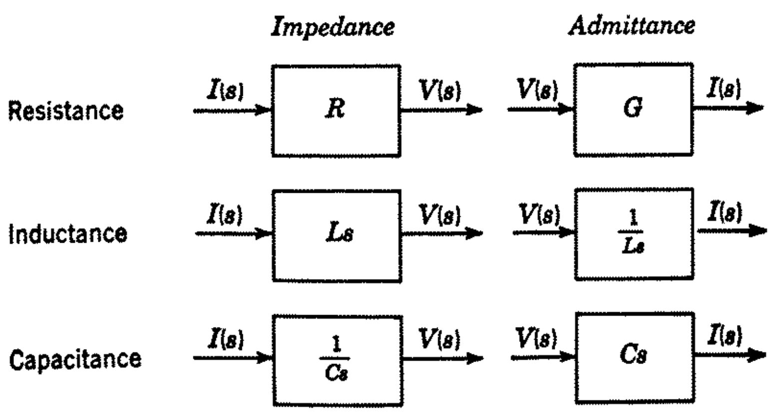


Fig. 15-9. Block diagrams and transfer functions for electrical elements.

reduce to the sinusoidal steady state with the substitution of $(j\omega)$ for s.

The use of this chart of block diagrams for the elements will be illustrated by an example. Consider the electric circuit of Fig. 15-10. The equation relating current and voltage is

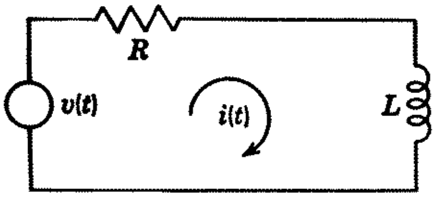


Fig. 15-10. *RL* circuit.

$$v(t) = L\frac{di(t)}{dt} + Ri(t) \qquad (15\text{-}6)$$

or, in terms of voltage and current transforms,

$$V(s) = LsI(s) + RI(s) \qquad (15\text{-}7)$$

if $i(0+) = 0$. This last equation tells us that the applied voltage is equal to the voltage drop across the inductor added to the voltage drop across the resistor; that is,

$$V(s) = V_L(s) + V_R(s) \qquad (15\text{-}8)$$

Equations 15-7 and 15-8 may be written in the following form and order:

$$V(s) - V_R(s) = V_L(s) \qquad (15\text{-}9)$$

$$I_L(s) = \frac{1}{Ls}V_L(s) \qquad (15\text{-}10)$$

$$V_R(s) = RI(s) \qquad (15\text{-}11)$$

There remains one task before we draw the block diagram equivalent of these equations. We must identify the input and output quantities. For this problem, let us identify the input as $V(s)$ and the output as $I(s)$, corresponding to the *excitation* and the *response*. Now, following the pattern suggested by Eq. 15-1 and the equivalent block of Fig. 15-1,

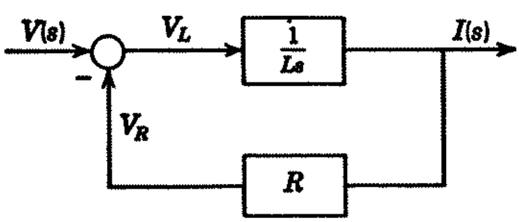


Fig. 15-11. Block diagram of network of Fig. 15-10.

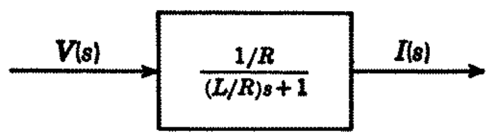


Fig. 15-12. Block diagram for *RL* network.

we arrive at the system of Fig. 15-11. Instead of this diagram with a block for each element in the network, a single block representation may be found by solving Eq. 15-7 for the ratio of output to input, that is, current to voltage. Thus

$$\frac{I(s)}{V(s)} = \frac{1}{Ls + R} = \frac{1/R}{Ls/R + 1} \qquad (15\text{-}12)$$

Thus an equivalent of the block diagram of Fig. 15-11, simplified in form, is shown in Fig. 15-12. The two diagrams for one network suggest that a method can be found for reducing one to the other. Rules for such manipulations will be given in a later section.

15-3. Open-loop and closed-loop block diagram equivalents

An open-loop system is represented in Fig. 15-13 with $E(s)$ as the input (which will later be called the error input) and $V_2(s)$ as the output. This block diagram is equivalent to the algebraic equation

$$V_2(s) = G(s)E(s) \tag{15-13}$$

Suppose that a loop is closed around the system of Fig. 15-13 as shown

Fig. 15-13. Open-loop system.

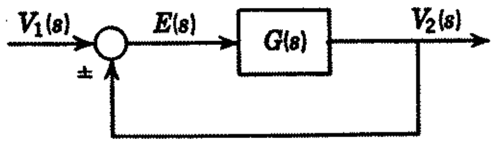


Fig. 15-14. Closed-loop system.

in Fig. 15-14, with a new input identified as $V_1(s)$. The new system is described by two equations:

$$E(s) = V_1(s) \pm V_2(s) \tag{15-14}$$

$$V_2(s) = G(s)E(s) \tag{15-15}$$

The quantity $E(s)$ is now identified as the *error transform*. Eliminating $E(s)$ from the two equations gives

$$\frac{V_2(s)}{V_1(s)} = L(s) = \frac{G(s)}{1 \mp G(s)} \tag{15-16}$$

In this equation $G(s)$ is sometimes spoken of as the *open-loop transfer function* and $L(s)$ as the *closed-loop transfer function*. The last equation relates the open-loop to the closed-loop transfer function.

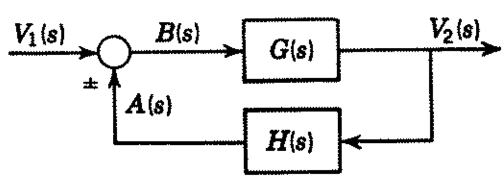


Fig. 15-15. Closed-loop system.

In a more generalized representation of a closed-loop system, an element or combination of elements is included in the *feedback loop* having a transfer function $H(s)$. Such a system is shown in Fig. 15-15.

The equations for this system are

$$B(s) = V_1(s) \pm A(s) \tag{15-17}$$

$$V_2(s) = G(s)B(s) \tag{15-18}$$

$$A(s) = H(s)V_2(s) \tag{15-19}$$

If the two quantities $A(s)$ and $B(s)$ are eliminated from these equations, there results

$$\frac{V_2(s)}{V_1(s)} = L(s) = \frac{G(s)}{1 \mp G(s)H(s)} \tag{15-20}$$

Thus, the two blocks of Fig. 15-16 are equivalent, and one may be substituted for the other.

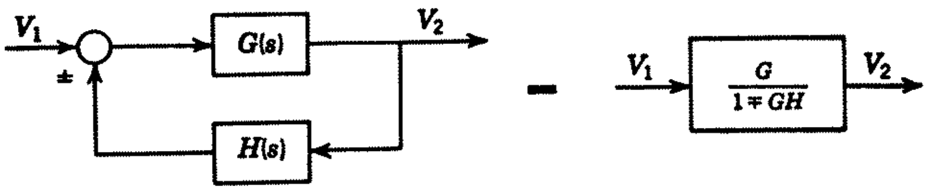


Fig. 15-16. Equivalent open- and closed-loop system.

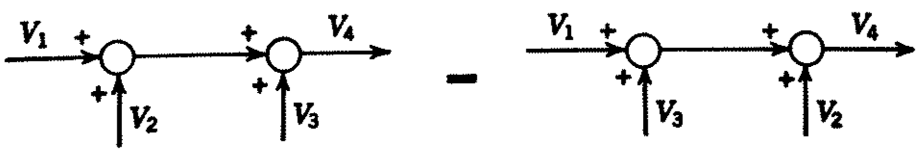
15-4. Block diagram transformations

Manipulations of block diagrams in the last two sections can be generalized into a system of block diagram algebra. The objective in the use of this algebra is simplification; that is, reduction in the number of blocks and the number of loops of a system. A large number of rules for manipulations are given in the literature;* a few of the most important are tabulated below.

Rule 1. Summation points may be interchanged without altering the system. As illustrated in the figure,

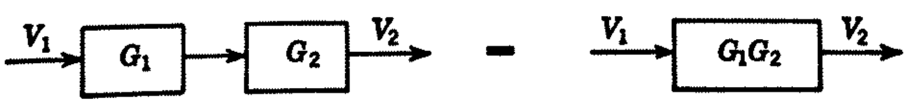
$$V_4 = V_1 + V_2 + V_3 = V_1 + V_3 + V_2$$

(by the associative law of algebra).



Rule 1

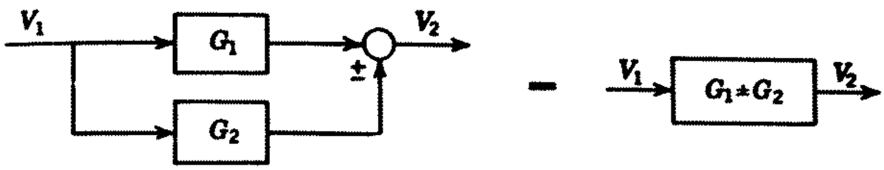
Rule 2. Blocks in tandem are multiplied.



Rule 2

* See references at the end of the chapter.

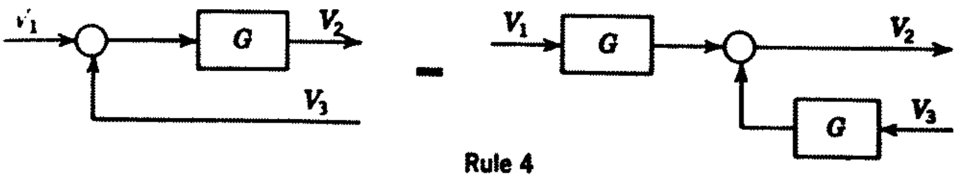
Rule 3. Blocks in parallel are added.



Rule 3

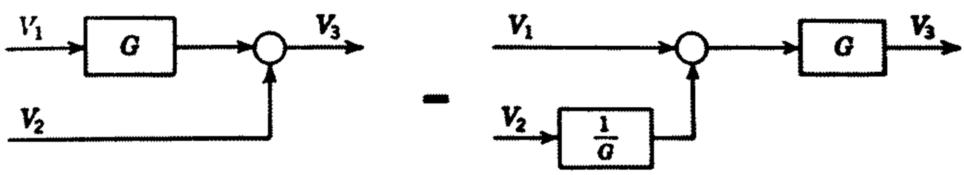
The following four rules indicate the procedure for shifting blocks past summing-points or take-off points.

Rule 4.



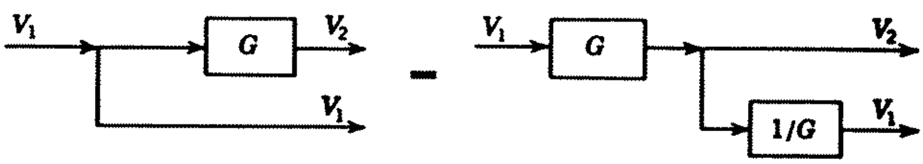
Rule 4

Rule 5.



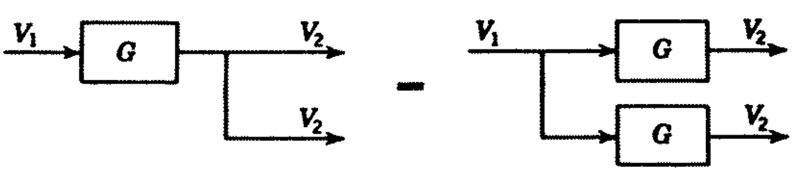
Rule 5

Rule 6.



Rule 6

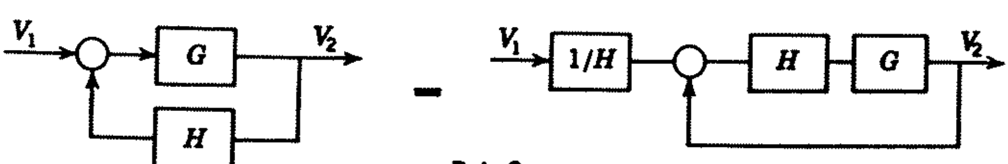
Rule 7.



Rule 7

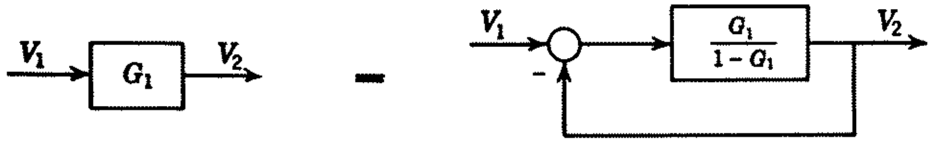
The following rules are applications of the rules given above.

Rule 8. Removing a block in the feedback loop.



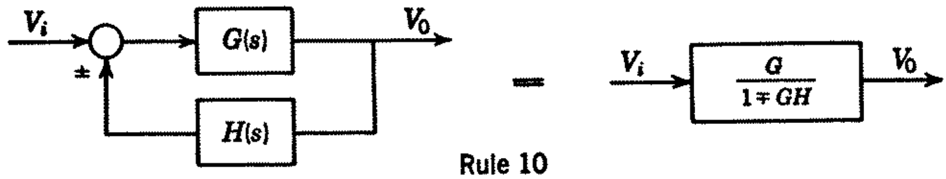
Rule 8

Rule 9. Transformation from open-loop to closed-loop.



Rule 9

Rule 10. Transformation from closed-loop to open-loop.



Rule 10

15-5. Limitations in the block diagram representation of physical systems

Care must be exercised in dividing a system into parts to be represented by block diagrams. This limitation may be illustrated by reference to Fig. 15-17. Two systems are shown, characterized by transfer functions G and H. The two systems can be connected together in

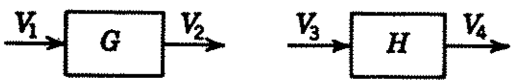


Fig. 15-17.

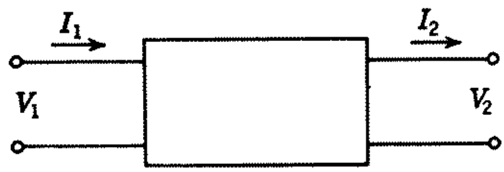


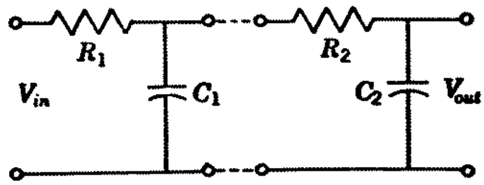
Fig. 15-18. Two-terminal-pair network.

tandem, making $V_2 = V_3$, only if in making this connection V_2 is unaffected. This is the assumption of *negligible loading* of one system by another. In terms of the electrical network shown in Fig. 15-18, the assumption of negligible loading means that the output current I_2 must equal zero or be so small that it may be neglected. If this is not the case, the interaction must be considered in writing the dynamic equations to describe the system. In other words, a section of a dynamic system cannot be separated from the system for analysis without considering the interaction of this section of the system with the rest of the system.

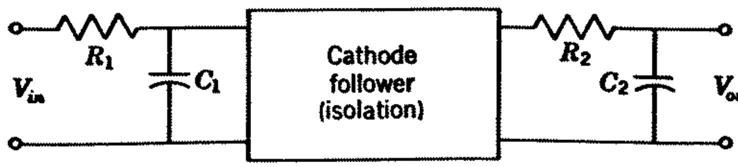
To illustrate, consider the double RC network of Fig. 15-19, shown separated into two separate RC networks. The transfer function for the entire system cannot be found by multiplying the transfer functions of the parts of the system, since the second network *loads* the first; that is, it causes a current to flow in the output of the first RC network. This will be discussed in Example 5 of this section. However, if some isolating device, such as an amplifier, is connected between the two networks, then

$$G_t = G_1G_2G_{cf} \tag{15-21}$$

where G_t is the over-all transfer function, G_1 is the transfer function of the first RC section, G_2 is the transfer function of the second RC section, and G_{cf} is the transfer function of the cathode follower amplifier



(a)



(b)

Fig. 15-19. (a) Double RC network; (b) modification of (a).

(considered a constant). This follows because the cathode follower has high input impedance and draws negligible current at its input.

Several examples will be given to illustrate the concept of the transfer function, of block diagram representation of physical systems, and of the restrictions on tandem connection of blocks.

Example 1

The electrical network shown in Fig. 15-20 is known as a *lag network*. It finds application as a compensating network in servomechanisms. The relationships between the instantaneous voltages and currents are given by Kirchhoff's law as

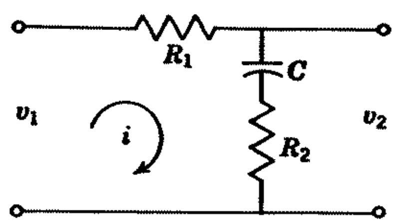


Fig. 15-20. Lag network.

$$v_1 = (R_1 + R_2)i + \frac{1}{C}\int i\,dt \tag{15-22}$$

$$v_2 = \frac{1}{C}\int i\,dt + R_2 i \tag{15-23}$$

From these equations, the transfer function is found to be

$$\frac{V_2(s)}{V_1(s)} = G(s) = \frac{R_2Cs + 1}{(R_1 + R_2)Cs + 1} = \frac{R_2}{R_1 + R_2}\frac{(s + 1/R_2C)}{[s + 1/(R_1 + R_2)C]} \tag{15-24}$$

The block diagram equivalent of the lag network of Fig. 15-20 is shown in Fig. 15-21 where $T_1 = R_2C$ and $T_2 = (R_1 + R_2)C$.

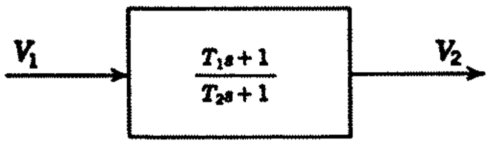


Fig. 15-21. Lag network block diagram.

Example 2

An electronic amplifier is shown in Fig. 15-22. From the equivalent circuit, the transfer function is found to be

$$\frac{V_2}{V_1} = \frac{\mu R_L}{r_p + R_L} \tag{15-25}$$

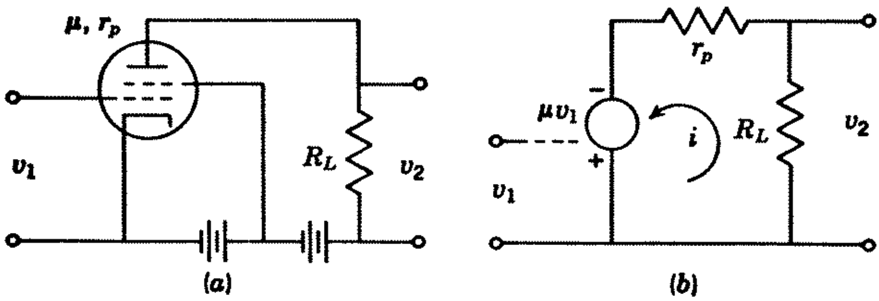


Fig. 15-22. Vacuum tube amplifier: (a) schematic; (b) equivalent circuit.

Example 3

A cathode follower schematic and the equivalent circuit are shown in Fig. 15-23. The transfer function can be determined from the

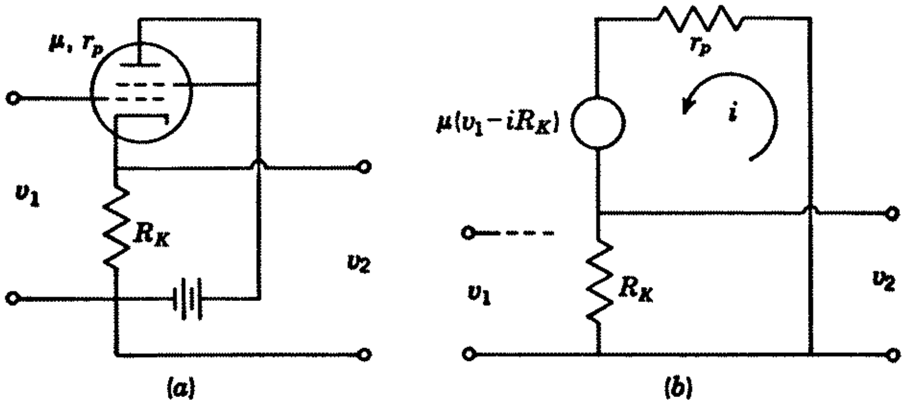


Fig. 15-23. Cathode follower: (a) schematic; (b) equivalent circuit.

equivalent circuit to be

$$\frac{V_2}{V_1} = \frac{(\mu/1 + \mu)R_K}{r_p/(1 + \mu) + R_K} \tag{15-26}$$

The value of the transfer function is a constant equal to the *gain* of the electronic circuit.

Example 4

The network for this example is shown in Fig. 15-24. The resistance R_g has a very high value such that the current flowing through it is negligible compared with that through R_2 (in other words, R_g does *not*

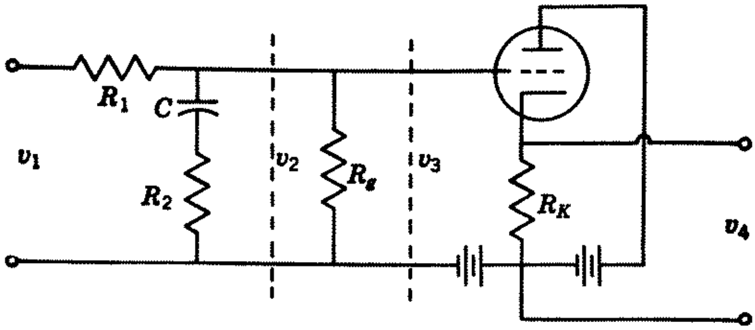


Fig. 15-24. Amplifier network.

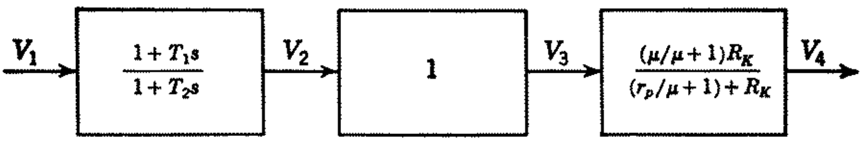


Fig. 15-25. Block diagram equivalent of Fig. 15-24.

load the remainder of the network). Also, it may be assumed that the grid draws negligible current. With these assumptions, it is seen that the network is actually composed of a lag network, shown in Fig. 15-20, and a cathode follower, shown in Fig. 15-23. The block diagram representation of the system is shown in Fig. 15-25.

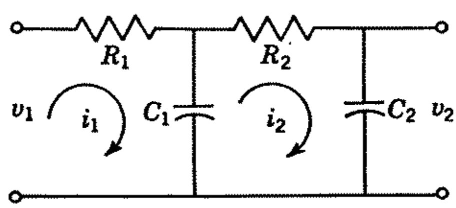


Fig. 15-26. Double *RC* network.

Example 5

For this example, consider the double *RC* network shown in Fig. 15-26. The two currents are marked i_1 and i_2. By Kirchhoff's law, the system equations are

$$
\begin{aligned}
v_1 &= R_1 i_1 + \frac{1}{C_1}\int (i_1 - i_2)\,dt \\
0 &= \frac{1}{C_1}\int (i_2 - i_1)\,dt + R_2 i_2 + \frac{1}{C_2}\int i_2\,dt \qquad (15\text{-}27) \\
v_2 &= \frac{1}{C_2}\int i_2\,dt
\end{aligned}
$$

The corresponding transform equations are, if the capacitors are initially uncharged,

$$V_1(s) = R_1 I_1(s) + \frac{1}{C_1 s}[I_1(s) - I_2(s)]$$

$$0 = \frac{1}{C_1 s}[I_2(s) - I_1(s)] + R_2 I_2(s) + \frac{1}{C_2 s} I_2(s) \tag{15-28}$$

$$V_3(s) = \frac{1}{C_2 s} I_2(s)$$

Considering V_1 as the input and V_3 as the output, the block diagram of Fig. 15-27 is constructed from the algebraic equations of Eq. 15-28. The block diagram reduction procedure, following the rules stated in

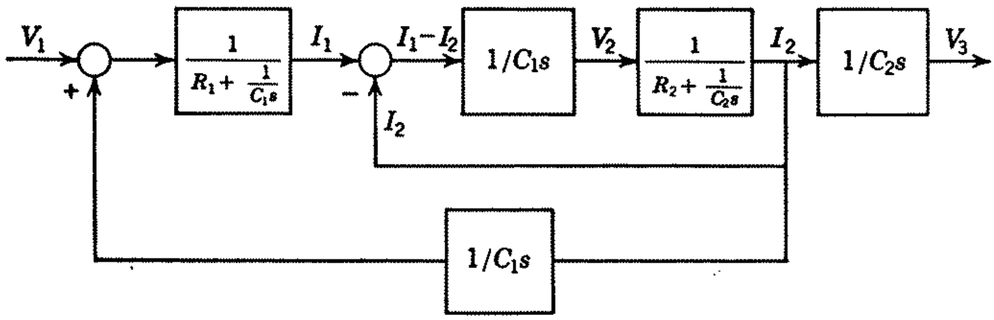


Fig. 15-27. Double *RC* network block diagram.

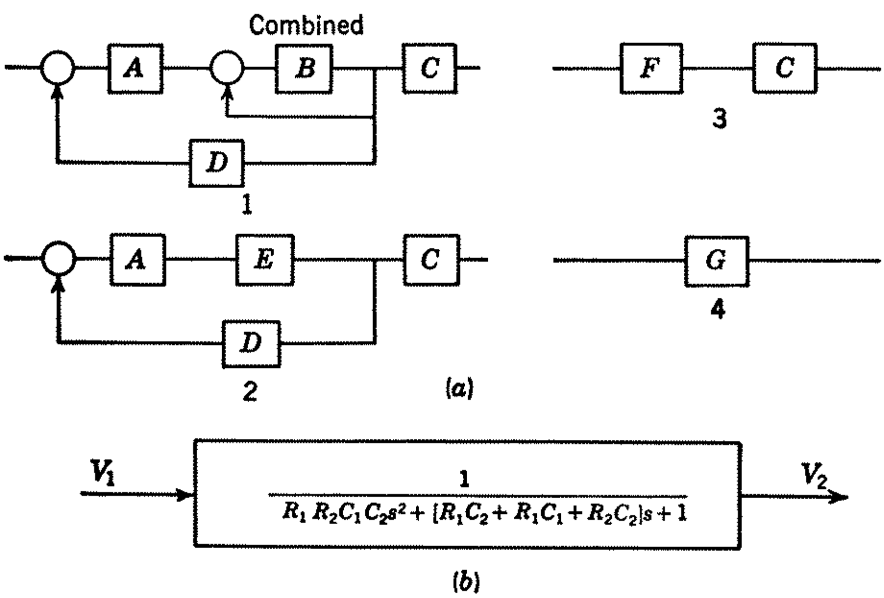


Fig. 15-28. Block diagram reduction.

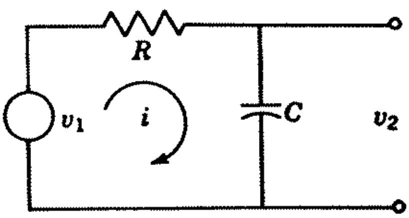
Art. 15-4, is outlined by the sequence of blocks in Fig. 15-28(a). The simplified block diagram and composite transfer function are shown in Fig. 15-28(b). The same result could be found by algebraic manipulation of the equations of Eq. 15-28 instead of the manipulation of blocks.

FURTHER READING

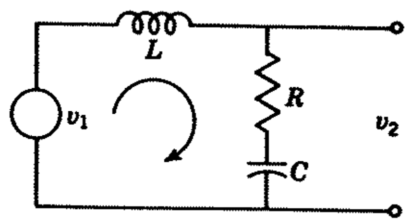
Excellent discussions of the use of block diagrams are found in the literature by T. M. Stout, "A block diagram approach to network analysis," *Trans. AIEE* (*Applications and Industry*), **3**, 255 (1952), and "Block-diagram solutions for vacuum-tube circuits," *Trans. AIEE* (*Communication and Electronics*), **9**, 561 (1953), and by T. D. Graybeal, "Block diagram network transformation," *Elec. Eng.*, **70**, 985 (1951). For a different approach to the same subject, the paper by S. J. Mason, "Feedback theory—some properties of signal flow graphs," *Proc. IRE*, **41**, 1144 (1953) is especially recommended.

PROBLEMS

15-1. For the network shown in the accompanying figure, (a) draw a block diagram with one block for *each* element, considering $V_1(s)$ to be the excitation (or input) and $I(s)$ the response (or output). (b) Repeat part (a) with $V_1(s)$ as the excitation but $V_2(s)$ as the response.

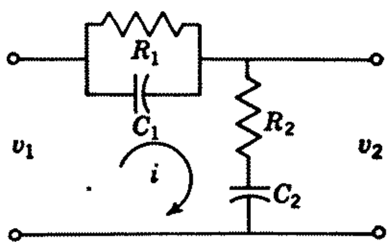


Prob. 15-1.

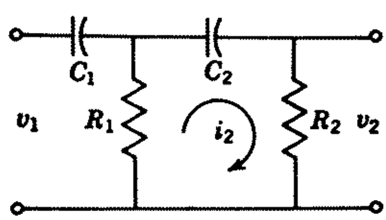


Prob. 15-2.

15-2. Repeat Prob. 15-1 for the network shown in the figure.
15-3. Repeat Prob. 15-1 for the network shown in the figure.

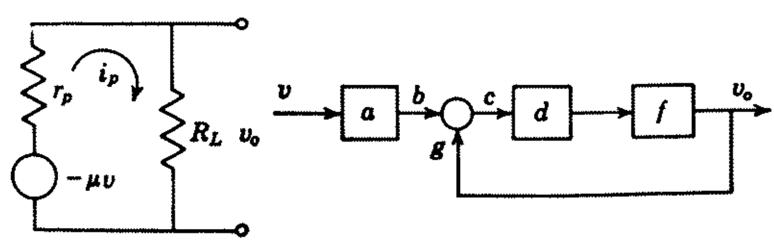


Prob. 15-3.



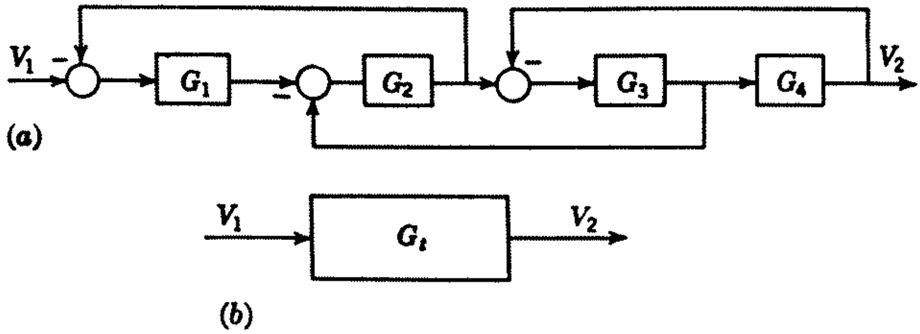
Prob. 15-4.

15-4. Repeat Prob. 15-1 for the network shown in the figure.
15-5. The schematic shown in the figure is the equivalent circuit of a vacuum tube amplifier. The accompanying block diagram representation shows a number of blocks and arrows labeled a, b, . . ., g for identification. Give the transfer function for each block and identify the quantity associated with all marked arrows.



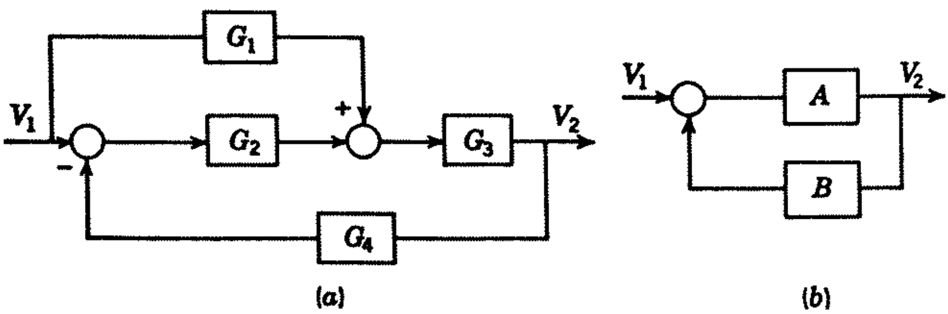
Prob. 15-5.

15-6. Simplify the block diagram shown as (a) to the form of (b). Give the value of G_t in terms of G_1, G_2, G_3, and G_4.



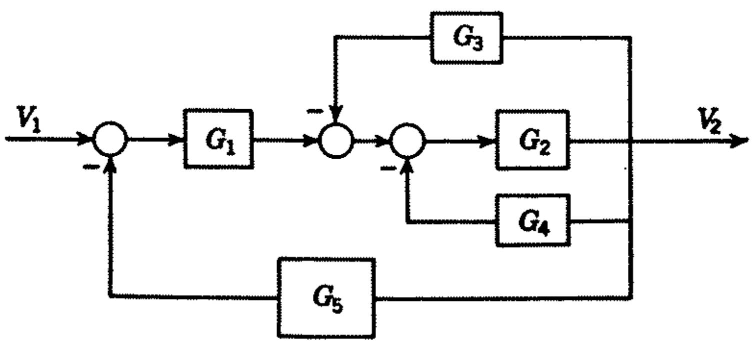
Prob. 15-6.

15-7. Reduce the block diagram shown in (a) to the form shown in (b), giving the value for A and B in the feed-forward and feedback loops.



Prob. 15-7.

15-8. Repeat Prob. 15-7 for the block diagram shown in the accompanying figure.

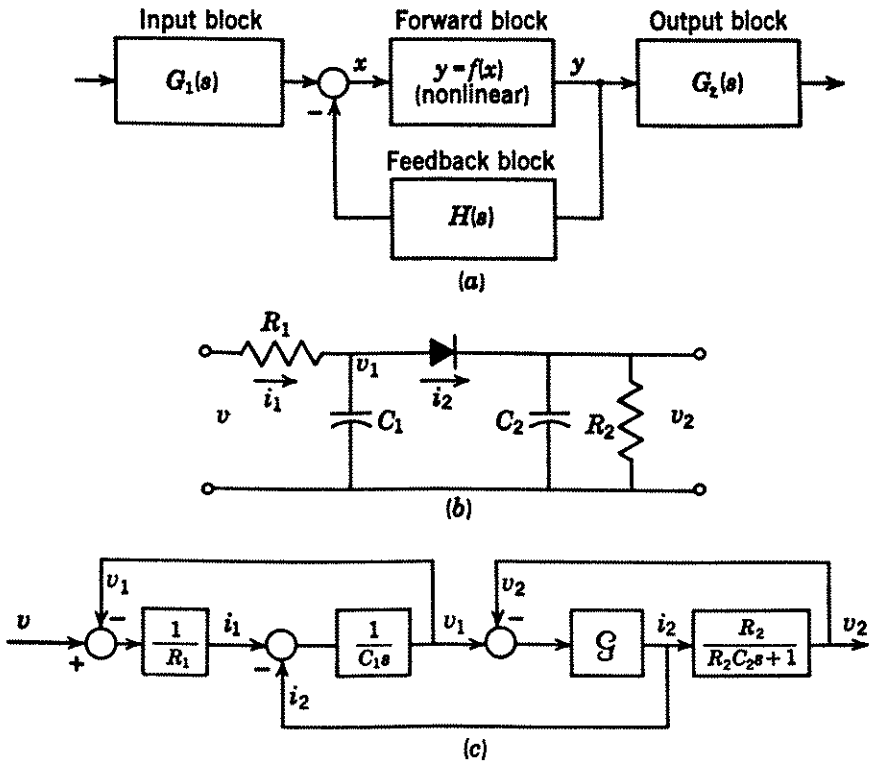


Prob. 15-8.

15-9. T. M. Stout* has shown that many networks containing one nonlinear element can be represented by a block diagram having the form shown below in (a). The nonlinear element in the circuit of (b) is a diode described by a conductance function, $i_2 = \mathcal{G}(v_1 - v_2)$. The equivalent block diagram for the system is shown in (c). Do not attempt

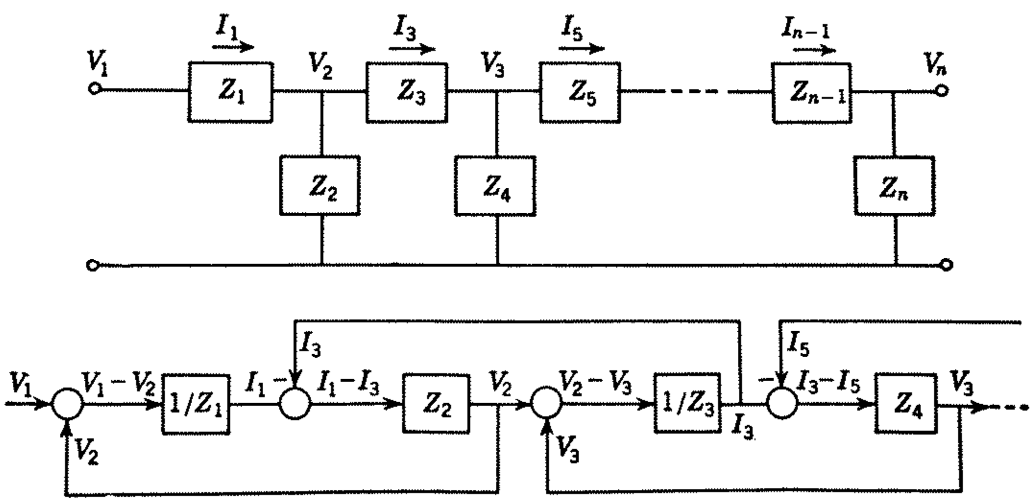
* T. M. Stout, "Block diagram simplification of some nonlinear circuits," Report No. 9, E. E. Dept., University of Washington, 1952.

to combine the nonlinear block with other blocks (this is not permitted since superposition does not hold). By manipulation and combination of the blocks, reduce the block diagram of (c) to the standard form shown in (a). Give the values for $G_1(s)$, $G_2(s)$, and $H(s)$.



Prob. 15-9.

15-10. The accompanying figure shows a general ladder network with alternating series and shunt elements or combinations of elements. Show by discussion and equations that the ladder network can be represented by the block diagram of the figure.



Prob. 15-10.

CHAPTER 16

STABILITY IN FEEDBACK SYSTEMS

16-1. Feedback systems

Many electric and electromechanical systems incorporate a so-called feedback path by means of which a part of the output is reintroduced at the input. For example, Fig. 16-1(a) shows a plate-tuned oscillator.

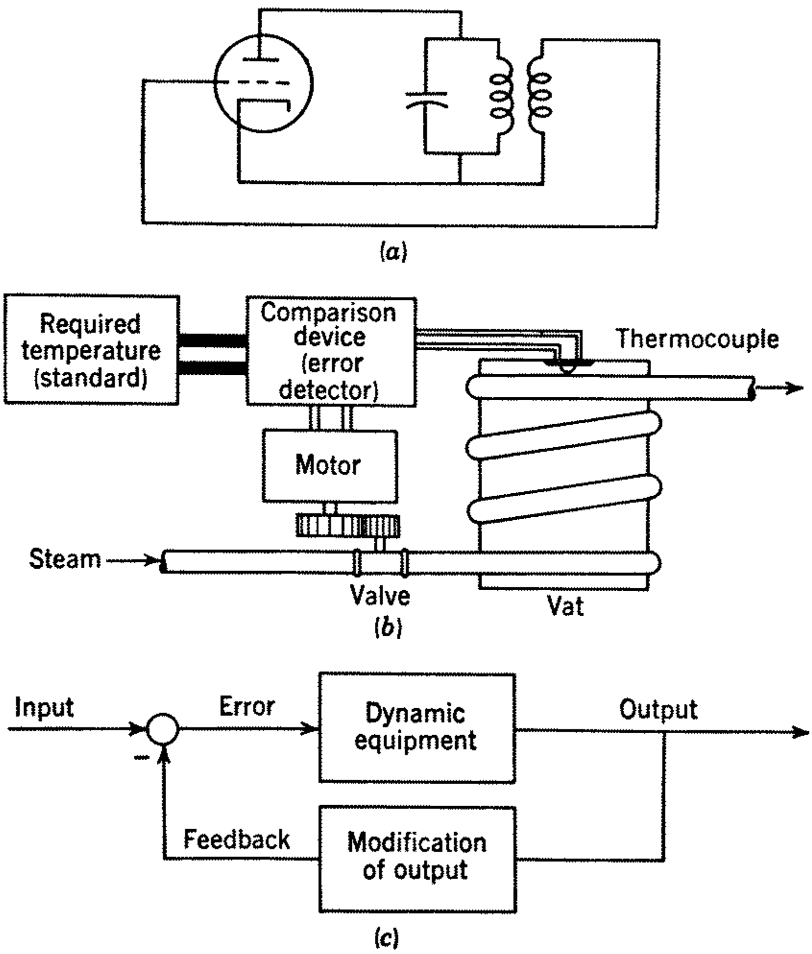


Fig. 16-1. Examples of feedback systems: (a) plate-tuned oscillator; (b) temperature regulating system; (c) feedback system representation.

Feedback, which is essential to the operation of an oscillator, is accomplished by means of the coupled coils returning the output from the plate to the grid. Figure 16-1(b) shows an electromechanical temperature-regulating system. Feedback is accomplished in this system by means of the thermocouple, which produces a voltage proportional to the temperature in the vat. This feedback is compared with the standard to produce an "error voltage" which in turn controls the amount of steam introduced into the system.

By means of block diagram algebra, discussed in the last chapter, such systems as the oscillator and the temperature control system can be reduced to a standard form of a feedback system shown in Fig. 16-1(c).

Feedback systems afford many advantages such as accurate control and improved time of response. These advantages are partially offset by such systems being capable of *instability*.

A system is said to be *stable* if, for small values of input, the output remains small or does not increase with time. We do not ordinarily think of a linear system as being capable of instability by this definition. If the system is made up entirely of passive linear elements, there is no energy source to supply an output which increases with time and thus without limit. There will be a distinct relationship between input and output as expressed by the transfer function. Linear elements can at most modify the form or the power level of the input.

If the output is to increase without limit, energy must be supplied to the system. This supply must be *within* the system closed by the feedback loop for the output to increase with the input remaining small. The feedback path is required for an unstable system to introduce a new input into the system from the output to override the initial small input.

Thus there are two essential features of a system capable of being unstable: (1) *there must be a source of energy within the system*, and (2) *there must be at least one feedback path.*

These are necessary but not sufficient conditions for instability. That is to say, feedback systems are, with proper design, not only stable but superior in many respects to systems without feedback. An important problem for the engineer is to meet specifications and yet avoid instability. This can be done from transfer functions by a mathematical or a graphical procedure. It is these procedures we will consider in this chapter.

16-2. System stability in terms of the characteristic equation

The question of stability is, by the definition just given, fundamentally transient in nature, and is related to the transient response of the closed-loop system. The transfer function of the closed loop written in terms of the open-loop transfer function was studied in Chapter 15, and is

$$\frac{V_2(s)}{V_1(s)} = L(s) = \frac{G(s)}{1 + G(s)H(s)} \tag{16-1}$$

where $G(s)$ and $H(s)$ are defined in Fig. 15-15; $V_2(s)$ and $V_1(s)$ are the output and input, respectively. This transfer function has the form of a quotient of polynomials in s. Expanding the denominator polynomial, we write

$$\frac{V_2(s)}{V_1(s)} = \frac{G(s)}{a_0 s^n + a_1 s^{n-1} + \ldots + a_{n-1}s + a_n} \tag{16-2}$$

or

$$(a_0 s^n + a_1 s^{n-1} + \ldots + a_n)V_2(s) = G(s)V_1(s) \tag{16-3}$$

Written in this form, the denominator polynomial, that is $[1 + G(s)H(s)]$, when set equal to zero is recognized as the *characteristic equation* of the system.* If the characteristic equation is factored into its n roots, Eq. 16-2 can be written

$$\frac{V_2(s)}{V_1(s)} = \frac{G(s)}{a_0(s - s_a)(s - s_b)\ldots(s - s_n)} \tag{16-4}$$

$$= \frac{G(s)}{a_0} \prod_{j=a}^{n} \frac{1}{(s - s_j)} \tag{16-5}$$

In solving for the time-domain solution of this equation with $v_1(t)$ specified, the usual procedure is to expand this equation by partial fractions, thus evaluating the arbitrary constants of the transient portion of the solution. (In using the transfer function, we assume that all initial conditions are zero in this particular time domain solution.) If there are no repeated roots in the characteristic equation, the expression for $v_2(t)$ will be

$$v_2(t) = v_{2_{ss}} + \sum_{j=1}^{n} K_j e^{s_j t} \tag{16-6}$$

In other words, the transient portion of the time-domain solution is determined in form by the roots of the characteristic equation. For the general case, the roots of the equation are complex and written as

$$s = \sigma + j\omega \tag{16-7}$$

The form of the time domain solution corresponding to each root depends on the magnitude and *signs* of σ and ω. Several cases are of interest:

Case 1. σ negative for a finite ω. For this case, the solution will be (for the combined terms of the conjugate pair $\sigma \pm j\omega$),

$$Ke^{-\sigma t} \sin (\omega t + \phi) \tag{16-8}$$

* Compare this equation with Eq. 6-88 and Eqs. 7-37 and 7-38.

This expression is termed a *damped sinusoid;* the magnitude reduces to zero as time t becomes large.

Case 2. $\sigma = 0$ and finite ω. For the conjugate pair of roots $\pm j\omega$, the solution will be similar in form to Eq. 16-8 with $\sigma = 0$; that is,

$$K \sin (\omega t + \phi) \tag{16-9}$$

This function oscillates at constant amplitude as a function of time.

Case 3. σ positive and a finite ω. Again, the solution is similar to Eq. 16-8 and is

$$Ke^{+\sigma t} \sin (\omega t + \phi) \tag{16-10}$$

In this case, the magnitude of oscillations increases exponentially and *without limit* for large values of time t.

Case 4. $\sigma = 0$ and $\omega = 0$. For this limiting case, the solution is a constant and does not change with time.

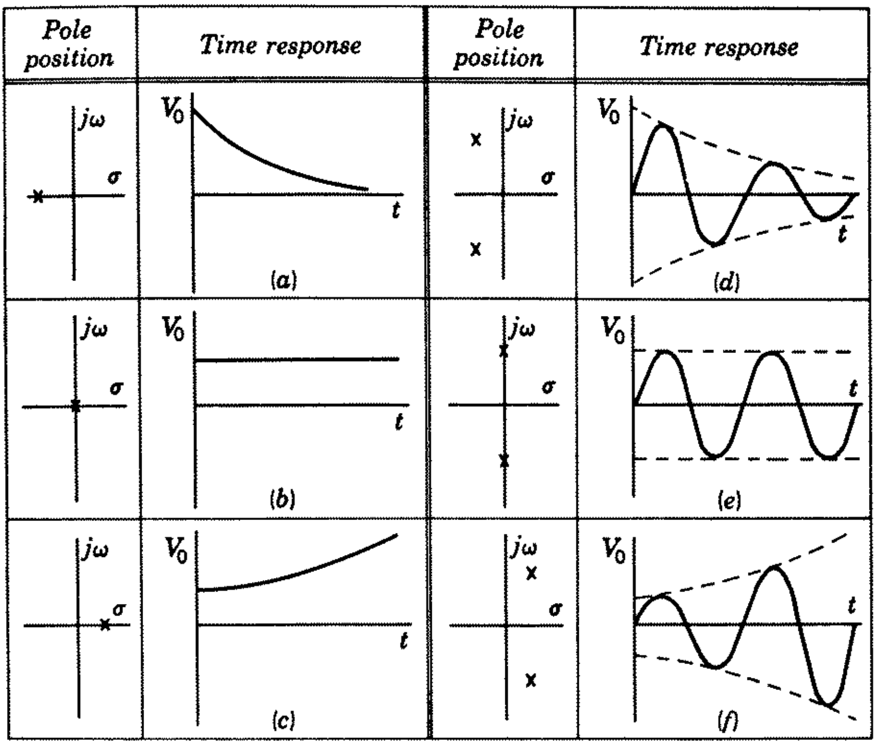


Fig. 16-2. Comparison of the time response of a system (V_0) with pole position in the complex s plane.

Case 5. σ either positive or negative, but $\omega = 0$. The solution for this case (for a simple root) has the form

$$Ke^{\pm\sigma t} \tag{16-11}$$

For negative values of σ, the magnitude of the term diminishes with time. For positive σ, the term increases exponentially. These several cases are illustrated in Fig. 16-2. From the figure, the effect of the sign of σ and the role of ω can be seen. A finite nonzero value of ω corresponds to oscillation. For negative values of σ, the response decreases with time, but for positive values of σ, the response increases without limit with time. Our concept of stability evidently ties in directly with the sign of the real part of the roots of the characteristic equation.

By our definition of stability, that the output should remain small or not increase with time, the responses of parts (c) and (f) of Fig. 16-2 are identified with unstable systems. These both correspond to positive values of σ. The transition case shown as part (e) of Fig. 16-2 is, strictly speaking, a stable case. The output does not increase without limit. It corresponds to sustained oscillations as, for example, in an electronic oscillator. With negative values of σ, there is *damping*, and the system is stable. Just as stability is related directly to the sign of the real part of the roots of the characteristic equation, so the basic problem in determining stability is finding an answer to this question: *Does the characteristic equation, $1 + G(s)H(s) = 0$, have any roots with positive real parts?* This is the basis of all stability studies.

The problem of determining stability is thus one of finding the roots of the equation

$$1 + G(s)H(s) = a_0s^n + a_1s^{n-1} + a_2s^{n-2} + \dots + a_{n-1}s + a_n = 0 \tag{16-12}$$

or some equivalent to finding the roots. For first- and second-order equations, $n = 1$ and $n = 2$, this is a simple matter: merely factor the equation and so find the roots. But as the order of the equation increases, the task of finding the roots becomes very tedious and time consuming (unless computing machines are available), and alternate methods are used.

Suppose that the characteristic equation is of first order,

$$a_0s + a_1 = 0 \tag{16-13}$$

The root of this equation is

$$s = -\frac{a_1}{a_0} \tag{16-14}$$

The real part of this root is negative as long as a_1 and a_0 are positive and real; this requirement is met for all physical systems. Likewise a second-order characteristic equation,

$$a_0s^2 + a_1s + a_2 = 0 \tag{16-15}$$

will have two roots given by the equation

$$s_1, s_1^* = -\frac{a_1}{2a_0} \pm \sqrt{\frac{a_1^2}{4a_0^2} - \frac{a_2}{a_0}} \tag{16-16}$$

The two roots s_1 and s_1^*, will have negative real parts only as long as the coefficients a_0, a_1, and a_2 are positive.

For both the first- and second-order system, a requirement that the sign of all coefficients be positive is sufficient to guarantee that the real part of the roots will be negative. But for higher-ordered systems, this rule is not sufficient. To illustrate, consider the third-order polynomial formed as follows.

$$(s + 4)(s^2 - 2s + 10) = s^3 + 2s^2 + 2s + 40 \tag{16-17}$$

In this example, the coefficients of the third-order polynomial are al positive even though there are clearly two roots with positive real parts. There are further requirements that must be satisfied in addition to the positive coefficient requirement. The requirements take the form of relationships of the magnitudes of the coefficients of the polynomial given by Routh's criterion. Before turning to a study of this criterion, let us summarize our findings thus far:

(1) In order that there be no roots of a polynomial with positive real parts, it is necessary but not sufficient that the coefficients of the polynomial be positive.
(2) If a coefficient of a polynomial is negative, the polynomial has at least one positive real root (since that is the only way the coefficient could be made negative).

16-3. The Routh criterion or Routh rule*

The Routh rule gives a procedure for determining the number of roots of a polynomial with positive real parts without actually finding the roots. Consider the polynomial

$$a_0s^n + a_1s^{n-1} + a_2s^{n-2} + \dots + a_{n-1}s + a_n = 0 \tag{16-18}$$

As the first step in the application of the rule, form two rows made up of alternate coefficients of the equation; that is, from the first, third, fifth, etc. coefficient form one row, and from the second, fourth, sixth, etc. coefficients form a second row as follows:

row 1

$$a_0s^n + a_1s^{n-1} + a_2s^{n-2} + a_3s^{n-3} + a_4s^{n-4} + a_5s^{n-5} + a_6s^{n-6} + \dots \tag{16-19}$$

row 2

* E. J. Routh, Advanced Part of *Dynamics of a System of Rigid Bodies*, Vol. II (6th ed.), (Macmillan & Co. Ltd., London, 1930).

Write the two rows in the form

$$\begin{array}{cccccc} a_0 & a_2 & a_4 & a_6 & a_8 & \cdots \\ a_1 & a_3 & a_5 & a_7 & a_9 & \cdots \end{array} \tag{16-20}$$

As the next step, complete the following array of numbers (shown for a sixth-order system):

$$\begin{array}{cccc} \text{use} & & & \\ a_0 & a_2 & a_4 & a_6 \\ a_1 & a_3 & a_5 & \\ b_1 & b_3 & b_5 & \\ c_1 & c_3 & & \\ d_1 & d_3 & & \\ e_1 & & & \\ f_1 & & & \end{array}$$

where the b, c, d, e, f, and g coefficients are defined in terms of the a coefficients by the following pattern:

$$b_1 = \begin{matrix} a_0 & a_2 \\ a_1 & a_3 \end{matrix} = \frac{a_1a_2 - a_0a_3}{a_1}; \quad b_3 = \begin{matrix} a_0 & a_4 \\ a_1 & a_5 \end{matrix} = \frac{a_1a_4 - a_0a_5}{a_1}$$

$$c_1 = \begin{matrix} a_1 & a_3 \\ b_1 & b_3 \end{matrix} = \frac{b_1a_3 - b_3a_1}{b_1}; \qquad \text{etc.}$$

In general, any new element is found from the two elements above the element in the same column and the two elements above but in the column to the right. These elements form a determinant-like structure. The elements joined by a line with positive slope have a positive sign, while the elements joined by a line with negative slope have a negative sign (just the opposite of the rule for determinants). We subtract these two products and divide this difference by the element on the lower left-hand corner of the array. This process is continued to give the Routh array.

The number of changes in sign in elements of the first column (marked *use*) indicates the number of roots of the equation with positive real parts. For there to be no roots with positive real parts, all elements of the first column must have the same sign.

For the rule as given to hold, it is necessary that no powers of s in the equation be missing. However, if any such terms are missing, the equation has at least one root with a positive real part and so fails the test by inspection. An exception occurs when the equation contains terms which are all even powers or all odd powers, indicating that all roots are purely imaginary.

Example 1

Consider the identity

$$(s + 1)(s + 2)(s + 3)(s + 4) = s^4 + 10s^3 + 35s^2 + 50s + 24 \quad (16\text{-}21)$$

The Routh array is formed to give the following:

$$\begin{array}{rrr} 1 & 35 & 24 \\ 10 & 50 & \\ 30 & 24 & \\ 42 & & \\ 24 & & \end{array}$$

From the first column, it is seen that there are no roots with positive real parts (agreeing with the known roots).

Example 2

As a second example, consider Eq. 16-17.

$$s^3 + 2s^2 + 2s + 40 = 0 \quad (16\text{-}22)$$

which is known to have two roots with positive real parts.

$$\begin{array}{rr} 1 & 2 \\ 2 & 40 \\ -18 & \\ 40 & \end{array}$$

There are two changes of sign (2 to -18 and -18 to 40) as required.

Example 3

Consider a third-order equation,

$$a_0s^3 + a_1s^2 + a_2s + a_3 = 0 \quad (16\text{-}23)$$

The Routh array for this equation is

$$\begin{array}{cc} a_0 & a_2 \\ a_1 & a_3 \\ \dfrac{a_1a_2 - a_0a_3}{a_1} & \\ a_3 & \end{array}$$

From the array we conclude that it is necessary that all coefficients in the equation be positive and, in addition, that $a_1a_2 > a_0a_3$ in order that there be no roots with positive real parts in a third-order equation.

16-4. The Hurwitz criterion

A Hurwitz polynomial is a polynomial having roots with negative real parts only. Polynomials representing the characteristic equations of stable systems are therefore Hurwitz polynomials. To apply the Hurwitz criterion to a polynomial, carry out the following steps:

(1) Separate the polynomial into even and odd parts (that is, parts with even powers in s and with odd powers in s). Form a quotient of these two polynomials with the part of higher degree as the numerator polynomial.

(2) Expand the quotient of polynomials as a Stieltjes *continued fraction;* thus

$$\frac{P(s)}{Q(s)} = \alpha_1 s + \cfrac{1}{\alpha_2 s + \cfrac{1}{\alpha_3 s + \cfrac{1}{\alpha_4 s + \cfrac{1}{\alpha_5 s + \cfrac{1}{\alpha_6 s + \cdots}}}}} \tag{16-24}$$

For the polynomial to be Hurwitz, it is necessary that all of the α-coefficients *be positive.** However, the test we have described does *not* rule out the possibility of roots without real parts (i.e., on the $j\omega$ axis of the s plane) in the polynomial under test. Such roots correspond to $(s^2 + \omega_1^2)$ factors in the polynomial which are incorporated in both the even and odd parts of the polynomial. Thus

$$\begin{aligned}(s^2 + \omega_1^2)[P(s) + Q(s)] \\ = (s^2 + \omega_1^2)P(s) + (s^2 + \omega_1^2)Q(s) = P_1(s) + Q_1(s)\end{aligned} \tag{16-25}$$

Hence, terms of the type $(s^2 + \omega_1^2)$ cancel when the quotient $P_1(s)/Q_1(s)$ is formed in applying the test. The test does assure that the polynomial is either Hurwitz or a Hurwitz polynomial multiplied by factors of the form, $(s^2 + \omega_1^2)$.

Formation of the continued fraction is most easily accomplished by an "invert-and-divide" procedure. After completion of step (1) listed above, divide the part of lower degree into the other part and complete one step only. Invert the remainder and continue the process until it comes to an end (as it must). This is best illustrated by an example. Consider the polynomial

$$4s^4 + 2s^3 + 8s^2 + 3s + 1.5 = 0 \tag{16-26}$$

* A derivation of this criterion is given by Guillemin; see reference at end of chapter.

The quotient of polynomials is formed as follows.

$$\frac{4s^4 + 8s^2 + 1.5}{2s^3 + 3s} \tag{16-27}$$

Dividing one step gives

$$
\begin{array}{rrl}
2s^3 + 3s) & 4s^4 + 8s^2 + 1.5 & (2s \\
 & \underline{4s^4 + 6s^2 \qquad\quad} & \\
 & 2s^2 + 1.5 &
\end{array}
$$

Again, we invert and divide as

$$
\begin{array}{rrl}
2s^2 + 1.5) & 2s^3 + 3s & (s \\
 & \underline{2s^3 + 1.5s} & \\
 & 1.5s &
\end{array}
$$

and again,

$$
\begin{array}{rrl}
1.5s) & 2s^2 + 1.5 & (\tfrac{4}{3}s \\
 & \underline{2s^2 \qquad\quad} & \\
 & 1.5 &
\end{array}
$$

and finally,

$$
\begin{array}{rrl}
1.5) & 1.5s & (1s
\end{array}
$$

Hence the continued fraction expansion is

$$\frac{4s^4 + 8s^2 + 1.5}{2s^3 + 3s} = 2s + \cfrac{1}{s + \cfrac{1}{\frac{4}{3}s + \cfrac{1}{s}}} \tag{16-28}$$

Since all the α-coefficients are positive, the polynomial is Hurwitz. The operations shown in four steps above are conveniently carried out as a continued operation as illustrated below.

$$
\begin{array}{rrrrrl}
2s^3 + 3s) & 4s^4 + 8s^2 + 1.5 & (2s & & & \\
 & \underline{4s^4 + 6s^2 \qquad\quad} & & & & \\
 & 2s^2 + 1.5) & 2s^3 + 3s & (s & & \\
 & & \underline{2s^3 + 1.5s} & & & \\
 & & 1.5s) & 2s^2 + 1.5 & (\tfrac{4}{3}s & \\
 & & & \underline{2s^2 \qquad\quad} & & \\
 & & & 1.5) & 1.5s & (s \\
 & & & & \underline{1.5s \qquad} & \\
 & & & & 0 &
\end{array}
$$

16-5. The Nyquist criterion

The stability criterion we will study next was developed by Nyquist* of the Bell Telephone Laboratories in 1932. While the objective of this criterion is the same as the Routh criterion and the Hurwitz criterion, the approach differs in several respects.

* H. Nyquist, "Regeneration theory," *Bell System Tech. J.*, **11,** 126 (1932).

(1) Analysis is made in terms of the open-loop transfer function rather than the closed-loop characteristic equation.
(2) The method is partially graphical and, as will be shown, inspection of the graphical plot gives more than the "yes or no" answer of the Routh and Hurwitz criteria.
(3) Analysis is made in terms of the sinusoidal steady state where the concepts of phase and magnitude ratio are readily related to experiments.

The basic operation in applying the Nyquist criterion is a *mapping* from the s plane to the $F(s)$ plane. By the term mapping, we mean that a set of values of s (for example, s_1, s_2, and s_3) have, for a given $F(s)$, a corresponding set of values of $F(s)$, [namely, $F(s_1)$, $F(s_2)$, and $F(s_3)$]. These three—and an infinite number of other—points are shown in Fig. 16-3. Here an arbitrary contour in the s plane is "mapped" into

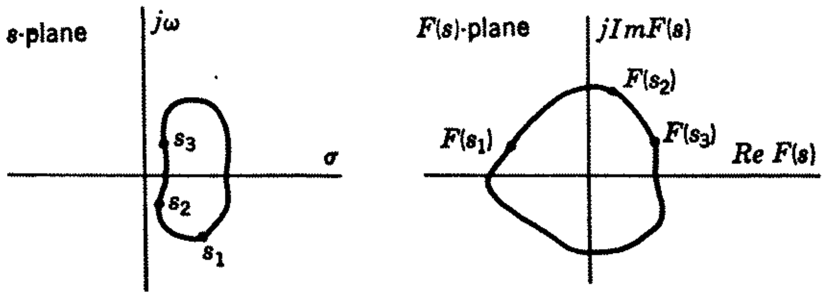


Fig. 16-3. Mapping illustration.

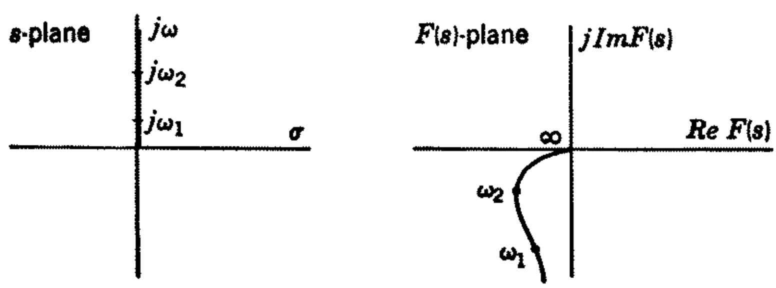


Fig. 16-4. Mapping example for $F(s) = 1/s(sT + 1)$.

a corresponding contour in the $F(s)$ plane. A specific example is shown in Fig. 16-4. The mapping is made for imaginary values of s; that is, $s = j\omega$ for $\omega \geqq 0$. The specific function is

$$F(s) = \frac{1}{s(sT + 1)} \tag{16-29}$$

As another example of a mapping operation, not so difficult as the one given above, suppose that two function are related by the equation

$$F(s) = P(s) + 1 \tag{16-30}$$

that is, the function $P(s)$ plus a constant (unity) is equal to the function $F(s)$. A typical plot in the two planes is shown in Fig. 16-5. The transformation evidently moves the plot one unit to the left.

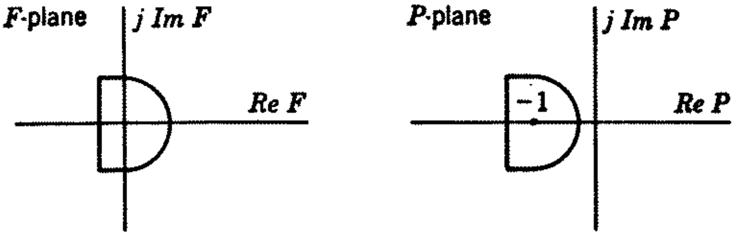


Fig. 16-5. Mapping of $F(s) = P(s) + 1$.

Next, suppose that $F(s)$ is factored to find its poles and zeros which are given in the equation

$$F(s) = K\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \tag{16-31}$$

where $s_1, s_2, \ldots, s_n$ are the zeros and $s_a, s_b, \ldots, s_m$ are the poles. These poles and zeros are displayed on a plot of the s plane shown in Fig. 16-6(a) (an arbitrary array for purposes of illustration). A single zero,

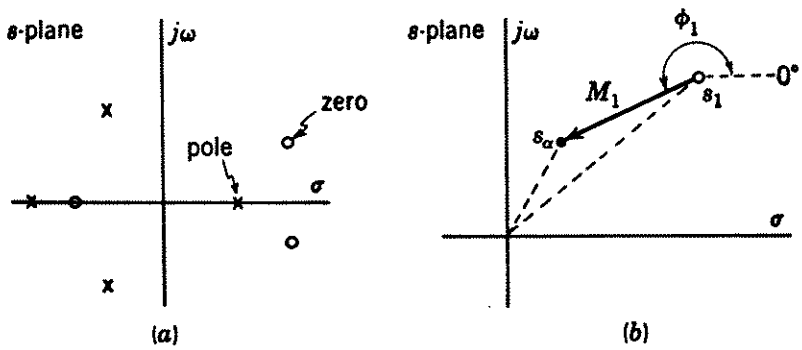


Fig. 16-6. Pole-zero configuration.

s_1, is isolated in Fig. 16-6(b). This zero comes from the term $(s - s_1)$ in Eq. 16-31. At some particular frequency s_α, this term has a value $(s_\alpha - s_1)$ which may be expressed in polar form as

$$(s_\alpha - s_1) = M_1 e^{j\phi_1} \tag{16-32}$$

where M_1 is the magnitude and ϕ_1 is the phase angle of the phasor $(s_\alpha - s_1)$. This magnitude and phase are shown on the s plane in Fig. 16-6(b). Any other term in Eq. 16-31 can be similarly expressed; for example,

$$(s_\alpha - s_b) = M_b e^{j\phi_b} \tag{16-33}$$

When all terms are so expressed, Eq. 16-31 takes the form

$$M_t e^{j\phi_t} = |F(s)|e^{j\,\mathrm{Ang}\,F(s)} = \frac{KM_1M_2M_3M_4\ldots}{M_aM_bM_cM_d\ldots}\,e^{j\phi_t} \tag{16-34}$$

where $$\phi_t = \phi_1 + \phi_2 + \ldots - \phi_a - \phi_b - \ldots \tag{16-35}$$

This last equation tells us that the total phase at some frequency s_α for the function $F(s)$ may be found by adding the phase of the "zero phasors" and subtracting the phase of the "pole phasors"; in other words,

$$\operatorname{Ang} F(s) = \operatorname{Ang}(s - s_1) + \operatorname{Ang}(s - s_2) + \ldots - \operatorname{Ang}(s - s_a) - \ldots \tag{16-36}$$

for any value of s.

Figure 16-7 shows the s plane with two zeros, s_1 and s_2, and a mapping contour C (s_α is a point on the contour). Consider the effect of s_1 (ignoring s_2) as s_α moves along C in a *clockwise* direction. After one complete traversing of the closed contour C, the phase of the phasor term $(s - s_1)$ has increased by -2π radians. Next, consider the effect of s_2 on the factor $(s - s_2)$, this time ignoring s_1, as the same closed contour C is traversed in the same clockwise direction. There is *no* net gain in phase of the phasor term $(s - s_2)$. In summary, if the closed contour encircles a zero in traversing a closed path in the clockwise direction, the function changes in phase by -2π radians; if no zero is encircled, there is no change in phase.

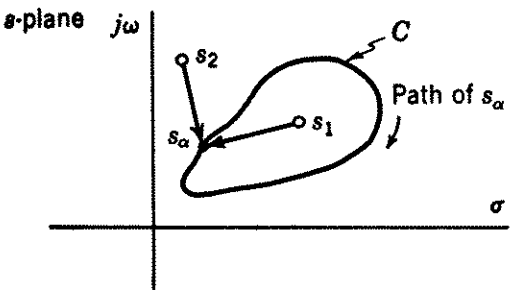


Fig. 16-7. Mapping from s-plane.

Exactly the same conclusion may be reached in the case of a pole except that the phase is changed by $+2\pi$ radians.

Suppose next that a contour is selected in the s plane of Fig. 16-6(a) such that P poles and Z zeros are encircled as the contour is traversed in a clockwise direction. The net change in the phase of the function $F(s)$ will be given by the equation

$$\Delta \operatorname{Ang} F(s) = 2\pi(P - Z) \quad \text{radians} \tag{16-37}$$

Return next to the mapping of the s plane into the $F(s)$ plane. Let us examine the behavior of the $F(s)$ plot in the complex plane as the closed contour in the s plane is traversed. An increase in the phase of $F(s)$ manifests itself in the $F(s)$ plane by an encirclement of the origin for every 2π radian increase. Further, every zero encircled will cause one counterclockwise encirclement of the origin just as every pole will cause a clockwise encirclement. Should the contour not encircle any poles or zeros—or if it encircles equal numbers of poles and zeros—the contour in the $F(s)$ plane will not encircle the origin. In summary, if the closed contour C in the s plane encircles in a clockwise (or negative) direction P poles and Z zeros, the corresponding contour in the $F(s)$ plane encircles the origin $(P - Z)$ times in a counterclockwise (or

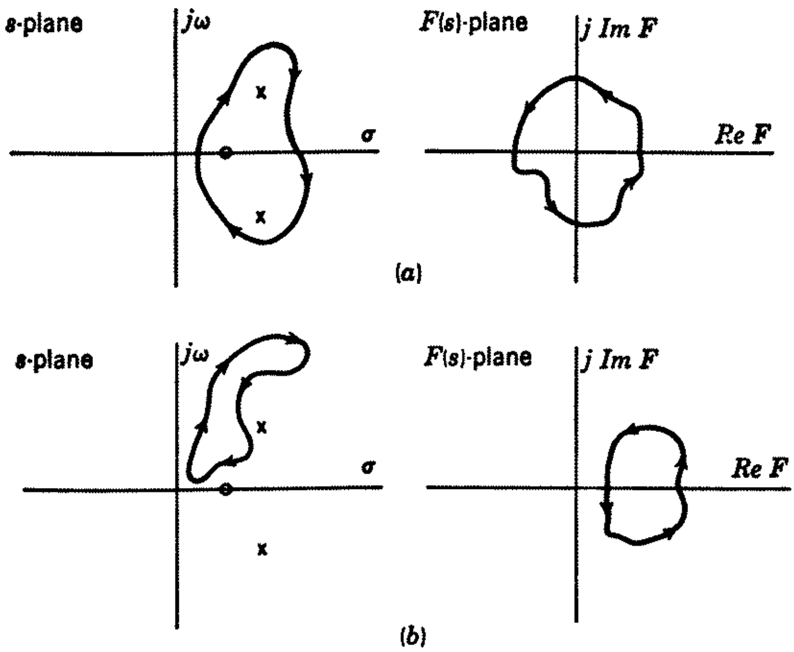


Fig. 16-8. Illustration of the rule $\Delta\phi = 2\pi(P - Z)$.

positive) direction. Two examples are given in Fig. 16-8 to illustrate this conclusion.

16-6. Application to a closed-loop system

The concepts reviewed in the last section will next be applied to a closed-loop system having a feed-forward transfer function $G(s)$ and a feedback transfer function $H(s)$, shown in Fig. 16-9. The input and output are related by the closed-loop transfer function, written in terms of $G(s)$ and $H(s)$ by the equation

$$\frac{V_2(s)}{V_1(s)} = \frac{G(s)}{1 + G(s)H(s)} \qquad (16\text{-}38)$$

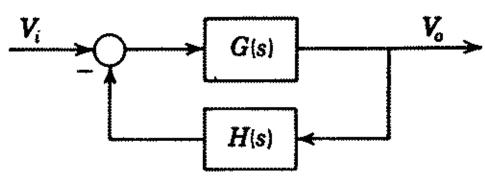


Fig. 16-9. Closed-loop system.

The poles and zeros of the two functions $(1 + GH)$ and (GH) must be considered in the derivation of the Nyquist criterion. Let

$$1 + G(s)H(s) = \frac{P(s)}{Q(s)} = K\frac{(s - s_1)(s - s_2)\ldots(s - s_n)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \qquad (16\text{-}39)$$

$$G(s)H(s) = K'\frac{(s - s_\alpha)(s - s_\beta)\ldots(s - s_\omega)}{(s - s_a)(s - s_b)\ldots(s - s_m)} \qquad (16\text{-}40)$$

The two functions have the *same poles*. In Eq. 16-39, the order of the polynomial $P(s)$ is n, and the order of $Q(s)$ is m. In deriving the Nyquist criterion, the orders are restricted to the case $n \leq m$, such that

$$\lim_{s \to \infty} G(s)H(s) = 0 \quad \text{or a constant} \tag{16-41}$$

It is important to distinguish the various poles and zeros. They are tabulated as follows:

$s_1, s_2 \ldots s_n$ are *zeros* of $[1 + G(s)H(s)]$.
$s_a, s_b \ldots s_m$ are *poles* of $[1 + G(s)H(s)]$.
$s_a, s_b \ldots s_m$ are *also* the *poles* of $G(s)H(s)$.
$s_\alpha, s_\beta \ldots s_\omega$ are the *zeros* of $G(s)H(s)$.

The $s_1, s_2, \ldots, s_n$ roots are of vital concern to us because these *zeros* are zeros of the equation $1 + GH = 0$, which is the characteristic equation of the closed-loop system. These roots must not have positive real parts for the system they represent to be stable. Note that the *zeros* of $(1 + GH)$ are, by Eq. 16-38, the *poles* of (V_2/V_1).

In studying stability, our specific interest is the zeros of the polynomial $(1 + GH)$ with positive real parts. This suggests that we choose a contour in the s plane to include the entire right half plane as shown in Fig. 16-10. This contour will enclose all the zeros of interest. The contour is traced in the direction 1-2-3-4-1, starting at $s = -j\infty$, avoiding the origin $(s = 0)$ for the time being, and continuing to $s = +j\infty$, thence on a circle of infinite radius to the point of beginning. This contour is traversed in a clockwise (or negative) direction. The contour in the s plane can be mapped in either the $(1 + GH)$ plane or the GH plane (the simple relationship between these mappings was considered in Eq. 16-30). If any poles or zeros of $(1 + GH)$ are encircled in the right half of the s plane, then (1) the locus in the $(1 + GH)$ plane will encircle the origin, or (2) the locus in the GH plane will encircle the point $(-1 + j0)$.

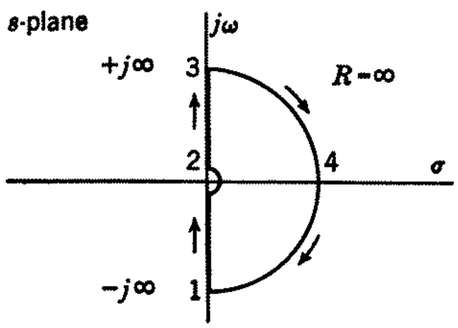


Fig. 16-10. Contour in the s plane.

Let Z = the zeros of $(1 + GH)$ with positive real parts, P = the poles of $(1 + GH)$ with positive real parts (*also* the poles of (GH) with positive real parts), R = the net counterclockwise encirclements of the point $(-1 + j0)$ in the (GH) plane or the origin in the $(1 + GH)$ plane. Then

$$R = P - Z \tag{16-42}$$

Since Z, the zeros of $(1 + GH)$ and the poles of V_2/V_1 with positive real parts, *must be equal to zero* for the system to be stable, the system with the characteristic equation $(1 + GH) = 0$ *is stable if and only if*

$$R = P \tag{16-43}$$

In most cases $P = 0$, and the criterion reduces to the requirement that $R = 0$ for stability.

To apply the Nyquist criterion, plot the $G(s)H(s)$ locus for the range of frequencies, $-\infty < \omega < \infty$. If R is the net counterclockwise encirclements* of the point $(-1 + j0)$ and P is the number of poles of $G(s)H(s)$ with positive real parts (and so in the right half plane), the system is stable if and only if $R = P$.

We have thus far avoided any problems that might arise because of a pole of $G(s)H(s)$ at the origin or several poles at the origin. Actually, there is a practical matter involved in taking into account these poles at the origin deserving of special attention. To illustrate the problem, consider a transfer function,

$$G(s)H(s) = \frac{K}{s(sT + 1)} \tag{16-44}$$

which is plotted in Fig. 16-11 for frequencies in the range $-\infty < \omega < +\infty$. The plot is complete except for one detail. The points $(+0)$ and (-0) should be joined together (as the same point). If the locus closes through the right half plane, the system is stable, since $R = 0$; however, if the locus closes the other direction in the left half plane, then $R = 1$ and the system is unstable. This is, as we see, a vital point.

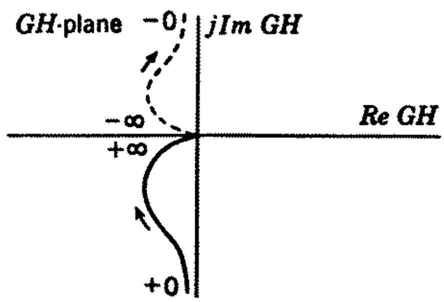


Fig. 16-11. Plot of $G(j\omega)H(j\omega)$.

As s becomes small, only the pole at the origin has an effect on the transfer function $G(s)H(s)$. Thus for small s, the transfer function can be written

$$G(s)H(s) = \frac{K}{s^n} \tag{16-45}$$

where n is the order (or multiplicity) of the poles at the origin. For

* To find the value of R, imagine a phasor with one end securely tied to the point $(-1 + j0)$ pointing away from this point. Let the end of this phasor trace the locus starting at $-\infty$ through -0 and $+0$ finally ending at $+\infty$. Count the net number of counterclockwise rotations of this phasor. This is the value of R. A clockwise rotation is designated by a negative number for R.

the semicircular path shown in Fig. 16-12, the equation of the s plane phasor locus is

$$(s - 0) = \delta e^{j\theta} \tag{16-46}$$

where δ is the radius of the semicircle and θ is the angle of the phasor $(s - 0)$ directed from the origin to a point on the circle. As $\delta \to 0$, the transfer function has the limiting value

$$\lim_{\delta \to 0} G(s)H(s) = \lim_{\delta \to 0} \frac{1}{\delta^n} e^{-jn\theta} = \infty e^{-jn\theta} \tag{16-47}$$

Hence as θ shown in Fig. 16-12 varies from $-\pi/2$ to $+\pi/2$, the phase of $G(s)H(s)$ ranges from $n\pi/2$ to $-n\pi/2$. In summary, the n poles at the origin in the transfer function $G(s)H(s)$ cause $(n/2)$ *clockwise* rotations at infinite radius of the phasor locus of $G(j\omega)H(j\omega)$.

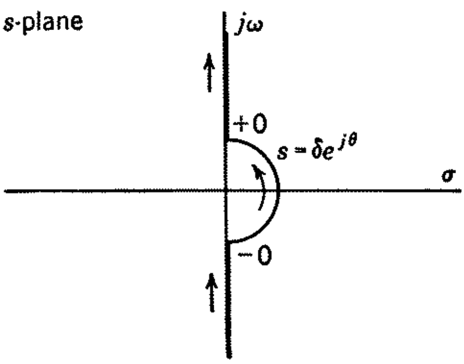


Fig. 16-12. Path at origin.

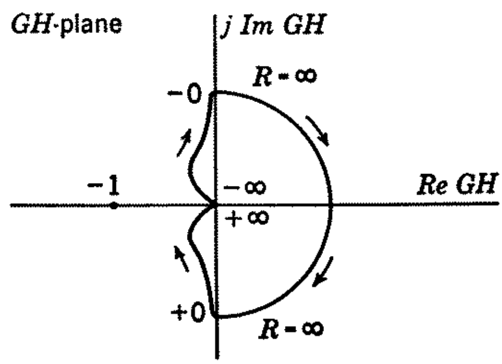


Fig. 16-13. Nyquist plot completed.

Applying this rule to the example of Eq. 16-44, we see that $n = 1$ causes $\frac{1}{2}$ clockwise rotation of the phasor locus of $G(j\omega)H(j\omega)$ in going from $s = -0$ to $s = +0$. Figure 16-11 is completed in Fig. 16-13.

In making the Nyquist plot, only positive values for ω need be considered. Because the real part of $G(j\omega)H(j\omega)$ is even and the imaginary part odd, it follows that

$$\begin{aligned} \text{Im}\, G(-j\omega)H(-j\omega) &= -\text{Im}\, G(+j\omega)H(+j\omega) \\ \text{Re}\, G(-j\omega)H(-j\omega) &= +\text{Re}\, G(+j\omega)H(+j\omega) \end{aligned} \tag{16-48}$$

The plot for negative values of ω can be made by reflecting the plot for positive frequency upon the real axis of the GH plane.

If the transfer function $G(s)H(s)$ has no poles in the right half plane (and the Routh or Hurwitz criteria can be used to advantage in making this determination) such that $P = 0$ in Eq. 16-43, a rule of thumb may be used to advantage. Trace ("walk") from $\omega = 0$ to $\omega = +\infty$ on the Nyquist plot. If the point $(-1 + j0)$ is on the *right* at the point (ω) of nearest approach of $G(j\omega)H(j\omega)$ to $(-1 + j0)$, the system is *unstable;* if on the *left* the system is *stable* $(P = 0$ only).

Several examples will next be considered to illustrate the application of the Nyquist criterion to the studies of system stability.

Example 4

For this example, consider the transfer function

$$G(j\omega)H(j\omega) = \frac{K}{j\omega T + 1} \tag{16-49}$$

The Nyquist plot is shown in Fig. 16-14, where K is the diameter of the circle. No matter how large K becomes, the locus cannot encircle the point -1. Hence the transfer function represents an *unconditionally stable system.*

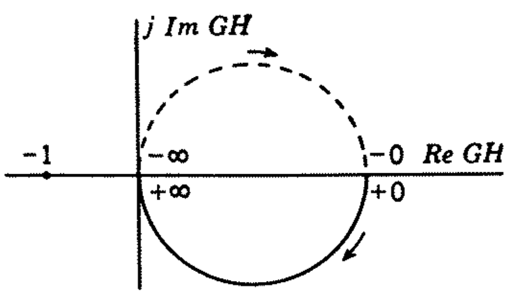


Fig. 16-14. $GH = K/(sT + 1)$ plotted.

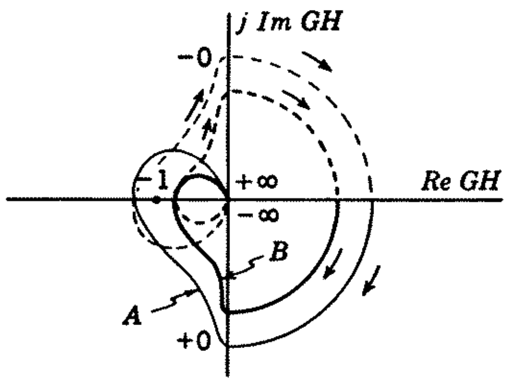


Fig. 16-15. Nyquist plot of Eq. 16-50.

Example 5

Let the locus plotted as Fig. 16-13 serve as a second example. Again, it is impossible for the locus to encircle the point -1 for any positive value of gain *except* infinite gain.

Example 6

The transfer function

$$G(j\omega)H(j\omega) = \frac{K}{j\omega(j\omega T_1 + 1)(j\omega T_2 + 1)} \tag{16-50}$$

is shown in Fig. 16-15 for two values of the constant K. For a value of K such that curve B results, the system is stable, since $R = 0$. However, if the value of K is increased to give the curve marked A, then $R = -2$, and since $P = 0$ by inspection of Eq. 16-50, the system is unstable. Such a system is described as a *conditionally* stable system.

Example 7

The exact nature of the transfer function for the plots shown in Fig. 16-16 is not given, but it is known that $P = 0$ for both cases. The locus of Fig. 16-16(a) represents a conditionally stable system. The locus shown in Fig. 16-16(b) is similar to that of (a) except the shape of part of the locus has been altered. Since there are no net rotations

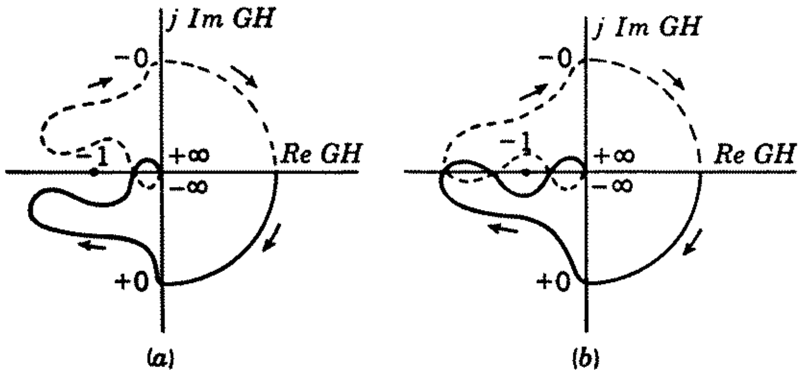


Fig. 16-16. Loci for conditionally stable systems: (a) and (b) $P = R = Z = 0$.

about the point -1, the system is stable. However, if the gain either increases or decreases corresponding to a shift of the -1 point into one of the two other loops, the system becomes unstable. This locus represents a system that is *conditionally stable.*

Example 8

Figure 16-17(a) shows a system having a transfer function

$$G(s)H(s) = \frac{K}{s(Ts - 1)} \tag{16-51}$$

For this locus, $P = 1$, $R = -1$ (one clockwise rotation) and so there

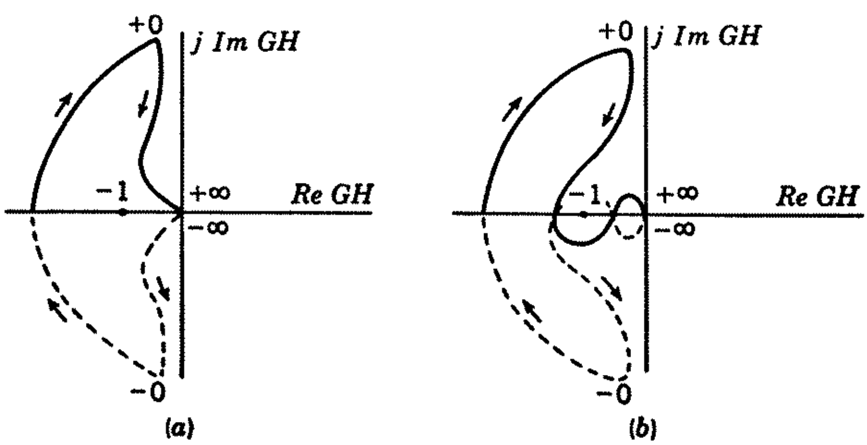


Fig. 16-17. Loci for transfer functions having poles with positive real parts: (a) $P = 1$, $R = -1$, $Z = 2$; (b) $P = 1$, $R = 1$, $Z = 0$.

are two zeros of $(1 + GH)$ with positive real parts and the system will be unstable for any value of gain. Such a system can be designated as *unconditionally unstable.*

Example 9

The locus plotted in Fig. 16-17(b) comes from a transfer function with one pole with a positive real part. For this particular system, however, the locus encircles the point -1 once in a counterclockwise direction such that $Z = 0$, and the system is stable. With the loop closed, this system is stable. However, with the loop open, the system is unstable. This open-loop instability is, of course, caused by another feedback path within the "open-loop" (as discussed at the beginning of the chapter).

FURTHER READING

For an interesting comparison of the various stability criteria, see F. E. Bothwell's article "Nyquist diagrams and the Routh-Hurwitz stability criterion," *Proc. IRE*, **38,** 1345 (1950). Bothwell points out that Nyquist, Routh, and Hurwitz all employed essentially the same procedures in their original writings. The articles of these three authors are: E. J. Routh, *Dynamics of a System of Rigid Bodies* (Macmillan & Co., Ltd., London, Part II, 1905), Chap. 6; H. Nyquist, "Regeneration theory," *Bell System Tech. J.*, **11,** 126 (1932); and A. Hurwitz, "Ueber die Bedingungen unter welchen eine Gleichung nur Wurzeln mit negativen reelen Teilen besitzt," *Math. Ann.*, **46,** 273 (1895). For additional reading on the Routh-Hurwitz criterion, see Guillemin, *The Mathematics of Circuit Analysis* (John Wiley & Sons, Inc., New York, 1949), pp. 395–409, or Tuttle, *Network Synthesis*, 2 vols. (John Wiley & Sons, Inc., New York, in preparation). In addition, see Chesnut and Mayer, *Servomechanisms and Regulating System Design* (John Wiley & Sons, Inc., New York, 1951), pp. 124–156.

For an interesting explanation of the process of organic evolution in terms of feedback system concepts and language, see p. 126 of Homer Jacobson, "Information, reproduction and the origin of life," *American Scientist*, **43,** 119–127 (1955).

PROBLEMS

16-1. Determine by means of Routh's stability criterion whether the systems having the following characteristic equations are stable or not. (a) $4s^3 + 7s^2 + 7s + 2 = 0$. (b) $2s^3 + s^2 - 5s + 2 = 0$. (c) $s^3 + 3s^2 + 4s + 1 = 0$. (d) $5s^3 + s^2 + 6s + 2 = 0$. *Answers.* (a) stable, (d) not stable.

16-2. Repeat Prob. 16-1 for the characteristic equations: (a) $5s^4 + 6s^3 + 4s^2 + 2s + 3 = 0$. (b) $s^4 + 3s^3 + 2s^2 + s + 1 = 0$. (c) $2s^4 + 3s^3 + 6s^2 + 7s + 2 = 0$. *Answer.* (a) not stable.

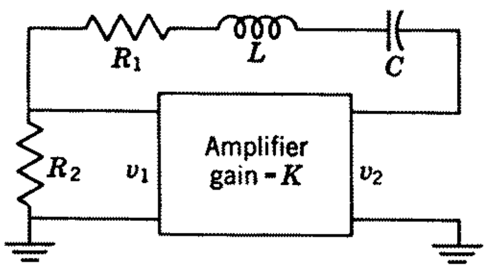
16-3. Repeat Prob. 16-1 for the characteristic equations: (a) $720s^5 + 144s^4 + 214s^3 + 38s^2 + 10s + 1 = 0$. (b) $25s^5 + 105s^4 + 120s^3 + 120s^2 + 20s + 1 = 0$. *Answer.* (a) stable.

16-4. A system has the characteristic equation,

$$s^3 + 5s^2 + Ks + 1 = 0$$

(a) Using the Routh criterion, determine the range of the values of K that will make the system stable. (b) Investigate system stability when $K = \frac{1}{5}$. Discuss your results.

16-5. The feedback system shown in the accompanying figure has been analyzed by J. F. Koenig in his paper "Stability diagrams for



Prob. 16-5.

feedback systems," *AIEE Conference Paper*, Baltimore, Oct., 1950. (a) Using Routh's criterion, find the relationship that must exist between R_1, R_2, and K for the system to be stable. (b) For the system to oscillate without damping, what must be the relationship between R_1, R_2, and K? From this equation, plot K as a function of R_1/R_2. On the same plot, show regions of stability and instability.

16-6. For a fourth-order characteristic equation,

$$a_0s^4 + a_1s^3 + a_2s^2 + a_3s + a_4 = 0$$

find a set of rules, similar to those given in Example 3, by Routh's criterion, that will insure that all roots of the characteristic equation have negative real parts such that the equation will represent a stable system. Assume that all coefficients must be positive.

16-7. Repeat Prob. 16-6 for a fifth-order characteristic equation.

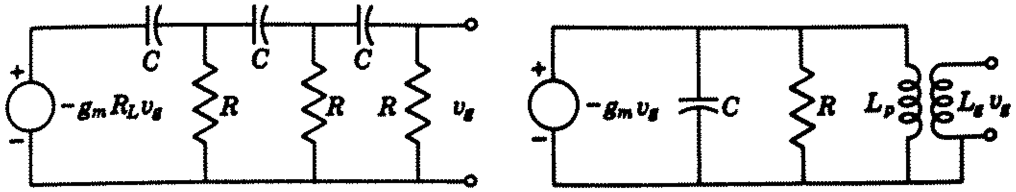
16-8. Classify the polynomials given in Prob. 16-1 as Hurwitz (having all roots in the left half plane) or not.

16-9. Apply the Hurwitz test to the polynomials given in Prob. 16-2.

16-10. Test the polynomials given in Prob. 16-3 using the Hurwitz criterion.

16-11. Rework Prob. 16-4 making use of the Hurwitz criterion.

16-12. The figure below shows an equivalent circuit of a *phase-shift oscillator* first described by Ginzton and Hollingsworth in *Proc. IRE*, **29,** 43 (1941). Show that the necessary condition for oscillation is $g_m R_L \geq 29$. (b) Show that the frequency of oscillation when $g_m R_L = 29$ is $\omega_0 = 1/\sqrt{6}\, RC$.



Prob. 16-12. **Prob. 16-13.**

16-13. The tuned-plate oscillator shown in Fig. 16-1 of the text may be represented by the equivalent circuit shown in the accompanying schematic. Show that the smallest value that the tube constant g_m can have if oscillations are to start is $g_m = L_p/MR$ and that the frequency of oscillation under this condition is $\omega_0 = 1/\sqrt{L_p C}$.

16-14. Consider the following transfer functions:

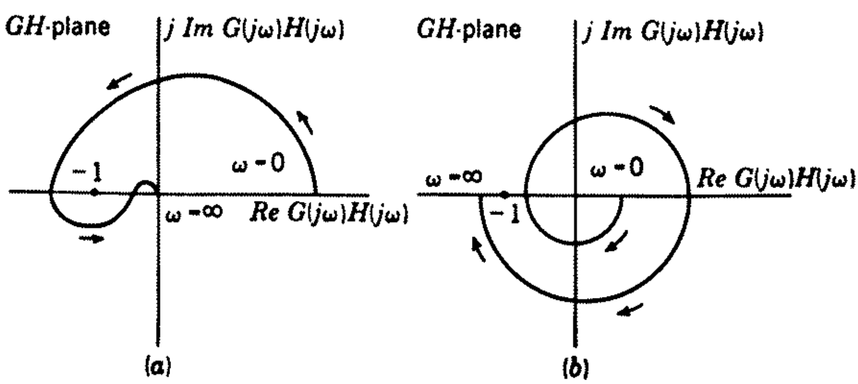
$$\text{(a)}\quad G(s)H(s) = K\frac{s-1}{s+1}$$

$$\text{(b)}\quad G(s)H(s) = K\frac{s+1}{s-1}$$

$$\text{(c)}\quad G(s)H(s) = \frac{K}{s(1+0.05s)}$$

For each of these functions: (a) plot $G(j\omega)H(j\omega)$ in the complex GH-plane from $\omega = 0$ to $\omega = \infty$ with $K = 1$. (b) Determine the range of values of K that will result in a stable system by means of the Nyquist criterion.

16-15. (a) The locus of $G(j\omega)H(j\omega)$ shown in the figure is for a system with two poles of $G(s)H(s)$ with positive real parts. Apply the



Prob. 16-15.

Nyquist criterion to determine if the system is stable with the loop closed. (b) The $G(j\omega)H(j\omega)$ locus shown in (b) is known to represent a system with no poles of $G(s)H(s)$ in the right half plane. Will the system be stable with the loop closed?

16-16. The following transfer functions relate to two servomechanisms:

$$\text{(a)}\quad G(s)H(s) = \frac{3000}{s^2(1 + 0.004s)}$$

$$\text{(b)}\quad G(s)H(s) = \frac{1500(1 + 0.04s)}{s^2(1 + 0.004s)^2}$$

Investigate the closed-loop stability of system 1 and system 2 by means of the Nyquist criterion. *Answer.* System (a) is unstable.

16-17. In the network of Prob. 16-12, let $R = 1$ ohm and $C = 1$ farad (these are normalized values). Plot the Nyquist diagram for (a) $g_mR_L = 10$ and (b) $g_mR_L = 40$. Which of the two conditions will represent a stable (nonoscillating) system?

16-18. A certain closed-loop system is described by the transfer functions

$$G(s) = \frac{K}{s(T_1s + 1)(T_2s + 1)} \quad \text{and} \quad H(s) = 1$$

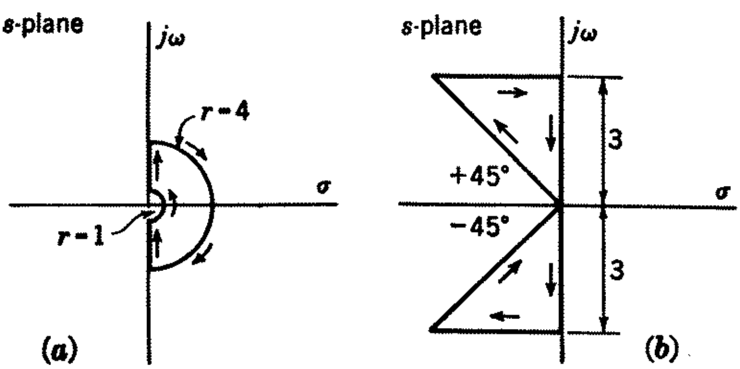
Determine the maximum value of K that may be used without making the system unstable.

16-19. (a) Consider two functions:

$$G(s) = \frac{s + 2}{s(s - 3)}, \qquad H(s) = 1$$

$$G(s) = \frac{s - 2}{s(s - 3)}, \qquad H(s) = 1$$

Plot these two functions for values of s along the contour shown for the s plane in (a) of the figure. Discuss how your results relate to the Nyquist criterion.



Prob. 16-19.

(b) Consider two functions:

$$G(s) = \frac{1}{s^2 + 2s + 5}, \qquad H(s) = 1$$

$$G(s) = \frac{1}{s^2 + 4s + 5}, \qquad H(s) = 1$$

Plot these functions for values of s along the contour shown in (b) of the figure. Does this suggest how the Nyquist criterion might be generalized as a criterion for other than poles in the right half plane? Discuss.

INDEX