part 1

FOUNDATIONS OF NONLINEAR NETWORK THEORY

1 TWO-TERMINAL NETWORK ELEMENTS

1-1 REVIEW OF BASIC PHYSICAL VARIABLES IN NETWORK THEORY

The advent of electrical science occurred with the discovery of the phenomenon that dry substances such as amber or rubber tend to repel or attract each other upon being rubbed by different materials such as silk or fur. This phenomenon was first explained by postulating the existence of a certain basic electrical quantity called the "electric charge" q, which may be either positive or negative, and which has the property that like charges exert a force of repulsion and unlike charges exert a force of attraction between each other. The quantity "charge" remains the most basic electrical quantity today, and its existence can now be explained by the atomic theory: a body is "charged" whenever there is an excess of the positive charges in the nucleus over the negatively charged electrons and vice versa. The practical unit of charge called the *coulomb* has been defined to be equivalent to the total charge possessed by 6.24×10^{18} electrons. The quantity of charge possessed by a body can be measured by various instruments such as the electroscope.

Since charged bodies exert forces on one another, energy or work is involved whenever one charged body is moved in the vicinity of another charged body. Hence if w is the work done by moving a charge q from point j to point k (assuming w is independent of the path taken),[1] then the potential difference, or voltage, between these points is defined as the work per unit charge; that is,

$$v_{jk} = \frac{w}{q} \tag{1-1}$$

[1] This assumption is only approximately satisfied in practice. The study of the conditions under which this assumption is valid belongs to a course in electromagnetic field theory. However, as far as network theory is concerned, the above assumption is automatically implied. Very roughly speaking, the above assumption is valid when the frequency of the signal is "not too high," that is, when the wavelength of the signals is long compared with the dimension of the physical network.

Observe that the magnitude of the charge is arbitrary; only the ratio between work and charge is important. Hence, the incremental work dw required to move an incremental test charge dq from point j to point k must also satisfy Eq. (1-1); thus

$$v_{jk} = \frac{dw}{dq} \tag{1-2}$$

When there is no possibility of confusion, we can delete the subscripts j and k and express the voltage simply as v. The practical unit of voltage is called the *volt*. The voltage v between two points can be measured by a voltmeter.

Charges can be caused to flow from one charged body into another by connecting a conducting wire between the two bodies. In 1819, Hans Christian Oersted discovered that the flow of charge through a wire produced a force on a compass needle in the vicinity of the wire and that force was proportional to the rate of flow of charge. Since the force on a compass needle can be easily determined by noting the deflection of the needle, the quantity "rate of flow of the charge" becomes very useful, and it has been given the name current, i. By definition,

$$i = \frac{dq}{dt} \tag{1-3}$$

The practical unit of current is the *ampere;* i.e., one ampere represents a charge flowing at a rate of one coulomb per second. The current i can be measured by an ammeter.

The deflection of a magnetic compass needle caused by the flow of charge, or current, in a conductor indicates that current produces a magnetic effect. This effect can be explained by the generation of a magnetic flux λ by the current. If the conductor is wound into a coil of n turns, then by defining $\varphi = n\lambda$ to be the flux linkage, Faraday discovered that the voltage between the two terminals of the coil is given simply by

$$v = \frac{d\varphi}{dt} \tag{1-4}$$

The practical unit of the flux linkage φ is called the *weber*. Flux linkage can be measured by a fluxmeter.

If we multiply together the left and right sides of Eqs. (1-2) and (1-3), we obtain

$$vi = \frac{dw}{dq}\frac{dq}{dt} = \frac{dw}{dt} \tag{1-5}$$

Since w represents energy, dw/dt represents the rate of change of energy, or the power p; hence

$$p(t) = v(t)i(t) \tag{1-6}$$

Summarizing, therefore, we find that the six basic electrical quantities of interest in network theory are the charge q, the voltage v, the current i, the flux linkage φ, the power p, and the energy w. The universal relationships between these quantities at any time t are

$$i(t) = \frac{dq(t)}{dt} \tag{1-7}$$

$$v(t) = \frac{d\varphi(t)}{dt} \tag{1-8}$$

$$p(t) = v(t)i(t) = \frac{dw(t)}{dt} \tag{1-9}$$

$$w(t) = \int_{-\infty}^{t} p(\tau)\,d\tau = \int_{-\infty}^{t} v(\tau)i(\tau)\,d\tau \tag{1-10}$$

$$q(t) = \int_{-\infty}^{t} i(\tau)\,d\tau \tag{1-11}$$

$$\varphi(t) = \int_{-\infty}^{t} v(\tau)\,d\tau \tag{1-12}$$

1-2 THE SIMULTANEITY POSTULATE IN LUMPED-NETWORK THEORY

The six basic electrical variables related by Eqs. (1-7) to (1-12) are assumed to be functions of only one independent variable, namely, the *time* of measurement t. Actually, to be exact, we must introduce another independent variable for specifying the relative location of the various terminals at which these electrical quantities are to be measured. This is the variable *length*, or *dimension*, in centimeters. The necessity for introducing this variable is due to the fact that it takes a finite amount of time for electrons to move from one point to another. For example, if we apply a voltage $v_s(t)$ across one end of a 30-cm lossless transmission line as shown in Fig. 1-1*a*, it will take 1 nsec (30 cm/3 $\times$ 10^{10} cm/sec = 10^{-9} sec) for the signal to arrive at the other end ($x = 30$ cm).[1] If

[1] For simplicity, we assume the electrons traverse down the line at the velocity of light. The actual electron velocity will, of course, depend on the characteristics of the transmission line.

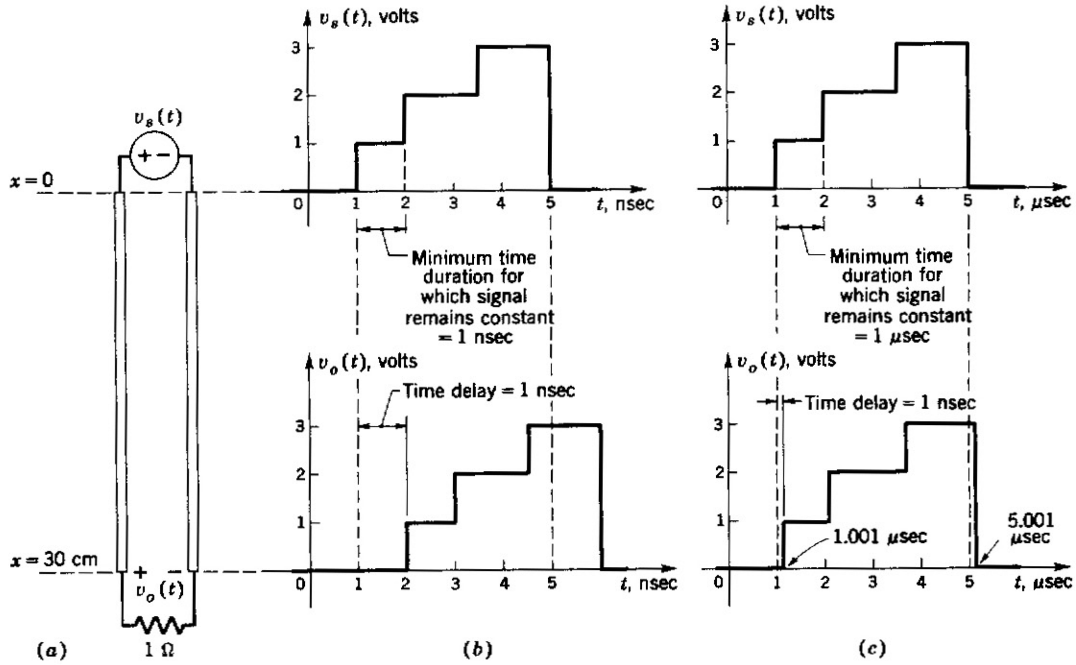


Fig. 1-1. The length of the transmission line introduces a time delay which is significant in (*b*) but may be neglected in (*c*).

the duration of time for which the signal level remains relatively unchanged is of the same order of magnitude (say, 2 nsec), then the time delay of the transmission line cannot be neglected. This is easily seen by comparing the signals $v_s(t)$ and $v_o(t)$ as shown in Fig. 1-1*b*. On the other hand, if the signal level does not change rapidly (relative to the time delay) as in Fig. 1-1*c*, then the time delay is insignificant and may therefore be neglected. Under this assumption, the output signal $v_o(t)$ may be considered to appear at the same instant as the input signal $v_s(t)$. This is equivalent to the assumption that the length of the transmission line is insignificant. In other words, the line can be *lumped* as one point so that *the current entering one terminal of a terminal pair appears instantaneously at the other terminal.* We will refer to this assumption as the *simultaneity postulate.*

The simultaneity postulate is a fundamental assumption in lumped-network theory that applies not only to transmission lines but also to all two-terminal black boxes considered in this book.[1] This postulate is valid whenever the physical dimension of each device inside the black box is small so that the time delay it introduces is insignificant compared with the minimum time duration for which the signals remain relatively constant. For periodic

[1] Networks which do not satisfy the simultaneity postulate are said to be distributed. The study of distributed networks belongs to a course in electromagnetic field theory.

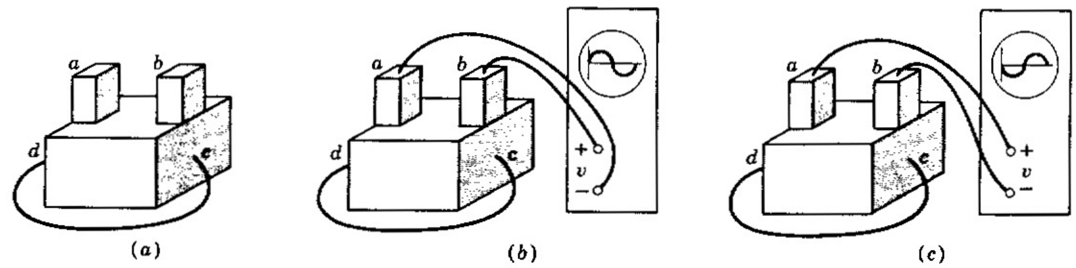


Fig. 1-2. An experiment demonstrating that regardless of which terminal of the black box is chosen to be positive, the actual voltage across terminals *a-b* can be unambiguously specified for all time.

signals, the reciprocal of the frequency is a good measure of this minimum time duration. Hence, roughly speaking, the higher the operating frequency, the smaller must be the device's physical dimension in order for the simultaneity postulate to be satisfied.[1] Fortunately, most nonlinear electronic circuits of interest do satisfy the simultaneity postulate. This is especially true with *integrated circuits* where the components are becoming so small that they can be seen only with the aid of a microscope.

1-3 SIGNIFICANCE OF THE REFERENCE CURRENT DIRECTION AND THE REFERENCE VOLTAGE POLARITY

One of the most basic concepts in physical science is that any physical quantity is invariably measured with respect to some "assumed" frame of reference. In electrical network theory, the frame of reference takes the form of an assumed reference direction of the current i and an assumed reference polarity of the voltage v. A thorough understanding of the concept of reference current direction and reference voltage polarity is absolutely essential in the study of nonlinear network theory. It is a fact that a large percentage of the mistakes committed by students of network theory can be traced to either the students' underestimation of the full significance of reference current directions and voltage polarities or the students' failure to maintain a consistent set of references.

Perhaps the simplest way to introduce the concept of assumed reference direction and polarity is through the following experiment. Suppose we are given a black box with a pair of terminals *a-b* and a wire *c-d* coming out of the box as shown in Fig. 1-2*a*. Suppose we are required to measure the voltage between terminals *a-b* and the current in the wire *c-d*.

Let us consider first measuring the voltage by connecting terminals *a-b* to the vertical input terminals of an oscilloscope.

[1] It can be justified on physical grounds that the simultaneity postulate is generally valid if the largest physical dimension of the device is much smaller than the *wavelength* of the highest anticipated frequency of operation.

Since one of the two vertical input terminals of any oscilloscope is marked with a positive sign while the other is marked with a negative sign, the question that immediately arises is which of the two terminals of the black box should we connect to the positive terminal of the oscilloscope in order to obtain the desired information. The answer is that it does not matter. In order to see this, suppose we *arbitrarily assume* terminal b to be connected to the positive terminal as shown in Fig. 1-2b. The assumption that terminal b is the positive terminal does not mean that the potential at b is higher than the potential at a. It does mean, however, that if at any time $t = t_1$, $v(t_1) > 0$, then the potential at b is higher than the potential at a. On the other hand, if $v(t_1) < 0$, then the potential at b at $t = t_1$ is actually lower than the potential at a. For example, if the voltage $v(t)$ displayed on the oscilloscope is given by

$$v(t) = 10 \sin \pi t \qquad \text{volts}$$

then terminal b is at a higher potential than terminal a during the time interval $0 < t < 1$ sec. But during the time interval $1 < t <$ 2 sec, terminal b is actually at a lower potential than terminal a.

Let us now consider what happens when we assume terminal a instead of terminal b to be the positive terminal, as shown in Fig. 1-2c. Since the connection in Fig. 1-2c is opposite to the connection in Fig. 1-2b, it is clear that the voltage $v(t)$ displayed on the oscilloscope is now given by

$$v(t) = -10 \sin \pi t \qquad \text{volts}$$

Since terminal a is now the assumed positive terminal, and since $v(t) < 0$ for $0 < t < 1$ sec, this means that during this time interval, terminal a is at a lower potential than terminal b. Similarly, we found that during the time interval $1 < t < 2$, terminal b is actually at a lower potential than terminal a.

In either case we found the final answers to be identical. We can, therefore, conclude that in order to specify the voltage between any pair of terminals unambiguously, we may arbitrarily assume any one of the two possible terminals to be the positive terminal.

By analogy, we can conclude that in order to specify the current in any wire unambiguously, we may arbitrarily assume any one of the two possible directions to be the positive direction. The actual direction in which the current $i(t)$ is flowing at any time $t = t_1$

will be in the assumed positive direction if $i(t_1) > 0$, and opposite to the assumed direction if $i(t_1) < 0$.

Let us consider next a two-terminal black box N and assume a reference direction for the terminal current i and a reference polarity for the terminal voltage v. Since the references for both i and v are arbitrary, there are four distinct sets of combinations of references. There is no reason to prefer any one combination over the others. However, in practice, it is usually convenient to choose the combination so that *positive* power

$$p(t) = v(t)i(t) > 0$$

represents power *entering* the black box. From basic electromagnetic principles, it can be shown that this condition is satisfied whenever the current is chosen to enter the assumed positive terminal of the black box. From the simultaneity postulate, the same current must leave the negative terminal. This means that the allowable reference combination must be either of the form shown in Fig. 1-3*a* or *b*.

In either case, observe that the current arrow either enters the positive terminal or leaves the negative terminal.

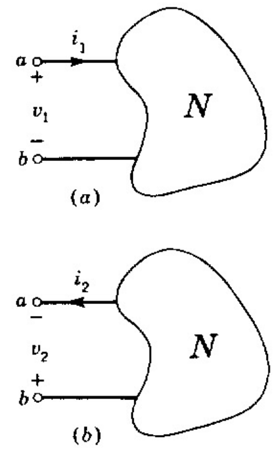


Fig. 1-3. Two possible sets of assumed reference direction and polarity for ensuring that positive power means power entering the black box.

1-4 INDEPENDENT SOURCES

Electrical energy must be supplied in order to move the charges which constitute the current i. Since energy can be neither created nor destroyed, it must be transformed from some other forms of energy. For example, a battery transforms chemical energy into electrical energy, a generator transforms mechanical energy into electrical energy. For convenience, we often refer to these energy-transforming devices, such as batteries or generators, as *sources of electrical energy* or simply *sources*. However, this statement should not be interpreted as implying that sources can create energy.

One of the earliest devices which serves as a source of electrical energy is the galvanic voltaic cell. Many other devices have been invented to function as sources of electrical energy, and, no doubt, many more will be invented in the future. Perhaps the simplest source of electrical energy today is the battery, which is capable of delivering a limited range of direct current to an external load connected with it, while maintaining an approximately *constant voltage* across its terminals. A less common source of electrical energy (but one gaining in popularity) is the solar cell, which develops a limited range of voltage drop across an external load connected with it, while maintaining an approximately *constant current* in the load. Observe that in the case of the battery, the output voltage is independent of the current drawn by the load (provided the current is not large). In contrast with this, it is the output current of the solar cell that is independent of the voltage drop across the load (provided the voltage drop is not large). It seems reasonable, therefore, to distinguish the above two types of electrical sources by calling the former a *voltage source* and the latter a *current source.*

Observe that in the above discussion, we have assumed that the current in the battery and the voltage drop across the solar cell are not large. This assumption is necessary because when a large current is drawn from the battery, its output voltage decreases and is no longer independent of the current. Similarly, a large voltage drop across the solar cell results in a decrease in the output current. The above phenomenon is a well-known experience; for example, the light dims whenever an appliance such as an air conditioner (which draws a large current) is turned on. In fact, no physical voltage source exists which is capable of developing a voltage that is entirely independent of its terminal current. Neither does there exist any physical current source which is capable of delivering a current that is entirely independent of its terminal voltage. While no such physical sources really exist, it is, nevertheless, extremely convenient to postulate the existence of the above sources as "ideal" sources. In other words, we are trying to "model" a physical source by an "ideal source" so that a network containing such sources can be conveniently analyzed. Observe that the above concept of modeling is analogous to that of the physicist who tries to represent the motion of a physical object by the motion of a "point" representing the center of gravity of the object. The concept of making a model to represent a physical system is so basic that we shall have many more occasions to use it in the future. With the above clarification, let us now render the concepts of independent sources more precise by the following discussions.

Independent voltage source An independent voltage source is a two-terminal device whose terminal voltage v is always equal to some given function of time $v_s(t)$, regardless of the value of the current flowing through its terminals; for example, $v_s(t) = 2 \sin t$. In particular, $v_s(t)$ may be a constant function such as $v_s(t) = E$, in which case, by analogy with direct current (dc) sources, we shall call the voltage source a dc voltage source. We shall use the symbols shown in Fig. 1-4 to denote an independent voltage source. Observe that a dc voltage source is denoted by the standard battery symbol in order to conform to popular usage.

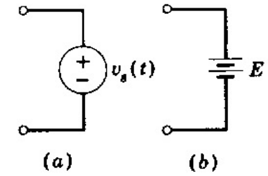


Fig. 1-4. Symbols for an independent voltage source.

Independent current source An independent current source is a two-terminal device whose terminal current i is always equal to some given function of time $i_s(t)$, regardless of the value of the voltage across its terminals. In particular, $i_s(t)$ may be a constant function such as $i_s(t) = I$, in which case we shall call the current source a dc current source. We shall use the symbols shown in Fig. 1-5 to denote an independent current source. Observe that a dc current source is denoted by the same symbol with the exception that $i_s(t)$ is replaced by a constant, I, independent of time.

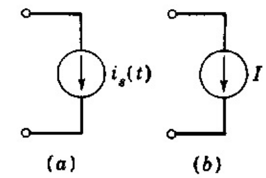


Fig. 1-5. Symbols for an independent current source.

Exercise: It is sometimes convenient to define an *independent flux-linkage source* and an *independent charge source* for the remaining two variables φ and q. State an analogous definition for each.

1-5 CHARACTERIZATION OF A TWO-TERMINAL BLACK BOX

Among the many physical devices of various complexities we shall be concerned in this chapter only with those which possess two *accessible* electrical terminals. Actually, the device may contain more than two terminals but only two of these are accessible to the external world in the sense that the device may be excited only through these terminals. For our purpose, it is convenient to imagine that the device is enclosed in a box and that the two accessible terminals are brought out by two connecting wires as shown in Fig. 1-6*a*. We shall call the resulting system a *two-terminal black box* and shall denote it by the symbol shown in Fig. 1-6*b*. It is important to emphasize that the *content* of the box may be as simple as a light bulb or as complicated as an arbitrary interconnection of other black boxes as shown in Fig. 1-6*c*.

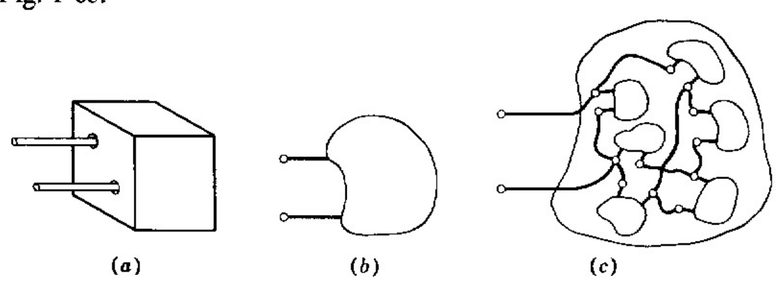


Fig. 1-6. Symbolic representation of a two-terminal black box.

The choice of the term "black box" is quite appropriate here because the box is really black inside in the sense that we cannot see its contents. As a matter of fact, unless we open the box and peep inside, there is no way of determining its contents. However, as engineers, we are not so much interested in the contents of the box as in knowing what the black box can do and how it behaves externally when it is connected with other black boxes into a network. In other words, we are primarily interested in predicting the external behavior of the black box without having to perform any tedious experiment. Our first step toward such an analytical approach is to "characterize" the black box. The concepts involved in characterizing a black box are so important that we pause here to consider a simple but illustrative analogy.

1-5-1 A MECHANICAL BLACK-BOX ANALOGY

Suppose we are given the mechanical black box containing a "spring" as shown in Fig. 1-7*a*. Suppose we did not know the contents of this black box and were asked to predict the behavior of the external terminals when an arbitrary force $f(t)$ is applied to terminal a of the black box with terminal b fixed against the wall. The mechanical variables of interest here are the displacement x (displacement to the right of the initial position 0 is assumed positive), the velocity v (of terminal a), and the force f (positive for tension and negative for compression). Clearly, the only way we can hope to characterize this black box (other than opening the box) is to start performing some experiments. Suppose we begin by applying a constant force $f = A$ and measure the corresponding velocity of terminal a. This would give us a point in the velocity-vs.-force plane (f-v plane). By repeating the above experiment with several values of force f, we obtain the data shown in Fig. 1-7*b*. We might be tempted to draw a smooth curve through

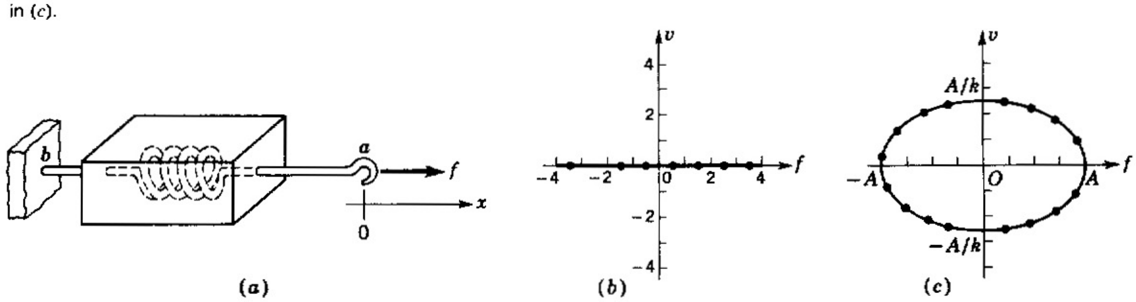


Fig. 1-7. An example illustrating the characterization of a mechanical black box. The data points in the v-vs.-f plane were found to lie on the horizontal axis in (*b*) and on the ellipse in (*c*).

these data points (which in this case happen to be the f axis) and claim to have characterized the black box in the sense that given any constant force f, we can analytically predict the associated velocity. However, a little thought will show that we have not really characterized the black box yet, for if, instead of applying a constant force we apply a slowly varying sinusoidal force, $f(t) = A \sin t$. The above characteristics would predict that $v(t) = 0$. This is of course contrary to what we expect to observe experimentally; namely, $v(t) = (A/k) \cos t$ where k is the "spring constant." We might hope that this inconsistency can be resolved by plotting all points (f,v) satisfying the above equations and obtaining an ellipse as shown in Fig. 1-7*c*. Observe, however, that the length of both axes of the ellipse depends on the amplitude A of the sinusoidal force, and for each value of A we would obtain a corresponding ellipse, so that eventually the entire f-v plane would be filled up with data points. Moreover, even if we can draw an infinite set of ellipses, we would be able to predict the velocity only if $f(t)$ is sinusoidal. Using these ellipses to predict v due to nonsinusoidal $f(t)$ would again yield erroneous answers. Reluctantly, we must admit that our efforts so far have been in vain and that just about the only useful information we obtained from the above experiment is that the black box cannot be characterized by a curve in the f-v plane.

Let us try another set of variables, say the force f and the displacement x, and repeat the experiments. As before, we begin by applying a constant force $f = A$ and measure the corresponding displacement x. Repeating this for various values of f, we obtain the data points shown in Fig. 1-8*a*. If we draw a smooth curve through these points, we obtain a single relationship

$$x = T(f)$$

Before we try to draw any conclusion, however, our previous experience suggests that we repeat the experiment with time-varying forces to see whether the above relationship still holds. Carrying out the proposed experiment with several low-frequency sinusoidal waveforms as before, we find that at any time $t = t_0$, the data point $[f(t_0),x(t_0)]$ always falls on the same curve $x = T(f)$. This is very encouraging, but to be sure, we must try some other nonsinusoidal waveforms for $f(t)$. Again, we find that, provided $f(t)$ does not change very rapidly,[1] the data point at any time also agrees with the curve in Fig. 1-8*a*. Hence, we can now draw the following conclusion: *For any $f(t)$ which does not change rapidly,*

[1] This condition is actually equivalent to the statement that the frequency of the sinusoidal waveform is not very high. This will become obvious after the reader studies signal analysis.

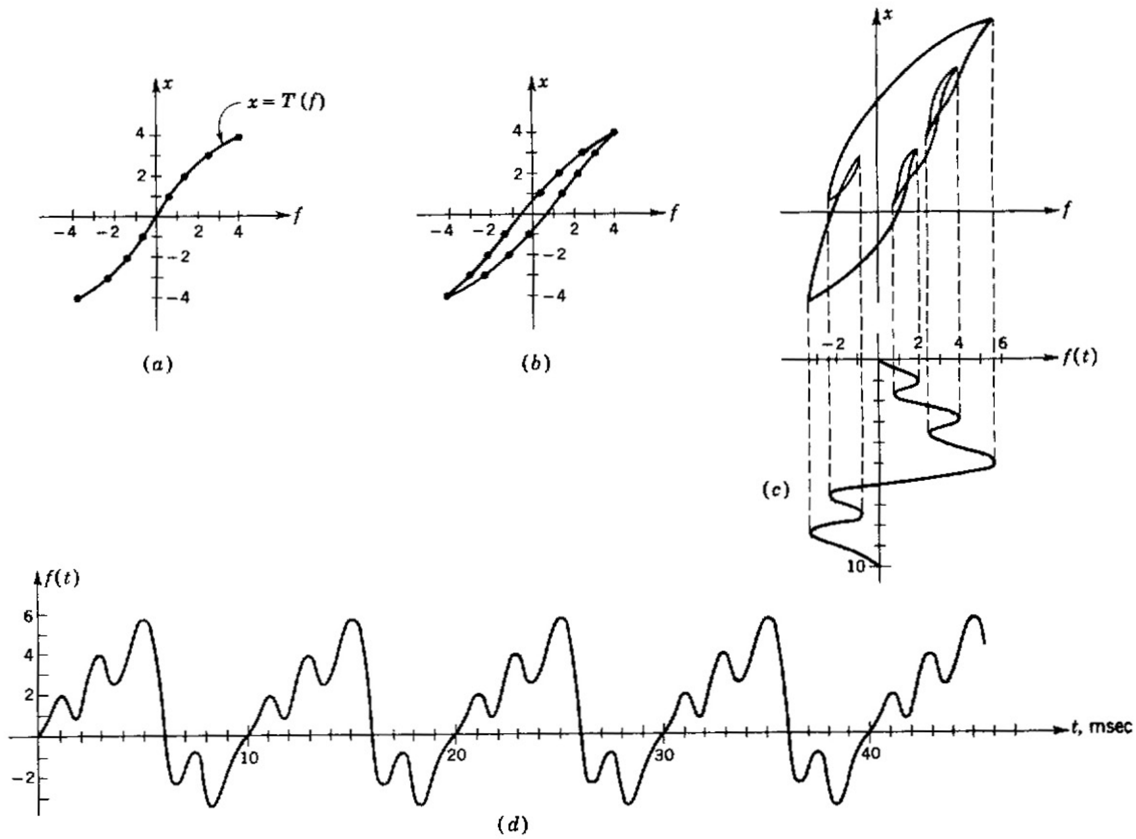


Fig. 1-8. As the frequency of the forcing function $f(t)$ increases, the x-vs.-f characteristic of the mechanical black box changes from a monotonic curve to a hysteresis loop.

the black box can be characterized by the displacement-vs.-force (f-x) curve shown in Fig. 1-8a.

After experiencing the length of time needed to carry out the above experiments, we can now begin to appreciate the utility of such a conclusion; namely, the characterization of the black box permits an analytical solution and thereby eliminates the need to carry out any further experiments.

Observe that our conclusion is based on the assumption that $f(t)$ does not change rapidly. Let us now repeat our experiment with higher-frequency sinusoidal waveforms, as well as with non-sinusoidal waveforms which change rapidly. The experiment shows that as we increase the frequency of the sinusoidal force $f(t)$, the data points begin to deviate (rather slowly at first) from the predicted curve $x = T(f)$. As we increase the frequency further, the data points begin to form a closed loop as shown in Fig. 1-8b, and the area enclosed by the loop tends to increase with frequency. Similarly, we find that if we apply a nonsinusoidal force which changes rapidly with time, the deviation from the curve in Fig. 1-8*a* is even worse. For example, Fig. 1-8*c* shows the *f*-*x* curve corresponding to the high-frequency nonsinusoidal waveform shown in Fig. 1-8*d*. The above experimental result shows that our earlier assumption, that $f(t)$ should not change very rapidly, is indeed necessary. In order to emphasize this restriction, it is a common practice to call the relationship obtained in Fig. 1-8*a* a *static* characteristic curve in contrast to the *dynamic* characteristic curve which corresponds to measurements at higher frequencies. Since the deviation of the measured characteristic curve from the static characteristic increases slowly with frequency, rather than abruptly, it is impossible to pick a definite frequency above which the static characteristic does not hold. Neither is it possible to find a single dynamic characteristic curve which would hold for all high frequencies. Hence, a certain amount of engineering judgment is involved in deciding whether a certain static characteristic curve can be used satisfactorily to solve a given problem. It is encouraging, however, to know that a large percentage of practical networks can indeed be analyzed by using only static characteristics. Moreover, even in cases when the static characteristic fails to give satisfactory solutions, we shall show in the future that we can often patch up the error by including "parasitic elements," namely, elements which are undesirable but which are invariably present in the black box in minute quantities. For the above example, the parasitic element consists of the *mass* associated with the spring. At low frequencies, the mass, being quite small, has relatively no effect on the measured *f*-*x* characteristic. However, as the frequency of the external force $f(t)$ increases, the acceleration of the spring increases, and the inertia force due to the mass becomes appreciable and, in fact, increases as acceleration increases. The deviation of the dynamic characteristic in Fig. 1-8*b* and 1-8*c* from the static characteristic in Fig. 1-8*a* can therefore be attributed to the inertia mass of the spring.

1-5-2 STATIC CHARACTERISTICS OF A TWO-TERMINAL BLACK BOX

The above discussion clearly shows the significance of static characteristics of a black box. Since *all characteristics to be considered in this book are assumed to be static characteristics,* we shall henceforth delete the adjective "static."

Let us now return to the main theme of this section, namely, the characterization of a two-terminal black box. Clearly, the only way we can hope to achieve this is to perform some meaningful external measurements. The only quantities of interest to us are those which can be measured externally. For example, the terminal voltage v and the terminal current i are of primary interest because they can be readily measured. The charge q and the flux linkage φ are also of interest because they can be indirectly measured by *integrating* the measured current waveform $i(t)$ and the measured voltage waveform $v(t)$ in accordance with Eqs. (1-11) and (1-12), respectively. From these measurements, we shall then try to establish a relationship, if there is any, between each pair of *independent* variables. Since the members of the pair of variables i and q are related by Eq. (1-7), they are not independent. Similarly, the variables v and φ are related by Eq. (1-8) and are also not independent. The only remaining combinations consist, therefore, of a relationship between the following variables:

1. Relationship between v and i
2. Relationship between v and q
3. Relationship between i and φ
4. Relationship between q and φ

The last relationship does not occur frequently in practice and has little practical significance. Hence, we shall restrict our attention throughout this book to only the first three cases. These correspond, respectively, to three basic types of two-terminal network elements, namely, a *two-terminal resistor*, a *two-terminal capacitor*, and a *two-terminal inductor*.

Our next step is therefore to plot the data in the v-i, v-q, and i-φ planes, in order to see if the points in any one of these planes can be connected to form a curve. In general, this may not be possible. For example, suppose the two-terminal black box happens to be a capacitance of 1 F. Then from elementary physics, we know that the relationship between v and i is $i = 1(dv/dt)$. But suppose we did not know that the black box contains a capacitance and proceeded to plot the data in the v-i plane. Clearly, it is impossible to expect that a curve can be found which passes through all data points in the v-i plane; in fact, if we take enough measurements, the data points will eventually fill the entire v-i plane. This is easily seen if we apply a voltage source of the form $v(t) = A \sin t$; since $i = dv/dt$, we obtain $i(t) = A \cos t$. Hence at any time $t = t_0$, we obtain a point $(A \sin t_0, A \cos t_0)$ in the v-i plane. Observe next that corresponding to each value of A, the above points form a circle of radius A since $v^2 + i^2 = A^2$. If we vary the value of A from 0 to ∞, we would eventually fill up the v-i plane with data points, and it would be impossible to find a curve passing through these points. On the other hand, if we choose to plot the points in the v-q plane, then these points can be connected by a smooth curve, namely, the line $q = v$. Therefore if a curve can be found which passes through all possible data points in either the v-i, the v-q, or the i-φ plane, then the two-terminal element is completely characterized by that curve.

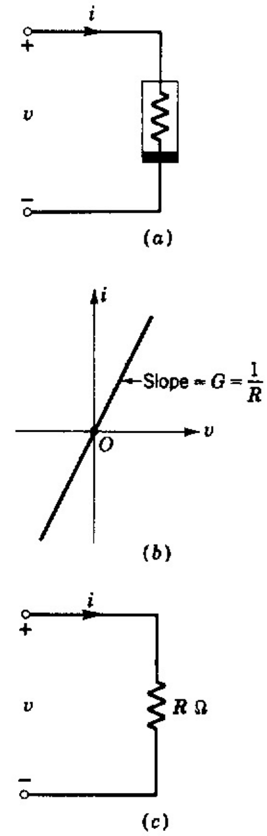


Fig. 1-9. Symbols for a two-terminal resistor.

1-6 TWO-TERMINAL RESISTORS

A two-terminal black box which can be characterized by a curve in the v-i plane is called a *two-terminal resistor* and will be denoted by the symbol shown in Fig. 1-9*a*. Observe that one edge of the symbol is darkened in order to distinguish between the two terminals. This is necessary because the v-i curve measured across the two terminals of a resistor is generally different from that measured across *the same* resistor but with the terminals interchanged (see Prob. 1-1).[1]

1-6-1 LINEAR RESISTORS

Among the infinite variety of v-i curves there is an important subclass which consists of *straight lines passing through the origin* as shown in Fig. 1-9*b*. Resistors of this subclass are called *linear resistors* and will be denoted by the standard symbol shown in Fig. 1-9*c*. Since the v-i curve of a linear resistor is a straight line *through the origin*, it can be described mathematically by $i = Gv$, or $v = Ri$. The constant G represents the slope of the line and is called the *conductance*. The constant R is defined as the reciprocal of G and is called the *resistance*. The practical unit of conductance is the *mho*. The practical unit of resistance is the *ohm* and will be denoted by Ω. A linear resistor is therefore completely characterized by one number, its conductance or its resistance. If the value of the resistance is positive, the linear resistor is said to be a *positive resistor*. Otherwise, it is said to be a *negative resistor*. If $R = 0$, the linear resistor is said to be a *short circuit*. If $R = \infty$, it is said to be an *open circuit*.

[1] In view of the nonsymmetrical nature of this symbol, we may avoid drawing voltage polarity and current direction signs beside the symbol *provided* we agree to assume that the darkened edge is the negative terminal and that the current enters the positive terminal. This convention will be followed in this book.

It is important to differentiate between the terms *resistor* and *resistance;* the former refers to a black box, but the latter refers to a property associated with the black box.[1]

Exercise 1: Explain why it is unnecessary to differentiate between the terminals of the symbol for a linear resistor.

Exercise 2: A certain v-i curve is described by an equation $v = 10i + 5$. Is this a linear resistor?

1-6-2 NONLINEAR RESISTORS

If a resistor is characterized by a v-i curve other than a straight line through the origin, it is called a *nonlinear resistor.* In this case, the resistor can no longer be described by a single number, and hence the entire v-i curve must be given. This may be specified either graphically by a curve or analytically by a mathematical relationship. For example, consider the set of practical two-terminal resistors listed in Table 1-1.[2] Since these components are all commercially available, they have been given names and symbols.[3] Each resistor in this table is characterized graphically by a typical v-i curve usually supplied by the manufacturer. In some cases, it may be possible to derive a mathematical relationship which closely approximates a certain v-i curve. For example, from physical principles one can show that the v-i curve of a vacuum diode can be represented approximately by a 3/2 power law, namely,[4]

$$i = kv^{3/2} \tag{1-13}$$

where k is a constant which depends on the physical dimensions of the internal structure of the diode. Similarly, a semiconductor junction diode can be represented approximately by an exponential law, namely,[5]

$$i = I_0(e^{kv} - 1) \tag{1-14}$$

where I_0 and k are constants which depend on the physical parameters of the diode. One can also sometimes derive an equation which approximates a v-i curve by interpolation and approximation techniques (see Appendix A). For example, the varistor shown in Table 1-1 can be represented approximately by the equation

$$v = \alpha i^{\beta} \tag{1-15}$$

[1] This subtle difference is not universally recognized. In many books, the terms resistor and resistance are used synonymously. In this book, the term resistance refers only to a linear resistor.

[2] To conform with the IEEE standard letter symbols for semiconductor devices (*IEEE Trans. Electron Devices,* vol. Ed-11, no. 8, pp. 392–397), we have chosen the *uppercase letters* V and I in favor of the *lowercase letters* v and i as used in the context. Whenever applicable, we shall also follow the latest IEEE standards for graphic symbols.

[3] In view of its relatively recent origin, the name and symbol for the constant-current diode are not universally used. The same device is sometimes referred to as a *current-limiting diode, a currector, a field-effect diode,* etc. A further discrepancy may be found in that portion of the v-i curve for negative voltages. Depending on how the device is made, the v-i curve for $v < 0$ either approximates an open circuit (horizontal line), *as will be assumed throughout this book,* or a short circuit (vertical line). Fortunately, this discrepancy is usually not important because, as will be shown later, only the portion of the v-i curve in the first quadrant is actually of practical interest. However, in any case, if the v-i curve for $v < 0$ approximates a short circuit, it can always be transformed into the v-i curve shown in Table 1-1 by connecting a junction diode in series

where α and β are constants which can be determined numerically from the curve. In all cases, we must remember that any mathematical relationship is at best an approximation to the actual v-i curve. Moreover, most v-i curves cannot be represented by such simple expressions as those given above. Therefore, the most general and common method to specify element characteristics is to describe the curve in graphical form.

1-6-3 CLASSIFICATION OF v-i CURVES

In order to be able to use the nonlinear resistors effectively in a practical design, it is necessary to classify v-i curves into various categories. For example, the v-i curves of the first three resistors in Table 1-1 have one property in common; namely, for each pair of points (v_1,i_1) and (v_2,i_2) on the curve, we observe that whenever $v_1 > v_2$, then $i_1 > i_2$. Such elements are said to be *strictly monotonically increasing* resistors. An examination of the v-i curves of the zener diode and the constant-current diode shows that they are not strictly monotonically increasing because if we pick a pair of points with voltages $v_1 > v_2$ along the horizontal portions of the v-i curve, then $i(v_1) \not> i(v_2)$. However, these v-i curves have another common property; namely, $i(v_1) \geq i(v_2)$ for any $v_1 > v_2$. Such elements are said to be monotonically (but not strictly) increasing resistors. The v-i curves of the tunnel diode and the remaining resistors below it are not monotonically increasing because each v-i curve has a portion having *negative slopes* ($di/dv < 0$). Such elements are sometimes called *negative-resistance elements*. Another common characteristic of a negative-resistance element is that either the voltage is a multivalued function of current (more than one voltage corresponds to some given value of current) or the current is a multivalued function of voltage (more than one current corresponds to some given value of voltage). In the first case, the current is a single-valued function of the voltage (but not vice versa); that is,

$$i = i(v) \tag{1-16}$$

and is therefore called a *voltage-controlled resistor*. In the second case, it is the voltage which is a single-valued function of current; that is,

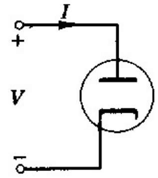
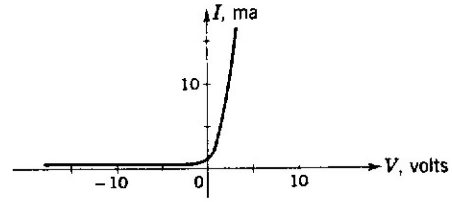
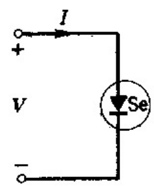
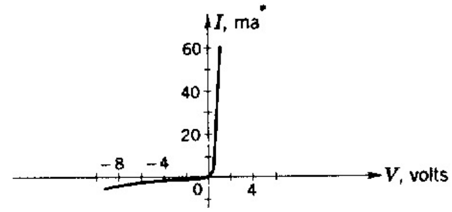
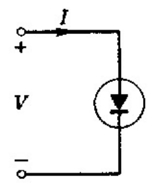
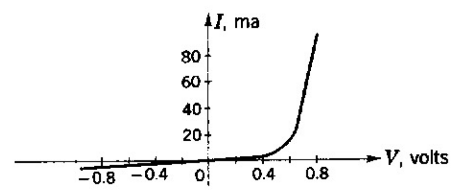
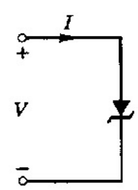
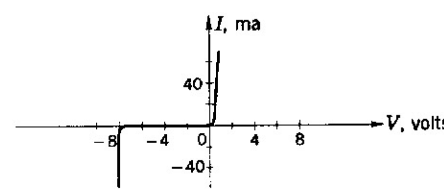
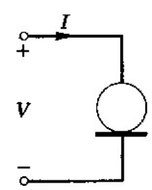
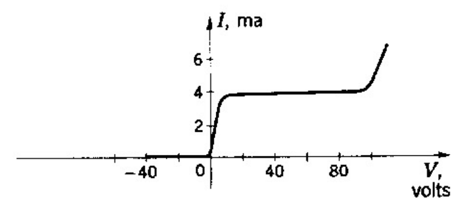
$$v = v(i) \tag{1-17}$$

with the constant-current diode. For more information concerning this device, the reader is referred to J. M. Carroll, "Microelectronic Circuits and Applications," pp. 234 and 235, McGraw-Hill Book Company, New York, 1965; and J. M. Carroll, "Tunnel-Diode and Semiconductor Circuits," pp. 122–128, McGraw-Hill Book Company, New York, 1963.

[4] J. Langmuir, The Effect of Space Charge and Residual Gases on Thermionic Currents in High Vacuum, *Phys. Rev.*, vol. 2, pp. 450–486, 1913.

[5] J. F. Gibbons, "Semiconductor Electronics," McGraw-Hill Book Company, New York, 1966.

TABLE 1-1 *Practical two-terminal resistors.*

Name	Symbol	v-i characteristic curve
Vacuum diode		
Selenium diode		
Semiconductor (junction) diode		
Zener (avalanche, breakdown) diode		
Constant-current diode		

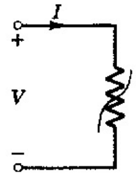
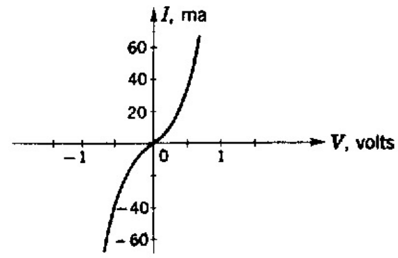
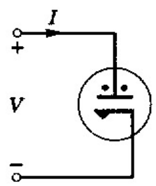
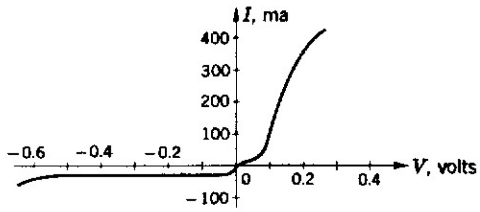
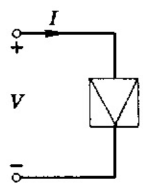
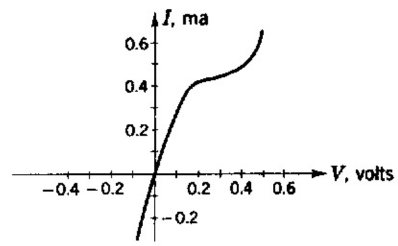
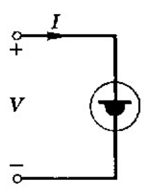
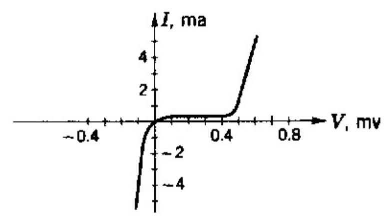
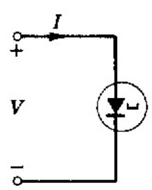
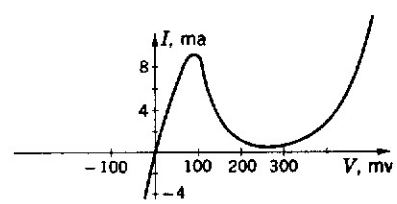
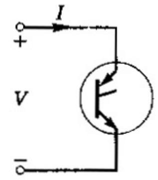
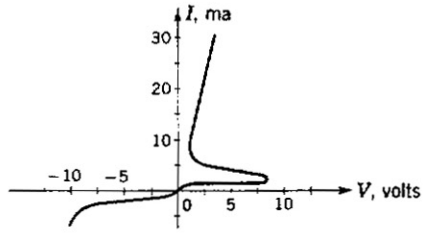
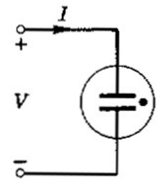
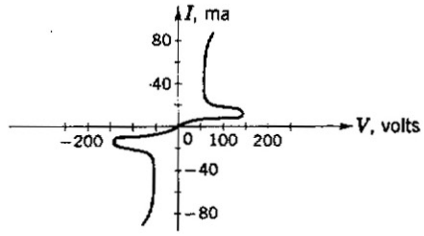
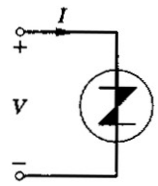
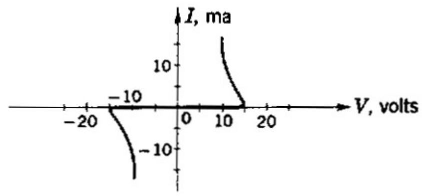
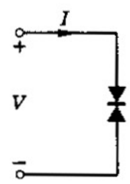
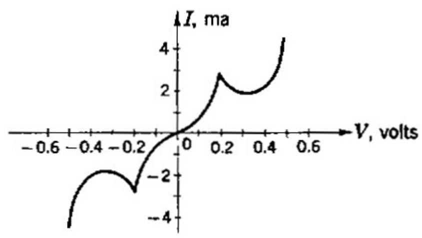
Name	Symbol	v-i characteristic curve
Varistor		
Solion liquid diode		
Tunnel resistor		
Back diode		
Tunnel diode		

TABLE 1-1 (*Continued*)

Name	Symbol	v-i characteristic curve
Four-layer diode		
Glow tube		
Trigger diode		
Superconducting tunnel junction		

and is therefore called a *current-controlled resistor.*[1] For example, the tunnel diode is a voltage-controlled resistor but the glow tube is a current-controlled resistor. Observe that a strictly monotonically increasing resistor is both voltage-controlled and current-controlled and can therefore be described either in the form of Eqs. (1-16) or (1-17).

[1] More generally, any curve in the x-y plane is said to be an x-controlled curve if it is a single-valued function of x and a y-controlled curve if it is a single-valued function of y.

Another important property shared by some v-i curves is their symmetry with respect to the origin. Such elements are called *bilateral resistors* because in this case the two terminals may be interchanged without affecting the v-i curve (see Prob. 1-1). For the resistors listed in Table 1-1, the varistor, the glow tube, the trigger diode, and the superconducting tunnel junction are the only bilateral resistors. The rest are nonbilateral.

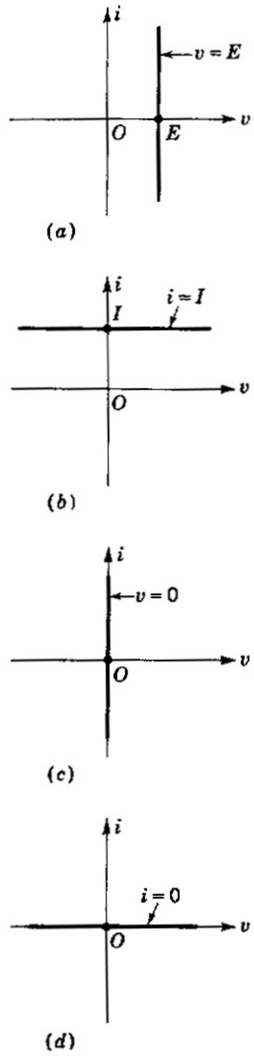


Fig. 1-10. The v-i curves of a dc-voltage source, a dc-current source, a short circuit, and an open circuit have one common property: they consist of either a vertical line or a horizontal line.

Exercise 1: It is sometimes convenient to describe a voltage-controlled resistor by an equation of the form $i = G(v)v$ and a current-controlled resistor in the form $v = R(i)i$. Find the functions $G(v)$ and $R(i)$ in terms of $i(v)$ in Eq. (1-16) and $v(i)$ in Eq. (1-17). Give a geometrical interpretation of $G(v)$ and $R(i)$.

Exercise 2: Is a monotonically (but not strictly) increasing resistor both current-controlled and voltage-controlled? If not, under what condition is it voltage-controlled? When is it current-controlled?

Exercise 3: A resistor which is neither voltage-controlled nor current-controlled is said to be a *multivalued resistor*. Give an example of a multivalued resistor. Can you describe a multivalued resistor in the form of Eq. (1-16) or (1-17)? Explain why. (See Appendix A.)

1-6-4 v-i CURVES OF DC SOURCES AND IDEAL DIODES

On many occasions we shall find it convenient to consider a dc voltage source and a dc current source as nonlinear resistors. This interpretation is valid because, by definition, a dc voltage source with terminal voltage E can be represented by a vertical line $v = E$ as shown in Fig. 1-10*a*. Similarly, a dc current source with terminal current I can be represented by a horizontal line $i = I$, as shown in Fig. 1-10*b*. In the special case where $E = 0$, the v-i curve of Fig. 1-10*a* becomes the $v = 0$ axis as shown in Fig. 1-10*c*. Since this coincides with the v-i curve of a short circuit, *a voltage source with zero terminal voltage is equivalent to a short circuit.* Similarly, when $I = 0$, the v-i curve of Fig. 1-10*b* becomes the $i = 0$ axis, as shown in Fig. 1-10*d*. Since this coincides with the v-i curve of an open circuit, *a current source with zero terminal current is equivalent to an open circuit.* Finally, a two-terminal resistor which does not exist in practice, but which is very useful conceptually, is the ideal diode whose symbol and v-i curve are shown in Fig. 1-11*a* and *b*, respectively. Analytically, an ideal diode is described by

$$
\begin{aligned}
i &= 0 && \text{for all } v < 0 \\
v &= 0 && \text{for all } i > 0 \qquad \text{(1-18)} \\
p &= vi = 0 && \text{for all } v \text{ and } i
\end{aligned}
$$

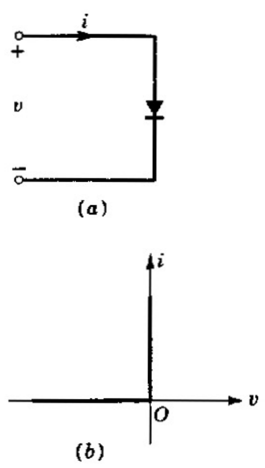


Fig. 1-11. The symbol and v-i curve of an ideal diode.

Observe that the last constraint is introduced to eliminate any point in the fourth quadrant from becoming a part of the v-i curve. It is also important to observe that an ideal diode becomes an open circuit for all $v < 0$ and a short circuit for all $i > 0$.

Before we leave this section, we wish to emphasize that the class of practical nonlinear resistors is not restricted to those listed in Table 1-1. In fact, when we reach Chap. 6, we shall be able to synthesize nonlinear resistors with almost any prescribed v-i curve of practical interest.

Exercise 1: Find the v-i curve of the ideal diode but with its terminals interchanged. Describe this curve analytically.

Exercise 2: A time-varying independent source may be represented by a family of v-i curves with the time t as a parameter. Sketch the v-i curves of a voltage source with terminal voltage $v_s(t) = 2t$ and a current source with terminal current $i_s(t) = 10 \sin \pi t$.

1-6-5 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR RESISTORS

What are nonlinear resistors good for? How do we make use of their v-i curves to design practical electronic gadgets? Do certain types of v-i curves seem more appropriate for one application than another? These are some of the questions that will be answered in the latter part of this book, after we have built up enough theory to understand the basic principles involved in a practical design. However, to satisfy the impatient reader, we shall present in this section a qualitative description of some typical applications. Needless to say, this oversimplified treatment will become more quantitative and precise as the reader gains more ground in the subsequent chapters.

Rectification In many practical applications such as electroplating, the power supply must be restricted to a *single-polarity* voltage or current source. Since the most economical power source is 60-Hz sinusoidal voltage, it is desirable to transform this alternating voltage into a single-polarity voltage. This conversion process is called *rectification,* and any network that carries out the desired transformation is called a *rectifier.* The simplest rectifier consists of an ideal diode in series with a linear resistor, as shown in Fig. 1-12. When the input voltage $v_i(t)$ is positive, the diode becomes a short circuit and $v_o(t) = v_i(t)$. However, when the input voltage $v_i(t)$ is negative, the diode becomes an open circuit and $v_o(t) = 0$.

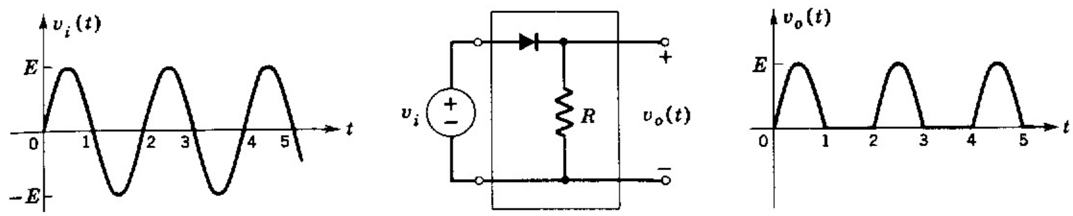


Fig. 1-12. An ideal rectifier converts a sinusoidal input voltage into a single-polarity output voltage.

The result is that the output voltage becomes zero during every other half cycle and is therefore a single-polarity voltage. Of course, this rectifier is an idealized circuit since it uses an ideal diode which does not exist in practice. However, an examination of Table 1-1 suggests that a practical rectifier may be constructed by replacing the ideal diode in Fig. 1-12 by a vacuum diode, a selenium diode, or a semiconductor junction diode.

The above procedure for arriving at a practical design by deriving first an idealized network (which is usually much easier to come by) and then approximating it by a practical circuit is a universal principle of creative design. This principle is based on the intuition that if two networks differ from each other only slightly (e.g., the v-i curves of corresponding resistors differ only slightly), then the corresponding voltage and current waveforms of the two networks must also differ only slightly. Mathematically, this is analogous to the variation of a *continuous* function; namely, a small variation in the value of the independent variable produces a correspondingly small variation in the value of the dependent variable. Because of its practical importance, we shall call the above assumption the small-variation postulate.

Frequency multiplication Another very common application of nonlinear resistors is to convert a low-frequency signal into a high-frequency signal. The ability to do this is instrumental in virtually all communication systems ranging from the simplest walkie-talkie to the most complex telemetry systems between artificial communication satellites. Amazingly, the principle for obtaining frequency multiplication is based on a simple observation from high school trigonometry; namely, the nth power of a sine or cosine function contains higher-harmonic components. For example, $\sin^3 x = \frac{3}{4} \sin x - \frac{1}{4} \sin 3x$, $\cos^4 x = \frac{3}{8} + \frac{1}{2} \cos 2x + \frac{1}{8} \cos 4x$, etc. Hence, if the v-i curve of a nonlinear resistor is described by a polynomial

$$i = a_0 + a_1 v + a_2 v^2 + a_3 v^3 + \cdots + a_n v^n \qquad (1\text{-}19)$$

then upon applying a voltage signal $v = A \sin \omega t$, we obtain, with the help of various standard trigonometric identities, the expression

$$\begin{aligned} i(t) &= a_0 + a_1(A \sin \omega t) + a_2(A \sin \omega t)^2 + \\ &\qquad a_3(A \sin \omega t)^3 + \cdots + a_n(A \sin \omega t)^n \\ &= b_0 + b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t + \cdots + b_n \sin n\omega t \\ &\quad + c_1 \cos \omega t + c_2 \cos 2\omega t + c_3 \cos 3\omega t + \cdots + c_n \cos n\omega t \end{aligned} \tag{1-20}$$

Observe that although the voltage consists of a sinusoidal signal of angular frequency ω, the resulting current contains a constant term b_0, a component of the same frequency ω, and a number of higher-harmonic components $2\omega, 3\omega, \ldots, n\omega$. In practice, any of these harmonic components can be extracted by interposing a network known as a *filter* which essentially suppresses all other components except the desired one.[1] In fact, we could even avoid the use of filters if we could obtain nonlinear resistors with suitable v-i curves. For example, to generate a third-harmonic signal, we apply a well-known trigonometric identity

$$\cos 3x = 4 \cos^3 x - 3 \cos x \tag{1-21}$$

to obtain the desired v-i curve,

$$i = 4v^3 - 3v \tag{1-22}$$

Hence, if $v = \cos \omega t$, then

$$i(t) = 4 \cos^3 \omega t - 3 \cos \omega t = \cos 3\omega t \tag{1-23}$$

which is the desired third harmonic. The next step then consists of finding a practical nonlinear resistor with a v-i curve which approximates Eq. (1-22). Unfortunately, no commercially available resistor is close enough even as an approximation. Hence, it would be necessary to *synthesize* this v-i curve using commercially available resistors as building blocks. The principles and techniques for synthesizing arbitrary v-i curves will be given in Chap. 8.

[1] The design of filters is a very well-developed subject and is usually given in a senior-level course called network synthesis.

Exercise 1: Find the values of the coefficients $b_0, b_1, b_2, \ldots, b_n$ and $c_1, c_2, \ldots, c_n$ in Eq. (1-20) in terms of the constant A and the coefficients $a_0, a_1, a_2, \ldots, a_n$, where $n = 5$.

Exercise 2: Using Eq. (1-21), find the desired v-i curve for converting a 100-volt, 60-Hz sinusoidal voltage into a 10-amp, 180-Hz sinusoidal current.

Exercise 3: Verify the trigonometric identity $\cos^5 x = \frac{5}{8}\cos x + \frac{5}{16}\cos 3x + \frac{1}{16}\cos 5x$ and find the desired v-i curve for converting a 1-ma, 1-kHz sinusoidal current waveform into a 1-volt, 5-kHz sinusoidal voltage waveform.

Frequency mixing Given two sinusoidal waveforms with *commensurate* angular frequencies ω_1 and ω_2 (that is, the ratio ω_1/ω_2 is a rational number), we are frequently interested in generating a new sinusoidal waveform with a frequency given by $(m\omega_1 \pm n\omega_2)$, where m and n are any integers, including zero. Each new frequency corresponding to a given combination (m,n) is called a *beat frequency* and will be denoted by $\omega_{(m,n)}$. One of the most common requirements in signal processing (e.g., a radio receiver or an electronic organ) is the generation of appropriate beat frequencies.[1] We shall now demonstrate that in order to generate beat frequencies, it is necessary to perform a nonlinear operation. Again, the basis for doing this is given by the well-known trigonometric identities:

$$\sin x \sin y = \tfrac{1}{2}[\cos(x-y) - \cos(x+y)] \tag{1-24}$$

and

$$\sin x \cos y = \tfrac{1}{2}[\sin(x+y) + \sin(x-y)] \tag{1-25}$$

To demonstrate how we generate beat frequencies, consider applying two voltage sources $v_1 = A\sin\omega_1 t$ and $v_2 = B\sin\omega_2 t$ in series with a nonlinear resistor with a v-i curve given by $i = v^3$. The current $i(t)$ is given by

$$\begin{aligned} i(t) &= (A\sin\omega_1 t + B\sin\omega_2 t)^3 \\ &= A^3\sin^3\omega_1 t + 3A^2B\sin^2\omega_1 t\sin\omega_2 t \\ &\qquad + 3AB^2\sin\omega_1 t\sin^2\omega_2 t + B^3\sin^3\omega_2 t \end{aligned}$$

If we now apply Eq. (1-25) and a number of standard trigonometric identities, we obtain, upon simplification, the expression

$$\begin{aligned} i(t) &= (a_1\sin\omega_1 t + b_1\sin\omega_2 t) + (a_2\sin 3\omega_1 t + b_2\sin 3\omega_2 t) \\ &\quad + [a_3\sin(\omega_2 - 2\omega_1)t + b_3\sin(\omega_2 + 2\omega_1)t] \\ &\qquad + [a_4\sin(\omega_1 - 2\omega_2)t + b_4\sin(\omega_1 + 2\omega_2)t] \end{aligned} \tag{1-26}$$

[1] The beat frequency is also called a *sideband frequency* and the collection of all beat frequencies is usually called *sidebands*. The definitions of beat frequency and sidebands are meaningful even if ω_1 and ω_2 are not commensurate with each other.

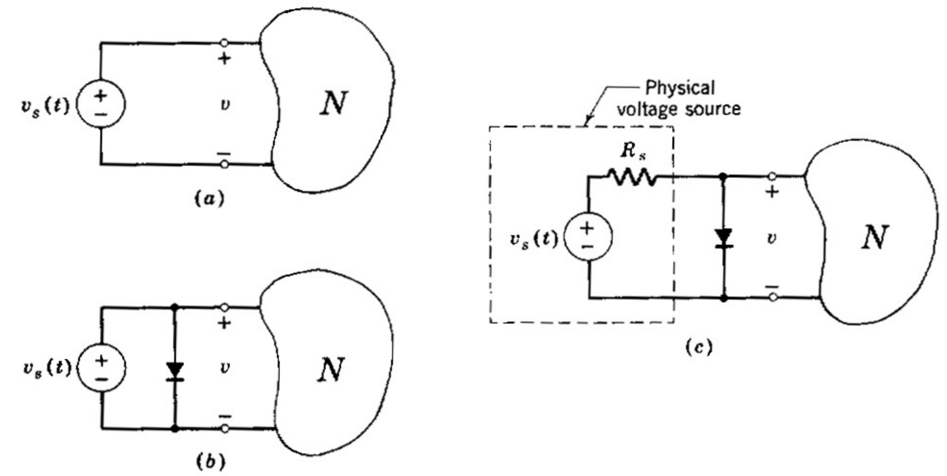
where the coefficients a_j, b_j are functions of A and B. Observe that in addition to sinusoidal terms having the same frequencies as the driving sources, the current $i(t)$ also contains the third-harmonic terms with frequencies $3\omega_1$ and $3\omega_2$ and the beat-frequency terms with frequencies $(\omega_2 \pm 2\omega_1)$ and $(\omega_1 \pm 2\omega_2)$. In the more general case where the v-i curve is described by a polynomial, we can expect, in general, sinusoidal terms with harmonic frequencies $m\omega_1$ and $n\omega_2$, as well as beat frequencies $m\omega_1 \pm n\omega_2$. In practice, any one of these beat frequencies may be extracted through a filter. This principle is widely used in telephone systems.

Exercise 1: Give an example of a pair of sinusoidal waveforms with incommensurate (i.e., not commensurate) frequencies. Is the sum of these two waveforms periodic?

Exercise 2: A speech synthesizer is an electronic system designed to simulate the human voice. An important component of this system is a mixer for generating as many beat frequencies as possible. Assuming that the v-i curve of the resistor is described by a polynomial, what must the general form of the polynomial be in order to generate beat-frequency terms with m and n equal to 0, ± 1, ± 2, and ± 3?

Limiting Any nonlinear resistor R with a v-i curve containing a (nearly) vertical segment can be used to *limit* the voltage across a two-terminal black box connected in parallel with R. For example, we can limit the terminal voltage across the black box N shown in Fig. 1-13a to nonpositive values by connecting an ideal diode across N as shown in Fig. 1-13b. This is because by definition, the voltage across an ideal diode is given by $v \leq 0$.

Fig. 1-13. The voltage across the black box N is constrained to nonpositive values by connecting an ideal diode in parallel with N.



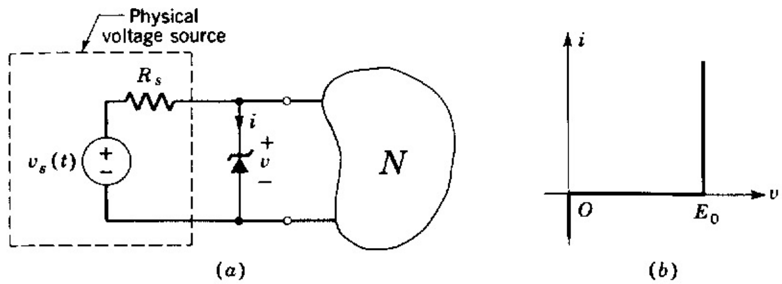


Fig. 1-14. The voltage across N is limited to a maximum value equal to the constant voltage E_0 of the zener diode.

A paradoxical situation arises when one questions what happens if a voltage source with a positive terminal voltage (say $v_s = 2$ volts) is connected across the network shown in Fig. 1-13*b*. By definition, the terminal voltage of this voltage source must remain constant regardless of the external network connected across it. But also by definition, the voltage across an ideal diode cannot be positive. The basic problem here is that we are connecting two incompatible ideal elements in parallel, thereby rendering the definitions inconsistent. In other words, this paradox arises because of an overidealization. It is no different from many paradoxes of a similar nature, most notably among which is the paradox: "What happens if one connects a short circuit across a voltage source with a nonzero terminal voltage?" The best way to resolve this type of paradox is to exclude all such incompatible connections. But how can we forbid anyone from making an incompatible connection in practice? The answer is that there is no such thing as an incompatible connection *in practice* because there are no such things as an ideal voltage source and an ideal diode. Any physical voltage source has a small internal resistance R_s in series with it, as shown in Fig. 1-13*c*. Once we introduce R_s, the paradox disappears because whenever $v_s(t)$ becomes positive, the diode becomes a short circuit and the entire voltage appears across R_s. Hence, the voltage across N can never be positive.

The same principle can be applied to limit the voltage across N from exceeding a prescribed value E_0. For example, if we connect a zener diode with a constant voltage $E_z = E_0$ across N as shown in Fig. 1-14*a*, then from the v-i curve of the zener diode shown in Fig. 1-14*b* (observe that the reference polarity and directions are opposite to those shown in Table 1-1) it is clear that $0 \leq v \leq E_0$. This circuit is commonly used for overload protection. For example, in a typical application, the black box N consists of a sensitive instrument (such as a voltmeter) whose maximum permissible voltage is equal to E_0.

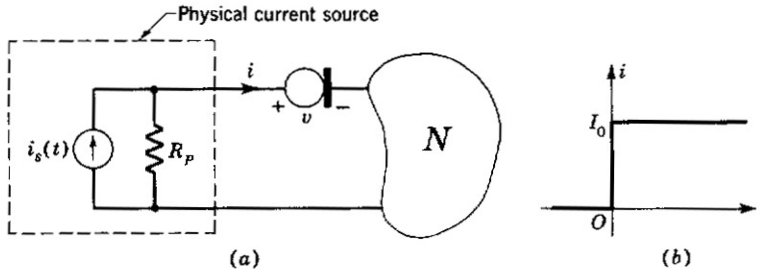
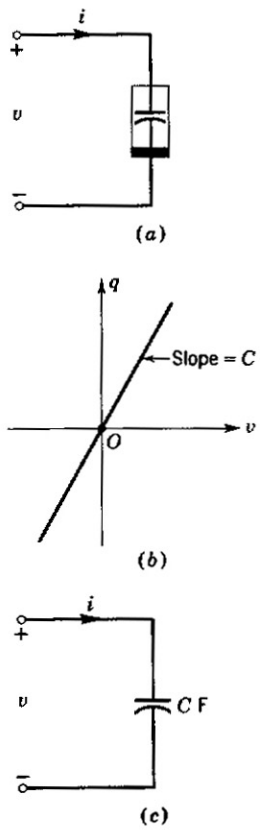


Fig. 1-15. The current entering N is limited to a maximum value equal to the constant current I_0 of the constant-current diode.

By analogous reasoning, any resistor R with a v-i curve containing a (nearly) horizontal segment can be used to *limit* the current entering a black box connected in series with R. For example, if we connect a constant-current diode with constant current I_0 in series with N as shown in Fig. 1-15*a*, then from the v-i curve of the constant-current diode shown in Fig. 1-15*b* it is clear that $0 \le i \le I_0$ (the resistor R_p is introduced to avoid a similar paradox).

Fig. 1-16. Symbols for a two-terminal capacitor.



Exercise 1: The maximum permissible range of voltages of a hypersensitive instrument is given by $-10 \le v \le 5$. Design an overload protection circuit and specify the v-i curve of any nonlinear resistor used in the circuit.

Exercise 2: Explain what happens if $i_s(t) > I_0$ in the circuit shown in Fig. 1-15. Replace the constant-current diode with an appropriate nonlinear resistor so as to limit the terminal current entering N to $|i| < 20$ ma.

1-7 TWO-TERMINAL CAPACITORS

A two-terminal black box which can be characterized by a curve in the v-q plane is called a *two-terminal capacitor* and will be denoted by the symbol shown in Fig. 1-16*a*. Observe that one edge of this symbol is darkened for the same reason as it was for the resistor.

1-7-1 LINEAR CAPACITORS

An important subclass of capacitors can be characterized by a straight line through the origin of the v-q plane, as shown in Fig. 1-16*b*. This subclass is called *linear capacitors* and will be denoted by the conventional symbol shown in Fig. 1-16*c*. A linear capacitor can be described analytically by

$$q = Cv \tag{1-27}$$

where the constant C represents the slope of the straight line and is called the *capacitance* associated with the capacitor. The unit of capacitance is the *farad* and will be denoted by F. To find the current entering a linear capacitor, we substitute Eq. (1-27) for q in Eq. (1-7) and obtain

$$i(t) = C\frac{dv(t)}{dt} \tag{1-28}$$

A linear capacitor is therefore completely characterized by one number, namely, its capacitance. Again, we would differentiate between the terms *capacitor* and *capacitance.*

1-7-2 NONLINEAR CAPACITORS

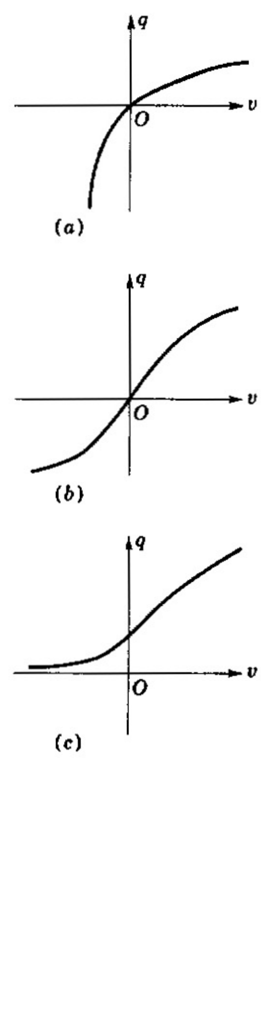
If a capacitor is characterized by a v-q curve other than a straight line through the origin, it is called a *nonlinear capacitor.* In this case, the capacitor can no longer be described by a single number, and hence the entire v-q curve must be given. An example of a practical, nonlinear capacitor is the metal-oxide-semiconductor (MOS) capacitor whose v-q curve is shown in Fig. 1-17*a*. This nonlinear capacitor is used quite extensively in integrated circuits, where the conventional linear capacitor becomes impractical to fabricate. Although there are at present only a few practical nonlinear capacitors available commercially, it is expected that more will become available in the near future. In fact, as will be shown in Chap. 3, it is possible to synthesize a capacitor with any prescribed v-q curve with the help of a new network component called the *mutator.*

There are other reasons for studying nonlinear capacitors. One reason is that components of many physical and biological systems behave in a manner analogous to that of a nonlinear capacitor. Hence the study of such systems can often be achieved by constructing an electrical network model to simulate the behavior of these systems. A simple example is the displacement-vs.-force curve of the nonlinear spring shown in Fig. 1-7*a*. This mechanical element is usually modeled by a nonlinear capacitor with a similar v-q curve, as shown in Fig. 1-17*b*. Another example is given by the volume-vs.-pressure curve of the ventilatory part of the human respiratory system. This biological component can be modeled by an analogous v-q curve as shown in Fig. 1-17*c*.

We shall denote the v-q curve of a nonlinear capacitor by

$$q = q(v) \tag{1-29}$$

Fig. 1-17. The v-q curves of three typical nonlinear capacitors.



if it is voltage-controlled, and by

$$v = v(q) \tag{1-30}$$

if it is charge-controlled. For a voltage-controlled capacitor, the current entering the capacitor can be expressed in a form analogous to Eq. (1-28); thus

$$i(t) = \frac{dq(t)}{dt} = \frac{dq(v)}{dv}\frac{dv(t)}{dt}$$

or

$$i(t) = C(v(t))\frac{dv(t)}{dt} \tag{1-31}$$

where

$$C(v) \equiv \frac{dq(v)}{dv} \tag{1-32}$$

is called the *incremental capacitance* of the capacitor. Notice that the incremental capacitance is a function of the capacitor voltage and becomes a constant only in the case of a linear capacitor.

Exercise 1: A typical nonlinear capacitor is characterized by the v-q curve $q = kv^{3/2}$, where k is a physical parameter. (*a*) Find the incremental capacitance $C(v)$. (*b*) If the applied voltage is given by $v(t) = \frac{1}{4}\cos^2 t$, find the charge $q(t)$ and the current $i(t) \equiv dq(t)/dt$. (*c*) Calculate $i(t)$ by using Eq. (1-31).

Exercise 2: An abrupt-junction diode is a semiconductor p-n junction which behaves like a capacitor, provided the voltage across the junction is less than 0.5 volt. Its incremental capacitance is given by $C(v) = k(\phi - v)^{-1/n}$, where k, ϕ, and n are constants which depend upon the parameters of the device. (*a*) Plot the incremental capacitance on logarithmic paper for the range $-100 < v < 0.5$ volt. (Assume $k = 80 \times 10^{-12}$, $\phi = 0.5$, and $n = 2$.) (*b*) What are the maximum and the minimum values of the capacitance (in picofarads or 10^{-12} F) within this range of applied voltage? (*c*) Do you have sufficient information to recover the v-q curve? If not, what additional information do you need?

1-7-3 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR CAPACITORS

What are nonlinear capacitors good for? Can they do useful things which nonlinear resistors cannot? The answer to the second question is obviously yes, for otherwise we would not be studying them. In addition to being able to do a number of things described earlier for resistors, a nonlinear capacitor can do better in certain cases. Although we do not yet have the background necessary to demonstrate this assertion, suffice it to say that both nonlinear resistors and capacitors are capable of generating higher harmonics. However, with an appropriate design, it is possible to extract more output power in any given harmonic component from a nonlinear capacitor than from a nonlinear resistor. This means that a nonlinear-capacitor-frequency multiplier has a higher efficiency than a nonlinear-resistor-frequency multiplier. In addition to this application, a few of the many other useful functions are briefly described as follows.

Frequency division In many practical systems, it is desirable to convert a given sinusoidal signal of frequency ω_1 into another sinusoidal signal of a lower frequency ω_2; namely, $\omega_2 = \omega_1/n$, where n is an integer. In this case, the lower-frequency output signal is said to be a *subharmonic* of the higher-frequency output signal. It can be shown that *a nonlinear resistor cannot generate subharmonics.* To demonstrate that a nonlinear capacitor can generate a subharmonic signal, consider a nonlinear capacitor whose incremental capacitance is given by

$$C(v) = \left[\frac{1 - \sqrt{1 - v^2}}{2(1 - v^2)}\right]^{1/2} \tag{1-33}$$

If we apply a voltage $v(t) = \sin \omega t$ across this capacitor, the current $i(t)$ can be calculated from Eqs. (1-31) and (1-33); thus

$$\begin{aligned} i(t) &= \left[\frac{1 - \sqrt{1 - \sin^2 \omega t}}{2(1 - \sin^2 \omega t)}\right]^{1/2} (\omega \cos \omega t) \\ &= \left(\frac{1 - \cos \omega t}{2 \cos^2 \omega t}\right)^{1/2} (\omega \cos \omega t) \\ &= \omega \sqrt{\frac{1 - \cos \omega t}{2}} = \omega \sin \frac{\omega}{2} t \end{aligned} \tag{1-34}$$

Hence, the output current is a sinusoid with frequency equal to half the original frequency. The phenomenon of subharmonic generation by a nonlinear capacitor has been utilized in many practical applications. One application consists of utilizing the two "distinct" frequencies as the two distinct states in designing a digital computer. Another interesting application consists of converting the high-frequency output of a laser beam into a lower-frequency signal.

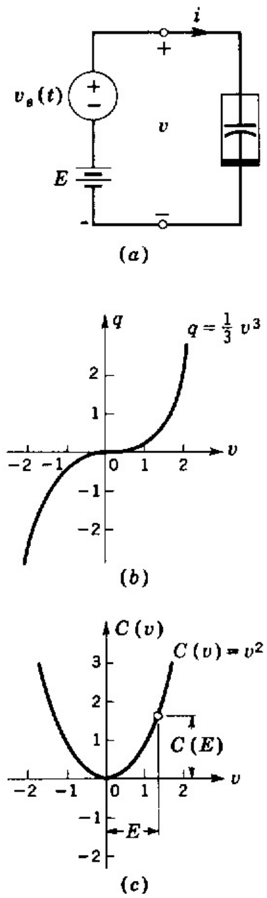


Fig. 1-18. A nonlinear capacitor can be used as a tuning element by varying the dc voltage E across the capacitor.

Parametric amplifier Just as with nonlinear resistors, it is possible to generate beat-frequency components by applying two sinusoidal signals of frequencies ω_1 and ω_2 in series with a nonlinear capacitor. It can be shown that if one of the two signals (say $v_1 = A \sin \omega_1 t$) is very weak while the other signal (say $v_2 = B \sin \omega_2 t$) is very strong, it is possible to extract (with the help of filters) the signal with frequency ω_1 and at the same time greatly amplify its amplitude, say $v_o = 1{,}000\, A \sin \omega_1 t$. The result is that we have an amplifier. For reasons that we are not equipped to elaborate here, this amplifier is called a *parametric amplifier.* It is widely used in artificial satellites because it has some definite advantages over conventional amplifiers.

Electronic tuning Suppose we connect a voltage source $v_s(t)$ and a battery with terminal voltage E in series with a nonlinear capacitor as shown in Fig. 1-18*a*. For simplicity, let the v-q curve be given by $q = \frac{1}{3} v^3$ as shown in Fig. 1-18*b*. Then its incremental capacitance is given by $C(v) = v^2$, as shown in Fig. 1-18*c*. Now in many electronic systems, such as a radio receiver, the signal $v_s(t)$ is very small (say, a few millivolts) compared with the value of the dc voltage E. Hence, for most practical purposes, the incremental capacitance

$$C(v) = C(v_s(t) + E) \approx C(E) \tag{1-35}$$

can be considered to depend *only* on the value of E. In this case, Eq. (1-31) becomes

$$i(t) = C(E)\frac{dv}{dt} \tag{1-36}$$

Since $C(E)$ is no longer a function of time, Eq. (1-36) is identical with Eq. (1-28) which describes a linear capacitor. The only difference is that we can change the value of the capacitance by simply changing the value of E. This observation is of great practical importance. One immediate application is in the area of *electronic tuning.* The conventional way to tune a radio receiver from one station to another is to turn a knob which moves the tuning dial. Any one who opens up the cover of a radio receiver would recognize that this tuning knob is used to rotate the plates of an air capacitor, thereby changing the value of its capacitance. In other words, the standard tuning process consists of adjusting the value of a capacitor *mechanically.* This operation can now be replaced by a nonlinear capacitor connected as shown in Fig. 1-18*a* where the tuning is accomplished by adjusting the voltage E. This method is clearly far superior to the use of bulky air capacitors. In fact, this technique of electronic tuning is fast becoming a standard method in electronic systems.

Exercise 1: Find the incremental capacitance $C(v)$ required to generate a 30-Hz subharmonic sinusoidal current waveform from an input voltage $v(t) = 100 \cos 120\pi t$. HINT: Make use of the trigonometric identity

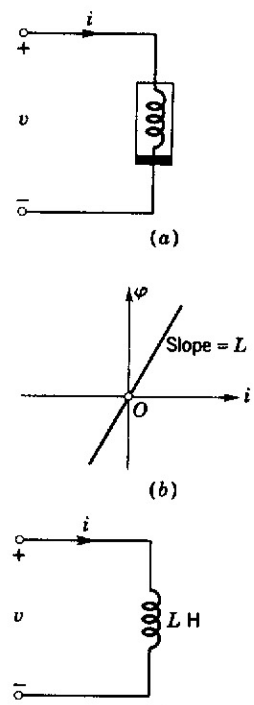
$$\cos\frac{x}{2} = \sqrt{\frac{1+\cos x}{2}}$$

Exercise 2: A common nonlinear capacitor used for electronic tuning is the *varactor diode*. It is characterized by a v-q curve $q(v) = -(3/2)C_0\phi_0(1 - v/\phi_0)^{2/3}$, where C_0 and ϕ_0 are constants which vary from device to device. When $v = 0$, the incremental capacitance was measured to be equal to 60 pF. (*a*) Derive the incremental capacitance $C(v)$. (*b*) If $\phi_0 = 0.35$, find the range of the input voltage required to tune the capacitance from 5 to 100 pF. To operate the varactor as a nonlinear capacitor, the voltage must not exceed 0.35 volt.

1-8 TWO-TERMINAL INDUCTORS

A two-terminal black box which can be characterized by a curve in the i-φ plane is called a *two-terminal inductor* and will be denoted by the symbol shown in Fig. 1-19*a*. The darkened edge of this symbol has the same significance as before.

Fig. 1-19. Symbols for a two-terminal inductor.



1-8-1 LINEAR INDUCTORS

An important subclass of inductors can be characterized by a straight line through the origin of the i-φ plane as shown in Fig. 1-19*b*. This subclass is called *linear inductors* and will be denoted by the conventional symbol shown in Fig. 1-19*c*. A linear inductor can be described analytically by

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

where the constant L represents the slope of the straight line and is called the *inductance* associated with the inductor. The unit of inductance is the *henry* and will be denoted by H. To find the voltage across a linear inductor, we substitute Eq. (1-37) for φ in Eq. (1-8) and obtain

$$v(t) = L\frac{di(t)}{dt} \tag{1-38}$$

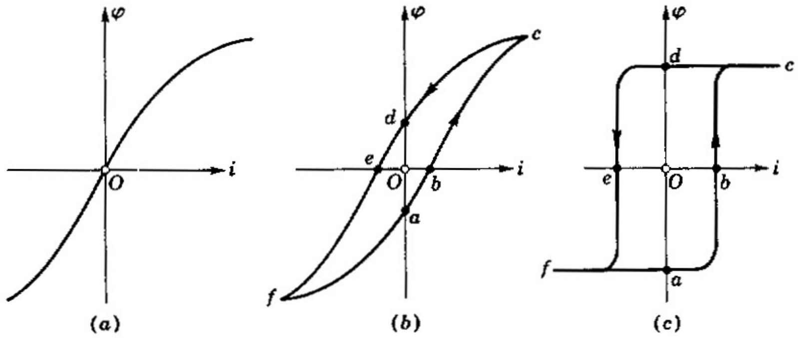


Fig. 1-20. The i-φ curve of three practical nonlinear inductors.

A linear inductor is therefore completely characterized by one number, namely, its inductance. Again, we would differentiate between the terms *inductor* and *inductance.*

1-8-2 NONLINEAR INDUCTORS

If an inductor is characterized by an i-φ curve other than a straight line through the origin, it is called a *nonlinear inductor.* In this case, the inductor can no longer be described by a single number, and hence the entire i-φ curve must be given. For example, Fig. 1-20a shows the i-φ curve of a typical nonlinear inductor.

Another common nonlinear inductor consists of a coil wound around an iron core. Its i-φ curve (obtained by applying a sinusoidal current excitation) is shown in Fig. 1-20b. This curve is a multivalued function of both i and φ and is commonly referred to as the *hysteresis loop.*[1] Observe that starting at point a with $i = 0$, the flux linkage φ increases with i along the path a-b-c. Upon reaching point c when φ attains its maximum value, the flux linkage φ does not retrace the original path. Instead, it decreases with the current i along the path c-d-e-f. Upon reaching point f when i attains its minimum value, the flux linkage φ returns to point a to complete the loop. The shape of the hysteresis loop depends on the type of material used for the core. For certain materials, the hysteresis loop is almost rectangular, as shown in Fig. 1-20c.

We shall denote the i-φ curve of a nonlinear inductor by

$$\varphi = \varphi(i) \tag{1-39}$$

if it is current-controlled, and by

$$i = i(\varphi) \tag{1-40}$$

[1] Actually, this hysteresis loop is a valid description only under the assumption that the current waveform is sinusoidal. For other periodic excitations, the hysteresis loop becomes much more complicated. A *complete* characterization of elements described by hysteresis loops is a very difficult and still unsolved problem.

if it is flux-controlled. In the case of a current-controlled inductor, the voltage across the inductor can be expressed in a form analogous to Eq. (1-38); thus

$$v(t) = \frac{d\varphi(t)}{dt} = \frac{d\varphi(i)}{di}\frac{di(t)}{dt}$$

or

$$v(t) = L(i(t))\frac{di(t)}{dt} \tag{1-41}$$

where

$$L(i) \equiv \frac{d\varphi(i)}{di} \tag{1-42}$$

is called the *incremental inductance* of the inductor. Notice that for a linear inductor, the incremental inductance coincides with the inductance, as it should.

Exercise 1: The i-φ curve of a certain nonlinear inductor can be represented approximately by the cubic equation $\varphi = i^3$. If the inductor is connected across a current source with terminal current $i_s(t) = \sin t$, find and sketch the incremental inductance $L(i)$ and the inductor voltage $v(t)$.

Exercise 2: An inductor is said to be the "dual" of a capacitor, and vice versa, because there exists a one-to-one correspondence between the two elements. Exhibit a list of corresponding quantities.

1-8-3 SOME PRACTICAL APPLICATIONS OF TWO-TERMINAL NONLINEAR INDUCTORS

What are nonlinear inductors good for? Where are they used in practice? To answer these questions would again require more background than we have at present. However, it is instructive to describe a few simple applications.

Frequency conversion Just as is true of capacitors, a nonlinear inductor is capable of generating both *harmonics* and *subharmonics* of a given sinusoidal signal. It can be shown to have the same efficiency as does a nonlinear capacitor. This property is widely used in telephone systems.

Memory and storage Consider the rectangular hysteresis curve shown in Fig. 1-20c. Observe that when $i = 0$, φ may assume either one of two distinct values (point a or point d) depending on the previous history of the excitation current. These two distinct states can be used to represent the two states (0 and 1) in a digital computer. When many of these elements are combined properly, the result is a "memory" or "storage" device to store present information for future use. While there are many other candidates, this memory device has some significant advantages. One is that in both states $i = 0$, and hence no power is being consumed. Since hundreds and thousands of these elements are used in a practical computer, the saving in power cost is enormous.

1-9 ENERGY AND POWER

The energy flow into a two-terminal black box during any time interval (t_0,t_1) is by definition the time integral of power from t_0 to t_1; namely,

$$w(t_0,t_1) = \int_{t_0}^{t_1} v(t)i(t)\,dt \tag{1-43}$$

Since $w(t_0,t_1)$ is a *relative* quantity depending on the time interval (t_0,t_1), it is convenient for us to define another related but *absolute* quantity by letting t_0 equal zero and t_1 approach infinity, and then take the average of the energy flow over the entire infinite time interval; namely,

$$P_{av} \equiv \lim_{t_1\to\infty} \frac{w(0,t_1)}{t_1} \tag{1-44}$$

Since the quantity P_{av} has the dimension of energy per second, it is called the average power. Substituting Eq. (1-43) into Eq. (1-44), we obtain the explicit expression

$$P_{av} = \lim_{t_1\to\infty} \frac{1}{t_1} \int_0^{t_1} v(t)i(t)\,dt \tag{1-45}$$

To illustrate the use of this formula, let us calculate the average power entering a 4-Ω linear resistor due to an applied voltage $v(t) = 2 \sin \pi t$; thus

$$P_{av} = \lim_{t_1\to\infty} \frac{1}{t_1} \int_0^{t_1} (2 \sin \pi t)\left(\frac{2 \sin \pi t}{4}\right) dt \tag{1-46}$$

$$P_{av} = \lim_{t_1\to\infty} \left(½ - \frac{\sin 2\pi t_1}{4\pi t_1}\right) = ½$$

In the case where the voltage $v(t)$ and current $i(t)$ are periodic functions with commensurate periods T_v and T_i, respectively, the power $p(t) = v(t)i(t)$ will also be periodic. However, the period of $p(t)$ is not necessarily equal to T_v or T_i. For the example considered in Eq. (1-46), $T_v = T_i = 2$, but the period of $p(t)$ is 1. If we denote the *minimum* period of $p(t)$ by T, then

$$p(t + nT) = v(t + nT)i(t + nT) = p(t) \tag{1-47}$$

In this case, it is more convenient to let $t_1 = nT$ and rewrite Eq. (1-45) in the equivalent form:

$$\begin{aligned} P_{av} &= \lim_{n\to\infty} \frac{1}{nT} \int_0^{nT} v(t)i(t)\,dt \\ &= \lim_{n\to\infty} \frac{1}{nT}\left[\int_0^{T} v(t)i(t)\,dt + \int_T^{2T} v(t)i(t)\,dt + \cdots + \int_{(n-1)T}^{nT} v(t)i(t)\,dt\right] \\ &= \lim_{n\to\infty} \frac{1}{nT}\left[n\int_0^{T} v(t)i(t)\,dt\right] \\ &= \lim_{n\to\infty} \frac{1}{T} \int_0^{T} v(t)i(t)\,dt \end{aligned}$$

Since the variable n no longer appears in this integral, the limit operation is superfluous and can be removed. Hence, *for periodic signals*, the average power can be written in the following simplified but equivalent form:

$$P_{av} = \frac{1}{T}\int_0^T v(t)i(t)\,dt = \frac{w(0,T)}{T} \tag{1-48}$$

where T is the minimum period of $v(t)i(t)$. Applying this formula to the same example considered in Eq. (1-46), we obtain

$$P_{av} = \frac{1}{1}\int_0^1 (2 \sin \pi t)\left(\frac{2 \sin \pi t}{4}\right) dt = \frac{1}{2}$$

as we should.

Exercise: The voltage and current waveforms of a two-terminal black box are given, respectively, by $v = \sin (3.14)t$ and $i = \sin \pi t$. (*a*) Show that even though both $v(t)$ and $i(t)$ are periodic, the power $p(t)$ is not periodic. (*b*) For most practical purposes, $p(t)$ is said to be "almost periodic." Explain why.

The three expressions given by Eqs. (1-43), (1-45), and (1-48) are valid for any two-terminal black box. Let us now consider the special cases where the black box consists of a single nonlinear resistor, capacitor, or inductor. In so doing, we shall be able to derive a number of useful relationships. We shall also be able to draw some very important physical interpretations. Let us consider the three cases one at a time.

Case 1: Two-terminal nonlinear resistor Consider the nonlinear resistor shown in Fig. 1-21*a* and the three common types of v-i curves shown in Fig. 1-21*b*, *c*, and *d*. The v-i curve can be described in the functional form by $i = i(v)$ if it is voltage-controlled, or by $v = v(i)$ if it is current-controlled. A strictly monotonically increasing v-i curve can obviously be described by either $i = i(v)$ or $v = v(i)$. Accordingly, the instantaneous power flow $p_R(t)$, energy flow $w_R(t_0,t_1)$, and average power $P_{R_{av}}$ can be determined and are tabulated in Table 1-2 for these three cases.

Observe that corresponding to any operating point Q at any time t, the instantaneous power $p_R(t)$ is simply equal to the area of the shaded rectangles shown in Fig. 1-21. This power must, of course, come from the energy supplied by the external circuit connected across the resistor. From Table 1-2 we observe that the expressions for $p_R(t)$, $w_R(t_0,t_1)$, and $P_{R_{av}}$ depend on two pieces of information, namely,

1. The v-i curve
2. The voltage waveform $v(t)$ or the current waveform $i(t)$

Hence, in order to find out what happens to the power that enters the resistor, we must be given these two pieces of information. For

Fig. 1-21. The instantaneous power absorbed by a nonlinear resistor at any time t_0 is equal numerically to the area of the rectangle formed by the v,i axes and a vertex Q with coordinates $(v(t_0),i(t_0))$.

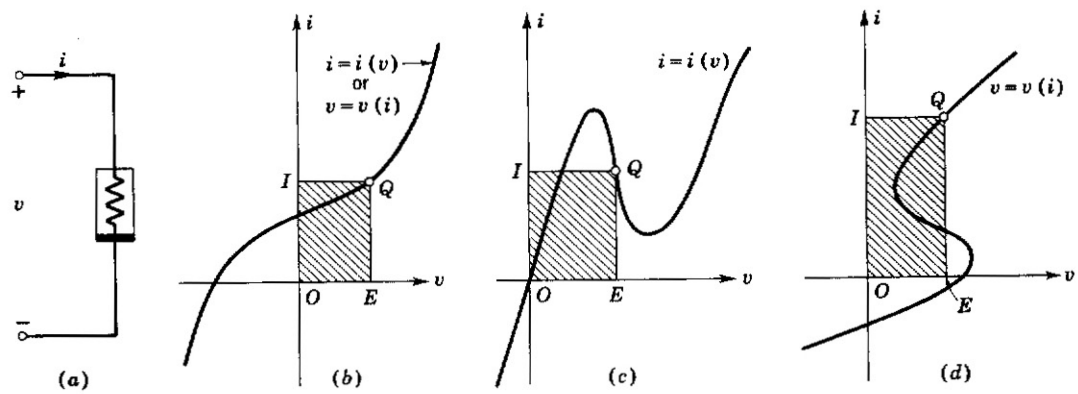


TABLE 1-2 *Instantaneous power, energy, and average power flow in a nonlinear resistor.*

	Strictly monotonically increasing v-i curve	Voltage-controlled v-i curve	Current-controlled v-i curve
$p_R(t)$	$= v(t)i(v(t))$ $= i(t)v(i(t))$	$= v(t)i(v(t))$	$= i(t)v(i(t))$
$w_R(t_0,t_1)$	$= \int_{t_0}^{t_1} v(t)i(v(t))\,dt$ $= \int_{t_0}^{t_1} i(t)v(i(t))\,dt$	$= \int_{t_0}^{t_1} v(t)i(v(t))\,dt$	$= \int_{t_0}^{t_1} i(t)v(i(t))\,dt$
$p_{R_{av}}$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} v(t)i(v(t))\,dt$ $= \lim_{t_1\to\infty} \frac{1}{t_1}\int_{t_0}^{t_1} i(t)v(i(t))\,dt$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} v(t)i(v(t))\,dt$	$= \lim_{t_1\to\infty} \frac{1}{t_1}\int_0^{t_1} i(t)v(i(t))\,dt$

example, suppose the v-i curve is represented by $i = v^3$, and the voltage is given by $v(t) = 2\sin\pi t$. The instantaneous power can then be calculated; thus

$$p_R(t) = (2\sin\pi t)(2\sin\pi t)^3 = 16(\sin\pi t)^4$$

Observe that $p_R(t)$ has a period $T = 1$. The energy flow during the time interval $(0,t_1)$ and the average power due to the periodic signal are given, respectively, by

$$w_R(0,t_1) = 6t_1 - \frac{3}{\pi}\sin 2\pi t_1 - \frac{4}{\pi}(\sin\pi t_1)^3(\cos\pi t_1)$$

and

$$p_{R_{av}} = \frac{w_R(0,T)}{T} = \frac{w_R(0,1)}{1} = 6 \tag{1-49}$$

Equation (1-49) shows that even though the voltage $v(t)$ changes from positive to negative values periodically, there is a net positive average power flow entering the resistor. Since this power is not returned to the external circuit whenever the voltage returns to its initial value during each period, it cannot be recovered and is therefore said to be "lost" or "dissipated" in the resistor. Since energy cannot be destroyed, this loss of electrical energy in the resistor is merely transformed into heat energy.

The average power for the above example is positive. Let us now consider another example where this is not true. Suppose the

v-i curve is represented by $i = v^3 - 2$ and suppose a constant voltage $v = 1$ volt is applied. The instantaneous and average power, respectively, are given by

$$p_R(t) = 1(1-2) = -1 \qquad \text{and} \qquad p_{R_{\text{av}}} = \frac{(-1)(T)}{T} = -1 \quad (1\text{-}50)$$

Since the average power is negative, energy is being supplied (instead of being absorbed) by the nonlinear resistor to the external circuit. Since energy cannot be created, this nonlinear resistor must have an external power source (e.g., a battery) associated with it, and is therefore called an *active resistor*. Without an external power source, a nonlinear resistor can only absorb power; namely, $p_R(t) \geq 0$. Such a resistor is said to be *passive*. It is easy to see that *a nonlinear resistor is passive if, and only if, its v-i curve lies entirely in the first and the third quadrants*. This follows from the fact that the instantaneous power is always nonnegative; namely,

$$p_R(t) = v(t)i(t) \geq 0 \quad (1\text{-}51)$$

Clearly, in its original form, a physical resistor must necessarily be passive. This is true, for example, with the commercial resistors listed in Table 1-1. Any of these resistors can, of course, be transformed into an active resistor by connecting a battery in series with it.

Exercise: The v-i curve of a certain nonlinear resistor is given by $i = 10(v^3 - 3v)$ma, and the voltage excitation is given by $v(t) = 10 \sin t$ volts. (*a*) Find the instantaneous power $p_R(t)$. (*b*) Find the energy flow $w_R(0,t_1)$ for all $t_1 > 0$. (*c*) Find the average power by using Eq. (1-48) and check by using Eq. (1-45). (*d*) Is this nonlinear resistor passive or active? Explain why.

Case 2: Two-terminal nonlinear capacitor Consider the nonlinear capacitor shown in Fig. 1-22*a* and the three typical types of v-q curves shown in Fig. 1-22*b* to *d*. The v-q curve can be described in the functional form by $v = v(q)$ if it is charge-controlled or by $q = q(v)$ if it is voltage-controlled. A strictly monotonically increasing v-q curve can obviously be described by either $v = v(q)$ or $q = q(v)$. The energy flow $w_C(t_0,t_1)$ into a capacitor during the time interval (t_0,t_1) is given by

$$w_C(t_0,t_1) = \int_{t_0}^{t_1} v(t)\,\frac{dq(t)}{dt}\,dt \quad (1\text{-}52)$$

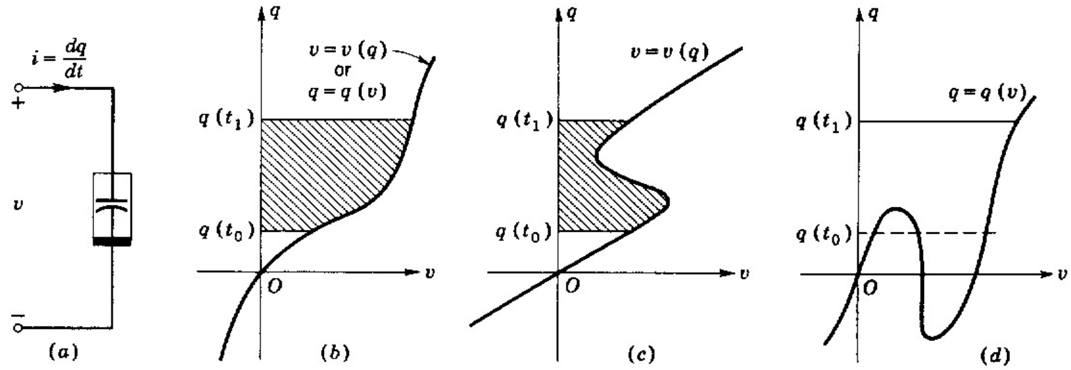


Fig. 1-22. The energy flow $w_C(t_0,t_1)$ from t_0 to t_1 into a nonlinear capacitor is equal numerically to the shaded area

In the case where the capacitor v-q curve is either strictly monotonically increasing or charge-controlled, Eq. (1-52) can be written as

$$w_C(t_0,t_1) = \int_{t_0}^{t_1} v(q(t))\,\frac{dq(t)}{dt}\,dt \tag{1-53}$$

By a standard change of variable, Eq. (1-53) becomes

$$w_C(t_0,t_1) = \int_{q(t_0)}^{q(t_1)} v(q)\,dq \tag{1-54}$$

In the special, but very important, case of a *linear* capacitor [$q = Cv$ or $v = (1/C)q$], Eq. (1-54) can be reduced to

$$w_C(t_0,t_1) = \int_{q(t_0)}^{q(t_1)} \frac{1}{C}q\,dq = \frac{1}{2C}\int_{q(t_0)}^{q(t_1)} d(q^2)$$

or

$$w_C(t_0,t_1) = \frac{1}{2C}[q^2(t_1) - q^2(t_0)] \tag{1-55}$$

Equation (1-55) can also be expressed in terms of v by substituting $q = Cv$ for q:

$$w_C(t_0,t_1) = \frac{C}{2}[v^2(t_1) - v^2(t_0)] \tag{1-56}$$

Referring to Fig. 1-22, Eq. (1-54) can be interpreted as follows: The energy flow $w_C(t_0,t_1)$ from t_0 to t_1 into a charge-controlled nonlinear capacitor is equal numerically to the area under the v-q curve (bounded by the q axis and the lines $q = q(t_0)$ and $q = q(t_1)$. This interpretation is significant because it shows that only three pieces of information are needed to determine $w_C(t_0,t_1)$, namely

1. The v-q curve
2. The initial value of the charge at $t = t_0$
3. The final value of the charge at $t = t_1$

Since no information is required of the waveforms of $q(t)$ and $v(t)$, the energy $w_C(t_0,t_1)$ is said to be independent of the excitation waveforms. This property is very different from the resistor case where the complete voltage and current waveforms are required to compute $w_R(t_0,t_1)$. Observe further from Fig. 1-22 that whenever the waveform $v(t)$ returns to the same initial point, i.e., when $q(t_1) = q(t_0)$, the energy $w_C(t_0,t_1) = 0$. For example, Eqs. (1-55) and (1-56) are both equal to zero under this condition. Hence, unlike the resistor case, there must be some form of "energy-swapping" mechanism between a capacitor and the external circuit connected across it. To investigate this mechanism, let us calculate the average power using Eq. (1-45); thus

$$P_{C_{av}} = \lim_{t_1 \to \infty} \frac{1}{t_1} \int_{q(0)}^{q(t_1)} v(q)\, dq \tag{1-57}$$

Now observe that except when $q(t)$ goes to infinity, a case that cannot occur in practice, the value of $q(t_1)$ will always be a finite number. This means that the area under the curve representing the integral in Eq. (1-57) will always be a finite number. But the value of t_1 in Eq. (1-57) must tend to infinity, therefore

$$P_{C_{av}} = 0 \tag{1-58}$$

Since this equation is derived only under the assumption that the v-q curve be charge-controlled (this includes clearly the special case of a monotonically increasing curve), it is a very general result. We can, therefore, conclude that *the average power entering a charge-controlled nonlinear capacitor is zero.* This condition is true for any capacitor current and voltage waveforms. In the special case where $q(t)$ and $v(t)$ are periodic, Eq. (1-57) can be simplified to

$$P_{C_{av}} = \frac{1}{T}\int_{q(0)}^{q(T)} v(q)\,dq \tag{1-59}$$

But $q(T) = q(0)$ for a periodic waveform of period T; therefore, Eq. (1-59) will integrate to zero, as it should.

From the preceding discussion, we can now conclude that a charge-controlled capacitor does not dissipate energy. Any energy entering it must be stored inside the capacitor and *may* eventually be returned. Because of this interpretation, a capacitor is often referred to as an *energy-storage element.* In the case of parallel-plate capacitors, it is possible to show, by electromagnetic field theory, that the energy is stored in the electric field between the plates. In view of this observation, the energy $w_C(t_0,t_1)$ in a capacitor is usually called the *electric stored energy.*

What happens if the v-q curve is neither monotonically increasing nor charge-controlled? In this case, it is no longer possible to describe the v-q curve by a function of q. It is not possible, therefore, to specify the area representing $v\,dq$ uniquely. To investigate this more general case, a new approach is required.[1]

Exercise 1: The v-q curve of a certain nonlinear capacitor is given by $q = \frac{1}{3} v^3$. Let the terminal voltage be given by $v(t) = e^{-t}$. (*a*) Find $w_C(0,t_1)$ for all $t_1 > 0$ by determining first $i(t) = (dq/dv)(dv/dt)$ and then using Eq. (1-43). (*b*) Repeat (*a*) by determining first $q(t)$ and then using Eq. (1-53). (*c*) Repeat (*a*) by using Eq. (1-54). (*d*) Let $v(t) = E \sin \omega t$ and verify that $P_{C_{av}} = 0$.

Exercise 2: Prove that the electric stored energy in a voltage-controlled capacitor is given by

$$w_C(t_0,t) = q(t_1)v(t_1) - q(t_0)v(t_0) - \int_{v(t_0)}^{v(t_1)} q(v)\,dv$$

HINT: Apply the integration-by-part theorem.

Case 3: Two-terminal nonlinear inductor Consider the nonlinear inductor shown in Fig. 1-23*a* and the three typical types of i-φ curves shown in Fig. 1-23*b* to *d*. The i-φ curve can be described in the functional form by $i = i(\varphi)$ if it is flux-controlled or by $\varphi = \varphi(i)$ if it is current-controlled. A strictly monotonically increasing i-φ curve can obviously be described by either $i = i(\varphi)$ or $\varphi = \varphi(i)$. The energy flow $w_L(t_0,t_1)$ into an inductor during the time interval (t_0,t_1) is given by

$$w_L(t_0,t_1) = \int_{t_0}^{t_1} i(t)\frac{d\varphi(t)}{dt}\,dt \tag{1-60}$$

[1] This approach is called the ***parametric approach*** and is discussed in Appendix A. See also L. O. Chua and R. A. Rohrer, On the Dynamic Equations of a Class of Nonlinear *RLC* Networks, *IEEE Trans. Circuit Theory,* vol. CT-12, no. 4, pp. 475–489, December, 1965.

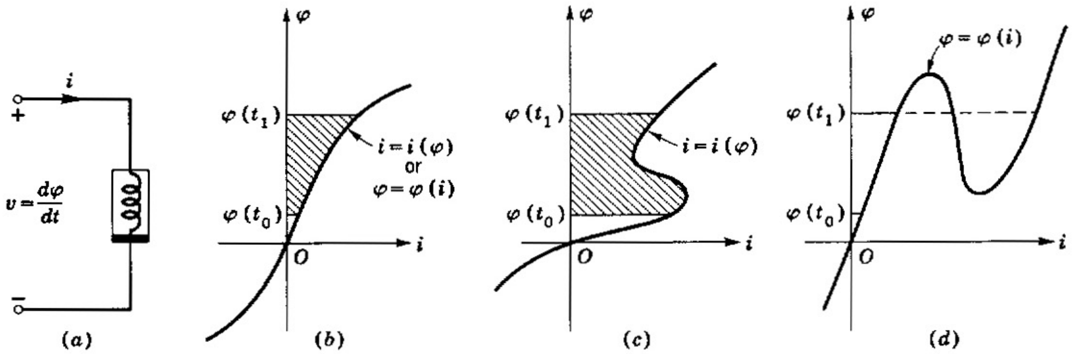


Fig. 1-23. The energy flow $w_L(t_0,t_1)$ from t_0 to t_1 into a nonlinear inductor is equal numerically to the shaded area.

Applying analogous procedure as in the capacitor case, we find that when the i-φ curve is either strictly monotonically increasing or flux-controlled, Eq. (1-60) can be written as

$$w_L(t_0,t_1) = \int_{\varphi(t_0)}^{\varphi(t_1)} i(\varphi)\, d\varphi \tag{1-61}$$

In the special case where the inductor is linear ($\varphi = Li$), Eq. (1-61) can be simplified further to

$$w_L(t_0,t_1) = \frac{1}{2L}[\varphi^2(t_1) - \varphi^2(t_0)] \tag{1-62}$$

or

$$w_L(t_0,t_1) = \frac{L}{2}[i^2(t_1) - i^2(t_0)] \tag{1-63}$$

Referring to Fig. 1-23, Eq. (1-61) can be interpreted as follows: The energy flow $w_L(t_0,t_1)$ from t_0 to t_1 into a flux-controlled nonlinear inductor is equal numerically to the area under the i-φ curve [bounded by the φ axis and the lines $\varphi = \varphi(t_0)$ and $\varphi = \varphi(t_1)$]. This interpretation has the same significance as for the capacitor; namely, only three pieces of information are needed to determine $w_L(t_0,t_1)$:

1. The i-φ curve
2. The initial value of the flux linkage at $t = t_0$
3. The final value of the flux linkage at $t = t_1$

By a similar procedure, we found *the average power in any flux-controlled inductor is zero;* thus

$$P_{L_{av}} = 0 \tag{1-64}$$

This means that a flux-controlled inductor cannot dissipate energy. In view of this observation, the inductor is also called an *energy-storage element.* In the case where the inductor is made of coils around an iron core, the energy can be shown, by electromagnetic principles, to be stored in the magnetic field around the coil. Hence, the energy stored in an inductor is usually called *magnetic stored energy.*

Exercise 1: Prove that Eq. (1-64) holds for a flux-controlled inductor. Verify this with $i(t) = I \cos \omega t$ and $\varphi = i^3$.

Exercise 2: Prove that the energy stored in a current-controlled inductor is given by

$$w_L(t_0,t_1) = \varphi(t_1)i(t_1) - \varphi(t_0)i(t_0) - \int_{i(t_0)}^{i(t_1)} \varphi(i)\, di$$

1-10 TIME-VARYING ELEMENTS

So far, the v-i, v-q, and i-φ curves characterizing a two-terminal resistor, capacitor, and inductor are assumed to remain unchanged for all times. These elements are said to be *time-invariant.* There exist some practical elements, however, whose v-i, v-q, or i-φ curves vary as functions of time. Such elements are said to be *time-varying resistors, capacitors,* or *inductors,* respectively.

Time-varying resistor The simplest example of a time-varying resistor is a potentiometer whose arm is being rotated by a motor as shown in Fig. 1-24*a*. At any time t, the potentiometer is simply a linear resistor with a straight-line v-i characteristic as shown in Fig. 1-24*b*. Hence, a time-varying linear resistor can be characterized by

$$v = R(t)i$$

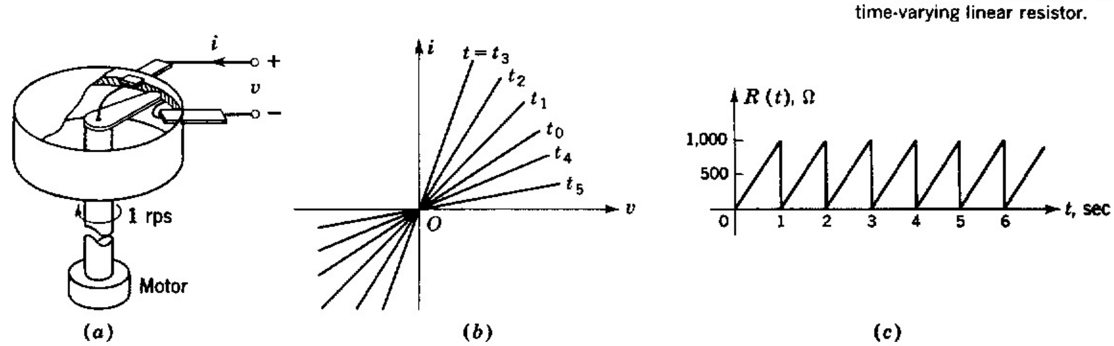


Fig. 1-24. An example of a time-varying linear resistor.

where $R(t)$ is the time-varying resistance representing the reciprocal of the slope of the straight line at any time t. For example, if the potentiometer has a resistance range of 0 to 1,000 Ω uniformly distributed around its rim and if the arm rotates at a speed of 1 rps, then the time-varying resistance is as shown in Fig. 1-24*c*.

A time-varying resistor need not be linear. For example, consider a resistor characterized by

$$i = v^3 + \sin t$$

The v-i curve of this time-varying nonlinear resistor is shown in Fig. 1-25*a* as a function of time. Observe that this resistor can be constructed in practice by connecting a sinusoidal current source in parallel with a time-invariant resistor (Fig. 1-25*b*) with the v'-i' curve shown in Fig. 1-25*c*. In general, a time-varying nonlinear resistor can be characterized by a relationship $i = i(v,t)$ if it is voltage-controlled, or $v = v(i,t)$ if it is current-controlled. A review of the power and energy expressions derived in the preceding section would show that these expressions remain valid for the time-varying case.

What are time-varying resistors good for? To give one simple application, let us consider the current waveform

$$i(t) = [1 + f(t)] \sin \omega t \qquad (1\text{-}65)$$

Equation (1-65) is called an *amplitude-modulated waveform* because the amplitude of the sine wave varies with time. This is the type

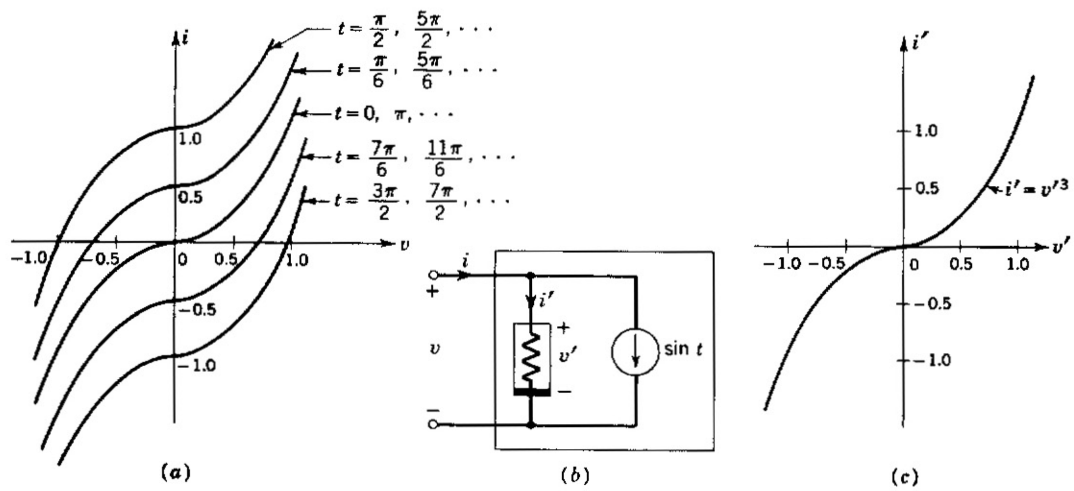


Fig. 1-25. An example of a time-varying nonlinear resistor.

of signal that an AM radio transmitter sends out. In practice, $f(t)$ represents a slowly changing signal and sin ωt represents a relatively high-frequency sine wave known as the "carrier." We are not equipped to explain why $f(t)$ cannot be transmitted directly, and why it must be "carried" by the sine wave. Suffice it to say that it takes a high-frequency waveform to traverse a long distance in space. Our objective here is to show how we may recover the signal $f(t)$ from Eq. (1-65). One possible method consists of applying this current to a time-varying linear resistor whose resistance changes at the *same* frequency as the carrier, namely,

$$R(t) = 1 + \sin \omega t$$

The voltage drop across this resistor is given by

$$\begin{aligned} v(t) &= R(t)i(t) \\ &= (1 + \sin \omega t)[1 + f(t)] \sin \omega t \\ &= \tfrac{1}{2} f(t) + \tfrac{1}{2} + [1 + f(t)] \sin \omega t - \tfrac{1}{2}[1 + f(t)] \cos 2\omega t \end{aligned} \tag{1-66}$$

Observe that Eq. (1-66) contains four terms; the first term is the signal that we would like to recover, the second term is a dc voltage, the third term is the carrier-frequency term, and the last term is at twice the carrier frequency. Through the use of a "filter," the last three components can be easily suppressed, thus leaving the desired signal $f(t)$. This recovering process is known as *synchronous detection* because the frequency of the time-varying resistance is synchronized at the same frequency as the carrier.

Exercise 1: Sketch the amplitude-modulated waveform given by Eq. (1-65) with $f(t) = \sin t$ and $\omega = 100$. What can you say about the "envelope" of this waveform?

Exercise 2: It is possible to rectify a sinusoidal current waveform $i(t) = I \sin t$ by applying this current to an appropriate time-varying linear resistance $R(t)$. Find $R(t)$ so that the resistor voltage is a rectified version of the current waveform; that is, $v(t) = i(t)$ whenever $i(t) \geq 0$, and $v(t) = 0$ whenever $i(t) \leq 0$.

Time-varying capacitor The simplest example of a time-varying linear capacitor is the air capacitor consisting of a fixed set of plates in mesh with a movable set of plates which is being rotated by a motor. A time-varying linear capacitor is therefore characterized by

$$q(t) = C(t)v(t) \tag{1-67}$$

where $C(t)$ is the time-varying capacitance. Unlike the resistor case, the expressions previously derived for the nonlinear capacitors do not apply in the time-varying case because when we differentiate $q(t)$ with respect to time, we obtain an additional term, namely,

$$i(t) = \frac{dq}{dt} = C(t)\frac{dv(t)}{dt} + v(t)\frac{dC(t)}{dt} \tag{1-68}$$

Since $C(t)$ is not a constant, the expressions given by Eqs. (1-55) and (1-56) are no longer applicable. Hence, to calculate the power or energy flow, we must resort to the original definitions.

Just as for the resistor, a time-varying capacitor may be nonlinear; in this case it is characterized by $q = q(v,t)$ if it is voltage-controlled or $v = v(q,t)$ if it is charge-controlled. Time-varying capacitors are useful in the study of parametric amplifiers. They are also useful in the modeling of many time-varying physical and biological systems. For example, the mass of a rocket during lift-off decreases rapidly with time as the rocket fuel is burned. This time-varying mass can be modeled by a time-varying capacitor.

Exercise 1: Find the average power $P_{C_{av}}$ entering a time-varying capacitor $C(t) = 2 - \cos \omega t$ and a terminal voltage $v(t) = E \sin \omega t$. Interpret whether this energy is being absorbed, delivered, or stored.

Exercise 2: Give an example for each of the following: (*a*) A time-varying linear capacitor, (*b*) a time-varying charge-controlled capacitor, and (*c*) a time-varying voltage-controlled capacitor.

Time-varying inductor By exact analogy to the capacitor, a time-varying linear inductor is characterized by

$$\varphi(t) = L(t)i(t) \tag{1-69}$$

where $L(t)$ is the time-varying inductance. Since $L(t)$ is no longer a constant, the expressions derived previously in the preceding sections are no longer valid. In particular, the inductor voltage is now given by

$$v(t) = \frac{d\varphi(t)}{dt} = L(t)\frac{di(t)}{dt} + i(t)\frac{dL(t)}{dt} \tag{1-70}$$

A time-varying inductor may be nonlinear; in this case it is characterized by $\varphi = \varphi(i,t)$ if it is current-controlled and $i = i(\varphi,t)$ if it is flux-controlled.

The analysis of a nonlinear network containing time-varying elements is a very difficult problem requiring advanced mathematics. Hence, we shall not consider any time-varying elements in the rest of this book. The above discussion is included mainly to emphasize the fact that most of the equations we derived in the previous sections are not valid for time-varying elements.

Exercise 1: Give an example for each of the following: (*a*) A time-varying linear inductor, (*b*) a time-varying flux-controlled inductor, and (*c*) a time-varying current-controlled inductor.

Exercise 2: Prove or disprove the assertion that if the current into a time-varying current-controlled inductor is periodic, then the instantaneous power $P_L(t)$ is also periodic.

1-11 CONCEPTS OF MODELING

One of the most basic principles in scientific analysis is that of modeling. Engineers and scientists seldom analyze a physical system in its original form. Instead, they construct a model which approximates the behavior of the system. By analyzing the behavior of the model, they hope to predict the behavior of the actual system. The primary reason for constructing models is that physical systems are usually too complex to be amenable to a practical analysis. In most cases, the complexity of a system is due in part to the presence of many nonessential factors. One basic principle of modeling consists, therefore, of extracting only the essential factors.

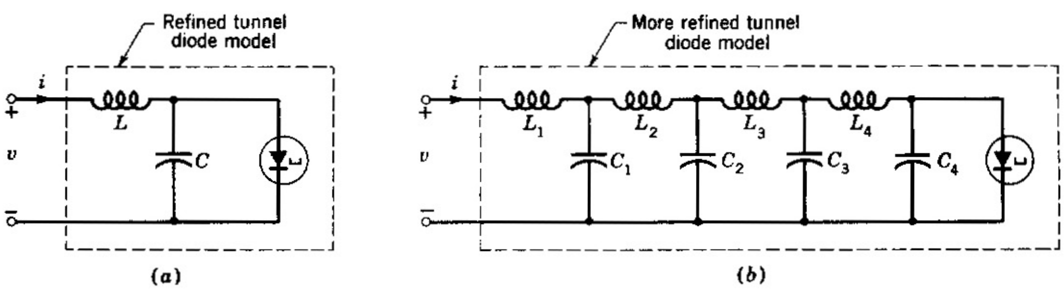
To illustrate the process of modeling, let us consider the problem of predicting the trajectory of a ballistic missile. This problem cannot be analyzed exactly because an exact analysis would require inclusion of all possible factors that may affect the trajectory. Some of these factors may be the weight and shape of the missile, the amount of thrust, the atmospheric drag, the deformation of the missile during flight, the distribution of weights of the internal components, the wind velocity, the impurity of the fuel, and the color of the missile. From experience, we know that the first three factors have a more significant influence on the trajectory than the remaining factors. This leads us to replace the missile by a model which includes only the first three factors. Obviously, the predicted trajectory based on this model is not going to be identical with that of the actual system. But as engineers, we are interested only in an "accurate" solution, not the exact solution. Hence, as long as the discrepancy between the predicted and the actual trajectories is tolerable, the model serves the purpose. Of course, in their desire to simplify analysis, engineers are often tempted to overidealize the model by stripping away some essential factors. In this case, the predicted solution may not be satisfactory.

The point we are driving at is that as engineers, we analyze the model which approximates an actual system. A model is always an idealization of a physical system. The more complex the model, the more accurate will the predicted solution be. Unfortunately, the analysis will also become more complicated. Hence, *a model is always a compromise between reality and simplicity.*

In the light of the above discussion, our definitions of a resistor, capacitor, and inductor must also be interpreted as models representing a physical device. For example, the v-i curve of the tunnel diode shown in Table 1-1 is a good model of a physical tunnel diode so long as the frequency at which we are operating is not very high. However, as the frequency increases, the static characteristic becomes less accurate, and a more realistic model must be found. For example, at very high frequencies, the connecting wires begin to behave like an inductance, and the capacitance between the wires gradually becomes significant. These elements are called "parasitic" or "stray" elements because they are invariably present, even though they are undesirable. A more realistic tunnel diode model must therefore include the parasitic elements such as the refined model shown in Fig. 1-26*a*. As the frequency gets higher, a still more complicated model such as shown in Fig. 1-26*b* may be chosen.

In this book, we shall be primarily interested in low-frequency models. In Chap. 11 we shall learn some basic techniques for constructing models of three-terminal devices in terms of two-terminal models. These low-frequency models can usually be refined for high-frequency analysis upon inclusion of appropriate parasitic elements.

Fig. 1-26. The static model of a tunnel diode must be refined for high-frequency analysis by the inclusion of appropriate parasitic inductances and capacitances at appropriate locations.



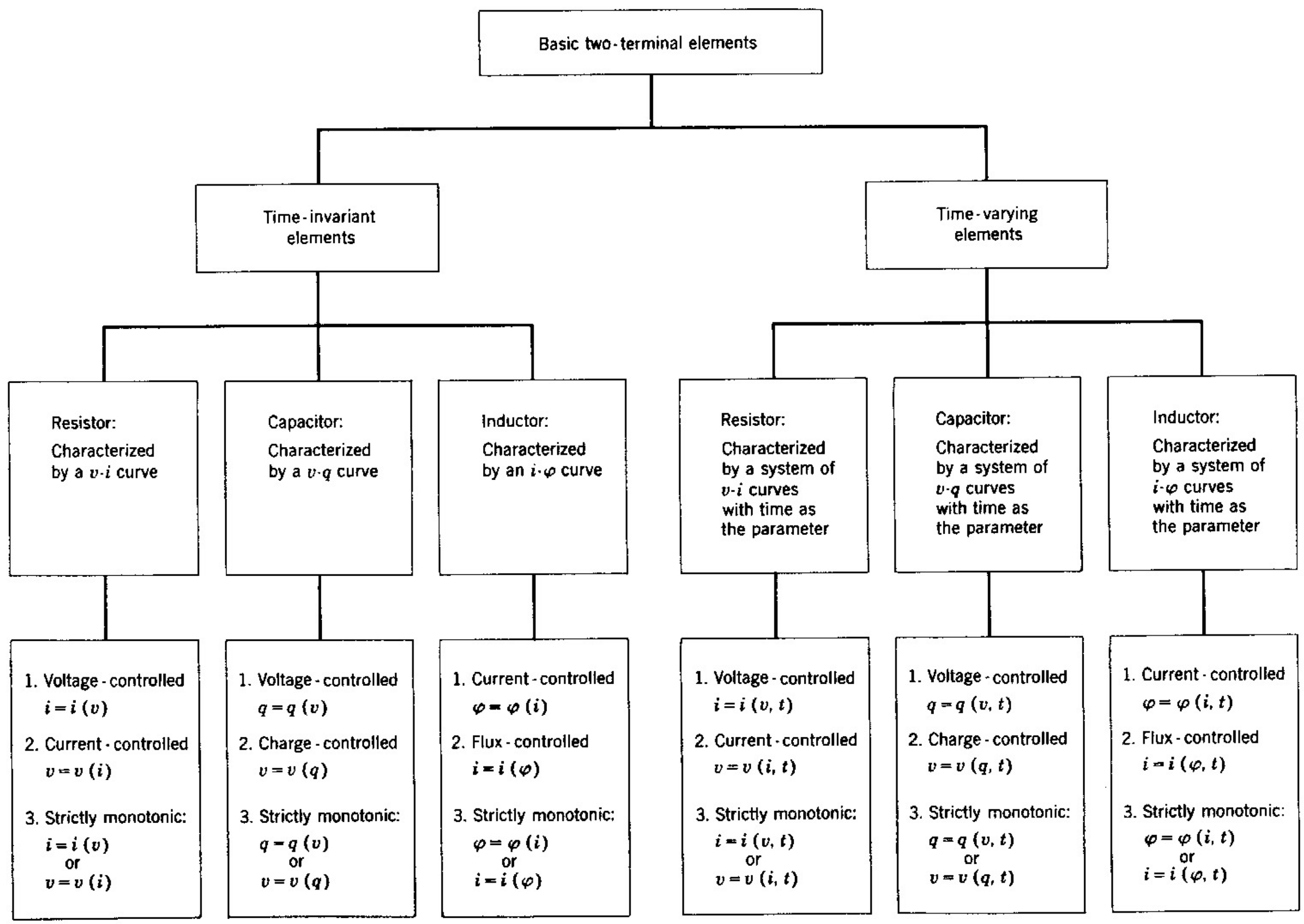

Basic two-terminal elements
Time-invariant elements
Time-varying elements
Resistor: Characterized by a v-i curve
Capacitor: Characterized by a v-q curve
Inductor: Characterized by an i-φ curve
Resistor: Characterized by a system of v-i curves with time as the parameter
Capacitor: Characterized by a system of v-q curves with time as the parameter
Inductor: Characterized by a system of i-φ curves with time as the parameter
1. Voltage-controlled $i = i(v)$ 2. Current-controlled $v = v(i)$ 3. Strictly monotonic: $i = i(v)$ or $v = v(i)$
1. Voltage-controlled $q = q(v)$ 2. Charge-controlled $v = v(q)$ 3. Strictly monotonic: $q = q(v)$ or $v = v(q)$
1. Current-controlled $\varphi = \varphi(i)$ 2. Flux-controlled $i = i(\varphi)$ 3. Strictly monotonic: $\varphi = \varphi(i)$ or $i = i(\varphi)$
1. Voltage-controlled $i = i(v, t)$ 2. Current-controlled $v = v(i, t)$ 3. Strictly monotonic: $i = i(v, t)$ or $v = v(i, t)$
1. Voltage-controlled $q = q(v, t)$ 2. Charge-controlled $v = v(q, t)$ 3. Strictly monotonic: $q = q(v, t)$ or $v = v(q, t)$
1. Current-controlled $\varphi = \varphi(i, t)$ 2. Flux-controlled $i = i(\varphi, t)$ 3. Strictly monotonic: $\varphi = \varphi(i, t)$ or $i = i(\varphi, t)$

Fig. 1-27.

It stores *electric* energy and may be returned to its external circuit ($P_{C_{av}} = 0$). A flux-controlled inductor does not dissipate energy. It stores *magnetic* energy and may be returned to its external circuit ($P_{L_{av}} = 0$).

Basic two-terminal elements (See Fig. 1-27.)

PROBLEMS

1-1 In order to demonstrate the importance of reference direction and polarity, consider the following:

(a) Sketch the v-i curve of each of the two-terminal black boxes shown in Fig. P1-1*a* to *e*. (The element inside the black box is a tunnel diode whose *V-I* curve is given in Table 1-1.)

(b) Give a simple rule for sketching the v-i curve of a resistor whose terminals have been reversed as in Fig. P1-1*e*.

(c) There exists a certain class of nonlinear elements in which it is unnecessary to distinguish between the two terminals. Such elements are called *bilateral elements*. All other elements are said to be *nonbilateral*. Give an example of a bilateral and a nonbilateral resistor, inductor, and capacitor.

(d) Find the necessary and sufficient condition for a nonlinear resistor, capacitor, or inductor to be bilateral.

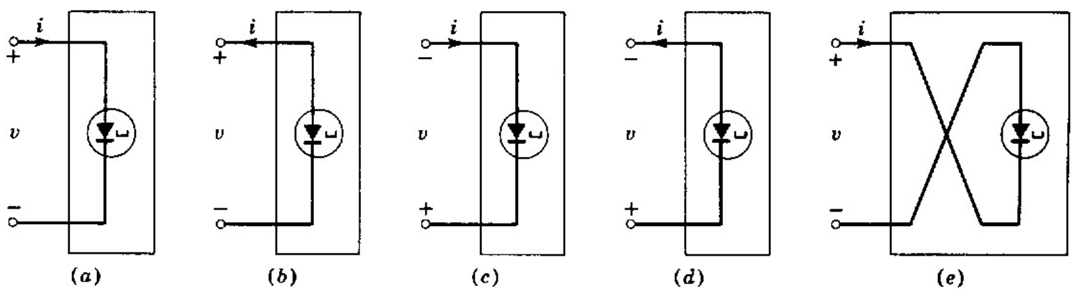
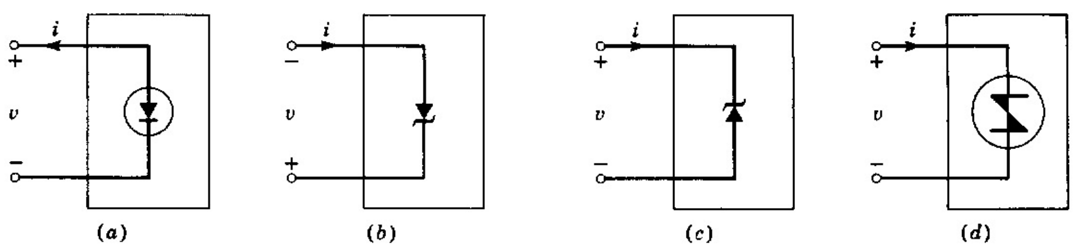


Fig. P1-1.

1-2 Sketch the v-i curve of each of the two-terminal black boxes shown in Fig. P1-2*a* to *h*. Refer to Table 1-1 for the *V-I* curves of the corresponding elements.

Fig. P1-2.



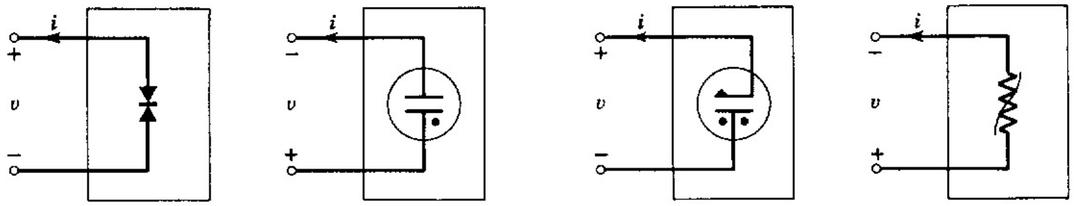


Fig. P1-2 (*Continued*).

1-3 The most accurate method for determining the v-i curve of a nonlinear resistor in the laboratory is the point-by-point method. Each point (E,I) on the curve is obtained by applying a voltage $v = E$ across the resistor and measuring the resulting current $i = I$. However, this method is rather tedious, and for most practical purposes it is desirable to design a v-i "curve tracer" that can display the v-i curve directly on an oscilloscope.

(a) Devise a simple circuit to carry out the above task using the ordinary 60-Hz ac line voltage and a Variac (variable voltage transformer) to provide the desired range of input voltage required by the given resistor. You may use a linear resistor whose voltage drop can be used to sense the magnitude of the current in the nonlinear resistor. To avoid grounding problems, you may use a 1:1 isolation transformer.

(b) Suppose that instead of using the line voltage as the energy source, we use the output voltage from a certain signal generator whose frequency can be changed from 10 Hz to 100 MHz. Do you expect the same v-i curve to be traced on the scope at all frequencies? If not, what frequency range must be chosen so that the v-i curve will agree approximately with the static curves supplied by the manufacturer?

1-4 It is sometimes convenient to define the dc resistance R_{dc} and the ac resistance R_{ac} at each point P of a v-i curve by

$$R_{dc} = \frac{v}{i}\bigg|_P$$

$$R_{ac} = \frac{dv}{di}\bigg|_P$$

(a) Show that $R_{dc} = \cot \alpha$, where α is the angle between the v axis and the straight line from the origin to point P.

(b) Show that $R_{ac} = \cot \beta$, where β is the angle between the v axis and the straight line tangent at point P.

(c) Verify that for a passive resistor, the value of R_{dc} is always positive.

(d) Verify that the value of R_{ac} may be either negative, zero, positive, or even infinite for a passive resistor.

(e) Sketch the relationships R_{dc} versus v and R_{ac} versus v for the varistor type 1NXX5, the zener diode type 1NXX3, and the tunnel diode type 1NXX6. See Appendix D for the v-i curve of these elements.

(f) Repeat (*e*) for the relationships R_{dc} versus i and R_{ac} versus i.

1-5 Consider the definitions of the dc resistance R_{dc} and ac resistance R_{ac} given in Problem 1-4.

(a) If the resistor v-i curve is strictly monotonically increasing, what can you say about the curves R_{dc} versus v, R_{dc} versus i, R_{ac} versus v, and R_{ac} versus i? Are they monotonic, single-valued, or multivalued?

(b) Repeat (*a*) for a voltage-controlled resistor.

(c) Repeat (*a*) for a current-controlled resistor.

1-6 It is sometimes convenient to define the dc conductance G_{dc} and the ac conductance G_{ac} at each point P of a v-i curve by

$$G_{dc} = \frac{i}{v}\bigg|_P$$

$$G_{ac} = \frac{di}{dv}\bigg|_P$$

(a) Show that $G_{dc} = \tan \alpha$, where α is the angle between the v axis and the straight line from the origin to point P.

(b) Show that $G_{ac} = \tan \beta$, where β is the angle between the v axis and the straight line tangent at point P.

(c) Verify that for a passive resistor, the value of G_{dc} is always a finite, positive number.

(d) Verify that the value of G_{ac} may be either negative, zero, positive, or even infinite for a passive resistor.

(e) Sketch the relationships G_{dc} versus i and G_{ac} versus i for the varistor type 1NXX5, the zener diode type 1NXX3, and the tunnel diode type 1NXX6. See Appendix D for the v-i curve of these elements.

(f) Repeat (*e*) for the relationships G_{dc} versus v and G_{ac} versus v.

1-7 Consider the definitions of the dc conductance G_{dc} and ac conductance G_{ac} given in Prob. 1-6.

(a) If the resistor v-i curve is strictly monotonically increasing, what can you say about the curves G_{dc} versus i, G_{dc} versus v, G_{ac} versus i, and G_{ac} versus v? Are they monotonic, single-valued, or multivalued?

(b) Repeat (*a*) for a voltage-controlled resistor.

(c) Repeat (*a*) for a current-controlled resistor.

1-8 If a sinusoidal voltage waveform $v = A \sin \omega t$ is applied across a nonlinear resistor characterized by a polynomial

$$i = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n$$

the resulting current will contain, in addition to the fundamental frequency term $i = B \sin \omega t$, other higher harmonic terms. In many practical applications, these harmonic terms are usually filtered out, in which case it becomes meaningful to define the ratio between the amplitudes of the fundamental voltage and current components to be the average resistance R_{av}.

(a) Find the average resistance R_{av} with $n = 3$.

(b) Assuming $a_0 = a_1 = a_2 = a_3 = 1$, plot R_{av} versus the amplitude A.

(c) What is the significance of the R_{av}-vs.-A curve obtained in (*b*)?

1-9 If a sinusoidal voltage waveform $v = A \cos \omega t$ is applied across a voltage-controlled capacitor characterized by

$$q = a_0 + a_1 v + a_2 v^2 + \cdots + a_n v^n$$

the resulting current will contain, in addition to the fundamental frequency term $i = B \sin \omega t$, other higher harmonic terms. In many practical applications these harmonic terms are filtered out, in which case the ratio between the amplitudes of the fundamental voltage and current components is usually called the *describing function* Z_C.

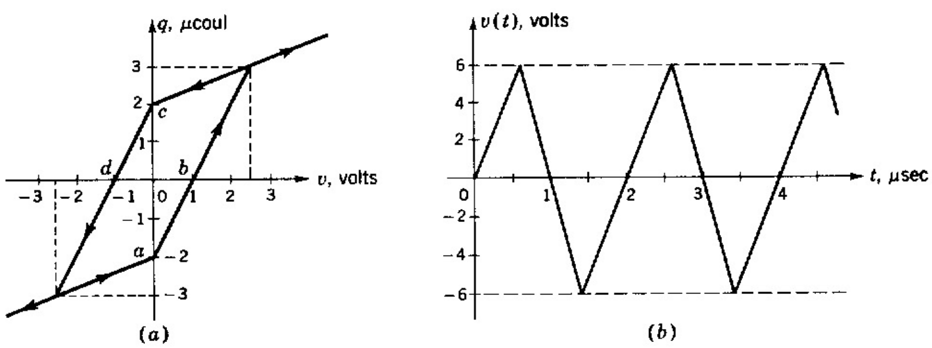
(a) Find the describing function Z_C with $n = 3$.

(b) Observe that unlike the average resistance in Prob. 1-8, the describing function Z_C of a capacitor is a function of the frequency ω. Assuming $a_0 = a_1 = a_2 = a_3 = A = 1$, plot the curve Z_C versus ω.

(c) What is the significance of the Z_C-vs.-ω curve?

1-10 The v-q curve of a practical, nonlinear capacitor with a barium titanate dielectric is shown in Fig. P1-10*a*. If the triangular voltage signal shown in Fig. P1-10*b* is applied across this capacitor, find the current waveform $i(t)$. Assume that the capacitor is operating initially at point *a* of the hysteresis curve. Assume also that the locus must follow the arrow directions.

Fig. P1-10.



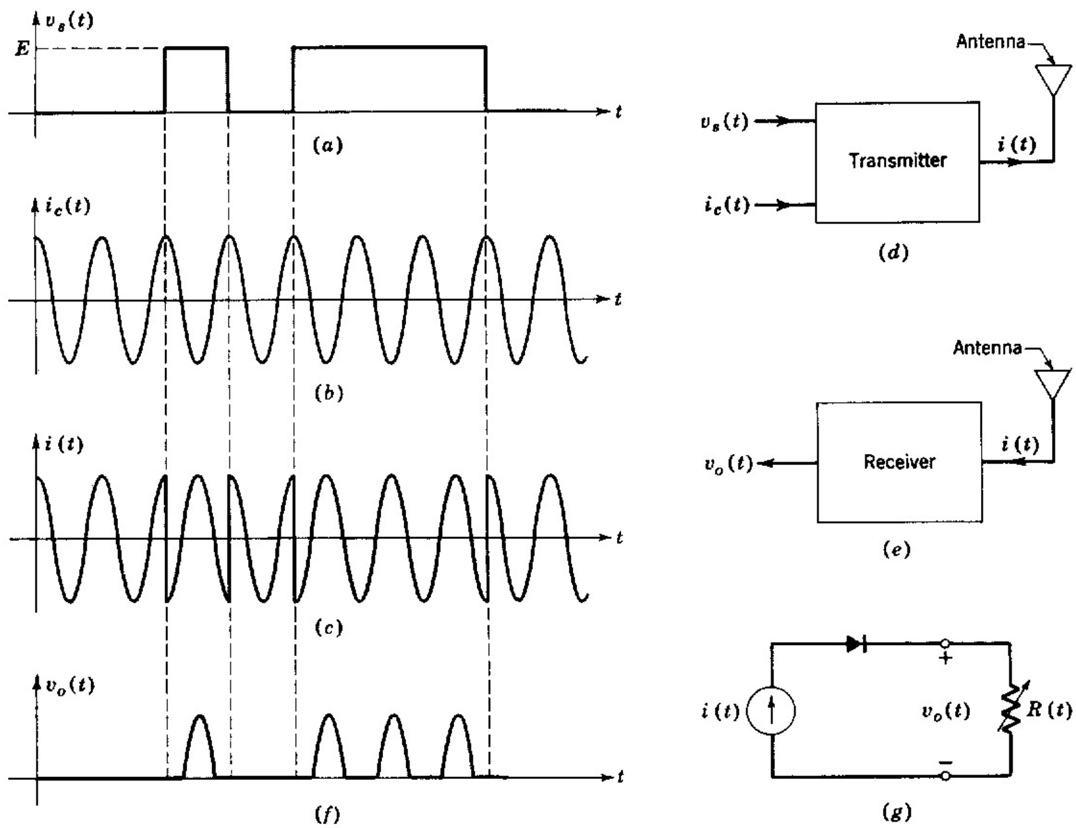
where sin ωt and cos ωt are very high-frequency (say 10 MHz) sine waves. This is the signal that will be received in Paris. Our problem is to recover $f_1(t)$ and $f_2(t)$ at the receiving end.

(a) Show that $f_1(t)$ can be recovered by applying $i(t)$ to a time-varying linear resistor with $R(t) = 1 + \sin \omega t$ and then suppressing the components with a frequency higher than ω by means of a filter.

(b) By a similar procedure, $f_2(t)$ can be recovered. Find the appropriate time-varying resistance $R(t)$ for accomplishing this.

1-14 One method for transmitting a telegraph signal over long distances is to modulate the "phase" of a high-frequency sinusoidal signal called the *carrier*. To be specific, suppose we wish to transmit the letter A in morse code by closing and opening the telegraph key at appropriate intervals. Corresponding to this code, the output voltage $v_s(t)$ shown in Fig. P1-14*a* will be generated. In order to transmit this waveform over long distances, an apparatus can be

Fig. P1-14.



designed to change the phase of the carrier signal $i_c(t)$ shown in Fig. P1-14*b* abruptly by 180° each time $v_s(t)$ changes its amplitude. For example, the resulting current waveform $i(t)$ is shown in Fig. P1-14*c*. This is the signal being transmitted and received, as shown in Fig. P1-14*d* and *e*. Our problem is to decode the received current waveform $i(t)$ so as to recover the message A. This can be accomplished by applying $i(t)$ (as a current source) to the time-varying circuit shown in Fig. P1-14*g* so as to produce the output voltage $v_o(t)$ shown in Fig. P1-14*f*. Observe that even though $v_o(t)$ is not identical with $v_s(t)$, the nature of the waveform is unmistakably similar to that of Fig. P1-14*g*. Hence the above decoding scheme would accomplish our objective.

(a) Specify the time-varying resistance $R(t)$ for accomplishing this task.

(b) What can you say about the frequency of $R(t)$ in comparison with that of the carrier signal $i_c(t)$?

(c) The above scheme for decoding the signal is known as the *synchronous phase detection*. Explain why.