harmonic-balance analysis and *Volterra-series* or *nonlinear transfer function* analysis. Harmonic-balance analysis is applicable primarily to strongly nonlinear circuits that are excited by a single large-signal excitation source; it can be applied to such circuits as transistor power amplifiers, mixer local oscillator circuits, and frequency multipliers using either diodes or transistors. Volterra-series analysis is applicable to the opposite problem: weakly driven, weakly nonlinear circuits having multiple small-signal excitations at noncommensurate frequencies. As such, it is most useful for evaluating intermodulation characteristics and other nonlinear phenomena in small-signal receiver circuits, such as amplifiers. With some modifications, the Volterra series can also be used to determine the IM properties of time-varying circuits such as mixers; similarly, harmonic-balance can be extended to certain situations involving noncommensurate signals. Demystifying the theory and practical use of these two techniques is the primary subject of this book.

All three methods—time-domain analysis, harmonic-balance analysis, and the Volterra series—require a circuit model consisting of lumped components and, for the latter two, impedance elements or multiports. Solid-state device models must consist of linear or nonlinear capacitors, inductors, resistors, and voltage or current sources (nonlinear inductors can be accommodated, although they are rarely encountered in solid-state devices or circuits). Underlying all the nonlinear models described in Chapter 2 is the *quasistatic assumption,* whereby all nonlinear elements are assumed to change instantaneously with changes in their control voltages. This assumption is also implicit in linear circuit theory; it requires, for example, that the charge on a capacitor is a function solely of the voltage at its terminals. If the capacitor is nonlinear, its incremental capacitance, as well as its charge, must change instantaneously with control voltage. A quasistatic circuit is not necessarily *memoryless;* a memoryless circuit is one in which no charge or flux storage elements (no capacitors or inductors) exist, so voltages and currents at any instant do not depend upon previous values of voltage or current. In a quasistatic circuit, the network voltages and currents may depend upon previous values of other voltages or currents, but the capacitances, inductances, resistances, and controlled sources do not depend directly upon their own histories.

The quasistatic assumption is critical to the entire business of both linear and nonlinear circuit analysis. It allows us, for example, to devise equivalent circuits for solid-state devices using only lumped linear and nonlinear elements, and makes many of the techniques of linear circuit theory applicable to at least the linear parts of nonlinear circuits. One of the nicest things about the quasistatic assumption is its range of validity.

Theoretical and experimental studies of silicon and gallium arsenide semiconductors and devices show that time-delay phenomena are usually on the order of picoseconds or are short compared to the inverse of the highest frequency at which we would attempt to use the device. Furthermore, the prohibition of time delays is not absolute; in some cases they can still be managed. For example, it is often possible to include in a quasistatic model the well known time delay between the gate voltage and drain current in a GaAs MESFET.

1.5 POWER AND GAIN DEFINITIONS

Although it is customary to speak loosely of gain and power in microwave circuits, these quantities can be defined in several different ways. The different definitions of gain are related to the concepts of available and dissipated power. These concepts are important in both linear and nonlinear circuits, although they are particularly important in nonlinear circuits where a waveform may have components at many frequencies that may or may not be harmonically related.

Available or *transferable power* is the maximum power that can be obtained from a source. The concept of available power is illustrated in Figure 1.6, in which a sinusoidal voltage source having a peak value V_s has an internal impedance of $R_1 + jX_1$ (unless we state otherwise, all frequency-domain voltages and currents in this book are phasor quantities; thus their magnitudes are equal to peak sinusoidal quantities, not rms).

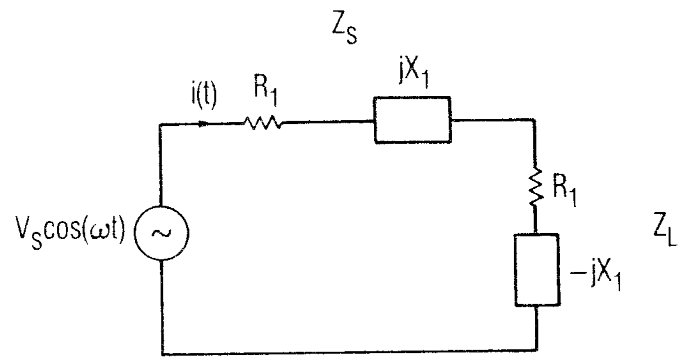

Figure 1.6 Circuit having a matched source and load, illustrating the concept of available power.

The maximum power is obtained from this source if the load impedance equals the conjugate of the source impedance, $Z_L = Z_s^* = R_1 - jX_1$. Under these conditions:

$$I = \frac{V_s}{2R_1} \tag{1.5.1}$$

where I is the peak value of the current, $i(t)$. The power dissipated in the load is

$$P_d = P_{av} = \frac{1}{2} I^2 R_1 = \frac{1}{2} I^2 \,\mathrm{Re}\{Z_s\} = \frac{V_s^2}{8R_1} \tag{1.5.2}$$

which is the maximum available from the source. *Dissipated*, or *transferred* power is the power dissipated in a load that may or may not be matched to the source. In Figure 1.7, the load is not conjugate-matched to the source, so the dissipated power is less than that given in (1.5.2). In this case,

$$I = \frac{V_s}{[(R_1 + R_2)^2 + (X_1 + X_2)^2]^{1/2}} \tag{1.5.3}$$

and the power dissipated in the load is

$$P_d = \frac{1}{2} I^2 R_2 = \frac{V_s^2 R_2}{2[(R_1 + R_2)^2 + (X_1 + X_2)^2]} \tag{1.5.4}$$

In a nonlinear circuit, the voltage source may contain many frequency components, and the source or load impedance may not be the same at each frequency. An example of this situation is the output circuit of a diode frequency multiplier. The multiplier generates many harmonics, only one of which is desired, so it has an output filter that allows only the desired harmonic to reach the output port. Thus, the impedance presented to the diode at the desired output frequency is the load impedance, but it is the out-of-band impedance of the filter at all other harmonics. The current in the loop is a function of frequency, as shown in Figure 1.8. Because the load and source are linear, each frequency component can be treated separately without concern for the others. Then, the available and transferred power are

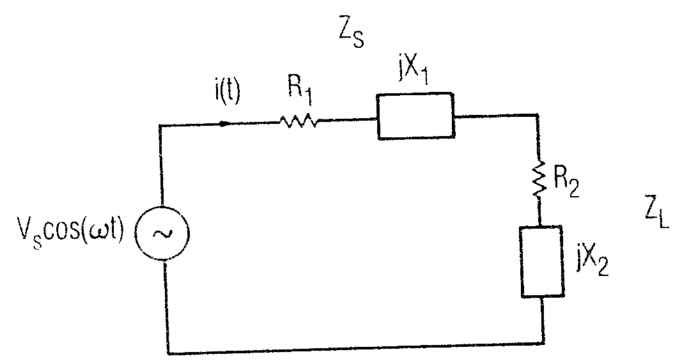

Figure 1.7 Circuit having an unmatched source and load.

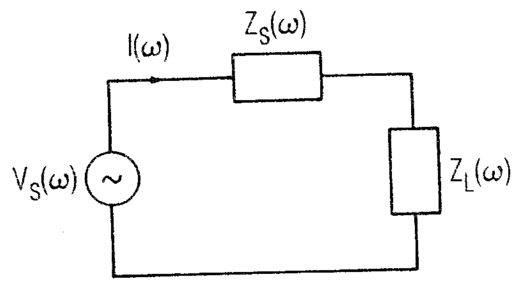

Figure 1.8 Unmatched circuit having a nonsinusoidal voltage-source excitation.

$$P_{av}(\omega) = \frac{|V_s(\omega)|^2}{8\,\mathrm{Re}\{Z_s(\omega)\}} \tag{1.5.5}$$

$$P_d(\omega) = \frac{1}{2} |I(\omega)|^2 \,\mathrm{Re}\{Z_L(\omega)\} \tag{1.5.6}$$

An equivalent representation uses a current source and admittances as shown in Figure 1.9. Similarly, the available and dissipated powers are found to be

$$P_{av}(\omega) = \frac{|I_s(\omega)|^2}{8\,\mathrm{Re}\{Y_s(\omega)\}} \tag{1.5.7}$$

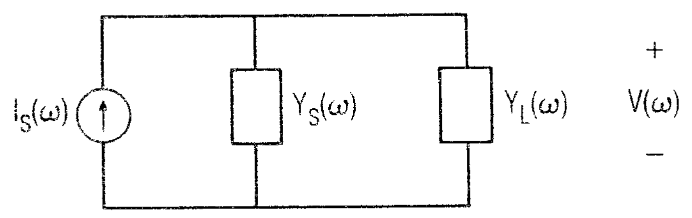

Figure 1.9 Unmatched circuit having a nonsinusoidal current-source excitation.

$$P_d(\omega) = \frac{1}{2}\,|V(\omega)|^2\,\mathrm{Re}\{Y_L(\omega)\} \tag{1.5.8}$$

Figure 1.10 shows a model often used for the situation wherein a voltage (or current) source has many discrete frequency components. The load impedance at each frequency is represented by an impedance in series with a filter. The filters $F_1, F_2, \ldots, F_N$ are ideal series-resonant circuits; that is, they are short circuits at their resonant frequencies and open circuits at all other frequencies. Thus, the current component at only one frequency circulates in each branch. One of these branches is the output circuit; the rest may be arbitrary impedances that represent the combined effects of out-of-band filter or matching circuit terminations, package or other circuit parasitics, or in some cases resonances (called *idlers*) that are purposely introduced to optimize performance. These terminations at frequencies other than the output frequency may have a strong effect upon the circuit's performance, so the design of the output network may have to take into account these terminations as well as that at the output frequency.

The gain of a two-port network can be defined in terms of available and dissipated powers. The two most important ways of specifying gain are *transducer gain* and *maximum available gain*. When a microwave engineer speaks loosely of "gain," he or she usually means transducer gain, whether he or she knows it or not. To see why this is so, imagine a technician using a signal generator and power meter to measure the gain of an amplifier. First, the technician connects the power meter to the carefully matched output of the signal generator and notes the power. Because the signal source and the power meter are matched, this is the available power. The technician then connects the signal generator to the amplifier input and the power meter to its output and notes the output power. The output power is the power dissipated in the load, which is not necessarily conjugate-matched to the amplifier's output. The technician

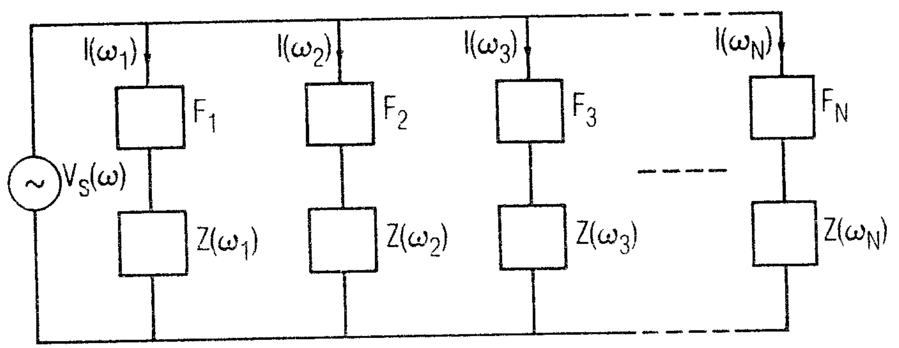

Figure 1.10 Model of a voltage source and load where the excitation has a number of discrete frequency components.

calls the ratio of these powers the *gain*, which in this case is the power delivered to the load divided by the power available from the source. This is precisely the definition of *transducer gain*.

Transducer gain is a very useful concept because, in microwave systems, the important thing to know is how much more or less power a circuit delivers to a standard load (e.g., a 50-Ω coaxial line), compared to the power that could have been obtained from the source alone. This is precisely what transducer gain tells us. Furthermore, transducer gain is almost always a defined quantity, because it requires only that the source and output powers be finite, and real sources always have finite available power. Thus, the concept is handy in nonlinear circuits where, as our discussion of large-signal S parameters illustrated earlier, it is often impossible to define input and output impedances or reflection coefficients.

Other gain definitions are often useless because they do not tell the engineer what he or she wants to know or occasionally result in meaningless or undefined quantities. One such concept is *power gain*, defined as power delivered to the load divided by power delivered to the two-port's input. We find that the power gain of a low-frequency MESFET amplifier, for example, is meaninglessly high: the FET's output power is modest but its input impedance is highly reactive, so the input power is close to zero. This result tells nothing about the way the amplifier works in a system. The concept of power gain can give even more bizarre results when applied to other circuits, such as a negative-resistance amplifier without a circulator. The input power of a negative-resistance device is difficult to define, but we could justifiably say that it is negative and equal to the output power. Thus, the power gain of a negative-resistance amplifier is always −1. Even with these strange results, however, the concept of power gain has some limited usefulness; one of these uses the design of linear amplifiers having prescribed values of transducer gain. This technique is described in Section 8.1.2.

Available gain is defined as the power available from the output divided by the power available from the source. It is intrinsically not a very useful concept (although it will costar with power gain in Section 8.1.2), but its maximum value, called the *maximum available gain*, which occurs when the input of the two-port is conjugately matched to the source, is very useful. The maximum available gain is, therefore, the highest possible value of the transducer gain, which occurs when both the input and output ports are conjugately matched. Maximum available gain is defined only if the two-port is unconditionally stable; that is, if the input and output impedances always have positive real parts when any passive load is connected to the opposite port.

1.6 STABILITY

The fundamental definition of a stable electrical network is that its response is bounded when its excitation is bounded. In the case of a linear two-port having a sinusoidal steady-state excitation, this definition leads to a stability criterion: if the input or output port has an impedance that is the negative of its terminating impedance, the two-port is unstable. If the termination is passive, instability (in fact, oscillation) occurs if the input or output impedance has a negative real part. If the input and output impedance have positive real parts when any passive load is connected to the opposite port, the network is unconditionally stable. If it is possible for the input or output impedance to have a negative real part when a passive termination is connected to the opposite port, the network is conditionally stable.

The situation is more complicated in the case of nonlinear circuits. Because the kinds of interactions that can occur in nonlinear circuits are more complex, such circuits often exhibit transient and steady-state phenomena other than sinusoidal oscillation that, although bounded, are loosely classed as instability. These include parasitic oscillations; spurious outputs that occur only under large-signal excitation; "snap" phenomena, in which the output level or bias conditions change abruptly as input level is varied; and the exacerbation of normal noise levels. Of course, plain, old-fashioned oscillation is also a possibility. Consequently, it is extremely difficult to devise a meaningful and practical stability criterion for nonlinear circuits.

Even without the academic advantage of a stability criterion, it is usually possible, with care, to design nonlinear or quasilinear circuits that are well behaved. For example, if a harmonic-balance analysis of a proposed circuit design converges without incident to a solution, one can be confident that it is stable, by all practical definitions of the term. (It is also stable in theory because harmonic-balance analysis is a process of perturbing the voltages across the nonlinear elements. If these perturbations do not cause larger perturbations, the circuit must be locally stable. The idea that a circuit is stable if such perturbations do not cause greater perturbations is equivalent to the concept of stability defined earlier.) The converse may not be true, however, because the failure of an iterative technique such as harmonic balance to converge may be caused by numerical problems, not by inherent instability. Other tricks and techniques for ensuring stability of individual types of circuits will be covered in their respective chapters.

REFERENCE

[1.1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1970.