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. 1-1a, 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.

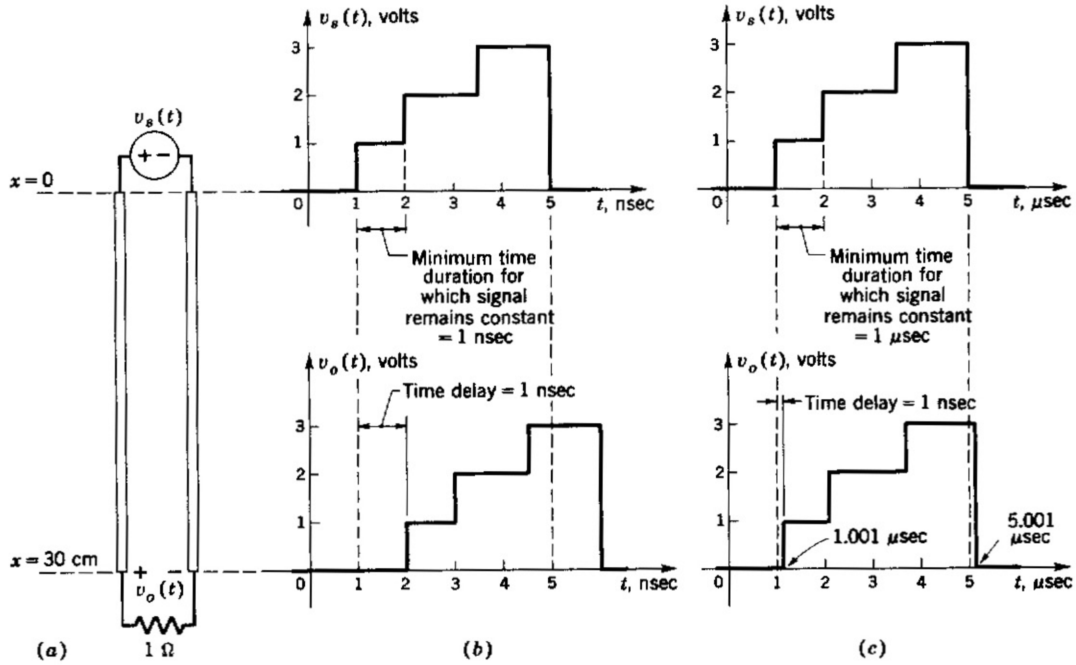

Fig. 1-1. The length of the transmission line introduces a time delay which is significant in (*b*) but may be neglected in (*c*).

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.

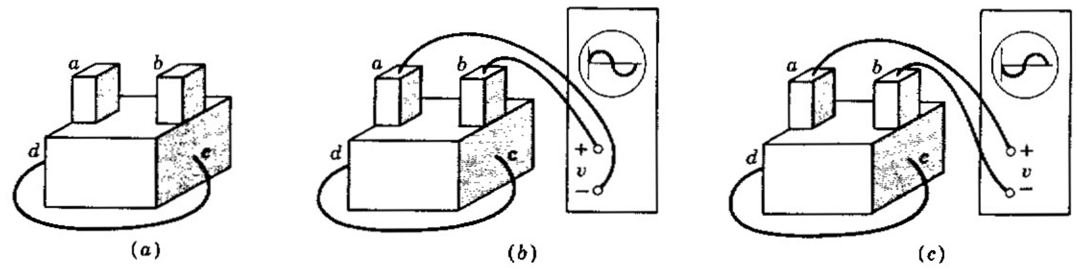

Fig. 1-2. An experiment demonstrating that regardless of which terminal of the black box is chosen to be positive, the actual voltage across terminals *a-b* can be unambiguously specified for all time.

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*.

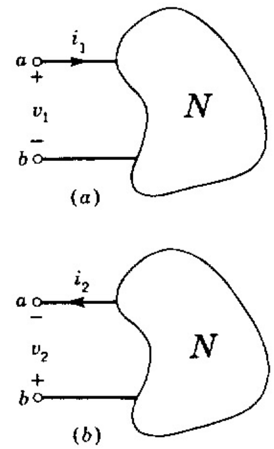

Fig. 1-3. Two possible sets of assumed reference direction and polarity for ensuring that positive power means power entering the black box.

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

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.

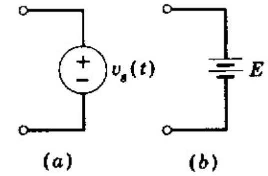

Fig. 1-4. Symbols for an independent voltage source.

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.

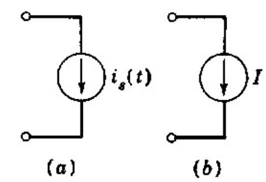

Fig. 1-5. Symbols for an independent current source.

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*.

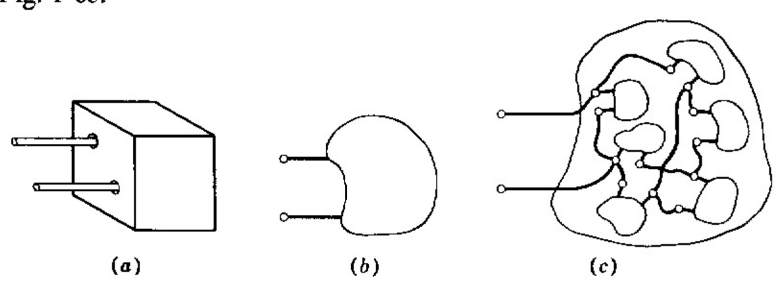

Fig. 1-6. Symbolic representation of a two-terminal black box.