CHAPTER 1

Measurement Theory

To understand the operation and use of electronic instruments, it is necessary to have a solid background in the electrical theory associated with electronic measurements. Although it is assumed that the reader understands basic electrical principles (voltage, current, Ohm's Law, etc.), these principles will be reviewed here with special emphasis on how the theory relates to electronic measurement. With this approach, the theory is used to lay the groundwork for discussing the use and operation of electronic instruments. Most of the fundamental concepts apply to multiple types of measurements and instruments.

1.1 Electrical Quantities

The purpose of electronic test instruments is to accurately measure electrical parameters. Before attempting to discuss electronic measurements, we will first make sure that the parameters to be measured are understood. Appendix A contains a table of electrical parameters, their units of measure, and standard abbreviations. The standard electrical units can be modified by the use of prefixes (milli, kilo, etc.), which are also covered in Appendix A. Our initial concern is with the measurement of voltage and current (with an emphasis on voltage); later we include other parameters such as capacitance and inductance.

Current (measured in units of amperes and often abbreviated to "amps") is the flow of electrical *charge* (measured in coulombs). The amount of charge flowing is determined by the number of electrons moving past a given point. An electron has a negative charge of 1.602×10^{-19} coulombs, or equivalently, a coulomb of negative charge consists of 6.242×10^{18} electrons. The unit of current (the ampere) is defined as the number of coulombs of charge passing a given point in a second (one ampere equals one coulomb per second). The more charge that moves in a given time, the higher the current. Even though current is usually made up of moving electrons, the standard electrical engineering convention is to consider current to be the flow of positive charge.[1] With this definition, the current is considered to be flowing in the direction opposite of the electron flow (since electrons are negatively charged).

Voltage (measured in units of volts), also referred to as *electromotive force* (EMF) or *electrical potential*, is the electrical force or pressure that causes the charge to move and the current to flow. Voltage is a relative concept, that is, voltage at a given point must be specified relative to some other reference point, which may be the system common or ground point.

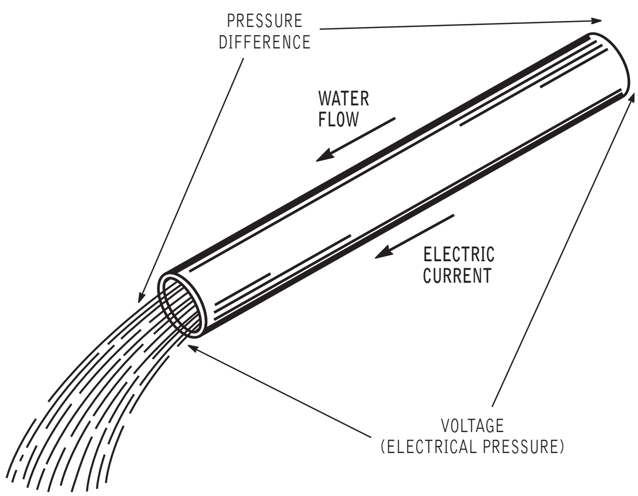


Figure 1.1 The water pipe analogy shows how water flow and pressure difference behave similarly to electrical current and voltage.

An often-used analogy to electrical current is a water pipe with water flowing through it (Figure 1.1). The individual water molecules can be thought of as electrical charge. The amount of water flowing is analogous to electrical current. The water pressure (presumably provided by some sort of external pump) corresponds to electrical pressure or voltage. In this case, the water pressure we are interested in is actually the difference between the two pressures at each end of the pipe. If the pressure (voltage) is the same at both ends of the

1. Current is also sometimes defined as electron flow (i.e., the current flows opposite to the direction used in this book).

pipe, the water flow (current) is zero. On the other hand, if the pressure (voltage) is higher at one end of the pipe, water (current) flows away from the higher pressure end toward the lower pressure end.

Note that while the water flows *through* the pipe, the water pressure is *across* the pipe. In the same way, current flows *through* an electrical device, but voltage (electrical pressure) exists *across* a device (Figure 1.2). This affects the way we connect measuring instruments, depending on whether we are measuring voltage or current. For voltage measurement, the measuring instrument is connected in parallel with the two voltage points (Figure 1.3a). Two points must be specified when referring to a particular voltage and two points are required for a voltage measurement (one of them may be the system common). Measuring voltage at one point only is incorrect. Often, we refer to a voltage at one point when the other point is implied to be the system common or ground point.[2] This is an acceptable practice as long as the assumed second point is made clear.

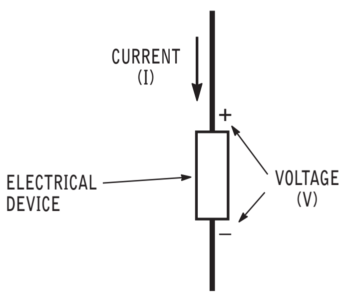


Figure 1.2 Electrical current flows through the device while the voltage exists across the device.

Current, similar to water flow, passes through a device or circuit. When measuring current, the instrument is inserted into the circuit that we are measuring (Figure 1.3b). (There are some exceptions to this, such as current probes, discussed in Chapter 4.) The circuit is broken at the point the current is to be measured and the instrument is inserted in series. This results in the current being measured passing through the measuring instrument. To preserve accuracy for both voltage and current measurements, it is important that the measuring instrument not affect the circuit that is being measured significantly.

2. The concept of an ideal system common or ground can sometimes be misleading as there can be small voltage variations between different locations in the system common.

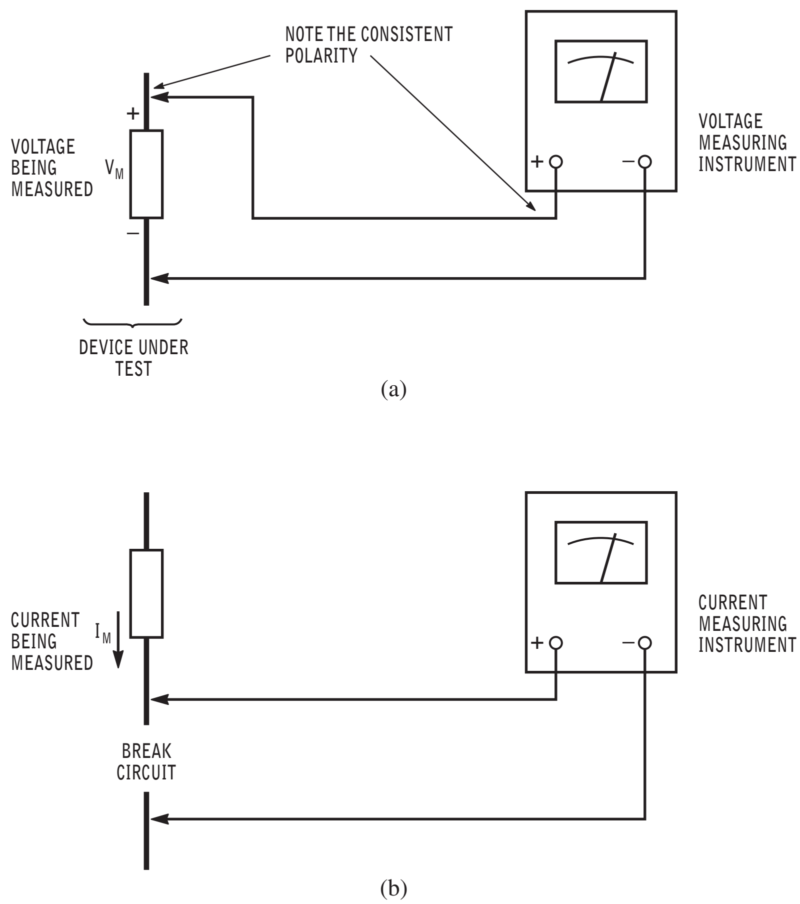


Figure 1.3 a) Voltage measurements are made at two points. Note that the polarity of the measuring instrument is consistent with the polarity of the voltage being measured. b) Current measurements are made by inserting the measuring instrument into the circuit such that the current flows through the instrument.

1.2 Resistance

A resistor is an electrical device that obeys Ohm's Law:

$$V = I \times R$$

$$I = \frac{V}{R}$$

$$R = \frac{V}{I}$$

Ohm's Law simply states that the current through a resistor is proportional to the voltage across that resistor. Returning to the water pipe analogy, as the pressure (voltage) is increased, the amount of water flow (current) also increases. If the voltage is reduced, the current is reduced. Carrying the analogy further, the resistance is related inversely to the size of the water pipe. The larger the water pipe, the smaller the resistance to water flow. A large water pipe (small resistor) allows a large amount of water (current) to flow for a given pressure (voltage). The name "resistor" is due to the behavior of the device: it resists current. The larger the resistor, the more it resists and the smaller the current (assuming a constant voltage).

1.3 Polarity

Some care must be taken to ensure that voltages and currents calculated or measured have the proper direction associated with them. The standard engineering sign convention (definition of the direction of the current relative to the polarity of the voltage) is shown in Figure 1.4. A voltage source (V) is connected to a resistor (R) and some current (I) flows through the resistor. The direction of I is such that for a positive voltage source, a positive current will leave the voltage source at the positive terminal and enter the resistor at its most positive end. (Remember, current is taken to be the flow of positive charge. The electrons would actually be moving opposite the current.) Figure 1.3 shows the measuring instrument connected up in a manner consistent with the directions in Figure 1.4.

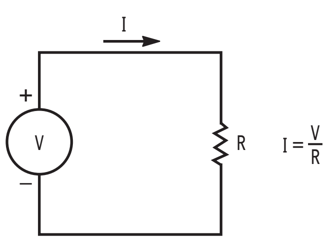


Figure 1.4 Ohm's Law is used to compute the amount of current (I) that will result with voltage (V) and resistance (R).

If the measuring instrument is connected up with the wrong polarity (backwards) when making direct current measurements, the instrument will attempt to measure the proper value, but with the wrong sign (e.g., –5 volts instead of +5 volts). Typically, in digital instruments this is not a problem since a minus sign is just added in front of the reading. In an analog instrument the reading will usually go off the scale and if an electromechanical meter is used it will deflect the meter in the reverse direction, often causing damage. The user should consult the operating manual to understand the limitations of the particular instrument.

Example 1.1

Calculate the amount of voltage across a 5 kΩ resistor if 3 mA of current is flowing through it.

Using Ohm's Law, V = I R = (.003) (5000) = 15 volts

1.4 Direct Current

Direct current (DC) is the simplest form of current. Both current and voltage are constant with respect to time. A plot of DC voltage[3] versus time is shown in Figure 1.5. This plot should seem rather uninformative, since it shows the same voltage for all values of time, but will contrast well with alternating current when it is introduced.

Batteries and DC power supplies produce DC voltage. Batteries are available with a variety of voltage and current ratings. DC power supplies convert AC voltages into DC voltages and are discussed in Chapter 7.

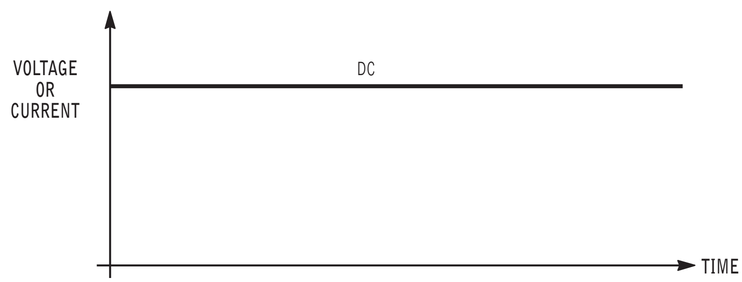


Figure 1.5 DC voltages and currents are constant with respect to time.

1.5 Power

Power is the rate at which energy flows from one circuit to another circuit. For DC voltages and currents, power is simply the voltage multiplied by the current and the unit is the Watt:

$$P = V \times I$$

Using Ohm's Law and some basic math, the relationships in Table 1.1 can be developed. Notice that power depends on both current and voltage. There can be high voltage

3. Even though the term "DC" specifically states direct *current*, it is used to describe both voltage and current. This leads to commonly used, but self-contradictory terminology such as "DC Voltage," which means, literally, "direct current voltage."

but no power if there is no path for the current to flow. Or there could be a large current flowing through a device with zero volts across it, also resulting in no power being received by the device.

Table 1.1 Basic equations for DC voltage, DC current, resistance, and power.

$V = I \times R$	Ohm's Law
$I = \frac{V}{R}$	Ohm's Law
$R = \frac{V}{I}$	Ohm's Law
$P = V \times I$	Power equation
$P = I^2 R$	Power in Resistor
$P = \frac{V^2}{R}$	Power in Resistor

1.6 Alternating Current

Alternating current (AC), as the name implies, does not remain constant with time like direct current, but instead changes direction, or alternates, at some frequency. The most common form of AC is the sine wave, as shown in Figure 1.6. The current or voltage starts out at zero, becomes positive for one half of the cycle, and then passes through zero to become negative for the second half of the cycle. This cycle repeats continuously.

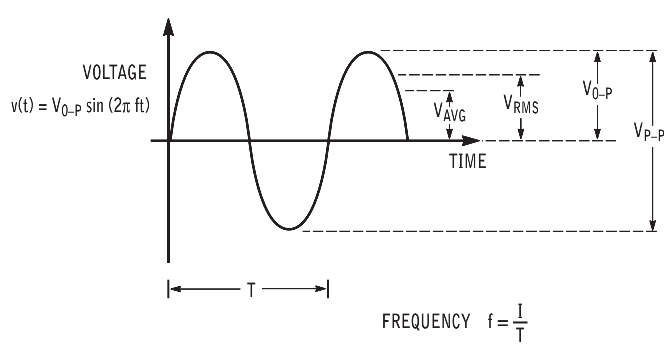


Figure 1.6 The most common form of alternating current is the sine wave. The voltage of the waveform can be described by the RMS value, zero-to-peak value, or the peak-to-peak value.

The sine wave can be described mathematically as a function of time.

$$v(t) = V_{0-P}\sin(2\pi ft)$$

The length of the cycle (in seconds) is called the *period* and is represented by the symbol *T*. The *frequency*, or *f*, is the reciprocal of the period and is measured in units of Hertz, which is equivalent to cycles per second.

$$f = \frac{1}{T}$$

The frequency indicates how many cycles the sine wave completes in one second. For example, the standard AC power line voltage in the United States has a frequency of 60 Hz, which means that the voltage goes through 60 complete cycles in one second. The period of a 60 Hz sine wave is T = 1/f = 1/60 = 0.0167 seconds.

Sometimes the sine wave equation is presented in the following form

$$v(t) = V_{0-P}\sin(\omega t)$$

where ω is the *radian frequency*, with units of radians/second.

By comparing the two equations for the sinusoidal voltage, we can see that

$$\omega = 2\pi f$$

and

$$f = \frac{\omega}{2\pi}$$

Since the sine wave voltage is not constant with time, it is not immediately obvious how to describe it. Sometimes the voltage is positive, sometimes it is negative, and twice every cycle it is zero. This problem does not exist with DC, since it is always a constant value. Figure 1.6 shows four different ways of referring to AC voltage. The *zero-to-peak value* (V_{0-P}) is simply the maximum voltage that the sine wave reaches. (This value is often just called the *peak value*.) Similarly, the *peak-to-peak value* (V_{P-P}) is measured from the maximum positive voltage to the most negative voltage. For a sine wave, V_{P-P} is always twice V_{0-P}.

1.7 RMS Value

Another way of referring to AC voltage is the *RMS value* (V_{RMS}). RMS is an abbreviation for *root-mean-square*, which indicates the mathematics behind calculating the value of an arbitrary waveform. To calculate the RMS value of a waveform, the waveform is first squared at every point. Then, the average or mean value of this squared waveform is found. Finally, the square root of the mean value is taken to produce the RMS (root of the mean of the square) value.

Mathematically, the RMS value of a waveform can be expressed as

$$V_{RMS} = \sqrt{\frac{1}{T}\int_{t_0}^{(t_0+\mathrm{T})} v^2(t)\,dt}$$

Determining the RMS value from the zero-to-peak value (or vice versa) can be difficult due to the complexity of the RMS operation. However, for a sine wave the relationship is simple. (See Appendix B for a detailed analysis.)

$$V_{RMS} = \frac{1}{\sqrt{2}}V_{0-P} = 0.707V_{0-P} \qquad \text{(sine wave)}$$

This relationship is valid only for a sine wave and does *not* hold for other waveforms.

1.8 Average Value

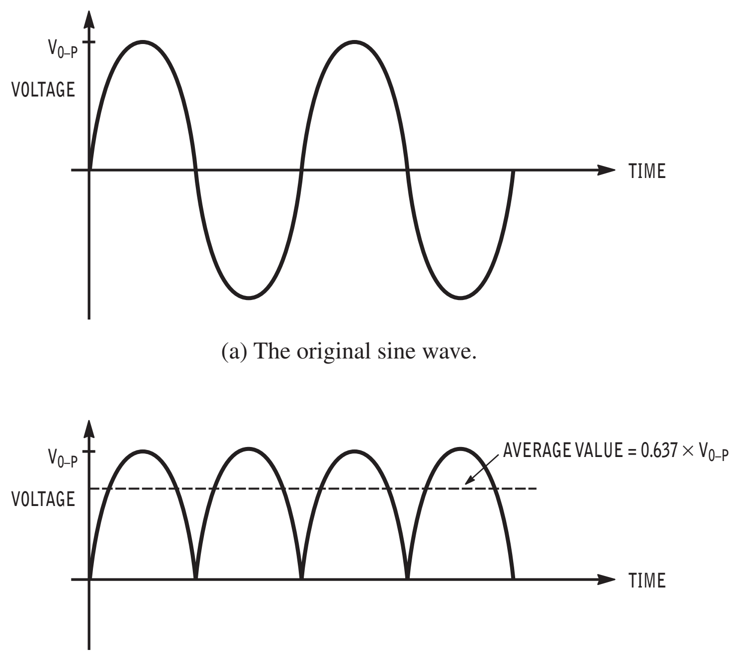
Finally, AC voltage is sometimes defined using an average value. Strictly speaking, the average value of a sine wave is zero because the waveform is positive for one half cycle and is negative for the other half. Since the two halves are symmetrical, they cancel out when they are averaged together. So, on the average, the waveform voltage is zero.

Another interpretation of average value is to assume that the waveform has been full-wave rectified. Mathematically this means that the absolute value of the waveform is used (i.e., the negative portion of the cycle has been treated as being positive).

$$V_{AVG} = \frac{1}{T}\int_{t_0}^{(t_0+T)} |v(t)|\,dt$$

This corresponds to the measurement method that some instruments use to handle AC waveforms, so this is the method that will be considered here. Unless otherwise indicated, V_{AVG} will mean the full-wave rectified average value. These averaging steps have been shown in Figure 1.7. Figure 1.7a shows a sine wave that is to be full-wave rectified. Figure 1.7b shows the resulting full-wave rectified sine wave. Whenever the original waveform becomes negative, full-wave rectification changes the sign and converts the voltage into a positive waveform with the same amplitude. Graphically, this can be described as folding the negative half of the waveform up onto the positive half, resulting in a humped sort of waveform. Now the average value can be determined and is plotted in Figure 1.7b. The relationship between V_{AVG} and V_{0-P} depends on the shape of the waveform. For a sine wave[4]

$$V_{AVG} = \frac{2}{\pi} V_{0-P} = 0.637 V_{0-P} \qquad \text{(sine wave)}$$



(a) The original sine wave.

(b) The full-wave rectified version of the sine wave.

Figure 1.7 The operations involved in finding the full-wave average value of a sine wave. a) The original sine wave. b) The full-wave rectified version of the sine wave. The negative half cycle is folded up to become positive. The resulting waveform is then averaged.

The details of the analysis are shown in Appendix B.

1.9 Crest Factor

The ratio of the zero-to-peak value to the RMS value of the waveform is known as the *crest factor*. Crest factor is a measure of how high the waveform peaks, relative to its RMS value. The crest factor of a waveform is important in some measuring instruments. Wave-

4. $\pi \approx 3.14159$.

forms with very high crest factors require the measuring instrument to tolerate very large peak voltages while simultaneously measuring the much smaller RMS value.

$$Crest\ factor = \frac{V_{0-P}}{V_{RMS}}$$

Example 1.2

What is the crest factor of a sine wave?
For a sine wave, $V_{RMS} = 0.707\ V_{0-P}$ so the crest factor is 1/0.707 = 1.414.
The sine wave has a relatively low crest factor. Its zero-to-peak value is not that much greater than its RMS value.

The *peak-to-average ratio*, also known as *average crest factor*, is similar to crest factor except that the average value of the waveform is used in the denominator of the ratio. It also is a measure of how high the peaks of the waveform are compared to its average value.

Although the preceding discussion has referred to AC voltages, the same concepts apply to AC currents. That is, AC currents can be described by their zero-to-peak, peak-to-peak, RMS, and average values.

1.10 Phase

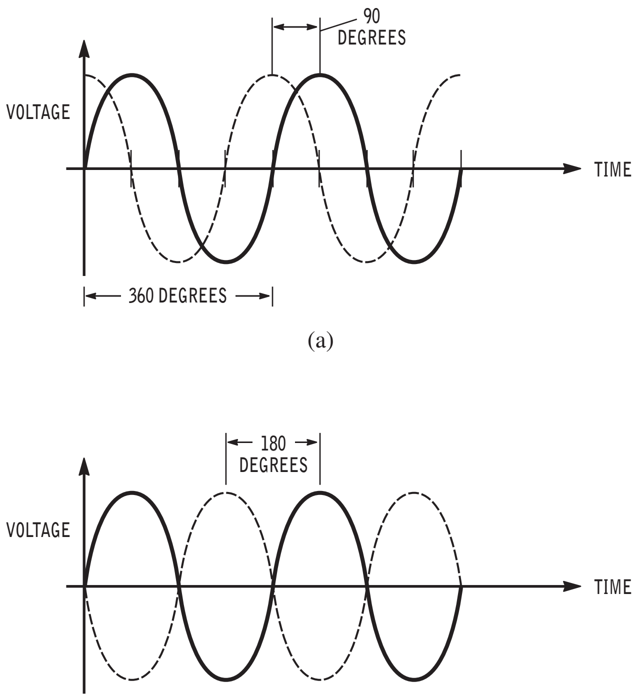
The voltage specifies the amplitude or height of the sine wave; and the frequency or the period specifies how often the sine wave completes a cycle. But two sine waves of the same frequency may not cross zero at the same time. Therefore, the *phase* of the sine wave is used to define its position on the time axis. The most common unit of phase is the degree, with one cycle of a sine wave divided up into 360 degrees of phase. Phase may also be expressed in units of radians, with one cycle corresponding to 2π radians.

The mathematical definition of the sine wave can be modified to include a phase term.

$$v(t) = V_{0-P}\sin(2\pi ft + \theta)$$

As written, the equation implies that phase is absolute. That is, there is some instant in time when $t = 0$—the reference for the phase angle. In practice, there usually is no such universal time and phase is a relative concept. In other words, we can refer to the phase between two sine waves, but not the phase of a single, isolated sine wave (unless some other time reference is supplied).

For example, the two sine waves in Figure 1.8a are separated by one-fourth of a cycle (they reach their maximum values one-fourth of a cycle apart). Since one cycle equals 360 degrees, the two sine waves have a phase difference of 90 degrees. To be more precise, the



(b)

Figure 1.8 The phase of a sine wave defines its relative position on the time axis. a) The phase between the two sine waves is 90 degrees. b) The two sine waves are 180 degrees out of phase.

second sine wave is 90 degrees behind the first or, equivalently, the second sine wave has a phase of –90 degrees relative to the first sine wave. (The first sine wave has a +90 degrees phase relative to the second sine wave.) It is also correct to say that the first sine wave *leads* the second one by 90 degrees or that the second sine wave *lags* the first one by 90 degrees. All of these statements specify the same phase relationship.

Figure 1.8b shows two sine waves that are one half cycle (180 degrees) apart. This is the special case where one sine wave is the negative of the other. The phase relationship between two sine waves simply defines how far one sine wave is shifted with respect to the other. When a sine wave is shifted by 360 degrees, it is shifted a complete cycle and is indistinguishable from the original waveform. Because of this, phase is usually specified over a 360-degree range, typically –180 degrees to +180 degrees.

1.11 AC Power

The concept of power gets more complicated when AC waveforms are considered. The *instantaneous power* is defined by multiplying the instantaneous voltage and instantaneous current:

$$p(t) = v(t) \cdot i(t)$$

In many cases, we are interested in the *average power* being transferred. The average power is determined by taking the instantaneous power and averaging it over a period of the waveform.

$$P = \frac{1}{T}\int_0^T p(t)dt = \frac{1}{T}\int_0^T v(t) \cdot i(t)dt$$

The average power dissipated by a resistor with an AC voltage across it is given by:

$$P = V_{RMS} I_{RMS} = \frac{V_{RMS}^2}{R} = I_{RMS}^2 \cdot R$$

This relationship holds for any waveform as long as the RMS value of the voltage and current are used. Note that these equations have the same form as the DC case, which is one of the reasons for using RMS values. The RMS value is often called the *effective value*, since an AC voltage with a given RMS value has the same effect (in terms of power) that a DC voltage with that same value. (A 10-volt RMS AC voltage and a 10-volt DC voltage both supply the same power, 20 watts, to a 5 Ω resistor.) In addition, two AC waveforms that have the same RMS value will cause the same power to be delivered to a resistor. This is *not* true for other voltage descriptions such as zero-to-peak and peak-to-peak. Thus, RMS is the great equalizer with respect to power.

Example 1.3

The standard line voltage in the United States is approximately 120 volts RMS. What are the zero-to-peak, peak-to-peak, and full-wave rectified average voltages? How much power is supplied to a 200 Ω resistor connected across the line?

$V_{RMS} = 0.707\ V_{0-P}$

so $V_{0-P} = V_{RMS}/0.707 = 120/0.707 = 169.7$ volts

$V_{P-P} = 2\ V_{0-P} = 2\ (169.7) = 339.4$ volts

$V_{AVG} = 0.637\ V_{0-P} = 0.637 \times 169.7 = 108.1$ volts

$P = V_{RMS}^2/R = 120^2\ /200 = 72$ watts

1.12 Nonsinusoidal Waveforms

There are other AC voltage and current waveforms besides sine waves that are commonly used in electronic systems. Some of the more common waveforms are shown in Table 1.2. Note that the values of V_{RMS} and V_{AVG} (relative to V_{0-P}) are unique for each waveform. The first three waveforms are symmetrical about the horizontal axis, but the half-sine wave and the pulse train are always positive. All of these waveforms are *periodic* because they repeat the same cycle or period continuously.

An example will help to emphasize the utility of RMS voltages when dealing with power in different waveforms.

Example 1.4

A sine wave voltage and a triangle wave voltage are each connected across two separate 300 Ω resistors. If both waveforms deliver 2 watts (average power) to their respective resistors, what are the RMS and zero-to-peak voltages of each waveform?

The two waveforms deliver the same power to identical resistors, so their RMS voltages must be the same. (This is *not* true of their zero-to-peak values.)

$$P = \frac{V_{RMS}^2}{R}$$

$$V_{RMS} = \sqrt{P \cdot R} = \sqrt{2 \cdot 300} = 24.5 \text{ volts RMS}$$

From Table 1.2, for the sine wave

$$V_{RMS} = 0.707 V_{0-P}$$

$$V_{0-P} = \frac{V_{\mathrm{RMS}}}{0.707} = \frac{24.5}{0.707} = 34.7 \text{ volts}$$

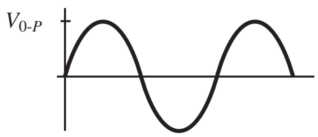
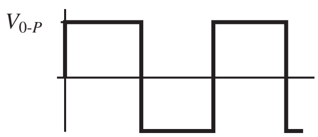
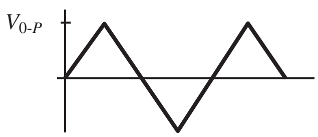
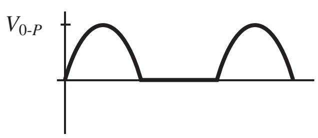
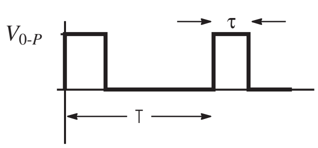
For the triangle wave

$$V_{RMS} = 0.577 V_{0-P}$$

$$V_{0-P} = V_{RMS}/0.577 = 24.5/0.577 = 42.5 \text{ volts}$$

So for the triangle wave to supply the same average power to a resistor, it must reach a higher peak voltage than the sine wave.

Table 1.2 Table of waveforms with peak-to-peak voltage (V_{P-P}), RMS voltage (V_{RMS}), zero-to-peak voltage (V_{0-P}), and crest factor for each waveform. V_{AVG} is the full-wave rectified average value of the waveform.

	Waveform	V_{P-P}	V_{RMS}	V_{AVG}	Crest Factor
	Sine Wave	$2\,V_{0-P}$	$\frac{1}{\sqrt{2}}V_{0-P}$ or $0.707V_{0-P}$	$\frac{2}{\pi}V_{0-P}$ or $0.637V_{0-P}$	$\sqrt{2}$ or 1.414
	Square Wave	$2\,V_{0-P}$	V_{0-P}	V_{0-P}	1
	Triangle Wave	$2\,V_{0-P}$	$\frac{1}{\sqrt{3}}V_{0-P}$ or $0.577V_{0-P}$	$\frac{1}{2}V_{0-P}$	$\sqrt{3}$ or 1.732
	Half Sine Wave	V_{0-P}	$\frac{1}{2}V_{0-P}$	$\frac{1}{\pi}V_{0-P}$ or $0.318V_{0-P}$	2
	Pulse Train	V_{0-P}	$\sqrt{\frac{\tau}{T}}V_{0-P}$	$\frac{\tau}{T}V_{0-P}$	$\sqrt{\frac{T}{t}}$

1.13 Harmonics

Periodic waveforms, except for absolutely pure sine waves, contain frequencies called harmonics. Harmonic frequencies are integer multiples of the original or *fundamental* frequency.

$$f_n = n \cdot f_{fundamental}$$

For example, a nonsinusoidal waveform with a fundamental frequency of 1 kHz has harmonics at 2 kHz, 3 kHz, 4 kHz, and so on. There can be any number of harmonics, out to infinity, but usually there is a practical limitation on how many need be considered. Each harmonic may have its own unique phase relative to the fundamental.

Harmonics are present because periodic waveforms, regardless of shape, can be broken down mathematically into a series of sine waves. However, this behavior is more than just mathematics. It is as if the physical world regards the sine wave as the purest, most simple sort of waveform with all other periodic waveforms being made up of collections of sine waves. The result is that a periodic waveform (such as a square wave) is exactly equivalent to a series of sine waves.

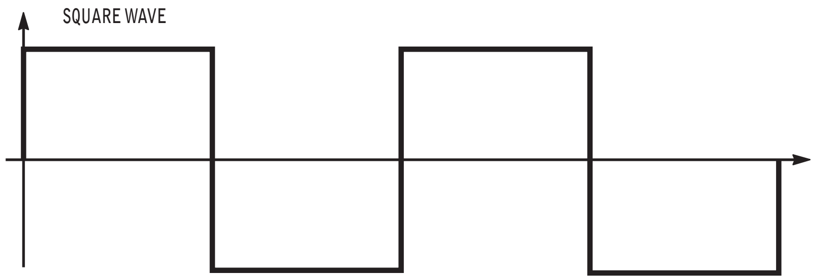
1.14 Square Wave

Consider the square wave shown in Figure 1.9a. Square waves are made up of the fundamental frequency plus an infinite number of odd harmonics.[5] In theory, it takes every one of those infinite numbers of harmonics to create a true square wave. In practice, the fundamental and several harmonics will approximate a square wave. In Figure 1.9b, the fundamental, third harmonic, and fifth harmonic are plotted. (Remember, the even harmonics of a square wave are zero.) Note that the amplitude of each higher harmonic is less than the previous one, so the highest harmonics may be small enough to be ignored. When these harmonics are added up, they produce a waveform that resembles a square wave. Figure 1.9c shows the waveform that results from combining just the fundamental and the third harmonic. Already the waveform starts to look somewhat like a square wave (well, a least a little bit). The fundamental plus the third and fifth harmonics is shown in Figure 1.9d. It is a little more like a square wave. Figures 1.9e and 1.9f each add another odd harmonic to our square wave approximation and each one resembles a square wave more closely than the previous waveform. If all of the infinite number of harmonics were included, the resulting waveform would be a perfect square wave. So the quality of the square wave is limited by the number of harmonics present.

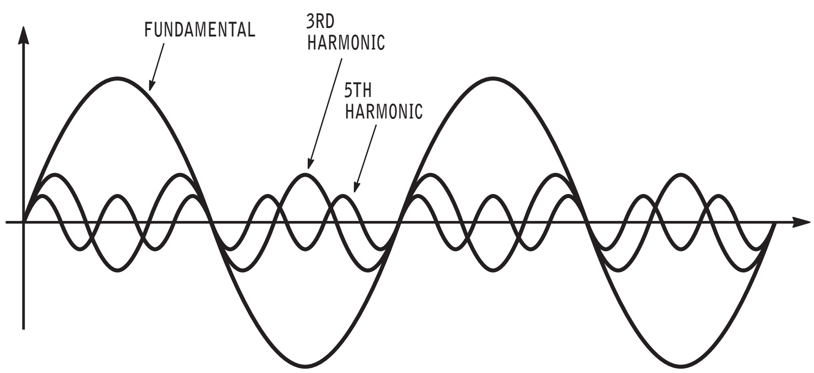
The amplitude of each harmonic must be just the right value for the resulting wave to be a square wave. In addition, the phase relationships between the harmonics must also be correct. If the harmonics are delayed in time by unequal amounts, the square wave will take on a distorted look even though the amplitudes of the harmonics may be correct. This phenomenon is used to advantage in square wave testing of amplifiers, as discussed in Chapter 5. It is theoretically possible to construct a square wave electronically by connecting a large number of sine wave generators together such that each one contributes the fundamental frequency or a harmonic with just the right amplitude. In practice, this may be difficult because the frequency and phase of each oscillator must also be precisely controlled.

So far, waveforms have been characterized using a voltage versus time plot, which is known as the *time domain* representation. Another way of describing the same waveform is with frequency on the horizontal axis and voltage on the vertical axis. This is known as the

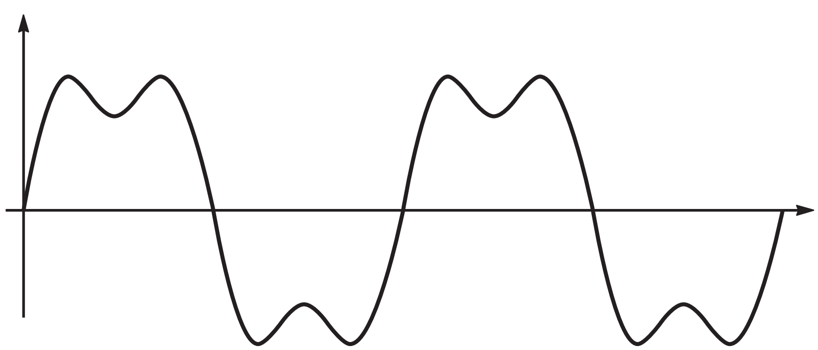
5. For a mathematical derivation of the frequency content of a square wave, see Witte, 2001.



(a)

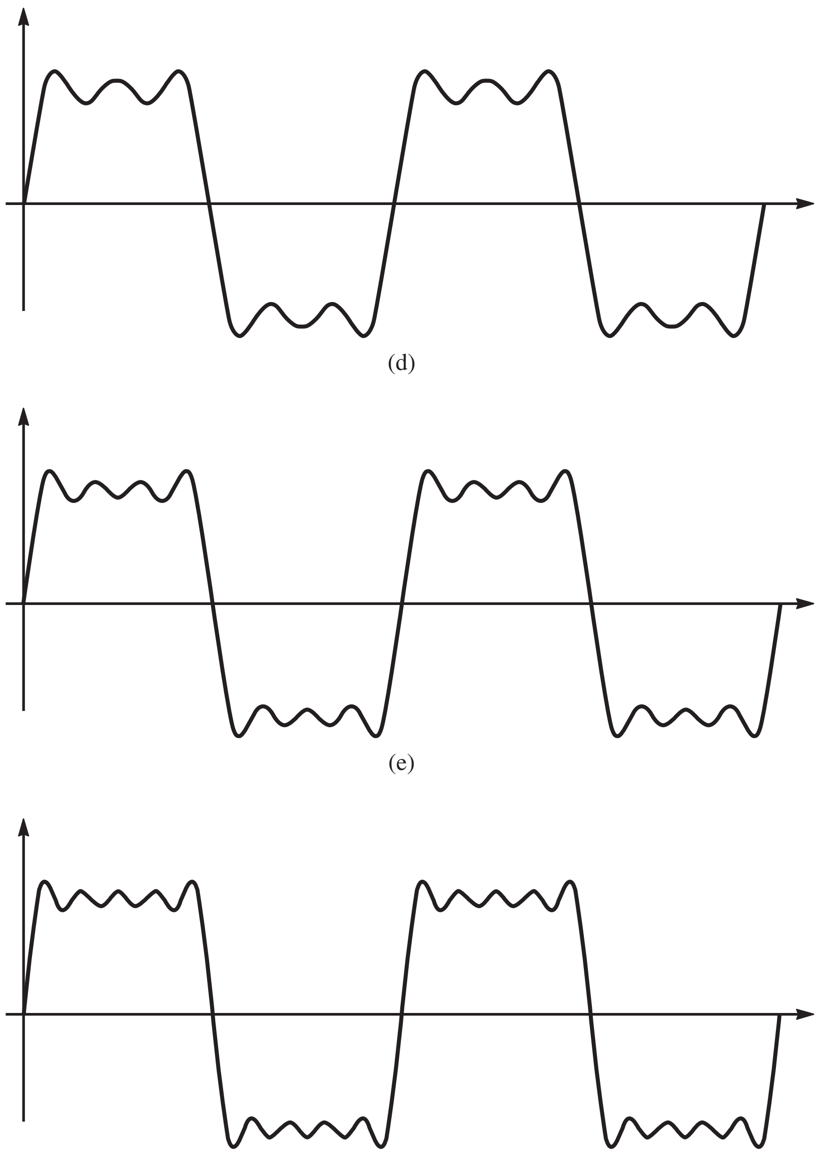


(b)



(c)

Figure 1.9 The square wave can be broken up into an infinite number of odd harmonics. The more harmonics that are included, the more the waveform approximates a square wave. a) The original square wave. b) The fundamental, third harmonic, and fifth harmonic. c) The fundamental plus third harmonic. c) The fundamental plus the third and fifth harmonics. d) The fundamental plus the third, fifth, and seventh harmonics. e) The fundamental plus the third, fifth, seventh, and ninth harmonics.



(f)

Figure 1.9 (Continued)

frequency domain representation or *spectrum* of the waveform. In the frequency domain representation, a vertical line (called a *spectral line*) indicates a particular frequency that is present (the fundamental or a harmonic). The height of each spectral line corresponds to the amplitude of that particular harmonic. A pure sine wave would be represented by one single spectral line. Figure 1.10 shows the frequency domain representation of a square wave. Notice that only the fundamental and odd harmonics are present and that each harmonic is smaller than the previous one. Understanding the spectral content of waveforms is important because the measurement instrument must be capable of operating at the frequencies of the harmonics (at least the ones that are to be included in the measurement).

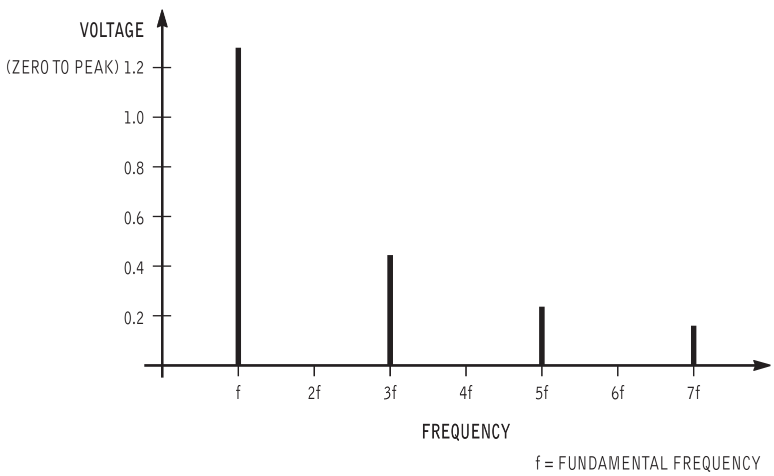


Figure 1.10 The frequency domain representation of a square wave, shown out to the seventh harmonic.

1.15 Pulse Train

The *pulse train*, or *repetitive pulse*, is a common signal in digital systems (Figure 1.11). It is similar to the square wave, but does not have both positive and negative values. Instead it has two possible values: V_{0-P} and 0 volts. The square wave spends 50% of the time at its positive voltage and 50% of the time at its negative voltage, corresponding to a 50% *duty cycle*. The pulse train's duty cycle may be any value between 0 and 100 percent and is defined by the following equation:

$$Duty\ Cycle = \frac{\tau}{T}$$

where τ is the length of time that the waveform is high and T is the period of the waveform.

The pulse train generates harmonic frequencies with amplitudes that are dependent on the duty cycle. A frequency domain plot of a typical pulse train (duty cycle of 25%) is shown in Figure 1.12. The envelope of the harmonics has a distinct humped shape that equals zero at integer multiples of $1/\tau$. Most of the waveform's energy is contained in the harmonics falling below this $1/\tau$ point. Therefore, it is often used as a rule of thumb for the bandwidth of the waveform.

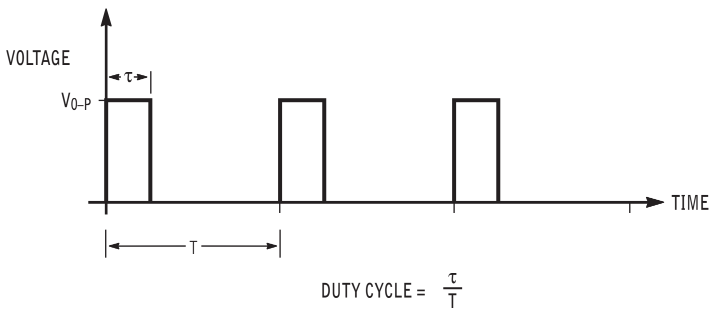


Figure 1.11 The pulse train is a common waveform in digital systems. The duty cycle describes the percent of the time that the waveform is at the higher voltage.

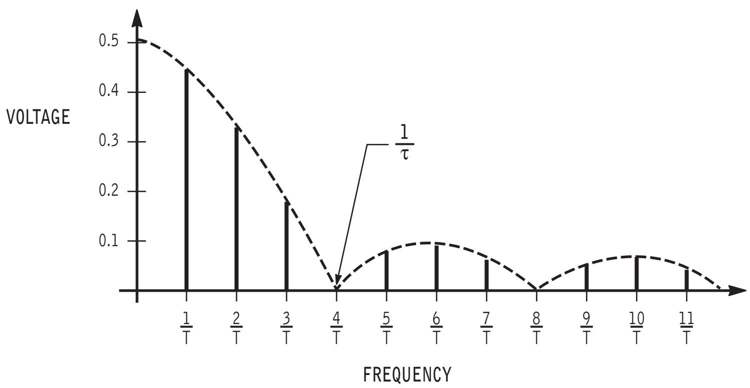


Figure 1.12 The frequency domain plot for a pulse train with a 25% duty cycle. The envelope of the harmonics falls to zero when the frequency equals 1/t.

Table 1.3 summarizes the harmonic content for each of the various waveforms with V_{0-P} = 1 volt. In all of the waveforms except the sine wave, there are an infinite number of harmonics but the amplitudes of the harmonics tend to decrease as the harmonic number increases. At some point, the higher harmonics can be ignored for practical systems because they are so small. As a measure of how wide each waveform is in the frequency domain, the last column lists the number of significant harmonics (ones that are at least 10% of the fundamental). The larger the number of significant harmonics, the wider the signal is in the frequency domain. The sine wave, of course, only has one significant harmonic (the fundamental can be considered the “first” harmonic). Notice that the amplitude value of the fundamental frequency component changes depending on the waveform, even though all of the waveforms have V_{0-P} = 1 volt.

Table 1.3 Table of harmonics for a variety of waveforms. All waveforms have a zero-to-peak value of 1. The number of significant harmonics column lists the highest harmonic whose amplitude is at least 10% of the fundamental.

Waveform	Fund.	Harmonics 2nd	3rd	4th	5th	6th	7th	Sig. Harm. (10%)	Equation
Sine wave	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1	
Square	1.273	0.000	0.424	0.000	0.255	0.000	0.182	9	$\frac{4}{n\pi}$ for odd n
Triangle	0.811	0.000	0.090	0.000	0.032	0.000	0.017	3	$\frac{8}{n^2\pi^2}$ for odd n
Pulse (50% duty cycle)	0.637	0.000	0.212	0.000	0.127	0.000	0.091	9	$\frac{2}{n\pi}\sin\left(\frac{n\pi}{2}\right)$
Pulse (25% duty cycle)	0.450	0.318	0.150	0.000	0.090	0.105	0.064	14	$\frac{2}{n\pi}\sin\left(\frac{n\pi}{4}\right)$
Pulse (10% duty cycle)	0.197	0.187	0.172	0.151	0.127	0.101	0.074	26	$\frac{2}{n\pi}\sin\left(\frac{n\pi}{10}\right)$

There are two important concepts to be obtained from the previous discussion of harmonics.

1. The more quickly a waveform transitions between its minimum and maximum values, the more significant harmonics it will have and the higher the frequency content of the waveform. For example, the square wave (which has abrupt voltage changes) has more significant harmonics than the triangle wave (which does not change nearly as quickly).
2. The narrower the width of a pulse, the more significant harmonics it will have and the higher the frequency content of the waveform.

Both of these statements should make intuitive sense if one considers that high frequency signals change voltage faster than lower frequency signals. The two conditions cited above both involve waveforms changing voltage in a more rapid manner. Therefore, it makes sense that waveforms that must change rapidly (those with short rise times or narrow pulses) will have more high frequencies present in the form of harmonics.

Example 1.5

What is the highest frequency that must be included in the measurement of a triangle wave that repeats every 50 microseconds (assuming that harmonics less than 10% of the fundamental can be ignored)?

The fundamental frequency, f = 1/T = 1/50 μsec = 20 kHz

From Table 1.3, the number of significant harmonics for a triangle wave (using the 10% criterion) is 3. The highest frequency that must be included is 3 f = 3 (20 kHz) = 60 kHz.

1.16 Combined DC and AC

There are many cases where a waveform is not purely DC nor purely AC, but can be thought of as being a combination of the two. For example, in transistor circuits, an AC waveform is often superimposed on a DC bias voltage, as shown in Figure 1.13. The DC component (a straight horizontal line) and the AC component (a sine wave) combine to form the new waveform.

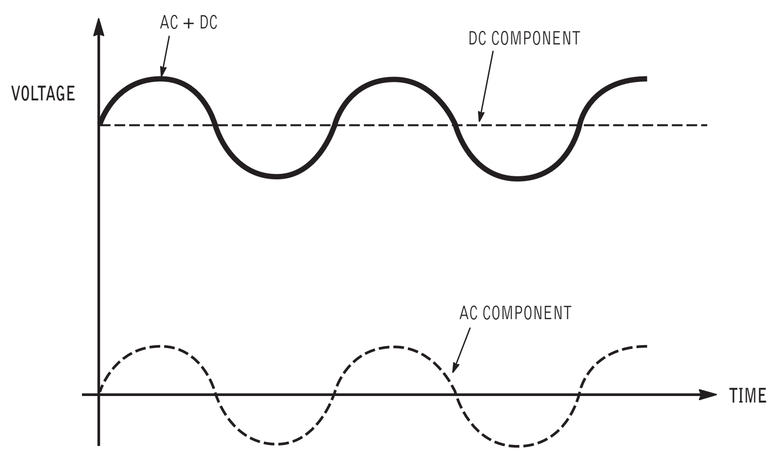


Figure 1.13 The waveform shown can be broken down into a DC component and an AC component. The DC component is just the average value of the original waveform.

The DC value of a waveform is also just the waveform's average value. In Figure 1.13, the waveform is above its DC value half of the time and below the DC value the other half, so on the average the voltage is just the DC value.

The pulse train shown in Figure 1.14 has a 50% duty cycle and is always positive in value. The average value of this waveform is, therefore, greater than zero. This waveform spends half of the time at V_{0-P} and the other half at 0, so the average or DC value =

$(V_{0-P} + 0)/2 = 1/2\ V_{0-P}$. The AC component left over when the DC component is removed is a square wave with half the zero-to-peak value of the original pulse train. In summary, the 50% duty cycle pulse train is equivalent to a square wave of half the zero-to-peak voltage plus a 1/2 V_{0-P} DC component.

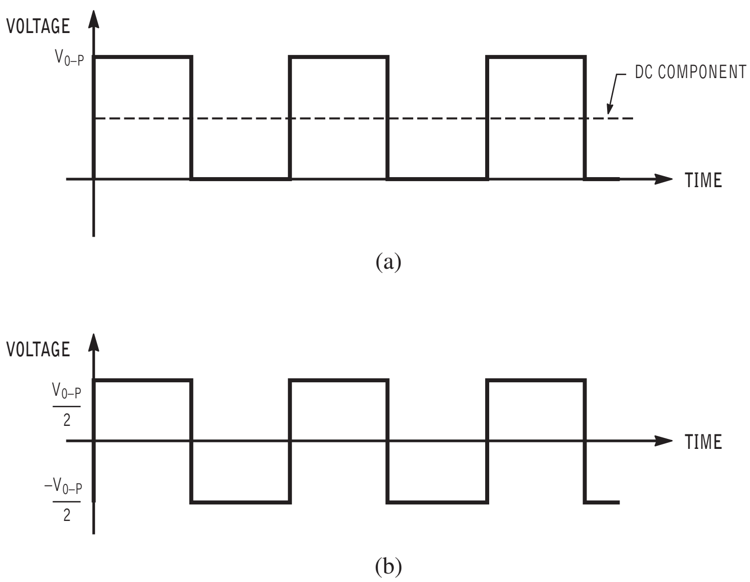


Figure 1.14 The 50% duty cycle pulse train is equivalent to a square wave plus a DC voltage. a) The pulse train with DC component shown. b) The square wave that results when the DC component is removed from the pulse train.

Example 1.6

A DC power supply has some residual AC riding on top of its DC component, as shown in Figure 1.15. Assuming that the AC component can be measured independently of the DC component, determine both the DC and AC values that would be measured (given the RMS value for the AC component).

The DC value is simply the average value of the waveform. Since the AC component is symmetrical, the average value can be calculated:

DC value = (10.6 + 10.2)/2 = 10.4 volts
If the DC is removed from the waveform, a triangle wave with V_{0-P} = 0.2 volts is left.

For a triangle wave, $V_{RMS} = 0.577\ V_{0-P}$ = 0.115 volts RMS

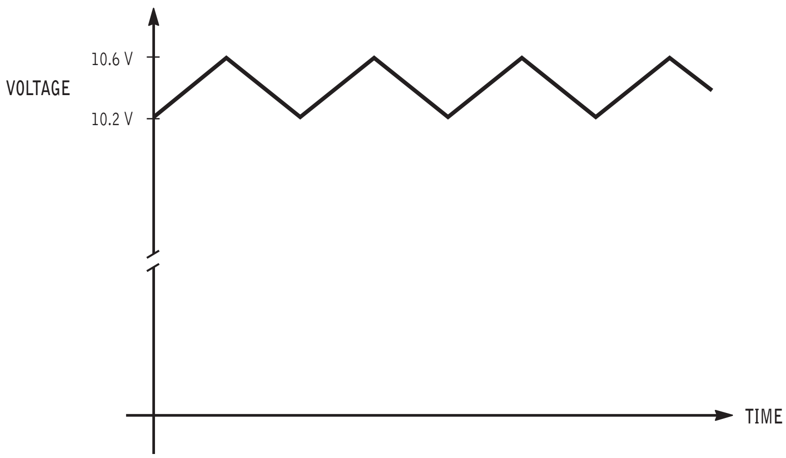


Figure 1.15 DC voltage with a residual triangle wave voltage superimposed on top of it (Example 1.6).

1.17 Modulated Signals

Sometimes sine waves are modulated by another waveform. For example, communication systems use this technique to superimpose low-frequency voice or data signals onto a high-frequency carrier that can be transmitted long distances. Modulation is performed by modifying some parameter of the original sine wave (called the *carrier*), depending on the value of the modulating waveform. In this way, information from the modulating waveform is transferred to the carrier. The most common forms of modulation used are amplitude, frequency, and phase modulation.

1.17.1 Amplitude Modulation

In *amplitude modulation* (AM), the amplitude of the carrier is determined by the modulating waveform. An amplitude modulated signal is represented by the equation

$$v(t) = A_c[1 + am(t)]\cos(2\pi f_c t)$$

where

A_c = the signal amplitude

a = modulation index ($0 \le a \le 1$)

$m(t)$ = normalized modulating signal (maximum value is 1)

f_c = carrier frequency

Figure 1.16a shows the original sine wave carrier with constant amplitude. Figure 1.16b is the modulating waveform. Figure 1.16c is the modulated carrier with the modulating waveform determining the amplitude of the carrier such that the modulating waveform

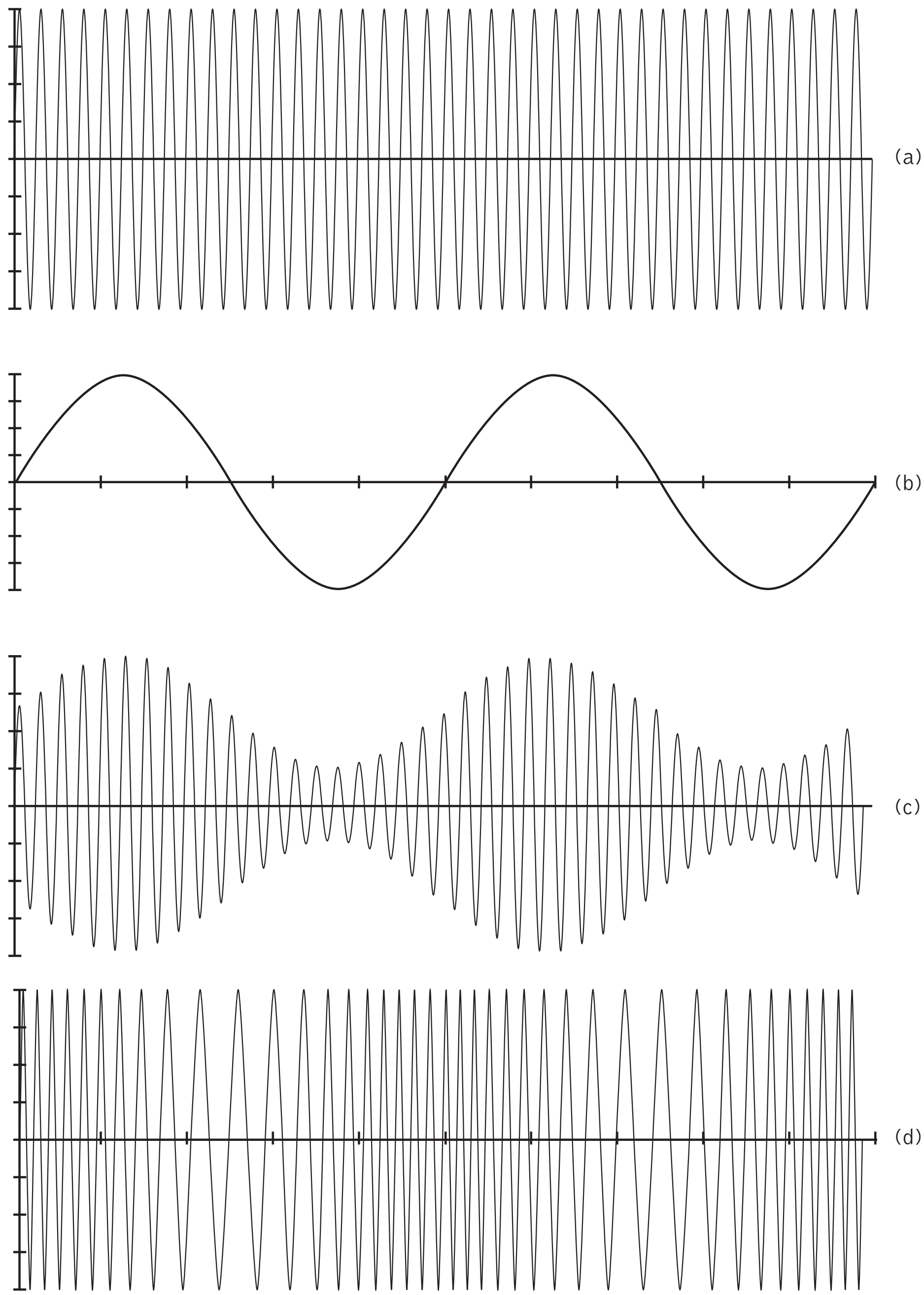


Figure 1.16 Amplitude and frequency modulation. a) The original carrier. b) The modulating signal. c) The resulting signal if amplitude modulation is used. d) The resulting signal if frequency modulation is used.

can be seen as the envelope of the modulated waveform. When the modulating waveform increases, the amplitude of the carrier increases and when the modulating waveform decreases, the amplitude of the carrier decreases.

1.17.2 Frequency Modulation

Frequency modulation (FM) also modulates a carrier, but the amplitude of the carrier remains constant while the frequency of the carrier changes. When the modulating waveform increases, the carrier frequency increases; when the modulating waveform decreases, the instantaneous carrier frequency decreases. A frequency-modulated signal is shown in Figure 1.16d.

Phase modulation is similar to frequency modulation. The carrier amplitude remains constant and the phase is changed according to the modulating waveform. Adjusting the phase of the carrier is somewhat like changing its frequency, so the effect on the carrier is similar. For many practical applications, phase and frequency modulation may be indistinguishable. FM and PM are generally considered to be variants of a type of modulation called *angle modulation*.

1.17.3 Modulated Signals in the Frequency Domain

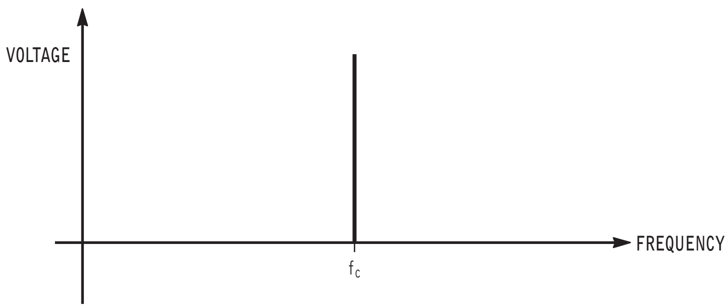
Modulating a carrier, either with AM or FM, has an effect in the frequency domain (Figure 1.17). With no modulation, the carrier is just a pure sine wave and exists only at one frequency (Figure 1.17a). When the modulation is added, the carrier is accompanied by a band of other frequencies, called *sidebands*. The exact behavior of these sidebands depends on the level and type of modulation, but in general the sidebands spread out on one or both sides of the carrier (Figure 1.17b). Often they are relatively close to the carrier frequency, but in some cases, such as wideband FM, the sidebands spread out much farther. In all cases, modulation has the effect of spreading the carrier out in the frequency domain, occupying a wider bandwidth.

When a carrier is amplitude modulated by a single sine wave with frequency f_m, two modulation sidebands appear (offset by f_m) on both sides of the carrier (Figure 1.17c). This can be shown mathematically by inserting a sinusoid with frequency f_m into the equation for an amplitude-modulated signal:

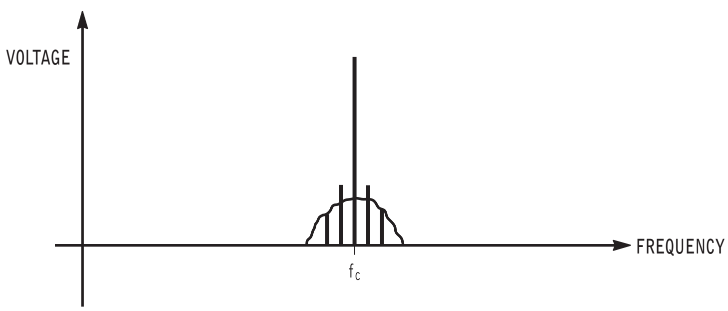
$$v(t) = A_c[1 + a\cos(2\pi f_m t)]\cos(2\pi f_c t)$$

Using trigonometric identities, this equation can be rewritten as

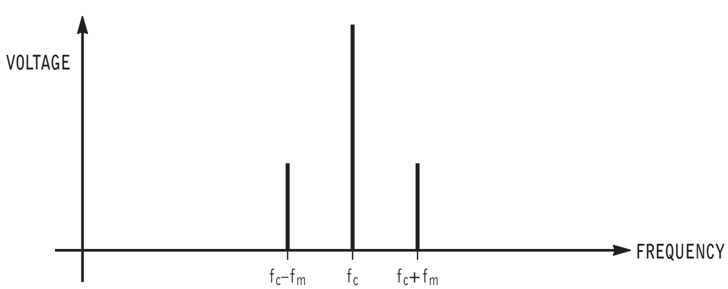
$$v(t) = A_c\cos(2\pi f_c t) + \frac{aA_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{aA_c}{2}\cos[2\pi(f_c - f_m)t]$$



(a)

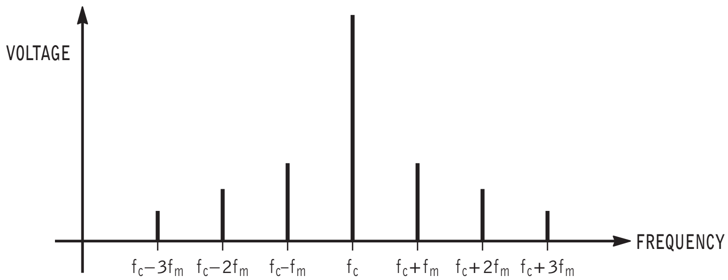


(b)



(c)

Figure 1.17 The effect of modulation in the frequency domain is to spread out the signal by creating sidebands. a) The unmodulated carrier is a single spectral line. b) Modulation on the carrier spreads the signal out in the frequency domain. c) An amplitude-modulated carrier (with single sine wave modulation) has a single pair of sidebands. d) A frequency-modulated carrier (with single sine wave modulation) may have many pairs of sidebands.



(d)

Figure 1.17 (Continued)

which gives us a frequency domain representation as consisting of the carrier frequency, f_c plus two sidebands at frequencies $f_c + f_m$ and $f_c - f_m$. The amplitudes of the two sidebands are the same and they depend on the modulation index, a.

Carriers that are frequency modulated may have a considerably larger number of sidebands.[6] With a modulating frequency of $f_{m,}$ the sidebands appear at integer multiples of f_m away from the carrier frequency. In the general case, FM is a wideband form of modulation, in that the modulated carrier has sidebands that extend out farther than fm away from the carrier. If the modulation occurs at a low level, FM may be narrowband, with sidebands very similar to the AM case. For both modulation schemes, when the modulating signal is complex (voice waveforms, multiple sine waves, etc.), the sidebands are also more complex, occupying a continuum of frequencies adjacent to the carrier.

Communications systems represent intentional use of modulation but sometimes modulation is produced unintentionally due to circuit imperfections. In this case, the modulation is undesirable, or at least unnecessary. The effect in the frequency domain is still the same, however, so a sine wave with some residual modulation on it have sidebands on it in the frequency domain, causing the signal to spread out.

6. The mathematical analysis of frequency modulation is complex and will not be covered in this book. For more information on FM, see Schwartz, 1993.

1.18 Decibels

The decibel (dB) is sometimes used to express electrical quantities in a convenient form. The definition of the decibel is based on the ratio of two power levels (log indicates the base ten logarithm):

$$dB = 10 \log\left[\frac{P_2}{P_1}\right]$$

Since $P = V^2/R$ (assuming RMS voltage),

$$dB = 10 \log\left[\frac{V_2^2/R_2}{V_1^2/R_1}\right]$$

And if $R_1 = R_2$ (the resistances involved are the same)

$$dB = 10 \log\left[\frac{V_2}{V_1}\right]^2$$

$$dB = 20 \log\left[\frac{V_2}{V_1}\right]$$

The voltages, V_1 and V_2, are normally RMS voltages. If the two waveforms are the same shape then the voltages can be expressed as zero-to-peak or peak-to-peak. Since many of our measurements are voltage measurements, calculating dB using voltages is the most convenient method. Strictly speaking, the voltage equation is valid only if the two resistances involved are the same. Misleading results can occur when $R_1 \neq R_2$ and the decibel equation is applied to both voltage and power. For example, the voltage gain calculation (in dB) of an amplifier often involves different resistances at the input and the output. However, it is common practice to apply the decibel equation to the voltage gain even though the two resistances are unequal.

To convert dB values back into a voltage or power ratio, a little reverse math is required:

$$\frac{P_2}{P_1} = 10^{(dB/10)}$$

$$\frac{V_2}{V_1} = 10^{(dB/20)}$$

The usefulness of decibels is sometimes questioned but they are widely used and must be understood by instrument users for that reason alone. In addition, they have at least two characteristics that make them very convenient:

1. Decibels compress widely varying electrical values onto a more manageable logarithmic scale. The range of powers extending from 100 watts down to 1 microwatt is a ratio of 100,000,000, but is expressed in dB as only 80 dB.
2. Gains and losses through circuits such as attenuators, amplifiers, and filters, when expressed in dB can be added together to produce the total gain or loss. To perform the equivalent operation without decibels requires multiplication.

Some cardinal decibel values are worth pointing out specifically:

- 0 dB corresponds to a ratio of 1 (for both voltage and power). A circuit that has 0-dB gain or 0-dB loss has an output equal to the input.
- 3 dB corresponds to a power ratio of 2. A power level that is changed by –3 dB is reduced to half the original power. A power level that is changed by +3 dB is doubled.
- 6 dB corresponds to a voltage ratio of 2. A voltage that is changed by –6 dB is reduced to half the original voltage. A voltage that is changed by +6 dB is doubled.
- 10 dB corresponds to a power ratio of 10. This is the only point where the dB value and the ratio value are the same (for power).
- 20 dB corresponds to a voltage ratio of 10. A voltage that is changed by +20 dB becomes 10 times larger. A voltage that is changed by –20 dB becomes 10 times smaller.

1.19 Absolute Decibel Values

Besides being useful for expressing ratios of powers or voltages, decibels can be used for specifying absolute voltages or powers. Either a power reference or voltage reference must be specified. For power calculations:

$$dB(absolute) = 10 \log\left(\frac{P}{P_{REF}}\right)$$

and for voltage calculations:

$$dB(absolute) = 20 \log\left(\frac{V}{V_{REF}}\right)$$

1.19.1 dBm

A convenient and often-used power reference for instrumentation use is the milliwatt (1 mW or 0.001 watt), which results in dBm:

$$dBm = 10 \log\left(\frac{P}{0.001}\right)$$

This expression is valid for any impedance or resistance value. If the impedance is known and specified, dBm can be computed using voltage. For a 50 Ω resistance, 1 mW of power corresponds to a voltage of 0.224 volts RMS:

$$P = \frac{V_{RMS}^2}{R}$$

$$V_{\text{RMS}} = \sqrt{P \cdot R} = \sqrt{0.001 \cdot 50} = 0.224 \text{ volts RMS}$$

Using this voltage as the reference value in the decibel equation results in:

$$\text{dBm } (50\ \Omega) = 20 \log (V_{RMS} / 0.224)$$

Similarly, for 600 ohms and 75 ohms:

$$\text{dBm } (600\ \Omega) = 20 \log (V_{RMS} / 0.775)$$

$$\text{dBm } (75\ \Omega) = 20 \log (V_{RMS} / 0.274)$$

These equations are valid only for the specified impedance or resistance.

1.19.2 dBV

Another natural reference to use in measurements is 1 volt (RMS). This results in the following:

$$\text{dBV} = 20 \log (\text{V}/1)$$

or simply

$$\text{dBV} = 20 \log (\text{V})$$

This equation is valid for any impedance level, since its reference is a voltage. Table 1.4 summarizes these equations and Figures 1.18 and 1.19 are plots of the basic decibel relationships. Table 1.5 shows voltage and power ratios along with their decibel values. The reader is encouraged to spend some time studying these figures to get a feel for how decibels work.

Table 1.4 Summary of equations relating to decibel calculations. Voltages (V) are RMS voltage and power (P) is in watts.

$dB = 10 \log (P_2/P_1)$

$dB = 20 \log (V_2/V_1)$

$dBm = 10 \log (P/0.001)$
$dBV = 20 \log (V)$

$dBm\ (50\ \Omega) = 20 \log (V/0.224)$
$dBm\ (75\ \Omega) = 20 \log (V/0.274)$
$dBm\ (600\ \Omega) = 20 \log (V/0.775)$
$dBW = 10 \log (P)$
$dBf = 10 \log (P/1 \times 10^{-15})$
$dBuV = 20 \log (V/1 \times 10^{-6})$
$dBc = 10 \log (P/P_{carrier})$
$dBc = 20 \log (V/V_{carrier})$

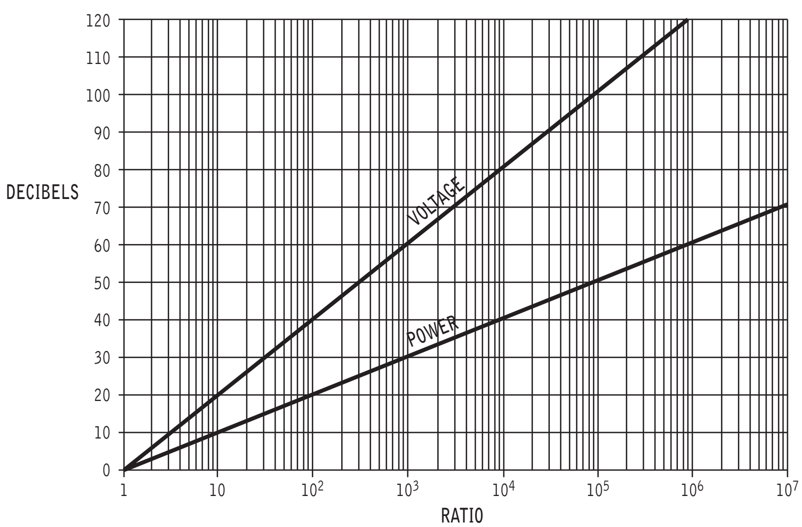


Figure 1.18 Graph for converting ratios to and from decibels for both voltage and power. Because of the logarithmic relationships, the plots are straight lines on log axes.

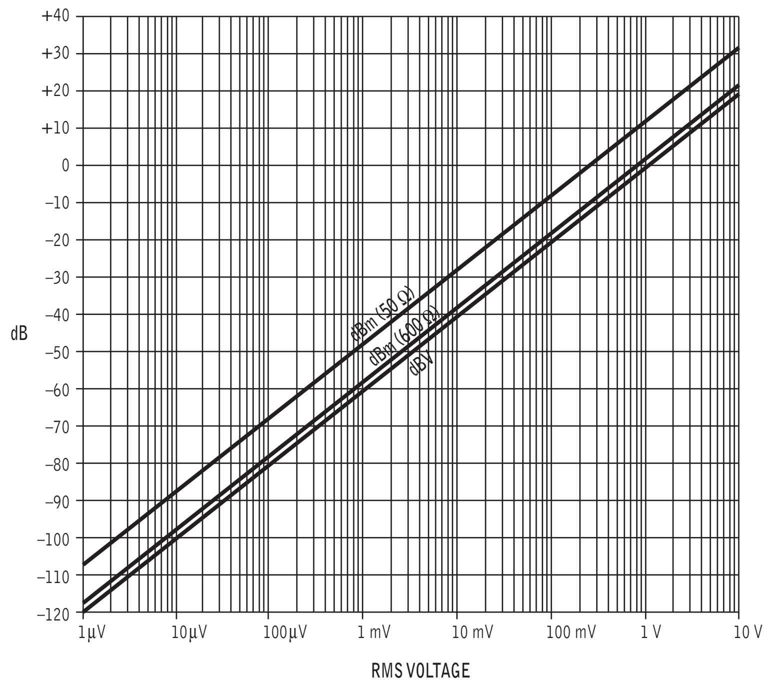


Figure 1.19 Graph of decibels versus voltage for dBV, dBm (50 Ω), and dBm (600 Ω).

Table 1.5 Table of decibel values for voltage ratios and power ratios.

Decibels	Power Ratio	Voltage Ratio
100	10000000000	100000
90	1000000000	31623
80	100000000	10000
70	10000000	3162
60	1000000	1000
50	100000	316.2
40	10000	100
30	1000	31.62
20	100	10
10	10	3.162
0	1	1.000
–10	0.1	0.3162

Table 1.5 Table of decibel values for voltage ratios and power ratios. (Continued)

Decibels	Power Ratio	Voltage Ratio
–20	0.01	0.1000
–30	0.001	0.03162
–40	0.0001	0.01000
–50	0.00001	0.003162
–60	0.000001	0.001000
–70	0.0000001	0.0003162
–80	0.00000001	0.0001000
–90	0.000000001	0.00003162
–100	0.0000000001	0.00001000
10	10.0000	3.1623
9	7.9433	2.8184
8	6.3096	2.5119
7	5.0119	2.2387
6	3.9811	1.9953
5	3.1623	1.7783
4	2.5119	1.5849
3	1.9953	1.4125
2	1.5849	1.2589
1	1.2589	1.1220
0.9	1.2303	1.1092
0.8	1.2023	1.0965
0.7	1.1749	1.0839
0.6	1.1482	1.0715
0.5	1.1220	1.0593
0.4	1.0965	1.0471
0.3	1.0715	1.0351
0.2	1.0471	1.0233
0.1	1.0233	1.0116

Table 1.5 Table of decibel values for voltage ratios and power ratios. (Continued)

Decibels	Power Ratio	Voltage Ratio
0	1.0000	1.0000
–0.1	0.9772	0.9886
–0.2	0.9550	0.9772
–0.3	0.9333	0.9661
–0.4	0.9120	0.9550
–0.5	0.8913	0.9441
–0.6	0.8710	0.9333
–0.7	0.8511	0.9226
–0.8	0.8318	0.9120
–0.9	0.8128	0.9016
–1	0.7943	0.8913
–2	0.6310	0.7943
–3	0.5012	0.7079
–4	0.3981	0.6310
–5	0.3162	0.5623
–6	0.2512	0.5012
–7	0.1995	0.4467
–8	0.1585	0.3981
–9	0.1259	0.3548
–10	0.1000	0.3162

Example 1.7

A 0.5-volt RMS voltage is across a 50 Ω resistor. Express this value in dBV and dBm. What would the values be if the resistor were 75 Ω?

50 Ω case:

dBV = 20 log (0.5) = –6.02 dBV

dBm (50 Ω) = 20 log (0.5/0.224) = +6.97 dBm

75 Ω case:

dBV is based on voltage so that value is the same as the 50 Ω case (–6.02 dBV).

Example 1.7 (Continued)

dBm is referenced to 1 mW of power. We must either use the voltage equation for dBm (75 Ω) or compute the power and use the power equation. We will compute the power.

$$P = \frac{V_{RMS}^2}{R} = \frac{(0.5)^2}{75} = 3.33\text{mW}$$

dBm = 10 log (P/0.001) = 10 log (0.00333/0.001)
dBm = 5.23 dBm

1.19.3 Other References

Other power reference values are 1 watt (dBW) and 1 femtowatt or 1×10^{-15} watt (dBf). A common voltage reference is the microvolt, resulting in dBμV. These values can be computed by the following equations:

$$\text{dBW} = 10 \log (P)$$

$$\text{dBf} = 10 \log (P/1 \times 10^{-15})$$

$$\text{dB}\mu\text{V} = 20 \log (V/1 \times 10^{-6})$$

Voltage or power measurements relative to a fixed signal or carrier are expressed as dBc.

$$\text{dBc} = 10 \log (P/P_{carrier})$$

or

$$\text{dBc} = 20 \log (V/V_{carrier})$$

The reader will undoubtedly encounter other variations. The number of possible permutations on the versatile decibel is limited only by the imagination of engineers worldwide.[7]

1.20 Measurement Error

A measuring instrument will always disturb the circuit being measured by removing energy from that circuit, no matter how small, and some error will always be introduced. Another way of saying this is that connecting an instrument to a circuit changes that circuit and changes the voltage or current that is being measured. This error can be minimized by careful attention to the loading effect, as discussed later.

7. It is rumored that some electrical engineers balance their checkbooks using dB$.

1.20.1 Internal Error

Errors internal to the instrument also degrade the quality of the measurement. These errors fall into two main categories:

Accuracy: the ability of the instrument to measure the true value to within some stated error specification.

Resolution: the smallest change in value that an instrument can detect.

Suppose a voltage measuring instrument has an accuracy of ±1% of the measured voltage and a 3-digit resolution. If the measured voltage was 5 volts, then the accuracy of the instrument is 1% of 5 volts or ±0.05 volts. The instrument could read anywhere from 4.95 to 5.05 volts with a resolution of 3 digits. The meter cannot, for instance, read 5.001 volts, since that would require 4 digits of resolution.

If the instrument had 4-digit resolution (but 1% accuracy), then the reading could be anywhere from 4.950 to 5.050 volts (the same basic accuracy, but with more digits). Assume for the moment that both meters read exactly 5 volts. If the actual voltage changed to 5.001 volts, the 3-digit instrument would probably not register any change, but the 4-digit instrument would have enough resolution to show that the voltage had indeed changed. (Actually, the 3-digit meter reading might change, but could not display 5.001 volts because the next-highest possible reading is 5.01 volts.) The accuracy of the 4-digit instrument is not any better, but it has finer resolution and can detect smaller changes. Neither meter guarantees that a voltage of 5.001 volts can be measured any more accurately than 1%.

The example given is a digital instrument, but the same concepts apply to analog instruments. Resolution is not usually specified in digits in the analog case, but all instruments have some fundamental limitation to their measurement resolution. Typically, the limitation is the physical size of the analog meter and its markings. For instance, an analog voltmeter with full-scale equal to 10 volts will have difficulty detecting the difference between two voltages, which differ by a few millivolts.

Usually, an instrument provides more measurement resolution than measurement accuracy. This guarantees that the resolution will not limit the obtainable accuracy and allows detection of small changes in readings, even if those readings have some absolute error in them. The relative accuracy between small steps may be much better than the full-scale absolute accuracy.

1.21 The Loading Effect

In general, when two circuits (a source and a load) are connected together, the voltages and currents in those circuits both change. The source might be, for example, the output of an amplifier, transmitter, or signal generator. The corresponding load might be a speaker, antenna, or the input of a circuit. In the case where an electronic measurement is being made, the circuit under test is the source and the measuring instrument is the load.

Many source circuits can be represented by a simple circuit model called a *Thevenin equivalent circuit*. A Thevenin equivalent circuit is made up of a voltage source, V_S, and a series resistance, R_S (Figure 1.20). Thus, complex circuits can be simplified by replacing them with an equivalent circuit model. In the same way, many loads can be replaced conceptually with a circuit model consisting of a single resistance, R_L. (Resistive circuits will be considered initially in this discussion, but later expanded to include AC impedances.)

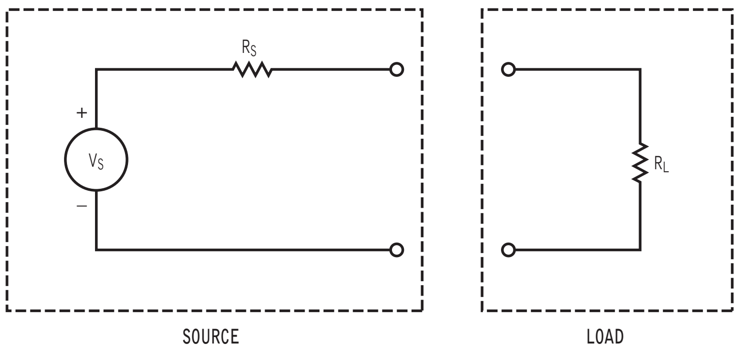


Figure 1.20 A voltage source with internal resistance and a resistive load (Thevenin equivalent circuit).

V_S is also known as the *open-circuit voltage*, since it is the voltage across the source circuit when no load is connected to it. This is easily proven by noting that no current can flow through R_S under open-circuit conditions. There is no voltage drop across R_S so the source voltage equals V_S.

1.22 The Voltage Divider

When the load is connected to the source, V_L, the voltage across the source and the load is no longer the open-circuit value, V_S. V_L is given by the voltage divider equation (named for the manner in which the total voltage in the circuit divides across R_S and R_L).

$$V_L = \frac{V_S R_L}{(R_S + R_L)}$$

Figure 1.21 shows a source and load connected in this manner. The resulting output voltage, V_L, as a function of the ratio R_L/R_S is plotted in Figure 1.22. If R_L is very small compared to R_S, then V_L is also very small. For large values of R_L (compared to R_S), V_L approaches V_S.

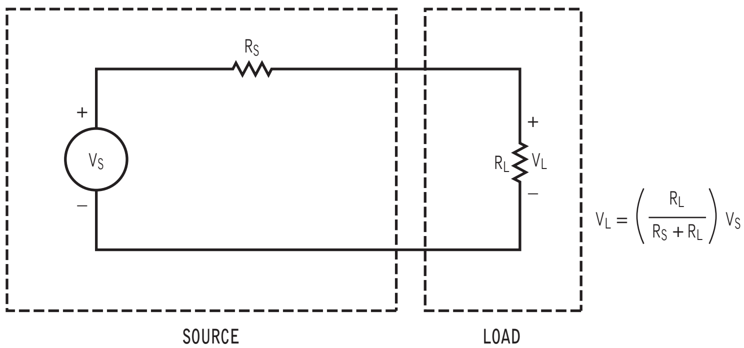


Figure 1.21 The source and load are connected together. The voltage across the load is given by the voltage divider equation. The loading effect causes this voltage to be less than the open circuit voltage of the source.

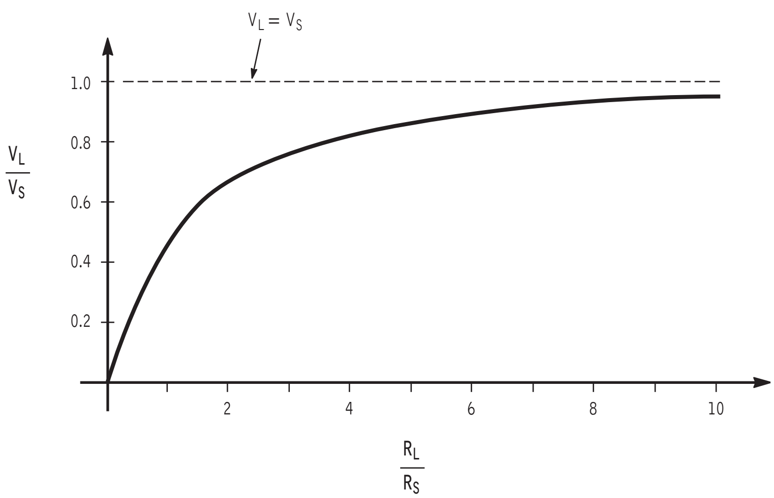


Figure 1.22 A plot of the output voltage due to the loading effect, as a function of the load resistance divided by the source resistance. The larger the load resistance, the better the voltage transfer.

1.23 Maximum Voltage Transfer

To get the maximum voltage out of a voltage source being loaded by some resistance, the ratio R_L/R_S should be as large as possible. From a design point of view, this can be approached from two directions: make R_S small or make R_L large. Ideally, we could make $R_S = 0$ and make R_L = infinity, resulting in $V_L = V_S$. In practice, this cannot be obtained but can be approximated. Figure 1.22 shows that making R_L 10 times larger than R_S results in a voltage that is 91% of the maximum attainable (V_S).

When making measurements, the source is the circuit under test and the load is the measuring instrument. We may not have control over the value of R_S (if it is part of the circuit under test) and our only recourse is choosing R_L (the resistance in our measuring instrument). Ideally, we would want the resistance of the instrument to be infinite, causing no loading effect. In reality, we will settle for an instrument that has an equivalent load resistance that is much larger than the equivalent resistance of the circuit.

1.24 Maximum Power Transfer

Sometimes the power delivered to the load resistor is more important than the voltage. Since power depends on both voltage and current ($P = V \times I$), maximum voltage does not guarantee maximum power. The power delivered to R_L can be determined as follows:

$$P = \frac{V_S R_L}{(R_L + R_S)} \frac{V_S}{(R_L + R_S)}$$

$$P = \frac{V_S^2 R_L}{(R_L + R_S)^2}$$

This relationship is plotted for varying values of R_L/R_S in Figure 1.23. For small values of R_L (relative to R_S), the power delivered to R_L is small, because the voltage across R_L is small. For large values of R_L (relative to R_S), the power delivered is also small, because the current through R_L is small. The power transferred is maximum when $R_L/R_S = 1$, or equivalently, when R_L equals R_S.

$$R_L = R_S \text{ for maximum power}$$

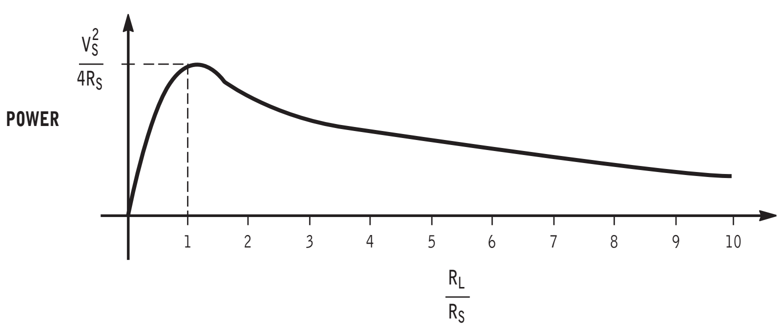


Figure 1.23 Plot of output power as a function of the load resistance divided by the source resistance. The output power is maximum when the two resistances are equal.

In many electronic systems, maximum power transfer is desirable. Such systems are designed with all source resistances and load resistances being equal to maximize the power transfer. Maximum power transfer is achieved while sacrificing maximum voltage transfer between the source and load.

1.25 Impedance

The previous discussion has assumed resistive circuits, but many circuit components are *reactive*, exhibiting a phase shift between their voltage and current. This type of voltage and current relationship is commonly represented through the use of *complex impedance*. The impedance of a device is defined as

$$Z = \frac{V_{0-P}\angle\theta_V}{I_{0-P}\angle\theta_I}$$

where

V_{0-P} is the AC voltage across the impedance
θ_V is the phase angle of the voltage
I_{0-P} is the AC current through the impedance
θ_I is the phase angle of the current.

Simplifying the equation,

$$Z = \frac{V_{0-P}}{I_{0-P}}\angle\theta_Z$$

where θ_Z is the phase angle of the impedance and is equal to $\theta_V - \theta_I$

The preceding equations show the complex impedance in magnitude and phase format. Alternatively, the impedance can be expressed in a rectangular format.[8]

$$Z = R + j\,X$$

where

R = the resistive component of the impedance
X = the reactive component of the impedance
j = the square root of –1

For the purposes of maximum power transfer, when the source impedance and load impedance are not resistive, maximum power transfer occurs when the load impedance has the same magnitude as the source impedance, but with opposite phase angle.[9] For instance, if the source impedance was 50 Ω with an angle of +45 degrees, the load impedance should

8. See Appendix B for information on rectangular and polar format numbers.
9. For a derivation of this, see Irwin, 2001.

be 50 Ω with an angle of –45 degrees to maximize the power transfer. Mathematically, this can be stated as

$$Z_L = Z_S^*$$

where * indicates the complex conjugate.

A special case occurs when the phase angle of the source impedance is zero. In that case, the load impedance should have the same magnitude as the source impedance and an angle of zero (load impedance equals source impedance). Note that this is the same as the purely resistive case.

1.26 Instrument Inputs

The characteristics of instrument input impedances vary quite a bit with each individual model, but in general they can be put into two categories: *high impedance* and *system impedance.*

1.26.1 High-Impedance Inputs

High impedance inputs are designed to maximize the voltage transfer from the circuit under test to the measuring instrument by minimizing the loading effect. As outlined previously, this can be done by making the input impedance of the instrument much larger than the impedance of the circuit. Typical values for instrument input impedance are between 10 kΩ and 1 MΩ. For instruments used at high frequencies, the capacitance across the input becomes important and therefore is usually specified by the manufacturer.

1.26.2 System-Impedance Inputs

Many electronic systems have a particular system impedance such as 50 Ω (Figure 1.24). If all inputs, outputs, cables, and loads in the system have the same resistive impedance, then, according to the previous discussion, maximum power is always transferred. At high frequencies (greater than about 30 MHz), stray capacitance and transmission line effects make this the only type of system that is practical. The system impedance is often called the *characteristic impedance* and is represented by the symbol Z_0.

At audio frequencies, a constant system impedance is not mandatory but is often used. It is sufficient for many applications to make all source circuits have a low impedance (less than 100 Ω) and all load circuits have a high impedance (greater than 1 kΩ). This results in near-maximum voltage transfer (power transfer is less important here). Some audio systems do maintain, or at least specify for test purposes, a system impedance, which is usually 600 Ω. This same impedance appears in telephone applications as well.

For radio and microwave frequency work, 50 Ω is by far the most common impedance. This impedance can be easily maintained despite stray capacitance and 50 Ω cable is easily realizable. Such things as amateur and commercial radio transmitters, transmitting

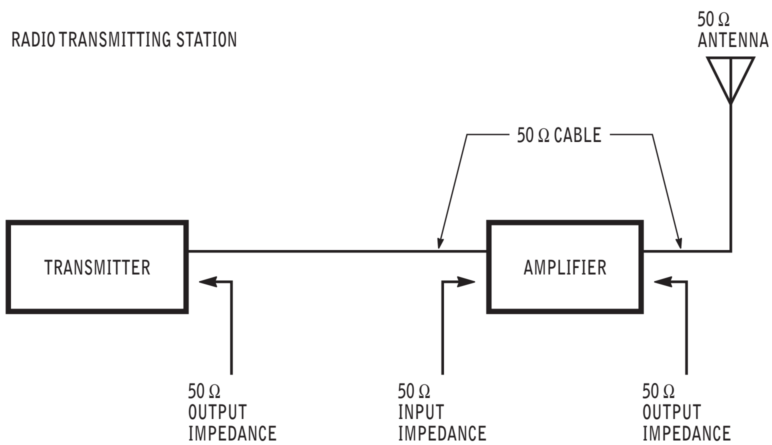


Figure 1.24 An electronic system that uses a common system impedance for maximum power transfer throughout.

antennas, communications filters, and radio frequency test equipment all commonly have 50 Ω input and output impedances. Runner up to 50 Ω in the radio frequency world is 75 Ω. This impedance is also used extensively at these frequencies, particularly in video-related applications such as cable TV. Other system impedances are used as special needs arise and will be encountered when making electronic measurements.

When measuring this type of system, most of the accessible points in the system expect to be loaded by the system impedance (Z_0). Because of this, many instruments are supplied with standard input impedance values. Thus, the instrument can be connected into the system and act as a Z_0 load during the measurement.

1.26.3 Connectors

Instruments use input and output connectors that are consistent with the signal or waveform being generated or measured. Usually, it is the frequency range of the instrument that determines the type of connector required.

For DC and low-frequency AC instruments, the *banana plug* and *banana jack* are often used, usually configured in pairs as a *dual-banana plug* or *jack*. The banana jack may be the "binding post" style that allows a bare wire to be connected to the jack without the need for a plug. For applications that could have high voltage present on the banana jack, such as with the input to a multimeter, the banana jack and plug are insulated to prevent hazardous voltages being exposed to the user.

The *BNC connector* is the most common connector used in electronic test equipment.[10] The BNC connector is compatible with coaxial cables and is easy to connect and disconnect due to its quarter-turn locking mechanism. The connector is available in 50 Ω and 75 Ω versions, consistent with the most common system impedances. These connectors can be used from DC to several GHz, depending on the specific application.

The *Type N connector* is considerably larger than the BNC connector and has a threaded locking mechanism.[11] The N connector is compatible with coaxial cables, is available in 50 Ω and 75 Ω versions, and is used for frequencies from DC to over 10 GHz. The threaded locking mechanism provides for a more repeatable connection than the BNC connector such that the N connector is sometimes used for precision measurements regardless of frequency.

1.27 Bandwidth

Instruments that measure AC waveforms generally have some maximum frequency above which the measurement accuracy is degraded. This frequency is the *bandwidth* of the instrument and is usually defined as the frequency at which the instrument's response has decreased by 3 dB. (Other values such as 1 dB and 6 dB are also sometimes used.) A typical response is shown in Figure 1.25. Note that the response does not instantly stop at the 3-dB bandwidth. It begins decreasing at frequencies less than the bandwidth and the response may still be usable for frequencies outside the bandwidth. Some instruments have their bandwidth implied in their accuracy specification. Rather than give a specific 3-dB bandwidth, the accuracy specification will be given as valid only over a specified frequency range.

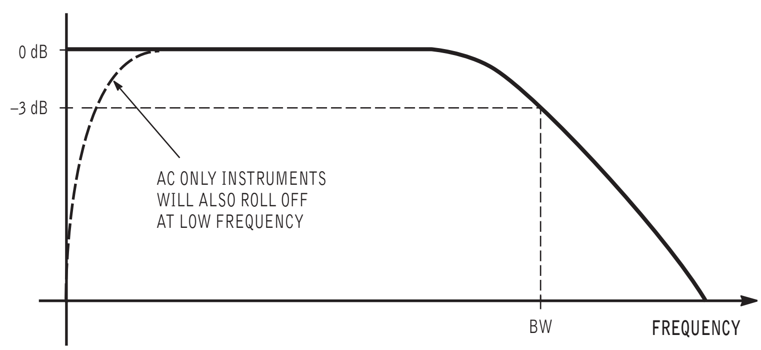


Figure 1.25 The frequency response of a typical measuring instrument rolls off at high frequencies. Some instruments that do not measure DC also roll off at low frequencies.

10. BNC stands for Bayonet Neill Concelman, named for the inventors of the connector, Paul Neill and Carl Concelman.
11. Type N stands for Neill, named for its inventor, Paul Neill.

In general, to measure an AC waveform accurately, the instrument must have a bandwidth that exceeds the frequency content of the waveform. For a sine wave, the instrument bandwidth must be at least as large as the sine wave frequency. For sine wave frequencies equal to the 3-dB bandwidth, the measured value will, of course, be decreased by 3 dB, so an even wider bandwidth is desirable. The amount of margin necessary will vary depending on how quickly the response rolls off near the 3-dB point.

For sine wave measurements, the frequency response near the 3-dB bandwidth can be adjusted for by recording the response at these frequencies using a known input signal. Assuming the response is repeatable, the measurement error due to limited bandwidth can be used to adjust the measured value to obtain the actual value.

Instruments that measure only AC voltages (and not DC) have a response that rolls off at low frequencies as well as high frequencies. When using instruments of this type (including many AC voltmeters), both the high-frequency and the low-frequency bandwidth limitations must be considered.

For waveforms other than sine waves, the harmonics must be considered. If the harmonics are outside the instrument bandwidth, their effect on the waveform will not be measured. This could be desirable if the harmonics outside some frequency range needed to be ignored. In general, their effect is usually desirable in the measurement. Waveforms with an infinite number of harmonics would require an instrument with infinite bandwidth to measure them. In reality, the higher harmonics have very little energy and can be ignored with proper regard to how much error this produces in the measurement.

Example 1.8

An electronic instrument is used to measure the voltage of a 2-kHz sine wave. What instrument bandwidth is required? What bandwidth would be required to measure a square wave having the same frequency (assume that any harmonic greater than 10% of the fundamental is to be included)?

Since a sine wave has only the fundamental frequency, the bandwidth of the instrument must be at least the frequency of the waveform. BW = 2 kHz. (In reality, we may want to choose a somewhat higher value since the instrument response is diminished by 3 dB at its bandwidth.)

From Table 1.3, the highest significant harmonic of a square wave (greater than 10% of the fundamental) is the ninth harmonic. Therefore, the bandwidth must be at least (and probably larger than):

$$BW = 9 \times 2 \text{ kHz} = 18 \text{ kHz}$$

1.28 Rise Time

Ideally, waveforms such as square waves and pulses change voltage level instantaneously. In reality, waveforms take some time to make an abrupt change, depending on the bandwidth of the system and other circuit parameters. The amount of time it takes for a waveform to transition from one voltage to another is called the *rise time*. (The rise time is normally measured at the 10% and 90% levels of the transition.) The bandwidth of a measuring instrument will limit the measured rise time of a pulse or square wave. For a typical instrument, the relationship between rise time and bandwidth is given by:

$$t_{RISE} = \frac{0.35}{BW}$$

BW = 3-dB bandwidth (in Hertz)

The validity of this relationship depends on the exact shape of the frequency response of the instrument (how fast it rolls off for frequencies above its bandwidth). It is exact for instruments with a single-pole roll-off[12] and is a good approximation for many instruments. The important point here is that the bandwidth, which is a frequency domain concept, limits the measurement of the rise time, which is a time domain concept. The two characteristics of the instrument (time domain and frequency domain) are intertwined. Fast changes (small rise times) correspond to high-frequency content in a waveform, so if the high-frequency content is limited by the bandwidth of a system, then the rise time will be larger.

The instrument should have a rise time significantly smaller than the rise time being measured. A rise time measurement using an instrument with a rise time two times smaller than the one being measured will result in an error of about 10%. This error drops to 1% when the instrument rise time is 7 times shorter than the measured rise time.

1.29 Bandwidth Limitation on Square Wave

As an example of how limited bandwidth can affect a waveform in the time domain, consider the square wave shown in Figure 1.26. The waveform is passed through a low-pass filter, which has a frequency characteristic similar to Figure 1.25. If the bandwidth of the filter is very wide (compared to the fundamental frequency of the waveform), the square wave appears at the output undistorted. If the bandwidth is reduced, some of the harmonics are removed from the waveform. The output still looks like a square wave, but it has some imperfections that would cause a measurement error. For very limited bandwidth, the square wave barely appears at all, and is very rounded due to the lack of high-frequency harmonics.

12. See Appendix B for a discussion of this type of frequency response and the derivation of the relationship between bandwidth and rise time.

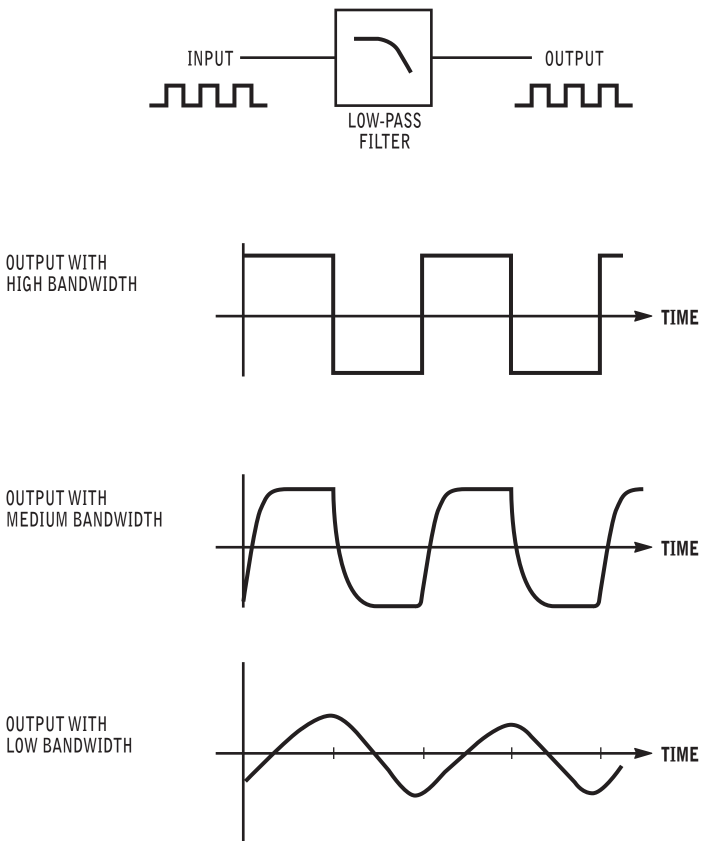


Figure 1.26 The effect of bandwidth on a square wave. With a very wide bandwidth, the square wave is undistorted; with a low bandwidth, the square wave is distorted.

Example 1.9

A pulse with zero rise time is passed through a low-pass Filter with a 3-dB bandwidth of 25 kHz. If the filter has a single-pole roll-off, what will the rise time be at the output?

Even though the rise time at the input is zero, due to the 25 kHz bandwidth limitation the rise time at the output will be

$$t_{RISE} = 0.35/\ BW = 0.35/\ 25\ kHz = 14\ \mu sec$$

Example 1.10

A square wave with a rise time of 1 usec is to be measured by an instrument. What 3-dB bandwidth is required in the measuring instrument to measure the rise time with 1% error?

For a 1% error in rise time, the measuring instrument must have a rise time that is 7 times less than the waveform's rise time.

For the instrument:

$$t_{RISE} = 1\ \mu\text{sec}/\ 7 = 0.14\ \mu\text{sec}$$

$$BW = 0.35/\ t_{RISE} = 0.35\ /\ 0.14\ \mu\text{sec} = 2.5\ \text{MHz}$$

1.30 Digital Signals

With the widespread adoption of digital logic circuitry, digital signals have become very common in electronic systems. The basic advantage and simplicity of a digital signal is that it has only two valid states: HIGH and LOW (Figure 1.27). The HIGH state is defined as any voltage greater than or equal to V_H and the LOW state is defined as any voltage less than or equal to V_L. Any voltage between V_H and V_L is undefined. V_H and V_L are called the *logic thresholds*.

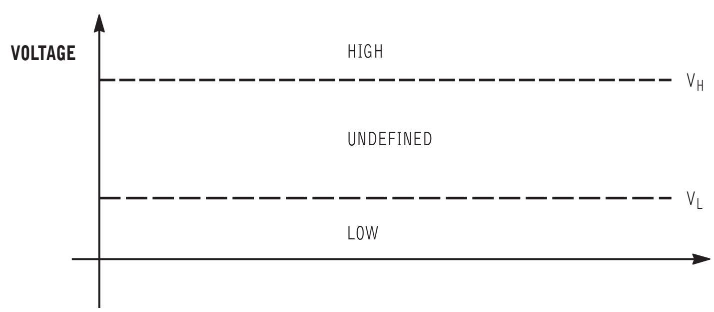


Figure 1.27 A digital signal must be one of two valid states, HIGH or LOW.

A typical, well-behaved digital signal (a pulse train) is shown in Figure 1.28a. Whenever the voltage is V_{0-P}, the digital signal is interpreted as HIGH. Whenever the voltage is zero, the digital signal is LOW. The digital signal in Figure 1.28b is not so well behaved. The signal starts out LOW, then enters the undefined region (greater than V_L but smaller than V_H), then becomes HIGH, then LOW and HIGH once more. Notice that in the last LOW state, the voltage is not zero but increases a small amount. It does stay less than V_L, so it is still a LOW signal. Similarly, during the last HIGH state the voltage steps down a small amount, but stays above the high threshold.

Although the region between the two logic thresholds is undefined, it does not follow that it should never occur. First of all, to get from LOW to HIGH, the waveform must pass through the undefined area. Also, there are times when one digital gate goes into an "off" or high impedance state while another gate drives the voltage HIGH or LOW. During this transition, the voltage could stay in the undefined area for a period of time. It is imperative that the digital signal settle to a valid logic level before the digital circuit following it uses the information, otherwise a logic error may occur.

Digital signals are mostly used to represent the binary numbers 0 and 1. If *positive logic* is being used, then a HIGH state corresponds to a logical 1 and a LOW state corresponds to a logical 0. Although positive logic may appear to be the most obvious convention, *negative logic* is also sometimes used. With negative logic, the HIGH state represents a logical 0 and the LOW state represents a logical 1. Thus, the relationship between the voltage of the digital signal and the binary number that it represents varies depending on the logic convention used. Fortunately, the concepts associated with digital signals remain the same but the instrument user may be left with a digital bookkeeping problem.

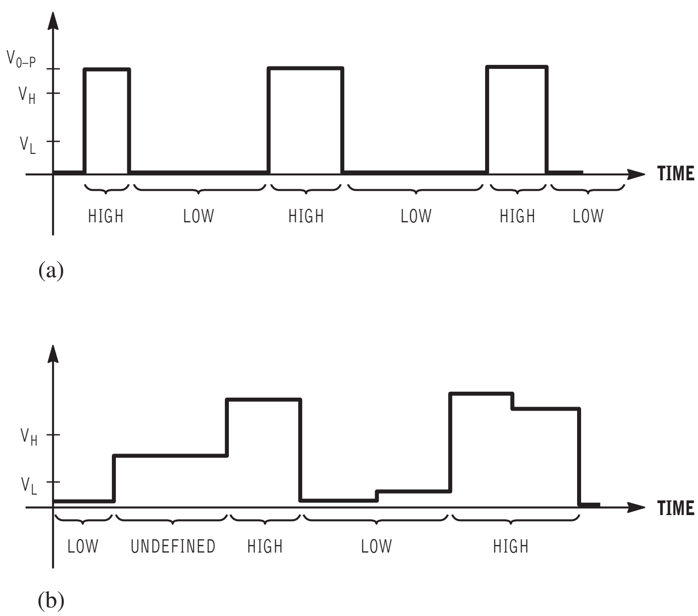


Figure 1.28 Two digital signals. a) A well-behaved digital signal that is always a valid logic level. b) A digital signal that is undefined at one point.

1.31 Logic Families

There are several different technologies used to implement digital logic circuits, with different speed and logic threshold characteristics. For any particular logic family, we need to know the correct values for the logic thresholds (V_H, V_L) so that the digital signal can be measured or interpreted correctly. Table 1.6 lists logic thresholds for the most popular logic families.

Table 1.6 Logic Levels for Standard Digital Logic Families

Logic Family	Supply Voltage V_{CC}	Input Threshold (Low), V_{IL}	Input Threshold (High), V_{IH}	Output Threshold (Low), V_{OL}	Output Threshold (High), V_{OH}
CMOS	5 V	1.5 V	3.5 V	0.5 V	4.4 V
TTL	5 V	0.8 V	2.0 V	0.4 V	2.4 V
LVTTL	3.3 V	0.8 V	2.0 V	0.4 V	2.4 V
LVCMOS	3.3 V	0.8 V	2.0 V	0.2 V	3.1 V
LVCMOS	2.5 V	0.7 V	1.7 V	0.4 V	2.0 V

Derived from "Benefits and Issues on Migration of 5-V and 3.3-V Logic to Lower-Voltage Supplies," Texas Instruments, September 1999.

The logic thresholds are different depending on whether you are concerned with the input of a gate or the output of a gate. The V_L for the output of a gate is lower than the V_L for the input to a gate. Similarly, the V_H for the output is higher than the V_H for the input. This slight amount of intentional design margin guarantees that the output of a digital circuit will drive the input of the next circuit well past its required logic threshold. The difference between the input threshold and the output threshold is called the *noise margin*, since the excess voltage is designed to overcome any electrical noise in the system. In general digital testing, the measurement instrumentation should usually be considered a digital input and the logic levels associated with the input should be used.

Standards and prevailing practice concerning logic signal levels are changing over time. The drive for more logic gates on an integrated circuit and higher-speed operation of those gates drives the technology toward smaller device size. As the number of devices on an integrated circuit increases, power consumption tends to also increase. To counteract both of these trends, integrated circuit manufacturers are specifying reduced power supply voltages on digital circuits. Lower supply voltages mean lower voltage stress on the devices, allowing them to be made smaller. A lower supply voltage also significantly decreases the power consumed by a given circuit.

The dominant logic levels in modern digital circuits were driven by the popularity of TTL (transistor–transistor logic) technology, implemented with a 5-volt power supply. In recent years, CMOS (Complementary Metal Oxide Semiconductor) has become the dominant digital technology. However, TTL logic levels are still commonly used. Logic gates may be described as "TTL-compatible," while actually implemented using CMOS technology. Historically, the most common power supply voltage for logic circuits is 5 volts, which was used on a series of popular TTL logic families. When CMOS technology gained in usage, the 5-volt power supply standard was adopted. More recent logic families use lower supply voltages, including 3.3-volt and 2.5-volt supplies. Backward compatibility with the TTL input high and low of 2.0 volts and 0.8 volts is possible with the 3.3-volt supply but not with the 2.5-volt supply.

These industry trends are creating confusion and change in digital logic levels. While there is a desire to maintain "TTL compatibility," the drive to lower supply voltages and higher speed is forcing changes in logic levels. Table 1.6 can be used as a guideline but users of test equipment may need to consult the data sheet for a particular part.

1.32 References

"2.5 V± 0.2 V (Normal Range), and 1.8 V to 2.7 V (Wide Range) Power Supply Voltage and Interface Standard for Nonterminated Digital Integrated Circuit," EIA/JEDEC Standard JESD8-5, Electronics Industry Alliance, October 1995.

"Interface standard for Nominal 3 V/3.3 V Supply Digital Integrated Circuits," EIA/JEDEC Standard JESD8-B, Electronics Industry Alliance, September 1999.

Irwin, J. David. *Basic Engineering Circuit Analysis*, 7th ed., New York: Wiley, 2001.

Oliver, Bernard M., and John M. Cage. *Electronic Measurements and Instrumentation*. New York: McGraw-Hill, 1971.

Schwartz, Mischa. *Information, Transmission, Modulation, and Noise*, 4th ed., New York: McGraw-Hill, 1990.

Schwarz, Steven E., and William G. Oldham. *Electrical Engineering: An Introduction*, 2nd ed., New York: Oxford University Press, 1993.

Wedlock, Bruce D., and James K. Roberge. *Electronic Components and Measurements*, Englewood Cliffs, NJ: Prentice Hall, 1969.

Witte, Robert A. *Spectrum and Network Measurements*, Norcoss, GA: Noble, 2001.