

ED 021 737

24

SE 004 635

By- LePage, Wilbur R.; Balabanian, Norman
INTRODUCTION TO ELECTRICAL SCIENCE.
Syracuse Univ., N.Y. Dept. of Electrical Engineering.
Bureau No- BR-5-0796
Pub Date 64

Contract- OEC-4-10-102

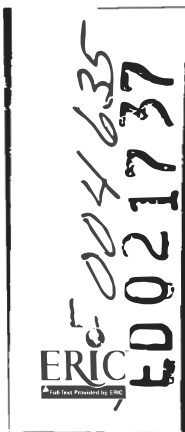
Note- 326p.

EDRS Price MF-\$1.25 HC-\$13.12

Descriptors- *COLLEGE SCIENCE, ELECTRICITY, *ENGINEERING EDUCATION, *INSTRUCTIONAL MATERIALS,
PHYSICAL SCIENCES, TEXTBOOKS, UNDERGRADUATE STUDY

Identifiers- United States Office of Education

This text (in mimeographed form) was developed under contract with the United States Office of Education and is intended as material of a first course in the electrical engineering sequence. Introductory concepts such as charge, fields, potential difference, current, and some of the basic physical laws are presented in Chapter I. Subsequent chapters develop concepts of (1) resistive and diode networks, (2) electrostatics, (3) electromagnetism, (4) steady state current analysis, (5) natural response of electric circuits, (6) electric motors, (7) semiconductor theory, (8) transistor amplifiers, and (9) magnetic coupling. (DH)



INTRODUCTION TO
ELECTRICAL SCIENCE

by

Wilbur R. LePage
and
Norman Balabanian

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

Electrical Engineering Department
Syracuse University

Contract No. OE 4-10-102
U.S. Office of Education

Copyright © 1964

"PERMISSION TO REPRODUCE THIS
COPYRIGHTED MATERIAL HAS BEEN GRANTED

BY *Norman Balabanian*

TO ERIC AND ORGANIZATIONS OPERATING
UNDER AGREEMENTS WITH THE U.S. OFFICE OF
EDUCATION. FURTHER REPRODUCTION OUTSIDE
THE ERIC SYSTEM REQUIRES PERMISSION OF
THE COPYRIGHT OWNER."

INTRODUCTION TO ELECTRICAL ENGINEERING SCIENCE

INDEX

- Ampere, 1-8, 4-28
- Ampere's Law, 4-9
- Armature, 4-26
- Biot-Savart Law, 4-15
- Capacitance, 3-16, 17
- Capacitors, 3-25
- Charge, 1-1, 1-22, 3-3
 - test, 1-3
 - point, 3-3
 - stationary test, 3-6
 - bound, 3-24
 - free, 3-24
- Coercivity, 4-25
- Conductance, 12-17
- Conducting materials, 3-14
- Coupling coefficient, 4-44
- Current, 1-8, 9
 - branch, 2-10
 - dividers, 2-3
 - loop, 2-10
 - source deactivation, 2-16
 - source equivalent, 2-5
- Diamagnetic material, 4-23
- Dielectric constant, 3-21
 - displacement vector, 3-25
 - breakdown strength, 3-27
- Diode circuits, 2-16
 - ideal, 2-17
 - model, 2-19
- Dipole, 3-5
 - field plot, 3-6
- Electromagnetism, 4-1
- Electromotive force, 4-12
- Electrostatic, 1-6, 7, 3-1
- Energy, 1-20
 - storage in elec. field, 3-33
 - storage in magnetic field, 4-54
- Farads, 3-17
- Faraday's law, 4-18
- Ferromagnetic material, 4-20, 23
- Field, conservative, 3-14
 - electric, 1-2, 3, 4, 3-1, 4, 6, 7
 - intensity, 4-22
 - magnetic, 1-7, 4-6
- Flux density, 4-5, 15
 - leakage, 4-33
 - linkage theorem, 4-51
 - remnant, 4-25
- Fringing, 4-32
- Gauss' law, 3-8, 11, 14, 16
- Gaussian surface, 3-9, 10, 11, 15, 16
- Henrys, 4-40
- Hysteresis
 - dielectric, 3-28
 - loss, 3-34
 - magnetic, 4-25
- Inductance
 - mutual, 4-43
 - self, 4-40
- Isotropic, 4-23, 25-29
- Joule, 1-20
- Kirchhoff's laws, 1-10, 11, 12, 13, 14, 15, 16
- Lenz's law, 4-20
- Load line, 2-20
- Loop equation, 2-8
- Lorentz force, 4-11
- Mmf, 4-28
- Node equations, 2-11
 - datum, 2-11
 - voltages, 2-11
- Norton equivalent, 2-7
- Ohm's law, 1-17, 2-2, 3
- Parallel circuits, 2-3
- Paramagnetic material, 4-23
- Permeability, 4-4
- Permeance, 4-38
- Permittivity, 3-4
 - relative, 3-21
- Polarization, 3-24
- Potential difference, 1-3, 5, 3-12, 14
- Power, 1-20
- Resistance, 1-17
 - forward, 2-17
 - reverse, 2-17
- Resistivity, 1-17, 19
- Saturation, 4-21
- Series circuits, 2-1
- Superposition principle, 2-15
- Thevenin's theorem, 2-7
 - equivalent voltage, 2-7
 - equivalent resistance, 2-7
- Transformer, 4-26, 44
- Transients
 - R-C circuits, 3-29
 - R-L circuits, 4-47
- Voltage, 1-3, 7, 20
 - divider, 2-1
 - motional induced, 4-12
 - source deactivation, 2-16
 - source equivalent, 2-5
- Watt, 1-21
- Weber, 4-5

Chapter 1

INTRODUCTORY CONCEPTS

The notion of electric charge has grown as a consequence of man's observations of a large number of phenomena, from lightning to the electric shock obtained by touching a metallic object after walking on a thick rug. Electric charge, as a property of electrons and protons, is a major building block of nature. The atomic theory of matter, at least in its simple form, is known to every schoolboy.

It has been known for more than two centuries that objects, which we now describe as being electrically charged, exert forces of attraction or repulsion on each other. The far-reaching consequences of these forces pervade almost all aspects of modern life -- from electrically-powered industrial and home machinery to telephonic communication, to radio and television, to devices for medical diagnosis and treatment, to the guidance and control of space vehicles. A highly satisfactory explanation of these forces has been achieved through inventing the concept of electric charge.

Quantitative laws have been discovered which relate the behavior of charged bodies to their configuration, their positions and orientations, and their states of motion. Application of these laws by engineers has led to the development and design of a host of useful accomplishments -- from communication using laser beams to microwave broiling of hamburgers.

1-1. Coulomb's Law

The basic amount of charge is that of an electron. However, this is so small compared to charges whose effects are observable in the world of the laboratory that it is not chosen as a unit of measure. In the rationalized MKSC (meter-kilogram-second-coulomb) system of measurements the unit of charge is called the coulomb, named after a Frenchman who first gave a quantitative relationship for the force of attraction or repulsion of one electric charge on another. This quantitative relationship, called Coulomb's law, was also named after him. It states that the algebraic value of the force on a small stationary electric charge Q_1 due to another small stationary electric charge Q_2 is proportional to Q_1 and Q_2 and inversely proportional to the square of the distance r between the charges.

$$F_{21} = K \frac{Q_1 Q_2}{r^2} \quad (1-1)$$

The constant of proportionality is dependent on the medium in which the charges are located. (We assume that they are located in a homogeneous medium.)

In the rationalized MKSC system the constant is chosen so that the force is $1/4\pi\epsilon$ newtons for two identical charges of 1 coulomb each separated by a distance of 1 meter in vacuum. ϵ is called the permittivity of the medium. When the medium is vacuum the symbol is written ϵ_0 and has the value 8.85×10^{-12} .*

Thus, Coulomb's law is written

$$F_{21} = \frac{Q_1 Q_2}{4\pi\epsilon r^2} \quad (1-2)$$

This expression represents the algebraic value of the force. But force is a vector quantity and has direction also. The direction of the force exerted on a charge Q_1 by charge Q_2 is along a line directed from Q_2 to Q_1 . It is not necessary to specify whether Q_1 and Q_2 are positive or negative; it is only necessary to treat the charges as algebraic quantities. Note the order of the subscripts on F_{21} : the first subscript specifies the charge, Q_2 , which is exerting the force; the second subscript specifies the charge, Q_1 , on which the force is exerted.

When more than two charges are present in a region the total force on a given charge is the resultant of the individual forces due to all of the other charges. Since each of these individual forces is a vector, the resultant must be composed as a vector sum. We shall designate vectors by putting an arrow above the symbol; a vector force is designated $\vec{F}$.

1-2. Electric Field

If we single out a charge Q_0 and consider the force $\vec{F}$ on this charge due to all other charges $Q_1, Q_2, \dots, Q_n$, we will have an expression

$$\vec{F}_0 = \vec{F}_{10} + \vec{F}_{20} + \dots + \vec{F}_{n0} \quad (1-3)$$

* Actually ϵ_0 is related to the speed of light, c , by the relationship $4\pi\epsilon_0 = 10^7/c^2$. Recognition of this relationship was an important factor in identifying electromagnetic radiation (radio waves) as the same thing as light.

1 where $\vec{F}_{10}$ is the force on Q_0 due to charge Q_1 , $\vec{F}_{20}$ is the force on Q_0 due to Q_2 , etc. Each of these partial forces is proportional to Q_0 , according to Coulomb's law. Hence Q_0 can be factored from each term of the sum and the result can be written

$$2 \quad \vec{F}_0 = Q_0 \vec{E} \quad (1.4)$$

3 where $\vec{E}$ is the vector summation of all the terms after Q_0 has been factored. We see that $\vec{E} = \vec{F}_0/Q_0$, the force on charge Q_0 per unit charge. It does not depend on Q_0 ; by Coulomb's law it depends on all the other charges and on their distances from Q_0 , as well as on the permittivity of the medium containing the charges. We call it the electric field. The presence of this
4 electric field is not dependent on the presence of Q_0 , which is simply a test charge placed somewhere to see if there is a force on an electric charge at that point. If we were to double the value of the test charge Q_0 , the value of the force on it would be doubled, by Coulomb's law, assuming that this
5 very doubling of the charge did not change the positions of all other charges. But a doubled force divided by a double charge leads to the same value of $\vec{E}$. If the location of Q_0 is changed without changing the locations of any other charge, the value of r to use in Coulomb's law for each charge will change
6 and, hence, the electric field will change. In a region containing charges, then, the electric field varies from point to point.

1-3. Potential Difference and Voltage

7 Now consider a charge Q_0 placed in an electric field. (This is a shorthand way of saying "placed in a region in which there is a force on an electric charge at any point.") Since there is a force on Q_0 , it will tend to move unless there is a restraining force preventing the charge from moving.
8 Suppose there is no other force except that due to the field $\vec{E}$, and the charge moves a certain distance. Work must be done in moving the charge and the amount of this work equals the distance moved times the component of force along the line of motion. Consider the diagram in Figure 1-1. Suppose
9 Q_0 moves a distance l from point P_1 to P_2 in a straight line.

1-4

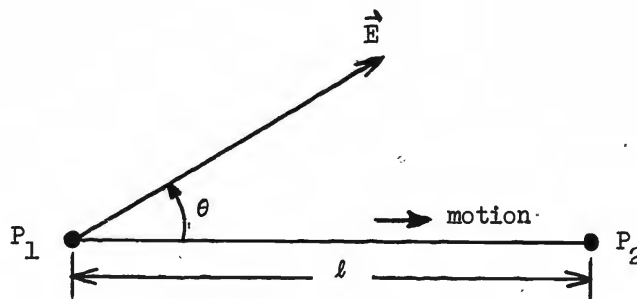


Figure 1-1.

We might be tempted to say the work done is l times $Q_0 E$ (the magnitude of the vector) times $\cos \theta$, where $Q_0 E \cos \theta$ is the component of the force on Q_0 along the line of motion. But since $\vec{E}$ may not have the same direction from point to point as the charge moves, its component ($E \cos \theta$) along the line of motion might change from point to point. To overcome the difficulty, we must find the differential work done in moving along a small displacement, $d\ell$, assuming that the direction of $\vec{E}$ does not change over this small distance. Then, we must integrate to find the total work. Thus,

$$\begin{array}{l} \text{Work done by the electric} \\ \text{field in moving a charge} \\ \text{from } P_1 \text{ to } P_2 \end{array} = \int_{P_1}^{P_2} Q_0 E \cos \theta \, d\ell \quad (1-5)$$

where E and $\cos \theta$ can change from point to point along the path.

Now, the potential energy that is available to do work by virtue of the fact that Q_0 is in an electric field has been decreased. (This is analogous to a body falling down a hill. The gravitational force exerted on the body when it is at the top of the hill is doing work as it falls. But all the while, the potential energy which was available is decreasing until, when the body reaches the bottom of the hill, it cannot fall any more and its potential energy is reduced to zero.) Returning to the integral in Eq. (1-5), we see that the work done represents the decrease in potential energy as Q_0 moves from P_1 to P_2 . If we divide by Q_0 , we find

$$\begin{array}{l} \text{Decrease in potential energy per unit} \\ \text{charge in moving a charge from } P_1 \text{ to } P_2 \end{array} = \int_{P_1}^{P_2} E \cos \theta \, d\ell \quad (1-6)$$

This quantity is extremely important. In order to avoid having to say all those words on the left side every time we want to refer to it, it is given a name; it is called the potential difference. More specifically it is the potential decrease from P_1 to P_2 .

Normally it is a difficult job to calculate the magnitude and direction of the electric field at all points along a path spanning two points between which it is desired to know the potential difference. Fortunately, in much of the study of electrical engineering it is not necessary to do this since other relationships can be used besides this integral.

We shall use the term voltage to stand for the potential decrease as calculated from the above integral. By definition, then, the voltage across two points P_1 and P_2 is the decrease in potential energy per unit charge when a charge moves from P_1 to P_2 .

Although the preceding discussion was carried out in terms of the electric field as the agency which exerted force and did work, the resulting definition of voltage is more general; the work can be done by the field or by an external agency, like gravity. Hence, the decrease in potential energy may turn out not to be an actual decrease at all, but an increase. Equation (1-6) takes this into account, as the resultant voltage is an algebraic quantity and may vary in sign. That is, when the potential energy per unit charge is actually increased, the decrease in potential energy is negative.

We shall use the symbol, v , to stand for voltage. Since the definition involves two points there must be some way of designating from which point to which point the integration in Eq. (1-6) is to be carried out. In Figure 1-2 two points a and b are shown, the voltage between which is under discussion.

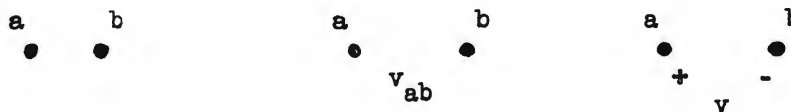


Figure 1-2.

If we go from a to b in the definition of voltage, we will label it v_{ab} . Thus, v_{ab} is the decrease in potential energy per unit charge when a charge

- 1 moves from a to b. In a particular situation this number may turn out to be positive or it may be negative. If the number is positive, it means that there is actually a decrease in potential energy in going from a to b. If negative, this means that there is actually an increase in potential energy in moving from a to b. In either case we get the required information about the state of potential energy.

Another way of designating from which point to which point in the integration of Eq. (1-6) is to use some marking, as in the third part of Figure 1-2.

- 3 A + sign is placed at the point from which we integrate and a - sign at the point to which we integrate. Thus, if we write v and put a plus sign at a and a minus sign at b we mean v_{ab} . (Of course, only the plus sign is enough; once you know which point carries the plus sign, then you know the other point carries the minus sign.)

- In discussing Coulomb's law we pointed out that the relationship holds for stationary charges. (The term electrostatics is used to designate this branch of the subject.) However, to determine the voltage we talked about moving a charge from one point to another. How can these two notions be reconciled? Actually, in moving a charge about in a field we must think of motions carried out with infinitesimally small velocities, otherwise kinetic energy will have to be considered and we no longer have an electrostatic situation. But just what is an infinitesimally small velocity? This is really hard to answer. By this phrase we shall mean, so small that the kinetic energy can be neglected compared with the potential energy involved. To imply that there is some non-zero velocity but that the results based on the static case still apply, the term quasi-static conditions is usually used. In everything we consider here we often use the term static although we shall be dealing with quasi-static conditions.

In discussing the moving of a charge from one point to another, no consideration was given to the path of motion. Thus, in Figure 1-3 a charge can move from

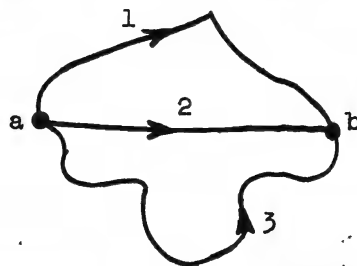


Figure 1-3.

- 1 a to b along either of the paths shown, or along any number of other paths.
 If the notion of voltage is to have a useful meaning, either one of two
 things must be true about the path taken in going from a to b; either the
 result must be independent of the path taken -- that is, it will be the
 2 same no matter which path is taken -- or a particular path is specified and
 we agree always to take that path.

- Although we shall not do so here, it is an easy matter to show that for
 the electrostatic case (charges stationary), the work done in moving a charge
 3 between two points is independent of the path taken. The same thing is true
 even in the case when charges are moving, provided that the resulting cur-
 rents are constant with time -- so-called direct current (dc). In these
 cases the term "voltage" will have an unambiguous meaning. (Note that in
 4 the case of gravitational force a similar result is obtained. That is, if
 we take a stone falling down a hill, its decrease in potential energy after
 falling a certain vertical distance is the same no matter how it bounded
 around in falling that distance.)

- 5 But we are also interested in situations where the currents do vary
 with time. In such cases the result of the integration in Eq. (1-6) will
depend on the path taken. But even in these cases we can make the term,
 "voltage", meaningful by agreeing as to the path taken. From previous studies
 6 you are no doubt familiar with the fact that electric currents are accompanied
 by magnetic fields. (Further discussion of this subject will take place in
 Chapter 4.) If the currents are time-varying, then so also will be the re-
 sulting magnetic field. In order to deal with such a situation, we assume
 7 that all time-varying magnetic fields are localized; that is, they are
lumped or concentrated in one or more devices, as shown in Figure 1-4, and
 the effects produced are accounted for at the terminals of the lumped device.

8

9

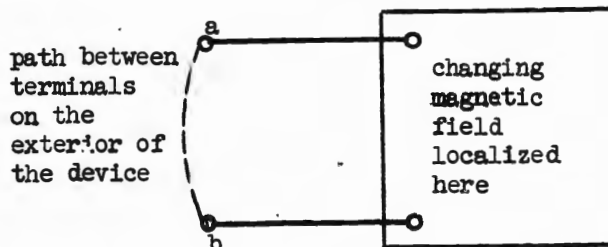


Figure 1-4.

- 1 We agree never to choose a path between the terminals a-b that goes inside the device, but always to take a path outside the device. With this agreement the notion of voltage between a and b will again be meaningful.

2 1-4. Current

- 3 If electric charges were all stationary, they would be of very little interest to anyone. They give rise to important observable electric phenomena when they are in motion. Electric charge in motion constitutes a current. Before we define current more carefully, note that both negative and positive charge can be moving, and that each can be moving in any direction.

- 4 In developing a definition of current, suppose we consider charges moving along a discrete path, like a conducting wire. Suppose further that only positive charges are moving and we count the amount of charge passing a particular cross section of the wire in a given time interval, say half an hour. We can then say how much charge came by in the half hour, but we won't know whether it came across faster over part of the time and slower over the rest of the time, or whether the rate at which the charge passed the cross section was the same for the whole interval. To remedy this situation we can make the observation interval smaller and smaller. Suppose we take the observation interval to be 2 seconds. We can add up the amount of charge coming by in each 2-second interval. The average rate of passing of the charge equals the amount per 2-second interval divided by 2. But still we can't say anything about the rate of flow within each interval, whether it speeded up or slowed down as the interval of time progressed from 0 to 2 seconds. Suppose we make the observation interval still smaller, say Δt , and we find the amount of charge passing a cross section during this interval to be Δq . The average rate over the time Δt at which charge is moving past a given cross section is then $\Delta q / \Delta t$. In the limit as we let the observation interval get smaller and smaller, we define the electric current i to be

$$8 \quad i = \frac{dq}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\Delta q}{\Delta t} \quad (1-7)$$

Current has the dimensions of coulombs per second and its unit is the ampere.

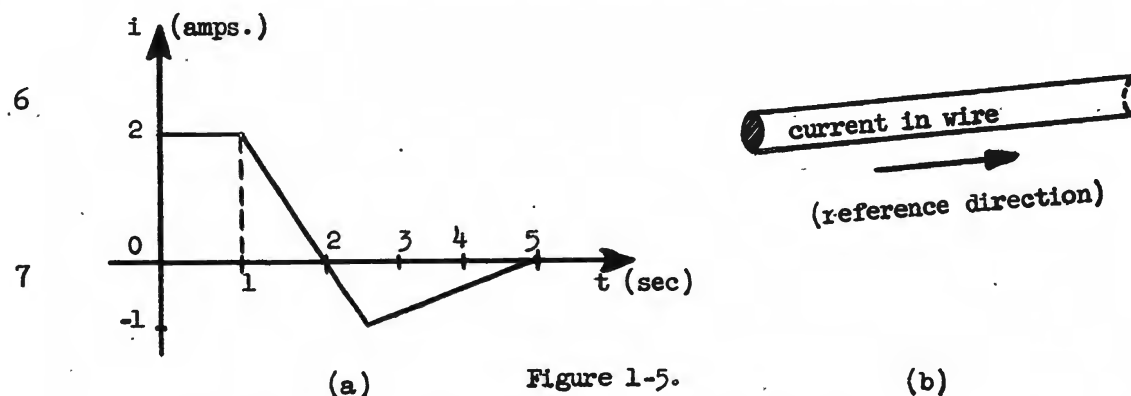
- 9 But, "moving past a cross-sectional area" can be in either one direction or the other. So we arbitrarily choose a particular direction as our reference. If positive charges go past a cross section in that direction we say the current is

1 positive; if they go in the opposite direction we say the current is negative.

2 But how about negative charges? Suppose a positive and a negative charge of equal magnitude move at the same rate in the same direction. The net flow of charge will be zero. This means that a negative charge moving in one
3 direction is equivalent to a positive charge of equal magnitude moving in the opposite direction. It is, therefore, unimportant to know whether the current is caused by the motion of positive charge or negative charge or both.

4 To summarize: when charges are flowing in a discrete path, we arbitrarily assign a direction to be the reference direction. We designate the current to be positive over a time interval if during this time positive charges are moving in the reference direction or negative charges are moving
5 opposite to the reference direction. The reference direction is indicated by an arrow drawn beside the path of charge flow.

6 In Figure 1-5 is shown a graph of current carried by a wire as a function of time. The reference direction of the current is shown as being to the right.



8 It is required to answer the following questions:

1. In what direction are positive or negative charges moving in the wire over the interval from 0 to 5 seconds?
2. How much charge and of what sign has passed a cross section of the wire toward the right after 1, 2 and 5 seconds?

(Try to answer these questions before reading on.)

1 1. The reference direction is to the right. When the current is positive,
positive charge is actually moving to the right. When the current is negative,
positive charge is actually moving to the left. Thus, from 0 to 2 seconds posi-
2 tive charge is actually moving to the right and from 2 to 5 seconds positive
charge is actually moving to the left. Since negative charge moving in one di-
rection is equivalent to an equal positive charge moving in the other direction,
the same result in each of the intervals can be described in terms of a negative
charge actually moving in a direction opposite to the direction in which the
3 positive charge is moving.

2. Since current is the time derivative of charge transferred, then the amount of charge must be equal to the integral of current.

$$4 \quad q = \int_0^t i \, dt \quad (1-8)$$

An integral can be interpreted as the area under a curve. Hence, after 1 second the area under the curve (Figure 1-5) is 2 times 1 ampere-seconds = 2 coulombs. It is positive, so a net positive charge has passed to the right, or a net negative charge has passed to the left. At 2 seconds, the current is zero but the area under the curve from 0 to 2 is not; it is $2 + 1 = 3$ coulombs. (Positive charge to the right or negative charge to the left.) From 2 to 5 seconds the current is negative, so the net area under the curve is being reduced; it is $3 - 3 \times 1/2 = 3/2$ coulombs.

In order to avoid having to say "positive charge to the left, negative charge to the right" in describing a current, we will henceforth assume that only positive charges are involved. So, when a current is positive, we will say that charge (meaning positive charge) is actually flowing in the reference direction; and when a current is negative, charge (meaning positive charge) is actually flowing opposite to the reference direction.

1-5. Kirchhoff's Laws

(Read the programmed booklet titled "Kirchhoff's Laws" for further instruction in the subject of this section.)

Kirchhoff's current law is a statement expressing the fact that electric charge cannot accumulate at a node or be generated there. Figure 1-6 shows four

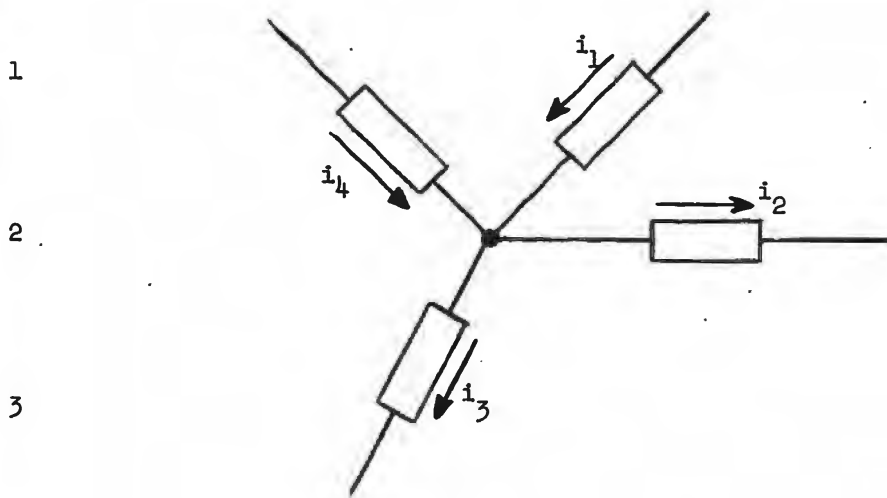


Figure 1-6.

branches joined at a node. The branch currents are each arbitrarily assigned a reference direction shown by an arrow alongside the branch.

Kirchhoff's current law can be stated in the following forms:

1. The algebraic sum of all current leaving any node (or junction) is zero at each instant of time (for the example, $-i_1 + i_2 + i_3 - i_4 = 0$); or

2. the algebraic sum of all currents entering any node is zero at each instant of time (for the example, $i_1 - i_2 - i_3 + i_4 = 0$); or

3. the sum of currents with references directed toward a node is equal at each instant of time to the sum of currents with references directed away from the node (for the example, $i_1 + i_4 = i_2 + i_3$).

In the first form, the node reference is 'leaving' while in the second form the node reference is 'entering'. In either case, if a branch reference coincides with the node reference, the corresponding term will carry a negative sign in the mathematical expression; if a branch reference is opposite to the node reference, the corresponding term will carry a negative sign.

Figure 1-7 shows a network having 4 nodes. (For simplicity the rectangles are omitted but every line shown represents a branch, not just a connection.) A reference direction for each branch current is arbitrarily

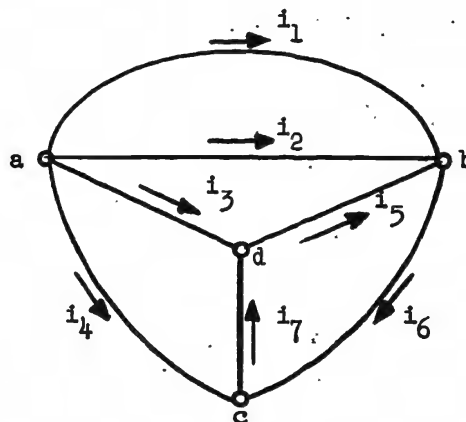


Figure 1-7.

picked. It is desired to apply Kcl and write a set of equations, one for each node. We shall do this, taking 'leaving' as the reference for each node. The result is the following. (You should write your own set of equations and check against these. It doesn't matter if your terms are not in the same order.)

$$\text{For node a: } i_1 + i_2 + i_3 + i_4 = 0$$

$$\text{For node b: } -i_1 - i_2 - i_5 + i_6 \neq 0$$

$$\text{For node c: } -i_4 - i_6 + i_7 = 0$$

$$\text{For node d: } -i_3 + i_5 - i_7 = 0$$

(1-9)

Additional Comments*

The objective in this problem was simply to give some practice in applying Kirchhoff's current law. However, having written the equations, it is possible to examine them to see if any additional information can be obtained. Notice how the equations have been written, with the terms for a given current appearing in a vertical column. Something very curious can be detected: each current appears twice in a column, once with a plus sign and once with a minus sign. Hence, if the equations are all added, the result will be identically zero! That is,

$$\text{Eq. a} + \text{Eq. b} + \text{Eq. c} + \text{Eq. d} = 0$$

(1-10)

*Subsections titled Additional Comment are for the purpose of introducing those who are interested to topics beyond the scope of the material for this course. No one is required to read these sections, but they will help any who do reach a deeper understanding.

1 From this it follows by solving for Eq. d that

$$\text{Eq. d} = -(\text{Eq. a} + \text{Eq. b} + \text{Eq. c}) \quad (1-11)$$

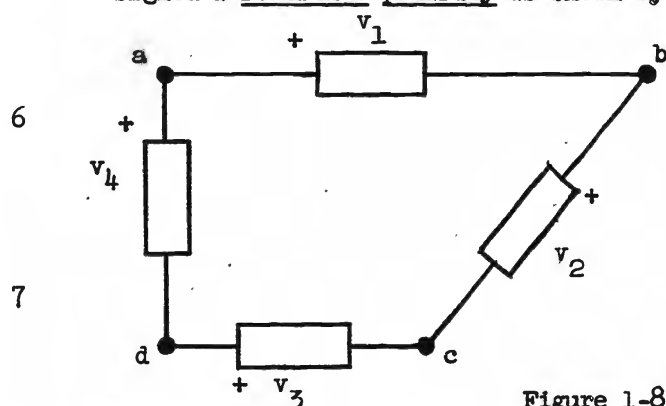
2 Actually, instead of Eq. d, we can solve for any one of the others and discover that

any one of the four equations = negative sum of all the others

3 which means that they are not all independent; if all but one are known, this one follows as the negative sum of the others. Verify this by adding the last three in Eq. (1-9) and comparing this sum with the first equation.

4 This result, which was found to be true for this example, is actually quite general and can be easily demonstrated. That is, for any network having N_n nodes, only $N_n - 1$ independent equations expressing Kirchhoff's current law can be written. There will be more about this in the next chapter.

5 We turn next to Kirchhoff's voltage law. Figure 1-8 shows four branches forming a closed path. The branch voltages have each been arbitrarily assigned a reference polarity as shown by the plus sign.



$$v_1 = v_{ab}$$

$$v_2 = v_{bc}$$

$$v_3 = v_{dc} = -v_{cd}$$

$$v_4 = v_{ad} = -v_{da}$$

Figure 1-8.

8 Kirchhoff's voltage law states that:

1. The algebraic sum of all voltages around any closed path in an electric network (traversed either clockwise or counterclockwise) is zero at each instant of time. (In the example, going clockwise, $v_1 + v_2 - v_3 - v_4 = 0$ or

2. around any closed path and at each instant of time, the sum of voltages with clockwise references is equal to the sum of voltages with counterclockwise references. (In the example, $v_1 + v_2 = v_3 + v_4$.)

- 1 In the first form, if a branch reference coincides with the loop reference
 (that is, the plus sign is encountered first when traversing the branch in the
 loop reference direction, which may be either clockwise or counterclockwise), the
 corresponding term will carry a positive sign in the mathematical expression;
 2 otherwise, a negative sign.

- Kirchhoff's voltage law can be taken as a basic postulate. But if we
 consider the electrostatic case only, Kvl follows from the definition of volt-
 age. This is easy to appreciate if we remember that in this case the 'voltage'
 3 is independent of the path. That is, in Figure 1-9 we can go from a to b along

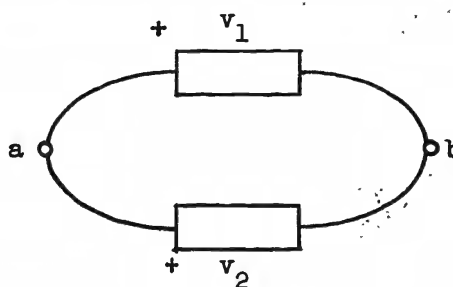


Figure 1-9.

- either the upper path or the lower path and the voltage (the decrease in poten-
 tial energy per unit charge) will be the same: $v_1 = v_2$ or $v_1 - v_2 = 0$, which
 6 is Kvl for Figure 1-9.

- For the general case of lumped networks we agree that in discussing voltage
 we will always take external paths between the terminals of the branches. As
 7 long as we do, it doesn't matter what combination of branches we traverse, the
 voltage between two points will be the same. Thus, in Figure 1-8 the voltage
 from a to b must be the same whether we go directly from a to b or go from a to d
 to c to b. Thus, $v_1 = v_4 + v_3 - v_2$, which can be written $v_1 + v_2 = v_3 + v_4$.

- 8 Note that the assumption of a lumped network means that there is no changing
 magnetic flux passing through the closed path in Figure 1-8.

- The diagram in Figure 1-10 shows a network having three closed paths, or
 loops, as shown by the dashed arrows. (To avoid confusion, the diagram is redrawn
 9 without the arrows.) A reference polarity for each branch voltage is arbitrarily

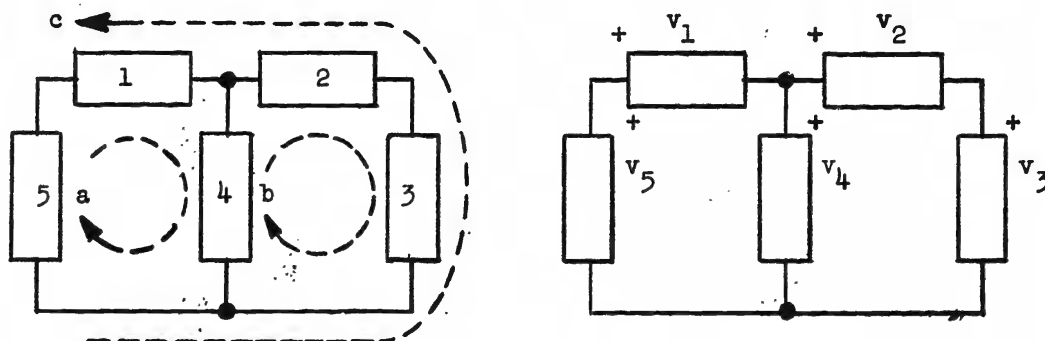


Figure 1-10

assigned. It is desired to apply Kirchhoff's voltage law to write a set of equations, one for each loop. (You should write your own set of equations before going on and check them against the ones below. Don't worry about the terms in your equations being in a different order from these.) Here is the result.

$$\text{loop a: } v_1 + v_4 - v_5 = 0$$

$$\text{loop b: } v_2 + v_3 - v_4 = 0 \quad (1-12)$$

$$\text{loop c: } -v_1 - v_2 - v_3 + v_5 = 0$$

Additional Comments

Let us again examine these equations for any additional information we can gather from them. Note again that each voltage appears twice in a vertical column, once with a plus sign and once with a minus sign. If the equations are all added, the result will, therefore, be identically zero. From this it again follows that any one of the equations can be obtained, once the other two are known. The third equation, for example, the one around the outside contour of the network, is just the negative sum of the other two. (Clearly, if this equation had been written in a clockwise sense instead of the other way, it would have been obtained as the positive sum of the other two equations.)

This result, unlike the corresponding one for Kirchhoff's current law, is not general. In more complicated networks, there are many more closed paths than there seem to be. For example, in Figure 1-11 only one additional branch has been added to the previous network. In addition to the previous

1-16

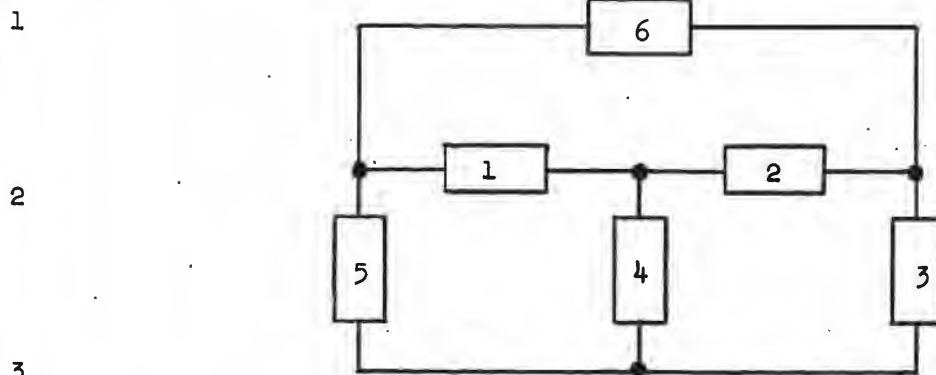


Figure 1-11

3 closed paths, there are now 4 more. Indicating these paths by listing the branches lying on them, these closed paths are 3-5-6, 1-2-6, 1-4-3-6 and 2-6-5-4. Hence, there are a total of 7 closed paths to which Kvl can be applied.

For a given network, it is easy to count the number of junctions to find how many total Kirchhoff current law equations can be written. Of these, all but one are independent. But the situation is different for the number of closed paths. In fact, there is no way of telling from the number of nodes or branches the total number of closed paths a network will have, short of actually finding them all. But fortunately we are not interested in the total number of Kvl equations in a network, only in the independent ones; and these it is possible to tell. If a network has N_n nodes and N_b branches, then there are $(N_b - N_n + 1)$ independent Kvl equations. (We shall not prove this result here.) In Figure 1-11, for example, there are 6 branches and 4 nodes; hence, there should be $N_b - N_n + 1 = 6 - 4 + 1 = 3$ independent Kvl equations. Verify this relationship also for Figure 1-10.

Write Kvl equations around loop 1-4-5, 2-3-4 and 1-2-6 in Figure 1-11 and notice that no one can be obtained from the other two, showing that all three are independent. Then write a Kvl equation for any one of the other closed paths and then try to obtain it by certain combinations of the first three equations.

1-6. Ohm's Law and Resistance

(Read the programmed text booklet titled "Ohm's Law and Sources" for further instruction in the subject of this section.)

1-18

- 1 The resistivity of a material depends on such things as the mass and charge of an electron, the density of free electrons in the material, the average velocity with which they move, and the average distance an electron moves before colliding with another particle. For our purposes, it is enough to know that there is considerable variation of resistivity among materials and that resistivities can be determined by measurement. Table 1-1 shows the resistivities of a number of materials.

- 2 Any condition that influences one or more of the quantities on which the resistivity depends (listed above) will have an influence on the resistivity, and hence on resistance. One clear condition that is likely to influence such things as the average distance traveled by an electron between collisions, or the electron density, is a change in temperature. Indeed, it is found empirically that the resistivity of materials does depend on temperature. For metals, the change in resistivity is approximately proportional to the change in temperature, at least near ordinary room temperatures. An adequate approximate expression between resistivity and temperature is the straight line:

$$\rho = \rho_0 (1 + \alpha T) \quad (1-16)$$

- where T is temperature in centigrade, ρ_0 , the resistivity at zero degrees and α is called the temperature coefficient of resistivity. Its value for some metals is also given in Table 1-1. Note that the same expression describes the temperature dependence of resistance as can be verified by multiplying both sides by l/A .

Example:

- 7 Find the length of aluminum wire having a cross-sectional area of .02 square millimeters which is needed to limit to 100 ma. at 0°C the current drawn from a 12-volt battery (assumed to have zero internal resistance). Also, find the range of the value of resistance during a year if the minimum and maximum temperatures are -20° and 35°C . The required resistance is $R = 12/.1 = 120$ ohms. From Eq. (1-15) the required length is $l = RA/\rho = \frac{120 \times .02 \times 10^{-6}}{2.62 \times 10^{-8}} = 91.6$ meters; the resistivity was taken from Table 1-1. Using Eq. (1-16) for resistance and taking α from Table 1-1 we find at the two extremes of temperature

$$R_{\min} = 120 (1 - 0.0039 \times 20) = 110.6 \text{ ohms}$$

$$R_{\max} = 120 (1 + 0.0039 \times 35) = 136.4 \text{ ohms}$$

1

Table 1-1

2

Resistivity in
ohm-meters
(at 0°C)Temperature Coefficient
of resistivity per
degree C at 0°

Silver

 1.47×10^{-8} 3.8×10^{-3} Copper (standard
annealed)

1.58

3.8

3

Aluminum

2.62

3.9

Tungsten

5.5

4.5

Zinc

5.8

3.7

Nickel

6.93

4.3

4

Iron

8.85

6.2

Platinum

11.0

4.2

Tin

11.5

4.2

5

Lead,

19.8

4.3

Carbon steel

20 to 50

2 to 5

Manganin

43.0

.003

Graphite

800

.075

6

1 1.7. Power and Energy

The concept of voltage was introduced by discussing the work done in moving a charge from one point to another. Specifically, the voltage is the work done when a charge moves in an electric field. This work, or energy, is either expended by the charge as it loses potential energy, or it is performed on the charge while moving it to a point of higher potential.

Consider the network branch shown in Figure 1-12. The branch is carrying

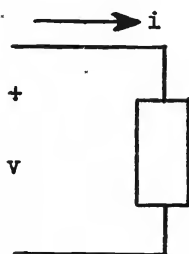


Figure 1-12

a current i with a voltage v across its terminals. After the passage of some time, a net charge q will have been transported through the branch from one terminal to the other. From the definition of voltage and the reference directions shown, an energy w is expended by the charge and this energy is

$$w = q v. \quad (1-17)$$

The unit of energy is the joule.

To determine the rate at which this energy is expended, let the charge in question be an incremental charge Δq and let its transfer between the terminals of the branch take place in Δt seconds. Then the incremental work done is $\Delta w = v \Delta q$. If we now divide both sides by the time increment Δt and let $\Delta t \rightarrow 0$, we will find the rate of expenditure of energy, or the power, to be

$$p = \frac{dw}{dt} = v \frac{dq}{dt} = vi. \quad (1-18)$$

Note again that this is energy expended in the branch by the charge. If either the voltage reference is reversed or the current reference is reversed, work will be done on the charges in moving them to points of higher potential. Thus the

- 1 power also has a reference direction related to those of current and voltage.
The unit of power is the watt, which is the same as a joule per second.

Example

- 2 Let the voltage and current in Figure 1-12 be given by the curves shown
in Figure 1-13. Find the energy expended at the end of 2 and 4 seconds.
Plot also a curve giving the amount of charge passing through the branch
as a function of time.

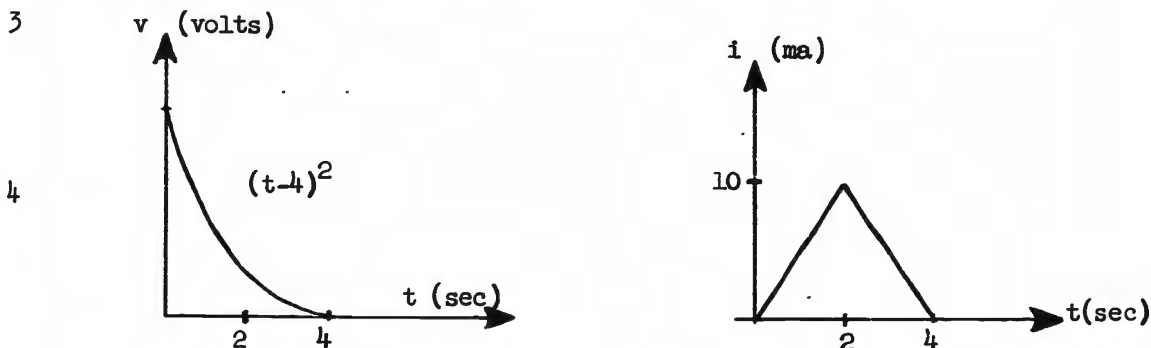


Figure 1-13.

The power is $p = vi$. From $t = 0$ to 2, $i = 5t$ ma. Hence

6
$$p = 5t(t-4)^2 \text{ milliwatts,}$$

and the energy expended after 2 seconds can be obtained by integrating the
power.

7
$$w \text{ (after 2 seconds)} = \int_0^2 5t(t-4)^2 dt = \int_0^2 (10t^3 - 40t^2 + 80t) dt = \frac{280}{3} \text{ millijoules.}$$

- 8 For the period from 2 to 4 seconds, it is first necessary to find the equa-
tion for i . The straight line has a slope of -5 and passes through the
point (4, 0). Hence, $i = -5t + 20$. At the end of 4 seconds the energy
will be $280/3$ plus the integral of the power from 2 to 4 sec.

9

1-22

$$\begin{aligned}
 1 \quad w \text{ (after 4 seconds)} &= \frac{280}{3} + \int_2^4 (t-4)^2 (-5t+20) dt = \\
 &= \frac{280}{3} + \int_2^4 (-5t^3 + 60t^2 - 240t + 320) dt = \frac{280}{3} + 20 = \frac{340}{3} \text{ millijoules.}
 \end{aligned}$$

The charge is the integral of the current. Thus

$$3 \quad q = \int_0^t 5t \, dt = \frac{5}{2} t^2 \text{ millicoulombs} \quad 0 \leq t \leq 2$$

$$\begin{aligned}
 4 \quad q &= \int_0^2 5t \, dt + \int_2^t t(-t+4) dt \\
 &= 10 + 5 \left(-\frac{t^2}{2} + 4t \right) \Big|_2^t = 10 + \frac{5}{2} (t-2)(6-t); \quad 2 \leq t \leq 4
 \end{aligned}$$

5 In each interval (from 0 to 2 and 2 to 4 secs.) the curves are parabolas. The complete curve is shown in Figure 1-14. (Verify that the slopes of the two parts are the same at $t = 2$ and that the slope is zero at $t = 4$, as they should be.)

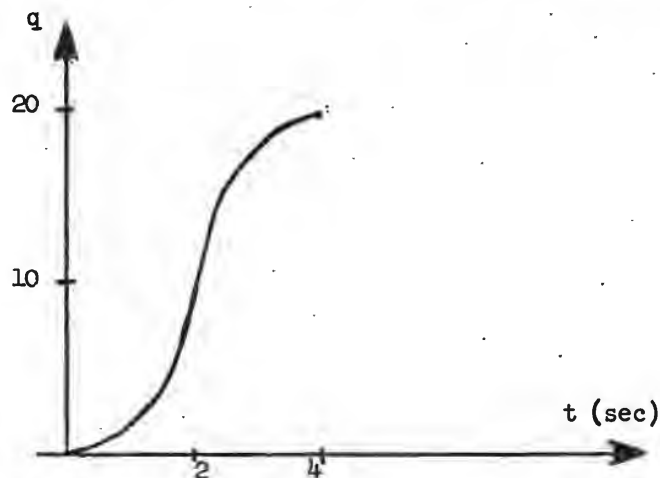


Figure 1-14.

9 If the branch in question is a resistor (ideal) the power expended (this is a short way of saying the rate at which energy is expended) becomes, in substituting Ohm's law into Eq. (1-18),

$$p = Ri^2 = \frac{v^2}{R} = G v^2 \quad (1-19)$$

Example

Figure 1-15 shows two batteries connected to a 10 ohm resistor. The batteries are each represented by an ideal voltage source in series with an internal resistance. Find the power expended in the 10 ohm resistor and the power supplied by each source. Is the principle of conservation of energy satisfied in this diagram?

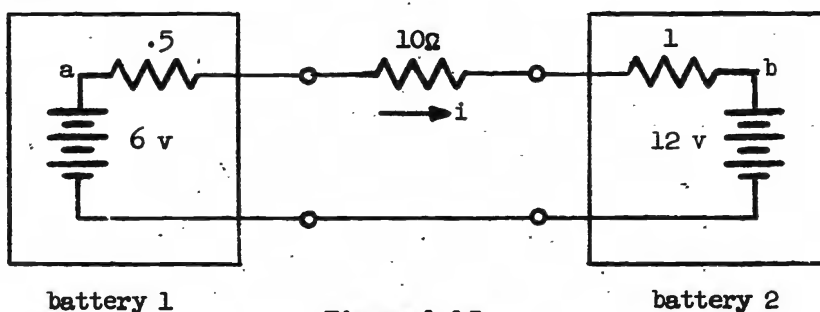


Figure 1-15.

The voltage v_{ab} is $6 - 12 = -6$ volts. This voltage appears across a combination of resistors whose total resistance is 11.5 ohms, so that the current i is $-6/11.5$ amp. Hence, the power dissipated in the 10 ohm resistor is

$$p = 10 i^2 = 10 \left(\frac{6}{11.5} \right)^2 = 2.72 \text{ watts.}$$

The power entering the ideal 12 volt source is $-12 \left(\frac{6}{11.5} \right) = -6.25$ watts. The negative sign indicates that the 12 volt battery is actually supplying power. In the case of the other battery, power entering the ideal 6 volt source is $6(6/11.5) = 3.13$ watts. Because the sign is positive, this power is actually absorbed.

To determine the power balance we must also compute the power expended in the two internal resistances as well. To summarize:

power absorbed by 6 volt ideal source	= 3.13 watts
power absorbed by 10 ohm resistor	= 2.72 watts
power absorbed by .5 ohm internal resistance	= 0.14 watts
power absorbed by 1 ohm internal resistance	= <u>0.27 watts</u>
Total	6.26 watts

This is equal to the power supplied by the 12 volt ideal source, as it must be if the principle of conservation of energy is to be satisfied.

Chapter 2

RESISTIVE AND DIODE NETWORKS

In the last chapter three hypothetical devices were introduced, and several "laws" relating to them. There was an ideal resistor, an ideal voltage source and an ideal current source. (The last two will often be abbreviated v-source and i-source.) Kirchhoff's two laws and Ohm's law determine the interrelations of voltage and current in a network containing interconnections of these three devices.

Practical resistance circuits involve the interconnection of devices which, in general, are non-ideal. That is, the $v-i$ curves of resistors are not exactly linear, the potential difference at the terminals of sources is never exactly independent of current (as required for an ideal voltage source) nor is the current of a source exactly independent of voltage (as required for an ideal current source). Nevertheless, there are many cases where resistors have very nearly linear properties, and where actual sources can be represented by equivalent circuits consisting of combinations of resistors and ideal sources. In these cases, actual circuits can be represented on paper by ideal circuits, and their behaviors can be analyzed and predicted by methods developed in this chapter.

We shall discuss procedures developed by the application of these basic relationships in various ways. Our interest will be in computing the voltage or current in a branch of a network, or the power dissipated in a resistor or supplied by a source, when the network itself is given. We are also interested in the converse process, that of determining what a specific resistor value, or source voltage or current, must be in order that a particular branch voltage or current or power take on a specified value. This is a problem in design, or synthesis, as opposed to the previous problem of analysis.

2-1. Series Circuits and Voltage Dividers

A number of branches are said to be in series if they are connected end-to-end such that the current in each branch is the same. Thus, Fig. 2-1 shows a series circuit of three resistors and two voltage sources (one constant, and one variable with time) connected so that the current in each element is the

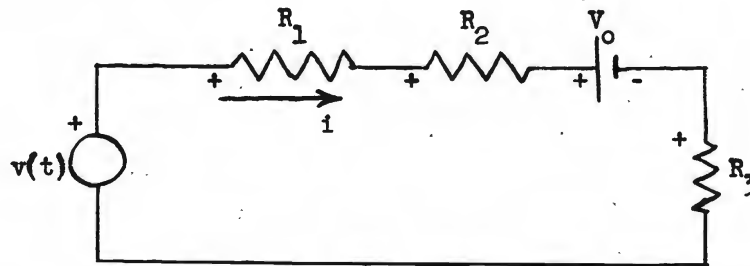


Fig. 2-1 A Series Circuit

same. This relationship of the branch currents automatically satisfies Kirchhoff's current law at the junctions between branches. There is a single closed path around which Kirchhoff's voltage law can be applied. As each voltage term is being written for a resistor, Ohm's law can be applied. With the current reference shown in Fig. 2-1, this simultaneous application of Kvl and Ohm's law leads to

$$R_1 i + R_2 i + V_0 + R_3 i - v = 0 \quad (2-1)$$

which can be solved for the unknown current. Thus,

$$i = \frac{v - V_0}{R_1 + R_2 + R_3} \quad (2-2)$$

Once the current is known, the voltage across any resistive branch follows from Ohm's law.

Now refer to Fig. 2-2, Applying the same technique of analysis, the current

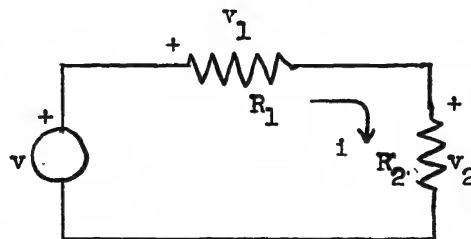


Fig. 2-2 A Voltage Divider

is easily found to be $i = v / (R_1 + R_2)$. The voltage across each of the resistors, with the references shown, can be written

$$v_1 = \frac{R_1}{R_1 + R_2} v \quad ; \quad v_2 = \frac{R_2}{R_1 + R_2} v \quad (2-3)$$

The structure shown in Fig. 2-2 is called a voltage divider; the branch voltage expressions in Eqs. (2-3) are said to be the voltage divider formula. It can be remembered as a proportionality as follows: "The voltage across one resistor of a series combination is to the total voltage what the value of that resistance is to the total resistance."

2-2. Parallel Networks and Current Dividers

A number of branches are said to be in parallel if their branches are connected so that the same voltage appears across each branch. Figure 2-3 shows a parallel network. This relationship of equal branch voltages automatically satisfies Kirchhoff's voltage law around the closed loops formed by the parallel branches. A single independent relationship is obtained by applying Kcl. As each current term is written for a resistive branch, Ohm's law can be applied. With the voltage reference shown in Fig. 2-3 this simultaneous application of

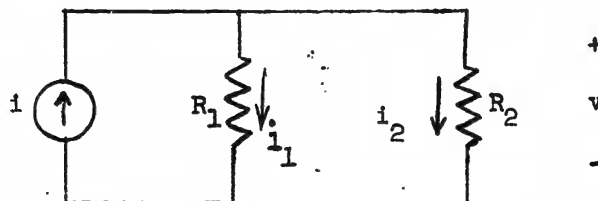


Fig. 2-3 A Current Divider

Kcl and Ohm's law leads to

$$-i + \frac{v}{R_1} + \frac{v}{R_2} = 0 \quad (2-4)$$

or, in terms of conductances,

$$-i + G_1 v + G_2 v = 0 \quad (2-5)$$

Solving for the voltage leads to

$$v = \frac{i}{G_1 + G_2} = \frac{R_1 R_2}{R_1 + R_2} i \quad (2-6)$$

2-4

1 It is seen that the equivalent resistance R of two resistors connected in parallel is given by

$$R = \frac{R_1 R_2}{R_1 + R_2} \quad (2-7)$$

2 In terms of conductances, the equivalent conductance G has the simpler form

$$G = G_1 + G_2 \quad (2-8)$$

3 The current in each resistive branch in Fig. 2-3 is easily found from the voltage in Eq. (2-6) to be

$$i_1 = \frac{G_1}{G_1 + G_2} v = \frac{R_2}{R_1 + R_2} v \quad (a)$$

$$i_2 = \frac{G_2}{G_1 + G_2} v = \frac{R_1}{R_1 + R_2} v \quad (b)$$

5 The structure of Fig. 2-3 is called a current divider. The current divider formula can be easily remembered as a proportionality: "The current in one resistor of a parallel combination is to the total current what the value of that conductance is to the total conductance."

2-3. Network Solution by Equivalent Source Transformations

7 We have found it a simple matter to find all branch voltages and currents in two network structures: a series circuit and a parallel combination. Suppose a network is given having a structure other than a simple series or parallel arrangement, and that a particular branch voltage or current is to be found. If the structure could be converted to a series or parallel arrangement containing the branch in question, the rest would be simple.

8 As one step in such a conversion, consider the two networks shown in Fig. 2-4: an ideal voltage source in series with a resistor and an ideal current source in parallel with a resistor. It is assumed that there is a branch (not shown) connected between terminals a and b in both cases, so that there is a current i and a voltage v at these terminals. It is desired to find the conditions on v_0 , i_0 , R_0 and R for these two configurations to be equivalent at the terminals. By this is meant that

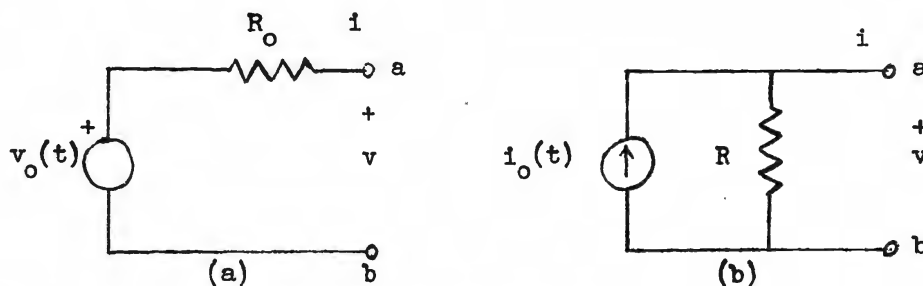


Fig. 2-4 Equivalent Sources

the relationship between the terminal voltage and current is to be the same for the two networks, independently of the load connected at the terminals.

Applying Kvl and Ohm's law in Fig. 2-4a, and Kcl and Ohm's law in Fig. 2-4b, there results

$$v = v_o - R_o i \quad (a) \quad (2-10)$$

$$i = i_o - \frac{v}{R} \quad \text{or} \quad v = R i_o - R i \quad (b)$$

Assuming identical loads, the two voltages should be equal if the two networks are to be equivalent. Equating them leads to

$$(v_o - R i_o) + i(R - R_o) = 0 \quad (2-11)$$

If the equivalence is to be independent of the load connected at the terminals, this relationship must be valid for all values of i . This will be true only if

$$R = R_o \quad (a) \quad (2-12)$$

$$v_o = R_o i_o \quad \text{or} \quad i_o = v_o / R_o \quad (b)$$

That is to say, the two configurations in Fig. 2-4 are equivalent if the two resistors are the same and the voltage source is related to the current source by $v_o = R_o i_o$. Figures 2-4a and b are respectively called a voltage source equivalent, and a current source equivalent. Note that these terms apply to the

- 1 ideal source together with the resistor, not the ideal source alone. Reference
 may be made to Chapter 1, to provide a reminder as to how these equivalents relate
 to actual sources.

- 2 With this equivalence, it is possible to reduce a given network to a series
 circuit or a parallel combination. The process will be illustrated by means of
 the network in Fig. 2-5. It is desired to find the voltage v across the 20-ohm

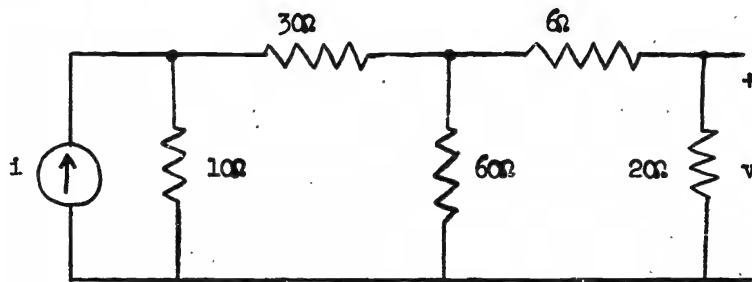


Fig. 2-5

- resistor. The approach will be to convert the network to a series circuit con-
 taining the 20-ohm resistor.

- The first step is to replace the combination of the current source and the
 10-ohm parallel resistor by its voltage source equivalent--an ideal voltage source
 $10i$ in series with 10 ohms. (If it is confusing to have a voltage source which
 seems to have a current designation, remember that $10i$ has the dimensions of
 resistance times current.) The 10 ohms in series with the 30-ohm resistor gives
 an equivalent resistance of 40 ohms. The series combination of this 40 ohms and
 the $10i$ voltage source is then converted to its current source equivalent, as
 shown in Fig. 2-6b. The 24 ohms equivalent resistance of the 40 ohms and 60 ohms

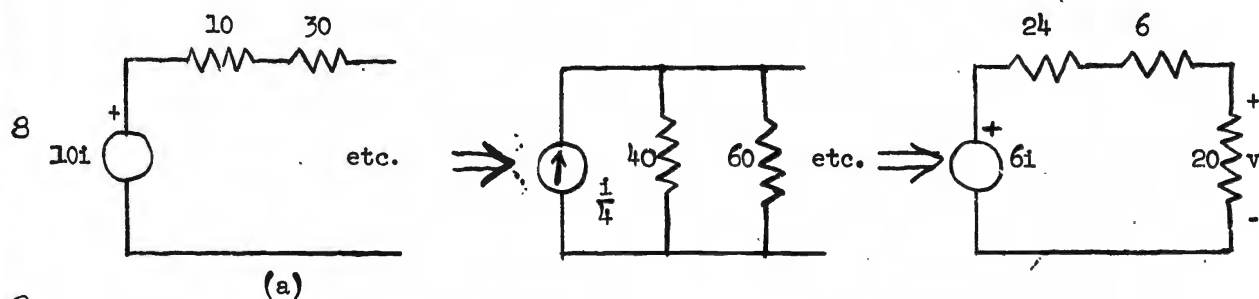


Fig. 2-6

- 1 in parallel together with the $i/4$ current source is now converted to their voltage
 source equivalent, as shown in Fig. 2-6c. Finally, application of the voltage
 divider formula leads to the desired voltage v .

2
$$v = \frac{20}{20 + 6 + 24} 6i = 2.4i \quad (2-13)$$

- Since the structure of the original network is destroyed, it is not possible
 from the final form in Fig. 2-6 to determine other branch voltages and currents.
 3 However, once the desired voltage has been computed, it is possible to return to
 the original network to find any other desired voltage or current. Thus, sup-
 pose it is desired to find the current in the 30-ohm resistor in Fig. 2-5. With
 v known, the current in the 20-ohm resistor ($v/20$) is known. But this is the
 4 same as the current in the 6-ohm resistor. The voltage across the 60-ohm resistor
 equals the sum of the voltages across the 6- and 20-ohm resistors ($6v/20 + v = 13v/10$)
 by Kvl. Hence, the current in the 60-ohm resistor becomes known by Ohm's law
 ($13v/10$ divided by 60). Finally, Kcl gives the desired current in the 30-ohm
 5 resistor ($i_{30} = i_{60} + v/20 = 13v/600 + v/20 = 43v/600$).

- Returning to Fig. 2-26c, note that everything but the 20-ohm resistor has
 been replaced by a voltage source in series with a resistor. Although this was
 demonstrated by an example, it is a general result which can be stated as follows

- 6 A network consisting of ideal current and voltage sources and linear resistors
 can be replaced at a pair of terminals by an equivalent consisting of a single
 voltage source and a single series resistance. This circuit is called a Thévenin
equivalent, the source being the Thévenin equivalent voltage and the resistor
 7 being the Thévenin equivalent resistance.

- Since it has already been demonstrated that a current source in parallel
 with a resistor can be made equivalent to a v -source in series with the same
 resistor, it follows that this configuration (an i -source in parallel with a
 8 resistor) can be equivalent to any network of sources and resistors at a pair of
 terminals. This new configuration is called a Norton equivalent. The equivalence
 is illustrated in Fig. 2-7.

- Only one process—converting from one source equivalent to another—
 9 has been described here for arriving at a Thévenin or Norton equivalent.
 However, other methods also exist but we shall not consider them here.
 The method described here depends on having each voltage source in a network

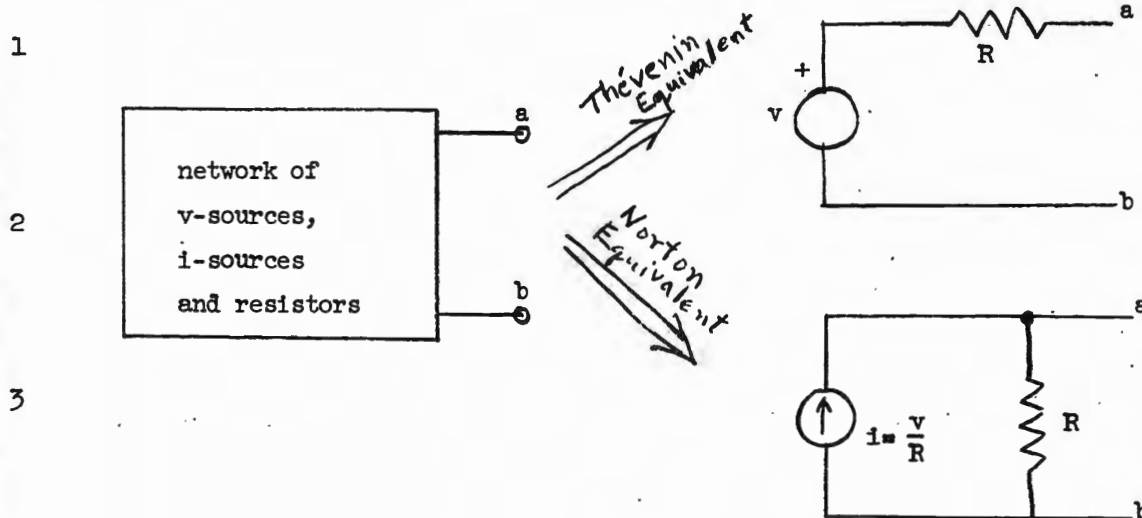


Fig. 2-7

appear with a series resistor and each current source with a parallel resistor. What happens if a source is initially "bare;" that is, no resistor in series with a voltage source or in parallel with a current source? Well, this question does have a favorable answer but discussion of it will be postponed to Sec. 2-6.

2-4. Loop Equations

The preceding method of solving network problems proceeds by converting the structure of the given network into a simple form. We shall now discuss a procedure that uses the three basic relationships--Kvl, Kcl and Ohm's law--in a particular order, thereby arriving at a set of equations. These equations are then solved, thereby determining all voltages and currents in the network. In this method, the structure remains intact,

The procedure will be illustrated in terms of the network in Fig. 2-8(p. 2-10) which is a slight modification of that in Fig. 2-5. The voltage source v_g takes the place of 10i in that network. There are 4 resistors in this network and so 4 resistive branch currents. However, by applying Kcl at the nodes of the network, two of the branch currents can be solved for in terms of the others. (It is trivially noted that the 6 ohm and 20 ohm resistors are in series so their currents are the same. If Kcl is applied to the node joining these two resistors, the same conclusion will follow.) An expression for the current in the 60 ohm resistor, labeled i_3 in the diagram, is obtained from kcl as $i_3 = i_1 - i_2$.

The next step is to apply Kvl around the closed paths, or loops, in the network. In the present case there are a total of three loops, the two inner "meshes" and the outside contour, but the Kvl equation for any one of them can be obtained from the other two, so only two of them are independent. To write Kvl, we need to choose voltage references. Let us agree to choose all resistive branch voltage and current references with the plus sign at the tail of the arrow. ($\overset{+}{\text{---}} \overrightarrow{\text{---}} \text{---}$) Then Ohm's law will always be written with a plus sign. Now, as we write Kvl around a loop, we mentally replace each voltage by a term of the form R_i with the appropriate R and i. Thus, writing Kvl around the two inner meshes, we get

$$40i_1 + 60(i_1 - i_2) - v_g = 0$$

$$6i_2 + 20i_2 - 60(i_1 - i_2) = 0$$

Upon collecting terms and transposing v_g , these become

$$\begin{aligned} 100i_1 - 60i_2 &= v_g \\ -60i_1 + 86i_2 &= 0 \end{aligned} \quad (2-15)$$

6 This pair of linear algebraic equations in two unknowns can be solved by Cramer's rule in terms of determinants, or by elimination. The solutions are

$$\begin{aligned} i_1 &= .0172 v_g \\ i_2 &= .012 v_g \end{aligned} \quad (2-16)$$

Once i_1 and i_2 are found, then all other branch currents become known; by Ohm's law, all branch voltages can then be determined. Thus, the voltage across the 20 ohm resistor will be $20i_2 = .24v$. This is to be compared with the value determined previously in Eq. (2-13), remembering that v_g here replaces $10i$ there.

9 The equations that result from this process are called loop equations since they come from applying Kvl around the loops of the network. The currents in terms of which the loop equations are written are called the loop currents.

- 1 To summarize: Given a network, first select a number of loop currents and express all branch currents in terms of these loop currents by Kirchhoff's current law. The next step is to write Kvl equations around as many closed paths in the network as there are loop currents while simultaneously substituting
- 2 Ri 's for the v 's for each resistance, where each branch current is expressed in terms of the loop currents. The resulting equations are the loop equations. A number of questions present themselves at this point.

1. Which branch currents should be chosen as loop currents and how many?

- 3 Except for a single-loop network, the selection of loop currents is not unique and a number of different sets of currents are equally satisfactory. There are well-defined criteria and procedures for selecting an adequate set of loop currents. However, for networks having no more than three loop currents,
- 4 which is the most we shall encounter, it is actually hard to make a mistake, even if one tries. Hence, no further attention will be given to the subject. It can be proved (although we shall not do so) that the number of independent loop equations in a network having N_b branches and N_n nodes is $N_b - N_n + 1$.
- 5 This expression can be used as a check to verify that the right number of loop currents have been chosen.

2. Which closed paths should be chosen for writing loop equations?

Here also a number of different possibilities are equally satisfactory.

- 6 For planar networks (those that can be drawn on a plane without crossing branches) an adequate set of loops are the "meshes", the internal "windows" of the network. Sometimes a different set of loops is more convenient. In any case, there should be as many equations as there are loop currents.

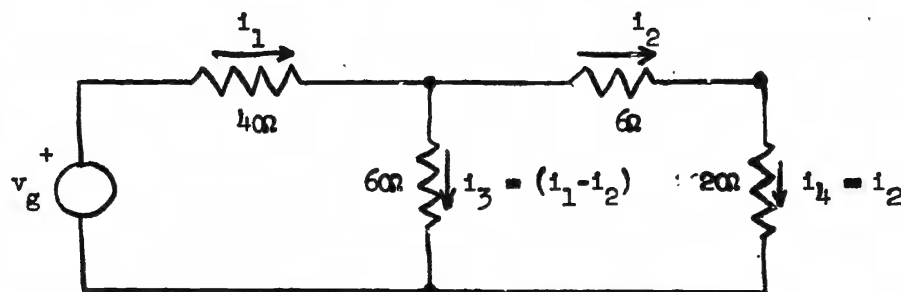


Fig. 2-8

2-5. Node Equations

In writing loop equations, a number of variables--called the loop currents-- are selected, and all branch currents are expressed in terms of these loop currents by Kcl. Let us now, instead, pick a number of voltage variables and express all branch voltages in terms of these variables by Kvl. The process will be illustrated by the network of Fig. 2-5 which is redrawn in Fig. 2-9.

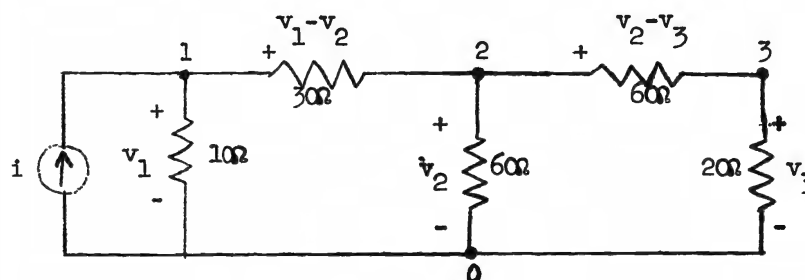


Fig. 2-9

The first step is to choose one node of the network as a datum node to which the voltages of all other nodes will be referred. In Fig. 2-9 let's choose the node labeled 0 as a datum. The voltages of the other nodes relative to that of the datum node, with references chosen plus at the nondatum nodes, are called the node voltages. The voltage of each branch between two nondatum nodes can be written as the difference between two node voltages by Kvl, as shown in Fig. 2-9. Now Kcl is applied at each nondatum node while at the same time replacing the currents by v/R (or Gv), with appropriate v 's. The result is a set of equations called the node equations. For Fig. 2-9, the node equations are

$$\text{node 1:} \quad -1 + \frac{v_1}{10} + \frac{v_1 - v_2}{30} = 0$$

$$\text{node 2:} \quad -\frac{(v_1 - v_2)}{30} + \frac{v_2}{60} + \frac{v_2 - v_3}{6} = 0 \quad (2-17)$$

$$\text{node 3:} \quad -\frac{(v_2 - v_3)}{6} + \frac{v_3}{20} = 0$$

2-12

- 1 Upon collecting terms, clearing fractions and transposing 1, these become

$$4v_1 - v_2 = 30i$$

2 $-2v_1 + 13v_2 - 10v_3 = 0$ (2-18)

$$-10v_2 + 13v_3 = 0$$

These node equations are three linear algebraic equations in 3 unknowns and

- 3 can be solved algebraically. The solutions are

$$v_1 = 8.281$$

4 $v_2 = 3.121$ (2-19)

$$v_3 = 2.41$$

- 5 Note that v_3 is the voltage across the 20 ohm resistor; it was previously labeled v in Fig. 2-5. The answer here agrees with the value found there.

With the node voltages known, all the branch voltages follow; from these the currents can all be determined

- 6 To summarize: Given a network, first select a datum node; any node of the network will do. Label the node voltages, which are the voltages of all nondatum nodes relative to that of the datum node. Express all branch voltages in terms of the node voltages by Kvl. Next write Kcl at all the nondatum nodes while simultaneously replacing the currents by voltage-over-resistance, with the branch voltages expressed in terms of node voltages.
- 7 The resulting equations are the node equations. If there are N_n nodes in the network there will be $N_n - 1$ node equations, all independent.

- 8 On comparing the two procedures--loop equations and node equations--we notice that each method utilizes all three of the basic relationships (Kcl, Kvl and Ohm's law) but in a different order. In a network there are initially both current and voltage unknowns. In the case of loop equations, the voltages are all eliminated and expressed in terms of currents; the resulting equations contain only loop currents as unknowns.
- 9 In the case of node equations, the branch currents are all eliminated and expressed in terms of voltages; the resulting equations contain only node voltages as unknowns.

- 1 But it is not essential to follow either of these two methods. One
 2 can keep a mixed set of variables--voltage and current--if this should
 3 prove more convenient in a given case. We shall not, however, develop
 4 detailed procedures for the use of such mixed variables in solving network
 5 problems.

*2-6. Additional Comments Concerning Equivalent Sources

- 6 The procedure that was used in Sec. 2-3 for obtaining a Thévenin equi-
 7 valent employs successive conversions from a voltage source in series with
 8 a resistor to an equivalent current source in parallel with the resistor,
 9 and vice versa. A nagging thought arises here: suppose there is a voltage
 10 source without a series resistor or a current source without a parallel
 11 resistor, what then? Such a situation is shown in Fig. 2-10(a); no single
 12 resistor is in series with the voltage source. Now consider the modification
 13 shown in Fig. 2-10(b). The voltage source appears to have slid through the
 14 node at its upper terminal into both branches connected there. How are the
 15 loop equations modified by this change?

- 16 For the loop labeled 3 nothing has been changed so that this loop
 17 equation will be the same; only the other two are possibly different. But
 18 writing loop equations for the loops labeled 1 and 2 in both the
 19 original network and in the modified one in Fig. 2-10(b) (or its redrawn form
 20 in (c)) shows that these loop equations are also the same for both cases.

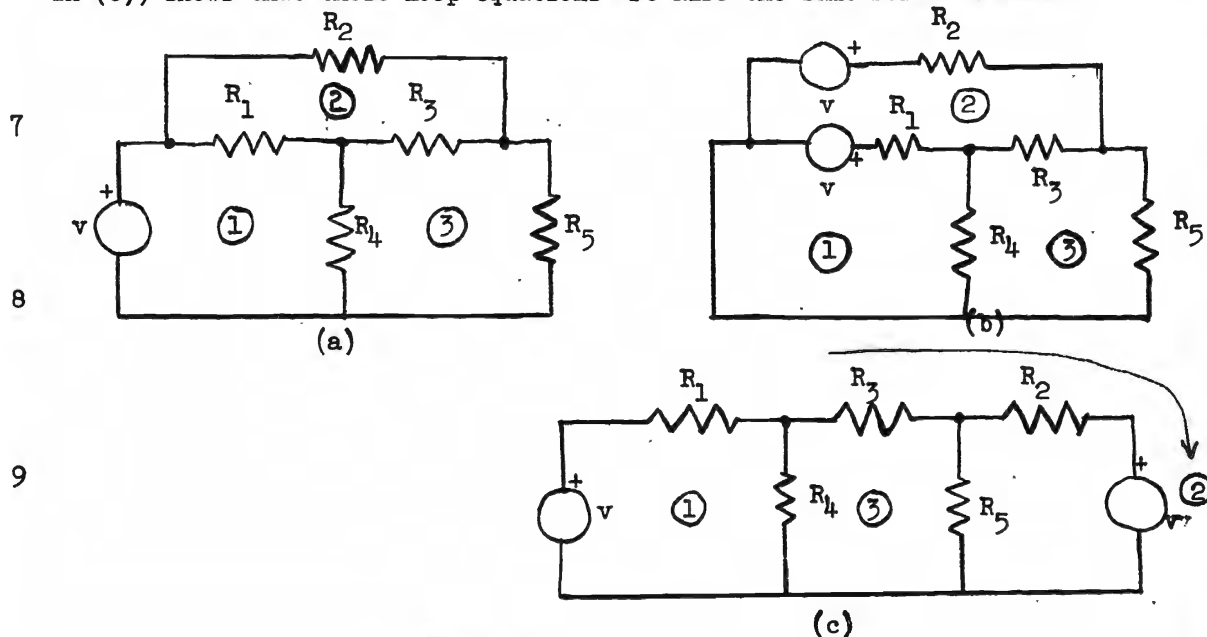


Fig. 2-10

- 1 These two cases are then equivalent since they lead to the same values for the loop currents, and thus for all currents and voltages in the network-- except for the current in the original source itself. This latter can be easily found by Kcl in original network once all other currents are determined.
- 2 This movement of the voltage source has now led to a network having two voltage sources. However, each source is now in series with a resistor and the combination can be replaced by a current source equivalent. Clearly, this result is general; it applies for any number of initially "bare"
- 3 voltage sources in a network and any number of branches connected at a terminal of each source. It leads to the following general statement
A voltage source can be moved through one of its terminals into each of the branches connected there, leaving its original position short-circuited without affecting the voltages and currents anywhere else in the network.
- 4

How about a "bare" current source, one without an accompanying parallel resistor? Such a case is shown in the network of Fig. 2-11(a) in which the

5

6

7

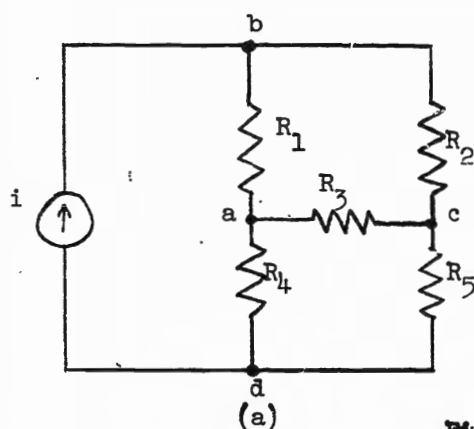
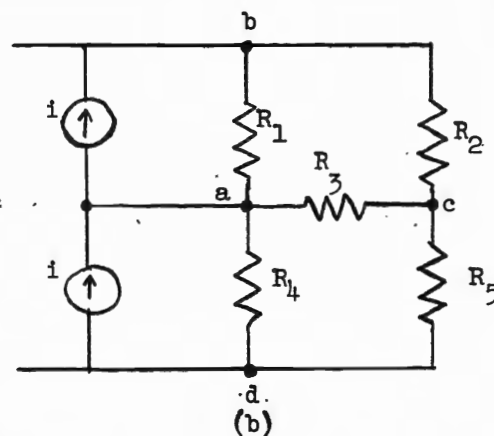


Fig. 2-11.



- 8 current source i is "bare" and forms a closed path with resistors R_1 and R_4 . Now consider the modification shown in Fig. 2-11(b). The current source has been replaced by two sources, both having current i , and their junction has been connected to the node labeled a . There is no current flowing in this connection, as verified by applying Kcl. Hence, the node equation at node a has not been changed. Hence, the two situations in the figure are equivalent, since they will lead to the same node equations and hence, to the
- 9

- 1 same values of voltage and current--except for the voltage across the original source itself. And this can be easily found from the original network, once all other voltages are known.

Although again the number of sources has increased, each current
2 source has now acquired a parallel resistance, and the combination can be replaced by a voltage source equivalent. Again the result is general and can be stated as follows.

- 3 A current source can be moved through any loop it forms with other branches and placed across each of these branches, leaving its original position open-circuited, without affecting the voltages and currents anywhere else in the network.

As a result of these two possibilities concerning the movement and
4 proliferation of sources, even when the sources originally appear "bare" in a network (that is, a v-source without an accompanying series resistor and an i-source without an accompanying parallel resistor) they can be made to acquire accompanying resistive branches, thereby permitting the conversion
5 to an equivalent source.

2-7. The Principle of Superposition

Very often it might be convenient to determine the total current or
6 voltage in a branch of a network containing several sources by finding what this current or voltage would be if each source was the only one in the network, then adding these results. The question is whether such a procedure is valid. The answer to the question is provided by the principle
7 of superposition which is a very general principle applying to a large number of situations in science and engineering. A general statement of the principle is:

Whenever an effect is linearly related to its cause, then the effect
8 owing to a combination of causes is the same as the sum of the effects owing to each cause acting alone, all other causes being inoperative, or deactivated.

In the case of an electric network the effects are currents and voltages in the branches of a network and the causes are the sources. We have seen
9 that any of the equations (loop equations, node equations) that result from applying the basic laws to networks of ideal resistors (and sources) are linear algebraic equations, in which effects are linearly related to causes. Hence, the principle of superposition applies to the calculation of voltage or current in such networks.

- 1 It only remains to clarify what it means to deactivate a source. A voltage source is an ideal device which maintains the voltage waveform at its terminals independent of the terminal current. To deactivate it, or cause it to become inoperative, means to make its voltage become zero. Zero voltage corresponds
- 2 to a short circuit. Hence, deactivating a voltage source means short circuiting it.

- Similarly, a current source will be rendered inoperative if its current is reduced to zero. Zero current corresponds to an open circuit. Hence, de-
- 3 activating a current source means open-circuiting it.

- To use the principle of superposition in a network containing several sources, either voltage or current, all sources but one are deactivated and a desired branch voltage or current due to the one remaining source is
- 4 determined. The process is repeated for each source. The sum of the results due to each source separately are then added to give the total due to all sources acting together.

- Note well that the principle of superposition is valid only if the effect
- 5 is linearly related to its cause. Thus, if the desired quantity is the power dissipated in a resistor due to more than one source, this power cannot be determined by superposition, since power is not linearly related to current. Thus, if two sources are present, and i_1 and i_2 represent the currents
- 6 in a resistor R due to each source acting alone, their sum being $i = i_1 + i_2$, the power dissipated in R when both sources are present is $R(i_1 + i_2)^2$. When the two sources are acting alone, the sum of the two powers will be $Ri_1^2 + Ri_2^2 = R(i_1^2 + i_2^2)$. These two expressions are not the same.

- 7 Similarly, the principle of superposition will not apply in a network containing devices whose voltage-current relation is not linear, such as the diodes to be discussed in the next section.

8 2-8. Diode Circuits

- A diode is a two-terminal device having a current-voltage curve approximately like that shown in Fig. 2-12. The symbol used for a diode is also shown in the figure. A relatively large current in one direction--called
- 9 the forward direction--is possible with a small voltage. Only a small amount of current is possible in the other, or reverse, direction.

- 1 The i - v curve is nonlinear. The greatest curvature, or deviation from a straight line, occurs near the origin, but even at other points the curve is not straight. However, the entire curve can sometimes be approximated by a combination of two straight lines as shown in Fig. 2-12(c).
- 2 The slope of each line represents a conductance, the reciprocal of a resistance. For positive voltage and current, the resistance is small (large slope) and is called the forward resistance. For negative voltage and current the resistance is large (small slope) and is called
- 3 the reverse current.

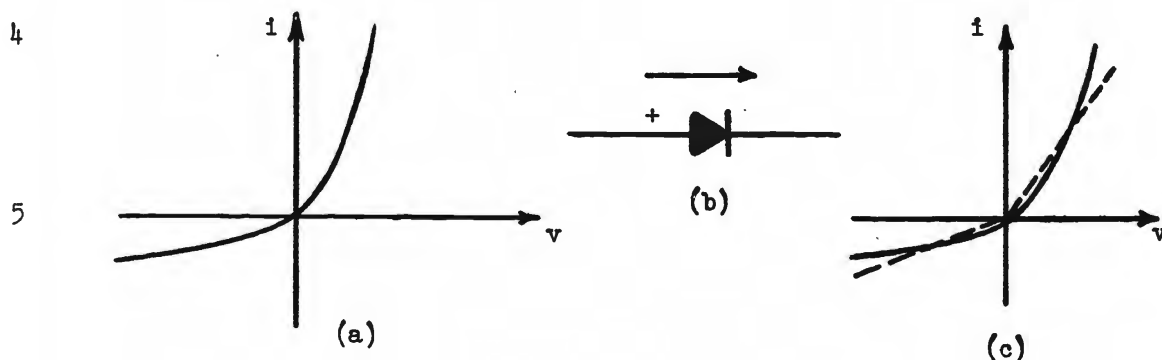
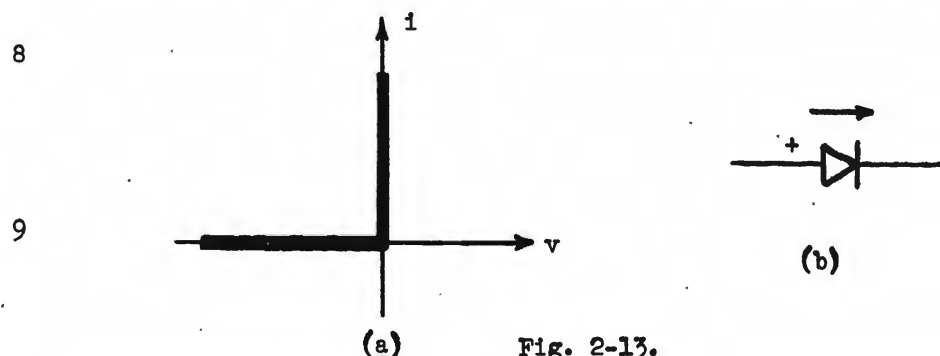


Fig. 2-12.

- It is often convenient to assume that the forward and reverse resistances take on the limiting values: zero forward resistance, $R_F = 0$, and infinite reverse resistance, $R_R = \infty$. The resulting i - v curve is shown in Fig. 2-13. The idealized device having this characteristic is called an ideal diode. To distinguish it from the physical diode, the symbol

Fig. 2-13.
Ideal Diode.

- 1 shown in Fig. 2-13(b) is used. (The arrow head is not black.) The ideal diode has the properties that forward (or positive) current is accompanied by zero voltage and reverse (or negative) voltage is accompanied by zero current. The amount of forward current in an ideal diode is limited
- 2 by the external network connected at its terminals. The same is true of the amount of reverse voltage.

- The ideal diode is seen to be a two-state device. When it is conducting, it is said to be on; when it is not conducting it is off. Whether or not it
- 3 is in one state or the other is determined by the external network. When making calculations in networks containing ideal diodes, it is not often possible to know beforehand whether a diode is on or off. We assume the diode to be in one state or the other, then calculate the diode voltage or
- 4 current and thereby determine whether the diode is actually in its assumed state. Thus, if the diode is assumed to be on, the value of its current can be calculated. If the current turns out to be positive, this verifies that the diode is actually on. If the current turns out to be negative, we
- 5 conclude that our first assumption about the diode being on was not correct; it must have been off under the conditions of the problem.

- Sometimes, of course, a source voltage or current may be varying with time and so the diode may switch its state as the source value changes. Thus,
- 6 a state of the diode may be assumed, say off. With the diode off (open circuited) an expression for its voltage is obtained. From this expression the critical value of the varying source voltage for which the diode voltage will turn positive can be determined. A similar condition exists when a
- 7 network parameter (the value of resistance, say) is not fixed but must be chosen to put the diode in one state or the other.

- In many applications sufficiently accurate results are obtained by representing a physical diode as an ideal diode. At other times more
- 8 accuracy can be obtained if the forward and reverse resistances are not allowed to take on their limiting values. The circuit shown in Fig. 2-14 represents the piecewise linear model of a diode, the one having the broken line i-v curve in Fig. 2-12(c). When the ideal diode in this equivalent
- 9 circuit is on, R_2 is shorted; hence, R_1 is the forward resistance R_F , a relatively low value. When the diode is off R_1 and R_2 are

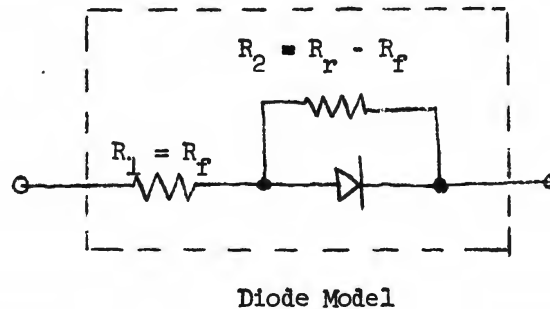


Fig. 2-14

in series and together equal the reverse resistance, R_R , a relatively high value.

With an actual diode replaced by a piecewise linear model, the same method of analysis as carried on before is valid. There is the difference, however, that, instead of switching state from open-circuit to short-circuit and back, the overall diode equivalent switches from its reverse resistance to its forward resistance.

When the accuracy provided even by the piecewise linear approximation is not adequate, use must be made of the actual, nonlinear diode characteristic. Consider the circuit shown in Fig. 2-15. The i - v curve of the diode is also shown. The curve provides one relationship between the

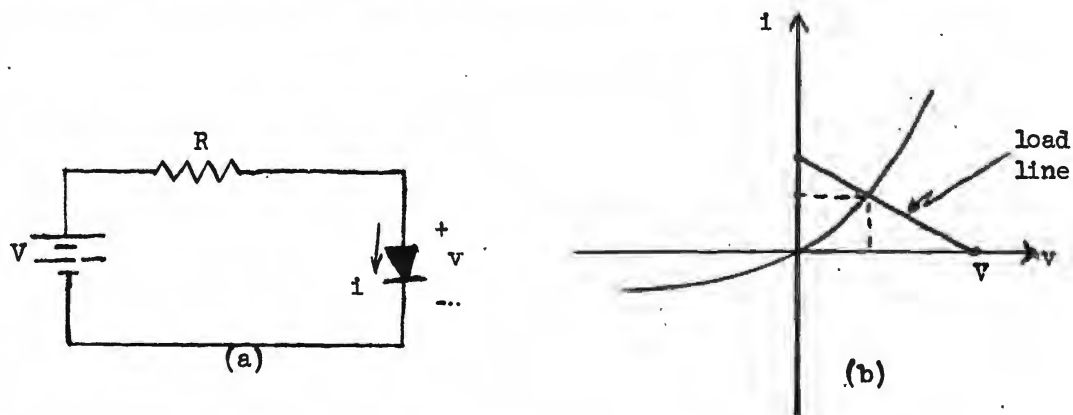


Fig. 2-15

diode voltage and current. Another relationship is provided by the rest of the circuit. Thus, the voltage across the resistor being $V-v$, the

- 1 current through it, which is the same as the current i in the diode, is

$$i = \frac{V}{R} - \frac{v}{R} \quad (2-20)$$

- 2 This is the equation of a straight line whose slope is $-1/R$ and whose intercept on the voltage axis equals the battery voltage. This line is also shown on the same axes as the diode i - v curve in Fig. 2-15(b). It is called the load line. The intersection of the load line with the diode i - v curve gives
3 the solution for the voltage and current of the diode.

- The same type of graphical solution can be followed even when the network in which a diode appears is more extensive, containing more ideal resistors and sources, so long as only a single nonlinear diode is present. The rest
4 of the network connected at the terminals of diode can be replaced by a Thévenin equivalent and the result will have the form of the simple series circuit in Fig. 2-15. After the diode voltage and current are determined in the modified circuit, other voltages and currents in the original network
5 can be found by returning to the original network and using the known values of the diode voltage and current.

- The graphical load line analysis described here for a network containing a diode can be used in many other cases as well when a single nonlinear device
6 is contained in a "network" of linear devices. This approach will be used later in the analysis of nonlinear magnetic circuits and of amplifier circuits.

7

8

9

Chapter 3

ELECTROSTATICS

1 Introduction

2 The subject of electrostatics has to do with electrical phenomena
which can be attributed to the location of stationary charges in space,
as distinct from phenomena associated with charges in motion at a constant
3 velocity (magnetism) or accelerating charges (radiation). Electrostatic
phenomena never exist completely alone. For example, forces on charged
clouds just prior to a lightning stroke are electrostatic forces, and
can be understood in terms of the principles of electrostatics. However,
4 the phenomena involved in the assembly of charges on a cloud, and the
subsequent lightning stroke are not electrostatic.

Another example of an essentially electrostatic phenomenon is found
in the capacitors that are used profusely in electrical circuits. A
5 capacitor charged to a constant voltage comprises a strictly electro-
static situation; it is of limited use or interest. This is the state
of a coupling capacitor in an audio amplifier when no signal is being
transmitted. In the presence of a signal, however, the voltage and
charge are continually changing with time, and a strictly electrostatic
situation does not exist. However, at any instant of time, the relation-
ship between voltage and charge is the same as if they were constant
(provided their rate of change is not too great) and so electrostatic
6 methods of analysis are appropriate.

Another case, where electrostatic principles apply even though
charge is in motion, occurs in the deflection of the electron beam in
a cathode ray tube, as the stream of electrons moves between a pair of
7 charged plates. The force on an electron due to these charged plates
is independent of the electron velocity, and hence electrostatic
principles apply.

3-1. Electric Field

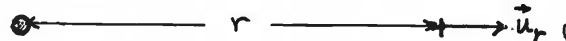
8 Electrostatic phenomena arise basically from the forces experienced
by electrical charges when they are in the presence of other charges.
The simplest possible case is illustrated in Fig. 3-1a, showing two
small charged spheres separated by a distance r . The charged spheres
9 may be regarded as "small" if their diameters are much less than r .

1 In terms of this figure, experiment yields the following results:*

- (1) The force is one of repulsion if the charges are of like sign;
- (2) The force is one of attraction if the charges are of opposite sign;
- (3) The magnitude of the force is proportional to the product of the magnitudes of the charges;
- (4) The magnitude of the force is inversely proportional to the square of the distance between the charges.



Figure 3-1.



6 The proportionalities expressed in (3) and (4) are taken care of by using the expression

$$F = \frac{kq_1q_2}{r^2}$$

7 for the force, where k is a proportionality constant. However, force is a vector quantity and this expression does not account for direction. Let us concentrate on the vector force $\vec{F}$ on q_2 . The direction of the vector $\vec{F}$ can be accounted for, in agreement with observations (1) and (2), by using a unit vector $\vec{u}_r$ directed radially away from q_1 , as in Fig. 3-1b. Then, the force is completely described by

9 *In these statements it is assumed that means are available for determination of amounts and signs of charges. This is not a trivial question, but is not essential to the description of this basic experiment. Suffice it to say here, therefore, that there are ways to measure amounts and signs of charge.

1

$$\vec{F} = k \frac{q_1 q_2}{r^2} \vec{u}_r \quad (3-1)$$

2

When q_1 and q_2 are of the same sign, the product $q_1 q_2$ is positive, and the above formula shows $\vec{F}$ in the direction of $\vec{u}_r$. If q_1 and q_2 are of opposite sign, $q_1 q_2$ is negative and so $\vec{F}$ is opposite to $\vec{u}_r$, in agreement with (2).

3

It has been stated that Eq. (3-1) applies to small charged spheres, without saying how small they should be. It is found that if the diameters are very large (say $r/2$) then Eq. (3-1) is not valid. In fact, it becomes increasingly accurate as the spheres approach mathematical points.

4

Accordingly, the proper interpretation is that Eq. (3-1) is a postulate, applying to hypothetical point charges. By this we mean that every evidence indicates this relationship is valid, but that it cannot be confirmed by direct experiment, because point charges cannot be attained in the laboratory.

5

The relationship described by Eq. (3-1) is called Coulomb's Law.

6

The factor k is experimentally determined, and in the MKSC system of units is found to have the approximate value 9×10^9 (8.988×10^9 is more accurate).^{*} It is a pertinent observation that charge is not defined by this equation, but is defined in terms of current, as its integral with respect to time.

7

Instead of making a direct substitution of this value of k in Eq. (3-1), for simplification of many subsequent formulas, it is more convenient to replace k by $1/4\pi\epsilon_0$, giving

$$\vec{F} = \frac{q_1 q_2}{4\pi\epsilon_0 r^2} \vec{u}_r \quad (3-2)$$

8

where

$$\epsilon_0 = \frac{1}{36\pi \times 10^9} = 8.85 \times 10^{-12}$$

9

^{*}The experiment described in Fig. 3-1 is not, however, the most accurate way to determine k . One way is to measure the capacitance of an accurately constructed capacitor. Another way, which depends on the development of field theory, is to measure the velocity of propagation of electromagnetic waves (radio or light) in vacuum. Theory shows that $k = c^2 \times 10^{-7}$, where c is the velocity of light.

3-4

- 1 The quantity ϵ_0 is called the permittivity of free space. The explicit appearance of 4π is entirely arbitrary, being introduced for later convenience.

We now observe that Eq. (3-2) can be written

2

$$\vec{F} = q_2 \vec{E} \quad (3-3)$$

if the vector quantity E is defined by

3

$$\vec{E} = \frac{q_1}{4\pi\epsilon_0 r^2} \vec{u}_r \quad (3-4)$$

4

$\vec{E}$ is called the electric field vector, and is quite fundamental to electrical theory. $\vec{E}$ is the ratio: the force on a test charge q_2 , divided by q_2 .*

5

The electric field due to a collection of point charges, like the two charges of Fig. 3-2a, can be determined by applying Eq. (3-4) to each, using respectively r_1 and r_2 for the distance variable r . Vector addition is used, as indicated in the figure. As has been pointed out, the point charge situation is hypothetical; in all practical cases charge is distributed over a surface (if the charged object is a conductor) or throughout a volume, as in an insulator. Equation (3-4) can be applied to such a situation, to as good an approximation as we like, by imagining the body to be broken up into a system of volume elements, as in Fig. 3-2b, with a point charge at the center of each element. The approximation improves as the number of elements is increased. Thus, in Fig. 3-2b, at a point P the field would be

7

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \left(\frac{\vec{u}_1}{r_1^2} + \frac{\vec{u}_2}{r_2^2} + \dots + \frac{\vec{u}_n}{r_n^2} \right) \quad (3-5)**$$

8

where n is the number of elements. Although Eq. (3-5) is given here for

9

* More accurately, the limit of this ratio as q_2 approaches zero. This qualification is necessary because in more general situations, where conductors are present, q_2 will have some effect on the charge distribution on these conductors, and therefore will affect $\vec{E}$ to some extent.

** In field theory, in the limit as each element is reduced to zero this sum becomes an integral.

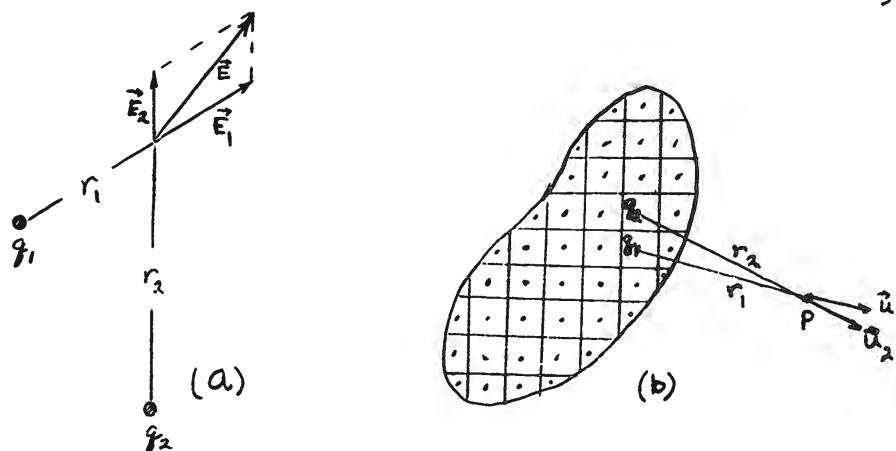


Figure 3-2.

the specific case of a single charged body like Fig. 3-2, it is a general expression for $\vec{E}$ due to any number of charged bodies, with the understanding that each body is treated in a similar manner. The process of obtaining the $\vec{E}$ vector arising from a distribution of charges by performing a vector addition of the individual contribution of each charge is an example of the principle of superposition. Its validity for finding $\vec{E}$ as described above can be approximately established experimentally with small numbers of charges, but the general applicability to any number of charges, including the distributed charge formulation of Eq. (3-5), is to be regarded as a postulate. As we shall see, this postulate leads to certain theoretical consequences which are amenable to experimental verification.

From the foregoing, it is evident that $\vec{E}$ exists in regions surrounding charged objects. It is sometimes helpful to use sketches of field plots, whereby $\vec{E}$ vectors are drawn at various points in space to show their directions and magnitudes. Some examples are shown in Fig. 3-3. The case shown at (a) is the plot for a point charge; the one in (b) is for a dipole (two charges of equal magnitude and opposite sign spaced a small distance apart). In these plots, vectors are actually shown. In many cases it is sufficient to show only a set of lines indicating the directions of $\vec{E}$ (i.e. radial lines in Fig. 3-3a). The state of the space in

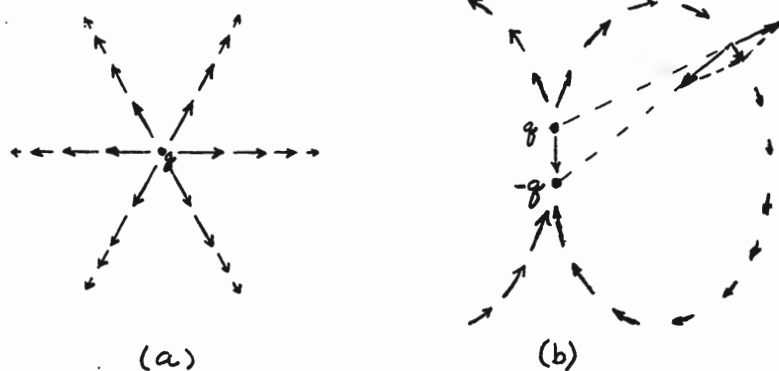


Figure 3-3.

which the field vector $\vec{E}$ is not zero is called an electric field.

So far, the discussion of $\vec{E}$ has been in terms of a force vector in free space, where it is easy to imagine measurement of the force on a test charge. A natural question arises as to whether $\vec{E}$ exists within material bodies, and if so, how it can be defined, inasmuch as it is impossible actually to place a test charge in a solid body and measure a force on it. The answer is the rather simple one of using Eq. (3-5) (or the more general integral form referred to in the footnote) to define $\vec{E}$ within material bodies. That is, if P in Fig. 3-2b should be inside the body, $\vec{E}$ at that point would be the vector given by Eq. (3-5). There is no inconsistency in doing this; when we come to a consideration of material bodies it will only be necessary to determine what are the consequences of this definition.

Note that an electric field refers to the state where there is a force on a stationary test charge. There are situations in which there is a force on a moving charge, due to its motion. This occurs when there is a magnetic

1 field, the subject of the next chapter.

2 3-2. Properties of an Electric Field

Return to the case of a point charge q , and imagine it to be in the center of an imaginary spherical surface, as in Fig. 3-4. As we have seen,

$$\vec{E} = E \vec{u}_r$$

where

$$3 \quad E = \frac{q_1}{4\pi\epsilon_0 r^2}$$

If we multiply E by the surface area of the sphere, we get

$$4 \quad 4\pi r^2 E = \frac{q_1}{\epsilon_0} \quad (3-6)*$$

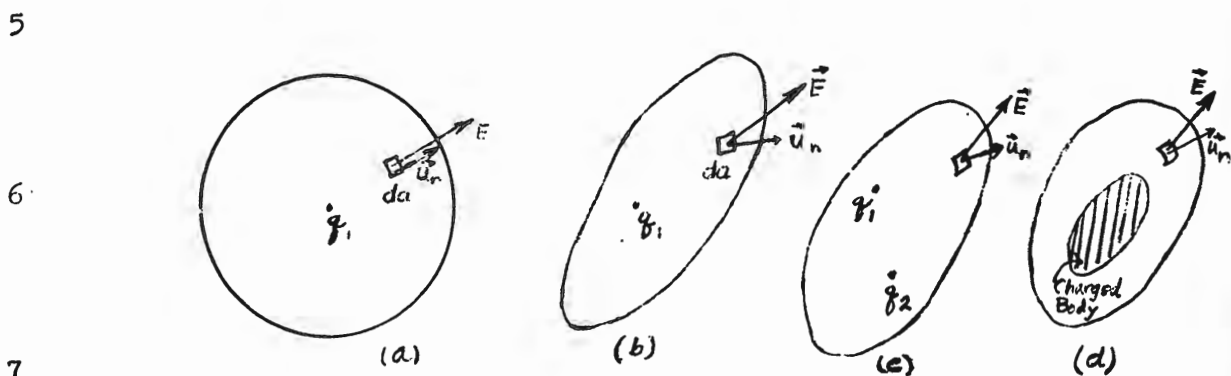


Figure 3-4.

8 Now note that the quantity on the left is a special case of the surface integral

$$9 \quad \oint \vec{E} \cdot \vec{u}_n da$$

* One of the reasons for arbitrarily introducing 4π in Eq. (3-2) was so that the right hand side of Eq. (3-5) would be free of 4π .

3-8

1 over the closed surface of the sphere, where $\vec{u}_n$ is a unit vector normal
 2 to the surface (identical with $\vec{u}_r$ in this case). The 0 through the integral
 symbols implies a closed surface. Thus, Eq. (3-6) can be written in the
 more general form

$$\oint \vec{E} \cdot \vec{u}_n da = \frac{q_1}{\epsilon_0} \quad (3-7)$$

3 which is identical with Eq. (3-6) so long as the surface integral is taken
 over a sphere of radius r . However, in Eq. (3-7) the surface can be distorted
 from spherical form, as in Fig. 3-4b. It is not very difficult to show that
 no matter how the surface is changed, so long as it remains closed with the
 charge q_1 inside, Eq. (3-7) will always be true.

4 Now suppose a surface encloses more than one charge, say two, as in
 Fig. 3-4c. The total vector $\vec{E}$ at a point on the surface is

$$\vec{E} = \vec{E}_1 + \vec{E}_2$$

5 where $\vec{E}_1$ and $\vec{E}_2$ are due, respectively, to q_1 and q_2 . Over the closed
 surface, then,

$$\oint \vec{E} \cdot \vec{u}_n da = \oint \vec{E}_1 \cdot \vec{u}_n da + \oint \vec{E}_2 \cdot \vec{u}_n da$$

6 and from Fig. 3-2a it is evident that the two integrals on the right are
 respectively q_1/ϵ_0 and q_2/ϵ_0 . Thus, we get

$$\oint \vec{E} \cdot \vec{u}_n da = \frac{q_1 + q_2}{\epsilon_0}$$

8 Now suppose a body with distributed charge is enclosed, as in Fig. 3-4d.
 In Sec. 3-1 it was stated (as a postulate) that bodies with distributed
 charges can be treated like aggregates of discrete points, to yield Eq.
 (3-5), and so the above process can be applied repeatedly, each time
 9 adding an additional charge represented by a term in Eq. (3-5), to give

1

$$\oint \vec{E} \cdot \vec{u}_n da = \frac{q}{\epsilon_0} \quad (3-8)$$

2

where $q = q_1 + q_2 + \dots + q_n$ is the total charge on all bodies inside the enclosing surface. This last equation differs from Eq. (3-7) only to the extent that in Eq. (3-8) the charge is not necessarily at a point.

3

The statement that Eq. (3-8) is generally valid for all situations of charge distributions, and for all enclosing surfaces, is known as Gauss' law for electric fields. The surface used in an application of Gauss' law is often called a Gaussian surface. Observe that Gauss' law is not a postulate. Although it was not proved here in detail, such a proof is possible, being based on the postulates leading to Eq. (3-5). Parenthetically, it is appropriate to add that one of the purposes in presenting Eq. (3-5) was to make possible the development of Gauss' law. Except in rare instances, Eq. (3-5) is not useful for calculation, but it provides a key step in the development of Gauss' law, which is a very powerful tool in solving problems.

5

6

Gauss' law will now be used to investigate a particular situation, that of an isolated spherical shell made of conducting material, of radius R , carrying a charge q , as illustrated in Fig. 3-5. We shall use Gauss' law to determine a formula for $\vec{E}$. Because of symmetry, and the fact that like charges repel, we can conclude that charge is uniformly distributed on the surface. The fact that each particle of charge tries to get as far from its neighbors as possible will prevent any build-up of charge concentration, and will cause the distribution to lie entirely on the surface of the sphere.

7

8

We construct a Gaussian surface A of radius r , as shown, and observe that due to symmetry $\vec{E}$ must be radial at each point on the surface, and therefore can be written $\vec{E} = E \vec{u}_r$. Also, observe that $\vec{u}_n$ is also radial (being identical with $\vec{u}_r$) so that

9

$$\vec{E} \cdot \vec{u}_n = E$$

3-10

Gauss' law gives

$$\oint \vec{E} \cdot \vec{u}_n da = 4\pi r^2 E = \frac{q}{\epsilon_0}$$

and so

$$E = \frac{q}{4\pi\epsilon_0 r^2} \quad (3-9)$$

or

$$\vec{E} = \frac{q}{4\pi\epsilon_0 r^2} \vec{u}_r, \quad R < r \quad (3-10)$$

This equation differs from Eq. (3-4) only to the extent that we have q , the total charge on the sphere, in place of q_1 , a point charge, and in the restriction that $r > R$. The reason for this restriction can easily be seen by using another Gaussian surface inside the sphere, labeled B. The charge inside this surface is zero, and therefore at all points on this surface $\vec{E}$ is zero, otherwise the integral of $\vec{E}$ over this surface would not be zero, thereby violating Gauss' law.

Thus, we have seen that, for the region of space outside a charged sphere, the electric field is the same as for a point charge located at the center.*

The next step in learning about fields is to consider two concentric spherical shells, as shown in Fig. 3-6a. The inside sphere carries a charge q , as before, and the outside sphere carries a charge q' . For any spherical Gaussian surface lying between the spheres, there is no change from Fig. 3-5. Therefore,

$$E = \frac{q}{4\pi\epsilon_0 r^2}, \quad R_1 < r < R_2 \quad (3-11)$$

* This may seem to contradict an earlier statement to the effect that Coulomb's law does not apply to spheres of finite size. However, here we are considering the electric field due to an isolated sphere, not the force between two spheres. In the presence of another sphere, the charge would not remain uniformly distributed, and this is why Coulomb's law would no longer apply.

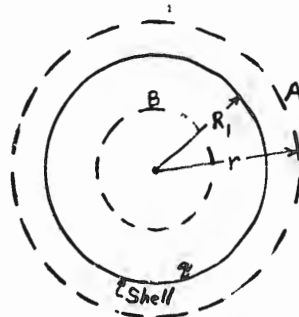


Figure 3-5.

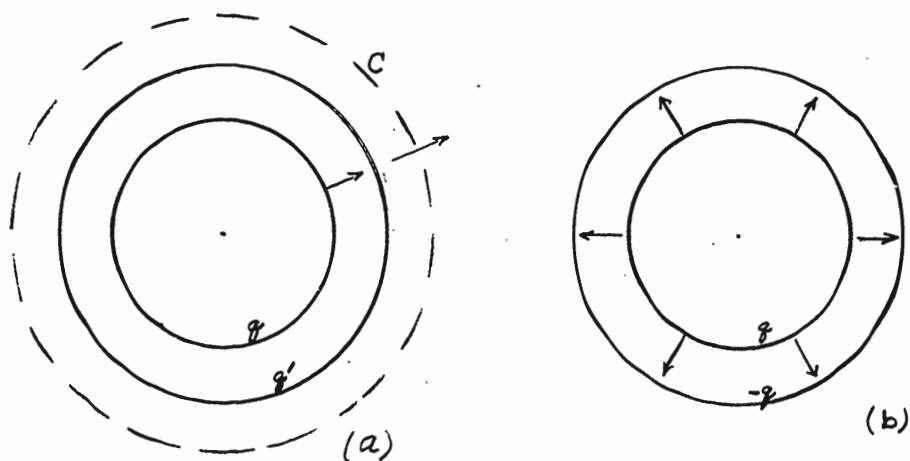


Figure 3-6.

which differs from Eq. (3-10) only in having an upper limit on r . When r is greater than R_2 , a spherical Gaussian surface C is used. The charge inside this surface is $q + q'$, and so we have

$$E = \frac{q + q'}{4\pi\epsilon_0 r^2}, \quad R_2 < r \quad (3-12)$$

The case where $q' = -q$ is particularly important and interesting. In such a case we learn from Eq. (3-12) that E is zero in the region outside the larger shell. In view of an earlier statement that $\vec{E}$ is zero in the region inside the smaller shell, it follows that for this case the field is confined to the region between the shells, as shown in Fig. 3-6b.

3-3. Potential Difference

Figure 3-7a shows a section of the concentric spheres considered in Fig. 3-6. In this new figure r is an integration variable to be used in calculating the work per unit charge by the field in moving a charge from the inside to the outside sphere. $\vec{E}$ is the force per unit charge, and is everywhere tangent to the straight line path covered by the variable r . Thus, the work is

$$V_{ab} = \int_{R_1}^{R_2} \vec{E} \cdot d\vec{r} \quad (3-13)$$

Although this is physical work, more specifically it is work per unit charge, and is called potential difference. This is a scalar quantity, requiring the specification of a reference, which is done by calling V_{ab} the potential of point a with respect to point b. This means that the quantity V_{ab} is positive when the field does work in moving a positive charge from a to b, which occurs when the inner sphere is more positively charged than the outer one.

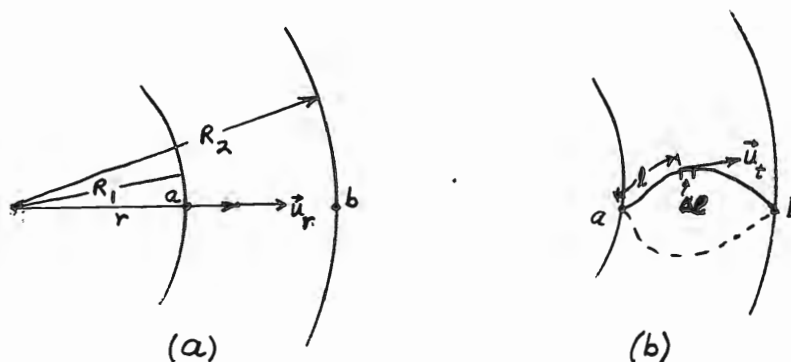


Figure 3-7.

Equation (3-13) can be evaluated for the case in question, using Eq. (3-10) for E , giving

$$V_{ab} = \frac{q}{4\pi\epsilon_0} \int_{R_1}^{R_2} \frac{dr}{r^2} = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right) \quad (3-14)$$

We shall have later use for this formula, but the purpose of presenting it here was to show a specific application of Eq. (3-13).

It can be shown mathematically from Eq. (3-5) that the work done by a static electric field in moving a charge between two points is independent of the path taken in going between the points. Thus, rather than the simple straight line path, a curved path might be taken, as in Fig. 3-7b.

For general use, it is convenient to have an integral expression for V_{ab} that can be used for any path. To see what it should be, in Fig. 3-7b let $\vec{u}_t$ be a unit vector tangent to the path, at some point where there is an increment Δl . The vectors $\vec{E}$ and $\vec{u}_t$ are not necessarily in the same direction, and so to get the work done in moving distance Δl we want the component of $\vec{E}$ along $\vec{u}_t$. Thus, the increment of work is

$$\vec{E} \cdot \vec{u}_t \Delta l$$

and the potential difference is obtained by summing these in the form of an integral, giving

$$V_{ab} = \int_a^b \vec{E} \cdot \vec{u}_t dl \quad (3-15)$$

The fact that the above integral is independent of the path between two points is equivalent to a statement that no work is done by a static field when a charge is carried around a closed path. To see this, in Fig. 3-7b, suppose there is a second path, shown by the dotted line. The

1 work done (V_{ab}) by the field is the same in carrying a charge from a to b
 by either path. Suppose, however, the charge goes from a to b along the top
 route, and is carried back to a along the bottom path. The work done in
 going from b back to a along the bottom path is the negative of the work
 2 done in going from a to b along this same path.* But this is $-V_{ab}$, and so
 the work done in covering the closed path is $V_{ab} - V_{ab} = 0$.

A field having the property described above is called a conservative
field. Since potential difference is what we mean by "voltage" in circuit
 3 analysis, it is evident that the above distinctive property of a conservative
 field is equivalent to Kirchhoff's voltage law.

3-4. Conducting Materials

The idea of potential difference provides a means for defining what we
 4 mean by a conductor. A conductor is a material such that under static conditions
 all points on and within it are at the same potential. In this statement "at
 the same potential" means that the potential difference between any two points
 will be zero. Referring to Eq. (3-15), it is seen that this is equivalent to
 5 saying that $\vec{E}$ is zero everywhere within a conductor, when charges are motionless.

The fact that $\vec{E}$ must be zero in a conductor has a very interesting
 consequence. In Fig. 3-8a, the geometrical figure represents a solid conducting
 body which carries a charge. The dotted figure is a Gaussian surface over which
 6 we have

$$\oint \vec{E} \cdot \vec{u}_n \, da = 0$$

7 since $\vec{E}$ is everywhere zero. However, from Gauss' law we know that the right
 hand side of the above equation is the total charge inside the Gaussian
 surface. The conclusion is that this charge is zero. This analysis applies
 to any Gaussian surface we might choose within the body. Thus, the charge
 is zero everywhere inside a solid conductor. Whenever a conductor carries a
 8 charge, the charge appears on the surface. (Previously we had presented an
 argument supporting this conclusion only in the case of a sphere.)

9 *It is commonly said that in this case the external force required to move the
 charge from b to a does work on the field. (i.e. the work done by the field
 is negative.)

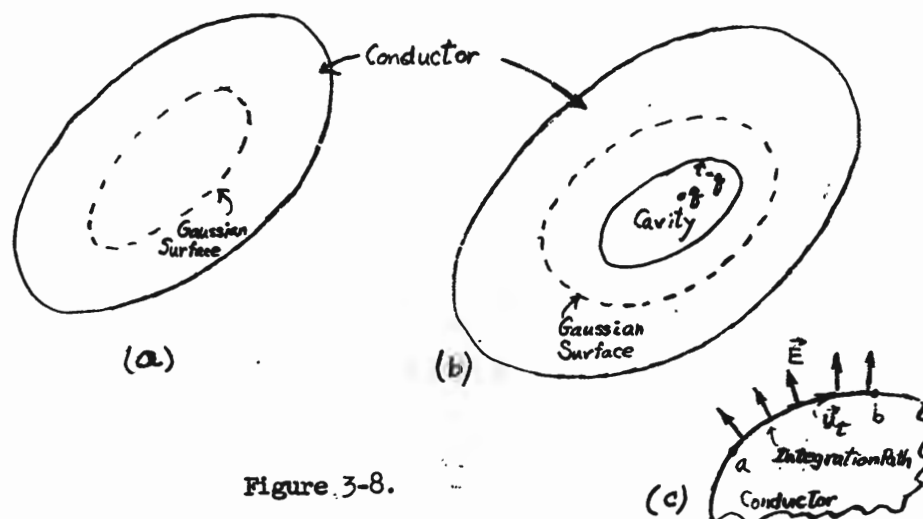


Figure 3-8.

If the conductor is not solid but has an internal cavity as shown in Fig. 3-8b, it is possible that there shall be a charge on the inside surface. This will occur if there is a charged object, insulated from the conductor, inside the cavity. Again using a Gaussian surface, as indicated by the dotted figure, it is known that the charge inside this surface must be zero. Thus, the inside surface must carry a charge $-q$. If the conducting body is uncharged, there must then be a charge q on the outside surface to make the net charge zero. However, this is not necessary; the outside surface can have zero charge, in which case the conducting body will have a net charge of $-q$.

The property that $\vec{E}$ is zero in a conductor is a consequence of the existence of "free" electrons, electrons which are free to move about. Thus, any attempt to create an electric field in a conductor results in motion of electrons until they reach a surface and distribute themselves in such a way as to cancel the original cause. This description emphasizes why it is important to say $\vec{E}$ must be zero under static conditions. While electrons are distributing themselves, currents are flowing, and $\vec{E}$ is

not zero during the process of charge flow.

Another consequence of the fact that a conducting body is at constant potential is obtained by considering Eq. (3-15) applied between two points on the surface of a conductor, as in Fig. 3-8c. For any two points a and b we must have

$$\int_a^b \vec{E} \cdot \vec{u}_t \, dl = 0$$

where the path of integration is along the surface. In general, $\vec{E}$ is not zero, and so the only way for this integral to be zero is for $\vec{E} \cdot \vec{u}_t$ to be zero. Since, $\vec{u}_t$ is tangent to the surface, $\vec{E} \cdot \vec{u}_t$ can be zero only if $\vec{E}$ is normal to the surface, as indicated in Fig. 3-8c.

Since we now know that charge always resides on the surface of a conductor, under static conditions, it can be concluded that in the arrangement of concentric spherical shells described in Fig. 3-7, it makes no difference how thick the shells are. In fact, the inside shell can be replaced by a solid sphere, with no change in the electric field.

We have deduced the property of charge appearing on the surface of a conductor from Gauss' law, which in turn depended upon two postulates concerning $\vec{E}$. At this point it is appropriate to observe that a very accurate experiment is possible to confirm that charge really does reside on the surface.* Thus, this experiment provides a confirmation of the original postulates.

3-5. Capacitance

Return now to a consideration of the two concentric spheres of Fig. 3-7, when the inside sphere carries a charge q , and the outside sphere a charge $-q$. Equation (3-14) gives

$$V_{ab} = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{R_1} - \frac{1}{R_2} \right)$$

as the potential difference between the spheres. Although this equation was derived for the potential difference between points a and b, in the

*This is the famous "ice pail" experiment of Kelvin.

light of Sec. 3-4 it is evident that V_{ab} is the potential difference between any point a on the inside sphere and any point b on the outside sphere.

Whenever two charged bodies carry charges of equal magnitude and opposite sign, the potential difference between them is proportional to the charge, with a proportionality factor which depends on the geometrical arrangement. Equation (3-14) is one such example, for the concentric sphere arrangement. This general proportionality can be written

$$V_{ab} = \left(\frac{1}{C}\right)q \quad (3-16)$$

where $1/C$ is the proportionality factor. C is called the capacitance, and when charge is in coulombs and V is in volts, C is in farads. The dependence of C on geometrical arrangement is illustrated by the present example. From Eq. (3-14) we get

$$C = \frac{4\pi\epsilon_0(R_1R_2)}{R_2 - R_1} \quad (3-17)$$

for concentric spheres.

This formula is not of great practical importance because the arrangement of concentric spheres is not practical. However, as we shall see in the next paragraph, it provides a means of analyzing the more practical ~~case of two parallel plates.~~

We now modify the Eq. (3-14) to give

$$V_{ab} = \frac{q}{4\pi\epsilon_0} \left(\frac{1}{R_2} - \frac{1}{R_1}\right) \quad (3-18)$$

and refer to Fig. 3-9 which shows two sections cut out of the concentric spherical shells. This is an hypothetical operation in which it is assumed that the charge distribution is not affected; that is, that the charge that was originally on the cut out sections remains on them. The entire spherical surfaces remain in place, but the two sections are insulated from them. The field distribution will remain undisturbed, and V_{ab}

will be the potential difference between the sections. In the figure, the outside section is only partly shown, but its edges are determined by extending radial lines from the edges of the inside section. Let A be the area of the inside section and q_s its charge, which will be q times the ratio of A to $4\pi R_1^2$, the area of the entire sphere. This calculation gives

$$q_s = \frac{qA}{4\pi R_1^2} \quad (3-19)$$

Likewise, recognizing that the area of the outside section is $(R_2/R_1)^2 A$ and that the area of the entire outside sphere is $4\pi R_2^2$, we have

$$\text{Charge on the outside section} = -q \left(\frac{R_2}{R_1} \right)^2 A \left(\frac{1}{4\pi R_2^2} \right) = - \frac{qA}{4\pi R_1^2}$$

The significance of the last calculation is that the charge on the outside section is the negative of Eq. (3-19), and thus that the two spherical

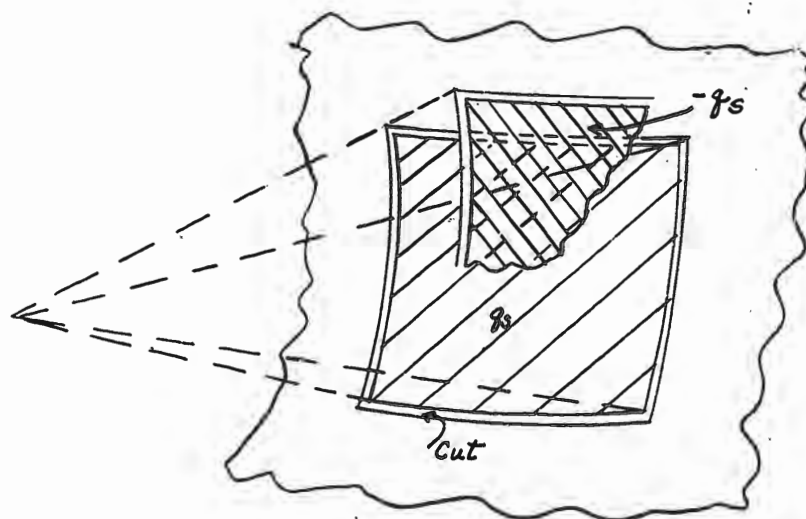


Figure 3-9.

sections have charges of equal magnitude and opposite sign, satisfying the condition for obtaining capacitance between the two sections, as the ratio of charge to potential difference. Equation (3-19) can be solved for q which

can be substituted into Eq. (3-18) to give

$$V_{ab} = \frac{q_s}{A\epsilon_0} \left(\frac{R_1}{R_2} \right) h$$

where h has been used for the separation between the shells. Finally, the capacitance is

$$C = \frac{\epsilon_0 A}{h} \left(\frac{R_2}{R_1} \right) \quad (3-20)$$

This result can be extended to yield the capacitance between two parallel plane plates by allowing R_1 and R_2 to approach infinity, while keeping $R_2 - R_1 = h$ constant. Geometrically this approaches the situation portrayed in Fig. 3-10a, where the plates are assumed to extend to infinity. In practice, there will be very little change in the central isolated portions if the plates do not extend to infinity. This makes it possible actually to construct the device and to attach wires to it, as in Fig. 3-10b. In this figure, V_{ab} is the battery voltage, and q_s is the integral with respect to time of current i .

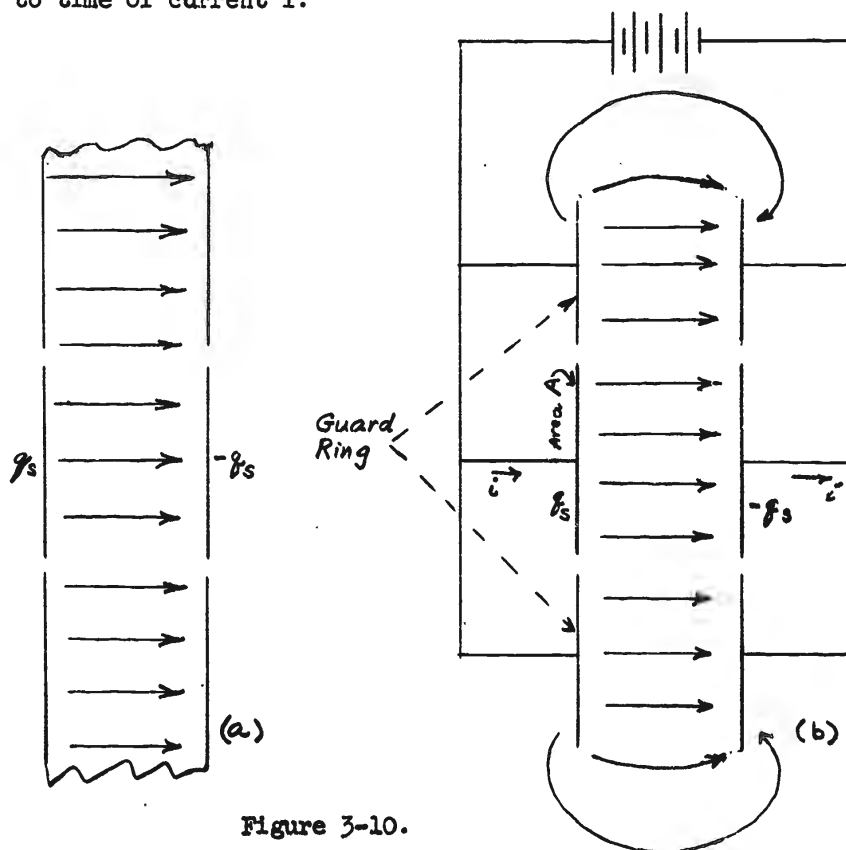


Figure 3-10.

3-20

Referring to Eq. (3-20), as R_1 and R_2 both approach infinity, their ratio approaches

$$\frac{R_1}{R_2} = \frac{R_1}{R_1 + h} = \frac{1}{1 + h/R_1} = 1$$

and so Eq. (3-20) becomes

$$C = \frac{\epsilon_0 A}{h} \quad (3-21)$$

for the capacitance between two parallel plates of area A , and spacing h . However, this is accurate only for the arrangement of Fig. 3-10b, where there is a "guard ring" around the plates which serves to keep the field lines parallel between the plates.

We can see that Eq. (3-21) cannot be exactly accurate for an isolated pair of plates, for it requires the E lines to be parallel between the plates, this being the limiting condition of the radial lines for the spherical case. Thus, referring to Fig. 3-11a, assume the plates are isolated and that the lines are parallel between the plates and zero outside. Then the integral of $\vec{E} \cdot \vec{u}_t$ along path (1) would be V_{ab} , but along path (2) the integral would be zero because $\vec{E}$ was assumed to be zero in the region of this path. However, this is impossible, since the potential difference must be the same when calculated by any path. In reality, the field lines must "fringe" out at the edges, as indicated in Fig. 3-11b, in order to make the integral of $\vec{E} \cdot \vec{u}_t$ independent of path. In practice, h is usually very small compared with linear dimensions of the plates, and in that case this fringing has

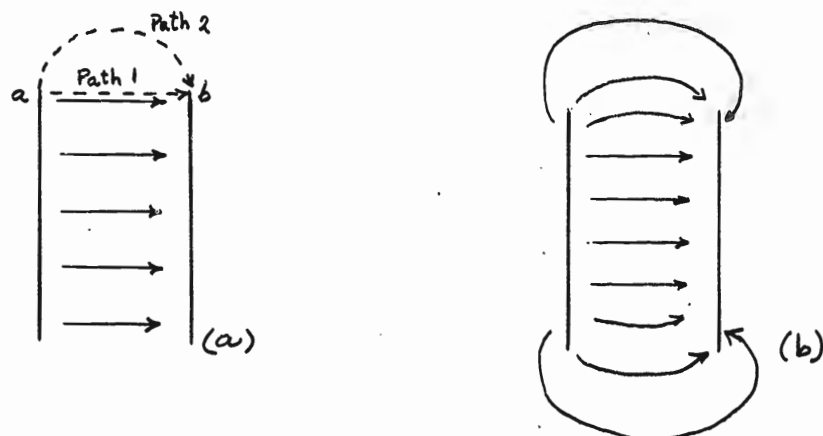


Figure 3-11.

1 a very small effect, and Eq. (3-21) can be used for the capacitance,
 2 to a high degree of accuracy.

3-6. Dielectric Materials

3 The experiment shown in Fig. 3-10b provides a means for investigating
 4 the electrical properties of insulating materials such as oils, glass,
 5 paper, plastics, etc. When the space between the plates is filled with
 6 one of these materials, it is found that for a given potential difference,
 7 the charge (integral of i with respect to t) will be greater than when the
 8 space between the plates is occupied by air. Let k_e be the factor by which
 9 the charge increases. The capacitance will then be given by

$$C = \frac{k_e \epsilon_0 A}{h} \quad (3-22)$$

4

The factor k_e is a numerical ratio, and is called the relative permittivity
 (or dielectric constant). Obviously, the relative permittivity of air is
 unity.

5

We shall now consider the consequences of applying Gauss' law to the
 parallel plate capacitor, when the space between the plates is filled
 with a dielectric material. Reference is made to Fig. 3-12, where the
 situations at (a) and (b) are identical, except for the inclusion of
 dielectric material at (b). Batteries of identical voltage V are used
 so that $\vec{E} = E \vec{u}_n$ (where $E = V/h$) is the same for both. The guard rings
 force the fringe flux to be far away from the central plates, so that $\vec{E}$
 is essentially zero except in the region between the plates.

7

The dotted rectangle shown in each figure is a side view of a hypo-
 thetical rectangular box which serves as a Gaussian surface. Lines
 labeled ΔA represent surface areas which are normal to $\vec{E}$. Let σ be the
 density of the charge on the surface of the plate in Fig. 3-12a. Applying
 Gauss' law to the Gaussian surface in Fig. 3-12a, we observe that the
 surface integral is zero over surfaces b , c , and d ; it is zero over b
 and c because $\vec{E}$ is parallel to these surfaces, and it is zero over d
 because $\vec{E}$ is zero on this surface. Thus, Gauss' law gives

9

$$\oint \vec{E} \cdot \vec{u}_n dA = (\vec{E} \cdot \vec{u}_n) \Delta A = \frac{\sigma \Delta A}{\epsilon_0} \quad (3-23)$$

3-22

and

$$\vec{E} \cdot \vec{u}_n = \frac{\sigma}{\epsilon_0} \quad (3-24)$$

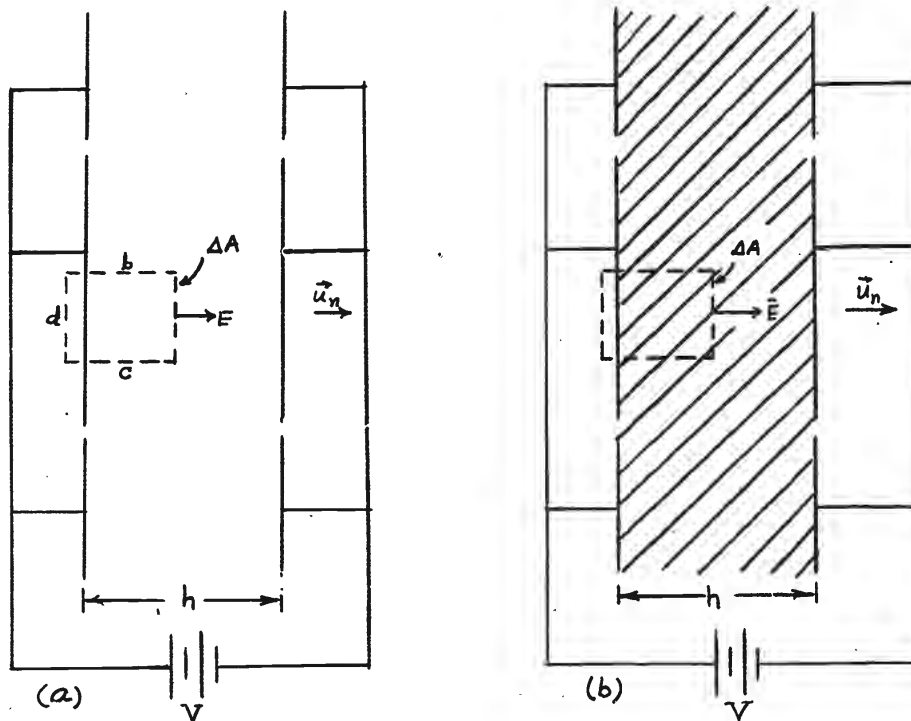


Figure 3-12.

Now observe that $E = V/h$ must be the same for Fig. 3-12b, since V and h are the same. Thus, knowing E for Fig. 3-12b, Gauss' law can be used to find the charge enclosed. When Eq. (3-24) is substituted into the surface integral of $\vec{E} \cdot \vec{u}_n$, again observing that $\vec{E} \cdot \vec{u}_n$ is zero except on face ΔA , we get

$$\iint \vec{E} \cdot \vec{u}_n da = \frac{\sigma}{\epsilon_0} \Delta A \quad (3-25)^*$$

which shows that the enclosed charge is $\sigma \Delta A$, and that the surface charge density is σ , the same as in Fig. 3-12a. However, recalling Eq. 3-22, the

*Equation (3-25) is identical with Eq. (3-23), but is obtained from a different starting point. In Eq. (3-23) Gauss' law is used to obtain E from σ , whereas in Eq. (3-25) Gauss' law is used to find σ from E .

capacitance of Fig. 3-12b must be k_e times the capacitance of Fig. 3-12a, and since $Q = CV$, it follows that the surface charge density plate in Fig. 3-12b should be

$$\sigma' = k_e \sigma \quad (3-26)$$

rather than σ , as predicted by Gauss' law. The conflict implicit in these separate conclusions is resolved by postulating that σ' , the charge density on the conducting plate, is modified by a surface charge density σ_b on the surface of the dielectric. Thus, in Fig. 3-12b the total surface to be used in Gauss' law, which we know from Eq. (3-25) to be σ , is also $\sigma_b + \sigma'$. Thus

$$\sigma = \sigma_b + \sigma'$$

and with Eq. (3-26) we have

$$\sigma_b = -(k_e - 1)\sigma \quad (a) \quad (3-27)$$

$$= -(1 - \frac{1}{k_e})\sigma' \quad (b)$$

Since $\sigma = \epsilon_0 E$, σ_b is also given by

$$\sigma_b = -(k_e - 1)\epsilon_0 E \quad (3-28)$$

Experiment confirms this conclusion. Charge σ_b can be explained as follows: due to the force of the E field, the positive and negative atoms of the dielectric are separated slightly, as shown in exaggerated form in Fig. 3-13. This phenomenon results in the appearance of a layer of negative charge on the left hand face, and a similar layer of positive charge

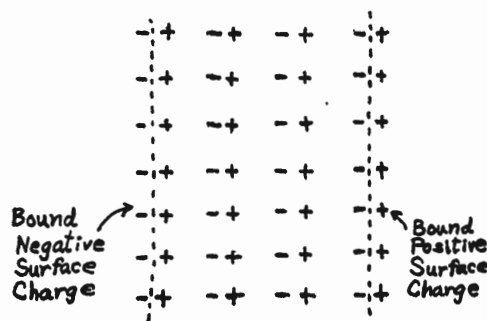


Figure 3-13.

charge on the right hand face. Observe that the negativeness of the charge on the left is in agreement with the negative sign in Eq. (3-25). The atoms represented in this figure are said to be polarized, and the phenomenon is called polarization. Since σ_b is due to displaced charges which are parts of neutral atoms of a non-conducting medium, they cannot actually be removed from the surface, say by conduction to a plate. Accordingly, σ_b is called a surface density of bound charge (as opposed to free charge). If the E field is reduced to zero, the positive and negative charges of each atom "spring back" together, and the surface charge disappears.

It appears that Gauss' law, as expressed by Eq. (3-8), is universally valid, even when dielectrics are present, if the charge enclosed within the surface is the sum of all free and bound charges.* It is convenient and useful to have a modified form of Gauss' law in which only free charge will be included. To this end, suppose a new vector

$$\vec{D} = k_e \epsilon_0 \vec{E} \quad (3-29)$$

is defined, and consider the surface integral of $\vec{D} \cdot \vec{u}_n$ over the Gaussian surface in Fig. 3-12b. In this case, referring to Eq. (3-24), we have

$$\vec{D} \cdot \vec{u}_n = k_e \epsilon_0 \left(\frac{\sigma}{\epsilon_0} \right) = k_e \sigma = \sigma' \quad (3-30)$$

and thus

$$\oiint \vec{D} \cdot \vec{u}_n da = \sigma' \Delta A$$

which is the free charge inside. Although this was shown for a particularly simple case, it can be shown that

$$\oiint \vec{D} \cdot \vec{u}_n da = q \quad (3-31)$$

is true for any closed surface, where q is the free charge inside. It is immaterial whether or not the surface passes through a dielectric object.

*It can be shown that in the general case, where $\vec{E}$ is not necessarily normal to a dielectric surface, Eq. (3-28) becomes

$$\sigma_b = -(k_e - 1) \epsilon_0 E_n$$

where E_n is the normal component of E directed into the dielectric.

Equation (3-30) gives additional information about how D and the free charge are related at a dielectric surface. In terms of Fig. 3-12b, to which that equation applies, we have $\vec{D} \cdot \vec{u}_n = \sigma'$, or

$$D_n = \sigma' \quad (3-32)$$

where $D_n = \vec{D} \cdot \vec{u}_n$ is the component of $\vec{D}$ inward and normal to the surface. This is generally true, even if the surface is not plane, or if $\vec{D}$ is not normal to the surface. Then the relationship is written

$$D_n = \sigma' \quad (3-33)$$

where the subscript (n) is a reminder that this is a normal component. Thus it is seen that $\vec{D}$ has a very simple relationship to the surface density of free charge.

Note from Eq. (3-29) that in air

$$\vec{D} = \epsilon_0 \vec{E}$$

so that in air Eq. (3-30) reduces to Eq. (3-8). Thus, Eq. (3-29) is generally valid for all cases, and is a second form of Gauss' law. It is usually the preferred form for the solution of field problems.

The vector $\vec{D}$ is called the dielectric displacement vector. From Eq. (3-30) it is apparent that in the MKSC system of units, $\vec{D}$ is in coulombs/m².

3-7. Composite Capacitors

To show that the invention of D is not merely an intellectual exercise, let us use it to determine the capacitance of the parallel plate arrangement shown in Fig. 3-14a. We make the simplifying assumption that the E lines are parallel, as they would be if guard rings were present. Using a Gaussian surface consisting of a rectangular box represented by the dotted rectangle, we note that there is no free charge on the dielectric surface (an assumption of the problem statement). Thus,

$$\vec{D}_1 \cdot \vec{u}_n \Delta A = \vec{D}_2 \cdot \vec{u}_n \Delta A$$

and so $\vec{D}$ is the same in air as in the dielectric, and we shall call it $\vec{D}$.

3-26.

Now, referring to Eq. (3-29), we have

$$\vec{E}_1 = \frac{\vec{D}}{\epsilon_0}, \quad \vec{E}_2 = \frac{\vec{D}}{k_e \epsilon_0}$$

and, finally,

$$V = \frac{h}{2}(\vec{E}_1 + \vec{E}_2) \cdot \vec{u}_n = \frac{h}{2}\left(\frac{1}{\epsilon_0} + \frac{1}{k_e \epsilon_0}\right) \vec{D} \cdot \vec{u}_n \quad (3-34)$$

But $\vec{D} \cdot \vec{u}_n$ equals the surface charge density σ , and the total charge on the surface of area A , is

$$q = \sigma A = (\vec{D} \cdot \vec{u}_n) A$$

Finally, from Eq. (3-34)

$$\frac{h}{2}\left(\frac{1}{\epsilon_0} + \frac{1}{k_e \epsilon_0}\right) \frac{q}{A} = V$$

and the capacitance (q/V) is

$$C = \left(\frac{\epsilon_0 A}{h}\right) \frac{2k_e}{1+k_e} \quad (3-35)$$

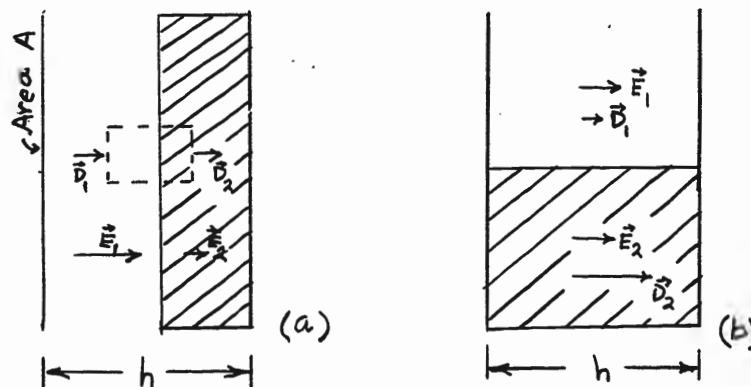


Figure 3-14.

1 Observe that if $k_e = 1$ this reduces to the previous formula for an air capacitor, given by Eq. (3-21), as we should expect.

2 Another example is shown in Fig. 3-14b. Here the dielectric material extends all the way between the plates, but covers only half of area A. In this case $\vec{E}_1 = \vec{E}_2$, since $(\vec{E}_1 \cdot \vec{u}_n)h = (\vec{E}_2 \cdot \vec{u}_n)h$, by virtue of each plate being at constant potential. Let $\vec{E}$ be this common value of $\vec{E}_1$ and $\vec{E}_2$. Now, in the two sections,

$$3 \quad \vec{D}_1 = \epsilon_0 \vec{E} \quad \text{and} \quad \vec{D}_2 = k_e \epsilon_0 \vec{E}$$

The total charge on the left hand plate, from Eq. (3-21) is

$$4 \quad q = \frac{A}{2}(\vec{D}_1 \cdot \vec{u}_n + \vec{D}_2 \cdot \vec{u}_n) = \frac{A}{2}(\epsilon_0 + k_e \epsilon_0) \vec{E} \cdot \vec{u}_n$$

But $\vec{E} \cdot \vec{u}_n = V/h$, giving

$$5 \quad q = \frac{A}{2} \epsilon_0 (1 + k_e) \frac{V}{h}$$

and for C we get

$$6 \quad C = \frac{\epsilon_0 A (1 + k_e)}{h} \quad (3-36)$$

7 It should be fairly obvious that dielectric materials can be useful to provide higher values of capacitance than would otherwise be possible. They do this in two ways; first, by the introduction of the factor k_e in the formula for C, and second, by permitting smaller values of h, for the reason we shall now describe. Dielectric materials, including air, have a property known as dielectric breakdown strength. This is the value of E at which a spark will jump through the material. It is not a precisely determinable quantity because it depends on many factors, such as humidity, smoothness of surface, presence of internal voids, and the like. For air a reasonable figure is 3,000 volts/mm, while solid materials can go as high as 28,000 volts/mm (the handbook figure for polystyrene, a common dielectric material). Thus, if a sheet of dielectric material is used, the plate spacing for a given voltage can be much smaller without fear of dielectric

breakdown. Also, reducing h increases C , as can be seen for the cases investigated here, because C is inversely proportional to h . In this way, dielectric materials can be very useful in saving space in capacitor design.

Some typical values of dielectric constants (relative permittivities) are given in the following table.

Material	k_e
Air (760 mm pressure)	1.0006 (usually taken as 1)
Cellulose Nitrate	11.4
Pyrex Glass	4.5
Mica	7.2
Phenol	5.5
Polyethelene	2.26
Polystyrene	2.56
Neoprene	6.7
Benzene	2.15
Petroleum oils	2.2
Ethyl Alcohol	25
Methyl Alcohol	31
Distilled Water	81

3-8. Dielectric Hysteresis

The relationship

$$\vec{D} = k_e \epsilon_0 \vec{E}$$

is approximately true for most materials. However, there are some materials in which there is what might be called a "sluggishness" in the return of the atoms to the unpolarized state when the external polarizing influence has been removed. In other words, the condition portrayed in Fig. 3-13 will persist to some extent, resulting in a partial retention of bound surface charge. Such materials are called electrets. The bound surface charge creates an electric field of its own, and this fact makes it possible to detect this state of "permanent electrification". This phenomenon has possible application for memory devices

1 in computers, although because the similar action in magnetic materials
 is so much more pronounced, they are more often employed for such a
 purpose.

2 The phenomenon just described can be portrayed as a graph of
 D vs. E, as in Fig. 3-15. If E is increased, starting with unpolarized
 material, D will increase along the straight line. However, if E is then
 reduced from point P, D will not return to zero when E is zero, and if
 3 E is carried into the negative region, and then back to point P, a closed
 loop will be formed. This is called a dielectric hysteresis loop. It
 can be shown that the area of this loop is proportional to the energy
 lost in the process. This energy loss can be viewed as being due to
 internal "friction" of the molecules as they react to the changing electric
 4 field.

One other possible deviation from Eq. (3-27) should be mentioned.
 Some substances do not have $\vec{D}$ and $\vec{E}$ in the same direction, and are
 called nonisotropic.* Advanced mathematics is required to treat this
 5 situation, and so no further consideration of it can be given here.

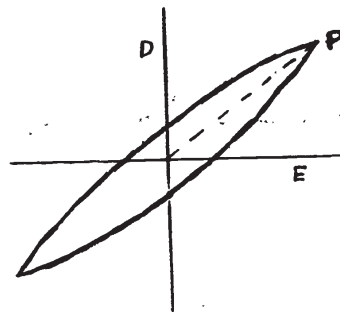


Figure 3-15.

3-9. Resistance Capacitance Circuits

8 Many practical applications involve circuits which include capacitors.
 We shall now briefly consider the transient phenomena which arise when an
 attempt is made to change the voltage across a capacitor. The essential

9 *Nonisotropic materials have different properties along different axes. This
 also applies to properties other than electrical, such as mechanical
 deformation and thermal conduction properties.

problem is exemplified by the circuit of Fig. 3-16a. A capacitor is uncharged ($v_c=0$) and then the switch is closed connecting a battery of voltage V , through a resistor R . We are asked to determine how v_c changes with time, subsequent to the closing of the switch. Zero on our time scale is arbitrarily chosen as the instant when the switch is closed.

We recall that the charge q accumulated on the top plate is related to v_c by

$$q = C v_c$$

and also that

$$i = \frac{dq}{dt}$$

Thus, i and v_c are related by

$$i = C \frac{dv_c}{dt} \quad (3-37)$$

After the switch is closed, Kirchhoff's voltage law gives

$$i R + v_c = V$$

or, in terms of the variable v_c from the previous equation,

$$RC \frac{dv_c}{dt} + v_c = V \quad (3-38)$$

This is a differential equation, to be solved for v_c . By writing this as an explicit expression for dt/dv_c we have

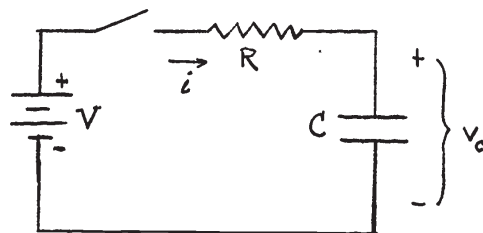
$$\frac{dt}{dv_c} = \frac{RC}{V-v_c}$$

From this form we can recognize the antiderivative

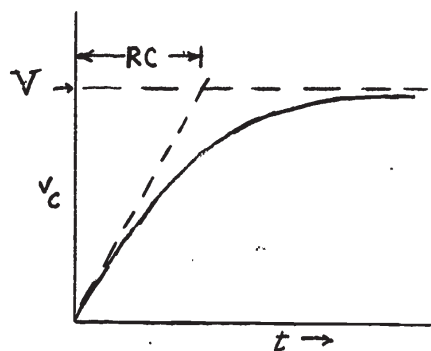
$$t = -RC \ln(V-v_c) + \ln K \quad (3-39)$$

where K is an arbitrary constant. Since v_c is wanted, we can write

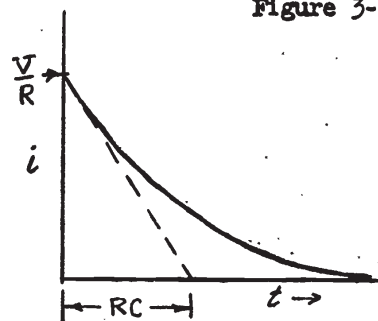
$$-\frac{t}{RC} = \ln\left(\frac{V-v_c}{K}\right)$$



(a)



(b)



(c)

Figure 3-16

and

$$\frac{V - v_c}{K} = e^{-t/RC}$$

or

$$v_c = V - K e^{-t/RC} \quad (3-40)$$

This is not yet the required solution, because K is an unknown constant which must be evaluated by introducing the initial condition; the value of v_c when $t = 0$. The original statement of the problem gave the information that $v_c = 0$ before the switch is closed. However, Eq. (3-40) begins to apply at the instant when the switch is closed (i.e. for $0 \leq t$), since it was derived from an equation written for a closed circuit. In connection with this question, observe that if v_c were to experience a sudden jump at $t = 0$, this would mean an infinite derivative (dv_c/dt) and hence Eq. (3-31) shows that the current would be infinite. However, with a finite source voltage, this is impossible, and hence we conclude that a sudden change in v_c is impossible. Thus,

3-32

1 $v_c = 0$ is the appropriate value to use in Eq. (3-40) to correspond to $t = 0$.
Substituting these gives

$$0 = V - K$$

2 and thus the required equation for v_c is

$$v_c = V(1 - e^{-t/RC}) \quad (3-41)$$

3 A graph of this function is shown in Fig. 3-16b. This figure includes
a geometrical interpretation of the parameter RC , as the time it would take
for the voltage to change to its final value, if it continued to change at its
initial rate. RC is called the time constant of the circuit, and the final
value attained by v_c is called the steady state value.

4 In addition to learning that in such a circuit the voltage across a
capacitor changes exponentially, one of the important observations to be made
is that the time required for a change of capacitor voltage to take place
depends on the product RC . An interpretation slightly different from the
5 graphical one of Fig. 3-16b involves finding the time interval required
for v_c to go through 90% of its total change. This is the value of t at
which $e^{-t/RC} = .1$, which occurs when $t = 2.3RC$, approximately. Time constants
in practical circuits vary from microseconds, in pulse transmission circuits,
6 to many seconds in certain filter applications.

Equation (3-38) can readily be used to obtain the current,

$$i = C \frac{dv_c}{dt}$$

$$= \frac{V}{R} e^{-t/RC}$$

8 A graph of this exponential is shown in Fig. 3-16c.

9 The idea of a change in v_c is further illustrated by the example of
Fig. 3-17a. Battery V_1 has been connected for a long period of time so that
at $t = 0$ (when the switch is opened) v_c has the value V_1 . Opening the switch
yields a situation in which the total battery voltage is $V_1 + V_2$, and the
total resistance is $R_1 + R_2$. These can be substituted, respectively, for V
and R in Eq. (3-40), to give

1

$$v_c = V_1 + V_2 - K e^{-t/(R_1+R_2)C} \quad (3-42)$$

In this case the initial condition is $v_c = V_1$, when $t = 0$, and so

2

$$V_1 = V_1 + V_2 - K$$

which determines that $K = V_2$. Thus, Eq. (3-42) becomes

3

$$v_c = V_1 + V_2 \left[1 - e^{-t/(R_1+R_2)C} \right] \quad (3-43)$$

4

A graph of this function is shown in Fig. 3-17b. The previous interpretation of the time constant still applies, if it is applied to the change of v_c between the initial and final value.

5

Equations (3-41) and (3-43) are both of the same form, where the change in voltage is represented by a quantity which varies like $(1 - e^{-\alpha t})$ x total change of v_c , with $1/\alpha$ as the time constant. It can be shown that this is typical of the result for any circuit having only one capacitor, and any number of resistors. D-c circuit analysis can be used to find the total change of v_c , and the time constant is C x the equivalent resistance of the circuit connected to C , which results when all sources are reduced to zero.

6

3-10. Energy Stored in an Electric Field

7

In Fig. 3-18, a parallel plate capacitor is charging from q_1 to q_2 , over a time interval from t_1 to t_2 . If q is the charge, and V is the voltage, the current is dq/dt , and thus the instantaneous power input to the capacitor is

8

$$p = v \frac{dq}{dt}$$

9

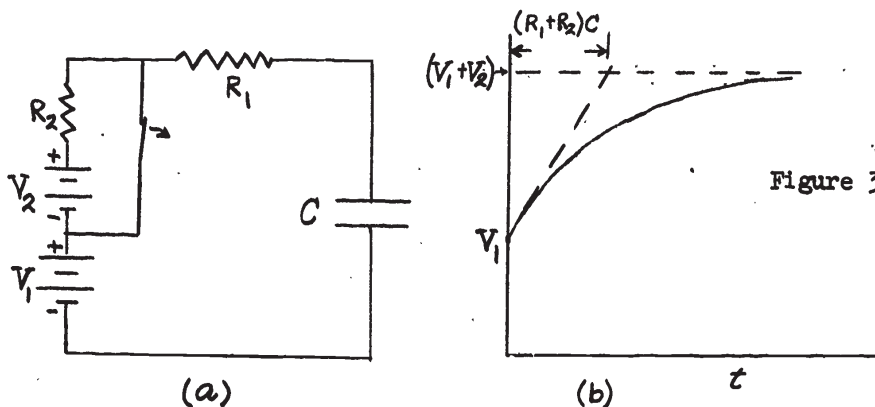


Figure 3-17

1 The energy supplied to the capacitor is

$$W = \int_{t_1}^{t_2} p \, dt = \int_{t_1}^{t_2} V \frac{dq}{dt} \, dt = \int_{q_1}^{q_2} V dq \quad (3-44)$$

2

In order to convert this to an expression in terms of E and D , we recall that

$$E = \frac{V}{h} \quad \text{and} \quad q = AD$$

3

The second of these expressions is obtained from Eq. (3-32), which is applicable because D is normal to the surface, and q is the free charge (i.e. charge on the plate). Using these in Eq. (3-44), we get

4

$$W = Ah \int_{D_1}^{D_2} E \, dD \quad (3-45)$$

5

The factor Ah is the dielectric volume, and thus

$$w = \int_{D_1}^{D_2} E \, dD \quad (3-46)$$

6

can be viewed as the density of energy storage in the dielectric.

7

The integral of Eq. (3-46) is interpreted in Fig. 3-20a, for an electret which is being electrified from an initially unelectrified state (i.e. along the curve which starts from the origin). The change of energy density associated with increasing D from D_1 to D_2 is represented by the shaded area projected back to the vertical axis, as in Fig. 3-19a. Now suppose D is decreased again to D_1 . In this case the energy density change decreases by an amount equal to the doubly shaded area. This energy is returned to the circuit, and thus the shaded area between the two curves is energy lost. Similarly, if electrification is carried around a complete loop, as in Fig. 3-19b, the energy lost for this cycle of operation is equal to the loop area. Of course, this energy loss, which is called hysteresis loss, causes heating. In many applications hysteresis loss is negligible in dielectrics

8

9

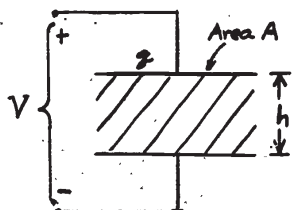


Figure 3-18.

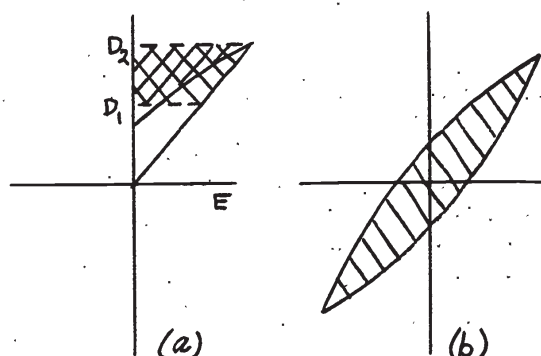


Figure 3-19.

because for most materials the loop area is very small. However, when capacitors are used in a-c circuits, hysteresis power loss can become very high because power loss is energy loss per cycle multiplied by frequency. Thus, a dielectric which behaves with negligible loss at 60 cps. may be totally unsatisfactory at 10^9 cps.

For the important case where $D = k_e \epsilon_0 E$ if we consider the energy increase in bringing D and E from zero to some specific values, from Eq. (3-46) we have

$$w = k_e \epsilon_0 \int_0^E E dE = \frac{k_e \epsilon_0 E^2}{2}$$

or

$$w = \frac{DE}{2} \quad (3-47)$$

This lost energy relationship is applicable to any dielectric in an electrified state. If it is part of a capacitor the energy can be given in terms of C, as we can see by observing that $i = C dv/dt$ so that the integral of the power is

$$W = \int_0^t p dt = C \int_0^t v_c \frac{dv}{dt} dt = C \int_0^v v dv$$

3-36

1 or

$$W = \frac{1}{2} C V^2 \quad (3-48)$$

2 This energy is completely stored, since in saying that C can be used we have implied that $q = C v_c$ and hence that D is proportional to E, which in turn means there is no hysteresis loop.

3

Chapter 4

ELECTROMAGNETISM

Introduction

In this chapter we shall deal with two basic physical phenomena. First, the phenomenon of mechanical forces (and/or torques) exerted on current-carrying circuits either in the vicinity of another current-carrying circuit, or a permanent magnet. Second, the creation of an electric potential difference as the result of motion of a conductor, or as a result of a changing current in another circuit. The first of these phenomena is the basis of the subject of magnetism. Since the second involves an interrelationship between electrical and magnetic effects, it is called an electromagnetic phenomenon.

Almost every practical electrical device involves electromagnetic phenomena. For example, the electric power we use is generated by the motion of a conductor in a magnetic field, the diaphragm of a telephone receiver is actuated by magnetic forces, magnetically operated relays are used to open and close circuits in many applications such as motor controls and telephone switching circuits, and, of course, electric motors operate on the principle that a force is exerted on a current-carrying conductor in a magnetic field. Thus, these few illustrations indicate that a study of electromagnetic phenomena is important.

4-1. Basic Magnetic Experiments

Consider an experiment in which two long straight parallel wires carry currents, as shown in Figure 4-1(a). Wire (1) is rigidly supported, and provision is made to measure the force on wire (2). The following observations can be made:

- 1) When i_1 and i_2 are in the same direction (either both in the reference directions shown, or both opposite), wire (2) experiences a force toward wire (1).
- 2) If either current is reversed (but not both), the force on (2) will be away from (1).
- 3) In either case, the magnitude of the force is proportional to the magnitude of the product $i_1 i_2$.

4-2.

4) The force is proportional to $1/r$, where r is the separation between wires.

A second experiment can be conducted as shown in Figure 4-1(b). Here, the wires are at right angles and in the same plane, one passing in back of the other by virtue of a small semicircular jog in one of them. In this case, there is found to be no net force on wire (2), but there is found to be a torque tending to rotate wire (2) into a position parallel with wire (1) so that i_1 and i_2 will be in the same direction. Thus, for currents in the reference directions indicated in Figure 4-1(b), the torque on wire (2) is found to be clockwise.

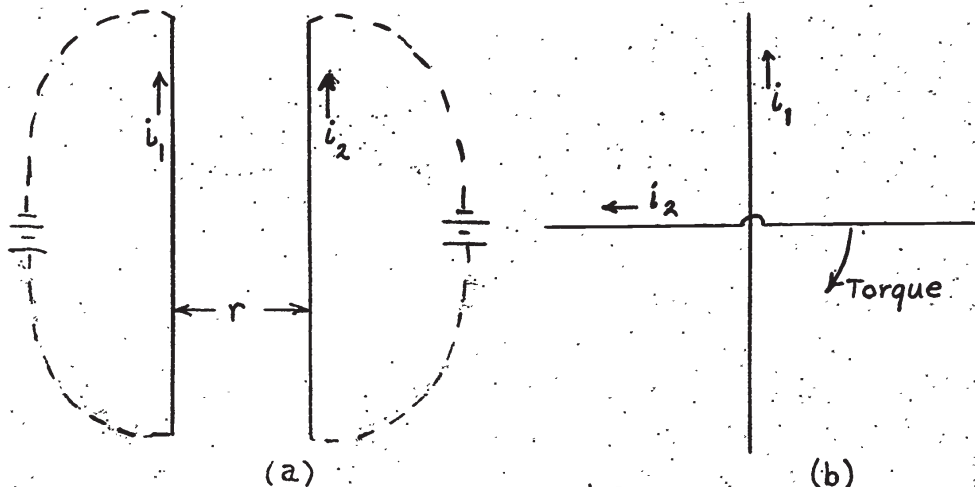


Figure 4-1.

4-2. Flux Density

In Figure 4-1(a), wire (2) experiences a force, and so it is reasonable to assume that there is a force on a short length Δl , as indicated in Figure 4-2(a). With fixed i_1 , the force is toward or away from wire (1), depending on the direction of i_2 . This much of the experiment suggests, perhaps, that the force on Δl can be described by defining a vector directed radially from wire (1). However, when we go to the crossed wires, as in Figure 4-2(b), we see that in order to provide the observed torque, the force on an element Δl must be parallel to wire (1); up or down depending on the direction of i_2 . The force on the element Δl is always perpendicular to the direction of current flow in Δl .

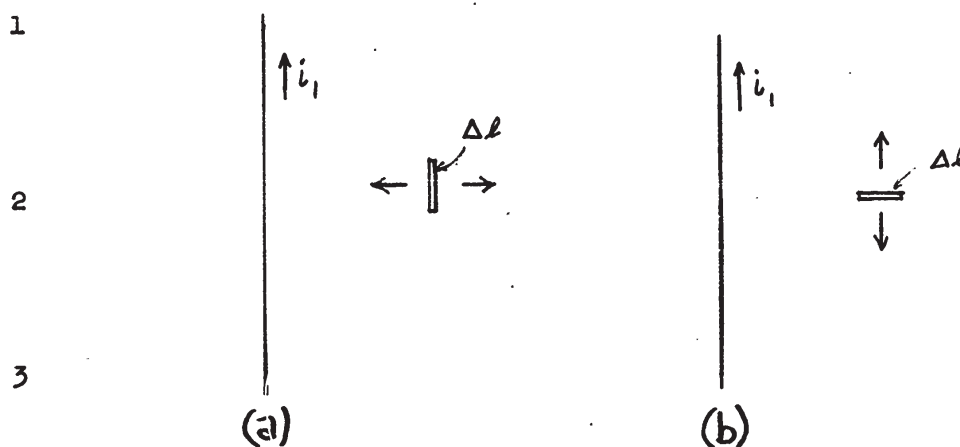


Figure 4-2.

The idea of a vector product of two vectors is used to describe the force on Δl . To see how, we shall consider the four conditions illustrated in Figure 4-2, repeated in perspective view in Figure 4-3. The direction of current i_2 , flowing in element Δl , is designated by a unit vector $\vec{u}$, and $\vec{F}$ is

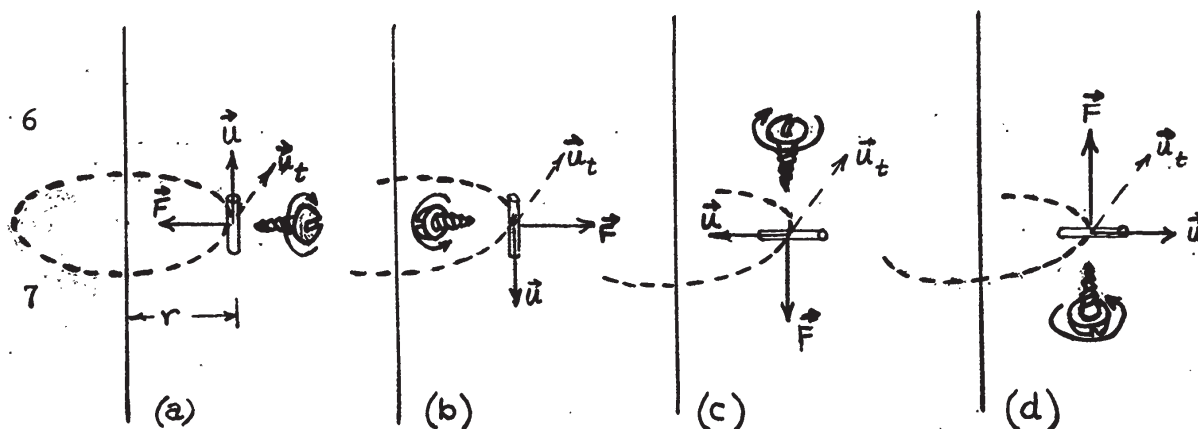


Figure 4-3.

the force vector. In each case a unit vector $\vec{u}_t$ is shown in perspective view, as a dashed line. This vector is drawn tangentially to a circle centered at wire (1), as indicated in Figure 4-3(a). It will be noted that in each case the right hand screw rule applied in the vector product

$$\vec{u} \times \vec{u}_t$$

- 1 will yield the direction of $\vec{F}$. Thus, a unique direction (that of $\vec{u}_t$) can be found, from which the direction of $\vec{F}$ can be derived, when the current direction (defined by $\vec{u}$) is known.

- 2 Regarding magnitude of the force, careful analysis of the mechanics involved will show that the magnitude is the same in all situations shown in Figure 4-3, provided that the perpendicular distance r from wire (1) to element $\Delta\ell$ is the same.

- 3 Since the force is proportional to $i_1 i_2$, this distance factor must appear explicitly in the equation for force. Also, the factor $1/r$ must appear, to account for the inverse relationship with distance obtained for Figure 4-1(a). Finally, we should expect force to be proportional to $\Delta\ell$. Thus, experimental results are satisfied by the relation

$$4 \quad \vec{F} = k \frac{i_1 i_2 \Delta\ell}{r} (\vec{u} \times \vec{u}_t) \quad (4-1)$$

- where k is a proportionality factor chosen to take care of units. This constant is usually written $k = \mu_0/2\pi$, where μ_0 is another constant called the permeability of free space. When force, current, and distance are in MKSC units (respectively newtons, amperes, meters), the value of μ_0 is

$$\mu_0 = 4\pi \times 10^{-7}$$

- 6 and so we get 2×10^{-7} for the value of k .*

- 7 *The appearance of the simple rational value 2×10^{-7} for k in Eq. (4-1) may be surprising, in view of the fact that Eq. (4-1) is presented as an experimental law. This is not an accident; it stems from the choice of e_0 which establishes indirectly the unit of charge at such a value that the coefficient is exactly 2×10^{-7} . This procedure of using a non-rational factor (e_0) to define charge, so that the factor comes out rational for B may seem to be somewhat puzzling; it would perhaps seem more logical to choose e_0 as a rational number, since electrostatic phenomena are simpler.
- 8 The answer lies in the historical development of systems of unit, and the fact that early measurements establishing the ampere were made by measurement of magnetic forces. Magnetic forces are easier to measure in the laboratory, although they are more difficult to analyze theoretically. In the MKSC system of units, adopted as the international standard in 1938, the theory was adjusted to agree with existing practical units wherever possible. Since current and charge are very fundamental, their units were retained.
- 9

1 Now we introduce this value of k , and rearrange to give

$$\Delta \vec{F} = i_2 \Delta l \left[\vec{u} \times \left(\frac{2 \times 10^{-7} i_1}{r} \vec{u}_t \right) \right]$$

2 Next, we introduce a symbol (B) to replace the quantity in parentheses and arrive at

$$\Delta \vec{F} = i_2 \Delta l (\vec{u} \times B \vec{u}_t)$$

3 where B is

$$B = \frac{2 \times 10^{-7} i_1}{r} = \frac{\mu_0 i_1}{4\pi r} \quad (4-2)$$

4 The quantity $\vec{B} = B \vec{u}_t$ is a vector in the direction of the unit vector $\vec{u}_t$ (or opposite if B is negative). It is called the flux density vector or, alternately, the induction vector. In the MKSC system of units, B is in webers per square meter.* Equation (4-2) is an expression of the Biot-Savart law, giving the flux density, at all points in space, due to a long current-carrying conductor. Thus, using the vector $\vec{B}$, the force equation is

$$\Delta \vec{F} = i_2 \Delta l (\vec{u} \times \vec{B}) \quad (4-3)$$

7 The validity of this expression has not been completely established by the experiments described, because only two orientations of the Δl elements have been considered. However, by conducting further experiments, it is found that Eq. (4-3) is indeed valid regardless of the orientation of Δl with respect to $\vec{B}$. Interpretation of the vector product shows that the magnitude of the force is

$$|\Delta \vec{F}| = \Delta l |i_2| |\vec{B}| \sin \theta$$

9 * In the older electromagnetic system of units, B is in gauss (1 gauss = 10^{-4} webers/sq.m.). In that system, if i_1 is in amperes, r is in cm., then $\mu_0 = 0.4\pi$ and in Eq. (4-3) force is in dynes.

4-6

- 1 where θ is the angle between the B vector and the conductor, and the force is normal to the plane of B and the conductor.

- 2 Plots of $\vec{B}$ at different distances from the wire yield a picture like Figure 4-4(a). At increasing distances from the wire, the arrows become shorter, illustrating, by their length, the $1/r$ relationship. The state of space in which forces are exerted on current elements (that is, regions in which $\vec{B}$ is not zero) is called a magnetic field. Figure 4-4(a) is a partial plot of the magnetic field surrounding a straight wire. An alternate method of plotting a field is shown in
- 3 Figure 4-4(b), in which only the directions are indicated by "flow lines".

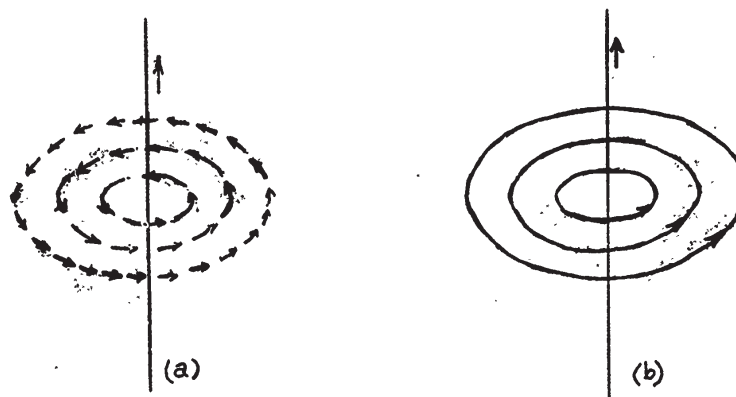


Figure 4-4.

- 6 When the force on a long conductor (as distinct from an incremental length) carrying a current i_2 is to be found, incremental forces as given by Eq. (4-3) can be summed to give the total force. Of course, this becomes the integral

$$\vec{F} = i_2 \int_{\text{conductor}} (\vec{u} \times \vec{B}) d\ell$$

- 7 where $\vec{u}$ is a unit vector tangent to the conductor, in the direction of i_2 , at each
- 8 point on the curve. The meaning of this integral can be illustrated by the example in Figure 4-5. The central dot represents the long conductor, and PQ is one side of a square loop. We are to find the force on this piece of conductor. In this case, $d\ell = dy$,

$$B = \frac{\mu_0 i_2}{4\pi \sqrt{b^2 + y^2}} \quad \text{and} \quad \sin \theta = \frac{y}{\sqrt{b^2 + y^2}}$$

giving

$$F = \frac{\mu_0 i_2}{2\pi} \int_0^h \frac{y \, dy}{b^2 + y^2} = \frac{\mu_0 i_2}{4\pi} \ln\left(\frac{b^2 + h^2}{b^2}\right)$$

This force is directed out of the paper, in accordance with the right-hand screw rule as applied to the vector product $\vec{u} \times \vec{B}$.

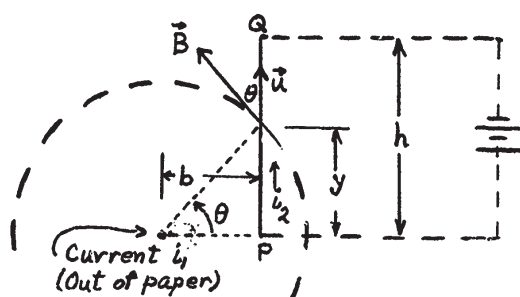


Figure 4-5.

The example chosen to introduce $\vec{B}$, the straight wire, is particularly simple. Problems of finding $\vec{B}$ in the space surrounding other wire shapes are generally more difficult, and are among the topics covered in the subject of field theory. Some illustrations of the structures of the magnetic fields around coils of wire are shown in Figure 4-6.

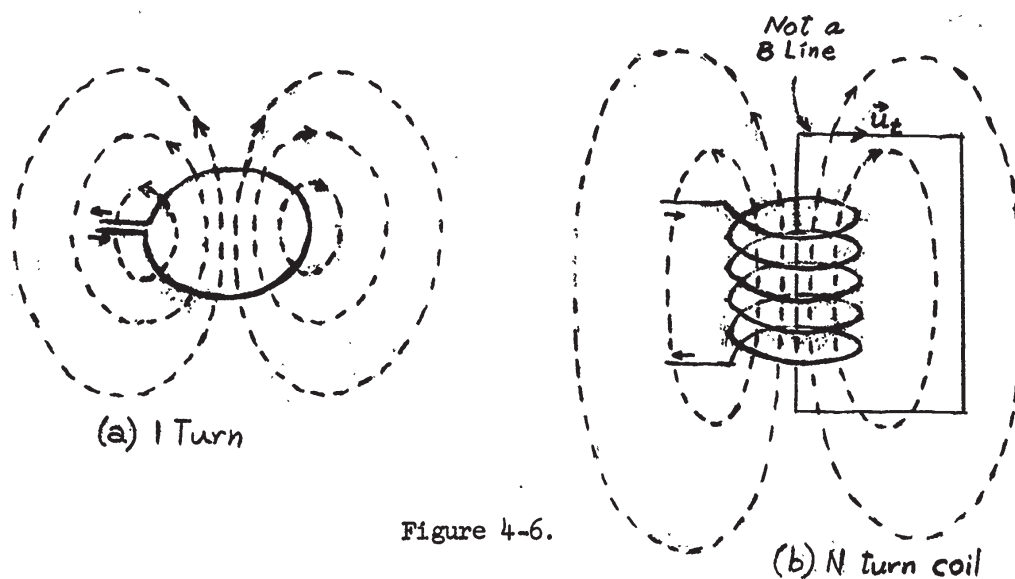


Figure 4-6.

4-8

One principle can be given in relatively simple terms here, however, which provides some idea of the relationship between B and the current which causes it. For this, we return to the relationship

$$B = \frac{\mu_0 i_1}{2\pi r}$$

for the straight wire. If we take any circle of radius r and multiply B on that circle by the circumference, we get $\mu_0 i_1$. But this product (B times circumference) is a special case of the more general formula

$$\oint B \, d\ell = \mu_0 i_1 \quad (4-5)$$

Increments $d\ell$ are along a closed B line (the 0 on the integral sign merely symbolizes the closed curve). The essential observation to be made in Eq. (4-5) is that i_1 is the current encircled by the closed curve. Furthermore, Eq. (4-5) is a special case of a more general one in which the curve along which the integral is taken is not necessarily identical with a B line. In that case B and $d\ell$ will not necessarily be in the same direction (see Figure 4-7), and so to agree with Eq. (4-5) the component $\vec{B} \cdot \vec{u}_t$ must be used, giving

$$\oint \vec{B} \cdot \vec{u}_t \, d\ell = \mu_0 i_1 \quad (4-6)$$

Up to this point, we have only shown that Eq. (4-5) is a special case of Eq. (4-6) and also that the relation $2\pi r B = \mu_0 i_1$ is in turn a special case of

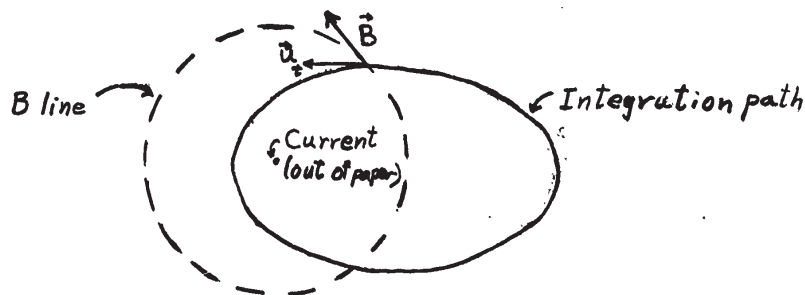


Figure 4-7.

Eq. (4-5). One further generalization has to do with the possibility of having a coil of N turns, like Figure 4-6(b). If the closed curve encircles all turns, the current encircled is Ni_1 . Thus, using the solid line in Figure 4-6(b), we have

$$\oint \vec{B} \cdot \vec{u}_t \, dl = \mu_0 Ni_1 \quad (4-7)$$

as an equation which reduces to Eq. (4-6) when $N = 1$.

In the absence of direct proof for the general case represented by Eq. (4-7), this equation will be taken as a basic postulate. By this statement we mean that Eq. (4-7) has not been proved experimentally for all possible coil shapes and sizes, but its consequences are consistent with all observed phenomena. Equation (4-7) is one form of Ampere's circuital law. Relative to Eqs. (4-6) and (4-7), a word of explanation is needed concerning the relationship between the direction of $\vec{u}_t$, which is along the curve of integration and the direction of i_1 . They are related in accordance with the right-hand screw rule; the reference direction of i_1 is the direction a right-hand screw will advance when rotated around the curve in the direction of $\vec{u}$. Figure 4-7 shows this relationship.

Ampere's circuital law permits the solution of another simple problem, involving the toroidal coil shown in Figure 4-8. From the symmetry of the

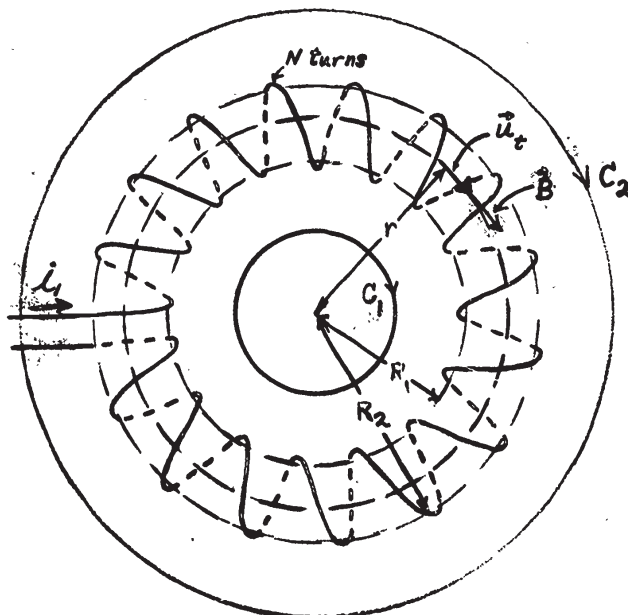


Figure 4-8.

4-10

- 1 figure, it is evident that, at every point within the coil, B will be tangent to
 2 a circle, as indicated. This circle, of radius r, encircles Ni_1 amperes, enter-
 ing the paper. Thus, the right-hand screw rule shows that u_t should be in the
 direction shown, and Eq. (4-7) becomes

$$2\pi r B = \mu_0 Ni_1$$

or

$$R_1 < r < R_2 \quad (4-8)$$

3

$$B = \frac{\mu_0 Ni}{2\pi r}$$

- 4 With the exception of the introduction of N, this is the same as Eq. (4-3). How-
 ever, Eq. (4-8) applies only for the region inside the coil. For points outside
 the coil, B is zero. This we can see by observing that any other circles, such
 as C_1 or C_2 , will encircle no current.

5 4-3. Force on a Moving Charge

- The phenomenon of force on a current carrying conductor can be extended to
 the situation where free charges (as distinct from charges within a conductor) are
 moving in a magnetic field. A common example is a television picture tube, in
 6 which control of the motion of a stream of electrons is partially provided by a
 magnetic field.

The current in a conductor can be viewed as due to the motion of a row of
 charges, symbolized by the dots in Figure 4-9. These charges are all of one sign,

7

8

9

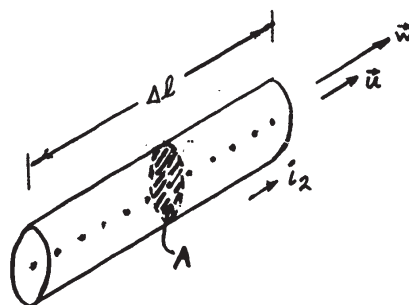


Figure 4-9.

1 and if the conductor is uncharged, there is an equal number of stationary
 charges of opposite sign, not shown in the figure. Let $\vec{w} = w \vec{u}$ be the average
 velocity of the moving charges, and let N be the number of them per unit
 length. Each moving charge has a charge q . In an increment of time Δt each
 2 charge will move a distance $w\Delta t$, and therefore in time Δt all charges in this
 length will pass a given cross section, such as A . Thus, the current is

$$i_2 = \frac{Nq w \Delta t}{\Delta t} = Nq w$$

3 If the conductor is in a magnetic field, we have seen that the force
 on it is

$$4 \quad \Delta \vec{F} = i_2 \Delta \ell (\vec{u} \times \vec{B})$$

Now we postulate that the force on the wire is actually due to force on the
 moving charges, and substitute for $i_2 \Delta \ell$ in the above, to give

$$5 \quad \begin{aligned} \Delta \vec{F} &= (Nq w \Delta \ell)(\vec{u} \times \vec{B}) \\ &= (Nq \Delta \ell)(\vec{w} \times \vec{B}) \end{aligned}$$

6 since $w\vec{u} = \vec{w}$. Finally, in length $\Delta \ell$ there are $N\Delta \ell$ charges, and so dividing
 by $N\Delta \ell$ will give

$$\vec{F} = q(\vec{w} \times \vec{B}) \quad (4-9)$$

7 as the force on one charge. The above postulate implies that this formula
 will give the force on a moving charge even when it is not in a conductor.
 Experiment on the deflection of the electron beam of an oscilloscope bears
 8 this out, and so Eq. (4-9) should be regarded as giving the force on a moving
 charge under all circumstances, whether it be a free electron or ion, a
 physically moving charged object, or the charges within a conductor or semi-
 conductor.

9 The force described here is known as the Lorentz force.

1 4-4. Motional Induced Voltage

Having established the idea of a magnetic flux density, and the Lorentz force, we are prepared to consider what happens when a conducting body moves in a magnetic field. Consider the experiment illustrated in Figure 4-10(a), in which a wire PQ is perpendicular to a pair of conducting rails along which it slides with velocity $\vec{w}$ in a direction parallel to the rails. Furthermore, there is a magnetic flux density B which we consider to be uniform over PQ and perpendicular to the plane of PQ and $\vec{w}$.

3 The conducting wire contains charges which are forced to have a velocity $\vec{w}$ because they must move with the wire. They therefore experience a Lorentz force which pushes them along the wire. Since the free charges are electrons, they move toward P, in the direction of $-(\vec{w} \times \vec{B})$. As a result, a charge separation takes place, as indicated by the polarity markings in Figure 4-10(a). By virtue of this charge separation, there is a potential difference v , which can be measured by a voltmeter. The charge separation takes place very quickly, and the motion of charge stops when the electrostatic force due to the separated charges is exactly equal to the Lorentz force, as indicated in Figure 4-10(b). Since the conductor is perpendicular to the plane of $\vec{w}$ and $\vec{B}$, the Lorentz force acts along the wire, and the work that it does in moving a unit of charge from P to Q is the force per unit charge times Δl . In the notation $\vec{w} = w \hat{u}_w$ and $\vec{B} = B \hat{u}_B$, since $\vec{w}$ and $\vec{B}$ are perpendicular to each other, on a positive charge this force, of amount Bw , is directed from P to Q. Thus, the work per unit positive charge is

$$e_{PQ} = Bw\Delta l \quad (4-10)$$

7 This quantity (e_{PQ}) is called the electromotive force (frequently abbreviated emf) acting from P to Q.* However, from the energy conservation postulate, the work done by the Lorentz force in moving a unit charge from P to Q must equal the work the electric field caused by the charge separation would do in moving it from Q back to P. But this is the potential difference v (or v_{QP}). Thus, for this situation,

$$v = v_{QP} = e_{PQ} \quad (4-11)$$

9 *This terminology is a misnomer, because electromotive "force" is a scalar quantity (energy) whereas force is a vector. This incorrect terminology is a heritage from the past which persists due to long usage.

1 Either of these is loosely termed induced voltage. In view of the identity
 2 of v_{QP} and e_{PQ} , one might naturally wonder why a distinction is made between
 3 them. As we shall see later, the reason is that when current flows in the wire,
 4 v_{QP} is no longer equal to e_{PQ} . They are the same only on open circuit.

5 In the above example, the moving conductor, B , and w are mutually per-
 6 pendicular. The result is an equation for e_{PQ} completely in the scalar quan-
 7 tities B , w , and Δl . It is not difficult to show that in the absence of
 8 mutual perpendicularity of B and w , the general formula is

$$e_{PQ} = \vec{u}_c \cdot (\vec{v} \times \vec{B}) \Delta l \quad (4-12)$$

9 where $\vec{u}_c$ is a unit vector from P to Q along Δl . This result will not be proved
 10 here.

11 In the case of a conductor which is extensive enough so that $\vec{B}$ cannot be
 12 considered uniform over its length, and may not necessarily be straight, Eq.
 13 (4-12) can be regarded as the emf induced in an incremented element of length
 14 Δl . An integration can then be performed with respect to a variable l measured
 15 along the conductor, giving

$$e_{PQ} = \int_P^Q \vec{u}_c \cdot (\vec{v} \times \vec{B}) dl \quad (4-13)$$

16 Reference to Figure 4-10(c) will help to clarify the meaning of this integral.
 17 Note that at each point on the curve $\vec{u}_c$ is tangent to the curve and directed
 18 toward Q .

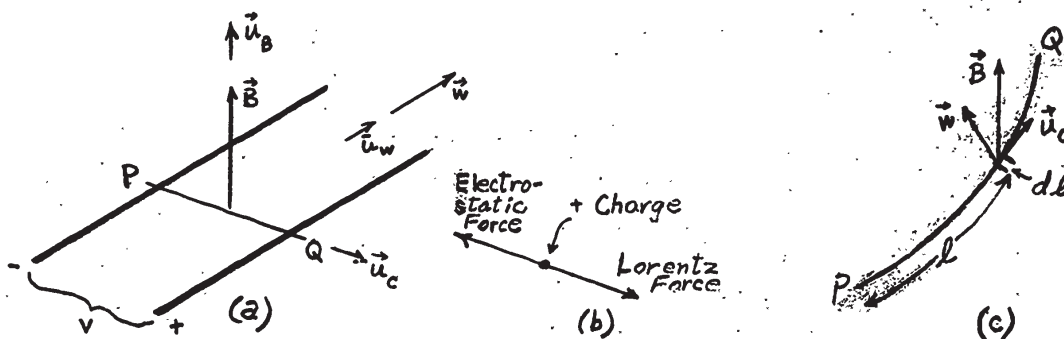


Figure 4-10.

4-14

A simple application is shown in Figure 4-11, where PQ is moving with velocity $\vec{w}$, and $\vec{B}$ is due to current i_1 in a long straight conductor. In the region of conductor PQ, $\vec{B}$ is into the paper, as indicated by crosses. It will be found that $\vec{u}_c \cdot (\vec{w} \times \vec{B})$ is negative, and since

$$B = \frac{\mu_0 i_1}{4\pi y}$$

we get

$$\begin{aligned} e_{PQ} &= - \frac{\mu_0 i_1 w}{4\pi} \int_b^{b+h} \frac{dy}{y} \\ &= - \frac{\mu_0 i_1 w}{4\pi} \ln \frac{(b+h)}{b} \end{aligned}$$

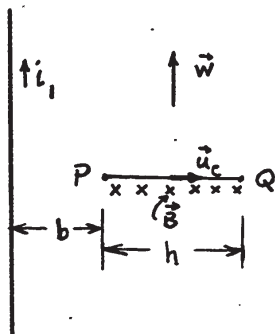


Figure 4-11.

4-5. Faraday Induced Voltage

Next we consider an experiment, illustrated in Figure 4-12(a), which involves the effect on one circuit of a changing current in another circuit. A voltmeter is connected to the open ends of a loop of wire, and the loop is near a long wire carrying a current i_1 which varies with time. The wire is in the plane of the loop.

Voltage indications on the meter are described by the following observations:

- (1) When i_1 is decreasing, the voltmeter will show a negative indication (indicating a potential difference opposite to the meter polarity marks).
- (2) When i_1 is increasing, a voltage of the indicated meter polarity will be observed.
- (3) The magnitude of the voltage is proportional to the magnitude of the rate of change of current.

- 1 (4) For a given rate of change of current, the magnitude of the voltage
varies inversely with distance from the wire, and at any given distance
is proportional to the loop area.*

2 Although the above observations involve current i_1 , a mathematical relation-
ship describing these phenomena can be obtained by using the magnetic flux den-
sity vector ($\vec{B}$), since B is related to i_1 by the Biot-Savart law, $B = \mu_0 i_1 / 4\pi r$.
If $\vec{u}$ is a unit vector directed out of the paper, from the center of the loop,
then $\vec{B} = B\vec{u}$, where $B = \mu_0 i_1 / 4\pi r$. The observed phenomena are accounted for by
3 the formula

$$v_{QP} = - \frac{d(AB)}{dt}$$

4 where B is the average value of the flux density within the loop, and A is
the loop area. The above observations (1), (2), and (3) are accounted for by
the negative sign and the fact that B is proportional to i_1 ; and observation
(4) is accounted for by the way B varies with r , and by the inclusion of the
5 factor A .

A further experiment using a coil of N turns of wire will lead to the
observation that the voltage is also proportional to N . Thus, for this case,
the above equation would include the factor N .

6 The quantity $\phi = AB$ is called the magnetic flux linking the loop. If the
loop had not been normal to the B vector, the induced voltage would be smaller,
and investigation would have shown that the component ($\vec{B} \cdot \vec{u}$) of $\vec{B}$ normal to the
loop should be used in obtaining ϕ . Thus, a general formula for flux is

$$\phi = A \vec{B} \cdot \vec{u} \quad (4-14)$$

7 where $\vec{u}$ is normal to the loop, and $\vec{B} \cdot \vec{u}$ is averaged over the loop surface. This
8 relationship between ϕ and B is, in fact, the origin of the designation of $\vec{B}$
as flux density.

9 *

The functional relationships described in (4) are accurately true only for a
loop of infinitesimal size.

In similarity with the case of motionally induced voltage, it is convenient to introduce $e_{PQ} = v_{QP}$ as the work done by the changing magnetic field in moving a unit positive charge from P to Q. Thus, using Eq. (4-14) for ϕ , and assuming a loop of N turns, the equation for the electromotive force is

$$e_{PQ} = -N \frac{d\phi}{dt} \quad (4-15)$$

Since, on open circuit, $v_{PQ} = -v_{QP} = -e_{PQ}$, the above equation can be written in terms of potential difference, as

$$v_{PQ} = N \frac{d\phi}{dt} \quad (4-16)$$

It is to be emphasized that Eq. (4-16) gives the terminal voltage only on open circuit.

A word of explanation is needed concerning the sign in this equation. As defined in Eq. (4-14), ϕ is a scalar quantity having a reference direction (like current). This involves a somewhat fictitious view of ϕ as something that "flows", but it is a useful idea. Obviously, the negative sign in Eq. (4-15) has no meaning unless reference directions for scalar quantities e_{PQ} and ϕ are properly defined. The reference for e_{PQ} is indicated by the order of subscripts (i.e., work done in moving positive charge from P to Q). The reference for ϕ is obtained from the order PQ by the right-hand screw rule, by stating that the unit vector $\vec{u}$ in Eq. (4-14) shall point away from area A in the direction a right-hand screw will move when rotated according to a progression from P to Q along the loop. This is illustrated in Figure 4-12(b). The reference direction of ϕ can be regarded as a direction from one side of the loop to the other, in the direction of $\vec{u}$.

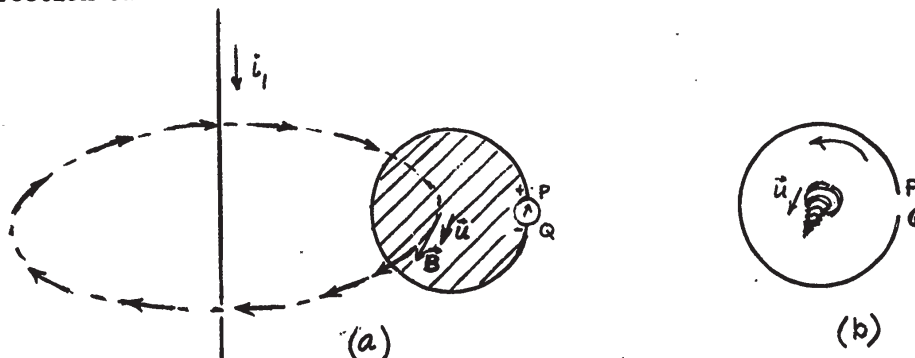


Figure 4-12.

In the above discussions, variation of B within the loop was not considered, having been tacitly accounted for by saying that B in the formulas is the average value. Where B varies over the area, an integral is used to obtain ϕ . This is done by observing that $\Delta\phi = \vec{B} \cdot \vec{u} \Delta a$ is an increment of flux through an area Δa , where $\vec{u}$ is normal to Δa . Thus, in general

$$\phi = \iint_{\text{over surface}} \vec{B} \cdot \vec{u} \, da \quad (4-17)$$

should be used for flux whenever B is not constant.

For example, for the rectangular loop in Figure 4-13, the area element

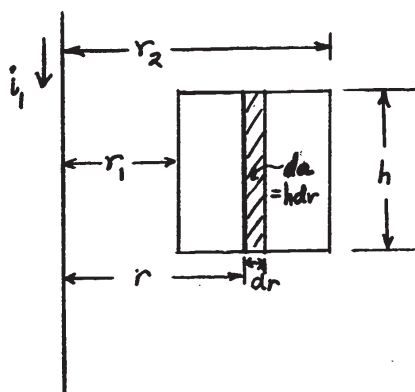


Figure 4-13.

is $h dr$, and $\vec{B} \cdot \vec{u} = \mu_0 i_1 / 4\pi r$, giving

$$\phi = \frac{\mu_0 i_1 h}{4\pi} \int_{r_1}^{r_2} \frac{dr}{r} = \frac{\mu_0 i_1 h}{4\pi} \ln \frac{r_2}{r_1}$$

The unit of ϕ in the MKSC system of units is the weber*

*In the electromagnetic system, ϕ is in maxwells (gauss $\times$ area in cm^2). One maxwell = 10^{-8} weber.

Equation (4-15) or its equivalent, Eq. (4-16), is an expression of the Faraday law of induced voltage. It is presented here for a simple geometrical arrangement, but it is postulated as a generally valid law, regardless of the source of ϕ or the cause of its variation with time. In this example, ϕ varies with time by virtue of the time variation of current i_1 . However, ϕ may change with time due to the motion of a fixed current, due to motion of a permanent magnet, or simply due to the motion of a piece of iron in the field.

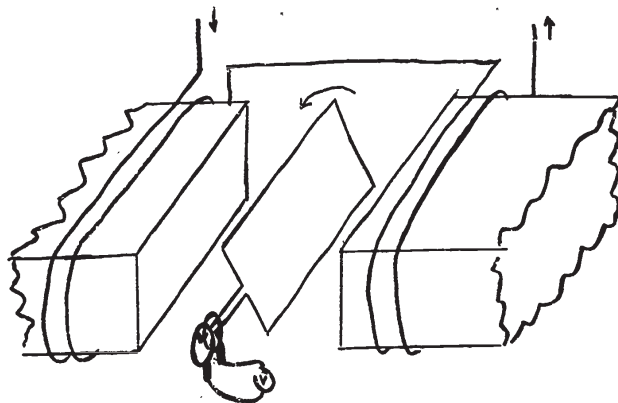
In some applications a loop moves in a changing magnetic field. In such a case care must be exercised to be sure that Eqs. (4-13) and (4-15) are used correctly. Carelessness in this respect once lead to an interesting historical mistake.*

4-6. Elementary Energy Relations

The previous two sections have shown how induced voltage (emf) can be derived from $\vec{B}$ or ϕ , or both, and this was related to potential difference when the device in question was on open circuit. That is, no current was allowed to flow, except for the slight momentary current flow during the short period of time while charge collects on the terminals.

Now consider Figure 4-14 in which the loop between P and Q is subject to

*It was once proposed that a commutatorless d-c generator could be made in the manner shown in the figure. It was proposed that the stationary coils should be supplied by alternating current, arranged to reverse once each half revolution of the rotating coil. The argument went that the voltage induced in each side of the rotating coil would always be in the same direction because as the direction of motion reversed at the top or bottom of its circular path the direction of flux would also reverse. However, when the effect of the time variation of flux is included, it is found that d-c is not obtained; and this is in agreement with experiment.



1 an induced voltage, either by virtue of conductor motion or time varying flux.

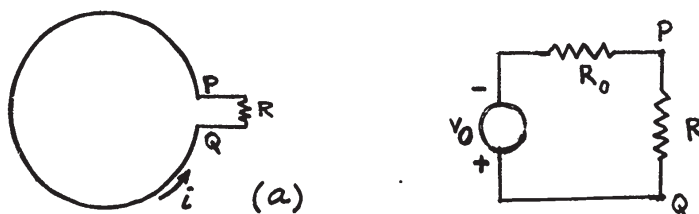


Figure 4-14.

We have seen how to determine e_{PQ} , and it has been recalled that this quantity is the work done per unit charge by the magnetic field in moving this charge from P to Q around the loop. In the present case, the terminals are connected to a resistor R and so a current will flow and the emf e_{PQ} will continually do work. Since e_{PQ} is work per unit charge, and i is units of charge per second, this work will be at the rate $e_{PQ}i$. Using the conservation of energy postulate, this energy rate will equal the sum of $i^2R + i^2R_0$, where R_0 is the resistance of the loop. Thus,

$$e_{PQ}i = i^2(R + R_0)$$

or

$$e_{PQ} = i(R + R_0)$$

But, $iR = v_{QP}$, the potential of Q with respect to P. Thus,

$$v_{QP} = e_{PQ} - iR_0 \quad (4-18)$$

Note that if $i = 0$, this reduces to the previously obtained equation $v_{QP} = e_{PQ}$.

Equation (4-18) shows the reason for making a distinction between potential difference and emf. Whereas the term "induced voltage" may be used to designate an open circuit potential difference, v_{QP} in Eq. (4-18) should not be called an induced voltage. An emf can be measured by a voltmeter on open circuit (neglecting voltmeter current); or by applying a short circuit to make $v_{QP} = 0$ so that $e_{PQ} = iR_0$, which determines e_{PQ} if i and R_0 are known.

- 1 Equation (4-18) is the general relationship between terminal potential difference and current for any device in which there is an electromagnetically induced voltage. If we define the quantity $v_0 = e_{PQ}$ (where v_0 is the open circuit value of v_{QP}) and use it as a voltage source in the circuit of Figure 4-14(b), we see that
- 2 Eq. (4-18) will be satisfied by this circuit. Thus, this is an equivalent circuit for the source device. It looks like a Thevenin equivalent, but is derived here from different concepts.

- 3 When an induced voltage causes a current to flow, as in Figure 4-14(a), a check of the current direction will show that it will always be in such a direction that the magnetic field caused by that current will act to oppose the action creating the emf. This principle is known as Lenz's law. If the emf is due to motion, as in Figure 4-10(a), the force on the conductor due to the current flowing
- 4 in the B field (the same field which causes the emf) will oppose the motion. Thus, due to this opposition, mechanical work must be done to maintain velocity w . In the case of an emf due to a changing flux, as in Figure 4-12(a), the current caused by this emf when the circuit is closed will create a flux opposing the flux change
- 5 which causes the emf.

Lenz's law is a consequence of conservation of energy, being a statement that there is an opposing action to any action which will cause a current to flow, such opposing action necessitating an expenditure of energy on the circuit.

6 4-7. Ferromagnetic Material

The next experiment to be considered is illustrated in Figure 4-15, where $\vec{B}$

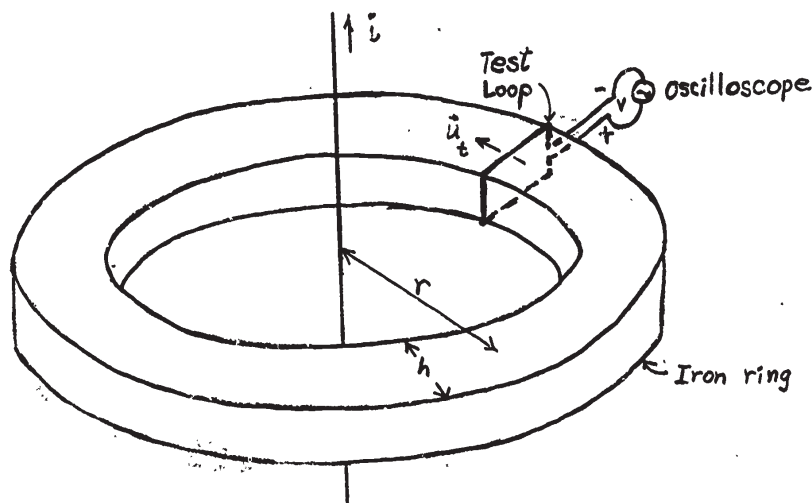


Figure 4-15.

is to be found within the circular iron ring encircling a long straight current-carrying conductor, as shown in the figure. The cross-sectional area of the ring is uniform, and dimension h is very small compared with r .* Obviously, the force on a test conductor cannot be measured within the iron, and so a different method of measuring B must be used.

The total flux through a cross section of the iron can be determined by placing a loop of conducting wire in the manner shown and connecting this to a voltage recording device such as an oscilloscope. Although such an arrangement cannot measure ϕ directly, it can be used to measure a change in ϕ . Thus, suppose i is increased in a sequence of jumps, in the manner illustrated graphically in Figure 4-16(a). With each increase in current, the flux will increase an amount $\Delta\phi$. Now, since

$$v = \frac{d\phi}{dt}$$

at any one of the jumps, say the second, we get

$$\Delta\phi_2 = \int_{t_1}^{t_1 + \Delta t} v \, dt \quad (4-19)$$

This integral can be evaluated from the oscilloscopic record of v as a function of time.** By making a succession of such measurements, the data points on the solid curve in Figure 4-16(b) are obtained.*** It is found that as the current increases by equal increments, successive increments of ϕ become smaller, causing the curve to flatten out. This phenomenon is called saturation.

* The reason for this requirement is that in view of Eq. (4-2) we should expect B to vary with radius. By restricting h to a small value, there cannot be much variation of r to all points in the ring, and thus B will be nearly constant over a cross section.

** This integral is normally evaluated by the use of a ballistic galvanometer, but in the interest of maintaining a direct attack on the immediate problem, we avoid this detail.

*** Readers with awareness of the phenomenon of hysteresis will realize that this curve will be obtained only if the iron is initially unmagnetized. This state can be attained by starting with a large value of i , reversing i to a slightly smaller value, and repeating this process until the current is zero.

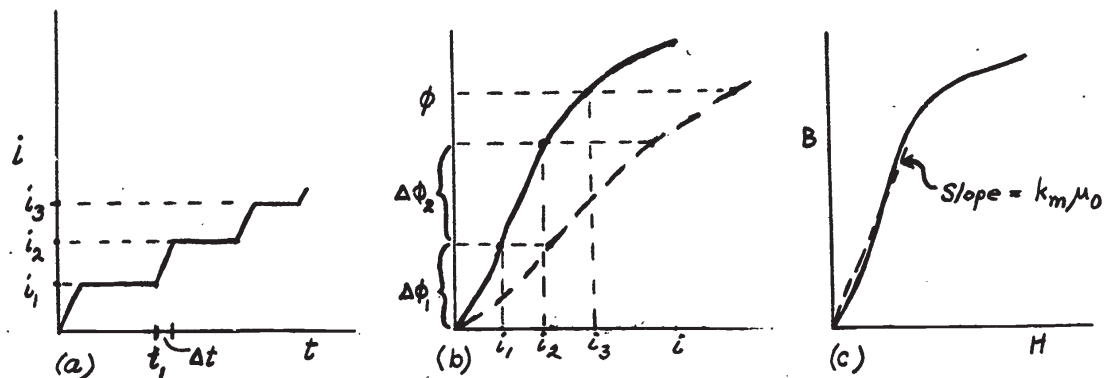


Figure 4-16.

The curve of Figure 4-16(b) is a property of a particular ring. In order to attain a curve which yields intrinsic properties of the magnetic material, independently of physical dimensions, a further experiment is needed, using a ring of different diameter, say twice as large. A new curve will then be obtained, like the one shown by the dashed line in Figure 4-16(b). The amount of current i required to yield a specific value of ϕ will be doubled. Thus, if we had plotted $i/2\pi r$ instead of i as the abscissa, the same curve would have been obtained in both cases. Thus, by using $i/2\pi r$ as the abscissa, the diameter (or circumference) parameter is eliminated. Furthermore, the cross-sectional area A can be eliminated by dividing ϕ by A , to give B .[Ⓢ] The result will be a similar curve with changed labels on the axes, as shown in Figure 4-16(c). A new variable (H) has been introduced such that

$$H = \frac{i}{2\pi r}, \quad \vec{H} = H\vec{u}_t \quad (4-20)$$

Since dimensions of the ring have been eliminated from consideration in Figure 4-16(c), the B - H curve is a property of the material, called a normal magnetization curve.

$\vec{H}$ is a fundamental magnetic quantity, called the field intensity vector. In the MKSC system of units, H is expressed in amperes per meter.

[Ⓢ]This neglects the slight variation of B over the cross section.

1 Let the formula

$$B = f_m(H) \quad (4-21)$$

2 describe the functional relationship between B and H , such as the graph in
Figure 4-16(c). In iron and in similar media, which are called ferromagnetic,
the relationship is non-linear. In air, the relationship can be determined
by comparing Eqs. (4-2) and (4-20); it is seen to be

$$B = \mu_0 H \quad (4-22)$$

This is a linear relationship showing that B and H are proportional.

4 There are other materials than air in which B is proportional to H .
For these, we write

$$B = k_m \mu_0 H \quad (4-23)$$

5 where k_m is a constant called the relative permeability. If $k_m < 1$, the mate-
rial is called diamagnetic; if $k_m > 1$, it is called paramagnetic. However,
the value of k_m for such materials is but little different from 1; for example,
6 one of the larger values is 1.003 for liquid oxygen.

The linear relationship in Eq. (4-23) is sometimes used, even for ferro-
magnetic materials, as a linear approximation for the general relationship in
Eq. (4-21), as indicated by the dashed straight line in Figure 4-16(c). This
7 is often done in preliminary design calculations, where the advantages of a
linear equation (as distinct from a graph) outweigh the inaccuracies involved.

Examples of normal magnetization curves for several ferromagnetic materials
are shown in Figure 4-17. Also included is the straight line $B = \mu_0 H$. It is
8 apparent that B can be increased very appreciably by the introduction of iron,
an important consideration in the construction of many items of electrical
apparatus.

In most cases (and all cases to be considered in this text) $\vec{H}$ is a vector
9 in the direction of $\vec{B}$, where the relationship between their magnitudes is
either given graphically or by Eqs. (4-22) or (4-23). The exception occurs in
certain materials which are non-isotropic (materials which exhibit different

4-24

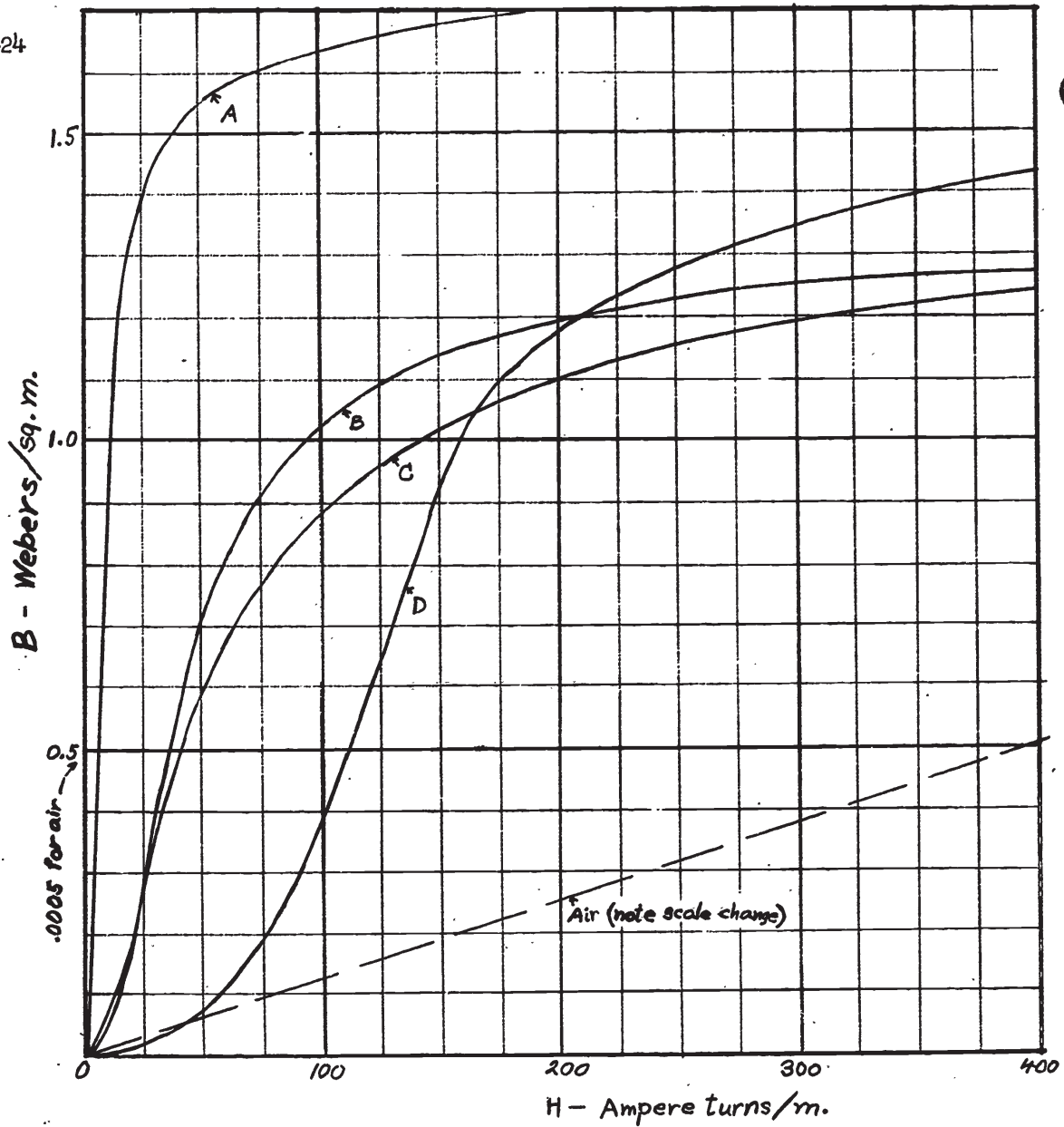


Figure 4-17. Normal magnetization curves (from U.S. Steel Handbook)

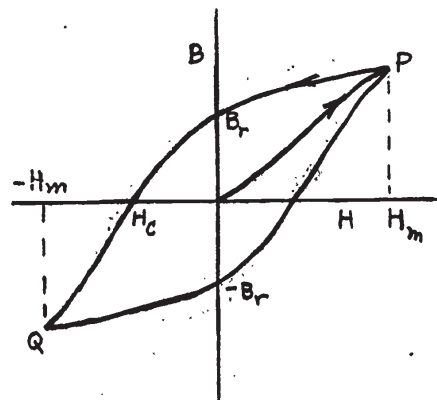


Figure 4-18.

Legend

- | | |
|--------------------------|------------------|
| A Grain oriented Trans- | } Silicon Steels |
| former "66" | |
| B Radio Transformer "58" | |
| C Transformer "52" | |
| D Pure iron (annealed) | |

1 properties in different directions), in which $\vec{B}$ and $\vec{H}$ do not necessarily have
the same direction.

2 4-8. Hysteresis

3 Again referring to Figure 4-15, assume the current has been increased,
bringing the magnetic state to point P in Figure 4-18 and then let the current
be decreased in steps. With each jump downward, $\Delta\phi$ can again be obtained from
the integral of voltage. It will be found that demagnetization will take place
4 along a curve higher than the original one, as indicated by the arrow. If the
current is then carried to negative values, as far as $-H_m$, and then back to
 H_m , a closed curve called a hysteresis loop will be obtained.

Ordinate B_r , where the curve crosses the vertical axis represents the
4 flux density remaining in the iron when the current has been reduced to zero,
is called the remnant flux density, and abscissa H_c , the value of H required to
reduce B to zero is called the coercive force. The phenomenon of hysteresis
described here further complicates computations involving magnetism in iron.
5 Not only is the material non-linear (B not proportional to H) but the relation-
ship between B and H is multivalued. For a given H, the value of B depends on
the previous magnetic history of the iron. In many cases approximate calcula-
tions are made using the normal magnetization curve. However, there are ap-
6 plications in which the hysteresis phenomenon is of paramount importance. One
of these is in devices involving "magnetic memory," such as tape recorders and
memory devices used in computers. The memory capability arises from the above-
mentioned fact that magnetic state depends on what has happened in the past.
7 For example, by measuring the flux at zero current (giving either $+B_r$ or $-B_r$)
it is possible to know whether the magnetization was previously at P or Q.

8 4-9. Magnetic Circuits

A magnetic circuit is an arrangement of ferromagnetic materials forming
closed paths, sometimes including short gaps of air or of other non-magnetic
materials, so as to form an easy path for magnetic flux. The source of the flux
consists of one or more current-carrying coils of wire surrounding some portion
9 of the path.

Figure 4-19 shows some examples of magnetic circuits, taken from typical
electromagnetic devices. The iron ring shown at (a) might be the "core" of a

- 1 transformer on which there are two or more windings. With a single winding, the device is an inductor. To determine the amount of flux in the iron due to a given current in one of the windings would constitute a magnetic circuit problem. As another example, consider Fig. 4-19(b) which shows the magnetic circuit of a d-c generator or motor. Suppose it is a generator, to operate at a specified speed and voltage. This will determine the required value of B in the airgap between the rotor and the poles, by way Eq. (4-10), and thus a certain airgap flux is known to be required. Final solution of this problem would then determine the number of turns and current on the "field" windings required to produce this flux. This is also a magnetic circuit problem. Our final example, Figure 4-19(c), shows an electromagnetic relay. The force on the moving armature is a function of the airgap flux, and thus a certain flux will be required to cause the relay to close. If this required flux is known, a magnetic circuit problem is then solved to determine the specifications of the coil (current and number of turns.)

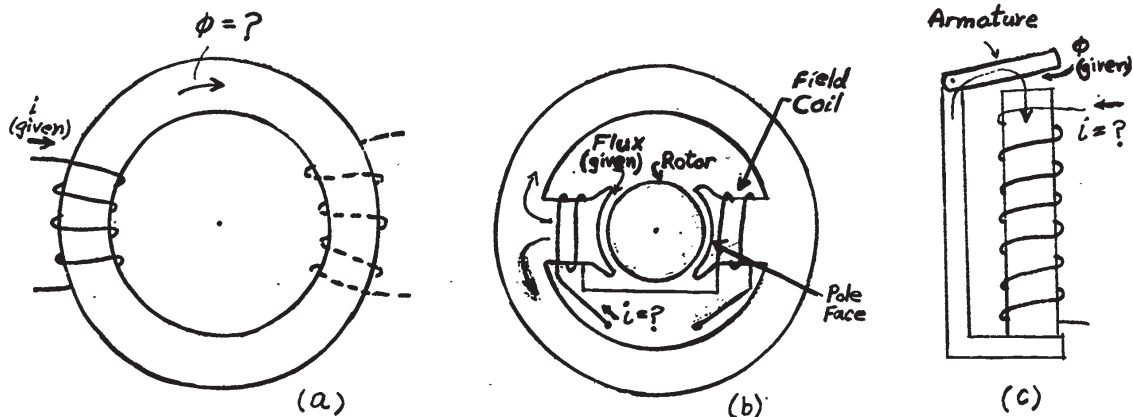


Figure 4-19.

- 8 We shall now consider how magnetic circuit problems like the above can be
 9 solved. As a start, we return to a consideration of a uniform iron ring, like
 Figure 4-15, but with a variety of coil arrangements as in the examples in Figure
 4-20. The case shown at (2) is the simplest. It is similar to Figure 4-8, except

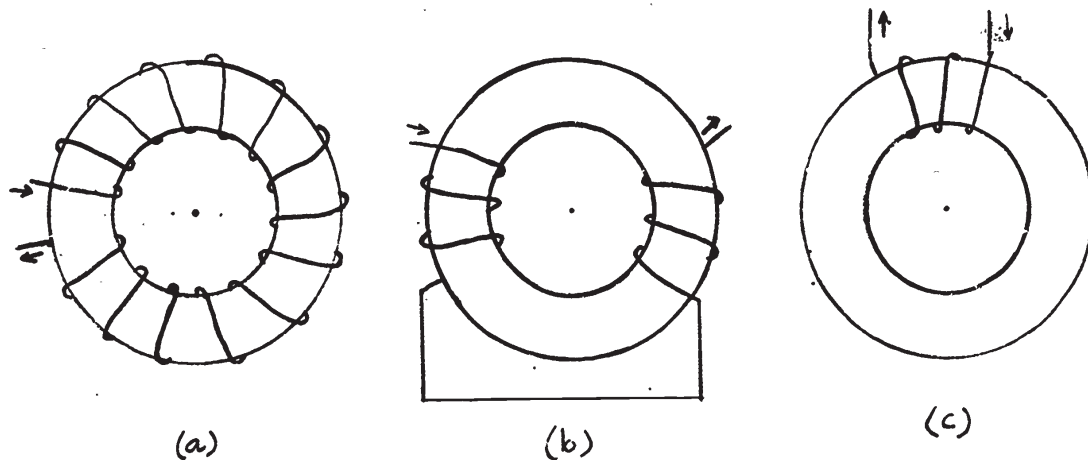


Figure 4-20.

that now the coil is wound on an iron core. It can be shown experimentally that if the product Ni in Figure 4-20 is the same as i in Figure 4-15, and if the iron ring dimensions are the same, then the flux in the iron will be the same. Now suppose the product Ni is also the same in the examples at (b) and (c) of Figure 4-20. It will be found that the flux in the iron is approximately the same as before, although by moving a test loop (like the test loop of Figure 4-15) to various points around the ring it will be found that there is a slight variation of flux, and that the flux is slightly larger near the windings.

This empirical result is significant, because it implies that the flux in a ring of magnetic material of given dimensions is approximately dependent only on the total current linking the ring (current in the wire times the number of turns, in the case of a coil). It is concluded that B and therefore H are nearly the same in each case, and that Eq. (4-20) approximately applies if we replace i by Ni , giving

$$2\pi rH = Ni \quad (\text{approximately}) \quad (4-24)$$

- 1 The value of H at a given point (fixed r) depends on the product Ni . The same H is obtained for different values of N and i , so long as the product is the same. Ni is called magnetomotive force,* abbreviated mmf, and is given the symbol U . Its unit is the ampere. (Its unit is sometimes taken to be an ampere-turn, but a turn has no dimensions.)

- 2 Let us tentatively leave this approximate relationship and consider the further experiment shown in Figure 4-21. In this case the ring shown at (a) includes an airgap. It is assumed that the airgap length l_a is small compared with the linear dimensions d and h of the cross section. It is also assumed that there is some means for closing the airgap temporarily, say by bending the ring slightly, as shown in Figure 4-21(b). Using a test loop like the one shown in Figure 4-15, two magnetization curves can be determined, one for the ring without the airgap, and the other with the airgap as shown in Figure 4-21(c).

- 4 Consider a specific value of flux as indicated by one of the horizontal lines in Figure 4-21(c). When there is no airgap, the required current is determined by U_1 , a specific value of Ni obtained from the magnetization curve. When the airgap is included, it is found that the mmf must be increased by an amount U_a in order to maintain the same flux in the iron.

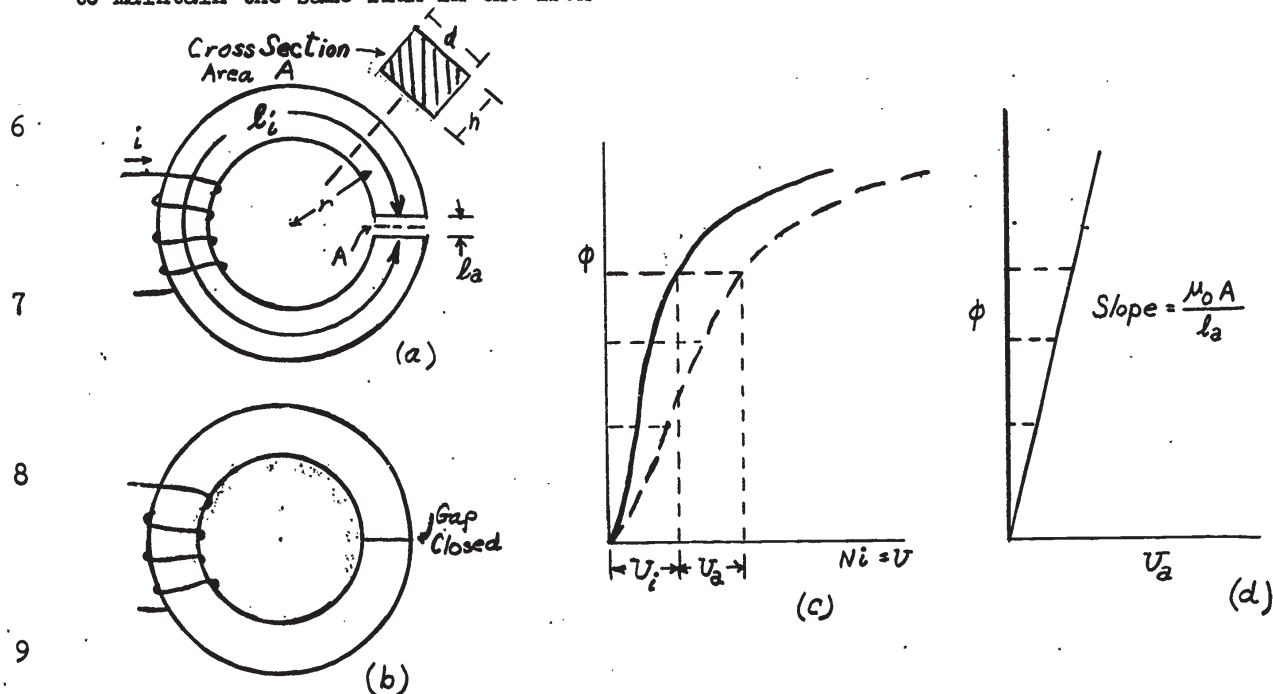


Figure 4-21.

*This is another example of incorrect terminology which persists due to long usage. Magnetomotive "force" is not a force; it is a scalar, not a vector.

1 Recalling that it is an approximation to assume that flux is the same every-
 2 where in the ring, it would be meaningless to retain the slight variation of H
 3 with r implied by Eq. (4-24). Instead, we will let l_1 be the average circumfer-
 4 ence of the entire ring when the gap is closed; this is the same as the average
 5 length of the iron when there is a short gap. Thus, for the closed-gap situation,
 6 Eq. (4-24) becomes

$$H_1 l_1 = U_1 \quad (4-25)$$

7 where H_1 signifies H in the iron. Since $B = \phi/A$ is the same for both cases,
 8 and since B and H_1 are related by a magnetization curve which is a property of
 9 the iron, H_1 is the same in both cases. If the quantity U_a is picked off this
 10 curve, for various values of ϕ , and plotted as a function of ϕ , as in Figure
 11 4-21(d), a straight line will be obtained. The slope of such a line can be ob-
 12 tained from experimental measurements, and will be found to be very close
 13 numerically to $\mu_0 A/l_a$. Thus, the equation of this line is

$$\phi = \frac{\mu_0 A}{l_a} U_a \quad (4-26)$$

14 Let us now imagine a flat surface of area A in the airgap, as indicated in
 15 Figure 4-21(a), and assume that the flux through this area is ϕ , the same as the
 16 flux in the iron. This is clearly an assumption at this point. Dividing both
 17 sides of Eq. (4-26) by A , gives B_a on the left, and so we get

$$B_a = \mu_0 \left(\frac{U_a}{l_a} \right)$$

18 But this is of the form $B = \mu_0 H$ (the relationship between B and H in air) and so,
 19 if the assumption about flux being the same in the iron and the airgap is cor-
 20 rect, then

$$U_a = H_a l_a \quad (4-27)$$

21 Now, since

$$U_1 + U_a = NI$$

- 1 from Figure 4-21(c), it follows that

$$H_1 l_1 + H_a l_a = Ni \quad (4-28)$$

- 2 This is a tentative result, based on an assumption that ϕ is the same on both sides of the gap interface between iron and air.

- 3 We shall now show that there is a certain similarity between Eq. (4-28) and Ampere's circuital law as given by Eq. (4-7). Referring to Eq. (4-7), if both sides are divided by μ_0 , we get

$$\oint \vec{H} \cdot \vec{u}_t \, d\ell = Ni \quad (4-29)$$

- 4 According to the original statement of Ampere's circuital law, the integration path for the above equation is wholly in air. Nevertheless, we shall tentatively apply it to the closed path in the iron ring and airgap of Figure 4-21(a). We find that in iron $\vec{H} \cdot \vec{u}_t = H_1$, and in the airgap $\vec{H} \cdot \vec{u}_t = H_a$ (H_1 and H_a are both constant) and so the integral on the left of Eq. (4-29) reduces to the left-hand side of Eq. (4-28).

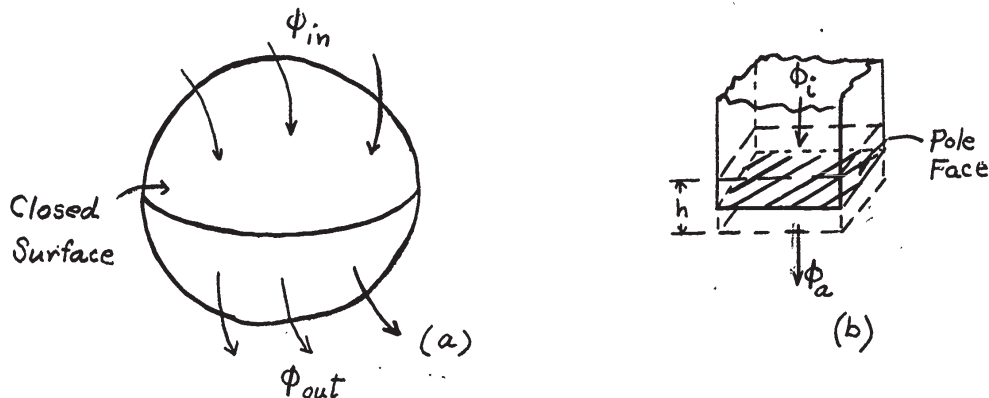
- 5 On the clue that Eq. (4-29) applies to a special case having part of the path in iron, we shall now introduce a postulate to the effect that Eq. (4-29) applies to all possible integration paths in all situations. This is the general Ampere's circuital law.

- 6 The condition $\phi_a = \phi_i$ used in arriving at the above must also be stated as a general postulate. This is done most easily in terms of the flux entering and leaving a closed surface. Specifically, referring to any hypothetical closed surface, as in Figure 4-22(a), it is postulated that the flux entering the volume equals the flux leaving. This postulate is known as Gauss' law for magnetic fields. To show that this postulate is consistent with $\phi_a = \phi_i$, we refer to Figure 4-22(b), in which a "pill box" shaped volume is used, with a side surface of infinitesimal height, so that any flux leaving this surface will be negligible (if B should have a component normal to this surface). One face of the surface is in the iron, and the other face is in the airgap. Thus, ϕ_i is the flux entering, and ϕ_a is the flux leaving. Gauss' law states they are equal, showing that the original assumption $\phi_a = \phi_i$ is consistent with Gauss' law.

1

4-31

2



3

Figure 4-22.

4

These two postulates are fundamental to the theory of magnetism and have far-reaching consequences which can be checked by experiment. All evidence indicates they are true. We may therefore regard them as fundamental laws, although in reality they remain postulates because their validity is based on evidence now available.*

5

Before proceeding to a description of methods of solving magnetic circuits, some interesting consequences of the above two laws can be shown, with reference to Figure 4-23(a). A closed integration path a-b-c is shown. Obviously, it does not link a current, and so for this path

6

$$\oint \vec{H} \cdot \vec{u}_t \, dl = 0$$

7

But that part of the integral due to the path in iron and airgap is $2H_i l_i + H_a l_a$, where the lengths l_i and l_a are shown in the figure. Thus,

8

$$\int_c^a \vec{H} \cdot \vec{u}_t \, dl = -(2H_i l_i + H_a l_a)$$

(air)

9

The quantity on the right is not zero, and so it is concluded that H along the air path c-a cannot be identically zero. Since $B = \mu_0 H$, B is also not

*The "law" of conservation of energy, which was really a postulate, is a good example of the tentativeness of a postulate. Until the discovery of nuclear reactions, this postulate satisfied all available evidence. However, it is known that it is not universally valid, and so can be used only in those situations in which there is no mass-energy interchange. During the 1940s there was a flurry of excitement in scientific circles concerning purported experimental evidence that "magnetic charge" had been discovered. If free magnetic charges exist, then Gauss' law for magnetic fields, as stated here, will not be true. The evidence for free magnetic charge was not conclusive, and so we continue to rely on Gauss' law.

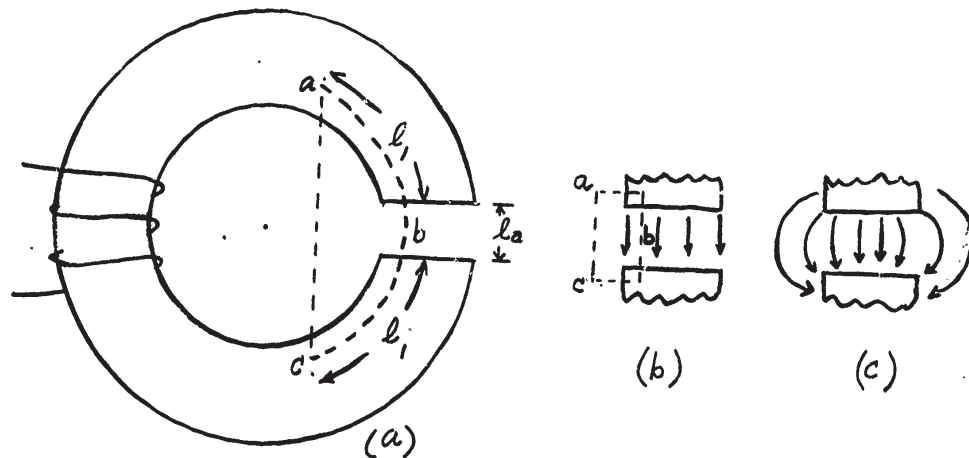


Figure 4-23.

identically zero.* We therefore conclude that there must always be some flux in the air surrounding a magnetic circuit. There is no such thing as a magnetic insulator, and in this sense a magnetic circuit differs from an electric circuit.

Let us pursue the same line of reasoning to investigate the airgap flux. If it is assumed that the field lines go straight across the gap as in Figure 4-23(b), an integration path a-b-c can again be used, showing that B cannot be zero everywhere along c-a, outside the gap. Accordingly, it is necessary for there to be fringing of the gap flux, as shown in Figure 4-23(c), by an amount sufficient to satisfy Ampere's circuital law.

The flux lines surrounding the ring will be somewhat as shown in Figure 4-24(a). Gauss' law gives additional information, if applied to the volume element shown at the top of this figure. If ϕ is the flux in the iron entering the volume at the left, an amount $\Delta\phi$ will leave the iron through the sides. Thus, by Gauss' law, the flux leaving the right, in the iron, is $\phi - \Delta\phi$. Thus, the flux in the iron gets progressively weaker with increasing distance from the coil. In many cases, however, the variation is negligible.

The difference between the flux through the coil, ϕ_t , and the flux ϕ_g at the

*We say "not identically zero" meaning not zero at all points. We cannot conclude, however, that there are no points where B and H are zero.

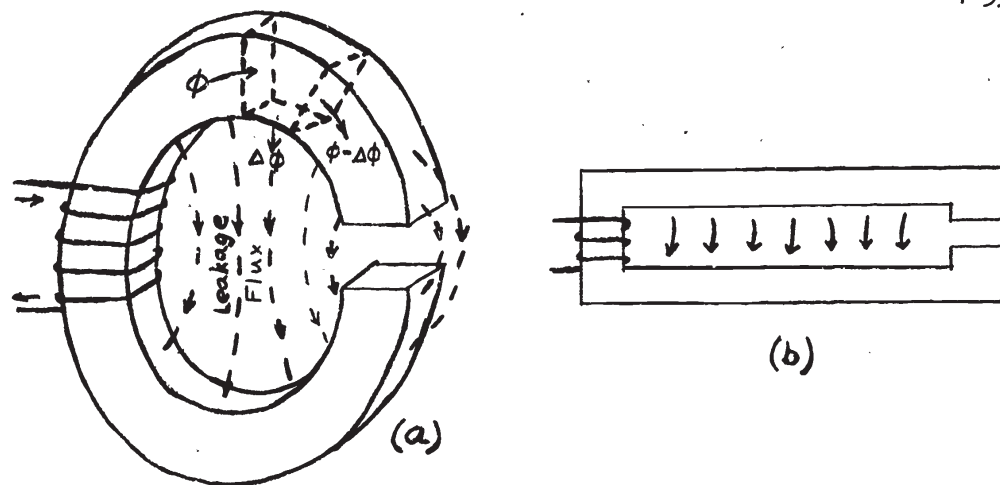


Figure 4-24.

gap, namely

$$\phi_l = \phi_t = \phi_g$$

is called the leakage flux. An arrangement like Figure 4-24(a) can be treated as a magnetic circuit only if ϕ_l is small compared with ϕ_g . Whether or not this is true depends on the permeability of the iron, and also on the geometrical shape. For example, an arrangement like Figure 4-24(b) can have a very large leakage flux.*

4-10. Solution of Magnetic Circuits

Reference should again be made to Figure 4-19, as illustrations of typical magnetic circuit problems. Particular note should be made of the fact that airgaps are often essential, as in the case of a motor or generator where mechanical clearance is necessary, or in the relay where opening and closing motion is possible only if there is an airgap. Also, it is to be noted that in the generator

* It can be shown that for a reasonably dimensioned circuit (not distorted like Figure 4-24(b)) without airgap the ratio of core flux to leakage flux is approximately

$$\frac{\phi_c}{\phi_l} = k_m \left(\frac{l_c}{l_i} \right)$$

where l_i and l_c are respectively the lengths of the iron path and the coil length, and k_m is the relative permeability of the iron. When there is an airgap, this ratio is smaller in proportion to the reduction in ϕ_c due to the airgap.

there are two magnetic circuits which share a common branch through the center.

The previous section established the principles upon which solution of magnetic circuits is based, including the notion that such solutions are always approximate. We reiterate these principles here, in terms of the example in

Figure 4-25.

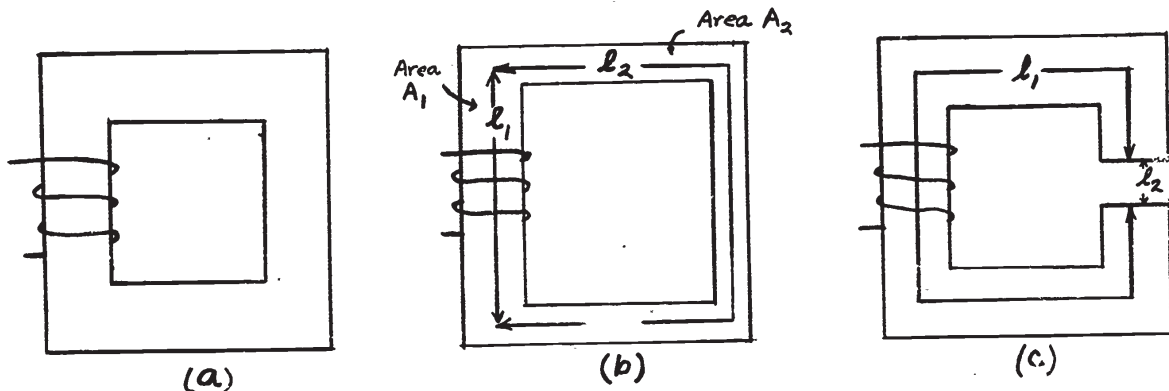


Figure 4-25.

- (1) Branches are identified such that in each branch the cross-sectional area is the same, the material is the same, and the flux is the same. Thus, in example (b), two branches are identified and labeled according to their lengths; l_1 , l_2 .
- (2) It is assumed that in an airgap the flux density is the same as in the iron, and that B is uniform across the airgap (i.e., fringing is neglected).
- (3) For each circuit the algebraic sum of the magnetomotive forces across individual branches is equal to the mmf supplied by coils linking the circuit. The magnetomotive force across a branch is the product of H times the length of the branch. A magnetomotive force term is positive in this summation if the reference direction of ϕ is related to the reference direction of the coil current by the right-hand screw rule. (This rule is a statement of Ampere's circuital law.)
- (4) At any junction of branches, the algebraic sum of fluxes entering equals the algebraic sum of the fluxes leaving. (This rule is a statement of Gauss' law for magnetic fields.)

- 1 (5) In any given branch, H is related to $B = \phi/A$ by the mag-
 2 netization curve of magnetic material, or in an airgap by

$$H = \frac{B}{\mu_0}$$

(4-30)

- 3 It is noted that rules (3) and (4) are quite similar to the Kirchhoff volt-
 4 age and current laws.

- For examples of solutions, we refer to the three simpler cases shown in
 Figure 4-25. At (a) there is only one branch, and the circuit has uniform cross
 section. Therefore, ϕ is everywhere the same, and since H and B are related by a
 magnetization curve, H is also uniform, and rule (3) gives

$$H\ell = Ni$$

- 5 We assume a magnetization curve is available, and from $H = Ni/\ell$ the correspond-
 ing B is obtained graphically. From this, $\phi = BA$ is known. Thus, if Ni is
 given, ϕ can be found. Also, if ϕ is given, B is known, H is determined from
 the curve, and finally Ni can be found.

- 6 The example of Fig. 4-25(b) is more complicated, because there are two
 branches, and the corresponding equation is

$$H_1\ell_1 + H_2\ell_2 = Ni \quad (4-31)$$

- 7 Although ϕ is the same in both branches, H_1 and H_2 are different because
 $B_1 = \phi/A_1$ and $B_2 = \phi/A_2$ are different.

- 8 Suppose we are to obtain Ni if a required value of ϕ has been specified.
 H_1 and H_2 are obtained from the same B - H curve (assuming both branches are of
 the same magnetic material) as indicated in Figure 4-26(a). B_2 is greater than
 B_1 because A_1 is greater than A_2 . Corresponding values of H_1 and H_2 are read
 off the curve, and substituted in Eq. (4-31) to give Ni as the required answer.

- 9 Now consider the more interesting case, where Ni is given and we are
 asked to find ϕ . The key to the solution of this problem is in the fact that
 ϕ must be the same in both branches. Accordingly, as in Figure 4-26(b), a

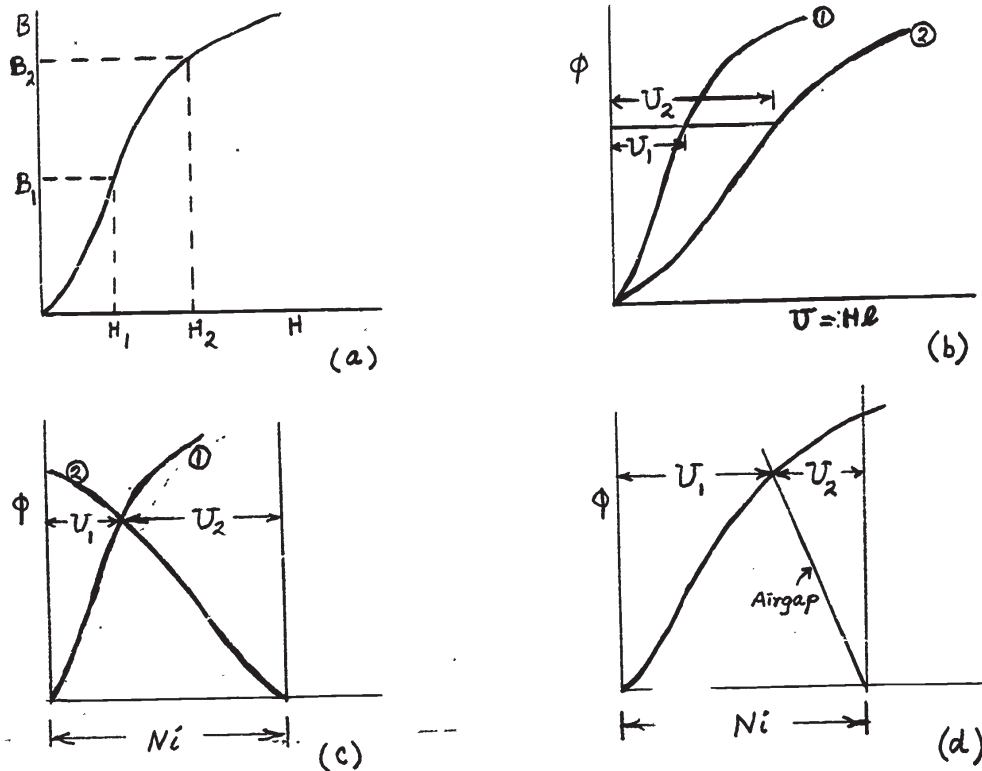


Figure 4-26.

magnetization curve for each branch is constructed as a curve of ϕ vs. magnetomotive force (Hl), and the value of ϕ is found such that the two abscissas add up to the given Ni as required by Eq. (4-31). That is, the horizontal line must be moved up and down until $U_1 + U_2 = Ni$. Note that these two curves are for this specific circuit, since curve (1) was obtained from the universal B-H curve by multiplying B by A_1 and H by l_1 , and curve (2) was obtained by using A_2 and l_2 in a similar way. This method involves a graphical procedure, incorporating a series of trials. The trial aspect can be eliminated, however, by using the scheme shown in Figure 4-26(c). One of the curves is reversed, and positioned so that the two origins are spaced an amount Ni . Then, $U_1 + U_2 = Ni$ is automatically satisfied at the point of intersection, which then yields the required value of ϕ .

The above method is restricted to circuits of not more than two branches. In many cases, the second branch is an airgap, as in Figure 4-25(c). This is merely a simple example of the previous case, in which ϕ vs. U for the airgap

1 is a linear relationship

$$U_2 = H_2 \ell_2 = \left(\frac{\mu_0 \ell_2}{A} \right) \phi$$

2 Thus, series circuits of uniform cross section, with an airgap, are easily solved graphically, as in Figure 4-26(d).

3 Many arrangements with two parallel circuits, like the generator of Figure 4-19(b), can be solved as a single series circuit, because of symmetry. Thus, the two circuits of Figure 4-27 are equivalent as regards the relationship between ϕ and Ni . We can see this by observing that $\phi = 2\phi_1$, in Figure 4-27(a), so that in a side branch $B = \phi/2A$. By increasing the area to $2A$ in Fig. 4-27(b), B , and therefore H , remain the same.

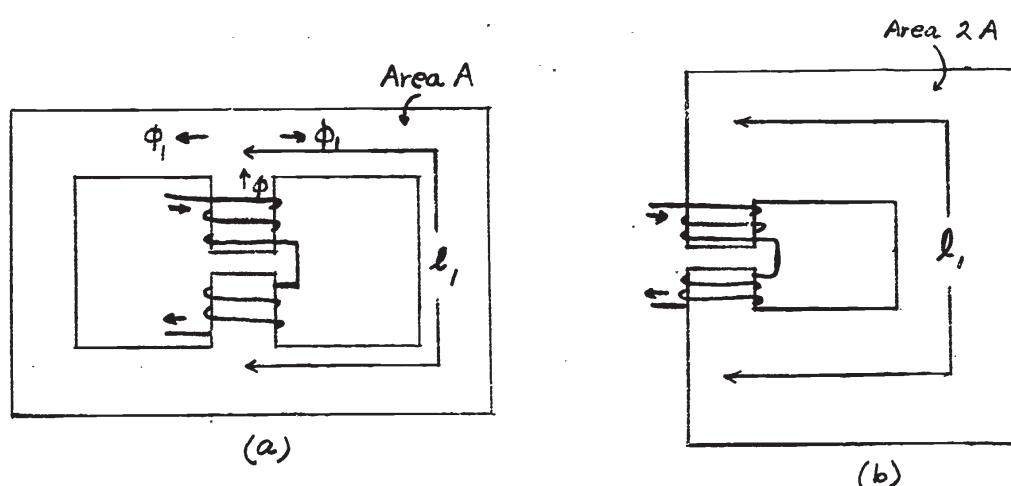


Figure 4-27.

7 This kind of simplification does not apply, however, to Figure 4-28, because the two sides are not symmetrical. The only way this circuit can be solved is to assume that ϕ_b is known. Then, H_5 can be obtained from properties of the air-gap, and magnetization curves yield H_3 and H_4 . Thus, in the equation

$$H_2 \ell_2 - (H_3 \ell_3 + H_5 \ell_5 + H_4 \ell_4) = 0 \quad (4-32)$$

9 obtained by summing $H\ell$ terms around the outside in accordance with rule (3), the quantity in parenthesis is known. From this H_2 can be determined, and from a magnetization curve B_2 and ϕ_c become known. Then, from rule (4) we have

4-38

$$\phi_a = \phi_b + \phi_c$$

(4-33)

and ϕ_a determines H_1 . Finally, the equation

$$H_1 l_1 + H_2 l_2 = Ni$$

for the left-hand loop yields the value of Ni . If the problem is stated with the coil mmf as the given quantity, the only way to solve for ϕ_b is to make a series of assumptions of ϕ_b , then to solve for corresponding values of Ni as indicated above. The results can be plotted as a curve of ϕ_b vs. Ni , from which the value of ϕ_b can be read, corresponding to the given Ni .

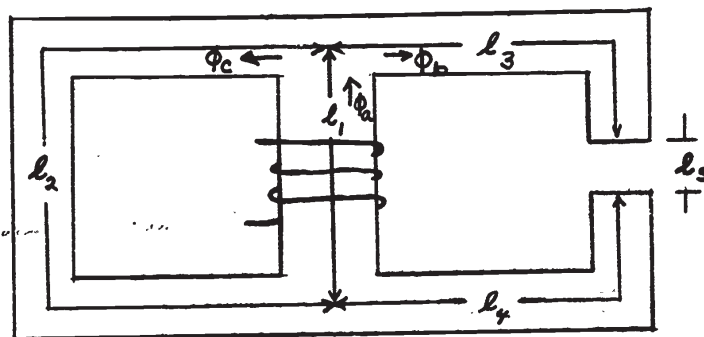


Figure 4-28.

4-11. Self Inductance

In the previous section we have seen how the relationship between flux through a coil and the mmf of the coil can be determined. In general, this relationship is non-linear but when airgaps are present the relationship may be approximately linear for a considerable range of the variables ϕ and Ni . Also, when no iron (or other ferromagnetic material) is present, the relationship is exactly linear, for all practical purposes. For the present, we shall consider that the relationship is linear, and write

$$\phi = P Ni$$

(4-34)

1 where P is a proportionality factor called the permeance of the magnetic circuit.

Recall now that the emf induced in a coil of N turns linked by a flux ϕ , using the reference directions shown in Figure 4-29, is

$$2 \quad e_{ab} = - \frac{d(N\phi)}{dt} \quad (4-35)$$

This equation has been referred to here to emphasize the importance of the quantity $N\phi$, called the flux linkage of the coil. Using Eq. (4-34) for ϕ , we get

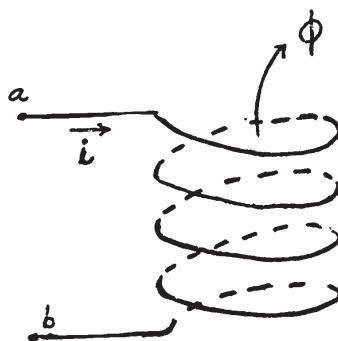
$$3 \quad N\phi = P N^2 i$$

4 and define a new quantity

$$L = \frac{N\phi}{i} \quad (4-36)$$

5 which, in view of the above expression for $N\phi$, can also be written

$$L = P N^2 \quad (4-37)$$



9 Figure 4-29.

L is a property of the coil, called the self inductance. When current is in amperes and emf is in volts, L is in henrys. The adjective "self" implies that the flux in question is due to current in the coil itself, as distinct from any flux due to external currents or magnets.

As an important observation, note that for a fixed magnetic circuit, Eq. (4-37) shows that inductance is proportional to the square of the number of turns.

Returning now to Eq. (4-35), we may use Li for $N\phi$, to give

$$e_{ab} = - \frac{d(Li)}{dt} \quad (4-38)$$

as an expression for the emf induced in a coil by its own current.

Of course, Figure 4-29 does not show the magnetic circuit, which can have any degree of complexity; or the coil may be isolated in air. If there is a moving part in the magnetic circuit, or if the dimensions of the coil change with time (i.e., by squeezing together or stretching out the turns as in a coil spring which is being pulled and released) P and therefore L can vary with time. For this reason, L is included under the derivative symbol in Eq. (4-38). In many cases, however, L is constant, and so in that case

$$e_{ab} = - L \frac{di}{dt} \quad (4-39)$$

Now we consider the terminal voltage v_{ab} in relation to the current, including the effect of coil resistance R_c . If R_c is negligible, recalling that $v_{ab} = -e_{ab}$, we have

$$v_{ab} = \frac{d(Li)}{dt} \quad \text{or} \quad v_{ab} = L \frac{di}{dt} \quad (4-40)$$

The phenomenon of voltage induced in a coil having resistance was treated in connection with Figure 4-14(a). Equation (4-18) applies, except that in that equation (for v_{qp}), current flows from P to Q through the coil. Now we want an equation for v_{ab} , with current flowing through the coil from a to b . Hence, the sign on the i term will change, giving

$$v_{ab} = iR_c + e_{ba} = iR_c - e_{ab}$$

1 or, inserting Eq. (4-37),

$$v_{ab} = iR_c + \frac{d(Li)}{dt} \quad (4-41)$$

2 The sign on iR_c can be confirmed by imagining the special case where i is constant. Then Eq. (4-40) becomes $v_{ab} = iR_c$, which is consistent with Ohm's law for the reference directions shown.

3 When a coil is connected to a source through an external resistor, as in Figure 4-30(a), we see that $v_{a'b} = iR_i + v_{ab}$, and so

$$v_{a'b} = i(R_i + R_c) + \frac{d(Li)}{dt} \quad (4-42)$$

4 Thus, the resistance of a coil can be treated like an external resistance, and therefore it is customary to use the equivalent diagram shown in Figure 4-30(b).

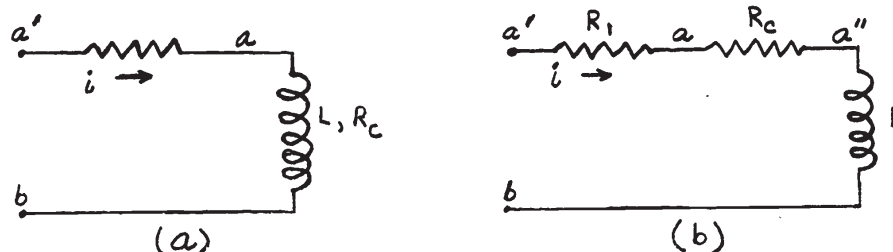


Figure 4-30.

7 The coil is then regarded as being ideal, having zero resistance. The fictitious voltage

$$8 \quad v_{a''b} = \frac{d(Li)}{dt}$$

9 is not accessible to laboratory measurement, but in circuit analysis may be treated as if it were measurable.

Concerning the computation of inductance, we refer to Eq. (4-36) which shows that if the flux due to a given current can be found by solving the

4-42

- 1 magnetic circuit, that equation will give the inductance. In general, inductance is a function of number of turns, coil dimensions, and sizes and permeabilities of the magnetic circuit components (if there is magnetic circuit). Formulas for coils of various shapes and sizes, in air, will be found in handbooks. Two ex-
- 2 amples are given below.

For a single layer air coil of length ℓ_c , diameter d , and N turns, as shown in Figure 4-31(a), the inductance is

$$3. \quad L = K_s d N^2 \quad (\text{henrys}) \quad (4-43)$$

where K_s is the function of the diameter to length ratio given graphically in Figure 4-31(b). Dimensions ℓ_c and d are in meters. A single circular loop of

- 4 wire of diameter d , and wire diameter w , has an inductance given approximately by

$$L = 2\pi d \left(\ln \frac{4d}{w} - \frac{7}{4} \right) \times 10^{-7} \quad (\text{henrys}) \quad (4-44)$$

5

6

7

8

9

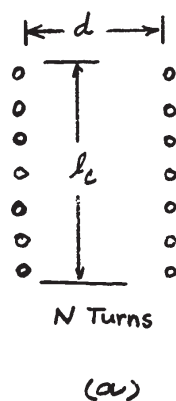
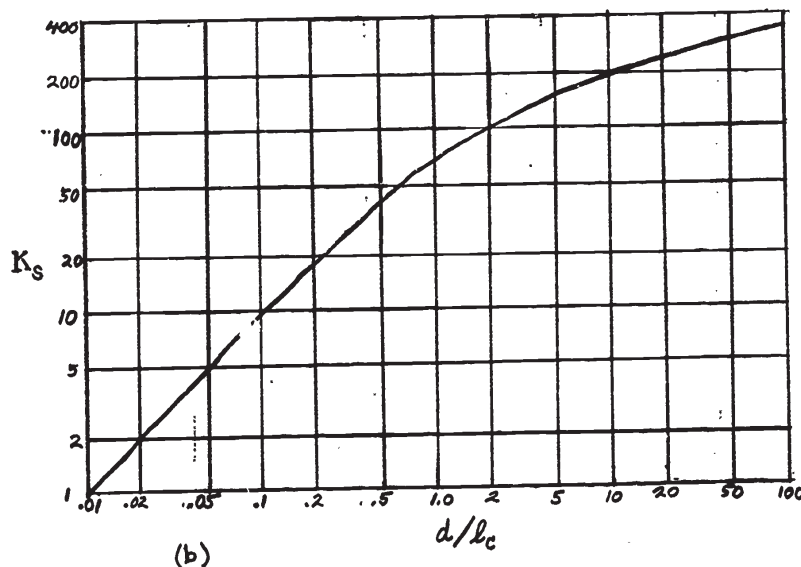


Figure 4-31.*

*For a more accurate graph, see Analysis of AC Circuits, p. 399, McGraw-Hill Book Co. 1952

4-12. Mutual Inductance

When two coils are in proximity, possible in air, or on a common magnetic core, a changing current in one will cause a changing flux in the other, and will therefore induce a voltage. Examples are shown in Figure 4-32.

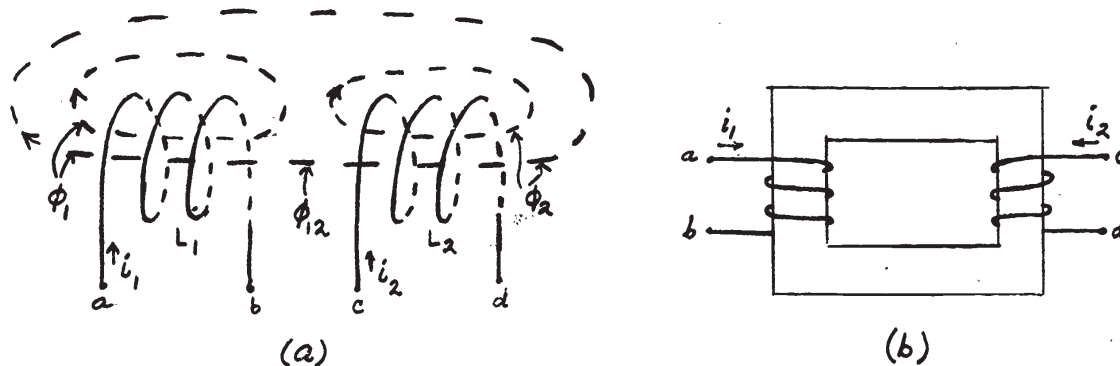


Figure 4-32.

Referring to case (a), let current i_1 be flowing in coil 1, creating flux ϕ_1 through itself and flux

$$\phi_{12} = k_{12}\phi_1$$

through coil 2 (the subscripts 12 imply flux through coil 2 due to current in coil 1). The factor k_{12} is that fraction of flux through coil 1 which also links coil 2, and is always less than unity. An emf will be induced in coil 2, given by

$$\begin{aligned} e_{cd} &= -N_2 \frac{d\phi_{12}}{dt} \\ &= -N_2 k_{12} \frac{d\phi_1}{dt} \end{aligned}$$

However, since $L_1 = N_1\phi_1/i_1$, where L_1 is the inductance of coil 1, we have

$$e_{cd} = -\left(\frac{N_2}{N_1} k_{12} L_1\right) \frac{di_1}{dt} \quad (4-45)$$

Now suppose the current in coil 1 is zero and that there is a varying current in coil 2. A similar analysis applies, this time involving a fraction k_{21}

4-44

- 1 (that fraction of the flux through coil 2 which also links coil 1). We get

$$e_{ab} = - \left(\frac{N_1}{N_2} \right) k_{21} L_2 \frac{di_2}{dt} \quad (4-46)$$

- 2 Experiment shows that the quantities in parentheses in Eqs. (4-45) and (4-46) are the same. This fact is an example of the principle of reciprocity. Thus, both equations can be written

$$3 \quad e_{ab} = - M \frac{di_2}{dt} \quad \text{and} \quad e_{cd} = - M \frac{di_1}{dt} \quad (4-47)$$

- where M , called the mutual inductance between the coils, stands for either of the quantities in parentheses. M is related to L_1 and L_2 in a manner that can be determined by referring to Eqs. (4-45) and (4-46), to get

$$M = \frac{N_2 k_{12} L_1}{N_1} \quad \text{and} \quad M = \frac{N_1 k_{21} L_2}{N_2}$$

- 5 Multiplying these two equations together gives

$$M^2 = k_{12} k_{21} L_1 L_2$$

- 6 and finally we have

$$M = k \sqrt{L_1 L_2} \quad (4-48)$$

- 7 where $k = \sqrt{k_{12} k_{21}}$ is called the coefficient of coupling. Since k_{12} and k_{21} are each less than unity, k is less than unity, giving the result

$$M \leq \sqrt{L_1 L_2}$$

8

Values of k very close to unity can be attained with iron cores, but in air values of k may be very small. Two coils placed at right angles will have no coupling, and k will then be zero.

- 9 A device consisting of two or more mutually coupled coils is called a transformer.

The possibility exists that M can vary with time, for example, if the coils are in relative motion. In that case, Eqs. (4-47) become

$$e_{ab} = - \frac{d(Mi_2)}{dt} \quad \text{and} \quad e_{cd} = - \frac{d(Mi_1)}{dt} \quad (4-49)$$

Equations (4-47) and (4-49) apply to the arrangements shown in Figure 4-32, where the winding directions are clearly shown. Reversal of the direction of either winding will change all signs to positive. The possibility of a sign ambiguity makes it necessary to show winding directions, or to have a symbol for signifying what signs to use. In Figure 4-32 each current (i_1 or i_2) enters the terminal (a or c) which appears first in the emf symbol (e_{ab} or e_{cd}). Note that these reference current directions are magnetically aiding. Each will produce a flux in the same direction through both coils. An alternate system which avoids showing actual winding directions is to place distinguishing marks (usually dots) on a diagram, as shown for two cases in Figure 4-33. These

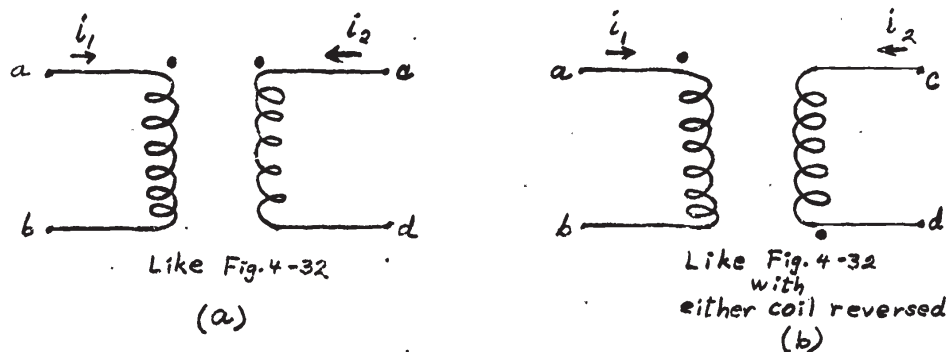


Figure 4-33.

dots mean that in the actual transformer symbolized by the diagram, currents entering the terminals with dots will be magnetically aiding. The pertinent equations for these two cases are

$$e_{ab} = \mp M \frac{di_1}{dt} \quad \text{and} \quad e_{cd} = \mp M \frac{di_2}{dt} \quad (4-50)$$

where the - signs are for case (a) and the plus signs for case (b).

Of course, in the above discussion it was assumed that only one coil carries a current at a time. If both coils carry changing currents, self inductance

4-46

1 must be included, giving (for the winding directions in Figure 4-32)

$$e_{ab} = -L_1 \frac{di_1}{dt} - M \frac{di_2}{dt}$$

$$e_{cd} = -M \frac{di_1}{dt} - L_2 \frac{di_2}{dt} \quad (4-50)$$

Finally, if we include coil resistances, and consider terminal voltages v_{ab} and v_{cd} , we get

$$v_{ab} = R_1 i_1 + L_1 \frac{di_1}{dt} + M \frac{di_2}{dt}$$

$$v_{cd} = M \frac{di_1}{dt} + R_2 i_2 + L_2 \frac{di_2}{dt} \quad (4-51)$$

It is not unusual for two mutually coupled coils to be connected in series as in either of the cases shown in Figure 4-34. The total emf ($e_{ad} = e_{ab} + e_{cd}$)

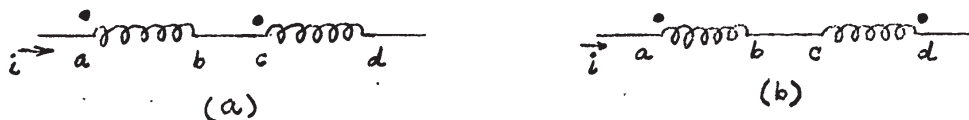


Figure 4-34.

will be obtained by adding Eqs. (4-50) and observing that $i_1 = i_2 = i$, giving

$$e_{ad} = -(L_1 + L_2 + 2M) \frac{di}{dt}$$

For case (b), the signs on both M terms in Eqs. (4-50) will be positive, and thus we then get

$$e_{ad} = -(L_1 + L_2 - 2M) \frac{di}{dt}$$

1 In each case, the quantity in parentheses is the inductance of the combination,
 2 and so we can say that when two coils are connected in series the combined
 inductance is

$$L = L_1 + L_2 \pm 2M \quad (4-52)$$

where the sign depends on the coil connections.

Transformers are among the most important of electrical devices. In power
 systems they provide means of changing voltage, and also are sometimes used to
 transmit power to circuits that are conductively isolated. They are also used
 in communication circuits, in filters and many other applications.

4-13. Transients in R-L Circuits

Consider the circuit shown in Figure 4-35(a). Resistance R includes the

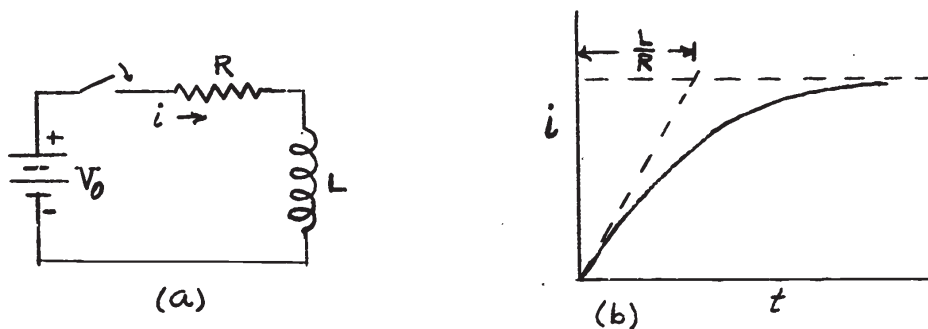


Figure 4-35.

coil resistance and any external resistance combined into one value. Assume
 the switch is closed at time $t = 0$, at which instant the current in the induc-
 tance is zero. As shown in the previous section, the Kirchhoff voltage law
 for this circuit is

$$V_0 = Ri + L \frac{di}{dt}$$

which can also be written

$$\frac{dt}{di} = \frac{L}{V_0 - Ri}$$

1 Recognizing the antiderivative, we have

$$\begin{aligned}
 t &= -\frac{L}{R} \ln(V_0 - Ri) + \ln K \\
 &= -\frac{L}{R} \ln\left(\frac{V_0 - Ri}{K}\right)
 \end{aligned}$$

where K is an arbitrary constant. Multiplying through by $-R/L$ and then writing the inverse of the logarithm to yield an exponential, we get

$$\frac{V_0 - Ri}{K} = e^{-Rt/L}$$

This is readily solved for i , as follows

$$i = \frac{V_0 - Ke^{-Rt/L}}{R} \quad (4-53)$$

This is a general solution which includes a constant K. It will be found that the original Kirchhoff voltage equation is satisfied by Eq. (4-53) regardless of the value of K. The solution to a specific problem, such as the one stated here, requires a specific value of K, and this is obtained by knowing the value of i at some particular value of t , this knowledge to be obtained from the statement of the problem. With a pair of corresponding values of i and t known, Eq. (4-53) can be solved for K. In most problems i is known at $t = 0$, and this value of i is called an initial condition.

Let us see how the initial condition is determined from a statement of the problem, which included the fact that the current is zero when the switch is closed (that is, for $t < 0$). We recognize that Eq. (4-53) is to apply for $0 \leq t$, and so we ask whether i as given by Eq. (4-53) should give $i = 0$ when $t = 0$. This would not be true if the current should experience a discontinuous jump as soon as the switch is closed. But such a jump is not possible because of the inductance. We see this by recalling that the voltage across L is $L(di/dt)$, and that with a finite source in the circuit this voltage must remain finite. On the other hand, a sudden jump of current would mean an infinite value of di/dt , and thus it is concluded that a sudden jump is impossible, and that the initial current is zero.

1 The substitution of $i = 0$ and $t = 0$ in Eq. (4-53) gives

$$K = V_0$$

2 and so the required solution is

$$i = \frac{V_0}{R}(1 - e^{-Rt/L}) \quad (4-54)$$

3 A graph of this function is shown in Figure 4-35(b). The factor L/R is the time constant, having the geometrical significance indicated. As the exponential term goes to zero, i approaches V_0/R , which is called the steady state value.

4 The significance of this result is the fact that a finite time interval is required to change the current in an inductor. Ninety per cent of the change will be accomplished when $e^{-Rt/L} = 0.1$, which is approximately when $t = 2.3 \times L/R$. In coils with very large inductance and low R , the time constant can be many seconds.

5 Perhaps the idea of change is better illustrated by Figure 4-36, where

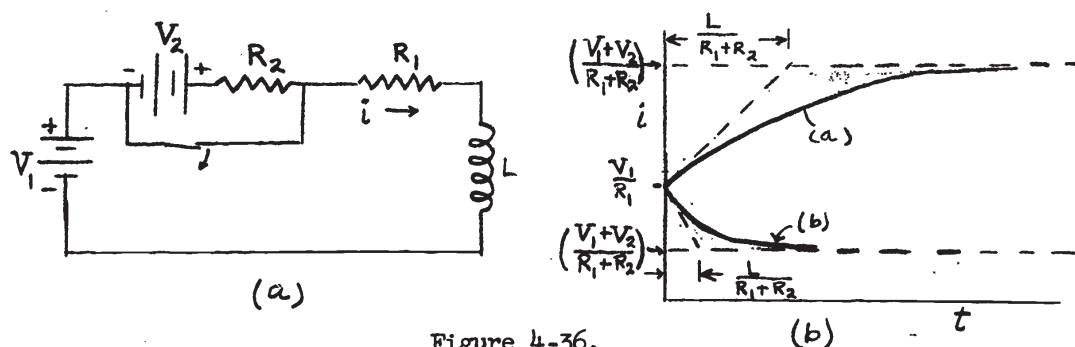


Figure 4-36.

the current initially has a non-zero value V_1/R , attained by virtue of source V_1 having been connected in the remote past. In this case, opening the switch at $t = 0$ inserts an additional source V_2 and resistor R_2 . Equation (4-53) applies if V_0 is replaced by $V_1 + V_2$, and R is replaced by $R_1 + R_2$, giving

$$i = \frac{V_1 + V_2}{R_1 + R_2} - K e^{-(R_1 + R_2)t/L}$$

4-50

- 1 We again get the initial condition from knowledge that the current cannot change suddenly, and so $i = V_1/R_1$ when $t = 0$, giving

2
$$\frac{V_1}{R_1} = \frac{V_1 + V_2 - K}{R_1 + R_2}$$

which, when solved for K , gives

3
$$K = \frac{V_2 R_1 - V_1 R_2}{R_1}$$

Direct substitution of this value of K yields

4
$$i = \frac{V_1 + V_2}{R_1 + R_2} - \frac{V_2 R_1 - V_1 R_2}{R_1 (R_1 + R_2)} e^{-(R_1 + R_2)t/L}$$

- Although this is a correct formula, it is not the most convenient form for interpretation because it does not place in direct evidence the initial value and the subsequent change. If we subtract V_1/R_1 from the constant term in the above equation we get

6
$$\frac{V_1 + V_2}{R_1 + R_2} - \frac{V_1}{R_1} = \frac{V_2 R_1 - V_1 R_2}{R_1 (R_1 + R_2)}$$

which is the same as the coefficient of the exponential in the equation for i . Thus, by adding and subtracting V_1/R_1 the result is

7
$$i = \frac{V_1}{R_1} + \frac{V_2 R_1 - V_1 R_2}{R_1 (R_1 + R_2)} \left[1 - e^{-(R_1 + R_2)t/L} \right] \quad (4-55)$$

- 8 By putting the solution in this last form, the initial ($t = 0$) value V_1/R_1 is clearly in evidence, since the bracketed term is zero when $t = 0$. The factor

9
$$\frac{V_2 R_1 - V_1 R_2}{R_1 (R_1 + R_2)}$$

is the total change in current, and can be either positive or negative depending on the relative magnitudes of $V_2 R_1$ and $V_1 R_2$. Both cases are illustrated in Figure 4-36(b). Similarity with Figure 4-35 is evident if it is noted that in each case a factor of the form $(1 - e^{-\alpha t})$ multiplies a factor which is the total change of the current. This principle can be used to solve a wide variety of switching problems in which there is a single inductor. Initial and final values of currents can be found from simple d-c circuit principles. The value of α is obtained by reducing all sources to zero, finding the equivalent resistance in series with the inductance, and dividing this by L .

An important special case is shown in Figure 4-37. Here, as a result of connection of the battery in the remote past, a current $V_0/(R_1 + R_2)$ is flowing in the inductor. At $t = 0$ the switch is closed, reducing the voltage between A and B to zero. Subsequent to $t = 0$, the current i changes from $V_0/(R_1 + R_2)$ to zero. Thus, the equation for the current is

$$i = \frac{V_0}{R_1 + R_2} - \frac{V_0}{R_1 + R_2} (1 - e^{-R_2 t/L})$$

Note that a negative sign appears on the second term because the change is a decrease, and also note that after the switch is closed, only R_2 is in series with L . The above reduces to the simpler form

$$i = \frac{V_0}{R_1 + R_2} e^{-R_2 t/L} \quad (4-56)$$

This can also be obtained directly from Eq. (4-53) by obtaining the appropriate value for K .

These sample solutions have all been given to find current. Of course, once i is known, the voltage across the inductor can readily be obtained from $L di/dt$.

4-14. Flux Linkage Theorem

In the previous section it was pointed out that $L di/dt$ must remain finite, because voltage across an inductor remains finite, and therefore that the current in an inductive circuit cannot experience a sudden jump. However, this principle applies only if L is constant with time, and if there is no changing flux due to external causes.

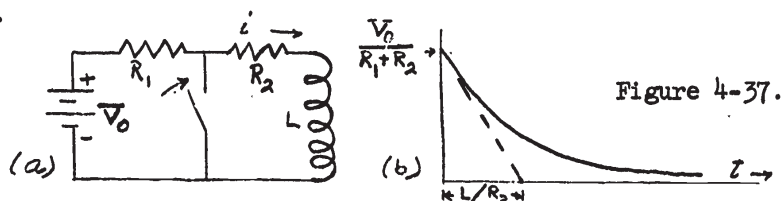


Figure 4-37.

Figure 4-38 shows the more general case which we shall consider briefly.

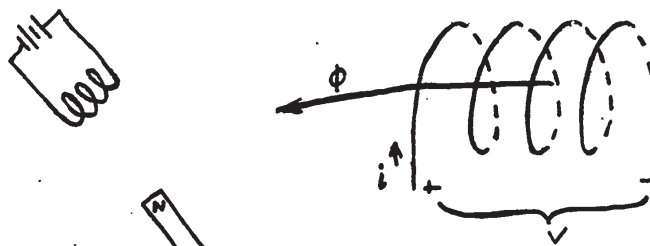


Figure 4-38.

The flux ϕ through the coil is due to current i , and also is in part due to external magnets or current-carrying coils. Thus, the total flux can be written

$$\phi = \phi_e + \frac{Li}{N} \quad (4-57)$$

where ϕ_e is due to causes external to the coil, and Li/N is the flux due to i , as determined from Eq. (4-42). The voltage v is

$$\begin{aligned} v &= \frac{d(N\phi)}{dt} \\ &= \frac{d}{dt} (N\phi_e + Li) \end{aligned} \quad (4-58)$$

In this case, again using the fact that v must remain finite, it is concluded that the quantity in parentheses cannot experience a sudden jump. However, any one of the quantities within the parentheses can change suddenly, provided there is a compensating change in one of the others.

Two cases are of practical importance. First, suppose there is a sudden change in the magnetic circuit associated with the coil, causing a sudden change ΔL in the inductance, and assume ϕ_e remains constant. Since $N\phi_e + Li$ must remain continuous (be free of a jump) it follows that there must be a corresponding sudden change in i such that

$$(L + \Delta L)(i + \Delta i) = Li$$

1 and consequently

$$\Delta i = -\left(\frac{\Delta L}{L + \Delta L}\right)i \quad (4-59)$$

2 As a second example, suppose ϕ_e suddenly changes by an amount $\Delta\phi_e$
due to a sudden mechanical displacement of one of the external flux
sources. Then we have

$$3 \quad N\phi_e + Li = N(\phi_e + \Delta\phi_e) + L(i + \Delta i)$$

and

$$4 \quad \Delta i = -\frac{N}{L}\Delta\phi_e \quad (4-60)$$

The principle discussed here, that the total flux linking a coil
($N\phi_e + Li$) cannot have sudden jumps is known as the constant flux
linkage theorem.*

5 This theorem is useful in solving circuit problems in which certain
types of switching operations cause sudden changes in inductance. Con-
sider, for example, Figure 4-39(a) in which there are two inductances
which are not magnetically coupled. Assuming the current has reached
6 the steady state value V/R , at time $t = 0$ the switch across one induc-
tance is closed, suddenly reducing the total inductance from $3L$ to $2L$.
In other words, $\Delta L = -L$, and so from Eq. (4-59) we find that the cur-
rent increases an amount.

$$7 \quad \Delta i = -\left(\frac{-L}{3L-L}\right)\frac{V}{R} = \frac{V}{2R}$$

8 Adding this change to the original current gives $(3/2)(V/R)$ as the
current immediately after the switch is closed.

Subsequent behavior of this circuit can be analyzed by methods
given in the previous section, using $(3/2)(V/R)$ as the initial value.
From inspection of the circuit it is evident that the current will
9 gradually go back to V/R along an exponential curve of time constant
 $L/2R$, as shown in Figure 4-39(b).

* In this statement the word "constant" means constant over a short period of time. Of course, flux linkage can change, but it must do so smoothly, in a continuous fashion, to use mathematical terminology. It would be better to call this the continuous flux linkage theorem.

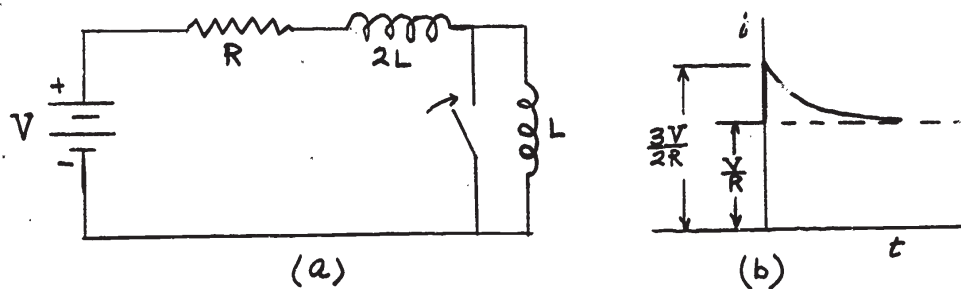


Figure 4-39.

4-15. Energy Stored in a Magnetic Field

Consider a torroidal inductor with iron core, as in Figure 4-40, and assume the core diameter d is small enough to allow the assumption of uniform B over the cross section, and assume the coil resistance is negligible. The power delivered to the coil during a change of current is

$$v_{ab} i = (N \frac{d\phi}{dt}) i$$

Now suppose that in time interval from t_1 to t_2 the flux changes from ϕ_1 to ϕ_2 .

The energy delivered from the source will be the time integral of power, namely

$$W = \int_{t_1}^{t_2} v_{ab} i dt = N \int_{t_1}^{t_2} i \frac{d\phi}{dt} dt = \int_{\phi_1}^{\phi_2} Ni d\phi$$

Now observe that, in terms of B and H , we have

$$\phi = BA \quad \text{and} \quad Ni = Hl$$

so that we get

$$W = lA \int_{B_1}^{B_2} H dB \quad (4-61)$$

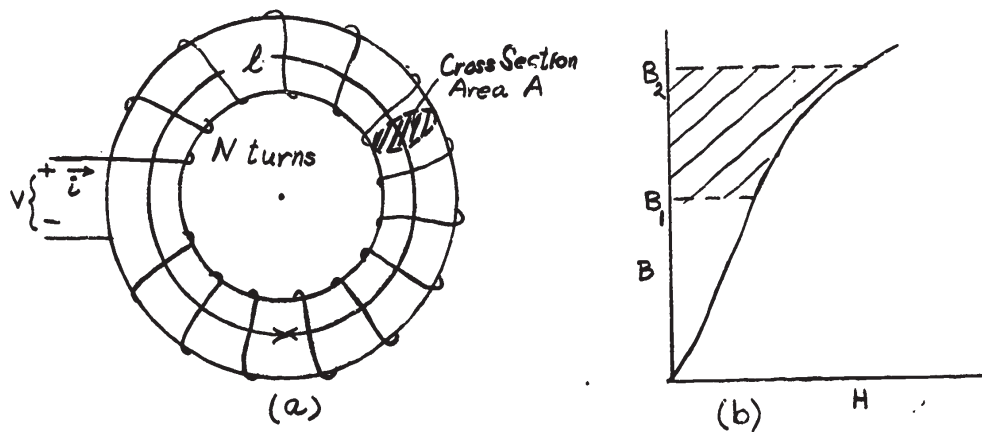


Figure 4-40.

This integral is interpreted as an area in Figure 4-40(b), for the case where the magnetism has been increased from B_1 to B_2 . If $B_1 = 0$, we get

$$w = \int_0^B H dB \quad (4-62)$$

which is interpreted as the energy density in the magnetic field when its strength is B . Actual energy is then obtained by multiplying the density by the volume lA . In accordance with this result for the special case of a toroid it is postulated that Eq. (4-62) gives the energy density for all field configurations.

No assumption as to linearity was made in arriving at Eqs. (4-61) and (4-62). Suppose magnetism is carried partially around a hysteresis loop, from 0 to P to Q, in Figure 4-41. In going from 0 to P the energy density increases

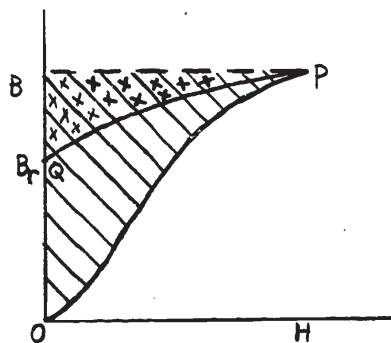


Figure 4-41.

by an amount

$$\int_0^{B_1} H dB$$

which is represented by the shaded area. In reducing B to B_r , the energy density will change by

$$\int_{B_1}^{B_r} H dB = - \int_{B_r}^{B_1} H dB$$

which is a decrease equal to the area covered by the small crosses. This energy is returned to the circuit. Thus, an amount of energy represented by the area between the curves has not been returned, even though the current has returned to zero. This represents an energy loss due to the molecular interactions which cause hysteresis. When magnetism is carried completely around a hysteresis loop, energy is lost in proportion to the area of the loop. This energy loss is an important consideration of the design of generators, motors, and transformers, since in all of these there are repeated reversals of flux.

When the magnetic material is linear, so that we may write $B = k_m \mu_0 H$, Eq. (4-62) reduces to

$$w = \frac{1}{k_m \mu_0} \int_0^B B dB = \frac{B^2}{2k_m \mu_0} = \frac{BH}{2} \quad (4-63)$$

Energy storage in linear media can also be expressed in terms of inductance. Using the reference directions of Figure 4-42, if the current increases from 0 to 1,



Figure 4-42.

1 as i goes from 0 to i , for the energy we have

$$W = \int_0^i i L \frac{di}{di} di = L \int_0^i i di$$

2 or

$$W = \frac{1}{2} L i^2 \quad (4-64)$$

3 We can get the same result from Eq. (4-63), for the torroidal case.
The total energy is

$$4 \quad W = \frac{BA H \ell}{2} = \frac{\phi N i}{2}$$

but $\phi N = Li$, from Eq. (4-42), and so Eq. (4-64) is obtained from the above.

5 4-16. Force Across an Airgap

Airgaps in magnetic circuits experience forces which tend to cause them to close. This force is utilized in a variety of applications, for example in electric relays.

6 The idea of energy density can be used to determine the force tending to close an airgap. Referring to Figure 4-43, suppose the gap is small, so that there will be negligible error in assuming B lines are parallel in the gap, and that B is negligible outside the gap. The pole face has an area A , and the gap length is ℓ , giving

$$7 \quad W = \frac{A \ell B^2}{2 \mu_0} \quad (4-65)$$

8 for the energy stored in the gap. Now suppose the gap is increased an amount $\Delta \ell$ by the application of an external force F . To do this in practice, there must be a hinge in the iron circuit; otherwise, the force will also have to overcome stress in the iron. Thus, F is only that force required to move the pole face against magnetic force.

9

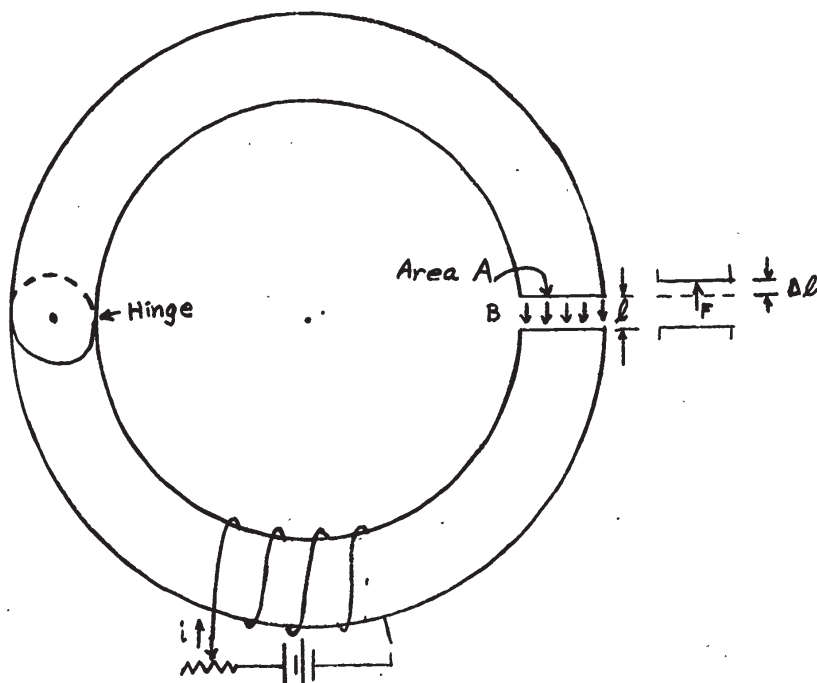


Figure 4-43.

If conditions of the external circuit remain constant, increasing the airgap would cause a reduction of ϕ which would complicate the analysis. To avoid this, assume that, coincident with the change in l , ~~this is gradually increased by~~ adjustment of Resistor R , so as to maintain ϕ constant. All mechanical energy expended will then go into the energy stored in the field.* This change is

$$\Delta W = \frac{AB^2}{2\mu_0} \Delta l$$

But from mechanics we know that this energy is also $F\Delta l$, and so we have

$$F = \frac{AB^2}{2\mu_0} \quad (4-61)$$

as the expression for the required force.

* No energy interchange occurs with the battery due to change in l because $d\phi/dt = 0$, and so the coil voltage is zero.

Chapter 5

STEADY STATE CIRCUIT ANALYSIS

Introduction

We begin this chapter with a description of the experiment illustrated in Fig. 5-1a, in which an alternating current generator is suddenly connected to a circuit consisting of R and L . Oscilloscopic records are taken of the voltage and current, yielding curves as shown in Fig. 5-1b. The voltage wave begins at

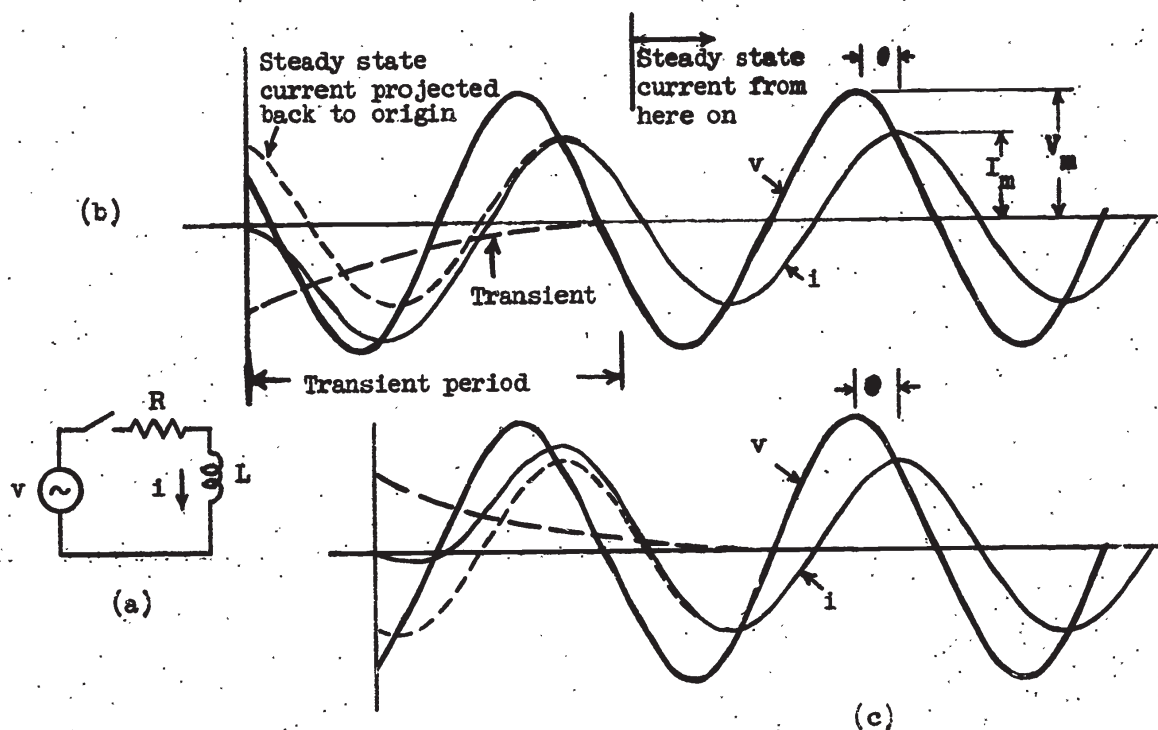


Figure 5-1.

an arbitrary point, $t = 0$, depending on the instant the switch was closed. The form of the current wave will depend upon this instant, and so the experiment will not yield the same result with each trial. Another example is shown in Fig. 5-1c. However, after a period of time, the current wave settles down to a form called the steady state wave. If the steady state wave is projected back to $t = 0$, as indicated by the dashed curves, it is found that the actual current can be viewed as the sum of this wave and the curve labelled transient. Whenever

the network includes resistance elements, as in the case under consideration, the transient will die out. In more complicated networks the form of the transient may be more complicated than the simple exponential shown.

This brief analysis illustrates the following:

(1) When a sinusoidal source is suddenly applied to a circuit, there is a transient period during which the response is not a repeating wave. (In the example above, the response was the current.)

(2) The actual form of the response during the transient period depends on the instant at which the switch is closed.

(3) After the transient has died out, the wave approaches a steady state wave having a magnitude and position on the time axis, relative to the source wave, which are independent of the time of switching. Thus, referring to (b) and (c) of Fig. 5-1, the ratio V_m/I_m and the phase difference, θ , are the same for both.

Although this has been presented as a description of an imagined experiment, theoretical confirmation of the conclusions is possible. However, such theoretical confirmation must be deferred until circuit analysis has been studied quite thoroughly, and therefore will not be considered at this time.

The fact that the steady state response is independent of what takes place during the transient period makes steady state analysis possible.

One may legitimately question whether a study of the steady state response of a network has practical value. In replying to this question, either of two different answers can be given, depending on whether a network is designed primarily to transmit power or communication signals. Power transmission usually takes place at a nearly constant rate, and so steady state analysis provides most of the needed answers. On the other hand, a communication system can be thought of as a power system in which signals are coded in the form of changes in the power transmitted (i.e., the turning on of a light can be a signal). Thus, transient considerations must also enter into the analysis of communication circuits. However, the analysis of networks for steady state response provides an important foundation for the study of both power and communication systems.

5-1. Description of a Sinusoid

Consider a plot of the function

$$i = I_m \cos(\omega t + \alpha) \quad (5-1)$$

which is shown in Fig. 5-2a. It is labelled as a current, but a similar expression could be written for a voltage or other quantity. In this equation, I_m , ω , and α are constants, and t is the variable time. Since ω is constant, ωt can

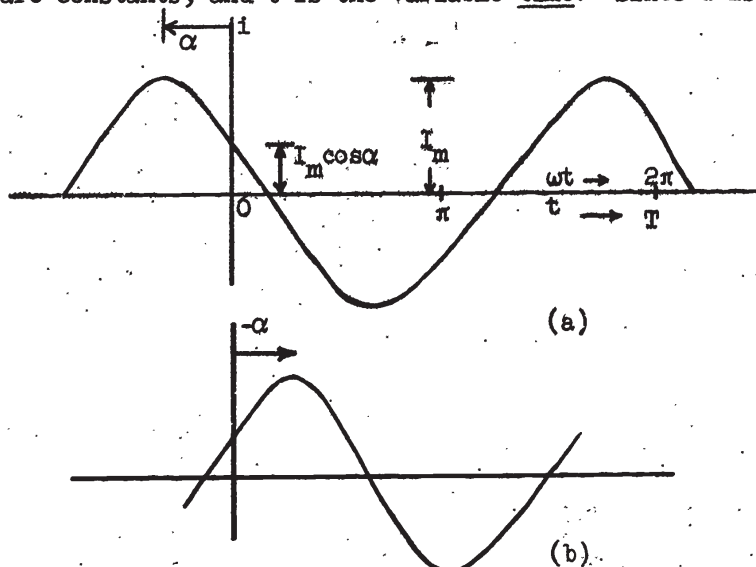


Figure 5-2.

also be regarded as a variable. Both these variables (t and ωt) are plotted along the horizontal axis in Fig. 5-2a. Since time has no beginning, it is necessary to choose some arbitrary instant as $t = 0$. At this instant, Eq. (5-1) gives $i = I_m \cos \alpha$, as indicated. Since α determines the value of $(\omega t + \alpha)$ when $t = 0$, it is called the initial angle of the wave. I_m is the maximum value, or amplitude.

To interpret the parameter ω , note that, since the cosine of an angle is unchanged if 2π is added to the angle, the wave will repeat itself when ωt has increased by 2π . Thus, the wave is said to be periodic. Let T be the increase in time t required to make ωt increase by 2π , so that

$$\omega T = 2\pi$$

or

$$\omega = \frac{2\pi}{T}$$

T is the period of the wave; and $1/T = f$, called the frequency, is the number of times the wave repeats in one second. (Earlier, the symbol T was used to

5-4

denote the time constant of an RC or RL network. Now, the same symbol is being used for the period of a periodic function.) For example, if $T = 0.1$ sec., there will be 10 repetitions of the wave in one second. In terms of f , Eq. (5-2) becomes

$$\omega = 2\pi f \quad (5-3)$$

Thus, ω has been established as a parameter related to the period (or frequency) of the wave. In view of ωt being an angle, ω is called angular frequency.

If I_m , α and ω are known as numerical values, the corresponding wave can be plotted unambiguously. Thus, these three parameters completely specify a wave. Furthermore, ω is determined by the frequency of the source; if this is given, the solution of steady state problems amounts to the finding of values of magnitudes and initial angles of various voltage and current waves in a network.

The decision to write Eq. (5-1) as a cosine is entirely arbitrary. If we let $\alpha = \alpha' - \frac{\pi}{2}$, Eq. (5-1) becomes

$$I_m \cos(\omega t + \alpha' - \frac{\pi}{2}) = I_m \sin(\omega t + \alpha')$$

Thus, any such wave can be written as a cosine function or a sine function. The cosine form has certain advantages, and so we shall use it in this text. Regardless of whether it is written as a cosine or sine function, the wave we are considering is called a sinusoid, and its shape is said to be sinusoidal.

Although the magnitude parameter I_m is entirely adequate to determine the size of a sinusoid, another value (called the rms value) is usually employed. This value, which we shall identify as I_{rms} or simply I , is the square root of the average (mean) value of the square of the wave. (The symbol rms is an abbreviation for the words "root", "mean", "square".) For simplicity, consider the current wave (with zero initial angle)

$$i = I_m \cos \omega t$$

for which the squared wave is

$$\begin{aligned} i^2 &= I_m^2 \cos^2 \omega t \\ &= \frac{I_m^2}{2} (1 + \cos 2\omega t) \end{aligned}$$

The second term shows that the i^2 curve varies in such a way as to have equal areas above and below the dashed line in Fig. 5-3. Therefore, this dashed line is the average value of the i^2 curve*, and so we can write

$$\text{Average of } i^2 = \frac{I_m^2}{2}$$

and the square root of this, the rms value, is

$$I = \frac{I_m}{\sqrt{2}} \quad (5-4)**$$

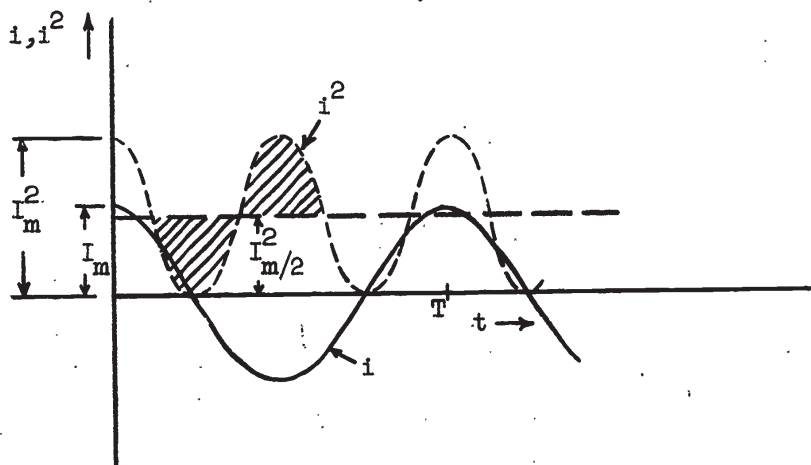


Figure 5-3.

For an interpretation of the rms value, assume $i = I_m \cos \omega t$ is the current in a resistor R . At any instant of time, power will be dissipated at a rate $i^2 R$, and from the above proof that the average of i^2 is $I_m^2/2$, it follows that the average power dissipation will be $(I_m^2/2)R$. In view of Eq. (5-4) this is also $I^2 R$. Thus, for finding average power, I can be used in the $I^2 R$ formula exactly as if it were

*Strictly speaking, the areas above and below the dashed curve are equal only if reckoned over an integral multiple of the period of the wave. For example, the dashed line would not be the average reckoned over a period and a half, but it can be shown that over a long time interval (many periods) the average is only slightly affected by the inclusion of a fraction of a period. That is, the averages over 100 periods and over $100 + 1/2$ periods are very nearly the same, as can be seen by making a simple calculation.

**Rms values can be defined for waveshapes which are not sinusoidal. However, the $1/\sqrt{2}$ relationship does not then generally apply. Equation (5-4) is given here for a sinusoidal wave.

a d-c current. Since I has this interpretation in terms of the average effectiveness of the wave in producing heat in a resistor, I is sometimes called the effective value of the wave. The rms designation is to be preferred, however.

Most a-c instruments for measuring voltage and current are calibrated to indicate rms values. Since the rms value is used more prevalently than the amplitude, in this chapter we shall use the rms value in writing an equation for a sinusoid. Accordingly, we write Eq. (5-1) in the alternate form

$$i = \sqrt{2} I \cos(\omega t + \alpha) \quad (5-5)$$

5-2 Complex Numbers

Complex numbers were invented by mathematicians to generalize the theory of quadratic equations. For example, the equation

$$x^2 - 2bx + c = 0$$

has roots given by

$$x_1 = b + \sqrt{b^2 - c}, \quad x_2 = b - \sqrt{b^2 - c}$$

If $b^2 \geq c$, x_1 and x_2 are real numbers. If $b^2 < c$, the square root of a negative number results. But there is no real number whose square is negative. Nevertheless, these same formulas for the roots can still apply if an imaginary unit j is defined such that

$$j^2 = -1 \quad (5-6)*$$

Then, if $b^2 < c$, the two roots can be written

$$x_1 = b + j\sqrt{c - b^2}, \quad x_2 = b - j\sqrt{c - b^2}$$

Each of these is a complex number consisting of a real part b and an imaginary part $\sqrt{c - b^2}$. But it is one thing merely to write $b \pm j\sqrt{c - b^2}$, and another to know what it means. In mathematics an "absolute" meaning is not required; meaning is

*The symbol i is used in mathematics, but due to the extensive electrical engineering literature in which i is used for current, the symbol j is standard in electrical engineering.

embodied in the rules of operations which are defined for a quantity. The rules of operating on a complex number are defined in such a way as to make them useful. For example, in order for x_1 to satisfy the given equation, it is necessary to define what is meant by x_1^2 and $-2bx_1$, when $x_1 = b + j\sqrt{c-b^2}$. The required definition is that a complex number shall be treated like an ordinary binomial, with j^2 being replaced by -1 whenever j^2 occurs. Thus

$$\begin{aligned} x_1^2 &= (b + j\sqrt{c-b^2})^2 = b^2 + j^2(c-b^2) + j2b\sqrt{c-b^2} \\ &= 2b^2 - c + j2b\sqrt{c-b^2} \end{aligned}$$

and

$$-2bx_1 = -2b^2 - j2b\sqrt{c-b^2}$$

The sum of these is $-c$, and so it is seen that $x^2 - 2bx + c = 0$ is satisfied for $x = x_1$.

Let the general notation

$$\bar{A} = A_1 + jA_2$$

$$\bar{B} = B_1 + jB_2 \quad (5-7)*$$

represent two complex numbers. (In this text a bar above a symbol will be used to denote that it is complex.) For reference, we shall state the rules of algebraic operations in terms of these examples. But first we must define what is meant by equality. By $\bar{A} = \bar{B}$ we mean that the real parts are equal to each other and the imaginary parts are equal to each other. That is,

$$A_1 = B_1 \quad \text{and} \quad A_2 = B_2$$

*A complex number is merely a combination of two numbers, which cannot themselves be combined. A good way to regard this pair of numbers is

$$\bar{A} = (1)A_1 + (j)A_2$$

This shows a symmetry between the two parts. One part is $A_1 \times$ (the real unit 1), and the other part is $A_2 \times$ (the imaginary unit j). A_1 and A_2 are real.

5-8

Conversely, if $A_1 = B_1$ and $A_2 = B_2$, we can write $\bar{A} = \bar{B}$. In other words, two complex numbers are equal if and only if their real parts are equal and their imaginary parts are equal.

For the operations of addition, multiplication and division, using the ordinary rules for binomials, we get

$$\begin{aligned}(A_1 + jA_2) + (B_1 + jB_2) &= (A_1 + B_1) + j(A_2 + B_2) \\(A_1 + jA_2)(B_1 + jB_2) &= A_1B_1 - A_2B_2 + j(A_2B_1 + A_1B_2) \quad (5-8) \\ \frac{A_1 + jA_2}{B_1 + jB_2} &= \frac{A_1B_1 + A_2B_2 + j(A_2B_1 - A_1B_2)}{B_1^2 + B_2^2}\end{aligned}$$

The validity of the formula for division can be confirmed by showing that an identity results when both sides of the equation are multiplied by $B_1 + jB_2$. The formula for division is rather complicated, but it can be reconstructed by multiplying numerator and denominator by the number $B_1 - jB_2$, as follows:

$$\frac{A_1 + jA_2}{B_1 + jB_2} = \frac{(A_1 + jA_2)(B_1 - jB_2)}{(B_1 + jB_2)(B_1 - jB_2)} = \frac{A_1B_1 + A_2B_2 + j(A_2B_1 - A_1B_2)}{B_1^2 + B_2^2}$$

The number $B_1 - jB_2$ used here is related to $\bar{B}$ by having the sign of the imaginary part changed. This number is called the complex conjugate of $\bar{B}$, and will be designated by $\bar{B}^*$. Using the conjugate, it follows that:

$$\begin{aligned}\bar{B} \bar{B}^* &= B_1^2 + B_2^2 \\ \frac{\bar{B} + \bar{B}^*}{2} &= B_1 \\ \frac{\bar{B} - \bar{B}^*}{2j} &= B_2\end{aligned} \quad (5-9)$$

The first of these follows directly from the rule for multiplication; the second and third are readily obtained by substituting $\bar{B} = B_1 + jB_2$ and $\bar{B}^* = B_1 - jB_2$. Thus, by combining a complex number with its conjugate, the real or imaginary part, or the sum of their squares can be obtained. (Each of these is a real number.)

1 The relationship

$$e^{j\theta} = \cos \theta + j \sin \theta \quad (5-10)^*$$

2 is called Euler's identity. The right-hand side of this equation is the definition
of the symbol $e^{j\theta}$. Thus, the formula requires no proof, but properties of $e^{j\theta}$ so
defined must be investigated. The properties of $e^{j\theta}$ we shall need are briefly
treated in the footnote, where it is shown that $e^{j\theta}$ can be multiplied and dif-
ferentiated like a real exponential.

3 Euler's identity can be used to provide an alternate notation for a complex
number

$$\bar{A} = A_1 + jA_2$$

4 This can be shown by multiplying both sides of Eq. (5-10) by a real number A,
to give

$$Ae^{j\theta} = A \cos \theta + j A \sin \theta$$

5 In view of the meaning of equality of complex numbers, we see that

$$Ae^{j\theta} = \bar{A}$$

6 if

$$A_1 = A \cos \theta \quad \text{and} \quad A_2 = A \sin \theta \quad (5-11)$$

7 *Equation (5-10) can be regarded as a definition because the number e raised to
an imaginary power has no meaning in the sense of a real exponent. It will be
found that all operations which can be carried out with real exponentials will
be true for Eq. (5-10). For example, consider the derivative of the left-hand
side with respect to θ . Using the rule for real exponents, we get

$$8 \quad \frac{de^{j\theta}}{d\theta} = je^{j\theta} = j(\cos \theta + j \sin \theta) = -\sin \theta + j \cos \theta$$

which is the same as the derivative of the right-hand side of Eq. (5-10). Also,
with the aid of the identities for $\cos(x+y)$ and $\sin(x+y)$ it can be shown that

$$e^{jx} e^{jy} = \cos(x+y) + j \sin(x+y) = e^{j(x+y)}$$

9 In other words, when exponentials are multiplied, the exponents add, in similar
ity with real exponentials.

5-10.

Assuming A_1 and A_2 were specified, these two equations can be solved for A and θ . Thus, squaring both equations and adding gives

$$A_1^2 + A_2^2 = A^2(\cos^2 \theta + \sin^2 \theta) = A^2$$

or

$$A = \sqrt{A_1^2 + A_2^2} \quad (5-12)$$

Also, dividing one equation by the other gives

$$\frac{A_2}{A_1} = \frac{\sin \theta}{\cos \theta}$$

and so

$$\theta = \arctan \frac{A_2}{A_1} \quad (5-13)$$

$Ae^{j\theta}$ is called the exponential form, whereas $A_1 + jA_2$ is called the rectangular form of $\bar{A}$. The latter terminology comes from Eqs. (5-11) which show that A_1 and A_2 can be viewed as the rectangular coordinates of a point whose position in a plane is specified in the polar coordinates A and θ . This is illustrated in Fig. 5-4. Since these rectangular coordinates are the components of a complex number, it is

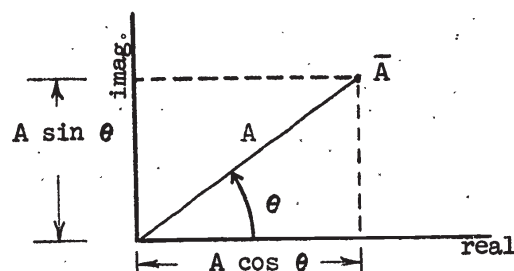


Figure 5-4.

customary to label the axes "real" (meaning real part) and "imaginary" (meaning imaginary part). A pair of axes so labeled constitutes a complex plane. A is called the magnitude of $\bar{A}$, and θ its angle.

The notation $Ae^{j\theta}$ is very convenient for analytical work, particularly when differentiation with respect to θ is involved. However, for specifying numerical

values, like $2e^{j\pi/3}$, it is more convenient to write $2/60^\circ$. This is called the polar form, since it merely involves specifying the polar coordinates of the point. (In the polar form, angles are usually specified in degrees.)

Observe that the exponential notation provides a simple interpretation for multiplication. Thus, if $\bar{A} = Ae^{j\theta_a}$ and $\bar{B} = Be^{j\theta_b}$ are two complex numbers, their product is

$$\overline{AB} = (AB)e^{j(\theta_a + \theta_b)}$$

In other words, magnitudes are multiplied and angles are added. Also note that in exponential notation the sign of the exponent is reversed in order to get a conjugate. Thus, if $\bar{A} = Ae^{j\theta}$, $\bar{A}^* = Ae^{-j\theta}$ (as can be seen by converting to the rectangular forms, since $\bar{A} = A \cos \theta + jA \sin \theta$ and $\bar{A}^* = A \cos \theta - jA \sin \theta$). In this notation, $\overline{AA^*} = A^2 e^{j(\theta - \theta)} = A^2$, which is the same as $A_1^2 + A_2^2$.

5-3 Complex numbers applied to circuit theory.

Let us now consider how complex numbers can be used to represent a sinusoidal wave like

$$i = \sqrt{2} I \cos(\omega t + \alpha) \quad (5-14)$$

Since

$$Ie^{j(\omega t + \alpha)} = I \cos(\omega t + \alpha) + j \sin(\omega t + \alpha)$$

$$Ie^{-j(\omega t + \alpha)} = I \cos(\omega t + \alpha) - j \sin(\omega t + \alpha)$$

it follows that⁺

$$i = \sqrt{2} I \cos(\omega t + \alpha) = \sqrt{2} \frac{Ie^{j(\omega t + \alpha)} + Ie^{-j(\omega t + \alpha)}}{2} \quad (5-15)$$

The advantage to be gained by making this change depends upon defining an additional complex number

$$\bar{I} = Ie^{j\alpha} \quad (5-16)$$

⁺This is merely an application of the second of Eqs. (5-9).

Then, Eq. (5-15) can be written

$$\begin{aligned}
 i &= \sqrt{2} \frac{I e^{j\alpha} e^{j\omega t} + I e^{-j\alpha} e^{-j\omega t}}{2} \\
 &= \sqrt{2} \frac{\bar{I} e^{j\omega t} + \bar{I}^* e^{-j\omega t}}{2} \quad (5-17)
 \end{aligned}$$

This formula is important. It shows that, given a sinusoid $i = \sqrt{2} I \cos(\omega t + \alpha)$, a complex number $\bar{I} = I e^{j\alpha}$ can be constructed from the two parameters I and α which describe the wave, and that an expression for the sinusoid can be given in terms of that complex number. The importance of this idea cannot be overestimated. Conversely, if Eq. (5-17) is given, the properties of the wave are apparent merely from looking at the complex number $\bar{I}$. Observe that the quantity $\bar{I}^*$ is superfluous insofar as relationship of Eq. (5-17) to the wave is concerned, because $\bar{I}^*$ is completely determined by $\bar{I}$. To illustrate numerically,

$$\sqrt{2} \frac{3e^{j\pi/3} e^{j\omega t} + 3e^{-j\pi/3} e^{-j\omega t}}{2}$$

is an expression for $\sqrt{2} (3) \cos(\omega t + \pi/3)$. Likewise,

$$\sqrt{2} \frac{(2+j2)e^{j\omega t} + (2-j2)e^{-j\omega t}}{2}$$

is an expression for $4 \cos(\omega t + \pi/4)$, since $2 + j2 = 2\sqrt{2} e^{j\pi/4}$.

Referring back to Eq. (5-17), since $\bar{I}$ carries all the information (except frequency) necessary to construct the sinusoidal wave, it is given a special name. $\bar{I}$ is called the phasor symbol for the wave.⁺ When a sinusoidal wave is expressed in complex form, the phasor symbol for that wave can always be recognized as the coefficient which multiplies $e^{j\omega t}$.

We shall now develop applications of Eq. (5-17) to the steady state analysis of networks. Figure 5-5a shows an inductor, for which the voltage and current are related by

$$v_L = L \frac{di}{dt}$$

⁺ In the older electrical engineering literature this is called a vector. However, calling $\bar{I}$ a vector is to be deprecated, because it does not have all the properties of a vector. Also, in some books it is called a sinor.

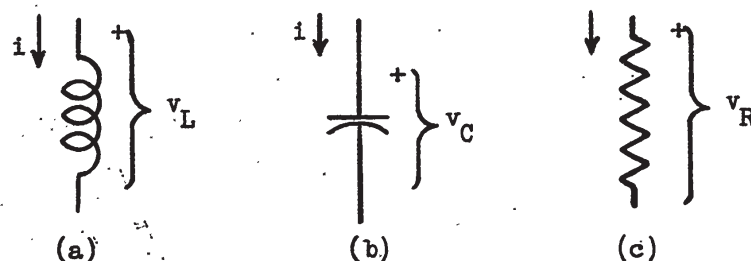


Figure 5-5.

Assume

$$i = \sqrt{2} I \cos(\omega t + \alpha) + I_{dc} \quad (5-18)$$

where I_{dc} is a constant (d-c) term. The voltage v_L is obtained by differentiating the current and multiplying by L , giving

$$v_L = -\sqrt{2} \omega L I \sin(\omega t + \alpha) \quad (5-19)$$

which is a sinusoidal wave. This can also be written as a cosine function by observing that $-\sin x = \cos(x + \pi/2)$, so that

$$v_L = \sqrt{2} \omega L I \cos(\omega t + \alpha + \frac{\pi}{2}) \quad (5-20)$$

Particular note should be made of the fact that, even if the current contains a d-c component, the voltage will be a sinusoidal wave, because the derivative of a constant is zero. In many practical circuits, particularly those containing transistors or vacuum tubes, there will be d-c components. However, these can be calculated by applying the principles of d-c circuit analysis and, as illustrated by this example, the sinusoidal component of the current is related to the sinusoidal voltage wave independently of the d-c component. Thus, in all subsequent consideration of inductor currents in this chapter we will deal only with a sinusoidal wave of current, recognizing that there may also be a d-c component, but that such a component will not be included in the expressions used for the inductor current.

Let us return to the sinusoidal component of current

$$i = \sqrt{2} I \cos(\omega t + \alpha)$$

and the voltage wave

$$v = \sqrt{2} \omega L I \cos(\omega t + \alpha + \frac{\pi}{2})$$

These can be written respectively in the complex notation

$$i = \sqrt{2} \frac{I e^{j\alpha} e^{j\omega t} + I e^{-j\alpha} e^{-j\omega t}}{2}$$

$$v = \sqrt{2} \frac{\omega L I e^{j(\alpha+\pi/2)} e^{j\omega t} + \omega L I e^{-j(\alpha+\pi/2)} e^{-j\omega t}}{2}$$

However, if we define the phasor $\bar{I} = I e^{j\alpha}$, and recognize that $e^{j\pi/2} = j$ and $e^{-j\pi/2} = -j$, these become

$$i = \sqrt{2} \frac{\bar{I} e^{j\omega t} + \bar{I}^* e^{-j\omega t}}{2} \quad (5-21)$$

$$v = \sqrt{2} \frac{j\omega L \bar{I} e^{j\omega t} - j\omega L \bar{I}^* e^{-j\omega t}}{2}$$

If we now define

$$\bar{V} = j\omega L \bar{I} \quad (5-22)$$

the expression for v can be rewritten as

$$v = \sqrt{2} \frac{\bar{V} e^{j\omega t} + \bar{V}^* e^{-j\omega t}}{2} \quad (5-23)$$

Equation (5-22) is important, for it gives the relationship between phasors for voltage and current in an inductor. It can be used in the form shown to find the parameters of the voltage wave if the current wave is given. If the voltage wave is specified, then $\bar{V}$ is known and $\bar{I}$ is found from the same relation, rewritten as

$$\bar{I} = \frac{\bar{V}}{j\omega L} = -j \frac{\bar{V}}{\omega L} \quad (5-24)$$

A similar analysis applies to the capacitor shown in Fig. 5-5b. The voltage-current relationship is

$$i = C \frac{dv_C}{dt}$$

Suppose the voltage v_C is given by

$$v_C = \sqrt{2} V_C \cos(\omega t + \beta) + V_{dc} \quad (5-25)$$

where V_{dc} represents a constant (d-c) component. Differentiating and multiplying by C gives

$$i = -\sqrt{2} \omega C V_C \sin(\omega t + \beta)$$

or

$$i = \sqrt{2} \omega C V_C \cos(\omega t + \beta + \frac{\pi}{2}) \quad (5-26)$$

Just as it is possible for an inductor current to have a d-c component, so also there can be a d-c component of capacitor voltage. Steady state analysis for sinusoidal waves pertains to the relationship between the sinusoidal component of capacitor voltage and current; the d-c component of capacitor voltage can be found independently. Equations (5-25) and (5-26) show that V_{dc} has no effect on the relationship between the sinusoidal waves. Subsequently in this chapter, any symbol for capacitor voltage will be understood to refer to the sinusoidal component.

In complex number notation,

$$v_C = \sqrt{2} \frac{V_C e^{j\beta} e^{j\omega t} + V_C e^{-j\beta} e^{-j\omega t}}{2}$$

$$i = \sqrt{2} \frac{\omega C V_C e^{j(\beta+\pi/2)} e^{j\omega t} + \omega C V_C e^{-j(\beta+\pi/2)} e^{-j\omega t}}{2}$$

By defining the phasor $\bar{V}_C = V_C e^{j\beta}$ for the voltage, we can write these

$$v_C = 2 \frac{\bar{V}_C e^{j\omega t} + \bar{V}_C^* e^{-j\omega t}}{2} \quad (5-27)$$

$$i = \sqrt{2} \frac{\bar{I} e^{j\omega t} + \bar{I}^* e^{-j\omega t}}{2} \quad (5-28)$$

5-16

where

$$\bar{I} = j\omega C \bar{V}_C \quad (5-29)$$

If the voltage wave is specified, $\bar{V}_C$ is known and Eq. (5-29) yields $\bar{I}$ and the current wave. If the current wave is specified, the voltage wave can be found from $\bar{V}_C$, where

$$\bar{V}_C = \frac{\bar{I}}{j\omega C} = -j \frac{\bar{I}}{\omega C} \quad (5-30)$$

The resistor case shown in Fig. 5-5c is quite simple, since no derivative is involved. If

$$i = \sqrt{2} I \cos(\omega t + \alpha)$$

then

$$v_R = \sqrt{2} IR \cos(\omega t + \alpha)$$

The respective complex number representations are

$$i = \sqrt{2} \frac{\bar{I} e^{j\omega t} + \bar{I}^* e^{-j\omega t}}{2}$$

and

$$v_R = \sqrt{2} \frac{R \bar{I} e^{j\omega t} + (R \bar{I})^* e^{-j\omega t}}{2}$$

Thus, if $\bar{I}$ and $\bar{V}_R$ are respectively the phasor symbols for i and v_R , then

$$\bar{V}_R = R \bar{I} \quad (5-31)$$

5-4 Phase Difference

In the previous section we found that, except for a resistor, current and voltage waves in a network element have different initial angles. To summarize, if in each case

$$i = \sqrt{2} I \cos(\omega t + \alpha) \quad (5-32)$$

then

$$v_R = \sqrt{2} V_R \cos(\omega t + \alpha) \quad (a)$$

$$v_L = \sqrt{2} V_L \cos(\omega t + \alpha + \frac{\pi}{2}) \quad (b) \quad (5-33)$$

$$v_C = \sqrt{2} V_C \cos(\omega t + \alpha - \frac{\pi}{2}) \quad (c)$$

The corresponding waves are shown in Fig. 5-6.

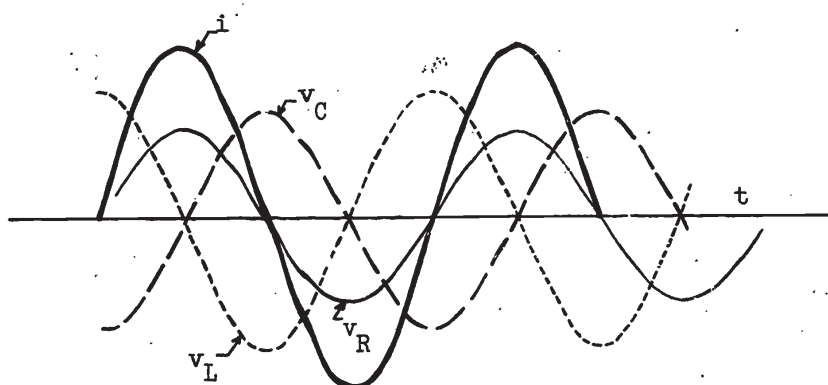


Figure 5-6.

When two waves like i and v_R have the same initial angles, the waves are said to be in phase. If the initial angles are different, the peak values of the wave occur at different instants of time. For example, in Fig. 5-6 wave v_L reaches a peak value sooner than the i wave. It is said that v_L leads i . The amount of separation between peak values of two waves is measured as an angle (on the ωt scale), and this quantity is called the phase difference. In the case of v_L and i , the phase difference is $\pi/2$. Thus, a complete statement of the phase relationship is to say that v_L leads i by $\pi/2$ radians.

If we compare v_C with i , we find that v_C lags i (or i leads v_C) meaning the peak of v_C occurs later in time than the peak of i . In fact, v_C lags i by $\pi/2$ radians. When specifying phase difference, in addition to specifying the angle, it is necessary to state whether it is an angle of lead or lag. The possibility of specifying a phase difference either way causes some redundancy. For example, in the capacitor case, we can replace the statement that v_C lags i by $\pi/2$ radians by the equivalent statement that i leads v_C by $(-\pi/2)$ radians. Phase difference

- 1 is an algebraic quantity for which the words "lead" or "lag" provide reference
 2 direction information. Since two reference directions are possible, there are
 3 two ways of specifying the phase difference.

Another kind of redundancy enters if angles greater than π are allowed.
 2 Thus, referring to Fig. 5-6, although the maximum labeled (2) of v_C occurs after
 3 maximum (1) of i , it is also true that maximum (3) of v_C occurs earlier than
 4 maximum (1). Thus, in addition to saying that v_C lags i by $\pi/2$ radians, it is
 5 correct to say that v_C leads i by $3\pi/2$ radians. This is merely an illustration
 6 of the fact that the phase difference can be increased or decreased by an
 7 integral multiple of 2π ; this principle being due to the fact that maxima occur
 8 at 2π intervals in the ωt scale.

Phase differences can be obtained by calculating the difference between
 4 initial angles. To state this in such a way as to get the proper sign, if we
 5 wish to determine the angle by which one wave (designated as the first wave)
 6 leads another wave (designated as the second wave), the initial angle of the
 7 second wave should be subtracted from the initial angle of the first wave. For
 8 example, to find the angle by which v_L leads i , observe that their initial angles
 9 are respectively $\alpha + \pi/2$ and α . Thus, the angle by which v_L leads i is
 $\alpha + \pi/2 - \alpha = \pi/2$.

Phase difference, as well as magnitude relationships, can conveniently be
 6 shown on phasor diagrams, as shown in Fig. 5-7. These are graphical portrayals

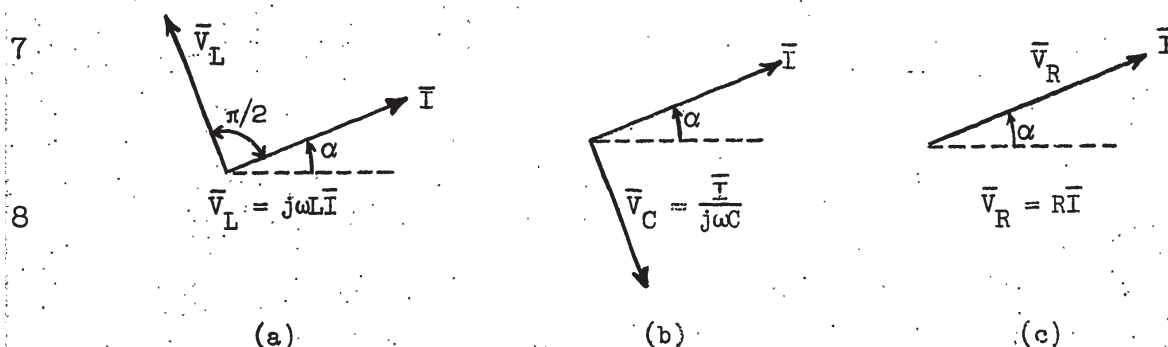


Figure 5-7.

of the three relations

$$\bar{V}_L = j\omega L \bar{I}$$

$$\bar{V}_C = \frac{\bar{I}}{j\omega C}$$

$$\bar{V}_R = R\bar{I}$$

where $\bar{I} = Ie^{j\alpha}$. We recall that v_L leads i by $\pi/2$ radians. This phase difference is evident as the angle between phasors $\bar{V}_L$ and $\bar{I}$, as indicated in Fig. 5-7a, and it is also evident that the phasor for the quantity which leads (v_L in this case) occupies a counter-clockwise position relative to the other one. This diagram also shows how it is possible to say that i leads v_L by $3\pi/2$, this being the counter-clockwise measured angle from $\bar{V}_L$ to $\bar{I}$.

Phase difference can be stated in terms of phasors. Thus, in addition to saying v_L leads i by $\pi/2$, it is also proper to say $\bar{V}_L$ leads $\bar{I}$ by $\pi/2$ radians. However, this latter statement is symbolic for the former.

5-5 Kirchhoff's Laws in Terms of Phasors

Consider a series circuit as shown in Fig. 5-8a, in which each box represents R, L or C, or a combination of these. The voltage across each box is sinusoidal and of the same frequency. That is, assume

$$v_1 = \sqrt{2} V_1 \cos(\omega t + \beta_1) \quad \text{and} \quad v_2 = \sqrt{2} V_2 \cos(\omega t + \beta_2)$$

In terms of the phasors

$$\bar{V}_1 = V_1 e^{j\beta_1} \quad \text{and} \quad \bar{V}_2 = V_2 e^{j\beta_2}$$

these voltages are

$$v_1 = \sqrt{2} \frac{\bar{V}_1 e^{j\omega t} + \bar{V}_1^* e^{-j\omega t}}{2} \quad \text{and} \quad v_2 = \sqrt{2} \frac{\bar{V}_2 e^{j\omega t} + \bar{V}_2^* e^{-j\omega t}}{2}$$

The Kvl equation for this figure is

$$v_T = v_1 + v_2 \quad (5-34)$$

- 1 Substituting the expressions for v_1 and v_2 and collecting terms that multiply $e^{j\omega t}$ and $e^{-j\omega t}$, gives

$$v_T = \sqrt{2} \frac{(\bar{V}_1 + \bar{V}_2)e^{j\omega t} + (\bar{V}_1 + \bar{V}_2)^* e^{-j\omega t}}{2} \quad (5-35)^+$$

- 2 The form of this shows that v_T is a sinusoidal wave of angular frequency ω , symbolically represented by

$$\bar{V}_T = \bar{V}_1 + \bar{V}_2 \quad (5-36)$$

- 3 In other words, if we define the phasor

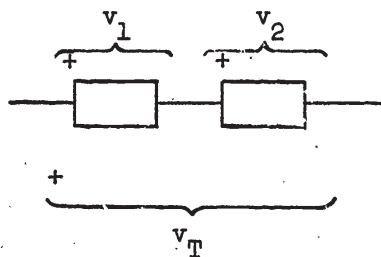
$$\bar{V}_T = V_T e^{j\beta_T}$$

- 4 the equation for the wave of v_T can also be written

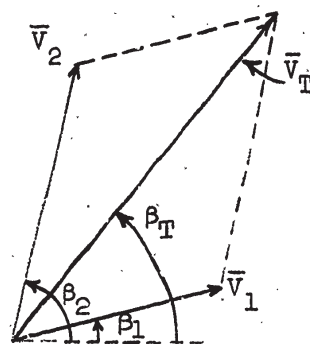
$$v_T = \sqrt{2} V_T \cos(\omega t + \beta_T)$$

5

6



(a)



(b)

Figure 5-8.

8

9

⁺In writing this equation, we have used the relation $\bar{V}_1^* + \bar{V}_2^* = (\bar{V}_1 + \bar{V}_2)^*$, which can readily be seen to be true by writing each phasor in rectangular form.

Equation(5-36) is an expression of the Kvl in terms of phasors. Of course, it is valid only when all quantities are sinusoidal. However, Eq. (5-35) shows that if v_1 and v_2 are each sinusoidal and of the same frequency, then v_T is also sinusoidal and of the same frequency. Equation (5-36) provides a means of finding V_T and β_T , which are the essential attributes of wave v_T . This can be done analytically, or graphically, as shown in Fig. 5-8b.

The Kirchhoff current law also can be written in terms of phasor quantities. Referring to Fig. 5-9a, if the currents i_1 and i_2 are sinusoidal and of the same frequency, namely,

$$i_1 = \sqrt{2} I_1 \cos(\omega t + \alpha_1) \quad \text{and} \quad i_2 = \sqrt{2} I_2 \cos(\omega t + \alpha_2)$$

they can be combined according to the Kcl equation

$$i_T = i_1 + i_2 \quad (5-37)$$

Again we shall perform the addition in terms of the complex notation, using the phasors $\bar{I}_1 = I_1 e^{j\alpha_1}$ and $\bar{I}_2 = I_2 e^{j\alpha_2}$ so that

$$i_1 = \sqrt{2} \frac{\bar{I}_1 e^{j\omega t} + \bar{I}_1^* e^{-j\omega t}}{2} \quad \text{and} \quad i_2 = \sqrt{2} \frac{\bar{I}_2 e^{j\omega t} + \bar{I}_2^* e^{-j\omega t}}{2}$$

The sum of these, after collecting terms, is

$$i_T = \sqrt{2} \frac{(\bar{I}_1 + \bar{I}_2) e^{j\omega t} + (\bar{I}_1 + \bar{I}_2)^* e^{-j\omega t}}{2} \quad (5-38)$$

which shows that i_T is a sinusoidal wave of angular frequency ω , and symbolically represented by a phasor $\bar{I}_T = I_T e^{j\alpha_T}$, where

$$\bar{I}_T = \bar{I}_1 + \bar{I}_2 \quad (5-39)$$

In trigonometric form,

$$i_T = \sqrt{2} I_T \cos(\omega t + \alpha_T)$$

where I_T and α_T are obtainable from phasor $\bar{I}_T$ as indicated above.

Equation (5-39) is a valid expression for the Kcl for situations where i_1 and i_2 are both sinusoidal waves of the same frequency. We have shown that the sum of

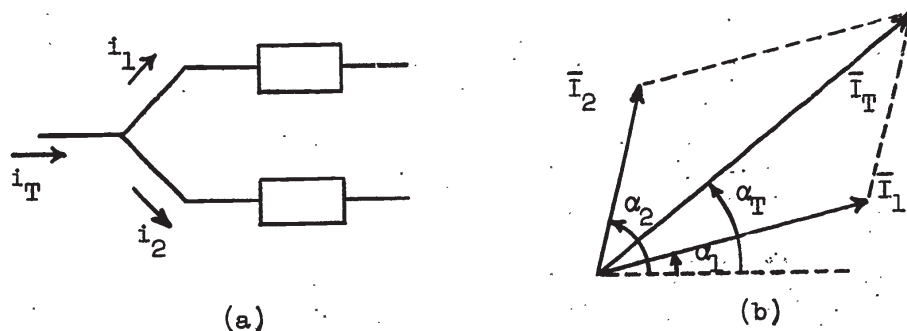


Figure 5-9.

two such sinusoidal current waves is also sinusoidal, and that the amplitude and initial angle of this wave can be obtained from Eq. (5-39). This equation can be used analytically; or a graphical interpretation like Fig. 5-9b can sometimes be useful.

The proofs just given, involving the addition of two voltages or two currents, can be extended to yield Kvl and Kcl equations in phasor form for any number of voltages or currents, subject, of course, to the condition that all voltages and currents are sinusoids of the same frequency. Referring to Fig. 5-10a, one form of the Kvl equation is

$$-v_4 = v_1 + v_2 + v_3 \quad (5-40)$$

It is assumed that v_1 , v_2 , and v_3 are each sinusoidal and of the same frequency, and that they are respectively represented by phasors $\bar{v}_1$, $\bar{v}_2$, and $\bar{v}_3$. By the previous proof the sum of two of them, say $v_1 + v_2$, is also sinusoidal, and so these can be placed in parentheses, as a reminder that they represent one sinusoidal wave, thus,

$$-v_4 = (v_1 + v_2) + v_3$$

But this is now reduced to the previous case, involving the addition of two sinusoidal waves $(v_1 + v_2)$ and v_3 . Following this plan, we first apply Eq. (5-36) to show that the phasor for $(v_1 + v_2)$ is $(\bar{v}_1 + \bar{v}_2)$. The same equation applies again to show that the phasor for $(v_1 + v_2) + v_3$ is $(\bar{v}_1 + \bar{v}_2) + \bar{v}_3$. Thus,

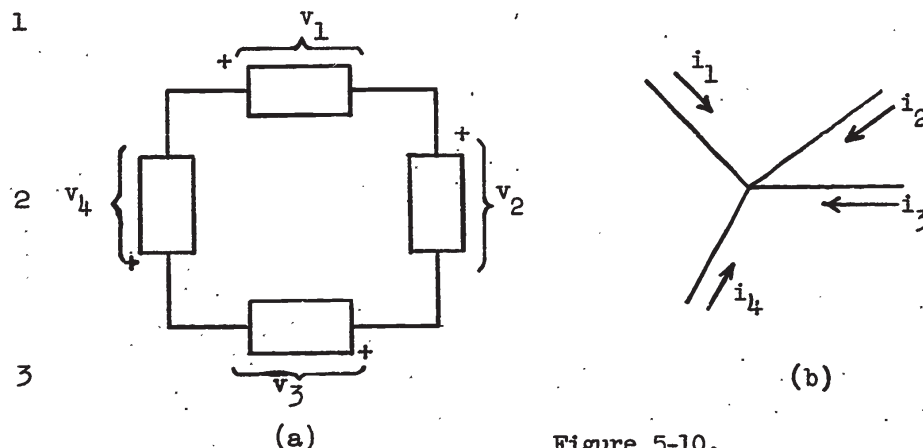


Figure 5-10.

from Eq. (5-40) we get

$$-\bar{v}_4 = (\bar{v}_1 + \bar{v}_2) + \bar{v}_3$$

or

$$0 = \bar{v}_1 + \bar{v}_2 + \bar{v}_3 + \bar{v}_4 \quad (5-41)$$

A similar development will apply to the Kcl equations when currents are sinusoidal and of the same frequency. Referring to Fig. 5-10b, we have

$$\begin{aligned} -i_4 &= i_1 + i_2 + i_3 \\ &= (i_1 + i_2) + i_3 \end{aligned} \quad (5-42)$$

If i_1 , i_2 , i_3 , and i_4 have the respective phasors $\bar{i}_1$, $\bar{i}_2$, $\bar{i}_3$, and $\bar{i}_4$, Eq. (5-39) applies to show that $(i_1 + i_2)$ has the phasor $\bar{i}_1 + \bar{i}_2$, and applies again to show that $(i_1 + i_2) + i_3$ has the phasor $(\bar{i}_1 + \bar{i}_2) + \bar{i}_3$. Thus, Eq. (5-41) yields

$$-\bar{i}_4 = (\bar{i}_1 + \bar{i}_2) + \bar{i}_3$$

or

$$0 = \bar{i}_1 + \bar{i}_2 + \bar{i}_3 + \bar{i}_4 \quad (5-43)$$

as the phasor expression of the Kcl equation.

The interpretation to be placed on the results of this section, culminating

5-24

in Eqs. (5-41) and (5-43), is that Kvl and Kcl equations can be written in terms of phasor symbols in exactly the same way as for actual time-varying quantities. The arguments given for summing three sinusoidal quantities can be extended to any number. This statement is, of course, subject to the condition that all quantities involved shall be sinusoidally varying with time, and shall have the same frequency.

5-6 Analysis of Elementary Circuits

In the previous section it was shown that phasor quantities can be used to write Kirchhoff's laws, when waves are sinusoidal, and we have also established phasor relationships for R, L, and C elements. The next step is to use these principles to analyze some simple circuits.

As a first example, consider Fig. 5-11a, for which we assume that the current i is sinusoidal and represented by the phasor $\bar{I}$. From Eqs. (5-23), (5-30) and (5-31) we have, respectively, $\bar{V}_L = j\omega L \bar{I}$, $\bar{V}_C = \bar{I}/j\omega C$, and $\bar{V}_R = R \bar{I}$. Also, by virtue of the discussion in Sec. 5-5, Kvl can be applied in terms of phasors, giving

$$\bar{V} = \bar{V}_R + \bar{V}_L + \bar{V}_C \quad (5-44)$$

or

$$\bar{V} = (R + j\omega L + \frac{1}{j\omega C}) \bar{I} = [R + j(\omega L - \frac{1}{\omega C})] \bar{I} \quad (5-45)$$

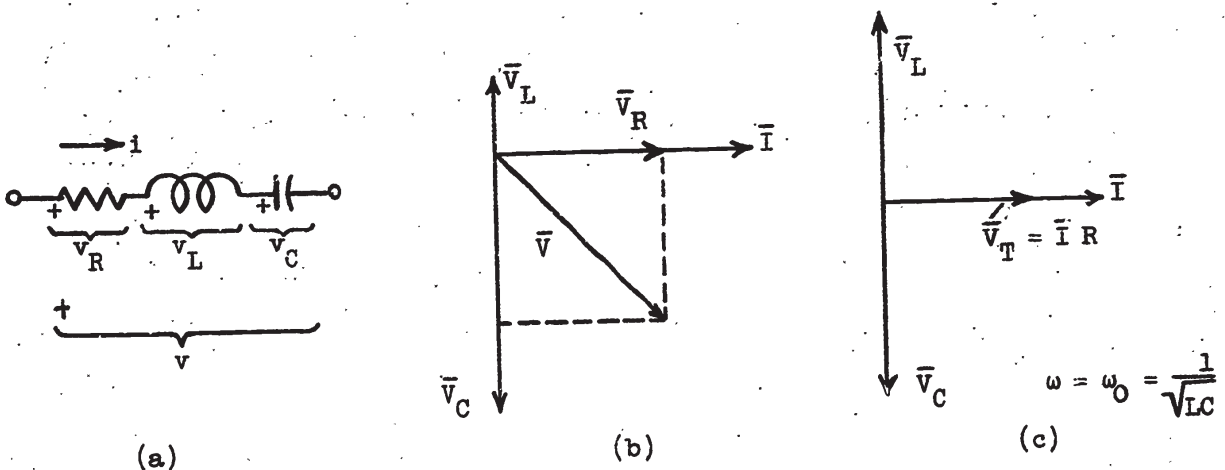


Figure 5-11.

The three phasors on the right of Eq. (5-44), and also their sum, are shown in Fig. 5-11b, for a particular value of ω . For this case, the terminal voltage lags the current. However, it can be seen that for a sufficiently large value of ω , ωL will be larger than $1/\omega C$, so that $|\bar{V}_L|$ will be larger than $|\bar{V}_C|$. Then $\bar{V}$ will lie in the first quadrant, and will lead the current. The transition from voltage lag to voltage lead will occur when $\omega L = 1/\omega C$. This particular value of ω , which we shall designate as ω_0 , given by

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (5-46)$$

is called the resonant angular frequency. The phasor diagram for this frequency is shown in Fig. 5-11c. In this case the phase difference between voltage and current is zero (the voltage and current are said to be in-phase).

This simple example has served to bring out several important aspects of alternating-current circuits: the application of complex numbers, resulting in Eq. (5-45) as a relationship between terminal voltage and current; the equivalent graphical analysis as displayed in the phasor diagram of Fig. 5-11b; and the notion that the relationship between current and voltage generally is dependent upon frequency (or ω), including the idea of a resonant frequency at which a circuit behaves like a pure resistance.

A similar analysis can be carried out for the parallel branch illustrated in Fig. 5-12a. Again referring to Eqs. (5-23), (5-30), and (5-31), we assume a

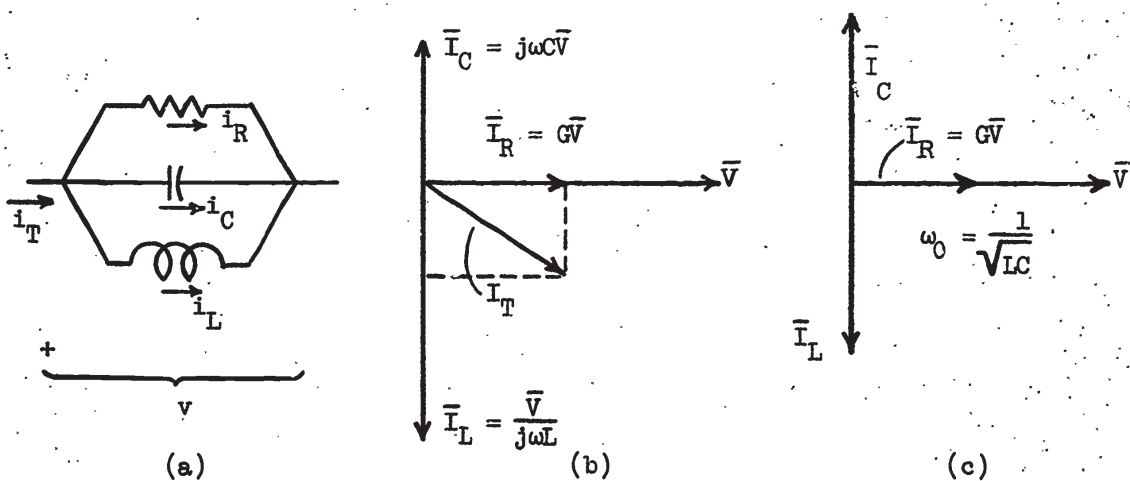


Figure 5-12.

5-26

sinusoidal voltage v , and have $\bar{I}_L = \bar{V}/j\omega L$, $\bar{I}_C = j\omega C\bar{V}$, and $\bar{I}_R = \bar{V}/R$. Also, from Kcl in phasor form, we have

$$\bar{I} = \bar{I}_R + \bar{I}_C + \bar{I}_L \quad (5-47)$$

or

$$\bar{I} = \left(\frac{1}{R} + j\omega C + \frac{1}{j\omega L} \right) \bar{V} = \left[\frac{1}{R} + j\left(\omega C - \frac{1}{\omega L} \right) \right] \bar{V} \quad (5-48)$$

Equation (5-48) is quite similar to Eq. (5-45), except that in the present case a voltage phasor is multiplied by a complex number to obtain the current phasor. A phasor diagram can be constructed in accordance with Eq. (5-47), as shown in Fig. 5-12b. Whether $\bar{I}$ lags or leads $\bar{V}$ depends on the frequency. In similarity with the series circuit, there is a resonant angular frequency, also given by Eq. (5-46), at which the voltage and current are in-phase.

5-7 Impedance and Admittance

Equations (5-45) and (5-48) apply to special cases of passive two-terminal networks. In general, these are interconnections of R, L and C elements, of any complexity that provides one or more current paths between two terminals. The word "passive" means that no sources can be included in any of the paths. The next step is to generalize to cases of greater complexity, such as the examples shown in Fig. 5-13.

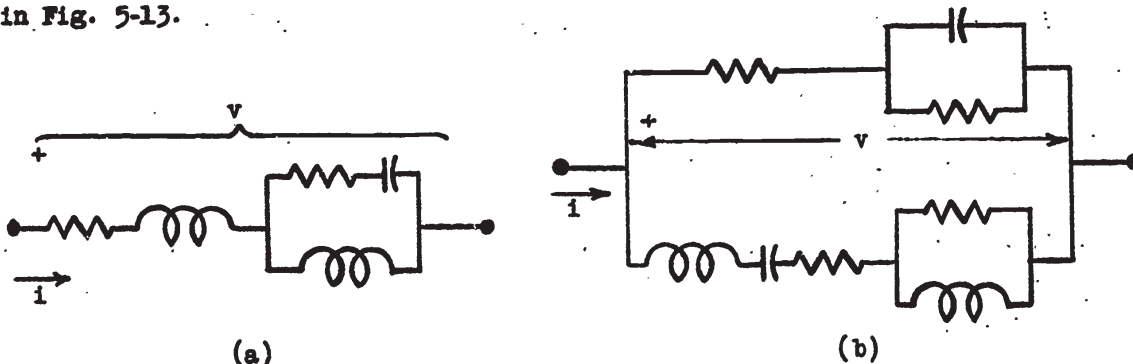


Figure 5-13.

First we note that for Fig. 5-11a we had the result

$$\bar{V} = \left[R + j\left(\omega L - \frac{1}{\omega C} \right) \right] \bar{I} \quad (a) \quad (5-49)$$

which can also be written

$$\bar{I} = \left[\frac{1}{R + j(\omega L - \frac{1}{\omega C})} \right] \bar{V} \quad (b) \quad (5-49)$$

and for Fig. 5-12a the similar results are

$$\bar{I} = \left[\frac{1}{R} + j(\omega C - \frac{1}{\omega L}) \right] \bar{V} \quad (a)$$

which can also be written

$$\bar{V} = \left[\frac{1}{\frac{1}{R} + j(\omega C - \frac{1}{\omega L})} \right] \bar{I} \quad (b) \quad (5-50)$$

In each case, one of these equations is superfluous (superfluous in the same sense that $v = Ri$ and $i = v/R$ are equivalent), but it is customary to adopt a notation and nomenclature which admits both forms. We shall now describe such a general notation. The above equations are examples of the following two forms

$$\bar{V} = \bar{Z}\bar{I} \quad (5-51)$$

$$\bar{I} = \bar{Y}\bar{V} \quad (5-52)$$

in which $\bar{Z}$ and $\bar{Y}$ are given by the appropriate bracketed quantities. $\bar{Z}$ is called the impedance of the circuit, and $\bar{Y}$ the admittance. In view of Eqs. (5-51) and (5-52), it is obvious that

$$\bar{Y} = \frac{1}{\bar{Z}} \quad (5-53)$$

so that if one of them is known, the other can easily be found. Although Eqs. (5-51) and (5-52) imply definitions for $\bar{Z}$ and $\bar{Y}$, it is worthwhile to define impedance and admittance explicitly as

$$\bar{Z} = \frac{\bar{V}}{\bar{I}} \quad (a) \quad \text{and} \quad \bar{Y} = \frac{\bar{I}}{\bar{V}} \quad (b) \quad (5-54)$$

In these definitions, and all the preceding formulas, it is understood that $\bar{V}$ is the phasor for the voltage across the terminals of the network in question,

and that $\bar{I}$ is the phasor for the current in the terminals, with reference directions as portrayed in Fig. 5-13.

Since $\bar{Z}$ and $\bar{Y}$ are complex numbers, each can be written in rectangular form as

$$\bar{Z} = R + jX \quad (5-55)$$

$$\bar{Y} = G + jB \quad (5-56)$$

R and G carry the same names as in resistance networks, namely resistance and conductance, respectively. X is called the reactance, and B the susceptance. It should be noted that the similarity of the names of R and G with those used in resistive circuits should not be construed to mean that R (or G) refers to a single resistor. To take R as an example, if we consider the impedance of Fig. 5-11a, we find R of Eq. (5-55) is identical with the resistance R of the circuit. However, the impedance of Fig. 5-12a is

$$\bar{Z} = \frac{1}{\frac{1}{R} + j(\omega C - \frac{1}{\omega L})} = \frac{\frac{1}{R} - j(\omega C - \frac{1}{\omega L})}{(\frac{1}{R})^2 + (\omega C - \frac{1}{\omega L})^2}$$

In this case the R of Eq. (5-55) is

$$\frac{\frac{1}{R}}{(\frac{1}{R})^2 + (\omega C - \frac{1}{\omega L})^2}$$

which is not identifiable as a resistive circuit element of the original network. In Eqs. (5-55) and (5-56), R and G are to be regarded respectively as the real parts of $\bar{Z}$ and $\bar{Y}$, rather than as circuit elements.

We turn now to the question of how to find $\bar{Z}$ and $\bar{Y}$ for circuits of more general nature, such as the examples of Fig. 5-13. We shall do so by making note of similarities with resistance networks. It will be recalled that Kirchhoff's laws are used to prove the following relationships for the resistive networks of Fig. 5-14.

$$\left. \begin{aligned} R_{eq} &= R_1 + R_2 \\ G_{eq} &= \frac{G_1 G_2}{G_1 + G_2} \end{aligned} \right\} \quad \text{for Fig. 5-14a}$$

$$G_{eq} = G_1 + G_2$$

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2}$$

for Fig. 5-14b

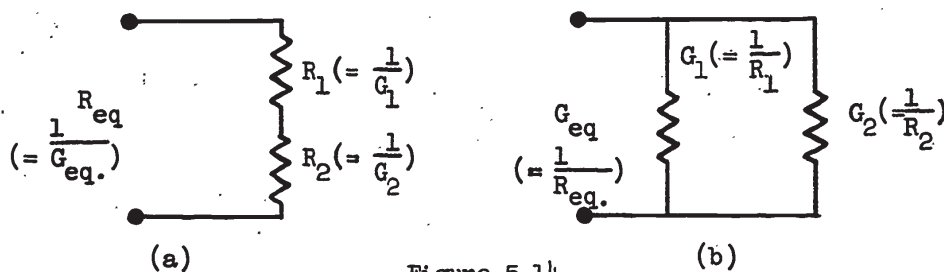


Figure 5-14.

These are the basic laws for combining resistances (or conductances) in series and parallel, which can be extended to more complicated cases merely by repetition.

By using Kirchhoff's laws in phasor notation, in proofs that are otherwise identical, it can be shown that impedances (or admittances) combine in similar ways. Referring to Fig. 5-15, the results are as follows:

$$\bar{Z}_{eq} = \bar{Z}_1 + \bar{Z}_2$$

$$\bar{Y}_{eq} = \frac{\bar{Y}_1 \bar{Y}_2}{\bar{Y}_1 + \bar{Y}_2}$$

for Fig. 5-14a (5-57)

$$\bar{Y}_{eq} = \bar{Y}_1 + \bar{Y}_2$$

$$\bar{Z}_{eq} = \frac{\bar{Z}_1 \bar{Z}_2}{\bar{Z}_1 + \bar{Z}_2}$$

for Fig. 5-14b (5-58)

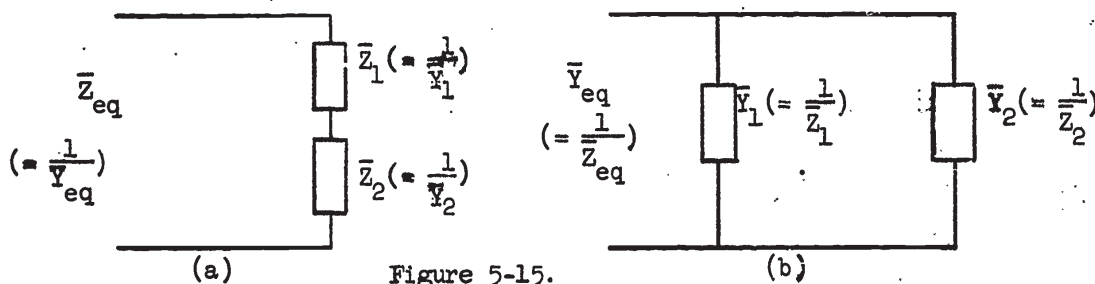


Figure 5-15.

These principles of combination can be used any number of times, starting with "rectangles" that represent branches that are simple enough to have known impedances or admittances (like Fig. 5-11a and Fig. 5-12a, for example, or special cases of these).

Although the procedure for combining impedances and admittances is algebraically identical with that for resistances (and conductances) the fact that complex numbers are involved in the present case means that numerical calculations are more laborious. Also, the translation from wave quantities (actual v and i) to phasor quantities and back must be included, a step which is not necessary for resistance networks. To illustrate these observations, the following two numerical examples are included:

Example 1: Referring to Fig. 5-16, assuming a terminal voltage $v_T = \sqrt{2} (75) \cos(3000t)$, the current i_T is to be found. The following numerical values will be used:

$$R_1 = 120 \text{ ohms}$$

$$R_2 = 2000 \text{ ohms } (G_2 = 5 \times 10^{-4} \text{ mho})$$

$$L = 0.5 \text{ henry}$$

$$C = 0.2 \times 10^{-6} \text{ farad}$$

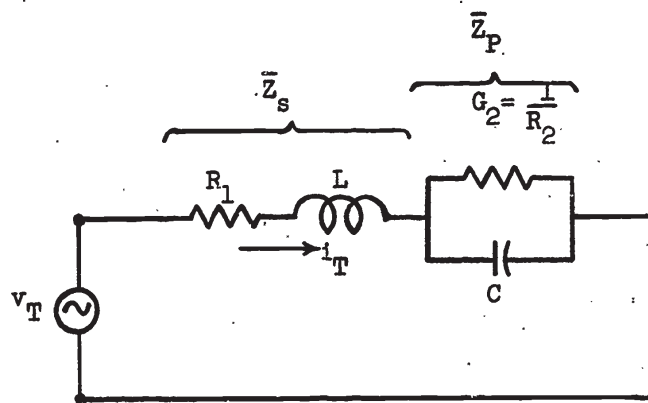


Figure 5-16.

Solution

Observe that the path between the source terminals can be viewed as two impedances (labeled $\bar{Z}_s$ and $\bar{Z}_p$) in series. Accordingly, these two impedances can be added, and the current I can be found from $\bar{I} = \bar{V}/\bar{Z}$. Furthermore, the

1 portion labeled $\bar{Z}_s$ is a special case of Fig. 5-11a (with C replaced by a short circuit, meaning infinite C in the formula), and the portion labeled $\bar{Z}_p$ is a special case of Fig. 5-12a in which L is absent (infinite). Thus, we can proceed as follows:

2 Beginning with the parallel combination of G_2 and C, we note that it has an admittance

$$3 \quad \frac{1}{\bar{Z}_p} = (5 + j6) \times 10^{-4}$$

where we have used $\omega C = 3000(0.2) \times 10^{-6} = 6 \times 10^{-4}$. The impedance of this parallel combination is therefore

$$4 \quad \bar{Z}_p = \frac{10^4}{5+j6} = \frac{10^4}{7.80/50.2^\circ} = 1282/-50.2^\circ$$

In rectangular form this is⁺

$$5 \quad \bar{Z}_p = 1282(\cos 50.2^\circ + j \sin 50.2^\circ)$$

$$\approx 1282(.641 + j.768)$$

$$6 \quad = 820 - j 983$$

The other portion of the circuit has an impedance

$$7 \quad \bar{Z}_s = 120 + j 1500$$

⁺The rectangular form can be found directly, without going to the polar form, thus

$$8 \quad \frac{10^4}{5+j6} = \frac{(5-j6)10^4}{(5+j6)(5-j6)} = \frac{(5-j6)10^4}{25+36} = \frac{5 \times 10^4}{61} - j \frac{6 \times 10^4}{61} = 820 - j 983$$

9 However, in most cases, where the sum of the squares of the real and imaginary parts of the denominator cannot be found as easily as in this example, converting to polar coordinates is the simpler procedure. Many modern slide rules are constructed as to facilitate conversions between rectangular and polar coordinates.

5-32

where $\omega L = 3000(0.5) = 1500$. The total impedance is

$$\begin{aligned}\bar{Z}_T &= \bar{Z}_p + \bar{Z}_s \\ &= 940 + j 517 = 1072 \angle 28.8^\circ\end{aligned}$$

The voltage phasor is $\bar{V} = 75 + j0$, and so, having found $\bar{Z}$, the current phasor is

$$\bar{I}_T = \frac{\bar{V}}{\bar{Z}_T} = \frac{75 + j0}{1072 \angle 28.8^\circ} = .0699 \angle -28.8^\circ$$

Finally, the equation for the current (not its phasor) is

$$i_T = \sqrt{2} (.0699) \cos(3000t - 28.8^\circ)$$

Observe that the voltage leads the current and that the total impedance Z_T has a positive angle. Voltage leads current whenever impedance has a positive angle. However, impedance can also have a negative angle, in which case the voltage will lag the current. In this circuit, by carrying out the same calculations for $\omega = 200$, it will be found that

$$\bar{Z}_T = 120 + j 100 + 1980 - j 159$$

$$= 3000 - j 59$$

which has a negative angle. In this case, the voltage will lag the current; by a small angle. [Note, in these statements, that voltage and current refer to voltage between the two terminals and current in one terminal (and out the other), using the customary reference markings for voltage and current]

Example 2: In the circuit of Fig. 5-17, assume that the current is known to be $i_T = \sqrt{2} (4) \cos \omega t$, where the frequency is 60 cps. Thus, $\omega = 2\pi(60) = 377$. The voltage v_T is required. The element values are

$$R_1 = 100 \text{ ohms } (G_1 = .01 \text{ mho})$$

$$R_2 = 150 \text{ ohms}$$

$$L_1 = 1.5 \text{ henry}$$

$$C_2 = 12 \times 10^{-6} \text{ farad}$$

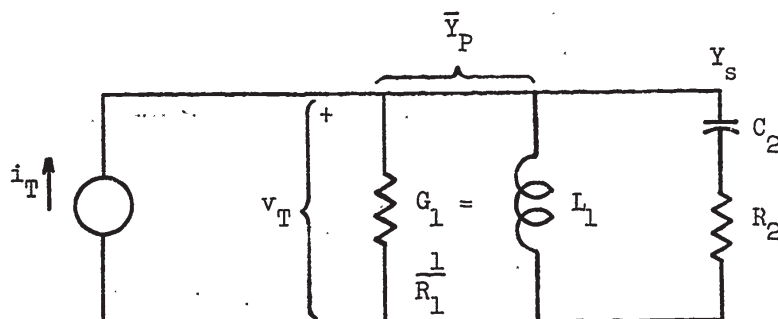


Figure 5-17.

Solution

This circuit has the general nature of a parallel connection, and can be viewed as having an equivalent admittance which is the sum of the two admittances $\bar{Y}_p$ and $\bar{Y}_s$. Thus, beginning with the R-C branch on the right, we write its impedance as a special case of Fig. 5-11a (with $L = 0$), giving

$$\begin{aligned}\bar{Z}_s &= 150 - j \frac{10^6}{(377)(12)} = 150 - j 221 \\ &= 268 \angle -55.8^\circ\end{aligned}$$

and

$$\bar{Y}_s = \frac{1}{\bar{Z}_s} = 3.74 \times 10^{-3} \angle 55.8^\circ = .0021 + j .0031$$

For the parallel combination of L and R_1 , the admittance is

$$\bar{Y}_p = .01 - j \frac{1}{(377)(0.5)} = .01 - j .00531$$

The total admittance is

$$\begin{aligned}\bar{Y}_T &= \bar{Y}_s + \bar{Y}_p = .0021 + .01 + j(.0031 - .00531) \\ &= .0121 - j .00221\end{aligned}$$

The phasor $\bar{V}_T$ (symbolic for v_T) is

5-34

$$\bar{V}_T = \frac{\bar{I}_T}{Y_T}$$

$$= \frac{4 + j0}{.0121 - j.00221} = \frac{4}{.01225/-10.4^\circ} = 326/-10.4^\circ$$

where $\bar{I}_T = 4 + j0$ is the phasor for the given current. Finally, the voltage wave is

$$v_T = \sqrt{2} (326) \cos(377t - 10.4^\circ)$$

5-8 Voltage and Current Source Equivalents

Following the general plan that is developing, whereby a-c circuits with R, L, and C elements can be treated like resistive circuits if phasor quantities are used, we now consider the equivalence of the two networks in Fig. 5-18.

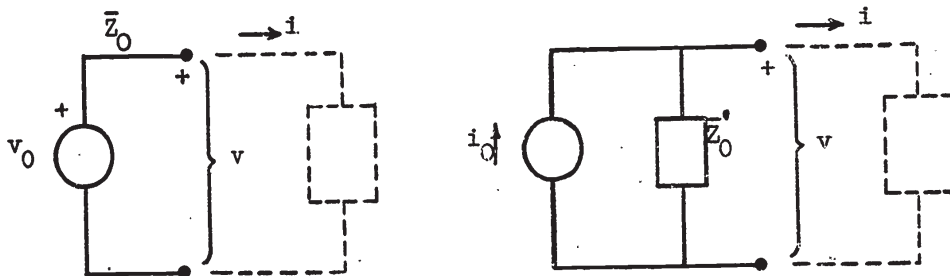


Figure 5-18.

The ideal voltage source v_0 maintains a sinusoidal voltage

$$v_0 = \sqrt{2} V_0 \cos(\omega t + \beta_0)$$

under all conditions, and the ideal current source maintains a sinusoidal current

$$i_0 = \sqrt{2} I_0 \cos(\omega t + \alpha_0)$$

under all conditions. These sources are respectively symbolically represented by phasors $\bar{V}_0 = V_0 e^{j\beta_0}$ and $\bar{I}_0 = I_0 e^{j\alpha_0}$. The dotted rectangles shown in each case represent identical two-terminal networks consisting of R, L, and C elements

and possibly sinusoidal sources, also of angular frequency ω . The presence of these external branches permits a current i to flow, and by restricting what can be in these branches we ensure that v and i will be sinusoidal (of angular frequency ω) and hence can respectively be represented by phasors $\bar{V}$ and $\bar{I}$.

Just as in the resistive network case, respectively for (a) and (b) of Fig. 5-18, we can write

$$\bar{V} = \bar{V}_0' - \bar{Z}_0' \bar{I} \quad (a) \quad (5-59)$$

$$\bar{I} = \bar{I}_0 - \frac{1}{\bar{Z}_0'} \bar{V} \quad (b)$$

The second of these can be written

$$\bar{V} = \bar{Z}_0' \bar{I}_0 - \bar{Z}_0' \bar{I} \quad (5-60)$$

In order for the two networks of Fig. 5-18 to be equivalent, they must be equivalent for all $\bar{V}$ and $\bar{I}$ (i.e., $\bar{V}$ and $\bar{I}$ must be the same for both). Thus, equating Eqs. (5-59) and (5-60), to ensure that $\bar{V}$ will be the same, we have

$$\bar{V}_0 - \bar{Z}_0 \bar{I} = \bar{Z}_0' \bar{I}_0 - \bar{Z}_0' \bar{I}$$

or

$$(\bar{V}_0 - \bar{Z}_0 \bar{I}_0) + (\bar{Z}_0' - \bar{Z}_0) \bar{I} = 0$$

This equation must be true for all values of $\bar{I}$, and so each term must be independently zero, giving

$$\bar{V}_0 = \bar{Z}_0 \bar{I}_0 \quad (5-61)$$

as the relationship between the sources, and

$$\bar{Z}_0' = \bar{Z}_0 \quad (5-62)$$

as the relationship between the impedances.

The equivalence between these two networks can be used to reduce circuit complexity in the same way as in resistance networks. However, it must be kept in mind that the present case is more restrictive in that the sources must be sinusoidal whereas in resistive networks any manner of time variation is permitted.

This is the "price we pay" for using phasor symbolism,[†] but many practical problems fall into this category.

There is one other important exception to the similarities between the solution of resistance networks and the steady state solution of RLC networks. In the case of resistance networks, it is possible to include ideal diodes, or piecewise linear equivalents, with due regard to the fact that the circuit changes whenever the diode current goes through zero. Phasors cannot be used for sinusoidal waves if diodes are included in a circuit, unless the diodes are biased sufficiently so that the current never reaches zero. Otherwise, if the diode goes from a conducting to a non-conducting condition during the swing of the signal, certain voltages and currents will become non-sinusoidal. This violates the conditions under which phasors can be used. Other methods of analysis must be used in such cases. There is no intent here to explain how to solve such cases, only to point out a limitation on the extension of resistance networks methods to sinusoidal steady state methods of analysis.

5-9 Thévenin and Norton Theorems

For any R, L, C network having only two terminals, such as the box on the left of Fig. 5-19, subject to the conditions that all sources within the box are of the same frequency, an equation of the form

$$\bar{V} = \bar{V}_0 - \bar{Z}_0 \bar{I} \quad (5-63)$$

will be found as the relationship between $\bar{V}$ and $\bar{I}$. In this case we do not find a source V_0 or an impedance $\bar{Z}_0$. These are quantities that might be derived from

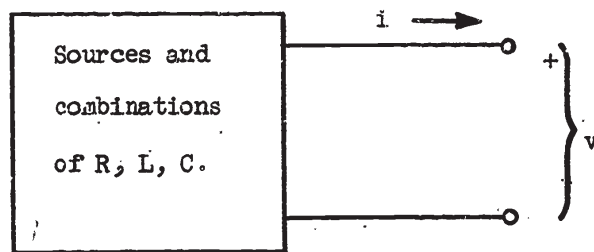


Figure 5-19.

[†]This is not to say that the general case of any type of source cannot be handled for R, L, C circuits, but to do this requires mathematical methods beyond those used in this text.

an analysis of the actual network (analysis by the solution of Kirchhoff equations, or perhaps by a succession of applications of the source equivalents). This means that $\bar{V}_0$ and $\bar{Z}_0$ are functions of the elements within the box, and how they are connected. A physical interpretation of them is possible, as follows: From Eq. (5-63), it is seen that $\bar{V} = \bar{V}_0$ when $\bar{I} = 0$. Thus, $\bar{V}_0$ is the open circuit value of $\bar{V}$. $\bar{V}_0$ is due to the internal sources, and will be zero if these sources are made zero. This fact provides an interpretation of $\bar{Z}_0$ as $-\bar{V}/\bar{I}$, when $\bar{V}_0 = 0$, (i.e., all internal sources are zero). Observe that $-\bar{I}$ is the current flowing into the + terminal, so that it is evident that $-\bar{V}/\bar{I}$ is the impedance of the network when all sources are zero.

This equivalence can be stated in the following form: A network having one pair of terminals and consisting of interconnections of R, L and C elements, and voltage and current sources, all of which are sinusoidal with the same frequency, is equivalent at the pair of terminals to a voltage source having phasor $\bar{V}_0$ in series with an impedance $\bar{Z}_0$. $\bar{V}_0$ is the value of the terminal voltage phasor when the terminals are open-circuited; and $\bar{Z}_0$ is the impedance at the terminals of the network when all the sources are deactivated.

Having already established the equivalence of circuits (a) and (b) of Fig. 5-18, it is evident that either of these circuits is equivalent to the original one, if $\bar{V}_0$, $\bar{Z}_0$, and $\bar{I}_0 = \bar{V}_0/\bar{Z}_0$ are interpreted as described above.

Figure 5-18a is called the Thévenin equivalent, and the theorem which states this equivalent is called Thévenin's Theorem. Also Fig. 5-18b is the Norton equivalent, and a statement of the equivalence is Norton's Theorem.

5-10 Methods of Network Analysis

Concerning steady state sinusoidal waves all of the same angular frequency ω , we can summarize the following:

- 1) Waves like $i = \sqrt{2} I \cos(\omega t + \alpha)$ and $v = \sqrt{2} V \cos(\omega t + \beta)$ can be treated symbolically in terms of the phasors $\bar{I} = I e^{j\alpha}$ and $\bar{V} = V e^{j\beta}$. These phasors do not equal i and v , however.
- 2) Kirchhoff's voltage and current equations can be written in terms of phasor symbols, to obtain the symbolic equivalent of writing the corresponding equations for the sinusoidal waves.
- 3) For individual R, L, or C elements, the voltage and current phasors are related by

$$\bar{V} = R \bar{I}$$

$$\bar{V} = j\omega L \bar{I}$$

$$\bar{V} = \frac{1}{j\omega C} \bar{I}$$

- 4) For any two-terminal network (branch) of R, L, and C elements the relationship between $\bar{V}$ and $\bar{I}$ can be written either as $\bar{V} = \bar{Z}\bar{I}$ or $\bar{I} = \bar{Y}\bar{V}$, and $\bar{Z}$ and $\bar{Y}$ can be found by methods similar to the series and parallel combination of resistances in resistance networks. Voltage and current divider principles also apply.
- 5) Thévenin and Norton theorems apply, using phasor quantities.

We shall use a practical example to illustrate further generalizations of methods first introduced in the chapter on resistance networks. In so doing, the important concepts of loop and node equations will be introduced. The fact that these are included only in an example should not be construed as a depreciation of their importance. They are not given a specific theoretical discussion because by now it seems that the close similarity with resistance networks should be accepted, so that it will be evident that resistance network principles will apply, with $\bar{V} = \bar{Z}\bar{I}$ replacing $v = Ri$, and $\bar{I} = \bar{Y}\bar{V}$ replacing $i = v/R$.

The example is the low-pass filter network shown in Fig. 5-20. Such a network gives preferential treatment to waves of low frequencies. This example will be

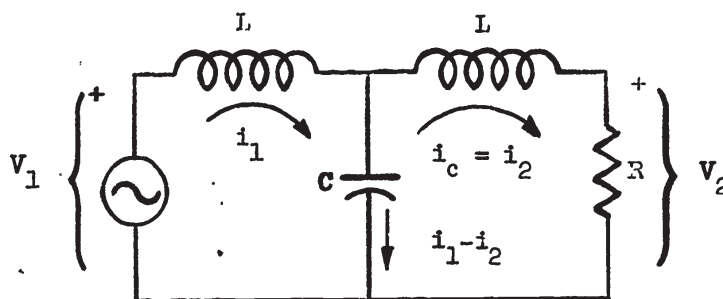


Figure 5-20.

used to illustrate various methods of determining the output voltage v_2 , when the input voltage v_1 is specified. The input voltage will be assumed to be

$$v_1 = \sqrt{2} V_1 \cos \omega t$$

so that its phasor is $V_1 e^{j0} = V_1 + j0$.

(a) Loop Analysis

Loop currents i_1 and i_2 (with phasors $\bar{I}_1$ and $\bar{I}_2$) will be used, so that the current i_c in the C element will be $i_1 - i_2$. Two Kvl equations will be obtained. For the left and right hand loops, respectively, these are

$$j\omega L \bar{I}_1 + \frac{1}{j\omega C} (\bar{I}_1 - \bar{I}_2) = \bar{V}_1 \quad (a) \quad (5-64)$$

$$-\frac{1}{j\omega C} (\bar{I}_1 - \bar{I}_2) + (R + j\omega L) \bar{I}_2 = 0 \quad (b)$$

The second of these equations can be solved for $\bar{I}_1$, to give

$$\begin{aligned} \bar{I}_1 &= j\omega C [R + j\omega L + \frac{1}{j\omega C}] \bar{I}_2 \\ &= (1 - \omega^2 LC + j\omega RC) \bar{I}_2 \end{aligned}$$

This can be substituted in Eq. (5-64a), giving

5-40

$$j(\omega L - \frac{1}{\omega C})(1 - \omega^2 LC + j\omega RC)\bar{I}_2 + \frac{j}{\omega C}\bar{I}_2 = \bar{V}_1$$

$$\left\{ \left(\frac{1}{\omega C} - \omega L \right)(\omega RC) + j \left[\left(\omega L - \frac{1}{\omega C} \right)(1 - \omega^2 LC) + \frac{1}{\omega C} \right] \right\} \bar{I}_2 = \bar{V}_1$$

$$[(1 - \omega^2 LC)R + j\omega L(2 - \omega^2 LC)] \bar{I}_2 = \bar{V}_1$$

and, finally, the voltage phasor ($\bar{V}_2 = R\bar{I}_2$) is

$$\bar{V}_2 = \left[\frac{1}{(1 - \omega^2 LC) + j \frac{\omega L}{R}(2 - \omega^2 LC)} \right] \bar{V}_1 \quad (5-65)$$

Before interpreting this result, we shall obtain Eq. (5-65) by two other methods.

(b) Node Analysis

The circuit under consideration is redrawn in Fig. 5-21, showing a node (3) in the center and corresponding voltage v_3 . First let us sum the currents entering node 3. This sum must be zero, and so we get

$$\frac{1}{j\omega L} (\bar{V}_1 - V_3) - j\omega C V_3 + \frac{1}{j\omega L} (\bar{V}_2 - V_3) = 0$$

or

$$\frac{1}{j\omega L} \bar{V}_1 + \frac{1}{j\omega L} \bar{V}_2 - \left(\frac{2}{j\omega L} + j\omega C \right) \bar{V}_3 = 0$$

Multiplying this through by $j\omega L$ gives

$$\bar{V}_1 + \bar{V}_2 - (2 - \omega^2 LC) \bar{V}_3 = 0 \quad (5-66)$$

This equation involves two unknowns ($\bar{V}_2$ and $\bar{V}_3$) since $\bar{V}_1$ is known. Thus, one other equation is needed. This equation is obtained from node 2. The current from (3) to (2) is

$$\frac{1}{j\omega L} (\bar{V}_3 - \bar{V}_2)$$

and the current leaving node 2, through R , is $\bar{V}_2/R$. We equate these, to give

$$\frac{\bar{V}_2}{R} = \frac{1}{j\omega L} (\bar{V}_3 - \bar{V}_2)$$

or

$$\bar{V}_3 = \left(\frac{j\omega L}{R} + 1 \right) \bar{V}_2 \quad (5-67)$$

Substituting this expression for $\bar{V}_3$ into Eq. (5-66) yields

$$\bar{V}_1 - \left[(2 - \omega^2 LC) \left(\frac{R + j\omega L}{R} \right) - 1 \right] \bar{V}_2 = 0$$

The quantity within the brackets simplifies to

$$(1 - \omega^2 LC) + \frac{j\omega L}{R} (2 - \omega^2 LC)$$

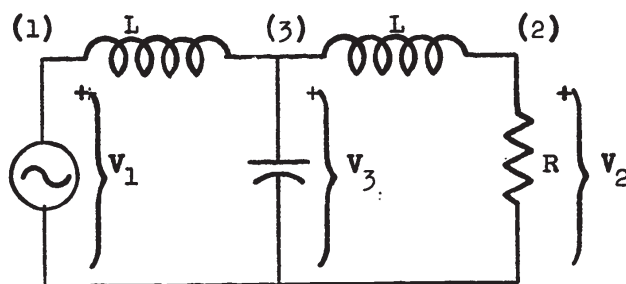


Figure 5-21.

and so the result is

$$\bar{V}_2 = \frac{\bar{V}_1}{(1 - \omega^2 LC) + \frac{j\omega L}{R} (2 - \omega^2 LC)}$$

which is in agreement with Eq. (5-65).

(c) Thévenin's Theorem

The same network can be solved by Thévenin's theorem by using the equivalence shown in Fig. 5-22. The section shown in solid lines at (a) is replaced by the circuit shown at (b). To accomplish this, we first refer to Fig. 5-22a, and observe that the open circuit value of $\bar{V}_{30}$ (which we shall designate as $\bar{V}_0$) can be obtained from the voltage divider principle, as

$$\bar{V}_0 = \frac{\frac{1}{j\omega C} \bar{V}_1}{j\omega L + \frac{1}{j\omega C}} = \frac{\bar{V}_1}{1 - \omega^2 LC} \quad (5-68)$$

5-42

The impedance $\bar{Z}_0$ is obtained from this figure by deactivating the source (replacing it by a short circuit) and computing the impedance between the terminals of the resulting network. This gives a network of L and C in parallel, for which the impedance is

$$\bar{Z}_0 = \frac{j\omega L \left(\frac{1}{j\omega C} \right)}{j\omega L + \frac{1}{j\omega C}} = \frac{j\omega L}{1 - \omega^2 LC} \quad (5-69)$$

The voltage divider principle can now be applied to the entire network of Fig. 5-21b,

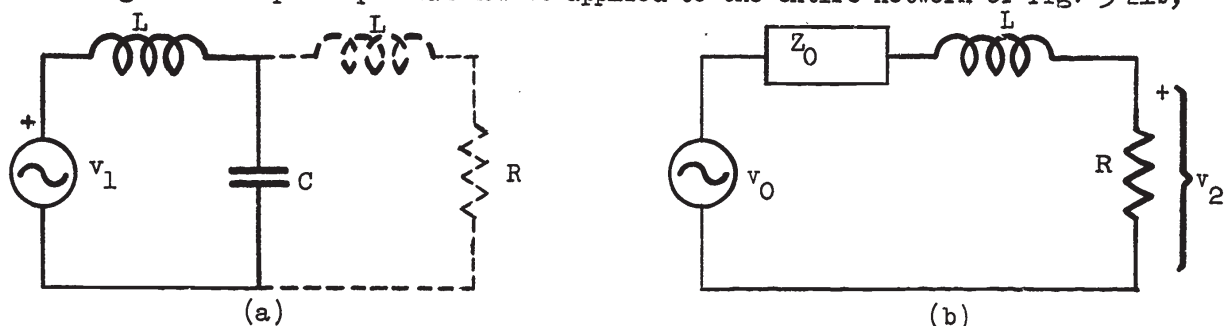


Figure 5-22.

giving

$$\begin{aligned} \bar{V}_2 &= \frac{R \left(\frac{\bar{V}_1}{1 - \omega^2 LC} \right)}{\frac{j\omega L}{1 - \omega^2 LC} + j\omega L + R} \\ &= \frac{\bar{V}_1}{1 - \omega^2 LC + j \frac{\omega L}{R} (2 - \omega^2 LC)} \end{aligned}$$

in agreement with Eq. (5-65).

We have now obtained the same formula for $\bar{V}_2$ by three different methods. In order to avoid leaving this as an abstract formula, let us consider its significance for some numerical cases. In so doing, we will not only add meaning to the formula, we shall be able to demonstrate more clearly what is meant by saying that the circuit in question is a low-pass filter.

The transmission at various frequencies can be determined from Eq. (5-65) by rewriting it as

$$\frac{\bar{V}_2}{\bar{V}_1} = \frac{1}{1 - \omega^2 LC + j \frac{\omega L}{R} (2 - \omega^2 LC)} \quad (5-70)$$

and assuming various numerical values for ω , from which numerical results can be obtained. Observe that there are only two combinations of parameters which enter into this expression, the product LC and the ratio L/R . Let these have the values $LC = 10^{-6} \text{ sec}^2$ and $L/R = 10^{-3} \text{ sec}$.

The following table can be constructed:

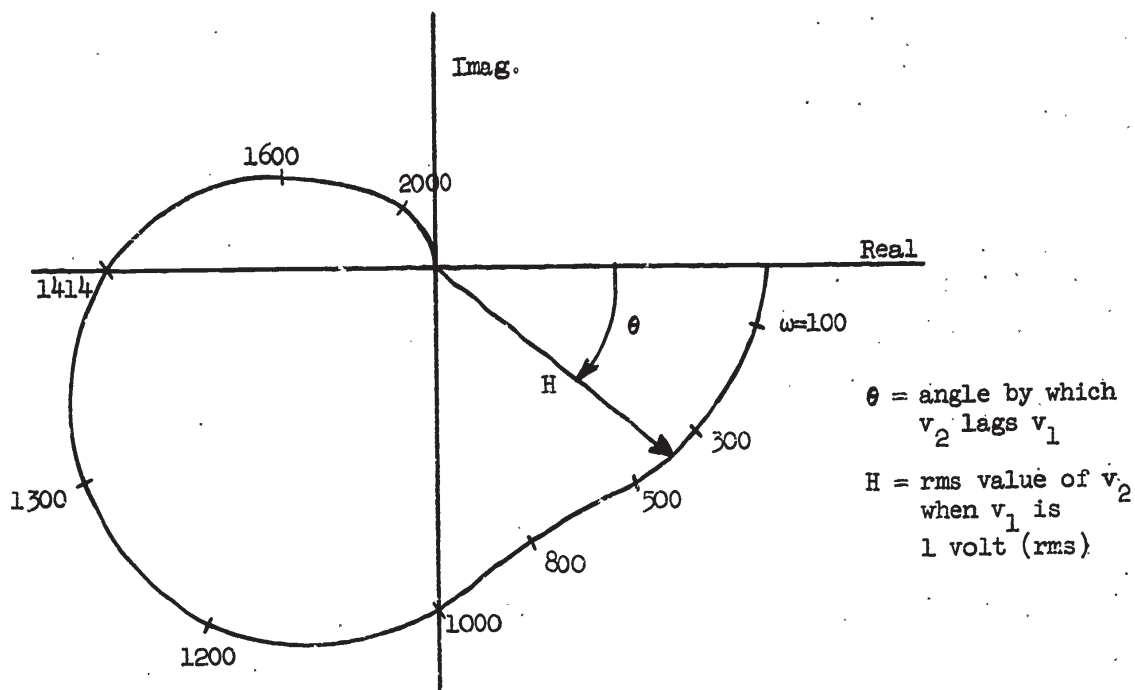
Table						
ω	$\omega^2 LC$	$1 - \omega^2 LC$	$2 - \omega^2 LC$	$.001 \omega (2 - \omega^2 LC)$	$1 - \omega^2 LC + j \frac{\omega L}{R} (2 - \omega^2 LC)$	$\bar{V}_2 / \bar{V}_1$
100	.01	.99	.99	.20	$1.00 + j.20 = 1.02 / 11^\circ$	$.98 / -11^\circ$
300	.09	.91	1.91	.57	$.91 + j.57 = 1.0 / 32^\circ$	$.93 / -32^\circ$
500	.25	.75	1.75	.88	$.75 + j.88 = 1.13 / 50^\circ$	$.88 / -50^\circ$
800	.64	.36	1.36	1.09	$.36 + j1.09 = 1.15 / 72^\circ$	$.87 / -72^\circ$
1000	1.00	0	1.00	1.00	$0 + j1 = 1 / 90^\circ$	$1 / -90^\circ$
1200	1.44	-.44	.56	.67	$-.44 + j.67 = .80 / 123^\circ$	$1.25 / -123^\circ$
1300	1.69	-.69	.31	.42	$-.69 + j.40 = .80 / 150^\circ$	$1.25 / -150^\circ$
1414	2.00	-1.00	0	0	$-1 + j0 = 1 / 180^\circ$	$1 / -180^\circ$
1600	2.56	-1.56	-.56	-.90	$-1.56 - j.90 = 1.80 / 210^\circ$	$.56 / -210^\circ$
2000	4	-3	-2	-4.0	$-3.0 - j4.0 = 5.0 / 233^\circ$	$.20 / -233^\circ$
3000	9	-8	-7	-21	$-8.0 - j21 = 22.4 / 249^\circ$	$.05 / -249^\circ$
4000	16	-15	-14	-56	$-5.0 - j56 = 56 / 265^\circ$	$.02 / -265^\circ$

The most direct way to portray the complex quantity $\bar{V}_2 / \bar{V}_1$ as a function of ω , is to plot the results carried in the right-hand column of the above tabulation, as in Fig. 5-23a. Imagine a unit voltage $\bar{V}_1$ (rms value = 1) so that its wave is described by

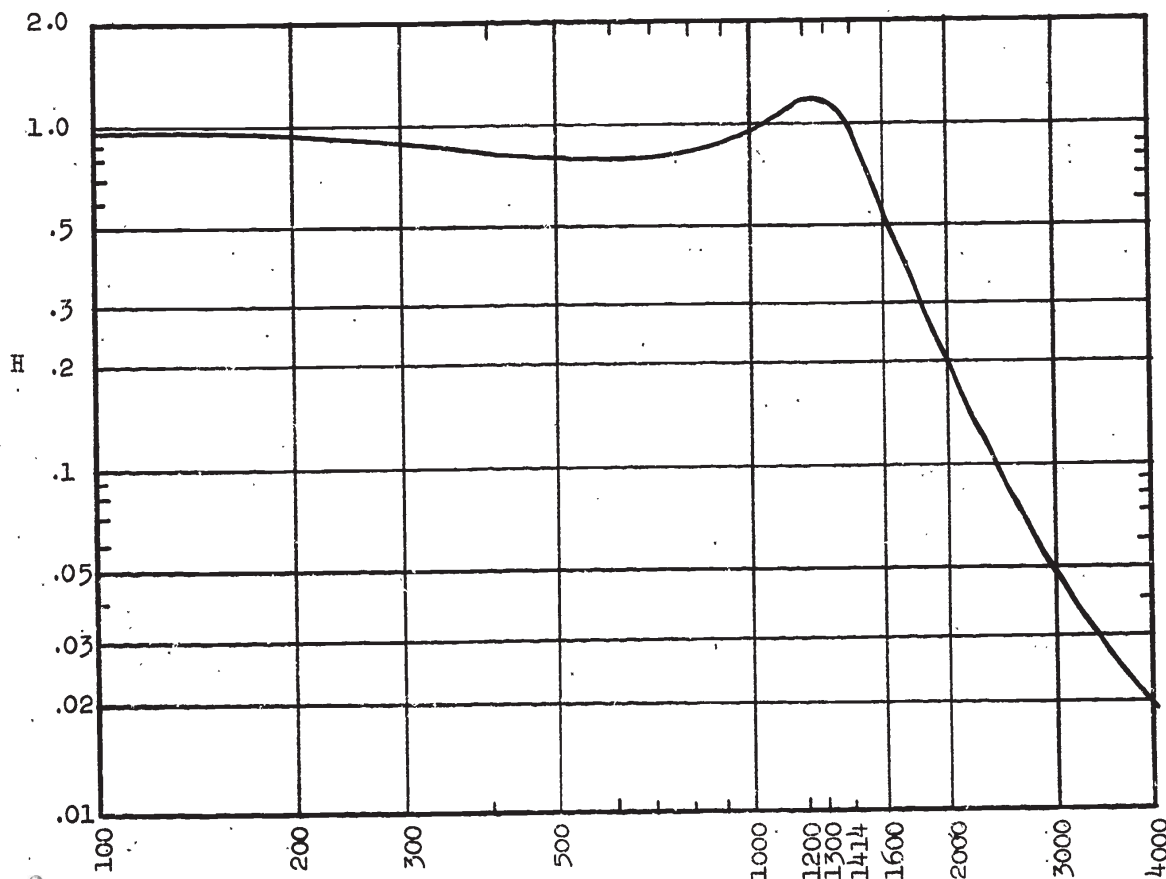
$$v_1 = \sqrt{2} \cos \omega t \quad ; \quad \text{and} \quad \bar{V}_1 = 1 + j0$$

where ω can have any value in the range covered by the table and Fig. 5-23a.

5-44



(a)



(b)

Figure 5-23.

At any frequency $\bar{V}_2/\bar{V}_1 = He^{-j\theta}$, where H and θ are indicated in the figure for $\omega = 300$. Since $\bar{V}_1 = 1 + j0$ for the unit voltage,

$$\bar{V}_2 = He^{-j\theta}$$

In other words, Fig. 5-23a portrays the response of the circuit for waves of different frequencies and unit rms value. At several frequencies the outputs are

$$v_2 = .93\sqrt{2} \cos(300t - 11^\circ) \quad \text{at} \quad \omega = 300$$

$$v_2 = .87\sqrt{2} \cos(800t - 72^\circ) \quad \text{at} \quad \omega = 800$$

$$v_2 = .20\sqrt{2} \cos(2000t - 233^\circ) \quad \text{at} \quad \omega = 2000.$$

It is evident that the output decreases at the frequency increases. This why this circuit is called a low pass filter.

A common method of portraying the variation of output with frequency is to plot H as a function of frequency, as in Fig. 5-23b. Note that logarithmic scales have been used for H and ω . This is usually done, because of the wide ranges of which they vary.

5-11 Power

Referring to Fig. 5-24a, let the box represent any combination of circuit

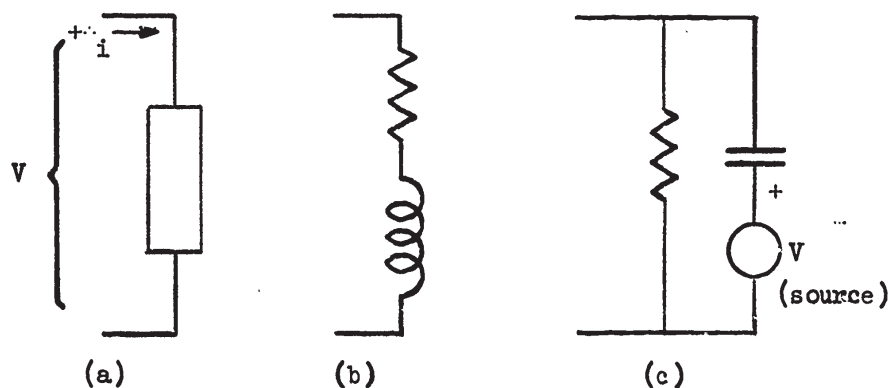


Figure 5-24.

elements, like the examples shown at (b) and (c). In general, there will be a phase difference between v and i , except in the case of a pure resistance. Thus, for the general case, v and i are respectively represented by

$$\begin{aligned} v &= \sqrt{2} V \cos(\omega t + \beta) \\ i &= \sqrt{2} I \cos(\omega t + \alpha) \end{aligned} \quad (5-71)$$

where the phase difference (the angle by which v leads i) is $(\beta - \alpha)$. In complex number representation, as introduced in Eqs. (5-17), these are

$$\begin{aligned} v &= \sqrt{2} \left(\frac{\bar{V} e^{j\omega t} + \bar{V}^* e^{-j\omega t}}{2} \right) \\ i &= \sqrt{2} \left(\frac{\bar{I} e^{j\omega t} + \bar{I}^* e^{-j\omega t}}{2} \right) \end{aligned}$$

where

$$\bar{V} = V e^{j\beta} \quad \text{and} \quad \bar{I} = I e^{j\alpha}$$

To find the power, we take the product of the above and get

$$\begin{aligned} p &= \frac{1}{2} (V e^{j\omega t} + \bar{V}^* e^{-j\omega t}) (I e^{j\omega t} + \bar{I}^* e^{-j\omega t}) \\ &= \left(\frac{\bar{V} \bar{I}^* + \bar{V}^* \bar{I}}{2} \right) + \left(\frac{\bar{V} \bar{I} e^{j2\omega t} + \bar{V}^* \bar{I}^* e^{-j2\omega t}}{2} \right) \end{aligned}$$

Referring to the first term on the right of the above equation, observe that

$$\bar{V} \bar{I}^* = V I e^{j(\beta - \alpha)} \quad \text{and} \quad \bar{V}^* \bar{I} = V I e^{-j(\beta - \alpha)}$$

so that for this term we have

$$\frac{\bar{V} \bar{I}^* + \bar{V}^* \bar{I}}{2} = V I \left(\frac{e^{j(\beta - \alpha)} + e^{-j(\beta - \alpha)}}{2} \right) = V I \cos(\beta - \alpha) \quad (5-72)$$

In the second term observe that $\bar{V} \bar{I} = V I e^{j(\beta + \alpha)}$ and that $\bar{V}^* \bar{I}^* = (\bar{V} \bar{I})^* = V I e^{-j(\beta + \alpha)}$, so that for this term we have

$$\frac{\bar{V} \bar{I} e^{j2\omega t} + \bar{V}^* \bar{I}^* e^{-j2\omega t}}{2} = V I \cos(2\omega t + \beta + \alpha)$$

Thus, starting with the product v_i , we have the formula for instantaneous

power

$$p = VI \cos(\beta - \alpha) + VI \cos(2\omega t + \beta + \alpha) \quad (5-73)$$

A graph of this is shown in Fig. 5-25, where the dashed line represents the constant

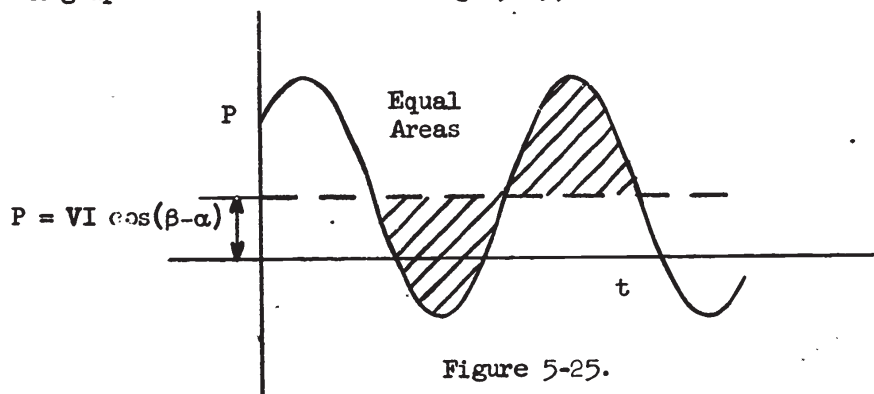


Figure 5-25.

term $P = VI \cos(\beta - \alpha)$. The $VI \cos(2\omega t + \beta + \alpha)$ term represents a sinusoidal wave of angular frequency 2ω , which therefore varies symmetrically with respect to this dashed line. Accordingly, over a long interval of time areas of the p curve above and below the dashed line nearly cancel, so that the dashed line is the steady state average value.⁺ (The words steady state mean "average taken over a long time interval.") If P is this average, we have the result

$$P = V I \cos(\beta - \alpha) \quad (5-74)$$

Observe from this that the power depends not only on the magnitude of the v and i waves, but also on the phase difference. In particular, if $\beta - \alpha = 90^\circ$, average power is zero, even if the voltage and current are not zero. This will be true of circuits having no resistance, since with only L and C elements there is a 90° phase difference between voltage and current waves.

Equation (5-72) shows that average power P can be obtained directly from $\bar{V}$ and $\bar{I}$, using

⁺Negative and positive areas cancel exactly only if reckoned over an integral number of cycles. If a fraction of a cycle is included in the interval over which the average is computed, the average will be different from $VI \cos \theta$. However, as the number of full cycles becomes large, the discrepancy becomes negligible.

$$\bar{P} = \frac{\bar{V}\bar{I}^* + \bar{V}^*\bar{I}}{2} \quad (5-75)$$

This form is useful when $\bar{V}$ and $\bar{I}$ are obtained from previous calculations in rectangular form. For example, suppose

$$\bar{V} = V_1 + jV_2 \quad \text{and} \quad \bar{I} = I_1 + jI_2$$

Then,

$$\bar{V}\bar{I}^* = (V_1 + jV_2)(\bar{I}_1 - j\bar{I}_2) = V_1\bar{I}_1 + V_2\bar{I}_2 + j(V_2\bar{I}_1 - V_1\bar{I}_2)$$

$$\bar{V}^*\bar{I} = (V_1 - jV_2)(I_1 + jI_2) = V_1I_1 + V_2I_2 - j(V_2I_1 - V_1I_2)$$

and thus, by substituting the sum of the above in Eq. (5-70), we have

$$P = V_1I_1 + V_2I_2 \quad (5-76)$$

This last formula is easily remembered. If voltage current phasors are known in rectangular form, (i.e., real and imaginary parts explicitly known) the average power is obtained by multiplying the real parts of $\bar{V}$ and $\bar{I}$, multiplying the imaginary parts, and adding the two quantities so obtained. As an alternate viewpoint, observe that the right-hand side of Eq. (5-76) is identical with the real part of either $\bar{V}\bar{I}^*$ or $\bar{V}^*\bar{I}$. Accordingly, Eq. (5-75) can be replaced by

$$P = \text{Re}(\bar{V}\bar{I}^*)$$

where the symbol $\text{Re}(\quad)$ means the real part of the quantity in the parentheses.

The product VI (which would be power in a d-c circuit) is called apparent power in steady state circuit analysis. The ratio

$$F = \frac{P}{VI} = \frac{\text{average power}}{\text{apparent power}} \quad (5-77)$$

is called the power factor of the device or circuit branch in question. From Eq. (5-74) it is evident that

$$F = \cos(\beta - \alpha)$$

where $(\beta - \alpha)$ is the phase difference between the v and i waves. Apparent power is measured in volt-amperes.

Chapter 6

NATURAL RESPONSE OF ELECTRIC CIRCUITS

1 Introduction

In the everyday world there are many opportunities to observe vibrations in mechanical systems. For example, a weight at the free end of a strip of springy steel clamped at the other end, as in Fig. 6-1a, will vibrate if it is deflected and then released. A plot of the deflection as a function of time is shown at (b). This is called a damped oscillation or a damped sinusoid, because the oscillations gradually die out. You will observe that the oscillation occurs in the absence of any kind of oscillatory source. The spring is merely deflected, and then let go. It is, therefore, appropriate that the resulting wave should also be called a natural response or a natural oscillation. There are many other examples: the oscillation of a piano string, the oscillation of the pointer on a weighing scale, when a weight is suddenly placed on it, etc. In the case of certain electrical circuits, currents, voltages, and charges can also oscillate in a similar manner.

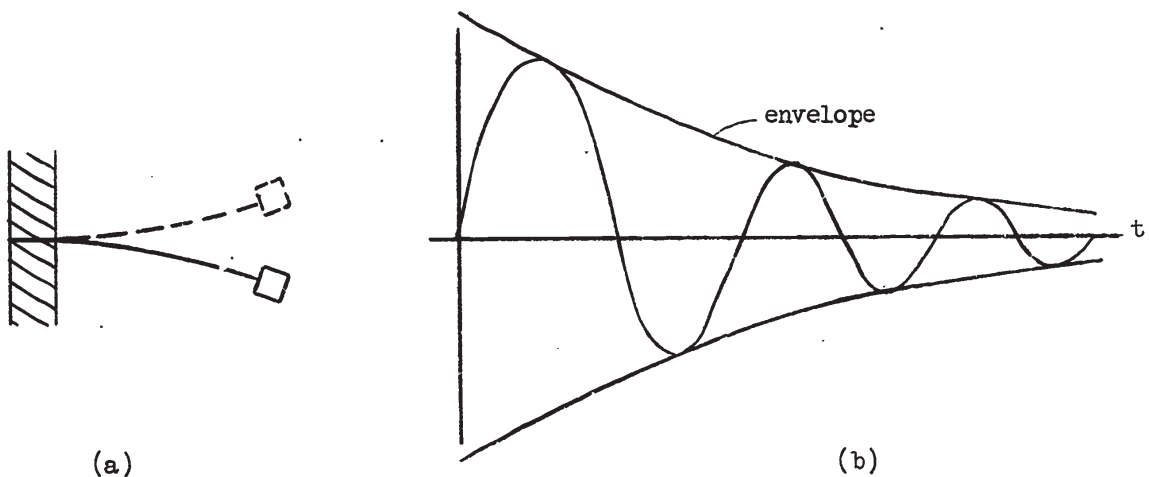


Figure 6-1.

- 1 A constant amplitude sinusoidal wave has the equation

$$v = V_m \cos(\omega t + \beta)$$

as we saw in Chap. 5. A damped sinusoid has the equation

2
$$v = V_m e^{-at} \cos(\omega t + \beta)$$

That is, it is a sinusoidal wave multiplied by a factor e^{-at} which forces the waves to be "pinched" down between the envelope curves shown in Fig. 6-1b. As time goes on, the amplitude of the wave decreases or decays.

- 3 In this chapter we shall show how, under certain conditions, responses of the above form are obtained from electric circuits. We shall also find that under certain conditions, if the rate of decay exceeds a certain value, the oscillatory character of the response disappears, and the above response
- 4 function reduces to a sum of two exponentials, like

$$v = V_a e^{-at} + V_b e^{-bt}$$

where V_a , V_b , a , and b are constants.

- In the body of this chapter we shall derive these equations, determine
- 5 how values of the circuit constant determine whether or not the response is oscillatory, and show how to obtain the various constants appearing in these equations.

6-1 The Circuit Equations

- 6 A number of different networks could be used as an introductory illustration. The series circuit shown in Fig. 6-2 is chosen. After the switch is closed, Kvl applies, and we have

$$v_L + v_R + v_C = V_B$$

- 7 and thus

6-3

1

$$L \frac{di}{dt} + Ri + v_C = V_B \quad (6-1)$$

In addition to this equation, i and v_C are related by

2

$$C \frac{dv_C}{dt} = i \quad (6-2)$$

3

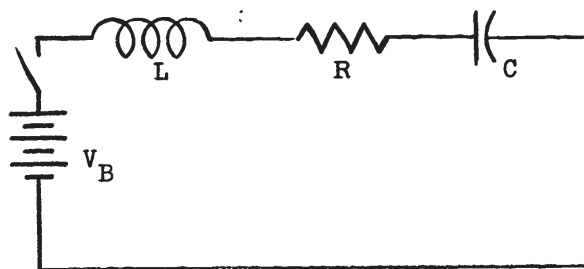


Figure 6-2.

4

Equations (6-1) and (6-2) constitute a pair of simultaneous equations in the two variables v_C and i . However, they are differential equations rather than algebraic equations. It is our objective to find solutions for these equations.

5

Ensuing steps will be simplified by introducing two new variables

$$x_1 = \sqrt{\frac{L}{C}} i \quad (a)$$

(6-3)

6

$$x_2 = v_C - V_B \quad (b)$$

and the parameters

$$\alpha = \frac{R}{L} \quad (a)$$

7

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad (c)$$

6-4

- 1 Incorporating these notational changes in Eqs. (6-1) and (6-2) yields the two new equations

$$\frac{dx_1}{dt} = -\alpha x_1 - \omega_0 x_2 \quad (6-5)$$

2
$$\frac{dx_2}{dt} = \omega_0 x_1 \quad (6-6)$$

as you can easily show by carrying out the steps yourself. These variable changes are not necessary to obtain a solution, but they have the advantage of reducing the original equations to the somewhat simpler forms shown above.

- 3 The remainder of our discussion will apply to these expressions, with the understanding that Eqs. (6-3) can be used to regain the original quantities v_C and i . You should observe that both x_1 and x_2 have the dimensions of voltage.

- 4 A two-pronged attack will be used. The first will be a graphical analysis of the equations, from which physical insight can be gained as to the nature of the solutions. The second will be the actual mathematical solution. While both are important, the first part of the discussion is particularly significant, from an engineering viewpoint.

6-3 The Phase Plane

- Since we are dealing with the two variables x_1 and x_2 , it is natural to think of plotting their relationship in a plane, called a phase plane. These quantities vary with time, and at any instant of time the pair of numbers (x_1, x_2) completely specifies the state of the circuit. Consequently, these variables are called the state variables. The point in the plane having the coordinates (x_1, x_2) is called the state point. In thinking about the phase plane we use the mental image of an ordinary geometrical plane surface. Thus, we talk about distances in the plane, even though the dimensions of the axis are not length, and we talk about velocities as the phase point moves around in the plane.

Let us determine how the state point varies with time, for the simplest possible case ($R = 0$, which means $\alpha = 0$), so that the equations reduce to

$$\frac{dx_1}{dt} = -\omega_0 x_2$$

(6-7)

$$\frac{dx_2}{dt} = \omega_0 x_1$$

To begin, we observe that an explicit algebraic solution is impossible, because these are differential equations. However, assume the state of the system is known at some particular instant of time; that is, assume x_1 and x_2 are known. Algebra can then be used to determine how the state is changing at that instant, by finding the velocity of the state point. This can be done because dx_1/dt and dx_2/dt are, respectively, the x_1 and x_2 components of that velocity, as shown by the two dashed vectors in Fig. 6-3.

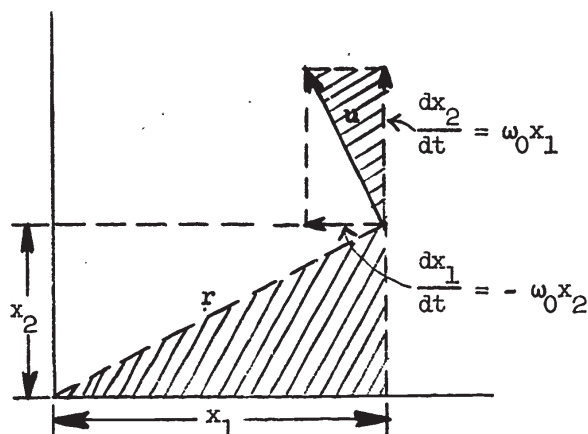


Figure 6-3.

The actual velocity of the point is indicated by the solid arrow, and this velocity has a magnitude which can be found from the figure to be

$$u = \sqrt{\left(\frac{dx_1}{dt}\right)^2 + \left(\frac{dx_2}{dt}\right)^2} = \omega_0 \sqrt{x_1^2 + x_2^2}$$

1 From Fig. 6-3 we make the following observations:

(1) $\sqrt{x_1^2 + x_2^2} = r$, the distance from the origin to the point;

(2) The shaded triangles are similar, and therefore the velocity is perpendicular to the radial line r .

2 Now observe that $\omega_0 (= 1/\sqrt{LC})$ is a constant, and that

$$u = r\omega_0 \quad (6-8)$$

Since the point moves with a velocity perpendicular to a radius, it follows that the point moves on a circle, with radius r . (Why?) Thus, we find that

3 the velocity of the state point has a constant magnitude. The point seems to be rigidly attached to a wheel rotating with constant angular velocity ω_0 .

The problem is not completely solved, but we have taken a big step. Although ω_0 is known from the circuit parameters, we do not yet know a value of r , nor is it clear where the point is at various instants of time.

4 Completion of the solution requires knowledge about the circuit at some particular instant of time, usually taken as $t = 0$. Thus, let $x_1(0)$ and $x_2(0)$ be the values of the state variables when $t = 0$. These values, which are called the initial conditions, must be determined from the specified problem. For example, suppose the actual circuit arrangement is shown in Fig. 6-4, in

5 which the switching arrangement is slightly more complicated than in Fig. 6-2, in order to provide the possibility of having an initial current $i(0)$ in the conductor. The capacitor is assumed to have a charge at $t = 0$, leading to an initial capacitor voltage $v_C(0)$. Thus, the initial values of the two state variables are

$$x_1(0) = \sqrt{\frac{L}{C}} i(0) \quad (a)$$

(6-9)

$$x_2(0) = v_C(0) - V_B \quad (b)$$

7 As battery A is disconnected from the inductance, at $t = 0$, the inductance cannot have a discontinuity of current, and so $i(0)$ must become the current in the L-C circuit. Although this may seem to be a rather artificial situation, it provides the possibility that $x_1(0)$ and $x_2(0)$ can both be different from zero.

8 We could have simply said assume both $x_1(0)$ and $x_2(0)$ are non-zero.

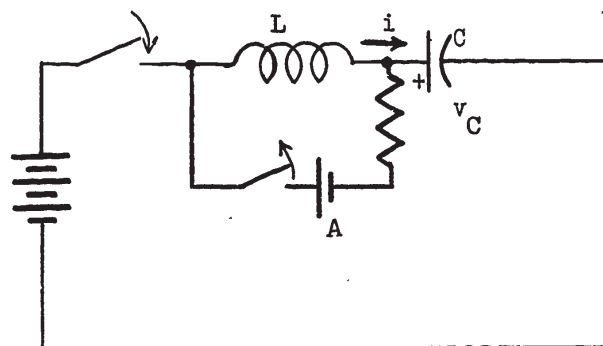


Figure 6-4.

With knowledge of $x_1(0)$, and $x_2(0)$, the position of the rotating point is known at $t = 0$, as indicated in Fig. 6-5, in which the state point is the tip of the radius line. Since we also know that the line rotates with angular velocity ω_0 , the motion of the state point becomes completely known. The time variation of $x_1(t)$ and $x_2(t)$ can now be obtained graphically, as shown in Fig. 6-5. This construction has an unusual orientation, with the time axis in a vertical position for x_2 .

Analytical expressions are easily obtained from this graphical analysis. The radius

$$r = \sqrt{[x_1(0)]^2 + [x_2(0)]^2}$$

and angle

$$\theta = \arctan \frac{x_2(0)}{x_1(0)} \quad (6-10)$$

are known from the initial conditions. Also, at any time t the angular position of the state point is $\omega t + \theta$, and so we have

$$x_1(t) = \sqrt{[x_1(0)]^2 + [x_2(0)]^2} \cos \left[\omega_0 t + \arctan \frac{x_2(0)}{x_1(0)} \right] \quad (a)$$

$$x_2(t) = \sqrt{[x_1(0)]^2 + [x_2(0)]^2} \sin \left[\omega_0 t + \arctan \frac{x_2(0)}{x_1(0)} \right] \quad (b) \quad (6-11)$$

1 It would have been rather easy to obtain these equations mathematically
 2 from the original equations. However, this graphical approach has the significant
 advantage of providing a pictorial indication of how the response is determined
 by the initial conditions, through their effect on both the radius of the
 rotating line and its initial angle.

This example has pedagogical value because, as the limiting case ($\alpha = 0$)
 in which there is no energy dissipation, it provides a base for consideration
 of the more general case. The $\alpha = 0$ case can also be of direct practical
 use in the analysis of transistor or vacuum tube oscillators. We shall not go
 into the details of this question here, except to point out that an essential
 function of the vacuum tube or transistor in an oscillator is to cancel the
 effect of the actual circuit resistance.

4 The fact that $x_1^2 + x_2^2$ is a constant has interesting implications in
 the case where $V_B = 0$, so that $x_2 = v_C$. Going back to the definitions of
 x_1 , and x_2 , in Eqs. (6-3), we then have

$$\frac{L}{C} i^2 + v_C^2 = r^2$$

or

$$\frac{Li^2}{2} + \frac{Cv_C^2}{2} = \frac{Cr^2}{2} \quad (6-12)$$

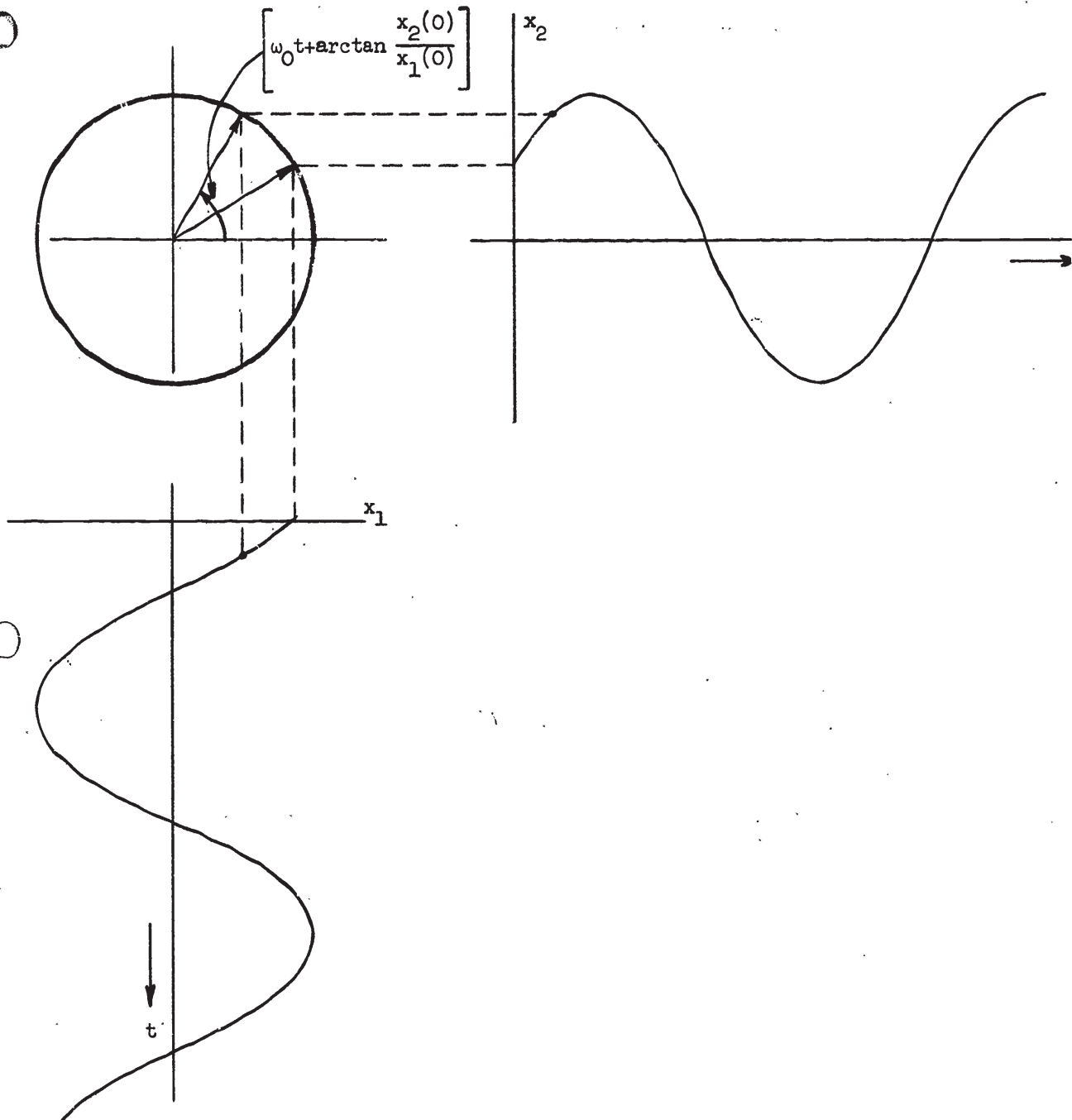
The right-hand side of this equation is a constant. On the left, $Li^2/2$ is
 the energy stored in the inductance at any instant, and $Cv_C^2/2$ is
 the energy stored in the capacitance. The total energy is constant (as we
 should expect, since no energy is dissipated in the absence of resistance),
 and so as one increases the other decreases. Twice during each cycle the
 energy is completely interchanged between the inductance and capacitance,
 being a maximum in one when it is zero in the other.

Now suppose $\alpha \neq 0$. Continuing to think of the velocity of the state
 point in the phase plane, referring to Eqs. (6-5) and (6-6) we see that
 the vertical velocity dx_2/dt is unchanged, but that the horizontal component
 now becomes

$$\frac{dx_1}{dt} = -\alpha x_1 - \omega_0 x_2$$

In other words, at any point not on the x_2 axis an additional component
 $-\alpha x_1$ must be added to the horizontal velocity, as indicated by two examples

6-9



6-10

in Fig. 6-6, one a positive value of x_1 and one at a negative value. The additional velocity component is always toward the x_2 axis, and is proportional to x_1 , the distance of the point from the x_2 axis. The sign on $-\alpha x_1$ ensures that the direction of this additional velocity component shall be toward the x_2 axis, regardless of the location of the state point.

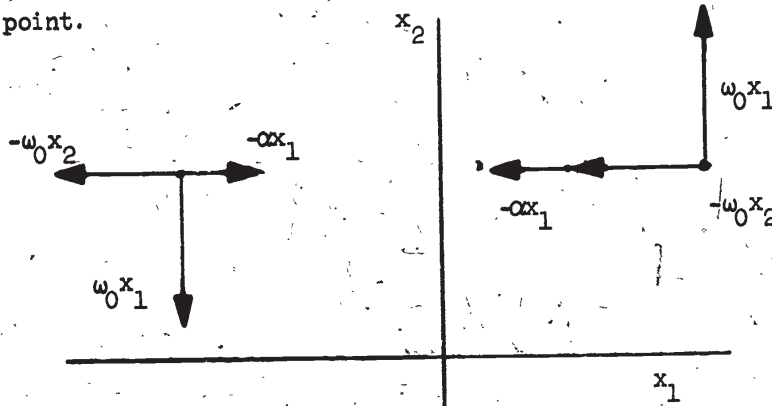


Figure 6-6.

For analysis of this case, it is convenient to write the original equations in the form

$$\left(\frac{1}{\omega_0}\right) \frac{dx_1}{dt} = -\left(\frac{\alpha}{\omega_0} x_1 + x_2\right)$$

(6-13)

$$\left(\frac{1}{\omega_0}\right) \frac{dx_2}{dt} = x_1$$

This shows that the ratio of the velocity components is

$$\frac{\alpha}{\omega_0} = \frac{R}{L} \sqrt{LC} = R \sqrt{\frac{C}{L}}$$

determines the extent to which the velocity is modified. A good idea of how the circuit will respond can be obtained by calculating dx_1/dt and dx_2/dt at a variety of points, and plotting the resulting velocity vectors. Figures 6-7 and 6-8 show these sets of vectors for the two cases $\alpha = \omega_0/2$ and $\alpha = 2\omega_0$.

6-11

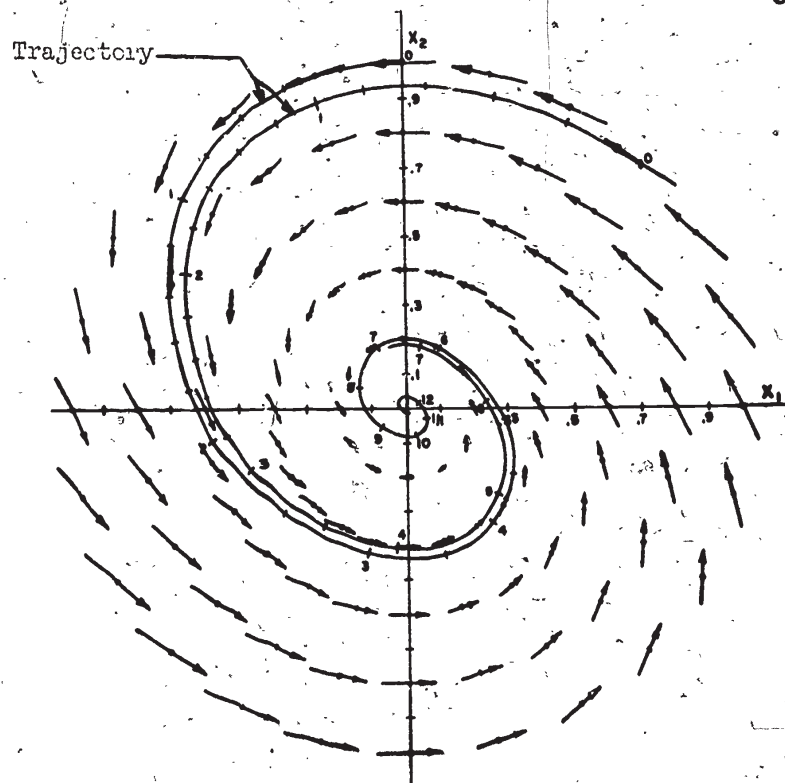


Fig. 6-7

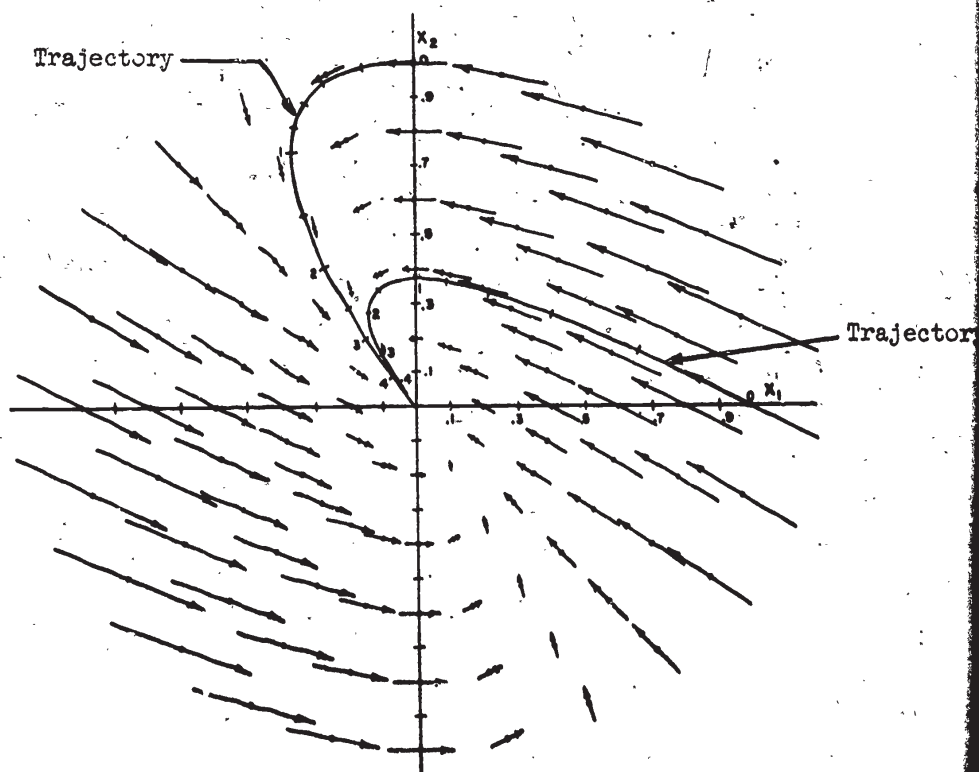
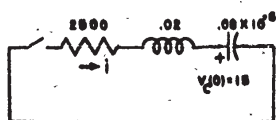


Fig. 6-8

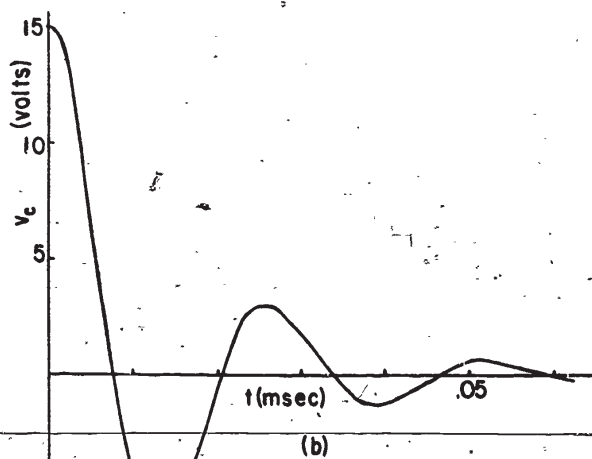
Such plots as these do not provide explicit solutions, but they do provide considerable insight as to what happens. As in the case when $\alpha = 0$, initial values $x_1(0)$ and $x_2(0)$ determine a starting point, from which the point moves with the appropriate velocity for that position. As the point moves, the velocity also changes and there is a new velocity appropriate for each position, and a path, or trajectory, is traced out. Because the velocities are tilted toward the vertical axis, these trajectories generally go toward the origin. Physically, this means that the circuit approaches a state of rest ($x_1 = x_2 = 0$, meaning $i = 0$ and $v_C = V_B$) as energy is gradually dissipated in the resistor. Several such trajectories are shown in Figs. 6-7 and 6-8 for different initial conditions.

These trajectories do not show the time variation directly, but this can be included (preferably in terms of values of $\omega_0 t$) by placing marks on the trajectory. In these figures, each arrow is of a length $0.2u$, meaning its length is the distance the point will move from the dot at the center of the arrow when $\omega_0 t$ increases by an amount 0.2, assuming a uniform velocity for this short length of time. Such time interval markings are included on the trajectories. Graphs of x_1 and x_2 as functions of time can be obtained by projecting points, as indicated in Figs. 6-9 and 6-10.

In comparing the cases $\alpha = \omega_0/2$ and $\alpha = 2\omega_0$, it is seen that in the former case the responses are basically oscillatory, but that this is not true when $\alpha = 2\omega_0$. The energy interpretation given for the dissipationless case provides some insight as to the reason for this difference. As previously stated, when there is no dissipation, energy is passed back and forth between capacitance and inductance, while the total energy remains constant. When there is dissipation ($\alpha > 0$), some energy is lost during each interchange, and hence each successive voltage and current peak of Fig. 6-9 is smaller than the previous one. However, when α becomes equal to or larger than a certain value (called the critical value, which we shall presently show is $2\omega_0$), the energy dissipation becomes so large that a complete cycle of operation never occurs, and the result is a response, as in Fig. 6-10, which is essentially non-oscillatory.



(a)



(b)

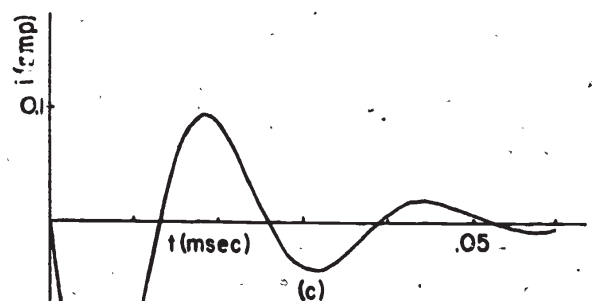
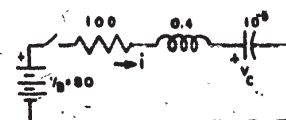
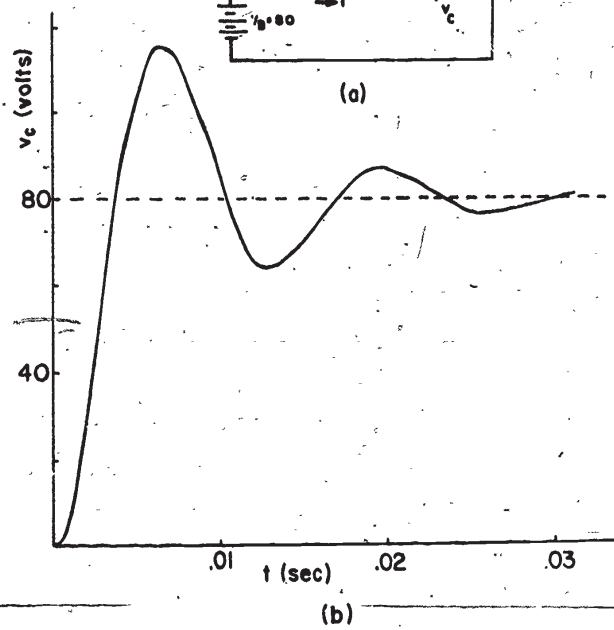


Fig. 6-11



(a)



(b)

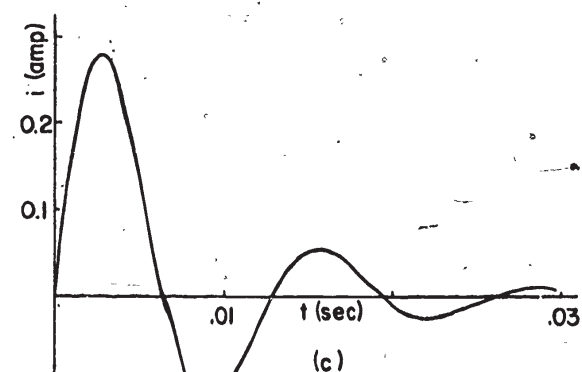


Fig. 6-12

- 1 This phase plane interpretation does not provide exact solutions, but it does show the effect of circuit resistance, and, as in the case of Fig. 6-5, shows how initial conditions determine the response curve, merely by locating the state point at $t = 0$. When R , L , and C are constants, as we
- 2 have tacitly assumed, exact analytical solutions are also possible, as we shall presently show. However, the phase plane approach can also be used for nonlinear equations, where α and/or ω_0 are functions of x_1 and/or x_2 . Such a functional relationship merely modifies the velocity vector at each
- 3 point in the phase plane. Graphical analyses are then generally necessary, or approximate point-by-point numerical calculations can be made to determine a succession of state points, perhaps by digital computer methods. The extensive velocity plots of Figs. 6-7 and 6-8 are not needed to obtain
- 4 solutions by this method, but they are included here for their pedagogical value, to aid in your visualization of a variety of trajectories from various starting points.

Illustrative Example

- 5 For the trajectories given in Figs. 6-7 and 6-8, the maximum radial coordinate for each curve (which occurs at the initial point) is unity. For this reason, they can be called normalized curves. Obviously, practical problems can have a value (r_m) for this initial radius which is different
- 6 from unity. Curves like Figs. 6-7 and 6-8 can then be regarded as plots of x_1/r_m vs. x_2/r_m (which will have a maximum value of unity). Actual values of x_1 and x_2 can then be obtained from the normalized curve merely by multiplying by r_m .

- 7 As an example, consider the circuit of Fig. 6-11a. It is assumed that the capacitor is initially charged to 15 volts, and we are to determine the variation of v_C and i after the switch is closed. Using the numerical values shown on the figure, we have

8

$$\alpha = \frac{2500}{.02} = 1.25 \times 10^5$$

$$\omega_0 = \frac{1}{\sqrt{(.02)(.08) \times 10^{-6}}} = 2.5 \times 10^5$$

6-15

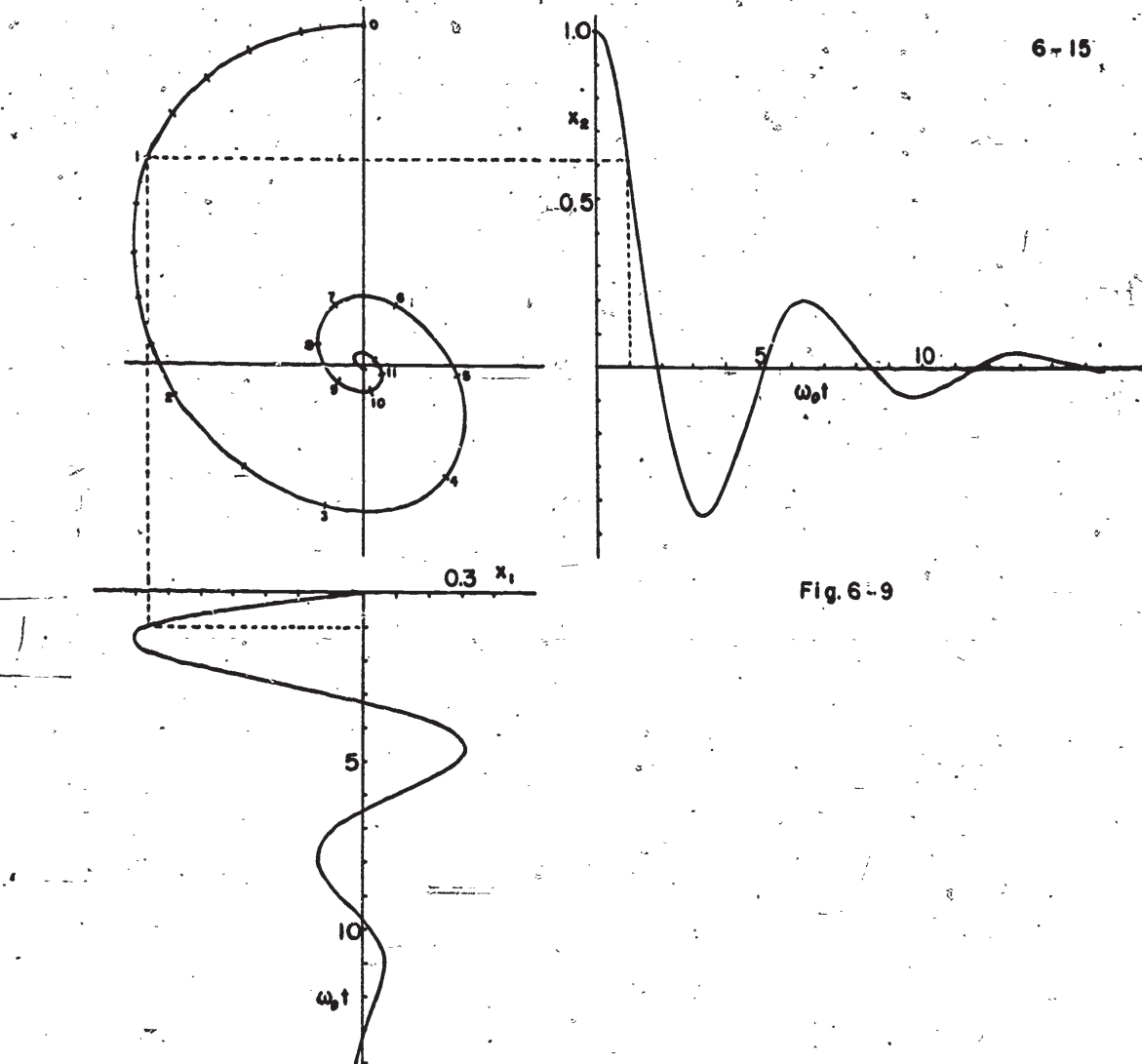


Fig. 6-9

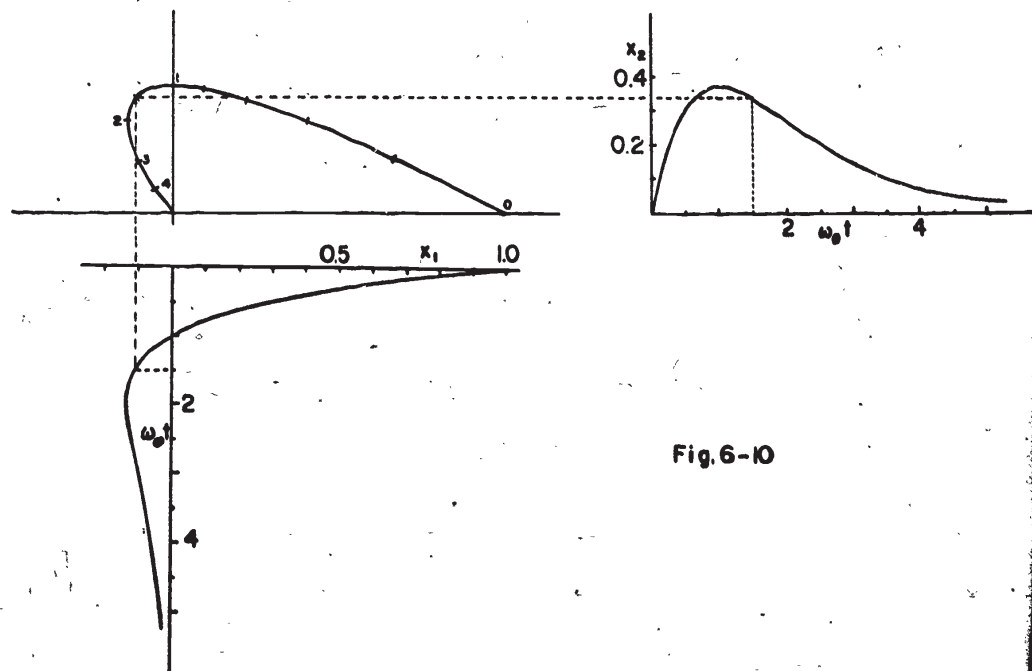


Fig. 6-10

- 1 Thus, we see that $\alpha = \omega/2$, and therefore that Fig. 6-7 is applicable.
 Also,

$$x_1 = \sqrt{\frac{.02}{.08 \times 10^{-6}}} i = 500i$$

$$x_2 = v_C$$

- and, from the stated initial conditions, $r_m = \sqrt{x_1(0)^2 + x_2(0)^2} = 15$.
 3 Since $x_1(0) = 0$, we see that Fig. 6-9, which was derived from Fig. 6-7, is the applicable curve, if values of x_2 are multiplied by 15. Values of x_1 are also multiplied by 15, but if we wish to obtain current i , we also divide by 500. Thus, to obtain i , the x_1 ordinates labels of Fig. 6-9 are multiplied by $15/500 = .03$. As mentioned previously, the time scale is obtained by dividing values of $\omega_0 t$ by $\omega_0 = 2.5 \times 10^5$ in this example. The resulting curves are shown in Fig. 6-11 (b) and (c). These are the same as the curves in Fig. 6-9, except for the scale changes noted above.

- 5 Another example is shown in Fig. 6-12a. In this case the capacitor is initially uncharged, and a battery of voltage $V_E = 80$ volts is switched into the circuit at $t = 0$. For this example

$$\alpha = \frac{100}{.4} = 250$$

$$\omega_0 = \frac{1}{\sqrt{(.40)(10) \times 10^{-6}}} = 500$$

- 7 Again $\alpha = \omega/2$, and so Fig. 6-7 can be used. The variables are

$$x_1 = \sqrt{\frac{.4}{10 \times 10^{-6}}} i = 200i$$

$$x_2 = v_C - 80$$

- 8 Observe that $x_1(0) = 0$ and $x_2(0) = -80$, and $r_m = 80$. We do not have a curve starting at a negative value of x_2 . However, Fig. 6-7 is symmetrical, and so the trajectory starting at $x_2(0) = 1$ can be used if signs on x_1 and
 9 x_2 are reversed. Furthermore, these variables are to be increased by the factor 80.

Thus, in effect x_1 and x_2 in Figs. 6-7 and 6-9 should be multiplied by -80 to yield the above values of x_1 and x_2 . Furthermore, to get i from the coordinates of Fig. 6-9, we divide x_1 by 200, and to get v_C we add 80. By inverting the x_2 of Fig. 6-9 and adding 80, we obtain a curve for v_C which starts from the origin and gradually settles down to the constant value 80, as shown in Fig. 6-12b. In other words, from the physical viewpoint, the capacitor is eventually charged to voltage $V_B = 80$. The current curve is shown in Fig. 6-12c, and this offers an interesting comparison with Fig. 6-11c. In Fig. 6-11 the current starts in a negative direction, because the capacitor is initially charged positively. When the switch is closed, positive charge begins to leave the left-hand plate, causing i to be negative. In Fig. 6-12, the capacitor voltage starts from zero, and so the applied voltage of the battery causes current to flow in the direction of the arrow, making i positive.

6-4 Mathematical Solution

We shall now consider the exact solution of the original equations, namely

$$\frac{dx_1}{dt} = -\alpha x_1 - \omega_0 x_2$$

$$\frac{dx_2}{dt} = \omega_0 x_1$$

for the linear case (α and ω_0 are independent of x_1 and x_2).

The form of the solutions $x_1(t)$ and $x_2(t)$ of these equations is not evident. However, it is a legitimate procedure to solve equations by a trial method, checking the trial solutions to see if they are satisfactory. The trial solutions will be

$$\begin{aligned} x_1 &= A_1 e^{st} \\ x_2 &= A_2 e^{st} \end{aligned} \tag{6-14}$$

where s is an undetermined constant introduced to provide generality to the exponential function. (That is, e^{st} is much like the exponentials $e^{-t/T}$ encountered in R-C and R-L networks, s being similar to $-1/T$.) These are

readily differentiated and substituted into the original equations, with the result

$$sA_1 e^{st} = -\alpha A_1 e^{st} - \omega_0^2 A_2 e^{st}$$

$$sA_2 e^{st} = \omega_0 A_1 e^{st}$$

The common factors e^{st} can be cancelled, giving

$$(s+\alpha)A_1 = -\omega_0 A_2 \quad (a)$$

$$sA_2 = \omega_0 A_1 \quad (b)$$

(6-15)

We now think of s as any constant, and attempt to solve these equations for A_1 and A_2 . The second equation gives $A_2 = (\omega_0/s)A_1$ which, when substituted in the first yields

$$(s+\alpha)A_1 = -\frac{\omega_0^2}{s}A_1 \quad \text{or} \quad (s+\alpha+\frac{\omega_0^2}{s})A_1 = 0$$

$A_1 = 0$ is a solution, but this is trivial and of no interest. If A_1 is not zero, we see that s must have a value which is a solution of the quadratic equation

$$s^2 + \alpha s + \omega_0^2 = 0 \quad (6-16)^*$$

This is called the characteristic equation of the network, and it has the two roots

*In more advanced treatments, where more than two equations is a possibility, Eqs. (6-15) are written

$$(s+\alpha)A_1 + \omega_0 A_2 = 0$$

$$-\omega_0 A_1 + sA_2 = 0$$

An attempt to solve these for A_1 and A_2 by using Cramer's rule, gives

$$A_1 = \frac{\begin{vmatrix} 0 & \omega_0 \\ 0 & s \end{vmatrix}}{\begin{vmatrix} s+\alpha & \omega_0 \\ -\omega_0 & s \end{vmatrix}}$$

Since the numerator is zero, A_1 can have a non-zero value only if

$$\begin{vmatrix} s+\alpha & \omega_0 \\ -\omega_0 & s \end{vmatrix} = 0$$

Expansion of this determinant yields Eq. (6-16).

$$s_a = -\frac{\alpha}{2} + \sqrt{\frac{\alpha^2}{4} - \omega_0^2} = -\frac{\alpha}{2} + \omega_0 \sqrt{\left(\frac{\alpha}{2\omega_0}\right)^2 - 1}$$

$$s_b = -\frac{\alpha}{2} - \sqrt{\frac{\alpha^2}{4} - \omega_0^2} = -\frac{\alpha}{2} - \omega_0 \sqrt{\left(\frac{\alpha}{2\omega_0}\right)^2 - 1} \quad (6-17)$$

which are different, unless $\alpha = 2\omega_0$. We shall exclude this exceptional case, for the present. Roots s_a and s_b are called the eigenvalues of the original equations. (The prefix eigen comes from a German word meaning characteristic.)

The two quantities s_a and s_b were obtained through our attempts to solve for A_1 and A_2 from Eqs. (6-15). From the second of these equations, we find

$$A_2 = \frac{\omega_0}{s_a} A_1 \quad \text{or} \quad A_2 = \frac{\omega_0}{s_b} A_1$$

depending on which eigenvalue is used. (The other equation also leads to the same results with a certain amount of mathematical manipulation. (Can you do it?) Let s_a be chosen. Then, from Eqs. (6-14) we see that

$$x_1 = A_1 e^{s_a t}$$

$$x_2 = \left(\frac{\omega_0}{s_a}\right) A_1 e^{s_a t} \quad (6-18)$$

are solutions of the original equations. Constant A_1 is arbitrary. (A direct check is worthwhile.) Since A_1 is arbitrary, and since s_b might just as well have been used in Eqs. (6-18), it follows that

$$x_1 = B_1 e^{s_b t}$$

$$x_2 = \left(\frac{\omega_0}{s_b}\right) B_1 e^{s_b t}$$

are also solutions, where B_1 is another arbitrary constant. When equations are linear, as is the case here, the sum of two solutions is also a solution. (An easy way to see this is to try it.) Thus, we can now say that the complete solutions are

$$x_1 = A_1 e^{s_a t} + B_1 e^{s_b t} \quad (a)$$

$$x_2 = \left(\frac{\omega_0}{s_a}\right) A_1 e^{s_a t} + \left(\frac{\omega_0}{s_b}\right) B_1 e^{s_b t} \quad (b) \quad (6-19)$$

These are solutions in the sense that identities are obtained upon substituting them into the original equations. However, the arbitrary constants A_1 and B_1 must be evaluated in order for Eqs. (6-19) to be solutions for a particular case. The fact that there are two constants in Eqs. (6-19) corresponds to the fact, as explained in the discussion of phase plane solutions, that there are two initial conditions, $x_1(0)$ and $x_2(0)$. Thus, for a specific case, placing $t = 0$ in Eqs. (6-19) gives

$$x_1(0) = A_1 + B_1 \quad (a)$$

$$x_2(0) = \left(\frac{\omega_0}{s_a}\right)A_1 + \left(\frac{\omega_0}{s_b}\right)B_1 \quad (b)$$

(6-20)

as two equations which can be solved for A_1 and B_1 .

This completes the formal steps in arriving at the solution, with the exception that the special case $s_a = s_b$ has not been treated. We can see that Eqs. (6-19) do not then provide a complete solution, because $e^{s_a t}$ becomes a factor of each term and, in effect, there is then only one arbitrary constant, which is not sufficient to accommodate two initial conditions. Treatment of that special case will be deferred until we develop some interpretive ideas about Eqs. (6-19). Referring to Eqs. (6-17) we see that if $\alpha \geq 2\omega_0$ the eigenvalues are real, and that they are complex if $\alpha < 2\omega_0$. Discussion of the implications of these two conditions will be an important part of our results. Another useful interpretation involves the relationship of the analytic solution to the phase plane solution presented earlier.

Concerning the phase plane, Fig. 6-13 shows the initial state point of a typical problem, and the obvious graphical interpretation of Eqs. (6-20).

Although s_a and s_b can be complex, and they appear in Eq. (6-20), it is necessary that $x_2(0)$ must be real. This is possible if A_1 and B_1 are also complex. In fact, since s_b is the conjugate of s_a , it can be shown that B_1 must be the conjugate of A_1 . This includes the possibility of A_1 and B_1 being equal and real, as a special case of the conjugate relationship,

- 1 when the imaginary part is zero.* Thus, in general, A_1 and B_1 can be complex and they are related by

$$\bar{B}_1 = \bar{A}_1^*$$

- 2 Bars are used over the symbols to indicate the possibility of their being complex:

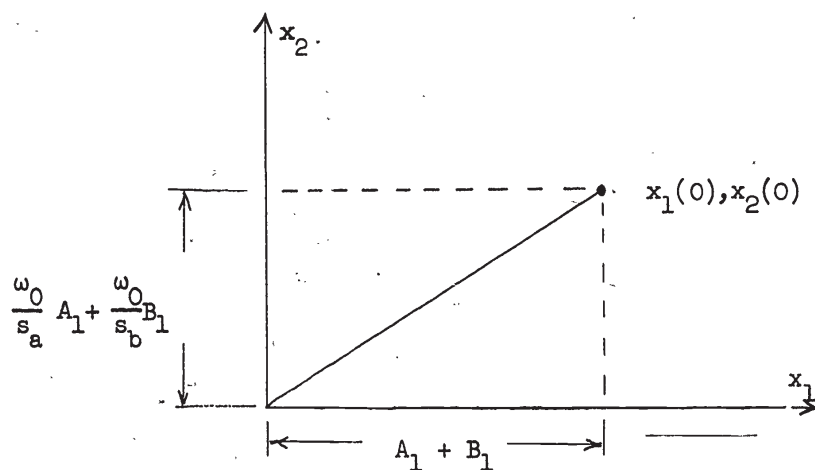


Figure 6-13

- 3
4
5 +To prove that $\bar{B}_1 = \bar{A}_1^*$ is necessary and sufficient for $x_2(0)$ to be real, adopt the notation $\bar{A}_1 = a_1 + ja_2$ and $\bar{B}_1 = b_1 + jb_2$. Also let $Me^{j\alpha} = (\omega_0/s_a)$, which implies that $(\omega_0/s_b) = Me^{-j\alpha}$. The equation for $x_1(0)$ is then

$$x_2(0) = M[(\cos \alpha + j \sin \alpha)(a_1 + ja_2) + (\cos \alpha - j \sin \alpha)(b_1 + jb_2)]$$

- 7 and its imaginary part is

$$M[a_1 \sin \alpha + a_2 \cos \alpha - b_1 \sin \alpha + b_2 \cos \alpha]$$

Setting this equal to zero gives

$$(a_2 + b_2) \cos \alpha = (a_1 - b_1) \sin \alpha$$

This equation must be true for any value of α . Therefore, the quantities in parentheses must be zero, and so

$$b_1 = a_1$$

$$b_2 = -a_2$$

Thus, $\bar{B}_1 = b_1 + jb_2 = a_1 - ja_2$, which is $\bar{A}_1^*$.

6-5 Interpretation of the Solutions

With the exception of the case $s_a = s_b$, Eqs. (6-20) can be solved in any numerical case. Thus, no additional analysis is needed to obtain numerical results for a specific problem. However, some general properties of the response curves can be determined. Also, the solution for the special case $s_a = s_b$ is yet to be determined. These are the topics to be considered in this section.

We shall treat three cases: (1) s_a and s_b are real; (2) $s_a = s_b$; (3) s_a and s_b are complex. Equations (6-19) and (6-20) serve as the starting point, the former providing the general form of the solution, and the latter containing the relationships from which A_1 and B_1 are determined.

Case (1): s_a and s_b are real.

Referring to Eqs. (6-17), it is seen that we can write

$$s_a = \sigma + \gamma \quad (6-21)$$

$$s_b = \sigma - \gamma$$

where

$$\left. \begin{aligned} \sigma &= -\frac{\alpha}{2} & (a) \\ \gamma &= \sqrt{\frac{\alpha^2}{4} - \omega_0^2} & (b) \end{aligned} \right\} \quad (6-22)$$

By straightforward algebraic manipulation, it is found that

$$\left. \begin{aligned} A_1 &= \frac{1}{2\gamma} [(\sigma + \gamma)x_1(0) - \omega_0 x_2(0)] & (a) \\ B_1 &= \frac{1}{2\gamma} [-(\sigma - \gamma)x_1(0) + \omega_0 x_2(0)] & (b) \end{aligned} \right\} \quad (6-23)$$

and

$$\left. \begin{aligned} A_2 &= \left(\frac{\omega_0}{\sigma + \gamma}\right)A_1 = \frac{1}{2\gamma} [\omega_0 x_1(0) - (\sigma - \gamma)x_2(0)] & (a) \\ B_2 &= \left(\frac{\omega_0}{\sigma - \gamma}\right)B_1 = \frac{1}{2\gamma} [-\omega_0 x_1(0) + (\sigma + \gamma)x_2(0)] & (b) \end{aligned} \right\} \quad (6-24)$$

1 It is rather easy to check that these are solutions by observing that the sum
of the first pair is $x_1(0)$, as required by Eq. (6-20), and the sum of the
second pair is $x_2(0)$, in agreement with Eq. (6-20).

2 Using Eqs. (6-23) and (6-24), we are now interested in interpretations
of

$$x_1 = A_1 e^{(\sigma+\gamma)t} + B_1 e^{(\sigma-\gamma)t} \quad (6-25)$$

$$x_2 = A_2 e^{(\sigma+\gamma)t} + B_2 e^{(\sigma-\gamma)t}$$

3 First consider the exponentials

$$e^{(\sigma+\gamma)t} \quad \text{and} \quad e^{(\sigma-\gamma)t}$$

4 observing that σ is negative, but greater in absolute value than γ . Thus,
for all positive t ,

$$(\sigma-\gamma)t < (\sigma+\gamma)t < 0$$

5 and also

$$e^{(\sigma-\gamma)t} < e^{(\sigma+\gamma)t}$$

6 For a typical case, these functions are related as in Fig. 6-14, illustrating
that $e^{(\sigma-\gamma)t}$ decays faster than $e^{(\sigma+\gamma)t}$. All solutions consist of suitable
combinations of these, according to Eqs. (6-23), (6-24), (6-25), (6-26).

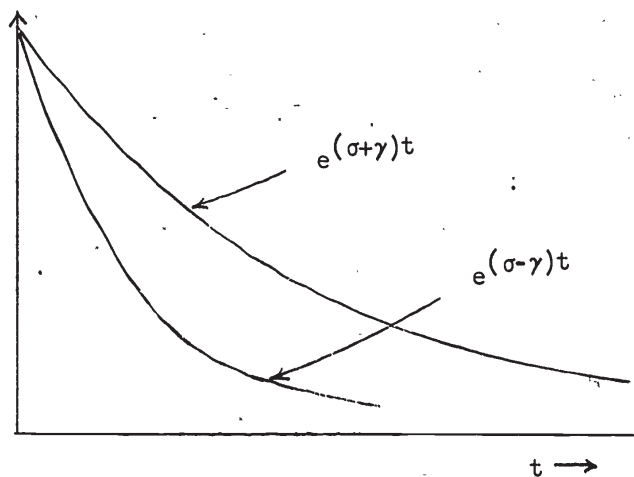


Figure 6-14.