Contents

Chapter 3 — Online Content

Articles

- Digital Electronic Basics by Dale Botkin, NØXAS
- Hands On Radio: Maxwell's Equations — Grad, Div, and Curl by Ward Silver, NØAX
- Hands On Radio: Maxwell's Equations — The Wave Emerges by Ward Silver, NØAX
- Hands On Radio: The Effects of Gain-Bandwidth Product by Ward Silver, NØAX
- Maxwell Without Tears by Paul Shuch, N6TX
- Untangling the Decibel Dilemma — Expanded by Ward Silver, NØAX

*Also see these articles with the **Electrical Fundamentals** chapter Online Content:*

- Radio Mathematics
- Radio Math Formulas and Notes

Tools and Data

- Frequency Response Spreadsheet

Chapter 3
Radio Fundamentals

Radio begins with an understanding of alternating current waveforms and how they are measured. This chapter also examines the relationship between ac voltage and current in energy-storing components like capacitors and inductors that defines reactance and impedance. We can then explore quality factor (Q) and the properties of resonant circuits.

Analog system concepts are introduced to explain the concepts and techniques of working with electronic circuits in radio. Finally, we discuss electromagnetic waves that carry information between stations.

As for the previous chapter, additional mathematics resources are available in the article "Radio Mathematics" in this book's online content.

3.1 AC Waveforms

A *waveform* is the pattern of *amplitudes* reached by voltage or current measured over time, including combinations of ac and dc voltage and current. For example, **Figure 3.1** shows two ac waveforms fairly close in frequency and their combination. **Figure 3.2** shows two ac waveforms dissimilar in both frequency and wavelength, along with the resultant combined waveform.

3.1.1 Sine Waves and Rotation

Not only is a sine wave the most fundamental ac waveform — energy at a single frequency — it also describes rotation. The cyclical nature of the sine wave is at the heart of much of radio technology, whether analog or digital. A good grasp of the sine wave and the closely related cosine wave are key to understanding the techniques that make up radio. (Sine and cosine functions as well as vectors and phasors are discussed in the "Radio Mathematics" article in this book's online content.)

Figure 3.3 illustrates the relationship. Imagine a rotating wheel with a visible dot at a point anywhere along the circumference (rim) of the wheel. If you spin the wheel at a constant rate and watch the wheel on-edge as in Figure 3.3A, the dot will just move up and down. If

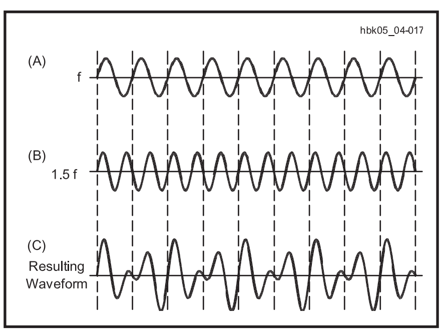

Figure 3.1 — Two ac waveforms of similar frequencies (f1 = 1.5 f2) with amplitudes added together to form a composite wave. Note the points where the positive peaks of the two waves combine to create high composite peaks at a frequency that is the difference between f1 and f2. The beat note frequency is 1.5f – f = 0.5f and is visible in the drawing.

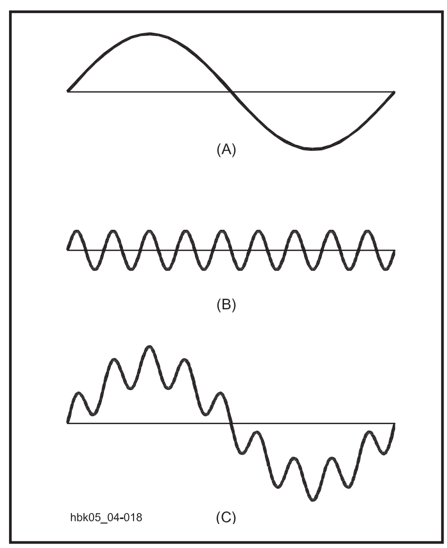

Figure 3.2 — Two ac waveforms of widely different frequencies and amplitudes form a composite wave in which one wave appears to ride upon the other.

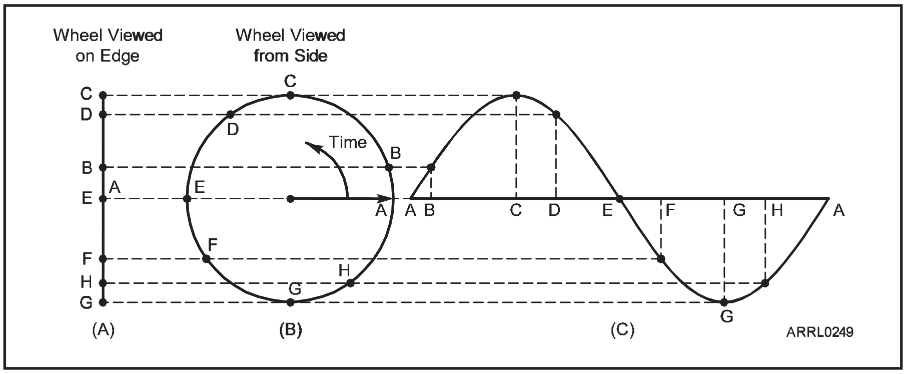

Figure 3.3 — This diagram illustrates the relationship between a sine wave and circular rotation. You can see how various points on the circle correspond to values on the sine wave.

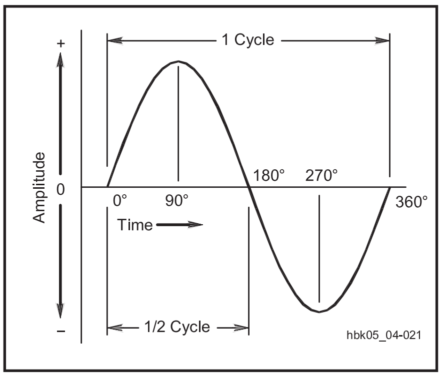

Figure 3.4 — An ac cycle is divided into 360 degrees that are used as a measure of time or phase.

we designate the height at point C as +1 and at point G as –1, then make a table of values as the wheel rotates through 360°, the height values will correspond exactly to the values of the sine function (*sin*). When the wheel is at 0° in position A, sin (0°) = 0; when at 90° in position C, sin (90°) = 1, and so forth all the way around.

Now look at the wheel from the side as in Figure 3.3B. Plotting the dot's height around its circular path against degrees on the horizontal axis will then trace out a sine wave as in Figure 3.3C. Each rotation of the wheel corresponds to one *cycle* of the sine wave. The amplitude (A) of the sine wave is equal to the sine of the wheel's *angular position* in degrees (θ):

$$A = \sin(\theta)$$

Figure 3.4 shows how each cycle of a sine wave is divided into 360º to measure angular position.

3.1.2 Frequency, Period, and Harmonics

With a continuously rotating generator, alternating current or voltage will pass through many equal cycles over time. This is a *periodic waveform*, composed of repeated identical cycles. An arbitrary point on any one cycle can be used as a marker of position on a periodic waveform. For this discussion, the positive peak of the waveform will work as an unambiguous marker. The number of times per second that the current (or voltage) reaches this positive peak in any one second is called the *frequency* of the waveform. In other words, frequency expresses the *rate* at which current (or voltage) cycles occur. The unit of frequency is *cycles per second*, or *hertz*—abbreviated Hz (after Heinrich Hertz, the 19th century physicist who demonstrated the existence of radio waves).

If the sine wave has a constant frequency, every complete cycle takes the same amount of time, the *period*, T, as in **Figure 3.5**. The signal's period is the reciprocal of its frequency:

$$\text{Frequency (f) in Hz} = \frac{1}{\text{Period (T) in seconds}}$$

and

$$\text{Period (T) in seconds} = \frac{1}{\text{Frequency (f) in Hz}}$$

Example: What is the period of 60 Hz ac current?

$$T = \frac{1}{f} = \frac{1}{60\ \text{Hz}} = 0.01666\ \text{s} = 16.6\ \text{ms}$$

To calculate the amplitude, A, of the sine wave at any point in time, t, we need to be able to convert time to the angle, θ:

$$\theta = 360\frac{t}{T}$$

The sine wave equation is now:

$$A = \sin\left(360\frac{t}{T}\right) = \sin\left(360 \times \frac{1}{T} \times t\right)$$
$$= \sin(360 \times f \times t)$$

HARMONICS

A *harmonic* is a signal with a frequency that is some integer multiple (2, 3, 4 and so on) of a signal at a *fundamental frequency*. Figure 3.5 shows a simple example of harmonics. The harmonic at twice the fundamental's frequency is called the *second harmonic*, at three times the fundamental frequency the *third harmonic*, and so forth. There is no "first harmonic." For example, if a complex waveform is made up of sine waves with frequencies of 10, 20, and 30 kHz, 10 kHz is the fundamental and the other two are harmonics. Signals at the frequency of harmonics are said to be *harmonically related* to the fundamental.

3.1.3 Phase

Now let's make the connection between angular position and time. Although time is measurable in parts of a second, it is more convenient to treat each cycle as a complete time unit divided into 360°. The conventional starting point for phase in a sinusoidal wave-

Degrees, Radians, and Angular Frequency

While most electronic and radio mathematics use degrees as a measure of phase, you will occasionally encounter *radians*. Radians are used because they are more convenient mathematically in certain types of equations and computations. There are 2π radians in a circle, just as there are 360°, so one radian = 360/2π ≈ 57.3°. In this book, unless it is specifically noted otherwise, the convention will be to use degrees in all calculations of phase or angle.

In engineering textbooks and other electronic references, you will often encounter the symbol ω used to represent *angular frequency*. The sine wave equation would then be written:

$$A = \sin(2\ \omega\ ft) = \sin(\omega\ t)$$

where $\omega = 2\pi f$. The use of angular frequency is more straightforward in many types of engineering calculations. Radians are used for angular position when angular frequency (ω) is being used.

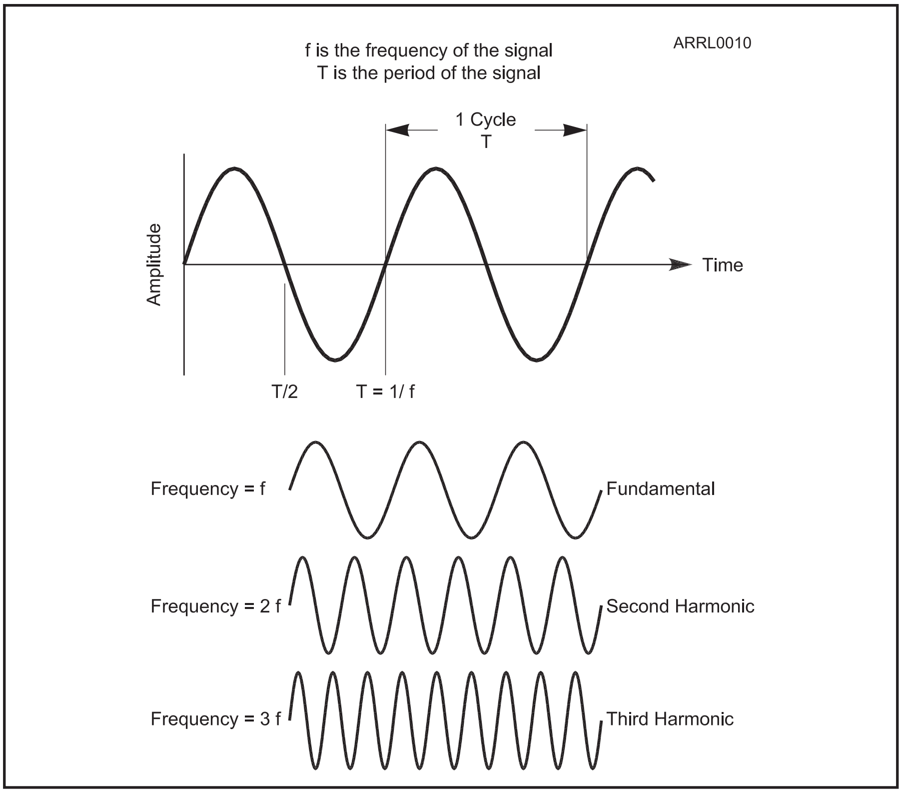

Figure 3.5 — The frequency of a signal and its period are reciprocals. A higher frequency means a shorter period and vice versa. Harmonics are signals with frequencies that are integer multiples of a fundamental frequency.

Phase versus Polarity

It is important to distinguish between polarity and phase. Polarity is the assigned conventions or directions for positive and negative voltage or current. Phase is a function of time or position in a waveform. It is quite possible for two signals to have opposite polarities, but still be in phase, for example. In a multi-phase ac power system, "phase" refers to one of the distinct voltage waveforms generated by the utility.

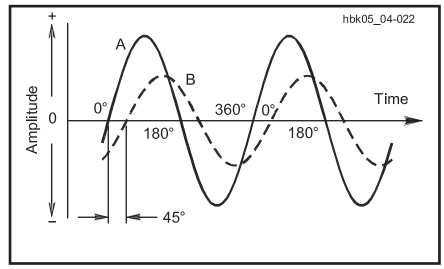

Figure 3.6 — When two waves of the same frequency start their cycles at slightly different times, the time difference or phase difference is measured in degrees. In this drawing, wave B starts 45° (one-eighth cycle) later than wave A, and so lags 45° behind A.

form is the *zero point* at which the positive *half cycle* begins as shown in Figure 3.4.

The advantage of treating the ac cycle in this way is that many calculations and measurements can be taken and recorded in a manner that is independent of frequency. The positive peak voltage or current occurs at 90° into the cycle. Relative to the starting point, 90° is the *phase* of the ac at that point. Phase is the position within an ac cycle expressed in degrees or *radians*. Thus, a complete description of an ac voltage or current involves reference to three properties: frequency, amplitude, and phase.

Each degree of phase represents the same amount of time. For example, a sine wave with a frequency of four cycles per second has a period T = 0.25 second and each degree of phase is equivalent to T/360 = 0.25 / 360 = 0.00069 second.

Phase relationships also permit the comparison of two ac voltages or currents at the same frequency. If the zero point of two signals with the frequency occur at the same time, there is zero phase difference between the signals and they are said to be *in phase.*

Figure 3.6 illustrates two waveforms with a constant phase difference. Since B crosses the zero point in the positive direction after A has already done so, there is a *phase difference* between the two waves. In the example, B *lags* A by 45°, or A *leads* B by 45°. If A and B occur in the same circuit, their composite waveform will also be a sine wave at an intermediate phase angle relative to each. Adding any number of sine waves of the same frequency always results in a sine wave at that frequency. Adding sine waves of different frequencies, as in Figure 3.1, creates a complex waveform with a *beat frequency* that is the difference between the two sine waves.

Figure 3.6 might equally apply to a voltage and a current measured in the same ac circuit. Either A or B might represent the voltage; that is, in some instances voltage will lead the current and in others voltage will lag the current.

Two important special cases appear in **Figure 3.7**. In Part A, line B lags 90° behind line A. Its cycle begins exactly one quarter cycle later than the A cycle. When one wave is passing through zero, the other just reaches its maximum value. In this example, the two sine waves are said to be *in quadrature*. If waveform B is a sine wave, waveform A is a *cosine wave*, leading waveform B by 90º. Quadrature signals form the basis of I/Q modulation, as is described in the **Modulation** chapter. (I and Q stand for In-Phase and Quadrature.)

In Part B, lines A and B are 180° *out of phase*, sometimes called *anti-phase*. In this case, it does not matter which one is considered to lead or lag. Line B is always positive

Vectors and Phasors

In Figure 3.3B, the arrow drawn from the center of the wheel to point A is a *vector* which has both an amplitude (its length) and direction (its angular position). The amplitude of the vector is equal to its length, which we arbitrarily decided would be 1 when constructing the table of sine values. Since the vector is pointing exactly along the horizontal axis, its direction is 0°. Thus, the vector is described as "1 at an angle of 0°" or "1 with a phase of 0°." In the *phasor notation* commonly used in radio, this is written $1\angle 0°$. The rotation of the wheel and the repeated cycles of the sine wave can also be described by a vector that is rotating (spinning around like the hand of a clock) at the frequency of the sine wave. You can find more information about vectors and phasors in the ARRL's "Radio Mathematics" article in the online content for this book.

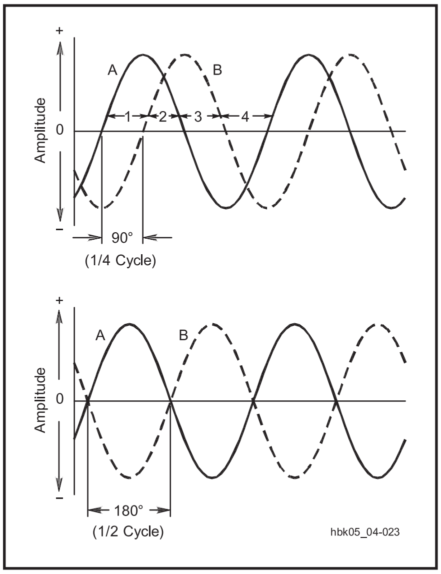

Figure 3.7 — Two important special cases of phase difference: In the upper drawing, the phase difference between A and B is 90°; in the lower drawing, the phase difference is 180°.

while line A is negative, and vice versa. If the two waveforms are of two voltages or two currents in the same circuit and if they have the same amplitude, they will cancel each other completely.

3.1.4 Time and Frequency Domain

To this point in the chapter, our discussions and illustrations have been in the *time domain* in which some characteristic of the signal (usually amplitude) is presented in relation to time. On a graph, this means the horizontal axis represents time. Events to the right take place later; events to the left occur earlier. For pure sinusoids (sine waves), this is enough to describe the signal.

For a complex signal with more than one sine wave, the time domain is insufficient to describe the necessary frequency and time information. The *frequency domain* is better for showing the characteristics of these signals as shown in **Figure 3.8**. A sine wave signal is shown as a vertical line, with the height of the line showing the signal's amplitude. Note that a sine wave signal occupies a single frequency.

To better understand the relationship between the time and frequency domains, refer to **Figure 3.9**. In Figure 3.9A, the three-dimensional coordinates show time (as the line sloping toward the bottom right); frequency (as the line sloping toward the top right); and amplitude (as the vertical axis). The two frequencies shown are harmoni-

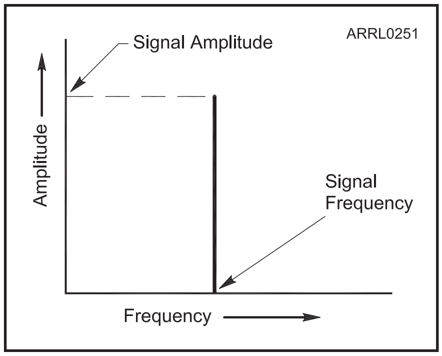

Figure 3.8 — A frequency domain or spectrum graph shows a sine wave as a single vertical line. The horizontal axis represents frequency and the vertical axis represents amplitude. The height of the line representing the sine wave shows its amplitude.

cally related (f_1 and $2f_1$). The time domain is represented in Figure 3.9B, in which all frequency components are added together. If the two frequencies were applied to the input of an oscilloscope, we would see the bold line that represents the amplitudes of the signals added together. The frequency domain contains information not found in the time domain, and vice versa.

The display shown in Figure 3.9C is typical of a spectrum analyzer's display of a complex waveform. (The spectrum analyzer is described in the **Test Equipment and Measurements** chapter.) In the figure the signal is separated into its individual frequency components, and a measurement made of the amplitude of each signal component. A signal's amplitude can be represented on the vertical scale as its voltage or as its power. You can see that using the frequency domain gives more information about the composition of the signal.

3.1.5 Complex Waveforms

A signal composed of more than one sine wave is called a *complex waveform*. A simple example of a complex waveform is the signaling waveform used by telephones when dialing. This waveform is composed of two different sine wave tones, thus the name "dual-tone multi-frequency" or DTMF for that signaling system. Listen carefully next time you dial and you will hear the two tones of different frequencies.

There are certain well-known and common complex waveforms that are made up of a sine wave and its harmonics. These are termed *regular* waveforms because the harmonic relationship of all the sine waves results in a waveform with a single overall frequency and period. A waveform that is made of sine

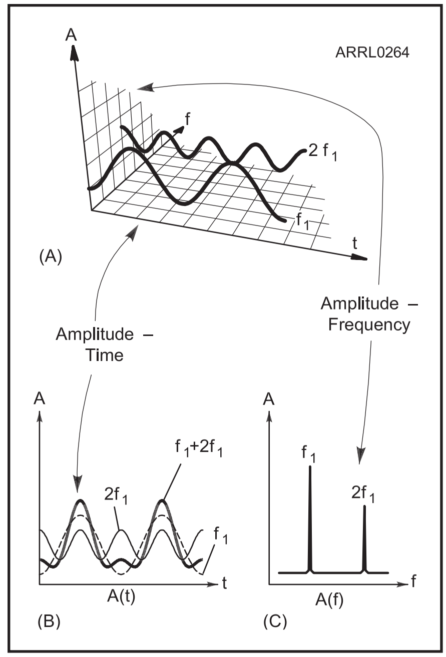

Figure 3.9 — This diagram shows how a complex waveform may be displayed in either the time domain or frequency domain. Part A is a three-dimensional display of amplitude, time, and frequency. At B, this information is shown in the time domain as on an oscilloscope. At C, the signal's frequency domain information

waves that are not harmonically related, such as human speech, is an *irregular* waveform. Whether regular or irregular, the sine waves that make up a complex waveform are called its *components*. (Analysis of complex signals in terms of its individual components is addressed in the **DSP and SDR Fundamentals** chapter.) A spreadsheet tool by Lou Ernst, WA2GKH, for experimenting with and comparing waveforms is available on this book's reference website in the Supplemental Information and Files section.

It is common for complex ac signals to contain a fundamental signal and a series of harmonics. Which harmonics are combined with the fundamental and the relative amplitude of each determine the final shape of the waveform as you will see in the following two sections. The set of all components that make up a signal is called the signal's *spectrum*. (More than one spectrum is *spectra*.)

SAWTOOTH WAVES

A *sawtooth* waveform, as shown in **Figure 3.10**, has a significantly faster *rise time* (the time it takes for the wave to reach a maximum value) compared to its *fall time* (the time it takes for the wave to reach a minimum value). A sawtooth wave is made up of a sine wave at its fundamental frequency and all of its har-

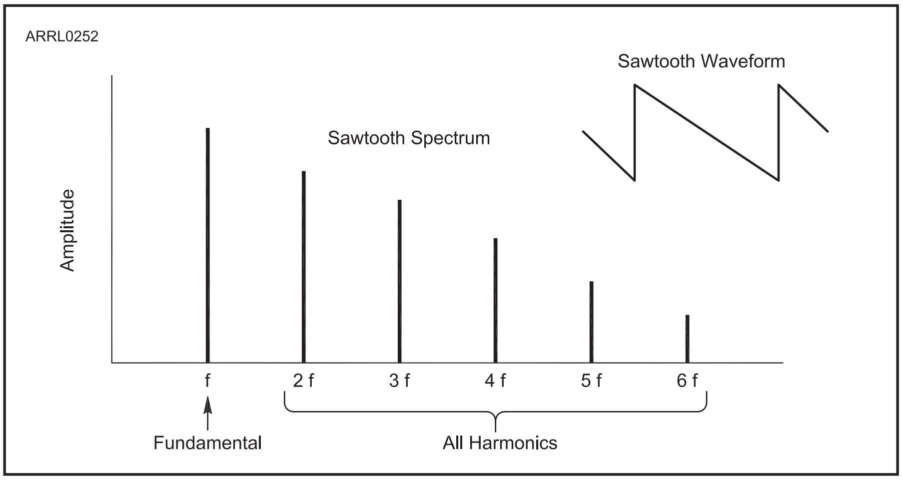

Figure 3.10 — The sawtooth waveform is made up of sine waves at the fundamental frequency and all of its harmonics. The amplitude of the harmonics decreases as their frequency increases.

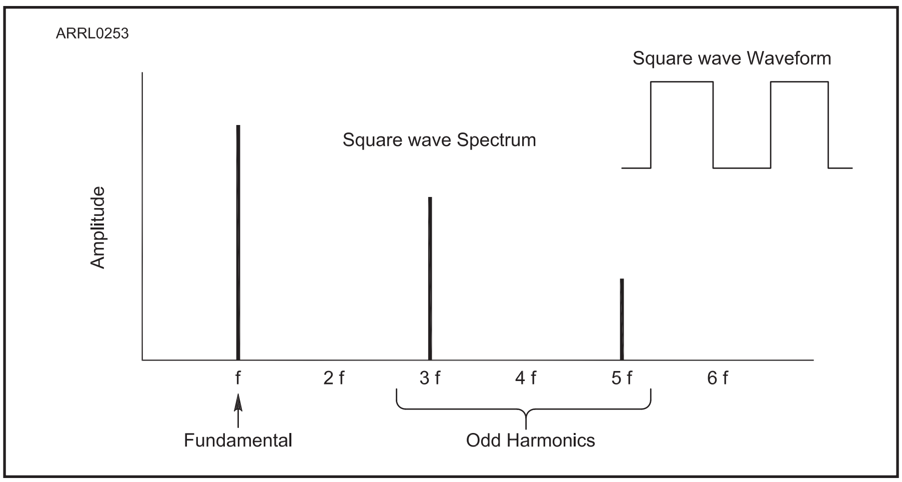

Figure 3.11 — The square wave is made up of sine waves at the fundamental frequency and only the odd harmonics. amplitude of the harmonics decreases as their frequency increases.

monics. The sawtooth's spectrum is shown in Figure 3.10. The *ramp* waveform is similar to the sawtooth but slowly rises (the ramp) then has a fast fall, the opposite of the sawtooth. Both the sawtooth and ramp waveforms are useful in timing circuits.

SQUARE WAVES

A square wave is one that abruptly changes back and forth between two voltage levels and remains an equal time at each level as in **Figure 3.11**. (If the wave spends an unequal time at each level, it is known as a *rectangular wave*.) A square wave is made up of sine waves at the fundamental and all the *odd* harmonic frequencies as shown in Figure 3.11.

3.2 Measuring AC Voltage, Current, and Power

Measuring the voltage or current in a dc circuit is straightforward, as **Figure 3.12A** demonstrates. Since the current flows in only one direction, the voltage and current have constant values until the resistor values are changed.

Figure 3.12B illustrates a perplexing problem encountered when measuring voltages and currents in ac circuits — the current and voltage continuously change direction and value. Which values are meaningful? How are measurements performed? In fact, there are several methods of measuring sine-wave voltage and current in ac circuits with each method providing different information about the waveform. Note that the following sections assume the waveform is a sine wave unless otherwise noted.

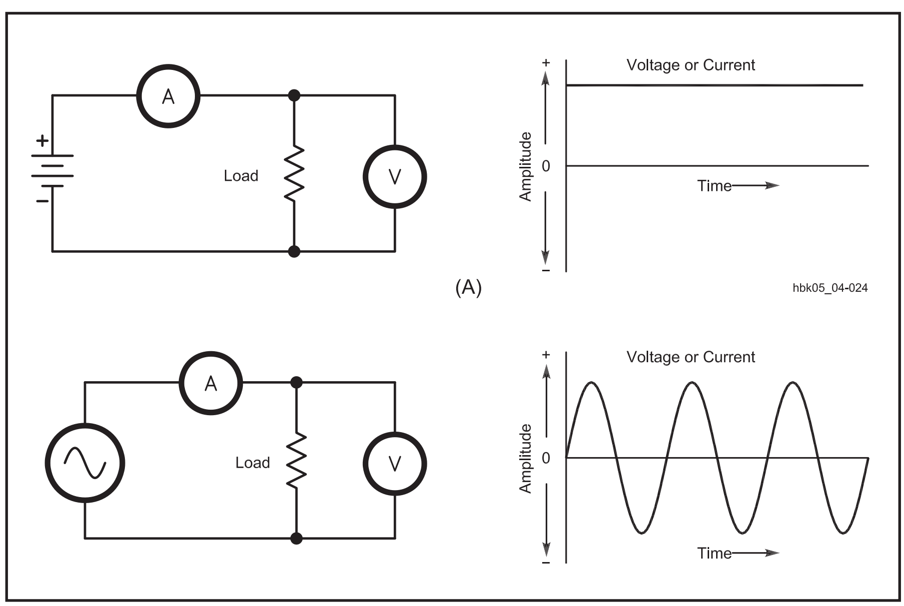

Figure 3.12 — Voltage and current measurements in dc (A) and ac circuits (B).

3.2.1 Instantaneous Values

By far, the most common waveform associated with ac of any frequency is the sine wave.

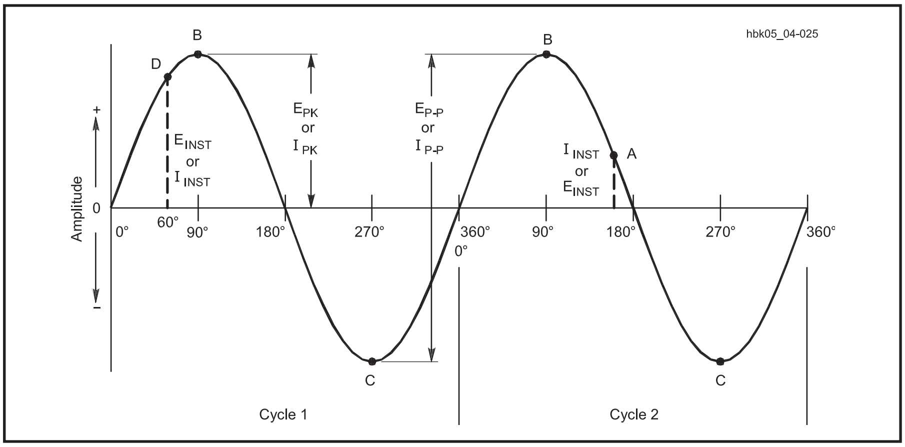

Figure 3.13 — Two cycles of a sine wave to illustrate instantaneous, peak, and peak-to-peak ac voltage and current values.

Unless otherwise noted, it is safe to assume that measurements of ac voltage or current are of a sinusoidal waveform. **Figure 3.13** shows a sine wave representing a voltage or current of some arbitrary frequency and amplitude. The *instantaneous* voltage (or current) is the value at one instant in time. If a series of instantaneous values are plotted against time, the resulting graph will show the waveform.

In the sine wave of Figure 3.13, the instantaneous value of the waveform at any point in time is a function of three factors: the maximum value of voltage (or current) along the curve (point B, E_{max}), the frequency of the wave, f, and the time elapsed from the preceding positive-going zero crossing, t, in seconds or fractions of a second. Thus,

$$E_{inst} = E_{max} \sin (ft)$$

assuming all sine calculations are done in degrees. (See the sidebar "Degrees, Radians and Angular Frequency." If the sine calculation is done in radians, substitute $2\pi ft$ for ft in the equation.)

If the point's phase is known — the position along the waveform — the instantaneous voltage at that point can be calculated directly as:

$$E_{inst} = E_{max} \sin \theta$$

where θ is the number of degrees of phase difference from the beginning of the cycle.

Example: What is the instantaneous value of voltage at point D in Figure 3.13, if the maximum voltage value is 120 V and point D's phase is 60.0°?

$$E_{inst} = 120\text{ V} \times \sin 60° = 120 \times 0.866 = 104\text{ V}$$

3.2.2 Peak and Peak-to-Peak Values

The most important of an ac waveform's instantaneous values are the maximum or *peak values* reached on each positive and negative half cycle. In Figure 3.13, points B and C represent the positive and negative peaks. Peak values (indicated by a "pk" or "p" subscript) are especially important with respect to component ratings, which the voltage or current in a circuit must not exceed without danger of component failure.

The *peak power* in an ac circuit is the product of the peak voltage and the peak current, or

$$P_{pk} = E_{pk} \times I_{pk}$$

The span from points B to C in Figure 3.13 represents the largest difference in value of the sine wave. Designated the *peak-to-peak* value (indicated by a "P-P" or "pk-pk" subscript), this span is equal to twice the peak value of the waveform. Thus, peak-to-peak voltage is:

$$E_{P\text{-}P} = 2\ E_{pk}$$

3.2.3 RMS Values

The *root mean square* or *RMS* values of voltage and current are the most common values encountered in electronics. Sometimes called *effective* values, the RMS value of an ac voltage or current is the value of a dc voltage or current that would cause a resistor to dissipate the same average amount of power as the ac waveform. This measurement became widely used in the early days of electrification when both ac and dc power utility power were in use. Even today, the values of the ac line voltage available from an electrical power outlet are given as RMS values. Unless otherwise specified, unlabeled ac voltage and current values found in most electronics literature are normally RMS values.

The RMS values of voltage and current get their name from the mathematical method used to derive their value relative to peak voltage and current. This procedure provides the RMS value for any type of periodic waveform, sinusoidal or not. Start by *squaring* the individual values of all the instantaneous values of voltage or current during an entire single cycle of ac. Take the average (*mean*) of these squares (this is done by computing an integral of the waveform) and then find the square *root* of that average.

SINE WAVE RMS VALUES

This section applies only when the waveform in question is a sine wave. The simple formulas and conversion factors in this section are generally *not* true for non-sinusoidal waveforms such as square or triangle waves (see the sidebar "Measuring Non-Sinusoidal Waveforms"). The following formulas are true *only* if the waveform is a sine wave and the circuit is *linear* — that is, raising or lowering the voltage will raise or lower the current

Measuring Nonsinusoidal Waveforms

Making measurements of ac waveforms is covered in more detail in the **Test Equipment and Measurements** chapter. However, this is a good point in the discussion to reinforce the dependence of RMS and values on the nature of the waveform being measured.

Analog meters and other types of instrumentation that display RMS values may only be calibrated for sine waves, those being the most common type of ac waveform. Using that instrumentation to accurately measure waveforms other than sine waves — such as speech, intermittent sine waves (such as CW from a transmitter), square waves, triangle waves or noise — requires the use of *calibration factors* or the measurement may not be valid.

To make calibrated, reliable measurements of the RMS or value of these waveforms requires the use of *true-RMS* instruments. These devices may use a balancing approach to a known dc value or if they are microprocessor-based, may actually perform the full root-mean-square calculation on the waveform. Be sure you know the characteristics of your test instruments if an accurate RMS value is important.

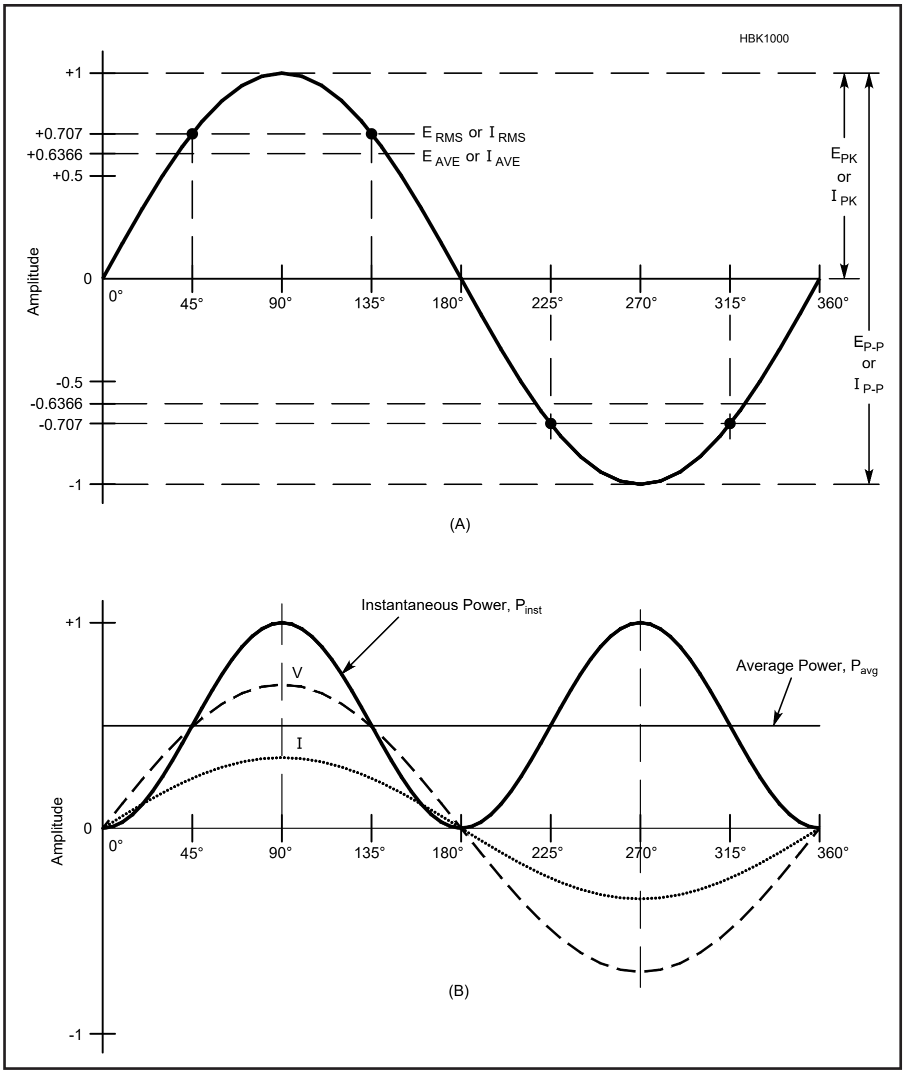

Figure 3.14 — (A) The relationships between RMS, average, peak, and peak-to-peak values of ac voltage and current for a sine wave. The numeric constants are different for non-sinusoidal waveforms. (B) The value of instantaneous power is the product of instantaneous voltage and current. Average power is the product of the RMS values of voltage and current.

Table 3.1
Conversion Factors for Sinusoidal AC Voltage or Current

From	*To*	*Multiply By*
Peak	Peak-to-Peak	2
Peak-to-Peak	Peak	0.5
Peak	RMS	$1/\sqrt{2}$ or 0.707
RMS	Peak	$\sqrt{2}$ or 1.414
Peak-to-Peak	RMS	$1/(2\times\sqrt{2})$ or 0.35355
RMS	Peak-to-Peak	$2\times\sqrt{2}$ or 2.828
Peak	Average	$2/\pi$ or 0.6366
Average	Peak	$\pi/2$ or 1.5708
RMS	Average	$(2\times\sqrt{2})/\pi$ or 0.90
Average	RMS	$\pi/(2\times\sqrt{2})$ or 1.11

Note: These conversion factors apply only to continuous pure sine waves.

proportionally. If those conditions are true, the following conversion factors have been computed and can be used without any additional mathematics.

For a sine wave to produce heat equivalent to a dc waveform the peak ac power required is twice the dc power. Therefore, the average ac power equivalent to a corresponding dc power is half the peak ac power.

$$P_{ave} = \frac{P_{pk}}{2}$$

A sine wave's RMS voltage and current values needed to arrive at average ac power are related to their peak values by the conversion factors:

$$E_{RMS} = \frac{E_{pk}}{\sqrt{2}} = \frac{E_{pk}}{1.414} = E_{pk} \times 0.707$$

$$I_{RMS} = \frac{I_{pk}}{\sqrt{2}} = \frac{I_{pk}}{1.414} = I_{pk} \times 0.707$$

RMS voltages and currents are what is displayed by most volt and ammeters.

If the RMS voltage is the peak voltage divided by $\sqrt{2}$, then the peak voltage must be the RMS voltage multiplied by $\sqrt{2}$, or

$$E_{pk} = E_{RMS} \times 1.414$$

$$I_{pk} = I_{RMS} \times 1.414$$

Example: What is the peak voltage and the peak-to-peak voltage at the usual household ac outlet, if the RMS voltage is 120 V?

$$E_{pk} = 120\text{ V} \times 1.414 = 170\text{ V}$$

$$E_{P\text{-}P} = 2 \times 170\text{ V} = 340\text{ V}$$

In the time domain of a sine wave, the instantaneous values of voltage and current correspond to the RMS values at the 45°, 135°, 225° and 315° points along the cycle shown in **Figure 3.14A**. (The sine of 45° is approximately 0.707.) The instantaneous value of voltage or current is greater than the RMS value for half the cycle and less than the RMS value for half the cycle.

Since circuit specifications will most commonly list only RMS voltage and current values, these relationships are important in finding the peak voltages or currents that will

Sine and Square Wave Measurement Definitions

Since square waves are very common waveforms, the following table provides definitions for the two types of waveforms. These are *not* conversion factors between two measurements unless a true-RMS instrument is used.

AC Measurements for Sine and Square Waves

	Sine Wave	*Square Wave*
Peak-to-Peak	2 × Peak	2 × Peak
Peak	0.5 × Peak-to-Peak	0.5 × Peak-to-Peak
RMS	0.707 × Peak	Peak
Peak	1.414 × RMS	RMS
Average	0 (full cycle)	0 (full cycle)
	0.637 × Peak (half cycle)	0.5 × Peak (half cycle)

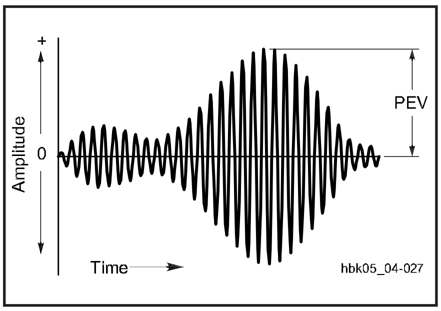

Figure 3.15 — The peak envelope voltage (PEV) for a composite waveform.

stress components.

Example: What is the peak voltage across a capacitor if the RMS voltage of a sinusoidal waveform signal across it is 300 V ac?

$$E_{pk} = 300 \text{ V} \times 1.414 = 424 \text{ V}$$

The capacitor must be able to withstand this higher voltage, plus a safety margin. (The capacitor must also be rated for ac use because of the continually reversing polarity and ac current flow.) In power supplies that convert ac to dc and use capacitive input filters, the output voltage will approach the peak value of the ac voltage rather than the RMS value. (See the **Power Sources** chapter for more information on specifying components in this application.)

3.2.4 Average Values of AC Waveforms

Certain kinds of circuits respond to the *average* voltage or current (not power) of an ac waveform. Among these circuits are analog electrodynamic meter movements and power supplies that convert ac to dc and use heavily inductive ("choke") input filters, both of which work with the pulsating dc output of a full-wave rectifier. The average value of each ac half cycle is the *mean* of all the instantaneous values in that half cycle. (The average value of a sine wave or any symmetric ac waveform over an entire cycle is zero!) Related to the peak values of voltage and current, average values for each half-cycle of a sine wave are 2/π (or 0.6366) times the peak value.

$$E_{ave} = 0.6366\ E_{pk}$$

$$I_{ave} = 0.6366\ I_{pk}$$

For convenience, **Table 3.1** summarizes the relationships between all of the common ac values. All of these relationships apply *only* to sine waves in linear circuits.

3.2.5 Average Power of AC Waveforms

The *instantaneous power*, P_{inst}, of an ac waveform as shown in **Figure 3.14B** is equal to the instantaneous values of voltage and current multiplied together:

$$P_{inst} = E_{inst} \times I_{inst}$$

Note that the frequency of the instantaneous power waveform is twice the frequency of voltage and current.

Average power, P_{avg}, is used for measuring the power of ac signals. It is determined by multiplying the RMS values of voltage and current. RMS values are calculated over an entire cycle, so average power applies only to an entire cycle:

$$P_{avg} = E_{RMS} \times I_{RMS}$$

This equation assumes the voltage and current are in-phase as shown. If reactance is present, there will be a phase shift between voltage and current — see the section "Reactive Power and Power Factor" later in this chapter. Although an RMS value can be calculated for the instantaneous power waveform, it is not equal to the average power and does not have any physical significance.

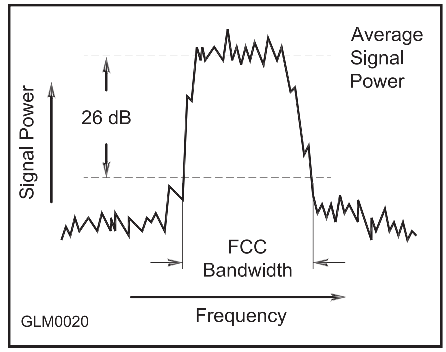

Figure 3.16 — The FCC defines bandwidth as "the width of a frequency band outside of which the mean (average) power of the transmitted signal is attenuated at least 26 dB below the mean power." (FCC §97.3(a)(8))

3.2.6 Complex Waveforms and Peak Envelope Values

Complex waveforms, as shown earlier in Figures 3.10 and 3.11, differ from sine waves. For speech signals, the peak voltage may vary significantly from one cycle to the next, for example. Therefore, other amplitude measures are required, especially for accurate measurement of voltage and power with transmitted speech or data waveforms.

An SSB waveform (either speech or data) contains an RF ac waveform with a frequency many times that of the audio-frequency ac waveform with which it is combined. Therefore, the resultant *composite* waveform appears as an amplitude envelope superimposed upon the RF waveform as illustrated by **Figure 3.15**. For a complex waveform such as this, the *peak envelope voltage* (PEV) is the maximum or peak value of voltage anywhere in the waveform.

Peak envelope voltage is used in the calculation of *peak envelope power* (PEP). The Federal Communications Commission (FCC) sets the maximum power levels for amateur transmitters in terms of peak envelope power. PEP is the *average* power supplied to the antenna transmission line by the transmitter during one RF cycle at the crest of the modulation envelope, taken under normal operating conditions. That is, the average power for the RF cycle during which PEV occurs.

Since calculation of PEP requires the average power of the cycle, and the deviation of the modulated RF waveform from a sine wave is very small, the error incurred by using the conversion factors for sine waves is insignificant. Multiply PEV by 0.707 to obtain an RMS value. Then calculate PEP by using the square of the voltage divided by the load resistance.

$$PEP = \frac{(PEV \times 0.707)^2}{R}$$

Example: What is the PEP of a transmitter's output with a PEV of 100 V into a 50-ohm load?

$$PEP = \frac{(100 \times 0.707)^2}{R} = \frac{(70.7)^2}{50} = 100 \text{ W}$$

COMPOSITE WAVEFORMS AND BANDWIDTH

Composite signals are groups of individual signals that combine to create a complex signal. Composite signals have *components* that generally cover a range of frequencies. The general definition of a signal's *bandwidth* is the difference in frequency between the two points at which the signal's amplitude falls to 3 dB below its peak value. For some kinds of very simple and very complex signals this may not be a useful definition.

The FCC has a more specific definition of bandwidth in section §97.3(a)(8): "*Bandwidth*. The width of a frequency band outside of which the mean [average] power of the transmitted signal is attenuated at least 26 dB below the mean power within the band." **Figure 3.16** illustrates how this measurement is made. This definition is used when evaluating a signal's occupied bandwidth to determine whether it satisfies FCC rules.

3.3 Effective Radiated Power

When evaluating total station performance, accounting for the effects of the entire system is important, including antenna gain. This allows you to evaluate the effects of changes to the station. Transmitting performance is usually computed as *effective radiated power* (*ERP*). ERP is calculated with respect to a reference antenna system — usually a dipole but occasionally an isotropic antenna — and answers the question, "How much power does my station radiate as compared to that if my antenna was a simple dipole?" Effective isotropic radiated power (EIRP) results when an isotropic antenna is used as the reference. If no antenna reference is specified, assume a dipole reference antenna.

ERP is especially useful in designing and coordinating repeater systems. The effective power radiated from the antenna helps establish the coverage area of the repeater. In addition, the height of the repeater antenna as compared to buildings and mountains in the surrounding area (*height above average terrain*, or *HAAT*) has a large effect on the repeater coverage. In general, for a given coverage area, with a greater antenna HAAT, less effective radiated power (ERP) is needed. A frequency coordinator may even specify a maximum ERP for a repeater, to help reduce interference between stations using the same frequencies.

ERP calculations begin with the *transmitter power output* (*TPO*). (This is assumed to be the output of the final power amplification stage if an external power amplifier is used.) Then the *system gain* of the entire antenna system including the antenna, the transmission line, and all transmission line components is applied to TPO to compute the entire station's output power.

System Gain = Transmission Line Loss – Transmission Components Loss + Antenna Gain

There is always some power lost in the feed line and often there are other devices inserted in the line, such as a filter or an impedance-matching network. In the case of a repeater system, there is usually a duplexer so the transmitter and receiver can use the same antenna and perhaps a circulator to reduce the possibility of intermodulation interference. These devices also introduce some loss to the system. The antenna system then usually returns some gain to the system. (See the **Antennas** chapter for information on antenna gain and the **Transmission Lines** chapter for information on feed line loss.)

ERP = TPO × System Gain

Since the system gains and losses are usually expressed in decibels, they can simply be added together, with losses written as negative values. System gain must then be converted back to a linear value from dB to calculate ERP.

$$ERP = TPO \times \log^{-1}\left(\frac{\text{System Gain (dB)}}{10}\right)$$

It is also common to work entirely in dBm and dB until the final result for ERP is obtained and then converted back to watts. (dBm represents "decibels with respect to one mW" such that 0 dBm = 1 mW, 30 dBm = 1 W, and so forth.)

ERP (in dBm) = TPO (in dBm) + System Gain (in dB)

Suppose we have a repeater station that uses a 50 W transmitter and a feed line with 4 dB of loss. There is a duplexer in the line that exhibits 2 dB of loss and a circulator that adds another 1 dB of loss. This repeater uses an antenna that has a gain of 6 dBd. Our total system gain looks like:

System gain = –4 dB + –2 dB + –1 dB + 6 dBd = –1 dB

Note that this is a loss of 1 dB total for the system from TPO to radiated power. The effect on the 50 W of TPO results in:

$$ERP = 50 \text{ W} \times \log^{-1}\left(\frac{\text{system gain (dB)}}{10}\right)$$

$$= 50 \times \log^{-1}(-0.1) = 50 \times 0.79 = 39.7 \text{ W}$$

This is consistent with the expectation that with a 1 dB system loss we would have somewhat less ERP than transmitter output power.

As another example, suppose we have a transmitter that feeds a 100 W output signal into a feed line that has 1 dB of loss. The feed line connects to an antenna that has a gain of 6 dBd. What is the effective radiated power from the antenna? To calculate the total system gain (or loss) we add the decibel values given:

System gain = – 1 dB + 6 dBd = 5 dB

and

$$ERP = 100 \text{ W} \times \log^{-1}\left(\frac{\text{system gain (dB)}}{10}\right)$$

$$= 100 \times \log^{-1}(0.5) = 100 \times 3.16 = 316 \text{ W}$$

The total system has positive gain, so we should have expected a larger value for ERP than TPO. Keep in mind that the gain antenna concentrates more of the signal in a desired direction, with less signal in undesired directions. So the antenna doesn't really increase the total available power. If directional antennas are used, ERP will change with direction.

Example: What is the effective radiated power of a repeater station with 150 W transmitter power output, 2 dB feed line loss, 2.2 dB duplexer loss and 7 dBd antenna gain?

The Decibel

The decibel (dB) is discussed in the section on Gain, later in this chapter. It is also covered in the "Untangling the Decibel Dilemma" item in the online content.

System gain = –2 dB – 2.2 dB + 7 dBd = 2.8 dB

$$\text{ERP} = 150\ \text{W} \times \log^{-1}\left(\frac{\text{system gain (dB)}}{10}\right)$$

$$= 150 \times \log^{-1}(0.28) = 150 \times 1.9 = 285\ \text{W}$$

Example: What is the effective radiated power of a repeater station with 200 W transmitter power output, 4 dB feed line loss, 3.2 dB duplexer loss, 0.8 dB circulator loss and 10 dBd antenna gain?

System gain = –4 – 3.2 – 0.8 + 10 = 2 dB

$$\text{ERP} = 200\ \text{W} \times \log^{-1}\left(\frac{\text{system gain (dB)}}{10}\right)$$

$$= 200 \times \log^{-1}(0.2) = 200 \times 1.58 = 317\ \text{W}$$

What is the effective isotropic radiated power of a repeater station with 200 W transmitter power output, 2 dB feed line loss, 2.8 dB duplexer loss, 1.2 dB circulator loss and 7 dBi antenna gain?

System gain = –2 – 2.8 – 1.2 + 7 = 1 dB

$$\text{ERP} = 200\ \text{W} \times \log^{-1}\left(\frac{\text{system gain (dB)}}{10}\right)$$

$$= 100 \times \log^{-1}(0.1) = 200 \times 1.26 = 252\ \text{W}$$

3.4 AC in Capacitors and Inductors

Both capacitors and inductors can store electrical or magnetic energy, respectively. When an ac signal is applied to them, the storing and releasing of energy results in a phase shift between the current and voltage waveforms. The phase shift varies with frequency and makes capacitors and inductors behave differently than resistors. They also interact to create frequency-sensitive tuned circuits, filters, impedance-matching circuits, and more.

3.4.1 Alternating Current in Capacitance

While a capacitor in a dc circuit will appear as an open circuit except for the brief charge and discharge periods, the same capacitor in an ac circuit will both pass and oppose current. A capacitor in an ac circuit does not handle electrical energy like a resistor, however. Instead of converting the energy to heat and dissipating it, capacitors store electrical energy when the applied voltage is greater than that across the capacitor and return it to the circuit when the opposite is true.

In **Figure 3.17** a sine-wave ac voltage having a maximum value of 100 V is applied to a capacitor. In the period OA, the applied voltage increases from 0 to 38, storing energy in the capacitor; at the end of this period the capacitor is charged to that voltage. In interval AB the voltage increases to 71; that is, by an additional 33 V. During this interval a smaller quantity of charge has been added than in OA, because the voltage rise during interval AB is smaller. Consequently the average current during interval AB is smaller than during OA. In the third interval, BC, the voltage rises from 71 to 92, an increase of 21 V. This is less than the voltage increase during AB, so the quantity of charge added is less; in other words, the average current during interval BC is still smaller. In the fourth interval, CD, the voltage increases only 8 V; the charge added is smaller than in any preceding interval and therefore the current also is smaller.

By dividing the first quarter-cycle into a very large number of such intervals, it can be shown that the current charging the capacitor has the shape of a sine wave, just as the applied voltage does. The current is largest at the beginning of the cycle and becomes zero at the maximum value of the voltage, so there is a phase difference of 90° between the voltage and the current. During the first quarter-cycle the current is flowing in the original (positive) direction through the circuit as indicated by the dashed line in Figure 3.17, since the capacitor is being charged. The increasing capacitor voltage indicates that energy is being stored in the capacitor.

In the second quarter-cycle — that is, in the time from D to H — the voltage applied to the capacitor decreases. During this time the capacitor loses charge, returning the stored energy to the circuit. Applying the same reasoning, it is evident that the current is small in interval DE and continues to increase during each succeeding interval. The current is flowing *against* the applied voltage, however, because the capacitor is returning energy to (discharging into) the circuit. The current thus flows in the *negative* direction during this quarter-cycle.

The third and fourth quarter-cycles repeat the events of the first and second, respectively, although the polarity of the applied voltage has reversed, and so the current changes to correspond. In other words, an alternating current flows in the circuit because of the alternate charging and discharging of the capacitance. As shown in Figure 3.17, the current starts its cycle 90° before the voltage,

Capacitive Reactance Timesaver

The fundamental units for frequency and capacitance (hertz and farads) are inconvenient for practical use in radio circuits. If the capacitance is specified in microfarads (µF) and the frequency is in megahertz (MHz), however, the reactance is calculated in ohms (Ω).

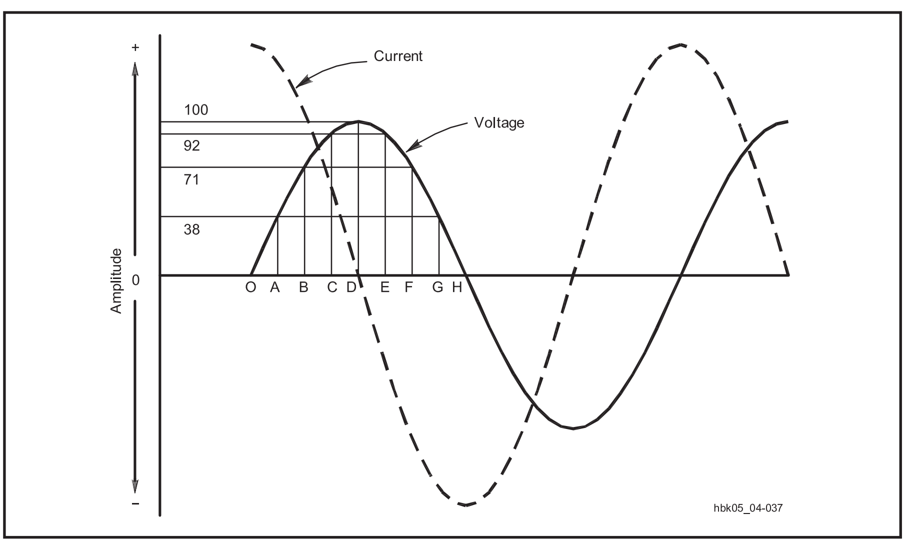

Figure 3.17 — Voltage and current phase relationships when an alternating current is applied to a capacitor.

so the current in a capacitor *leads* the applied voltage by 90°. You might find it helpful to remember the word "ICE" as a mnemonic because the current (I) in a capacitor (C) comes before voltage (E). (See the sidebar "ELI the ICE man" in the section on inductors.) We can also turn this statement around, to say the voltage in a capacitor *lags* the current by 90°.

3.4.2 Capacitive Reactance and Susceptance

The quantity of electric charge that can be placed on a capacitor is proportional to the applied voltage and the capacitance. If the applied voltage is ac, this amount of charge moves back and forth in the circuit once each cycle. Therefore, the rate of movement of charge (the current) is proportional to voltage, capacitance and frequency. Stated in another way, capacitor current is proportional to capacitance for a given applied voltage and frequency.

When the effects of capacitance and frequency are considered together, they form a quantity called *reactance* that relates voltage and current in a capacitor, similar to the role of resistance in Ohm's Law. Because the reactance is created by a capacitor, it is called *capacitive reactance*. The units for reactance are ohms, just as in the case of resistance. Although the units of reactance are ohms, there is no power dissipated in reactance. The energy stored in the capacitor during one portion of the cycle is simply returned to the circuit in the next.

The formula for calculating the magnitude of the capacitive reactance is:

$$X_C = \frac{1}{2\,\pi\,f\,C}$$

where:

X_C = magnitude of capacitive reactance in ohms,
f = frequency in hertz,
C = capacitance in farads, and
π = 3.1416.

By convention, capacitive reactance is assigned a negative value whereas inductive reactance (discussed below) is assigned a positive value.

Note: In many references and texts, angular frequency $\omega = 2\pi f$ is used and the equation would read:

$$X_C = \frac{1}{\omega\,C}$$

Example: What is the reactance of a capacitor of 470 pF (0.000470 µF) at a frequency of 7.15 MHz?

$$X_C = \frac{1}{2\,\pi\,f\,C}$$

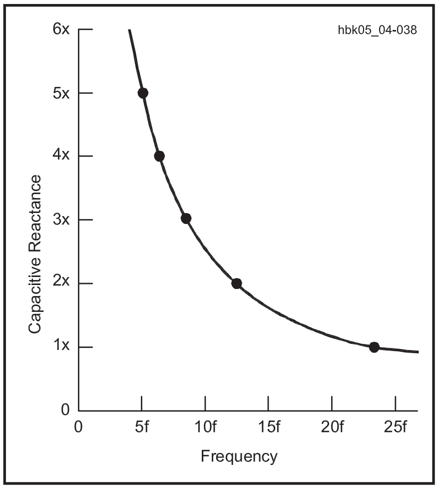

Figure 3.18 — A graph showing the general relationship of reactance to frequency for a fixed value of capacitance.

$$= \frac{1}{2\,\pi \times 7.15\text{ MHz} \times 0.000470\ \mu\text{F}}$$

$$= \frac{1\,\Omega}{0.0211} = 47.4\,\Omega$$

Example: What is the reactance of the same capacitor, 470 pF (0.000470 µF), at a frequency of 14.29 MHz?

$$X_C = \frac{1}{2\,\pi\,f\,C}$$

$$= \frac{1}{2\,\pi \times 14.3\text{ MHz} \times 0.000470\ \mu\text{F}}$$

$$= \frac{1\,\Omega}{0.0422} = 23.7\,\Omega$$

Current in a capacitor is directly related to the rate of change of the capacitor voltage. The maximum rate of change of voltage in a sine wave increases directly with the frequency, even if its peak voltage remains fixed. Therefore, the maximum current in the capacitor must also increase directly with frequency. Since, if voltage is fixed, an increase in current is equivalent to a decrease in reactance, the reactance of any capacitor decreases proportionally as the frequency increases. **Figure 3.18** illustrates the decrease in reactance of an arbitrary-value capacitor with respect to increasing frequency. The only limitation on the application of the graph is the physical construction of the capacitor, which may favor low-frequency uses or high-frequency applications.

CAPACITIVE SUSCEPTANCE

Just as conductance is sometimes the most useful way of expressing a resistance's ability to conduct current, the same is true for capacitors and ac current. This ability is called *susceptance* (abbreviated *B*). The units of susceptance are siemens (S), the same as that of conductance and admittance.

Susceptance in a capacitor is *capacitive susceptance*, abbreviated B_C. In an ideal capacitor with no resistive losses, susceptance is simply the negative reciprocal of reactance.

$$B_C = -\frac{1}{X_C}$$

where

X_C is the capacitive reactance, and
B_C is the capacitive susceptance.

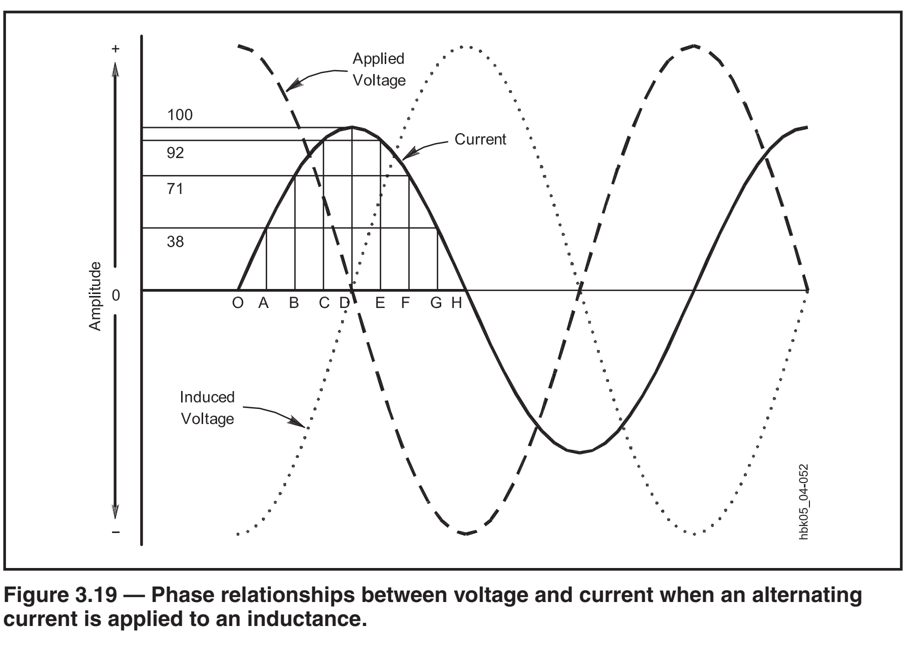

Figure 3.19 — Phase relationships between voltage and current when an alternating current is applied to an inductance.

3.4.3 Alternating Current in Inductors

For reasons similar to those that cause a phase difference between current and voltage in a capacitor, when an alternating voltage is applied to an ideal inductance with no resistance, the current is 90° out of phase with the applied voltage. In the case of an inductor, however, the current *lags* 90° behind the voltage as shown in **Figure 3.19**, the opposite of the capacitor current-voltage relationship. (Here again, we can also say the voltage across an inductor *leads* the current by 90°.) Interpreting Figure 3.19 begins with understanding that the cause for current lag in an inductor is the opposing voltage that is induced in the inductor and that the amplitude of the opposing voltage is proportional to the rate at which the inductor current changes.

In time segment OA, when the applied voltage is at its positive maximum, the rate at which the current is changing is also the highest, a 38% change. This means that the opposing voltage is also maximum, allowing the least current to flow. In the segment AB, as a result of the decrease in the applied voltage, current changes by only 33% inducing a smaller opposing voltage. The process continues in time segments BC and CD, the latter producing only an 8% rise in current as the applied and induced opposing voltage approach zero.

In segment DE, the applied voltage changes polarity, causing current to begin to decrease, returning stored energy to the circuit from the inductor's magnetic field. As the current rate of change is now negative (decreasing) the induced opposing voltage also changes polarity. Current flow is still in the original direction (positive), but is decreasing as less energy is stored in the inductor.

As the applied voltage continues to increase negatively, the current — although still positive — continues to decrease in value, reaching zero as the applied voltage reaches its negative maximum. The energy once stored in the inductor has now been completely returned to the circuit. The negative half-cycle then continues just as the positive half-cycle.

Similarly to the capacitive circuit discussed earlier, by dividing the cycle into a large number of intervals, it can be shown that the current and voltage are both sine waves, although with a difference in phase.

Compare Figure 3.19 with Figure 3.17. Whereas in a pure capacitive circuit, the current *leads* the voltage by 90º, in a pure inductive circuit, the current *lags* the voltage by 90º. These phenomena are especially important in circuits that combine inductors and capacitors. Remember that the phase difference between voltage and current in both types of circuits is a result of energy being stored and released as voltage across a capacitor and as current in an inductor.

ELI the ICE Man

If you have difficulty remembering the phase relationships between voltage and current with inductors and capacitors, you may find it helpful to think of the phrase, "ELI the ICE man." This will remind you that voltage across an inductor leads the current through it, because the E comes before (leads) I, with an L between them, as you read from left to right. (The letter L represents inductance.) Similarly, I comes before (leads) E with a C between them.

3.4.4 Inductive Reactance and Susceptance

The amount of current that can be created in an inductor is proportional to the applied voltage but inversely proportional to the inductance because of the induced opposing voltage. If the applied voltage is ac, the rate of change of the current varies directly with the frequency and this rate of change also determines the amplitude of the induced or reverse voltage. Hence, the opposition to the flow of current increases proportionally to frequency. Stated in another way, inductor current is inversely proportional to inductance for a given applied voltage and frequency.

The combined effect of inductance and frequency is called *inductive reactance*, which — like capacitive reactance — is expressed in ohms. As with capacitive reactance, no power is dissipated in inductive reactance. The energy stored in the inductor during one portion of the cycle is returned to the circuit in the next portion.

The formula for calculating the magnitude of the inductive reactance is:

$$X_L = 2 \pi f L$$

where

X_L = magnitude of inductive reactance in ohms,
f = frequency in hertz,
L = inductance in henrys, and
π = 3.1416.
(If $\omega = 2 \pi f$, then $X_L = \omega L$.)

Example: What is the reactance of an inductor having an inductance of 8.0 H at a frequency of 120 Hz?

$$X_L = 2 \pi f L$$

$$= 6.2832 \times 120 \text{ Hz} \times 8.0 \text{ H}$$

$$= 6{,}030\ \Omega$$

Inductive Reactance Timesaver

Similarly to the calculation of capacitive reactance, if inductance is specified in microhenrys (µH) and the frequency is in megahertz (MHz), the reactance is calculated in ohms (Ω). The same is true for the combination of mH and kHz.

Example: What is the reactance of a 15.0-microhenry inductor at a frequency of 14.0 MHz?

$$X_L = 2 \pi f L$$

$$= 6.2832 \times 14.0 \text{ MHz} \times 15.0\ \mu\text{H}$$

$$= 1{,}320\ \Omega$$

The resistance of the wire used to make the inductor has no effect on the reactance, but simply acts as a separate resistor connected in series with the inductor.

Example: What is the reactance of the same inductor at a frequency of 7.0 MHz?

$$X_L = 2 \pi f L$$

$$= 6.2832 \times 7.0 \text{ MHz} \times 15.0\ \mu\text{H}$$

$$= 660\ \Omega$$

The direct relationship between frequency and reactance in inductors, combined with the inverse relationship between reactance and frequency in the case of capacitors, will be of fundamental importance in creating resonant circuits.

INDUCTIVE SUSCEPTANCE

As a measure of the ability of an inductor to limit the flow of ac in a circuit, inductive reactance is similar to capacitive reactance in having a corresponding susceptance, or ability to pass ac current in a circuit. In an ideal inductor with no resistive losses — that is, no energy lost as heat — susceptance is simply the negative reciprocal of reactance.

$$B_L = -\frac{1}{X_L}$$

where

X_L = the inductive reactance, and
B_L = the inductive susceptance.

The unit of susceptance for both inductors and capacitors is the *siemens*, abbreviated S.

3.5 Working with Reactance

3.5.1 Ohm's Law for Reactance

Only ac circuits containing capacitance or inductance (or both) have reactance. Despite the fact that the voltage in such circuits is 90° out of phase with the current, circuit reactance does oppose the flow of ac current in a manner that corresponds to resistance. That is, in a capacitor or inductor, reactance is equal to the ratio of ac voltage to ac current and the equations relating voltage, current and reactance take the familiar form of Ohm's Law:

$$E = I\,X$$

$$I = E / X$$

$$X = E / I$$

where

E = RMS ac voltage in volts,

I = RMS ac current in amperes, and

X = inductive or capacitive reactance in ohms.

Example: What is the voltage across a capacitor of 200 pF at 7.15 MHz, if the current through the capacitor is 50 mA?

Since the reactance of the capacitor is a function of both frequency and capacitance, first calculate the reactance:

$$X_C = \frac{1}{2\,\pi\,f\,C}$$

$$= \frac{1}{2 \times 3.1416 \times 7.15 \times 10^6\ \text{Hz} \times 200 \times 10^{-12}\ \text{F}}$$

$$= \frac{10^6\ \Omega}{8980} = 111\,\Omega$$

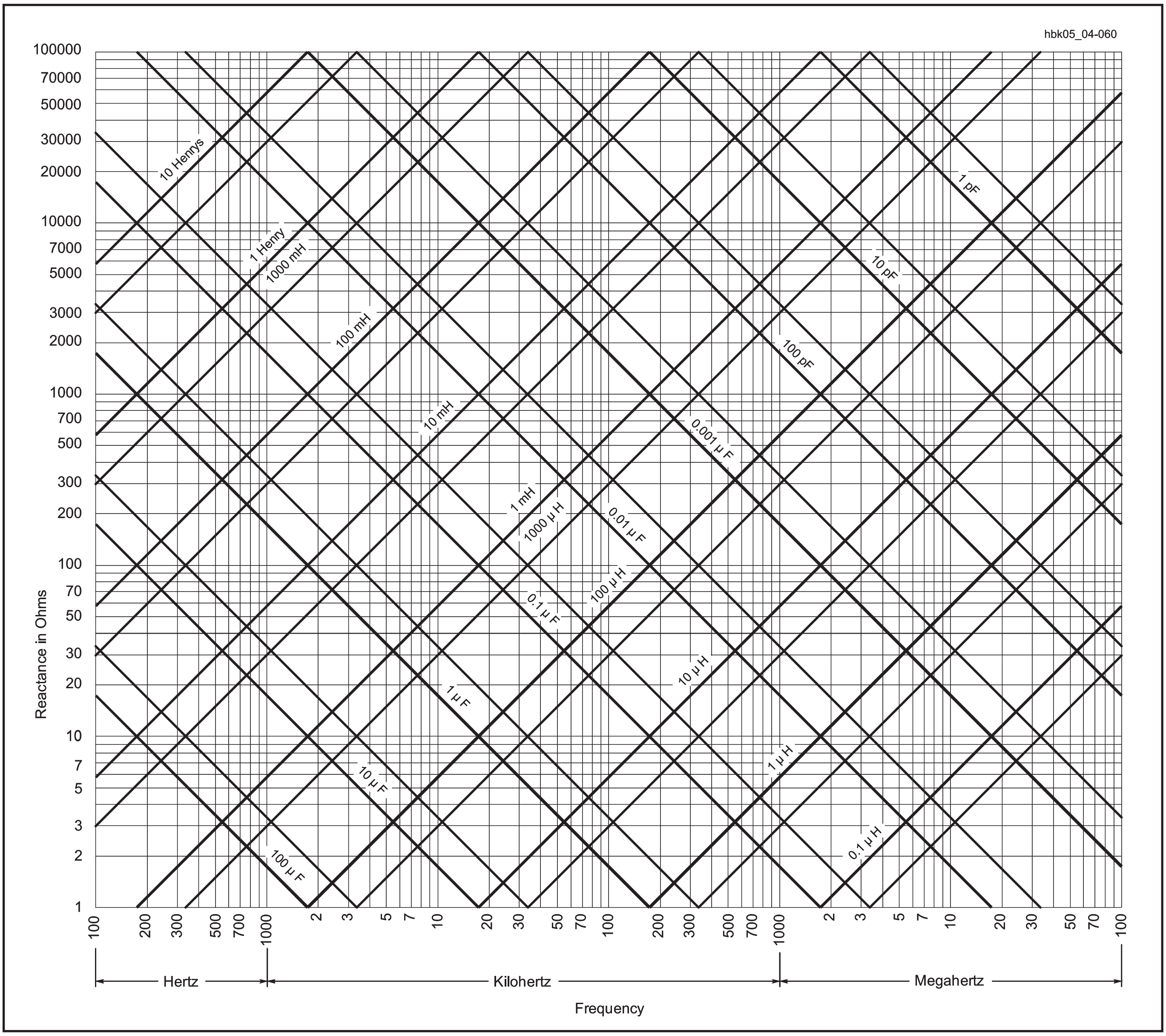

Figure 3.20 — Inductive and capacitive reactance vs frequency. Heavy lines represent multiples of 10, intermediate lines multiples of 5. For example, the light line between 10 µH and 100 µH represents 50 µH; the light line between 0.1 µF and 1 µF represents 0.5 µF, and so on. Other values can be extrapolated from the chart. For example, the reactance of 10 H at 60 Hz can be found by taking the reactance of 10 H at 600 Hz and dividing by 10 for the 10 times decrease in frequency. (Originally from Terman, *Radio Engineer's Handbook*. See references.)

Next, use:

$E = I \times X_C = 0.050\ A \times 111\ \Omega = 5.6\ V$

Example: What is the current through an 8.0-H inductor at 120 Hz, if 420 V is applied?

$X_L = 2\ \pi\ f\ L$

$= 2 \times 3.1416 \times 120\ Hz \times 8.0\ H$

$= 6030\ \Omega$

$I = E / X_L = 420 / 6030 = 69.6\ mA$

Figure 3.20 charts the reactances of capacitors from 1 pF to 100 µF, and the reactances of inductors from 0.1 µH to 10 H, for frequencies between 100 Hz and 100 MHz. Approximate values of reactance can be read or interpolated from the chart. The formulas will produce more exact values, however. (The chart can also be used to find the frequency at which an inductor and capacitor have equal reactances, creating resonance as described in the section "Reactances At and Near Resonance" below.)

Although both inductive and capacitive reactance oppose the flow of ac current, the two types of reactance differ. With capacitive reactance, the current *leads* the voltage by 90°, whereas with inductive reactance, the current *lags* the voltage by 90°. The convention for charting the two types of reactance appears in **Figure 3.21**. On this graph, inductive reactance is plotted along the +90° vertical line, while capacitive reactance is plotted along the –90° vertical line. This convention of assigning a positive value to inductive reactance and a negative value to capacitive reactance results from the mathematics used for working with impedance as described elsewhere in this chapter.

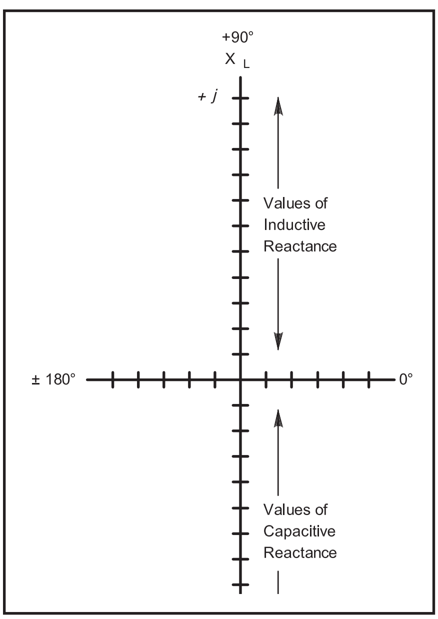

Figure 3.21 — The conventional method of plotting reactances on the vertical axis of a graph, using the upward or "plus" direction for inductive reactance and the downward or "minus" direction for capacitive reactance. The horizontal axis will be used for resistance in later examples.

3.5.2 Reactances in Series and Parallel

If a circuit contains two reactances of the same type, whether in series or in parallel, the resulting reactance can be determined by applying the same rules as for resistances in series and in parallel. Series reactance is given by the formula

$X_{total} = X1 + X2 + X3 ... + X_n$

Example: Two noninteracting inductances are in series. Each has a value of 4.0 µH, and the operating frequency is 3.8 MHz. What is the resulting reactance?

The reactance of each inductor is:

$X_L = 2\ \pi\ f\ L$

$= 2 \times 3.1416 \times 3.8 \times 10^6\ Hz \times 4 \times 10^{-6}\ H$

$= 96\ \Omega$

$X_{total} = X1 + X2 = 96\ \Omega + 96\ \Omega = 192\ \Omega$

We might also calculate the total reactance by first adding the inductances:

$L_{total} = L1 + L2 = 4.0\ \mu H + 4.0\ \mu H = 8.0\ \mu H$

$X_{total} = 2\ \pi\ f\ L$

$= 2 \times 3.1416 \times 3.8 \times 10^6\ Hz \times 8.0 \times 10^{-6}\ H$

$= 191\ \Omega$

(The fact that the last digit differs by one illustrates the uncertainty of the calculation caused by the limited precision of the measured values in the problem, and differences caused by rounding off the calculated values. This also shows why it is important to follow the rules for significant figures.)

Example: Two noninteracting capacitors are in series. One has a value of 10.0 pF, the other of 20.0 pF. What is the resulting reactance in a circuit operating at 28.0 MHz?

$$X_{C1} = \frac{1}{2\ \pi\ f\ C}$$

$$= \frac{1}{2 \times 3.1416 \times 28.0 \times 10^6\ Hz \times 10.0 \times 10^{-12}\ F}$$

$$= \frac{10^6\ \Omega}{1760} = 568\ \Omega$$

$$X_{C2} = \frac{1}{2\ \pi\ f\ C}$$

$$= \frac{1}{2 \times 3.1416 \times 28.0 \times 10^6\ Hz \times 20.0 \times 10^{-12}\ F}$$

$$= \frac{10^6\ \Omega}{3520} = 284\ \Omega$$

$$X_{total} = X_{C1} + X_{C2} = 568\ \Omega + 284\ \Omega = 852\ \Omega$$

Alternatively, combining the series capacitors first, the total capacitance is 6.67×10^{-12} F or 6.67 pF. Then:

$$X_{total} = \frac{1}{2\ \pi\ f\ C}$$

$$= \frac{1}{2 \times 3.1416 \times 28.0 \times 10^6\ Hz \times 6.67 \times 10^{-12}\ F}$$

$$= \frac{10^6\ \Omega}{1170} = 855\ \Omega$$

(Within the uncertainty of the measured values and the rounding of values in the calculations, this is the same result as the 852 Ω we obtained with the first method.)

This example serves to remind us that *series capacitance* is not calculated in the manner used by other series resistance and inductance, but *series capacitive reactance* does follow the simple addition formula.

For reactances of the same type in parallel, the general formula is:

$$X_{total} = \frac{1}{\frac{1}{X1} + \frac{1}{X2} + \frac{1}{X3} + ... + \frac{1}{X_n}}$$

or, for exactly two reactances in parallel

$$X_{total} = \frac{X1 \times X2}{X1 + X2}$$

Example: Place the capacitors in the last example (10.0 pF and 20.0 pF) in parallel in the 28.0 MHz circuit. What is the resultant reactance?

$$X_{total} = \frac{X1 \times X2}{X1 + X2}$$

$$= \frac{568\ \Omega \times 284\ \Omega}{568\ \Omega + 284\ \Omega} = 189\ \Omega$$

Alternatively, two capacitors in parallel can be combined by adding their capacitances.

$$C_{total} = C_1 + C_2 = 10.0\ pF + 20.0\ pF = 30\ pF$$

$$X_C = \frac{1}{2 \pi f C}$$

$$= \frac{1}{2 \times 3.1416 \times 28.0 \times 10^6 \text{ Hz} \times 30 \times 10^{-12} \text{ F}}$$

$$= \frac{10^6 \ \Omega}{5280} = 189 \ \Omega$$

Example: Place the series inductors above (4.0 μH each) in parallel in a 3.8-MHz circuit. What is the resultant reactance?

$$X_{total} = \frac{X_{L1} \times X_{L2}}{X_{L1} + X_{L2}}$$

$$= \frac{96 \ \Omega \times 96 \ \Omega}{96 \ \Omega + 96 \ \Omega} = 48 \ \Omega$$

Of course, a number (N) of equal reactances (or resistances) in parallel yields a reactance that is the value of one of them divided by N, or:

$$X_{total} = \frac{X}{N} = \frac{96 \ \Omega}{2} = 48 \ \Omega$$

All of these calculations apply only to reactances of the same type; that is, all capacitive or all inductive. Mixing types of reactances requires a different approach.

UNLIKE REACTANCES IN SERIES

When combining unlike reactances — that is, combinations of inductive and capacitive reactance — in series, it is necessary to take into account that the voltage-to-current phase relationships differ for the different types of reactance. **Figure 3.22** shows a series circuit with both types of reactance. Since the reactances are in series, the current must be the same in both. The voltage across each circuit element differs in phase, however. The voltage E_L *leads* the current by 90°, and the voltage E_C *lags* the current by 90°. Therefore, E_L and E_C have opposite polarities and cancel each other in whole or in part. The line E in Figure 3.22 approximates the resulting voltage, which is the *difference* between E_L and E_C.

Since for a constant current the reactance is directly proportional to the voltage, the net reactance is still the sum of the individual reactances. Because inductive reactance is considered to be positive and capacitive reactance negative, the resulting reactance can be either positive (inductive) or negative (capacitive) or even zero (no reactance).

$$X_{total} = X_L - X_C$$

The convention of using absolute values for the reactances and building the sense of positive and negative into the formula is the preferred method used by hams and will be used in all of the remaining formulas in this chapter. Nevertheless, before using any formulas that include reactance, determine whether this convention is followed before assuming that the absolute values are to be used.

Example: Using Figure 3.22 as a visual aid, let X_C = 20.0 Ω and X_L = 80.0 Ω. What is the resulting reactance?

$$X_{total} = X_L - X_C$$

$$= 80.0 \ \Omega - 20.0 \ \Omega = +60.0 \ \Omega$$

Since the result is a positive value, total reactance is inductive. Had the result been a negative number, the total reactance would have been capacitive.

When reactance types are mixed in a series circuit, the resulting reactance is always smaller than the larger of the two reactances. Likewise, the resulting voltage across the series combination of reactances is always smaller than the larger of the two voltages across individual reactances.

Every series circuit of mixed reactance types with more than two circuit elements can be reduced to this simple circuit by combining all the reactances into one inductive and one capacitive reactance. If the circuit has more than one capacitor or more than one inductor in the overall series string, first use the formulas given earlier to determine the total series inductance alone and the total series capacitance alone (or their respective reactances). Then combine the resulting single capacitive reactance and single inductive reactance as shown in this section.

UNLIKE REACTANCES IN PARALLEL

The situation of parallel reactances of mixed type appears in **Figure 3.23**. Since the elements are in parallel, the voltage is common to both reactive components. The current through the capacitor, I_C, *leads* the voltage by 90°, and the current through the inductor, I_L, *lags* the voltage by 90°. In this case, it is the currents that are 180° out of phase and thus cancel each other in whole or in part. The total current is the difference between the individual currents, as indicated by the line I in Figure 3.23.

Since reactance is the ratio of voltage to current, the total reactance in the circuit is:

$$X_{total} = \frac{E}{I_L - I_C}$$

In the drawing, I_C is larger than I_L, and the resulting differential current retains the phase of I_C. Therefore, the overall reactance, X_{total}, is capacitive in this case. The total reactance of the circuit will be smaller than the larger of the individual reactances, because the total current is smaller than the larger of the two individual currents.

In parallel circuits, reactance and current are inversely proportional to each other for a constant voltage and the equation for two reactances in parallel can be used, carrying the positive and negative signs:

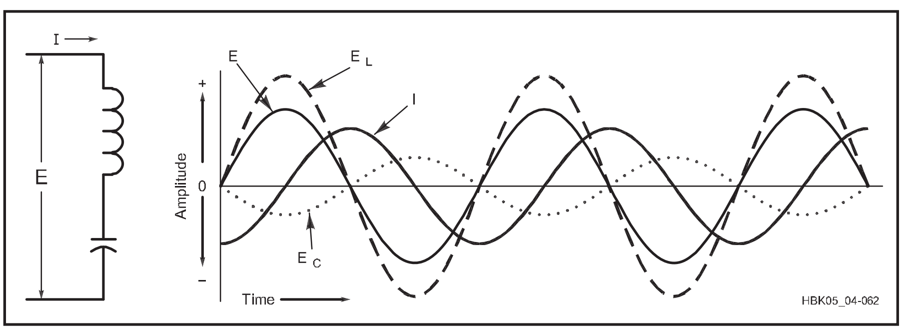

Figure 3.22 — A series circuit containing both inductive and capacitive components, together with representative voltage and current relationships.

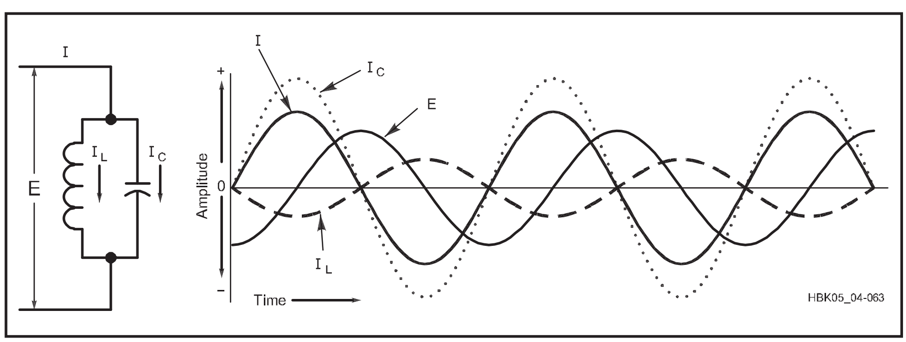

Figure 3.23 — A parallel circuit containing both inductive and capacitive components, together with representative voltage and current relationships.

$$X_{total} = \frac{X_L \times (-X_C)}{X_L - X_C} = \frac{-X_L \times X_C}{X_L - X_C}$$

As with the series formula for mixed reactances, follow the convention of using absolute values for the reactances, since the minus signs in the formula account for capacitive reactance being negative. If the solution yields a negative number, the resulting reactance is capacitive, and if the solution is positive, then the reactance is inductive.

Example: Using Figure 3.23 as a visual aid, place a capacitive reactance of 10.0 Ω in parallel with an inductive reactance of 40.0 Ω. What is the resulting reactance?

$$X_{total} = \frac{-X_L \times X_C}{X_L - X_C}$$

$$= \frac{-40.0\ \Omega \times 10.0\ \Omega}{40.0\ \Omega - 10.0\ \Omega}$$

$$= \frac{-400\ \Omega}{30.0\ \Omega} = -13.3\ \Omega$$

The reactance is capacitive, as indicated by the negative solution. Moreover, like resistances in parallel, the resultant reactance is always smaller than the larger of the two individual reactances.

As with the case of series reactances, if each leg of a parallel circuit contains more than one reactance, first simplify each leg to a single reactance. If the reactances are of the same type in each leg, the series reactance formulas for reactances of the same type will apply. If the reactances are of different types, then use the formulas shown above for mixed series reactances to simplify the leg to a single value and type of reactance.

3.5.3 Reactances At and Near Resonance

Any series or parallel circuit in which the values of the two unlike reactances are equal is said to be *resonant*. For any given inductance or capacitance, it is theoretically possible to find a value of the opposite reactance type to produce a resonant circuit for any desired frequency. This is discussed below in the section Resonant Circuits.

When a series circuit like the one shown in Figure 3.22 is resonant, the voltages E_C and E_L are equal and cancel; their sum is zero. This is a *series-resonant* circuit. Since the reactance of the circuit is proportional to the sum of these voltages, the net reactance also goes to zero. Theoretically, the current, as shown in **Figure 3.24**, can become infinite. In fact, it is limited only by losses in the components and other resistances that would exist in a real circuit of this type. As the frequency of operation moves slightly off resonance and the reactances no longer cancel completely, the net reactance climbs as shown in the figure. Similarly, away from resonance the current drops to a level determined by the net reactance.

In a *parallel-resonant* circuit of the type in Figure 3.23, the current I_L and I_C are equal and cancel to zero. Since the reactance is inversely proportional to the current, as the current approaches zero, the reactance becomes infinite. As with series circuits, component losses and other resistances in the circuit prevent the current from reaching zero. **Figure 3.25** shows the theoretical current curve near and at resonance for a purely reactive parallel-resonant circuit. Note that in both Figure 3.24 and Figure 3.25, the departure of current from the resonance value is close to, but not quite, symmetrical above and below the resonant frequency. This is discussed below in the section Resonant Circuits.

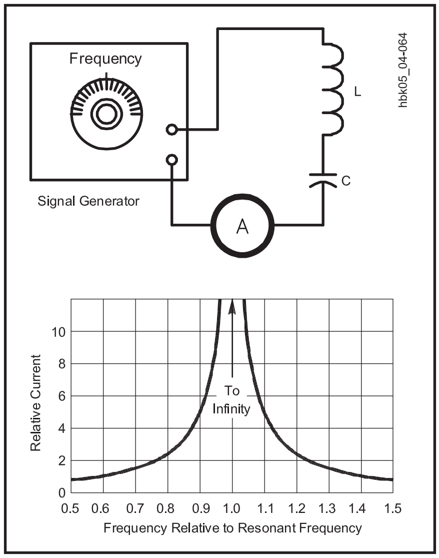

Figure 3.24 — The relative generator current with a fixed voltage in a series circuit containing inductive and capacitive reactances as the frequency approaches and departs from resonance.

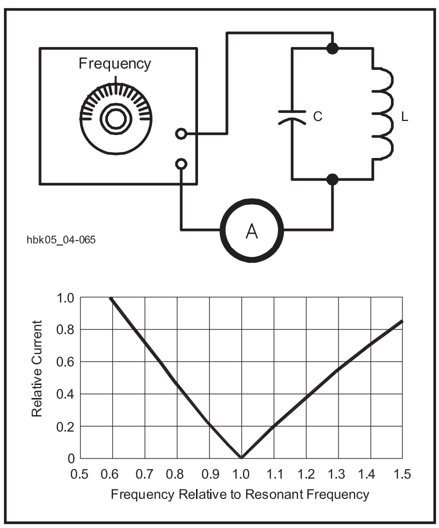

Figure 3.25 — The relative generator current with a fixed voltage in a parallel circuit containing inductive and capacitive reactances as the frequency approaches and departs from resonance. (The circulating current through the parallel inductor and capacitor is a maximum at resonance.)

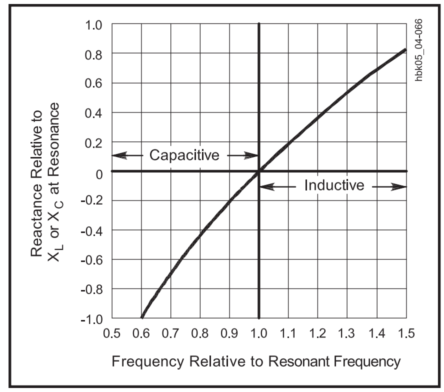

Figure 3.26 — The transition from capacitive to inductive reactance in a series-resonant circuit as the frequency passes resonance.

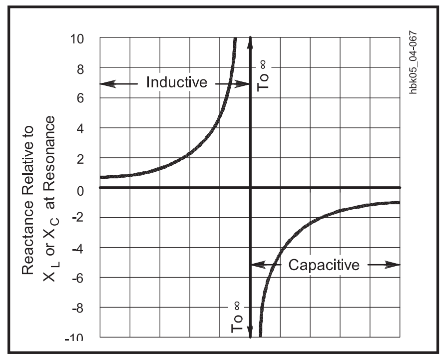

Figure 3.27 — The transition from inductive to capacitive reactance in a parallel-resonant circuit as the frequency passes resonance.

Example: What is the reactance of a series L-C circuit consisting of a 56.04-pF capacitor and an 8.967-µH inductor at 7.00, 7.10 and 7.20 MHz? Using the formulas from earlier in this chapter, we calculate a table of values:

Frequency (MHz)	X_L *(Ω)*	X_C *(Ω)*	X_{total} *(Ω)*
7.000	394.4	405.7	–11.3
7.100	400.0	400.0	0
7.200	405.7	394.4	11.3

The exercise shows the manner in which the reactance rises rapidly as the frequency moves above and below resonance. Note that in a series-resonant circuit, the reactance at frequencies below resonance is capacitive, and above resonance, it is inductive. **Figure 3.26** displays this fact graphically. In a parallel-resonant circuit, where the reactance becomes infinite at resonance, the opposite condition exists: above resonance, the reactance is capacitive and below resonance it is inductive, as shown in **Figure 3.27**. Of course, all graphs and calculations in this section are theoretical and presume a purely reactive circuit. Real circuits are never purely reactive; they contain some resistance that modifies their performance considerably. Real resonant circuits will be discussed later in this chapter.

3.5.4 Reactance and Complex Waveforms

All of the formulas and relationships shown in this section apply to alternating current in the form of regular sine waves. Complex wave shapes complicate the reactive situation considerably. A complex or nonsinusoidal wave can be treated as a sine wave of some fundamental frequency and a series of harmonic frequencies whose amplitudes depend on the original wave shape. When such a complex wave — or collection of sine waves — is applied to a reactive circuit, the current through the circuit will not have the same wave shape as the applied voltage. The difference results because the reactance of an inductor and capacitor depend in part on the applied frequency.

For the second-harmonic component of the complex wave, the reactance of the inductor is twice and the reactance of the capacitor is half their respective values at the fundamental frequency. A third-harmonic component produces inductive reactances that are triple and capacitive reactances that are one-third those at the fundamental frequency. Thus, the overall circuit reactance is different for each harmonic component.

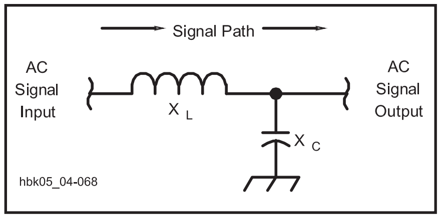

Figure 3.28 — A signal path with a series inductor and a shunt capacitor. The circuit presents different reactances to an ac signal and to its harmonics.

The frequency sensitivity of a reactive circuit to various components of a complex wave shape creates both difficulties and opportunities. On the one hand, calculating the circuit reactance in the presence of highly variable as well as complex waveforms, such as speech, is difficult at best. On the other hand, the frequency sensitivity of reactive components and circuits lays the foundation for filtering, that is, for separating signals of different frequencies or acting upon them differently. For example, suppose a coil is in the series path of a signal and a capacitor is connected from the signal line to ground, as represented in **Figure 3.28**. The reactance of the coil to the second harmonic of the signal will be twice that at the fundamental frequency and oppose more effectively the flow of harmonic current. Likewise, the reactance of the capacitor to the harmonic will be half that to the fundamental, allowing the harmonic an easier current path away from the signal line toward ground. The result is a low-pass filter that attenuates the harmonic more than the fundamental signal. (See the **Analog and Digital Filtering** chapter for detailed information on filter theory and construction.)

3.6 Impedance

When a circuit contains both resistance and reactance, the combined opposition to current is called *impedance*. Symbolized by the letter Z, impedance is a more general term than either resistance or reactance. Frequently, the term is used even for circuits containing only resistance or reactance. Qualifications such as "resistive impedance" are sometimes added to indicate that a circuit has only resistance, however.

The reactance and resistance comprising an impedance may be connected either in series or in parallel, as shown in **Figure 3.29**. In these circuits, the reactance is shown as a box to indicate that it may be either inductive or capacitive. In the series circuit at A, the current is the same in both elements, with (generally) different voltages appearing across the resistance and reactance. In the parallel circuit at B, the same voltage is applied to both elements, but different currents may flow in the two branches.

In a resistance, the current is in phase with the applied voltage, while in a reactance it is 90° out of phase with the voltage. Thus, the phase relationship between current and voltage in the circuit as a whole may be anything between zero and 90°, depending on the relative amounts of resistance and reactance.

As shown in Figure 3.21 in the preceding section, reactance is graphed on the vertical axis to record the phase difference between the voltage and the current. **Figure 3.30** adds resistance to the graph. Since the voltage is in phase with the current, resistance is recorded on the horizontal axis, using the positive or right side of the scale.

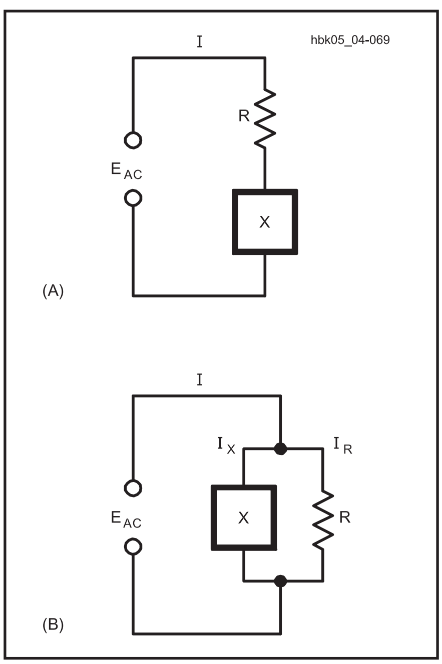

Figure 3.29 — Series and parallel circuits containing resistance and reactance.

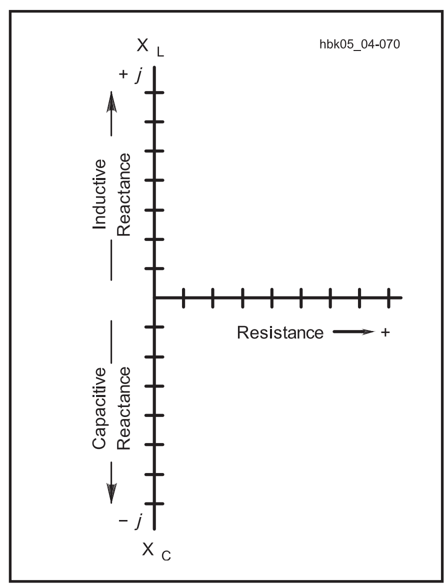

Figure 3.30 — The conventional method of charting impedances on a graph, using the vertical axis for reactance (the upward or "plus" direction for inductive reactance and the downward or "minus" direction for capacitive reactance), and using the horizontal axis for resistance.

3.6.1 Calculating Z from R and X in Series Circuits

Impedance is the complex combination of resistance and reactance. Since there is a 90° phase difference between resistance and reactance (whether inductive or capacitive), simply adding the two values does not correspond to what actually happens in a circuit and will not give the correct result. Therefore, expressions such as "Z = R + X" are incorrect because they show resistance and reactance being added directly. The correct expression is "Z = R + *j*X" showing that complex mathematics must be used. In pure mathematics, "*i*" indicates an imaginary number. Because *i* represents current in electronics, we use the letter "*j*" for the same mathematical operator, although there is nothing imaginary about what it represents in electronics. (References to explain imaginary numbers, rectangular coordinates, polar coordinates and how to work with them are provided in the "Radio Mathematics" article in this book's online content.) With respect to resistance and reactance, the letter *j* is normally assigned to those figures on the vertical axis, 90° out of phase with the horizontal axis. The presence of *j* means that impedance is a vector and calculating impedance from resistance and reactance involves *vector addition*.

As noted earlier, a vector is a value with both magnitude and direction, such as velocity; "10 meters/second to the north." Impedance also has a "direction" derived from the phase differences between voltage and current as described below. In vector addition, the result of combining two values with a phase difference is a quantity different from the simple *algebraic addition* of the two values. The result will have a phase difference intermediate between two vectors.

RECTANGULAR FORM OF IMPEDANCE

Because this form for impedances, Z = R ± *j*X, can be plotted on a graph using rectangular coordinates, this is the *rectangular form* of impedance. The rectangular coordinate system in which one axis represents real number and the other axis imaginary numbers is called the *complex plane* and impedance with both real (R) and imaginary (X) components is called *complex impedance*. Unless specifically noted otherwise, assume that "impedance" means "complex impedance" and that both R and X may be present.

Consider **Figure 3.31**, a series circuit consisting of an inductive reactance and a resistance. As given, the inductive reactance is 100 Ω and the resistance is 50 Ω. Using *rectangular coordinates*, the impedance becomes

$$Z = R + jX$$

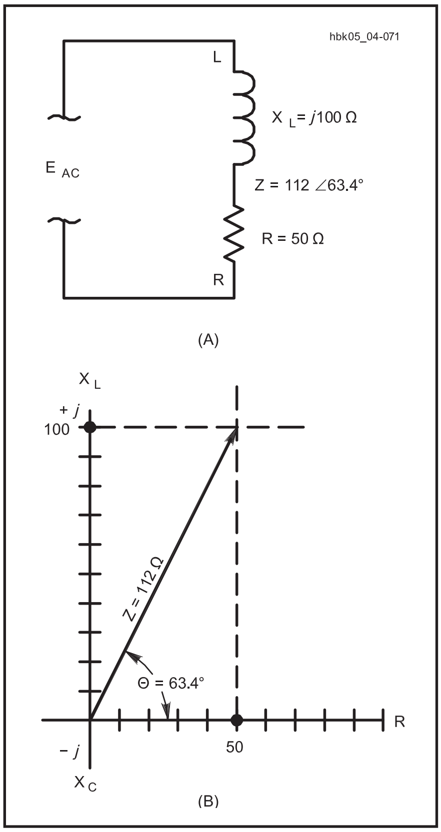

Figure 3.31 — A series circuit consisting of an inductive reactance of 100 Ω and a resistance of 50 Ω. At B, the graph plots the resistance, reactance, and impedance.

where
Z = the impedance in ohms,
R = the resistance in ohms, and
X = the reactance in ohms.

In the present example,

$$Z = 50 + j100\ \Omega$$

This point is located at the tip of the arrow drawn on the graph where the dashed lines cross.

POLAR FORM OF IMPEDANCE AND PHASE ANGLE

As the graph in Figure 3.31 shows, the impedance that results from combining R and X can also be represented by a line completing a right triangle whose sides are the resistance and reactance. The point at the end of the line — the complex impedance — can be described by how far it is from the origin of the graph where the axes cross (the *magnitude* of the impedance indicated by vertical bars around the variable, |Z|) and the angle made by the line with the horizontal axis representing 0° (the *phase angle* of the impedance, θ). This is the *polar form* of impedance and it is written in the form

$$Z = |Z| \angle\theta$$

Occasionally, θ may be given in radians. The convention in this handbook is to use degrees unless specifically noted otherwise.

The length of the hypotenuse of the right triangle represents the magnitude of the impedance and can be calculated using the formula for calculating the hypotenuse of a right triangle, in which the square of the hypotenuse equals the sum of the squares of the two sides:

$$|Z| = \sqrt{R^2 + X^2}$$

In this example:

$$|Z| = \sqrt{(50\ \Omega)^2 + (100\ \Omega)^2}$$

$$= \sqrt{2500\ \Omega^2 + 10000\ \Omega^2}$$

$$= \sqrt{12500\ \Omega^2} = 112\ \Omega$$

The magnitude of the impedance that results from combining 50 Ω of resistance with 100 Ω of inductive reactance is 112 Ω. From trigonometry, the tangent of the phase angle is the side opposite the angle (X) divided by the side adjacent to the angle (R), or

$$\tan\theta = \frac{X}{R}$$

where
X = the reactance, and
R = the resistance.

Find the angle by taking the inverse tangent, or arctan:

$$\theta = \arctan\frac{X}{R}$$

Calculators sometimes label the inverse tangent key as "$\tan^{-1}$". Remember to be sure your calculator is set to use the right angular units, either degrees or radians.

In the example shown in Figure 3.31,

$$\theta = \arctan\frac{100\ \Omega}{50\ \Omega} = \arctan 2.0 = 63.4°$$

Using the information just calculated, the complex impedance in polar form is:

$$Z = 112\ \Omega \angle 63.4°$$

This is stated verbally as "112 ohms at an angle of 63 point 4 degrees."

POLAR TO RECTANGULAR CONVERSION

The expressions R ± *j*X and |Z| ∠θ both provide the same information, but in two different forms. The procedure just given permits conversion from rectangular coordinates into polar coordinates. The reverse procedure is also important. **Figure 3.32** shows an impedance composed of a capacitive reactance and a resistance. Since capacitive reactance appears as a negative value, the impedance will be at a negative phase angle, in this case, 12.0 Ω at a phase angle of – 42.0° or Z = |12.0 Ω| ∠ – 42.0°.

Remember that the impedance forms a triangle with the values of X and R from the rectangular coordinates. The reactance axis forms the side opposite the angle θ.

$$\sin\theta = \frac{\text{side opposite}}{\text{hypotenuse}} = \frac{X}{|Z|}$$

Solving this equation for reactance, we have:

$$X = |Z| \times \sin\theta \text{ (ohms)}$$

Likewise, the resistance forms the side adjacent to the angle.

$$\cos\theta = \frac{\text{side adjacent}}{\text{hypotenuse}} = \frac{R}{|Z|}$$

Solving for resistance, we have:

$$R = |Z| \times \cos\theta \text{ (ohms)}$$

Then from our example:

X = 12.0 Ω × sin (–42°)
= 12.0 Ω × –0.669 = –8.03 Ω

R = 12.0 Ω × cos (–42°)
= 12.0 Ω × 0.743 = 8.92 Ω

Since X is a negative value, it is plotted on the lower vertical axis, as shown in Figure 3.32, indicating capacitive reactance. In rectangular form, Z = 8.92 Ω – *j*8.03 Ω.

In performing impedance and related calculations with complex circuits, rectangular coordinates are most useful when formulas require the addition or subtraction of values. Polar notation is most useful for multiplying and dividing complex numbers. (See the article "Radio Mathematics" in this book's online content for references dealing with the mathematics of complex numbers.)

All of the examples shown so far in this section presume a value of reactance that contributes to the circuit impedance. Reactance is a function of frequency, however, and many impedance calculations may begin with a value of capacitance or inductance and an operating frequency. In terms of these values, |Z| can be calculated in either of two ways, depending on whether the reactance is inductive or capacitive:

$$|Z| = \sqrt{R^2 + (2\pi f L)^2}$$

$$|Z| = \sqrt{R^2 + \left(\frac{1}{2\pi f C}\right)^2}$$

Example: What is the impedance of a circuit like Figure 3.31 with a resistance of 100 Ω and a 7.00-μH inductor operating at a frequency of 7.00 MHz? Using the equation appropriate for inductive reactance,

$$|Z| = \sqrt{R^2 + (2\pi f L)^2}$$

$$= \sqrt{(100\ \Omega)^2 + (2\pi \times 7.0 \times 10^{-6}\ \text{H} \times 7.0 \times 10^{6}\ \text{Hz})^2}$$

$$= \sqrt{10000\ \Omega^2 + (308\ \Omega)^2}$$

$$= \sqrt{10000\ \Omega^2 + 94900\ \Omega^2}$$

$$= \sqrt{104900\ \Omega^2} = 323.9\ \Omega$$

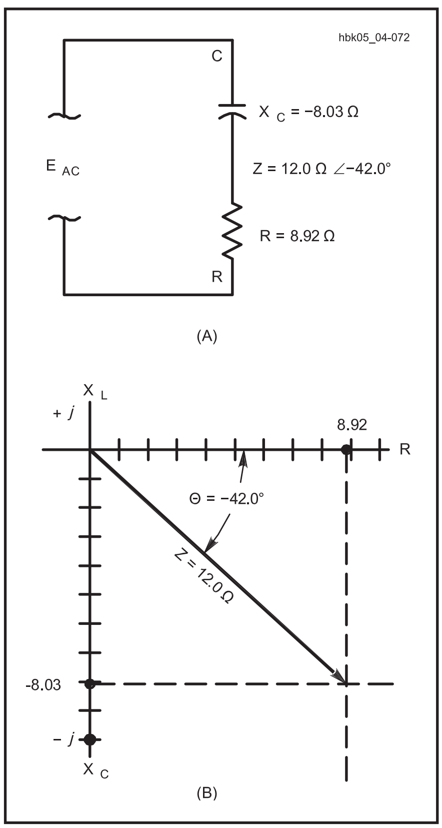

Figure 3.32 — A series circuit consisting of a capacitive reactance and a resistance: the impedance is given as 12.0 Ω at a phase angle θ of –42 degrees. At B, the graph plots the resistance, reactance, and impedance.

Since 308 Ω is the value of inductive reactance of the 7.00-μH coil at 7.00 MHz, the phase angle calculation proceeds as given in the earlier example:

$$\theta = \arctan\frac{X}{R} = \arctan\left(\frac{308.0\ \Omega}{100.0\ \Omega}\right)$$

$$= \arctan(3.08) = 72.0°$$

Since the reactance is inductive, the phase angle is positive.

3.6.2 Calculating Z from R and X in Parallel Circuits

In a parallel circuit containing reactance and resistance, such as shown in **Figure 3.33**, calculation of the resultant impedance from the values of R and X does not proceed by direct combination as for series circuits. The general formula for parallel circuits is:

$$|Z| = \frac{RX}{\sqrt{R^2 + X^2}}$$

where the formula uses the absolute (unsigned) reactance value. The phase angle for the parallel circuit is given by:

$$\theta = \arctan\left(\frac{R}{X}\right)$$

The sign of θ has the same meaning in both series and parallel circuits: if the parallel reactance is capacitive, then θ is a negative angle, and if the parallel reactance is inductive, then θ is a positive angle.

Example: A resistor of 30 Ω is in parallel with an inductor with a reactance of 40 Ω. What is the resulting impedance and phase angle?

$$|Z| = \frac{RX}{\sqrt{R^2 + X^2}} = \frac{30.0\ \Omega \times 40.0\ \Omega}{\sqrt{(30.0\ \Omega)^2 + (40.0\ \Omega)^2}}$$

$$= \frac{1200\ \Omega^2}{\sqrt{900\ \Omega^2 + 1600\ \Omega^2}} = \frac{1200\ \Omega^2}{\sqrt{2500\ \Omega^2}}$$

$$= \frac{1200\ \Omega^2}{50.0\ \Omega} = 24.0\ \Omega$$

$$\theta = \arctan\left(\frac{R}{X}\right) = \arctan\left(\frac{40.0\ \Omega}{30.0\ \Omega}\right)$$

$$\theta = \arctan(1.33) = 53.1°$$

In polar form,

$$Z = 24.0\angle 53.1°\ \Omega.$$

Since the parallel reactance is inductive, the resultant angle is positive.

Example: A capacitor with a reactance of 16.0 Ω is in parallel with a resistor of 12.0 Ω.

What is the resulting impedance and phase angle? (Remember that capacitive reactance is negative when used in calculations.)

$$|Z| = \frac{RX}{\sqrt{R^2 + X^2}} = \frac{12.0\ \Omega \times -16.0\ \Omega}{\sqrt{(12.0\ \Omega)^2 + (-16.0\ \Omega)^2}}$$

$$= \frac{-192\ \Omega^2}{\sqrt{144\ \Omega^2 + 256\ \Omega^2}} = \frac{-192\ \Omega^2}{\sqrt{400\ \Omega^2}}$$

$$= \frac{-192\ \Omega^2}{20.0\ \Omega} = -9.60\ \Omega$$

$$\theta = \arctan\left(\frac{R}{X}\right) = \arctan\left(\frac{12.0\ \Omega}{-16.0\ \Omega}\right)$$

$$\theta = \arctan(-0.750) = 36.9°$$

Because the parallel reactance is capacitive and the reactance negative, the resultant phase angle is negative.

3.6.3 Admittance

Impedance also has an inverse: *admittance* (*Y*), measured in siemens (S). Thus,

$$Y = 1 / Z$$

Since resistance, reactance and impedance are inversely proportional to the current (Z = E / I), conductance, susceptance and admittance are directly proportional to current. That is,

$$Y = I / E$$

Admittance can be expressed in rectangular and polar forms, just like impedance,

$$Y = G + jB = |Y| \angle\theta$$

The phase angle for admittance has the opposite sign convention as for impedance; if the susceptance component is inductive, the phase angle is negative, and if the susceptive component is capacitive, the phase angle is positive.

One handy use for admittance is in simplifying parallel circuit impedance calculations. Similar to series combinations, the admittance of a parallel combination of reactance and resistance is the vector addition of susceptance and conductance. In other words, for parallel circuits:

$$|Y| = \sqrt{G^2 + B^2}$$

where

|Y| = magnitude of the admittance in siemens,

G = conductance or 1 / R in siemens, and

B = susceptance or –1 / X in siemens.

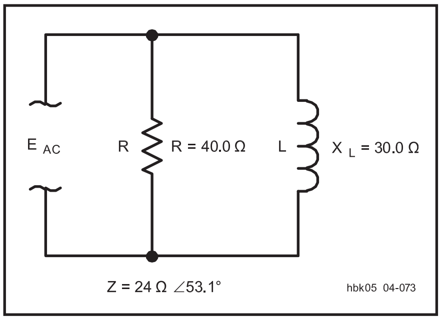

Figure 3.33 — A parallel circuit containing an inductive reactance of 30.0 Ω and a resistor of 40.0 Ω. No graph is given, since parallel impedances cannot be manipulated graphically in the simple way of series impedances.

Repeating the previous example for an inductor with a reactance of 30.0 Ω in parallel with a resistor of 40.0 Ω. Calculate the resulting impedance and phase angle by using admittance.

The inductor's susceptance is –1 / 30.0 Ω = –0.0333 S and the resistor's conductance is 1 / 40.0 Ω = 0.0250 S.

$$|Y| = \sqrt{(0.0250\,S)^2 + (-0.0333\,S)^2}$$

$$= \sqrt{0.00173\ S^2} = 0.0417\ S$$

The admittance's phase angle in terms of conductance and susceptance is:

$$\theta = \arctan\left(\frac{B}{G}\right)$$

$$\theta = \arctan\left(\frac{-0.0333\ S}{0.0250\ S}\right) = \arctan(-1.33) = -53.1°$$

In polar form,

$$Y = 0.0417\angle -53.1°$$

Converting back to impedance:

$$|Z| = \frac{1}{|Y|} = \frac{1}{0.0417\ S} = 24.0\ \Omega$$

The impedance's phase angle is the negative of the admittance's phase angle, so:

$$\angle Z = -\angle Y = \angle 53.1°$$

and in polar form

$$Z = 24.0\angle 53.1°\ \Omega$$

Conversion between resistance, impedance and admittance is very useful in working with complex circuits and in impedance matching of antennas and transmission lines. There are many on-line calculators that can perform these operations and many programmable calculators and suites of mathematical computer software have these functions built-in. Knowing when and how to use them, however, demands some understanding of the fundamental strategies shown here.

3.6.4 More than Two Elements in Series or Parallel

When a circuit contains several resistances or several reactances in series, simplify the circuit before attempting to calculate the impedance. Resistances in series add, just as in a purely resistive circuit. Series reactances of the same kind — that is, all capacitive or all inductive — also add, just as in a purely reactive circuit. The goal is to produce a single value of resistance and a single value of reactance that can be used in the impedance calculation.

Figure 3.34A illustrates a case in which a circuit consists of two different types of reac-

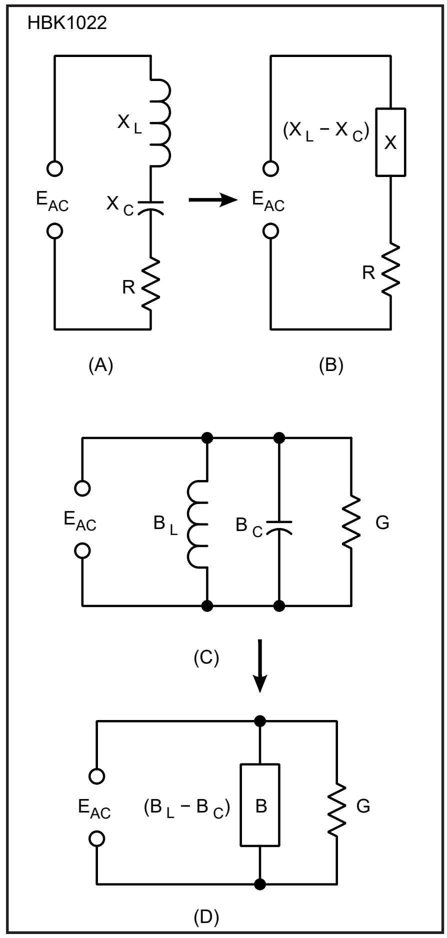

Figure 3.34 — A series circuit (A) containing both capacitive and inductive reactances can be simplified to a circuit with a single reactance (B) by adding the reactance values together. A parallel circuit (C) is simplified using the formulas in the previous section. Another method (D) is to convert the parallel reactances to susceptances, adding the values, then converting the resulting susceptance back to reactance.

tance in series, along with series resistance. The series combination of X_C and X_L reduce to a single value using the same rules of combination discussed in the section on purely reactive components. As Figure **3.34B** shows, the resulting reactance is the difference of the two reactances in series; $X_{TOTAL} = X_L - X_C$.

For parallel circuits with multiple resistances or multiple reactances of the same type, use the rules of parallel combination to reduce the resistive and reactive components to single elements. Where two or more reactive components of different types appear in the same circuit, they can be combined using formulas shown earlier for pure reactances. As **Figure 3.34C** suggests, however, they can also be combined as susceptances. Parallel susceptances of different types add, with attention to their differing signs. The resulting single susceptance can then be combined with the conductance to arrive at the overall circuit admittance whose inverse is the final circuit impedance.

3.6.5 Equivalent Series and Parallel Circuits

The two circuits shown in Figure 3.29 are equivalent if the same current flows when a given voltage of the same frequency is applied, and if the phase angle between voltage and current is the same in both cases. It is possible, in fact, to transform any given series circuit into an equivalent parallel circuit, and vice versa.

Using the conventions shown in **Figure 3.35**, a series RX circuit can be converted into its parallel equivalent by means of the formulas:

$$R_P = \frac{R_S^2 + X_S^2}{R_S}$$

$$X_P = \frac{R_S^2 + X_S^2}{X_S}$$

where the subscripts P and S represent the parallel- and series-equivalent values, respectively. If the parallel values are known, the equivalent series circuit can be found from:

$$R_S = \frac{R_P\, X_P^2}{R_P^2 + X_P^2}$$

and

$$X_S = \frac{R_P^2\, X_P}{R_P^2 + X_P^2}$$

Example: Let the series circuit in Figure 3.29 have a series reactance of –50.0 Ω (indicating a capacitive reactance) and a resistance of 50.0 Ω. What are the values of the equivalent parallel circuit?

$$R_P = \frac{R_S^2 + X_S^2}{R_S} = \frac{(50.0\ \Omega)^2 + (-50.0\ \Omega)^2}{50.0\ \Omega}$$

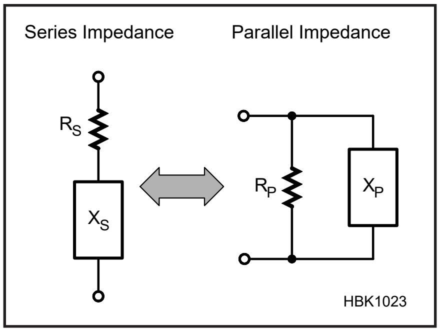

Figure 3.35 — Circuits containing reactance and resistance can be transformed into their series and parallel equivalents using the formulas in this section. Remember to use positive values for inductive reactance (X_L) and negative for capacitive reactance (X_C).

$$= \frac{2500\ \Omega^2 + 2500\ \Omega^2}{50.0\ \Omega} = \frac{5000\ \Omega^2}{50\ \Omega} = 100\ \Omega$$

$$X_P = \frac{R_S^2 + X_S^2}{X_S} = \frac{(50.0\ \Omega)^2 + (-50.0\ \Omega)^2}{-50.0\ \Omega}$$

$$= \frac{2500\ \Omega^2 + 2500\ \Omega^2}{-50.0\ \Omega} = \frac{5000\ \Omega^2}{-50\ \Omega} = -100\ \Omega$$

A capacitive reactance of 100 Ω in parallel with a resistance of 100 Ω is the equivalent circuit to the series circuit.

3.6.6 Ohm's Law for Impedance

Ohm's Law applies to circuits containing impedance just as readily as to circuits having resistance or reactance only. The formulas are:

$E = I\, Z$

$I = E / Z$

$Z = E / I$

where

E = voltage in volts,

I = current in amperes, and

Z = impedance in ohms.

Z must now be understood to be a complex number, consisting of resistive and reactive components. If Z is complex, then so are E and I, with a magnitude and phase angle. The rules of complex mathematics are then applied and the variables are written in boldface type as **Z**, **E**, and **I**, or an arrow is added above them to indicate that they are complex, such as,

$$\vec{E} = \vec{I}\, \vec{Z}$$

If only the magnitude of impedance, voltage, and currents are important, however, then the magnitudes of the three variables can be combined in the familiar ways without regard to the phase angle. In this case E and I are assumed to be RMS values (or some other steady-state value such as peak, peak-to-peak, or average). **Figure 3.36** shows a simple circuit consisting of a resistance of 75.0 Ω and a reactance of 100 Ω in series. From the series-impedance formula previously given, the impedance is

$$Z = \sqrt{R^2 + X_L^2} = \sqrt{(75.0\ \Omega)^2 + (100\ \Omega)^2}$$

$$= \sqrt{5630\ \Omega^2 + 10000\ \Omega^2}$$

$$= \sqrt{15600\ \Omega^2} = 125\ \Omega$$

If the applied voltage is 250 V, then

$$I = \frac{E}{Z} = \frac{250\ V}{125\ \Omega} = 2.0\ A$$

This current flows through both the resistance and reactance, so the voltage drops are:

$$E_R = I\, R = 2.0\ A \times 75.0\ \Omega = 150\ V$$

$$E_{XL} = I\, X_L = 2.0\ A \times 100\ \Omega = 200\ V$$

AC Component Summary

	Resistor	*Capacitor*	*Inductor*
Basic Unit	ohm (Ω)	farad (F)	henry (H)
Units Commonly Used		microfarads (μF) picofarads (pF)	millihenrys (mH) microhenrys (μH)
Time constant	(None)	RC	L/R
Voltage-Current Phase	In phase	Current leads voltage Voltage lags current	Voltage leads current Current lags voltage
Resistance or Reactance	Resistance	$X_C = 1 / 2\pi fC$	$X_L = 2\pi fL$
Change with increasing frequency	No	Reactance decreases	Reactance increases
Q of circuit	Not defined	X_C / R	X_L / R

Illustrating one problem of working only with RMS values, the simple arithmetical sum of these two drops, 350 V, is greater than the applied voltage because the two voltages are 90° out of phase. When phase is taken into account,

$$E = \sqrt{(150\ V)^2 + (200\ V)^2}$$

$$= \sqrt{22500\ V^2 + 40000\ V^2}$$

$$= \sqrt{62500\ V^2} = 250\ V$$

3.6.7 Reactive Power and Power Factor

The charge placed on a capacitor during part of an ac cycle is returned to the circuit during the next part of a cycle. Likewise, the energy stored in the magnetic field of an inductor returns to the circuit later in the ac cycle. A reactive circuit simply cycles and recycles energy into and out of the reactive components. If a purely reactive circuit were possible in reality, it would consume no energy at all.

In reactive circuits, circulation of energy accounts for seemingly odd phenomena. For example, in a series circuit with capacitance and inductance, the voltages across the components may exceed the supply voltage. That condition can exist because, while energy is being stored by the inductor, the capacitor is returning energy to the circuit from its previously charged state, and vice versa. In a parallel circuit with inductive and capacitive branches, the current circulating through the components may exceed the current drawn from the source. Again, the phenomenon occurs because the inductor's collapsing magnetic field supplies current to the capacitor, and the discharging capacitor provides current to the inductor.

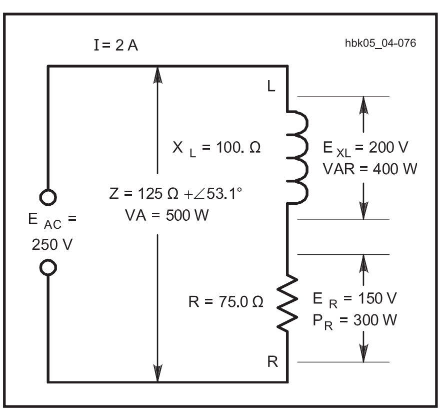

Figure 3.36 — A series circuit consisting of an inductive reactance of 100 Ω and a resistance of 75.0 Ω. Also shown is the applied voltage, voltage drops across the circuit elements, and the current.

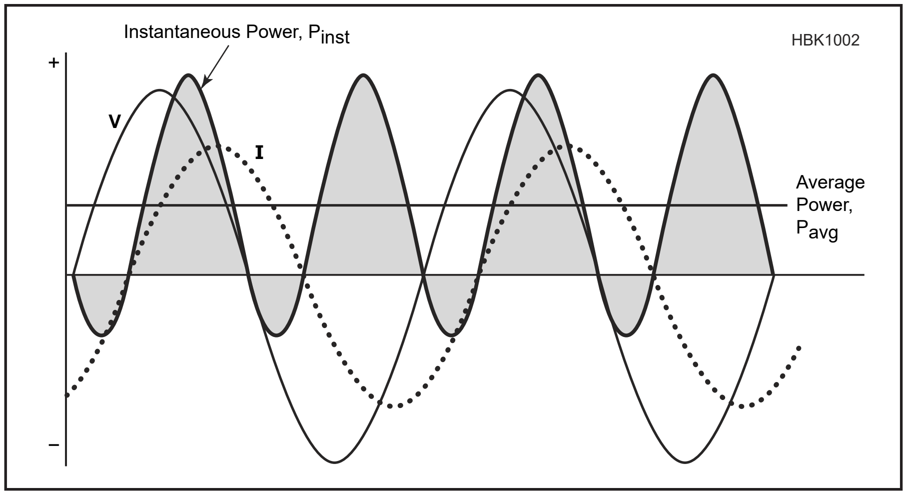

Figure 3.37 — Instantaneous power is the product of instantaneous voltage and current but may be positive or negative depending on the relative phase of voltage or current. This reduces average power which is calculated over the entire cycle.

Reactance creates a phase difference between voltage and current as shown in **Figure 3.37**, which shows current lagging behind voltage by about 60°. Instantaneous power is still calculated as the product of instantaneous voltage and current, but notice that it is negative for part of the cycle when voltage is positive and current negative, or vice versa. This causes average power dissipated in the resistance over the entire cycle to be lower than when voltage and current are in-phase. If current and voltage are 90° out of phase as in a purely reactive circuit, average power is reduced to zero. So, although reactive circuits show a measurable RMS value for ac voltage and current, we cannot simply multiply the values together as for the purely resistive circuit of Figure 3.14B.

To distinguish between the non-dissipated energy circulating in a purely reactive circuit and the dissipated or *real power* in a resistive circuit, the unit of *reactive power* is called the *volt-ampere reactive*, or *VAR*. The term watt is not used and sometimes reactive power is called "wattless" power. VAR has only limited use in radio circuits. Formulas similar to those for resistive power are used to calculate VAR:

$$VAR = I \times E$$

$$VAR = I^2 \times X$$

$$VAR = E^2 / X$$

where E and I are RMS values of voltage and current.

Real, or dissipated, power is measured in watts. *Apparent power* is the product of the voltage across and the current through an impedance. To distinguish apparent power from real power, apparent power is measured in *volt-amperes* (*VA*).

In the circuit of Figure 3.36, an applied voltage of 250 V results in a current of 2.00 A, giving an apparent power of 250 V × 2.00 A = 500 W. Only the resistance actually consumes power, however. The real power dissipated by the resistance is:

$$E = I^2\ R = (2.0\ A)^2 \times 75.0\ V = 300\ W$$

and the reactive power is:

$$VAR = I^2 \times X_L = (2.0\ A)^2 \times 100\ \Omega = 400\ VA$$

The ratio of real power to the apparent power is called the circuit's *power factor* (PF).

$$PF = \frac{P_{consumed}}{P_{apparent}} = \frac{R}{Z}$$

Power factor is frequently expressed as a percentage. The power factor of a purely resistive circuit is 100% or 1, while the power factor of a pure reactance is zero. In the example of Figure 3.36 the power factor would be 300 W / 500 W = 0.600 or 60%.

Apparent power has no direct relationship to the power actually dissipated unless the power factor of the circuit is known.

$$P = \text{Apparent Power} \times \text{Power Factor}$$

An equivalent definition of power factor is:

$$PF = \cos \theta$$

where θ is the phase angle of the circuit impedance.

Since the phase angle in the example equals:

$$\theta = \arctan\left(\frac{X}{R}\right) = \arctan\left(\frac{100\ \Omega}{75.0\ \Omega}\right)$$

$$\theta = \arctan(1.33) = 53.1°$$

and the power factor is:

$$PF = \cos 53.1° = 0.600$$

as the earlier calculation confirms.

Since power factor is always rendered as a positive number, the value must be followed by the words "leading" or "lagging" to identify the phase of the voltage with respect to the current. Specifying the numerical power factor is not always sufficient. For example, many dc-to-ac power inverters can safely operate loads having a large net reactance of one sign but only a small reactance of the opposite sign. Hence, the final calculation of the power factor in this example would be reported as "0.600, leading."

3.7 Quality Factor (Q) of Components

Components that store energy, such as capacitors and inductors, may be compared in terms of *quality factor* or *Q factor*, abbreviated *Q*. The concept of *Q* originated in 1914 (then dubbed *K*) and first appeared in print in 1923 when Kenneth S. Johnson used it to represent the ratio of reactance to resistance as a "figure of merit" for inductors in US patent 1,628,983. For a series or parallel representation of a reactive circuit element:

$$Q = \frac{X_S}{R_S} = \frac{R_P}{X_P}$$

where for series-connected reactance and its series loss resistance (such as an inductor)

Q = quality factor (no units),
X_S = series reactance of the component (in ohms), and
R_S = the sum of all series resistances associated with the energy losses in the component (in ohms).

For a parallel connected reactance and its parallel loss resistance (such as a capacitor)

Q = quality factor (no units),
X_P = parallel-connected reactance of the component (in ohms), and
R_P = the total parallel resistance associated with the energy losses in the component (in ohms).

Table 3.2
Equivalent Formulas for Expressing Q and Their Uses

[a]	$Q = \frac{\text{Series Reactance}}{\text{Series Resistance}}$	Johnson's historical definition of Q for inductors, used for series circuits.
[b]	$Q = \frac{\text{Parallel Resistance}}{\text{Parallel Reactance}}$	Parallel equivalent circuit definition of Q, useful for capacitors.
[c]	$Q = \frac{2\pi \times \text{Stored energy}}{\text{Energy lost in one cycle}}$	Fundamental energy definition, useful with antennas, reactive components, and mechanical systems.
[d]	$Q = \frac{\sqrt{f_U f_L}}{f_U - f_L} = \frac{\text{Frequency}}{\text{Bandwidth}}$	Bandwidth formula for simple resonant circuits. Impedance Z = R + *j*X, and f_U is the upper frequency where R = X, and f_L is the lower frequency where R = –X, and $f_U - f_L$ represents the –3 dB bandwidth.

Several exactly equivalent formulas for Q may be seen in **Table 3.2**. In Table 3.2, equation [a] most naturally represents the Q of an inductor, while equation [b] is useful for a capacitor. Both representations are equivalent to equation [c] which relates the energy storage to energy losses in inductors and capacitors. Note that in a series circuit representation, the series resistance is proportional to energy loss, and the series reactance is proportional to stored energy. In a parallel circuit, however, the reciprocal of the resistance is proportional to the lost energy and the reciprocal of the reactance is proportional to the stored energy. See the referenced paper by Ohira in the References section for a theoretical discussion of Q as applied to different circuit characteristics.

The Q of a tuned circuit may be found by measuring the upper and lower frequencies where the resistance equals the magnitude of the reactance, and applying equation [d]. The geometrical mean frequency is

$$f = \sqrt{f_U f_L}$$

and may be replaced by the center frequency for high-Q circuits. In circuits having several reactive components, such as the tuned circuits in Figure 3.42 later in this chapter, the circuit Q is the parallel combination of the individual Q factors. For example:

$$Q = 1 / \left(\frac{1}{Q_C} + \frac{1}{Q_L}\right)$$

where Q_C is the capacitor Q (sometimes specified by a manufacturer) and the inductor Q is Q_L.

The Q of capacitors is ordinarily high. Good quality ceramic capacitors and mica capacitors may have Q values of 1,200 or more. Microwave capacitors can have poor Q values — 10 or less at 10 GHz and higher frequencies because X_C will be low. Capacitors are subject to predominantly dielectric losses which are modeled as a parallel loss resistance across the capacitive reactance. Capacitors also have a series loss resistance associated with the conductor leads and capacitor plates, but this loss is often small enough to ignore.

Inductors are subject to several types of electrical energy losses such as wire resistance (including skin effect) and core losses. All electrical conductors have some resistance through which electrical energy is lost as heat. Wire conductors suffer additional ac losses because alternating current tends to flow on the conductor surface due to the skin effect discussed in the chapter on **RF Techniques**. If the inductor's core is iron, ferrite, or brass, the core will introduce additional losses of energy. Note that core losses for inductors are modeled as a resistor in parallel with the inductor (analogous to capacitor dielectric losses). The specific details of these losses are discussed in connection with each type of core material.

The sum of all core losses may be depicted by showing an equivalent series connected resistor with the inductor (see the section on Practical Inductors in the **Circuits and Components** chapter), although there is no separate component represented by the resistor symbol. As a result of inherent energy losses, inductor Q rarely approaches capacitor Q in a circuit where both components work together. Although many circuits call for the highest Q inductor obtainable, other circuits may call for a specific Q, even a very low one.

Q is also discussed in the following section on Resonant Circuits. For these circuits, Q is closely related to circuit bandwidth and selectivity. In addition, the circuit Q depends on whether or not a load is attached.

3.8 Resonant Circuits

A circuit containing both an inductor and a capacitor — and therefore, both inductive and capacitive reactance — is often called a *tuned circuit* or a *resonant circuit*. For any such circuit, there is a particular frequency at which the inductive and capacitive reactances are the same, that is, $X_L = X_C$. For most purposes, this is the *resonant frequency* of the circuit. At the resonant frequency — or at *resonance*, for short:

$$X_L = 2 \pi f L = X_C = \frac{1}{2 \pi f C}$$

By solving for f, we can find the resonant frequency of any combination of inductor and capacitor from the formula:

$$f = \frac{1}{2 \pi \sqrt{L C}}$$

where
f = frequency in hertz (Hz),
L = inductance in henrys (H),
C = capacitance in farads (F), and
π = 3.1416.

For most high-frequency (HF) radio work, smaller units of inductance and capacitance and larger units of frequency are more convenient. The basic formula becomes:

$$f = \frac{10^3}{2 \pi \sqrt{L C}}$$

where
f = frequency in megahertz (MHz),
L = inductance in microhenrys (μH),
C = capacitance in picofarads (pF), and
π = 3.1416.

Example: What is the resonant frequency of a circuit containing an inductor of 5.0 μH and a capacitor of 35 pF?

$$f = \frac{10^3}{2 \pi \sqrt{L C}} = \frac{10^3}{6.2832 \sqrt{5.0 \times 35}}$$

$$= \frac{10^3}{83} = 12 \text{ MHz}$$

To find the matching component (inductor or capacitor) when the frequency and one component is known (capacitor or inductor) for general HF work, use the formula:

$$f^2 = \frac{10^6}{4 \pi^2 L C}$$

where f, L, and C are in Hz, μH, and pF. To find L or C, the formula is converted to:

$$L = \frac{25{,}330}{f^2 C}$$

$$C = \frac{25{,}330}{f^2 L}$$

For most radio work, these formulas will permit calculations of frequency and component values well within the limits of component tolerances.

Example: What value of capacitance is needed to create a resonant circuit at 21.1 MHz, if the inductor is 2.00 μH?

$$C = \frac{25{,}330}{f^2 L} = \frac{25{,}330}{(21.1^2 \times 2.0)}$$

$$= \frac{25{,}330}{890} = 28.5 \text{ pF}$$

Figure 3.20 can also be used if an approximate answer is acceptable. From the horizontal axis, find the vertical line closest to the desired resonant frequency. Every pair of diagonals that cross on that vertical line represent a combination of inductance and capacitance that will resonate at that frequency. For example, if the desired frequency is 10 MHz, the pair of diagonals representing 5 μH and 50 pF cross quite close to that frequency. Interpolating between the given diagonals will provide more resolution — remember that all three sets of lines are spaced logarithmically.

Resonant circuits have other properties of importance, in addition to the resonant frequency, however. These include impedance, voltage drop across components in series-resonant circuits, circulating current in parallel-resonant circuits, and bandwidth. These properties determine such factors as the selectivity of a tuned circuit and the component ratings for circuits handling significant amounts of power. Although the basic determination of the tuned-circuit resonant frequency ignored any resistance in the circuit, that resistance will play a vital role in the circuit's other characteristics.

3.8.1 Series-Resonant Circuits

Figure 3.38 presents a basic schematic diagram of a *series-resonant circuit*. Although most schematic diagrams of radio circuits would show only the inductor and the capacitor, resistance is always present in such circuits. The most notable resistance is associated with the series resistance losses in the inductor at HF. The dominant losses in the capacitor may be modeled as a parallel resistance (not shown), but these losses are low enough at HF to be ignored. The current meter shown in the circuit is a reminder that in series circuits, the same current flows through all elements.

At resonance, the reactance of the capacitor cancels the reactance of the inductor. The voltage and current are in phase with each other, and the impedance of the circuit is determined solely by the resistance. The actual current through the circuit at resonance, and for frequencies near resonance, is determined by the formula:

$$I = \frac{E}{Z} = \frac{E}{\sqrt{R^2 + \left[2\,\pi\,f\,L - \frac{1}{(2\,\pi\,f\,C)}\right]^2}}$$

where all values are in basic units.

At resonance, the reactive factor in the formula is zero (the bracketed expression under the square root symbol). As the frequency is shifted above or below the resonant frequency without altering component values, however, the reactive factor becomes significant, and the value of the current becomes smaller than at resonance. At frequencies far from resonance, the reactive components become dominant, and the resistance no longer significantly affects the current amplitude.

The exact curve created by recording the current as the frequency changes depends on the ratio of reactance to resistance. When the reactance of either the coil or capacitor is of the same order of magnitude as the resistance, the current decreases rather slowly as the frequency is moved in either direction away from resonance. Such a curve is said to be *broad*. Conversely, when the reactance is considerably larger than the resistance, the current decreases rapidly as the frequency moves away from resonance, and the circuit is said to be *sharp*. A sharp circuit will respond a great deal more readily to the resonant frequency than to frequencies quite close to resonance; a broad circuit will respond almost equally well to a group or band of frequencies centered around the resonant frequency.

Both types of resonance curves are useful. A sharp circuit gives good selectivity — the ability to respond strongly (in terms of current amplitude) at one desired frequency and to discriminate against others. A broad circuit is used when the apparatus must give about the same response over a band of frequencies, rather than at a single frequency alone.

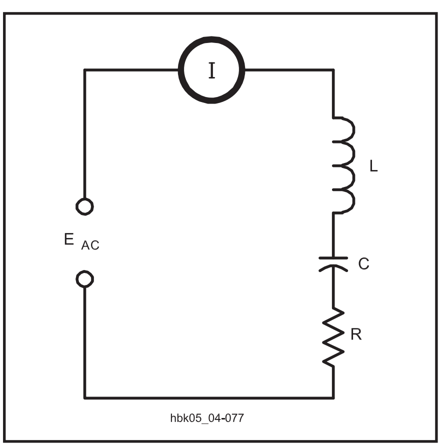

Figure 3.38 — A series circuit containing L, C, and R is resonant at the applied frequency when the reactance of C is equal to the reactance of L. The I in the circle is the schematic symbol for an ammeter.

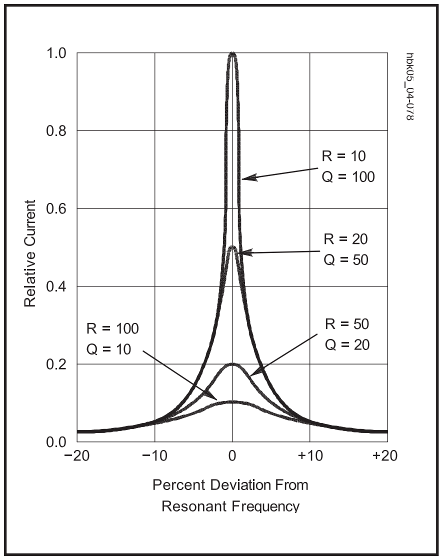

Figure 3.39A — Relative current in series-resonant circuits with various values of series resistance and Q. (An arbitrary maximum value of 1.0 represents current at resonance.) The reactance at resonance for all curves is 1000 Ω. Note that the current is hardly affected by the resistance in the circuit at frequencies more than 10% away from the resonant frequency.

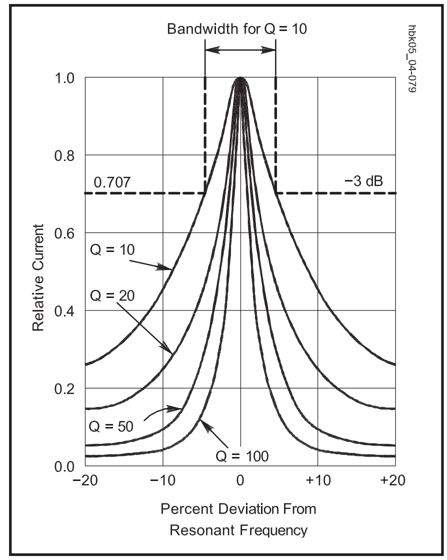

Figure 3.39B — Relative current in series-resonant circuits having different values of Q_U. The current at resonance is normalized to the same level for all curves in order to show the rate of change of decrease in current for each value of Q_U. The half-power points are shown to indicate relative bandwidth of the response for each curve. The bandwidth is indicated for a circuit with a Q_U of 10.

Figure 3.39A presents a family of curves, showing the decrease in current as the frequency deviates from resonance. In each case, the inductive and capacitive reactances are assumed to be 1,000 Ω. The maximum current, shown as a relative value on the graph, occurs with the lowest resistance, while the lowest peak current occurs with the highest resistance. Equally important, the rate at which the current decreases from its maximum value also changes with the ratio of reactance to resistance. It decreases most rapidly when the ratio is high and most slowly when the ratio is low.

UNLOADED Q

As noted in equation [a] of Table 3.2 earlier in this chapter, Q is the ratio of series reactance representing 2π times the stored energy (equation [c] in Table 3.2) to series resistance or consumed energy. Since both terms of the ratio are measured in ohms, Q has no units and is known as the *quality factor* (and less frequently, the *figure of merit* or the *multiplying factor*). The series resistive losses of the coil often dominate the energy consumption in HF series-resonant circuits, so the inductor Q largely determines the resonant-circuit Q. Since this value of Q is independent of any external load to which the circuit might transfer power, it is called the *unloaded Q* or Q_U of the circuit.

Example: What is the unloaded Q of a series-resonant circuit with a series loss resistance of 5 Ω and inductive and capacitive components having a reactance of 500 Ω each? With a reactance of 50 Ω each?

$$Q_{U1} = \frac{X1}{R} = \frac{500\,\Omega}{5\,\Omega} = 100$$

$$Q_{U2} = \frac{X2}{R} = \frac{50\,\Omega}{5\,\Omega} = 10$$

BANDWIDTH

Figure 3.39B is an alternative way of drawing the family of curves that relate current to frequency for a series-resonant circuit. By assuming that the peak current of each curve is the same, the rate of change of current for various values of Q_U and the associated ratios of reactance to resistance are more easily compared. From the curves, it is evident that the lower Q_U circuits pass current across a greater *bandwidth* of frequencies than the circuits with a higher Q_U. For the purpose of comparing tuned circuits, bandwidth is often defined as the frequency spread between the two frequencies at which the current amplitude decreases to 0.707 (or $\sqrt{2}\,/\,2$) times the maximum value. Since the power consumed by the resistance, R, is proportional to the square of the current, the power at these points is half the maximum power at resonance, assuming that R is constant for the calculations. The half-power, or –3 dB, points are marked on Figure 3.39B.

For Q values of 10 or greater, the curves shown in Figure 3.39B are approximately symmetrical. On this assumption, bandwidth (BW) can be easily calculated by inverting equation [d] in Table 3.2, and approximating the geometrical mean –3 dB frequency by f:

$$BW = \frac{f}{Q_U}$$

where BW and f are in the same units, that is, in Hz, kHz or MHz.

Example: What is the 3 dB bandwidth of a series-resonant circuit operating at 14 MHz with a Q_U of 100?

$$BW = \frac{f}{Q_U} = \frac{14\ \text{MHz}}{100} = 0.14\ \text{MHz} = 140\ \text{kHz}$$

The relationship between Q_U, f and BW provides a means of determining the value of circuit Q when inductor losses may be difficult to measure. By constructing the series-resonant circuit and measuring the current as the frequency varies above and below resonance, the half-power points can be determined. Then:

$$Q_U = \frac{f}{BW}$$

Example: What is the Q_U of a series-resonant circuit operating at 3.75 MHz, if the –3 dB bandwidth is 375 kHz?

$$Q_U = \frac{f}{BW} = \frac{3.75\ \text{MHz}}{0.375\ \text{MHz}} = 10.0$$

If the loss resistance of the inductor is much greater than of the capacitor (the usual case), BW is approximately R/L. The Q of a series resonant circuit can also be stated

$$Q_U = \frac{f_0}{BW} = \frac{1}{R}\sqrt{\frac{L}{C}}$$

The illustrations the relationship between Q and bandwidth at **hyperphysics.phy-astr.gsu.edu/hbase/electric/serres.html** are helpful in understanding the concept. **Table 3.3** provides some simple formulas for estimating the maximum current and phase angle for various bandwidths, if both f and Q_U are known.

Table 3.3
The Selectivity of Resonant Circuits

Approximate percentage of current at resonance[1] or of impedance at resonance[2]	*Bandwidth (between half-power or –3 dB points on response curve)*	*Series circuit current phase angle (degrees)*
95	f / 3Q	18.5
90	f / 2Q	26.5
70.7	f / Q	45
44.7	2f / Q	63.5
24.2	4f / Q	76
12.4	8f / Q	83

[1]For a series resonant circuit
[2]For a parallel resonant circuit

VOLTAGE DROP ACROSS COMPONENTS

The voltage drop across the coil and across the capacitor in a series-resonant circuit are each proportional to the reactance of the component for a given current (since E = I X). These voltages may be many times the applied voltage for a high-Q circuit. In fact, at resonance, the voltage drop is:

$$E_X = Q_U\, E_{AC}$$

where
E_X = the voltage across the reactive component,
Q_U = the circuit unloaded Q, and
E_{AC} = the applied voltage in Figure 3.38.

(Note that the voltage drop across the inductor is the vector sum of the voltages across the resistance and the reactance; however, for Q greater than 10, the error created by using this is not ordinarily significant.) Since the calculated value of E_X is the RMS voltage, the peak voltage will be higher by a factor of 1.414. Antenna couplers and other high-Q circuits handling significant power may experience arcing from high values of E_X, even though the source voltage to the circuit is well within component ratings.

CAPACITOR LOSSES

Although capacitor energy losses tend to be insignificant compared to inductor losses up to about 30 MHz, the losses may affect circuit Q in the VHF range. Leakage resistance, principally in the solid dielectric that forms the insulating support for the capacitor plates, appears as a resistance in parallel with the capacitor plates. Instead of forming a series resistance, capacitor leakage usually forms a parallel resistance with the capacitive reactance. If the leakage resistance of a capacitor is significant enough to affect the Q of a series-resonant circuit, the parallel resistance (R_P) may be converted to an equivalent series resistance (R_S) before adding it to the inductor's resistance.

$$R_S = \frac{{X_C}^2}{R_P} = \frac{1}{R_P \times (2\,\pi\, f\, C)^2}$$

Example: A 10.0 pF capacitor has a leakage resistance of 10,000 Ω at 50.0 MHz. What is the equivalent series resistance?

$$R_S = \frac{1}{R_P \times (2\,\pi\, f\, C)^2}$$

$$= \frac{1}{1.0 \times 10^4 \times (6.283 \times 50.0 \times 10^6 \times 10.0 \times 10^{-12})^2}$$

$$= \frac{1}{1.0 \times 10^4 \times 9.87 \times 10^{-6}}$$

$$= \frac{1}{0.0987} = 10.1\,\Omega$$

In calculating the impedance, current and bandwidth for a series-resonant circuit in which this capacitor might be used, the series-equivalent resistance of the unit is added to the loss resistance of the coil. Since inductor losses tend to increase with frequency because of skin effect in conductors, and capacitor dielectric losses also tend to increase with frequency, the combined losses in the capacitor and the inductor can seriously reduce circuit Q.

3.8.2 Parallel-Resonant Circuits

Although series-resonant circuits are common, the vast majority of resonant circuits used in radio work are *parallel-resonant circuits*. **Figure 3.40** represents a typical HF parallel-resonant circuit. As is the case for series-resonant circuits, the inductor is the chief source of resistive losses (that is, the parallel loss resistance across the capacitor is not shown), and these losses appear in series with the coil. Because current through parallel-resonant circuits is lowest at resonance, and impedance is highest, they are sometimes called *antiresonant* circuits. (You may encounter the old terms *acceptor* and *rejector* referring to series- and parallel-resonant circuits, respectively.)

Because the conditions in the two legs of the parallel circuit in Figure 3.40 are not the same — the resistance is shown in only one of the legs — all of the conditions by which

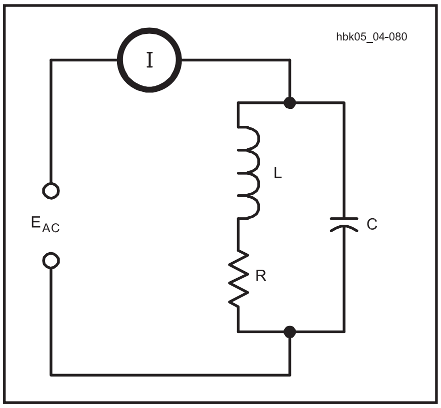

Figure 3.40 — A typical parallel-resonant circuit, with the resistance shown in series with the inductive leg of the circuit. Below a Q_U of 10, resonance definitions may lead to three separate frequencies which converge at higher Q_U levels. See text.

series resonance is determined do not occur simultaneously in a parallel-resonant circuit. **Figure 3.41** graphically illustrates the situation by showing the currents through the two components. (Currents are drawn in the manner of complex impedances shown previously to show the phase angle for each current.) When the inductive and capacitive reactances are identical, the condition defined for series resonance is met as shown at point (a). The impedance of the inductive leg is composed of both X_L and R, which yields an impedance greater than X_C and that is not 180° out of phase with X_C. The resultant current is greater than the minimum possible value and is not in phase with the voltage.

By altering the value of the inductor slightly (and holding the Q constant), a new frequency can be obtained at which the current reaches its minimum. When parallel circuits are tuned using a current meter as an indicator, this point (b) is ordinarily used as an indication of resonance. The current "dip" indicates a condition of maximum impedance and is sometimes called the *antiresonant* point or *maximum impedance resonance* to distinguish it from the condition at which $X_C = X_L$. Maximum impedance is achieved at this point by vector addition of X_C, X_L and R, however, and the result is a current somewhat out of phase with the voltage.

Point (c) in the figure represents the *unity-power-factor* resonant point. Adjusting the inductor value and hence its reactance (while holding Q constant) produces a new resonant frequency at which the resultant current is in phase with the voltage. The new value of inductive reactance is the value required for a parallel-equivalent inductor and its parallel-equivalent resistor (calculated according to the formulas in the last section) to just cancel the capacitive reactance. The value of the parallel-equivalent inductor is always smaller than the actual inductor in series with the resistor and has a proportionally smaller reactance. (The parallel-equivalent resistor, conversely, will always be larger than the coil-loss resistor shown in series with the inductor.) The result is a resonant frequency slightly different from the one for minimum current and the one for $X_L = X_C$.

The points shown in the graph in Figure 3.41 represent only one of many possible situations, and the relative positions of the three resonant points do not hold for all possible cases. Moreover, specific circuit designs can draw some of the resonant points together, for example, compensating for the resistance of the coil by retuning the capacitor. The differences among these resonances are significant for circuit Q below 10, where the inductor's series resistance is a significant percentage of the reactance. Above a Q of 10, the three points converge to within a percent of the frequency and the differences between them can be ignored for practical calculations. Tuning for minimum current will not introduce a sufficiently large phase angle between voltage and current to create circuit difficulties.

PARALLEL CIRCUITS OF MODERATE TO HIGH Q

The resonant frequencies defined above converge in parallel-resonant circuits with Q higher than about 10. Therefore, a single set of formulas will sufficiently approximate circuit performance for accurate predictions. Indeed, above a Q of 10, the performance of a parallel circuit appears in many ways to be simply the inverse of the performance of a series-resonant circuit using the same components.

Accurate analysis of a parallel-resonant circuit requires the substitution of a parallel-equivalent resistor for the actual inductor-loss series resistor, as shown in **Figure 3.42**. Sometimes called the *dynamic resistance* of the parallel-resonant circuit, the parallel-equivalent resistor value will increase with circuit Q, that is, as the series resistance value decreases. To calculate the approximate parallel-equivalent resistance, use the formula:

$$R_P = \frac{{X_L}^2}{R_S} = \frac{(2\,\pi\,f\,L)^2}{R_S} = Q_U X_L$$

for $R_S << X_C << R_P$ and $X_P \approx X_S$ in the equations for series-parallel conversion in the section on Impedance.

Example: What is the parallel-equivalent resistance for a coil with an inductive reactance of 350 Ω and a series resistance of 5.0 Ω at resonance?

$$R_P = \frac{{X_L}^2}{R_S} = \frac{(350\,\Omega)^2}{5.0\,\Omega}$$

$$= \frac{122{,}500\,\Omega^2}{5.0\,\Omega} = 24{,}500\,\Omega$$

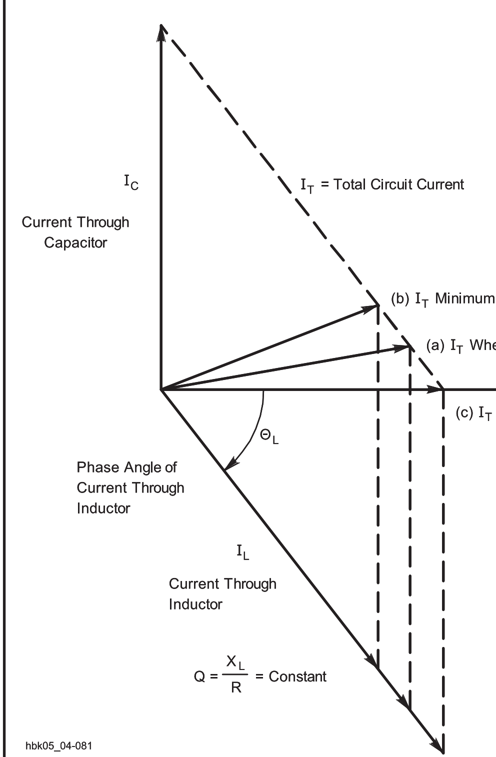

Figure 3.41 — Resonant conditions for a low- Q_U parallel circuit. Resonance may be defined as (a) $X_L = X_C$ (b) minimum current flow and maximum impedance or (c) voltage and current in phase with each other. With the circuit of Figure 3.40 and a Q_U of less than 10, these three definitions may represent three distinct frequencies.

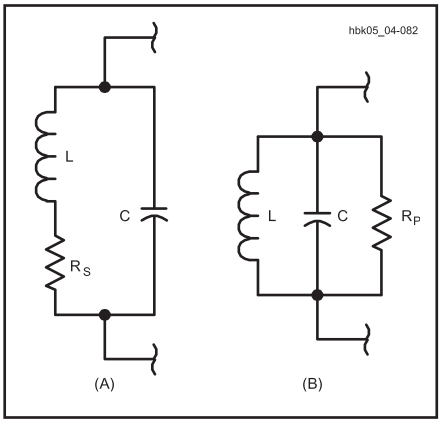

Figure 3.42 — Series and parallel equivalents when both circuits are resonant. The series resistance, R_S in A, is replaced by the parallel resistance, R_p in B, and vice versa. $R_P = X_L^2 / R_S$.

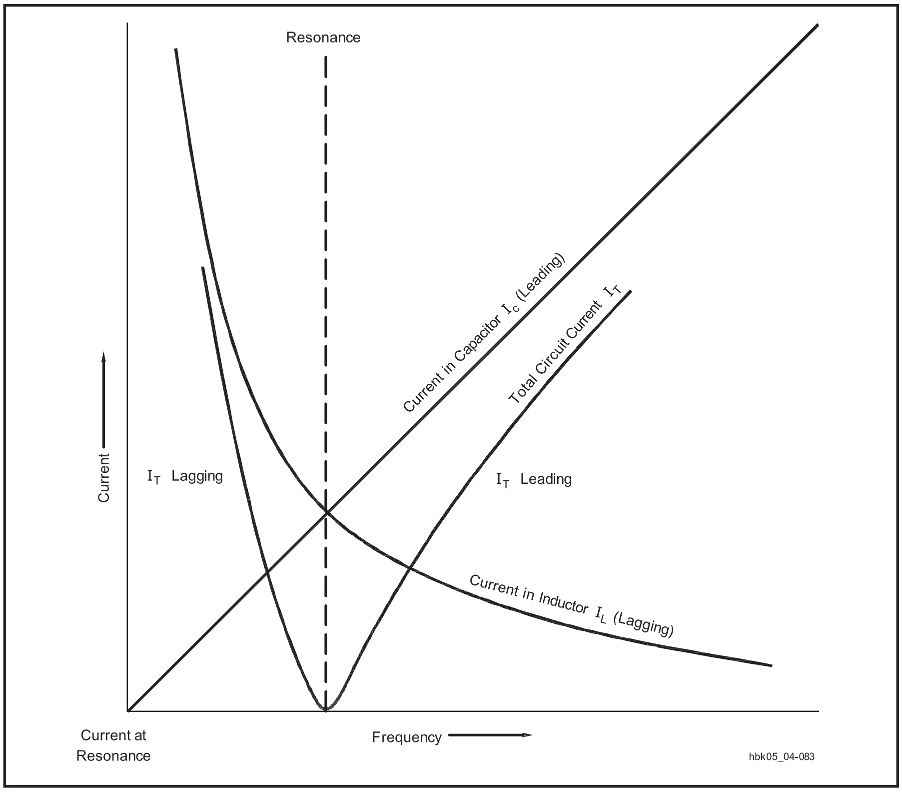

Figure 3.43 — The currents in a parallel-resonant circuit as the frequency moves through resonance. Below resonance, the current lags the voltage; above resonance the current leads the voltage. The base line represents the current level at resonance, which depends on the impedance of the circuit at that frequency.

Since the coil Q_U remains the inductor's reactance divided by its series resistance, the coil Q_U is 70. Multiplying Q_U by the reactance also provides the approximate parallel-equivalent resistance of the coil series resistance.

At resonance, where $X_L = X_C$, R_P defines the impedance of the parallel-resonant circuit. The reactances just equal each other, leaving the voltage and current in phase with each other. In other words, the circuit shows only the parallel resistance. Therefore, the equation for R_P can be rewritten as:

$$Z = \frac{{X_L}^2}{R_S} = \frac{(2 \pi f L)^2}{R_S} = Q_U X_L$$

In this example, the circuit impedance at resonance is 24,500 Ω.

At frequencies below resonance the current through the inductor is larger than that through the capacitor, because the reactance of the coil is smaller and that of the capacitor is larger than at resonance. There is only partial cancellation of the two reactive currents, and the total current therefore is larger than the current taken by the resistance alone. At frequencies above resonance the situation is reversed and more current flows through the capacitor than through the inductor, so the total current again increases.

The current at resonance, being determined wholly by R_P, will be small if R_P is large, and large if R_P is small. **Figure 3.43** illustrates the relative current flows through a parallel-tuned circuit as the frequency is moved from below resonance to above resonance. The base line represents the minimum current level for the particular circuit. The actual current at any frequency off resonance is simply the vector sum of the currents through the parallel equivalent resistance and through the reactive components.

To obtain the impedance of a parallel-tuned circuit either at or off the resonant frequency, apply the general formula:

$$Z = \frac{Z_C \, Z_L}{Z_S}$$

where

Z = overall circuit impedance

Z_C = impedance of the capacitive leg (usually, the reactance of the capacitor),

Z_L = impedance of the inductive leg (the vector sum of the coil's reactance and resistance), and

Z_S = series impedance of the capacitor-inductor combination as derived from the equation for current in a series-resonant circuit.

After using vector calculations to obtain Z_L and Z_S, converting all the values to polar form — as described earlier in this chapter — will ease the final calculation. Of course, each impedance may be derived from the resistance and the application of the basic reactance formulas on the values of the inductor and capacitor at the frequency of interest.

Since the current rises away from resonance, the parallel-resonant-circuit impedance must fall. It also becomes complex, resulting in an ever-greater phase difference between the voltage and the current. The rate at which the impedance falls is a function of Q_U. **Figure 3.44** presents a family of curves showing the impedance drop from resonance for circuit Q ranging from 10 to 100. The curve family for parallel-circuit impedance is essentially the same as the curve family for series-circuit current.

As with series-resonant circuits, the higher the Q of a parallel-tuned circuit, the sharper will be the response peak. Likewise, the lower the Q, the wider the band of frequencies to which the circuit responds. Using the half-power (–3 dB) points as a comparative measure of circuit performance as in series-resonant circuits, $BW = f / Q_U$ and $Q_U = f / BW$, where the resonant frequency and the bandwidth are in the same units. Also similarly to the series-resonant circuit:

$$Q_U = R\sqrt{\frac{C}{L}}$$

As a handy reminder, **Table 3.4** summarizes the performance of parallel-resonant circuits at high and low Q and above and below resonant frequency.

It is possible to use either series- or parallel-resonant circuits to do the same work in many circuits, thus giving the designer considerable flexibility. **Figure 3.45** illustrates this general principle by showing a series-resonant circuit

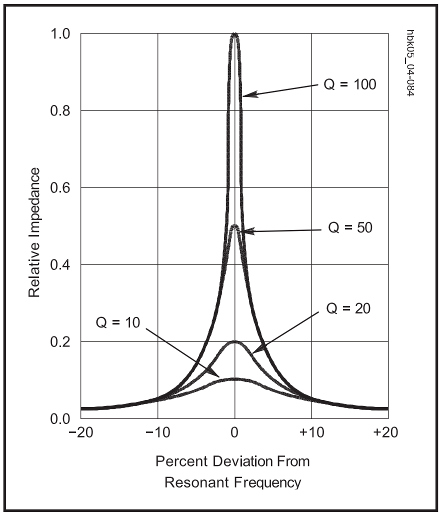

Figure 3.44 — Relative impedance of parallel-resonant circuits with different values of Q_U. The curves are similar to the series-resonant circuit current level curves of Figure 3.39A. The effect of Q_U on impedance is most pronounced within 10% of the resonance frequency.

Table 3.4

The Performance of Parallel-Resonant Circuits

A. High- and Low-Q Circuits (in relative terms)

Characteristic	*High-Q Circuit*	*Low-Q Circuit*
Selectivity	high	low
Bandwidth	narrow	wide
Impedance	high	low
Total current	low	high
Circulating current	high	low

B. Off-Resonance Performance for Constant Values of Inductance and Capacitance

Characteristic	*Above Resonance*	*Below Resonance*
Inductive reactance	increases	decreases
Capacitive reactance	decreases	increases
Circuit resistance	unchanged*	unchanged*
Relative impedance	decreases	decreases
Total current	increases	increases
Circulating current	decreases	decreases
Circuit impedence	capacitive	inductive

*This is true for frequencies near resonance. At distant frequencies, skin effect may alter the resistive losses of the inductor.

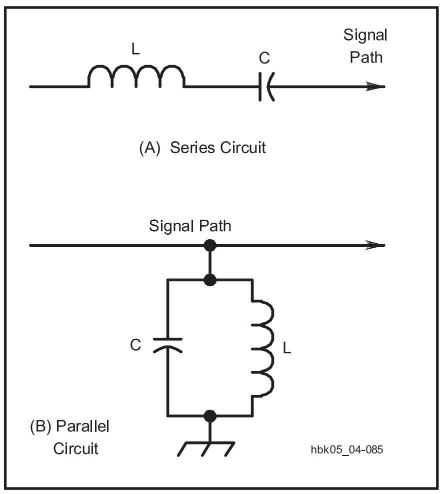

Figure 3.45 — Series- and parallel-resonant circuits configured to perform the same theoretical task: passing signals in a narrow band of frequencies along the signal path. A real design example would consider many other factors.

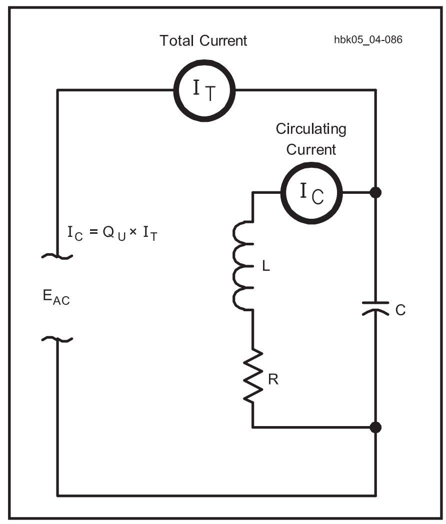

Figure 3.46 — A parallel-resonant circuit redrawn to illustrate both the total current and the circulating current.

in the signal path and a parallel-resonant circuit shunted from the signal path to ground. Assume both circuits are resonant at the same frequency, f, and have the same Q. The series-resonant circuit at A has its lowest impedance at f, permitting the maximum possible current to flow along the signal path. At all other frequencies, the impedance is greater and the current at those frequencies is less. The circuit passes the desired signal and tends to impede signals at undesired frequencies. The parallel circuit at B provides the highest impedance at resonance, f, making the signal path the lowest impedance path for the signal. At frequencies off resonance, the parallel-resonant circuit presents a lower impedance, thus presenting signals with a path to ground and away from the signal path. In theory, the effects will be the same relative to a signal current on the signal path. In actual circuit design exercises, of course, many other variables will enter the design picture to make one circuit preferable to the other.

CIRCULATING CURRENT

In a parallel-resonant circuit, the source voltage is the same for all the circuit elements. The current in each element, however, is a function of the element's reactance. **Figure 3.46** redraws the parallel-resonant circuit to indicate the total current and the current circulating between the coil and the capacitor. The current drawn from the source may be low, because the overall circuit impedance is high. The current through the individual elements may be high, however, because there is little resistive loss as the current circulates through the inductor and capacitor. For parallel-resonant circuits with an unloaded Q of 10 or greater, this *circulating current* is approximately:

$$I_C = Q_U \, I_T$$

where

I_C = circulating current in A, mA or µA,
Q_U = unloaded circuit Q, and
I_T = total current in the same units as I_C.

Example: A parallel-resonant circuit permits an ac or RF total current of 30 mA and has a Q of 100. What is the circulating current through the elements?

$$I_X = Q_U \, I = 100 \times 30 \text{ mA} = 3000 \text{ mA} = 3 \text{ A}$$

Circulating currents in high-Q parallel-tuned circuits can reach a level that causes component heating and power loss. Therefore, components should be rated for the anticipated circulating currents, and not just the total current.

LOADED Q

In many resonant-circuit applications, the only power lost is that dissipated in the resistance of the circuit itself. At frequencies below 30 MHz, most of this resistance is in the coil. Within limits, increasing the number of turns in the coil increases the reactance faster than it raises the resistance, so coils for circuits in which the Q must be high are made with relatively large inductances for the frequency.

When the circuit delivers energy to a load (as in the case of the resonant circuits used in transmitters), the energy consumed in the circuit itself is usually negligible compared with that consumed by the load. The equivalent of such a circuit is shown in **Figure 3.47**, where the parallel resistor, R_L, represents the load to which power is delivered. If the power dissipated in the load is at least 10 times as great as the power lost in the inductor and capacitor, the parallel impedance of the resonant circuit itself will be so high compared with the resistance of the load that for all practical purposes the impedance of the combined circuit is equal to the load impedance. Under these conditions, the load resistance replaces the circuit impedance in calculating Q. The Q of a parallel-resonant circuit loaded by a resistive impedance is:

$$Q_L = \frac{R_L}{X}$$

where

Q_L = circuit loaded Q,
R_L = parallel load resistance in ohms, and
X = reactance in ohms of either the inductor or the capacitor.

Example: A resistive load of 3000 Ω is connected across a resonant circuit in which the inductive and capacitive reactances are each 250 Ω. What is the circuit Q?

$$Q_L = \frac{R_L}{X} = \frac{3000\ \Omega}{250\ \Omega} = 12$$

The effective Q of a circuit loaded by a parallel resistance increases when the reactances are decreased. A circuit loaded with a relatively low resistance (a few thousand ohms) must have low-reactance elements (large capacitance and small inductance) to have reasonably high Q. Many power-handling circuits, such as the output networks of transmitters, are designed by first choosing a loaded Q for the circuit and then determining component values. See the chapter on **RF Power Amplifiers** for more details.

Parallel load resistors are sometimes added to parallel-resonant circuits to lower the circuit Q and increase the circuit bandwidth. By using a high-Q circuit and adding a parallel resistor, designers can tailor the circuit response to their needs. Since the parallel resistor consumes power, such techniques ordinarily apply to receiver and similar low-power circuits, however.

Example: Specifications call for a parallel-resonant circuit with a bandwidth of 400 kHz at 14.0 MHz. The circuit at hand has a Q_U of 70.0 and its components have reactances of 350 Ω each. What is the parallel load resistor that will increase the bandwidth to the specified value? The bandwidth of the existing circuit is:

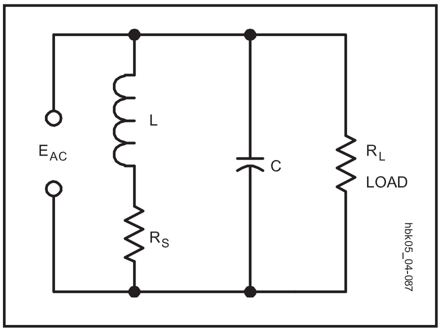

Figure 3.47 — A loaded parallel-resonant circuit, showing both the inductor-loss resistance and the load, R_L. If smaller than the inductor resistance, R_L will control the loaded Q of the circuit (Q_L).

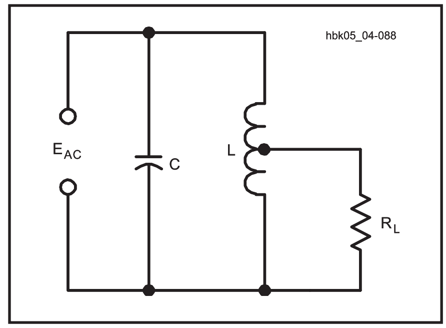

Figure 3.48 — A parallel-resonant circuit with a tapped inductor to effect an impedance match. Although the impedance presented to E_{AC} is very high, the impedance at the connection of the load, R_L, is lower.

$$BW = \frac{f}{Q_U} = \frac{14.0\ \text{MHz}}{70.0} = 0.200\ \text{MHz}$$

= 200 kHz

The desired bandwidth, 400 kHz, requires a circuit with a Q of:

$$Q = \frac{f}{BW} = \frac{14.0\ \text{MHz}}{0.400\ \text{MHz}} = 35.0$$

Since the desired Q is half the original value, halving the resonant impedance or parallel-resistance value of the circuit is in order. The present impedance of the circuit is:

$$Z = Q_U X_L = 70.0 \times 350\ \Omega = 24500\ \Omega$$

The desired impedance is:

$$Z = Q_U X_L = 35.0 \times 350\ \Omega$$

= 12250 Ω = 12.25 kΩ

or half the present impedance.

A parallel resistor of 24,500 Ω, or the nearest lower value (to guarantee sufficient bandwidth), will produce the required reduction in Q and bandwidth increase. Although this example simplifies the situation encountered in real design cases by ignoring such factors as the shape of the band-pass curve, it illustrates the interaction of the ingredients that determine the performance of parallel-resonant circuits.

IMPEDANCE TRANSFORMATION

An important application of the parallel-resonant circuit is as an impedance matching device. Circuits and antennas often need to be connected to other circuits or feed lines that do not have the same impedance. To transfer power effectively requires a circuit that will convert or "transform" the impedances so that each connected device or system can operate properly.

Figure 3.48 shows such a situation where the source, E_{AC}, operates at a high impedance, but the load, R_L, operates at a low impedance. The technique of impedance transformation shown in the figure is to connect the parallel-resonant circuit, which has a high impedance, across the source, but connect the load across only a portion of the coil. (This is called *tapping the coil* and the connection point is a *tap*.) The coil acts as an *autotransformer*, described in the following section, with the magnetic field of the coil shared between what are effectively two coils in series, the upper coil having many turns and the lower coil fewer turns. Energy stored in the field induces larger voltages in the many-turn coil than it does in the fewer-turn coil, "stepping down" the input voltage so that energy can be extracted by the load at the required lower voltage-to-current ratio (which is impedance). The correct tap point on the coil usually has to be experimentally determined, but the technique is very effective.

When the load resistance has a very low value (say below 100 Ω) it may be connected in series in the resonant circuit (such as R_S in Figure 3.42A, for example), in which case the series L-R circuit can be transformed to an equivalent parallel L-R circuit as previously described. If the Q is at least 10, the equivalent parallel impedance is:

$$Z_R = \frac{X^2}{R_L}$$

where

Z_R = resistive parallel impedance at resonance,
X = reactance (in ohms) of either the coil or the capacitor, and
R_L = load resistance inserted in series.

If the Q is lower than 10, the reactance will have to be adjusted somewhat — for the reasons given in the discussion of low-Q parallel resonant circuits — to obtain a resistive impedance of the desired value.

These same techniques work in either "direction" — with a high-impedance source and low-impedance load or vice versa. Using a parallel-resonant circuit for this application does have some disadvantages. For instance, the common connection between the input and the output provides no dc isolation. Also, the common ground is sometimes troublesome with regard to ground-loop currents. Consequently, a circuit with only mutual magnetic coupling is often preferable. With the advent of ferrites, constructing impedance transformers that are both broadband and permit operation well up into the VHF portion of the spectrum has become relatively easy. The basic principles of broadband impedance transformers appear in the **RF Techniques** chapter.

3.9 Analog Signal Processing

The term *analog signal* refers to voltages, currents and waves that make up ac radio and audio signals, dc measurements, even power. The essential characteristic of an analog signal is that the information or energy it carries is continuously variable. Even small variations of an analog signal affect its value or the information it carries. This stands in contrast to *digital signals* that have values only within well-defined and separate ranges called *states*. To be sure, at the fundamental level all circuits and signals are analog: Digital signals are created by designing circuits that restrict the values of analog signals to those discrete states.

Analog signal processing involves various electronic stages to perform functions on analog signals such as amplifying, filtering, modulation and demodulation. A piece of electronic equipment, such as a radio, is constructed by combining a number of these circuits. How these stages interact with each other and how they affect the signal individually and in tandem is the subject of sections later in the chapter.

3.9.1 Terminology

A similar terminology is used when describing active electronic devices. The letter V or v stands for voltages and I or i for currents. Capital letters are often used to denote dc or bias values (bias is discussed later in this chapter). Lower-case often denotes instantaneous or ac values.

Voltages generally have two subscripts indicating the terminals between which the voltage is measured (V_{BE} is the dc voltage between the base and the emitter of a bipolar transistor). Currents have a single subscript indicating the terminal into which the current flows (I_C is the dc current into the collector of a bipolar transistor). If the current flows out of the device, it is generally treated as a negative value.

Resistance is designated with the letter R or r, and impedance with the letter Z or z. For example, r_{DS} is resistance between drain and source of an FET and Z_i is input impedance. For some parameters, values differ for dc and ac signals. This is indicated by using capital letters in the subscripts for dc and lower-case subscripts for ac. For example, the common-emitter dc current gain for a bipolar transistor is designated as h_{FE}, and h_{fe} is the ac current gain. (See the section on transistor amplifiers in the **Circuits and Components** chapter for a discussion of the common-emitter circuit.) Qualifiers are sometimes added to the subscripts to indicate certain operating modes of the device. SS for saturation, BR for breakdown, ON and OFF are all commonly used.

Power supply voltages have two subscripts that are the same, indicating the terminal to which the voltage is applied. V_{DD} would represent the power supply voltage applied to the drain of a field-effect transistor.

Since integrated circuits are collections of semiconductor components, the abbreviations for the type of semiconductor used also apply to the integrated circuit. For example, V_{CC} is a power supply voltage for an integrated circuit made with bipolar transistor technology in which voltage is applied to transistor collectors.

3.9.2 Linearity

The premier properties of analog signals are *superposition* and *scaling*. Superposition is the property by which signals are combined, whether in a circuit, in a piece of wire, or even in air, as the sum of the individual signals. This is to say that at any one point in time, the voltage of the combined signal is the sum of the voltages of the original signals at the same time. In a *linear system* any number of signals will add in this way to give a single combined signal. (Mathematically, this is a

The Decibel

The decibel (dB) is the standard unit for comparing two quantities, such as power or voltage, as a ratio. It is logarithmic so very large and very small ratios are easy to work with. The formula for calculating decibels is:

$$dB = 10\log\left(\frac{P_2}{P_1}\right) = 20\log\left(\frac{V_2}{V_1}\right)$$

When calculating decibels using voltage (or current), remember that both voltages must be measured at the same value of circuit impedance. For example, the 50 Ω input and output impedance of a preamplifier. If this is not the case, convert each voltage to a power level (V^2/Z) and calculate decibels using a power ratio.

Another caution is rounding decibel values. Because decibels are a logarithmic value, a change of 1 dB, whether from 10 dB or 10,000 dB, represents a change of nearly 26% so that ±1 dB is also ±26%. Be careful when rounding or specifying tolerances in dB.

For more on working with decibels, read the articles "Untangling the Decibel Dilemma" found in this book's online content. Rohde & Schwarz has published Application Note IMA98 "dB or not dB? Everything you ever wanted to know about decibels but were afraid to ask…" at the website listed in the References and Bibliography section of this chapter.

linear combination.) For this reason, analog signals and components are often referred to as *linear signals* or *linear components*. A linear system whose characteristics do not change, such as a resistive voltage divider, is called *time-invariant*. If the system changes with time, it is *time-varying*. The variations may be random, intermittent (such as being adjusted by an operator) or periodic.

One of the more important features of superposition, for the purposes of signal processing, is that signals that have been combined by superposition can be separated back into the original signals. This is what allows multiple signals that have been received by an antenna to be separated back into individual signals by a receiver.

3.9.3 Linear Operations

Any operation that modifies a signal and obeys the rules of superposition and scaling is a *linear operation*. The following sections explain the basic linear operations from which linear systems are made.

AMPLIFICATION AND ATTENUATION

Amplification and *attenuation* scale signals to be larger and smaller, respectively. The operation of *scaling* is the same as multiplying the signal at each point in time by a constant value; if the constant is greater than one then the signal is amplified, if less than one then the signal is attenuated.

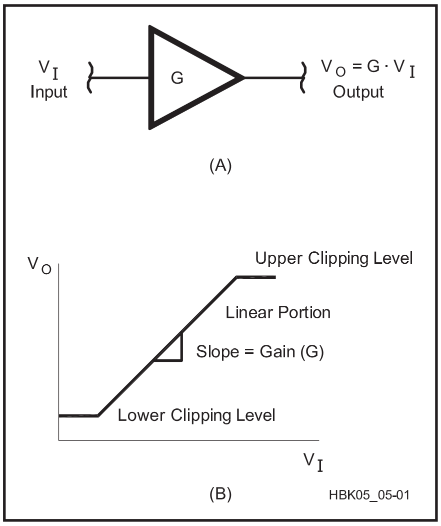

Figure 3.49 — Generic amplifier. (A) Symbol. For the linear amplifier, gain is the constant value, G, and the output voltage is equal to the input voltage times G; (B) Transfer function, input voltage along the x-axis is converted to the output voltage along the y-axis. The linear portion of the response is where the plot is diagonal; its slope is equal to the gain, G. Above and below this range are the clipping limits, where the response is not linear and the output signal is clipped.

An *amplifier* is a circuit that increases the amplitude of a signal. Schematically, a generic amplifier is signified by a triangular symbol, its input along the left face and its output at the point on the right (see **Figure 3.49**). The linear amplifier multiplies every value of a signal by a constant value. Amplifier gain is often expressed as a multiplication factor (× 5, for example).

$$\text{Gain} = V_o/V_i$$

where V_o is the output voltage from an amplifier when an input voltage, V_i, is applied.

An *attenuator* is a circuit that reduces the amplitude of a signal. Attenuators can be constructed from passive circuits, such as the attenuators built using resistors, described in the chapter on **Test Equipment and Measurements**. Active attenuator circuits include amplifiers whose gain is less than one or circuits with adjustable resistance in the signal path, such as a PIN diode attenuator or amplifier with gain is controlled by an external voltage.

GAIN AND TRANSCONDUCTANCE

The operation of an amplifier is specified by its *gain*. Gain in this sense is defined as the change (Δ) in the output parameter divided by the corresponding change in the input parameter. If a particular device measures its input and output as currents, the gain is called a *current gain*. If the input and output are voltages, the amplifier is defined by its *voltage gain*. *Power gain* is often used, as well. Gain is technically unit-less, but is often given in V/V. Decibels are often used to specify gain, particularly power gain. (Gain is often expressed in decibels (dB) — see the sidebar "The Decibel".)

If an amplifier's input is a voltage and the output is a current, the ratio of the change in output current to the change in input voltage is called *transconductance*, g_m.

$$g_m = \frac{\Delta I_o}{\Delta V_i}$$

Transconductance has the same units as conductance and admittance, siemens (S), but is only used to describe the operation of active devices, such as transistors or vacuum tubes.

Ideal linear amplifiers have the same gain for all parts of a signal. Thus, a gain of 10 changes 10 V to 100 V, 1 V to 10 V and –1 V to –10 V. (Gain can also be less than one.) The ability of an amplifier to change a signal's level is limited by the amplifier's *dynamic range*, however. An amplifier's dynamic range is the

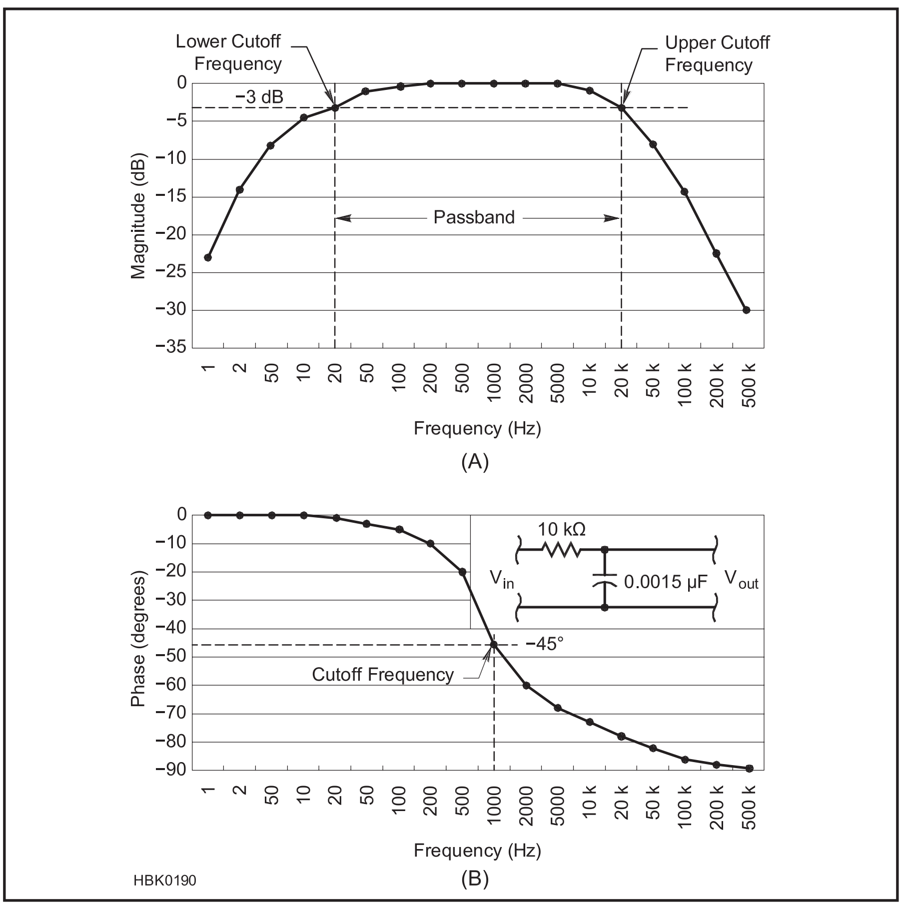

Figure 3.50 — Bode plot of (A) band-pass filter magnitude response and (B) an RC low-pass filter phase response.

range of signal levels over which the amplifier produces the required gain without distortion. Dynamic range is limited for small signals by noise, distortion and other nonlinearities.

Dynamic range is limited for large signals because an amplifier can only produce output voltages (and currents) that are within the range of its power supply. (Power-supply voltages are also called the *rails* of a circuit.) As the amplified output approaches one of the rails, the output cannot exceed a given voltage near the rail and the operation of the amplifier becomes nonlinear as described below in the section on Clipping and Rectification.

Another similar limitation on amplifier linearity is called *slew rate*. Applied to an amplifier, this term describes the maximum rate at which a signal can change levels and still be accurately amplified in a particular device. *Input slew rate* is the maximum rate of change to which the amplifier can react linearly. *Output slew rate* refers to the maximum rate at which the amplifier's output circuit can change. Slew rate is an important concept, because there is a direct correlation between a signal level's rate of change and the frequency content of that signal. The amplifier's ability to react to or reproduce that rate of change affects its frequency response and dynamic range.

FREQUENCY RESPONSE AND BODE PLOTS

Another important characteristic of a circuit is its *frequency response*, a description of how it modifies a signal of any frequency. Frequency response can be stated in the form of a mathematical equation called a *transfer function*, but it is more conveniently presented as a graph of gain vs frequency. The ratio of output amplitude to input amplitude is often called the circuit's *magnitude* or *amplitude response*. Plotting the circuit's magnitude response in dB versus frequency on a logarithmic scale, such as in **Figure 3.50A**, is called a *Bode plot* (after Henrik Wade Bode). The combination of decibel and log-frequency scales is used because the behavior of most circuits depends on ratios of amplitude and frequency and thus appears linear on a graph in which both the vertical and horizontal scales are logarithmic.

Most circuits also affect a signal's phase along with its amplitude. This is called *phase shift*. A plot of phase shift from the circuit's input to its output is called the *phase response*, seen in Figure 3.50B. Positive phase greater than 0° indicates that the output signal *leads* the input signal, while *lagging* phase shift has a negative phase. The combination of an amplitude and phase response plot gives a good picture of what effect the circuit has on signals passing through it.

TRANSFER CHARACTERISTICS

Transfer characteristics are the ratio of an output parameter to an input parameter, such as output current divided by input current, h_{FE}. There are different families of transfer characteristics, designated by letters such as h, s, y or z. Each family compares parameters in specific ways that are useful in certain design or analysis methods. The most common transfer characteristics used in radio are the h-parameter family (used in transistor models) and the s-parameter family (used in RF design, particularly at VHF and above). See the **RF Techniques** chapter for more discussion of transfer characteristics.

COMPLEX FREQUENCY

We are accustomed to thinking of frequency as a real number — so many cycles per second — but frequency can also be a complex number, s, with both a real part, designated by σ, and an imaginary part, designated by *j*ω. (ω is also equal to 2πf.) The resulting complex frequency is written as $s = \sigma + j\omega$. At the lower left of **Figure 3.51** a pair of real and imaginary axes are used to plot values of s. This is called the *s-plane*. Complex frequency is used in Laplace transforms, a mathematical technique used for circuit and signal analysis. (Thorough treatments of the application of complex frequency can be found in college-level textbooks on circuit and signal analysis.)

When complex frequency is used, a sinusoidal signal is described by Ae^{st}, where A is the amplitude of the signal and t is time. Because s is complex, $Ae^{st} = Ae^{(\sigma+j\omega)t} = A(e^{\sigma t})(e^{j\omega t})$. The two exponential terms describe independent characteristics of the signal. The second term, $e^{j\omega t}$, is the sine wave with which we are familiar and that has frequency f, where $f = \omega/2\pi$. The first term, $e^{\sigma t}$, represents the rate at which the signal increases or decreases. If σ is negative, the exponential term decreases with time and the signal gets smaller and smaller. If σ is positive, the signal gets larger and larger. If $\sigma = 0$, the exponential term equals 1, a constant, and the signal amplitude stays the same.

Complex frequency is very useful in describing a circuit's stability. If the response to an input signal is at a frequency on the right-hand side of the s-plane for which $\sigma > 0$, the system is *unstable* and the output signal will get larger until it is limited by the circuit's power supply or some other mechanism. If the response is on the left-hand side of the s-plane, the system is *stable* and the response to the input signal will eventually die out. The larger the absolute value of σ, the faster the response changes. If the response is precisely on the *j*ω axis where $\sigma = 0$, the response will persist indefinitely.

In Figure 3.51 the equation for the simple RC-circuit's transfer function is shown at the left of the figure. It describes the circuit's behavior at real-world frequencies as well as imaginary frequencies whose values contain *j*. Because complex numbers are used for f, the transfer function describes the circuit's phase response, as well as amplitude. At one such frequency, $f = -j/2\pi RC$, the denominator of the transfer function is zero, and the gain is infinite! Infinite gain is a pretty amazing thing to achieve with a passive circuit — but because this can only happen at an imaginary frequency, it does not happen in the real world.

The practical effects of complex frequency can be experienced in a narrow CW crystal or LC filter. The poles of such a filter are just to the left of the *j*ω axis, so the input signal causes the filters to "ring", or output a damped sine wave along with the desired signal. Similarly, the complex frequency of an oscillator's output at power-up must have $\sigma > 0$ or the oscillation would never start! The output amplitude continues to grow until limiting takes place, reducing gain until $\sigma = 1$ for a steady output.

Obtaining a Frequency Response

With the computer tools such as spreadsheets, it's easy to do the calculations and make a graph of frequency response. If you don't have a spreadsheet program, then use semi-log graph paper with the linear axis used for dB or phase and the logarithmic axis for frequency. An Excel spreadsheet set up to calculate and display frequency response is available in this book's online content. You can modify it to meet your specific needs.

Follow these rules whether using a spreadsheet or graph paper:

- Measure input and output in the same units, such as volts, and use the same conventions, such as RMS or peak-to-peak.
- Measure phase shift from the input to the output. (The **Test Equipment and Measurements** chapter discusses how to make measurements of amplitude and phase.)
- Use 10 log (P_O/P_I) for power ratios and 20 log (V_O/V_I) for voltage or current.

To make measurements that are roughly equally spaced along the logarithmic frequency axis, follow the "1-2-5 rule." Dividing a range this way, for example 1-2-5-10-20-50-100-200-500 Hz, creates steps in approximately equal ratios that then appear equally spaced on a logarithmic axis.

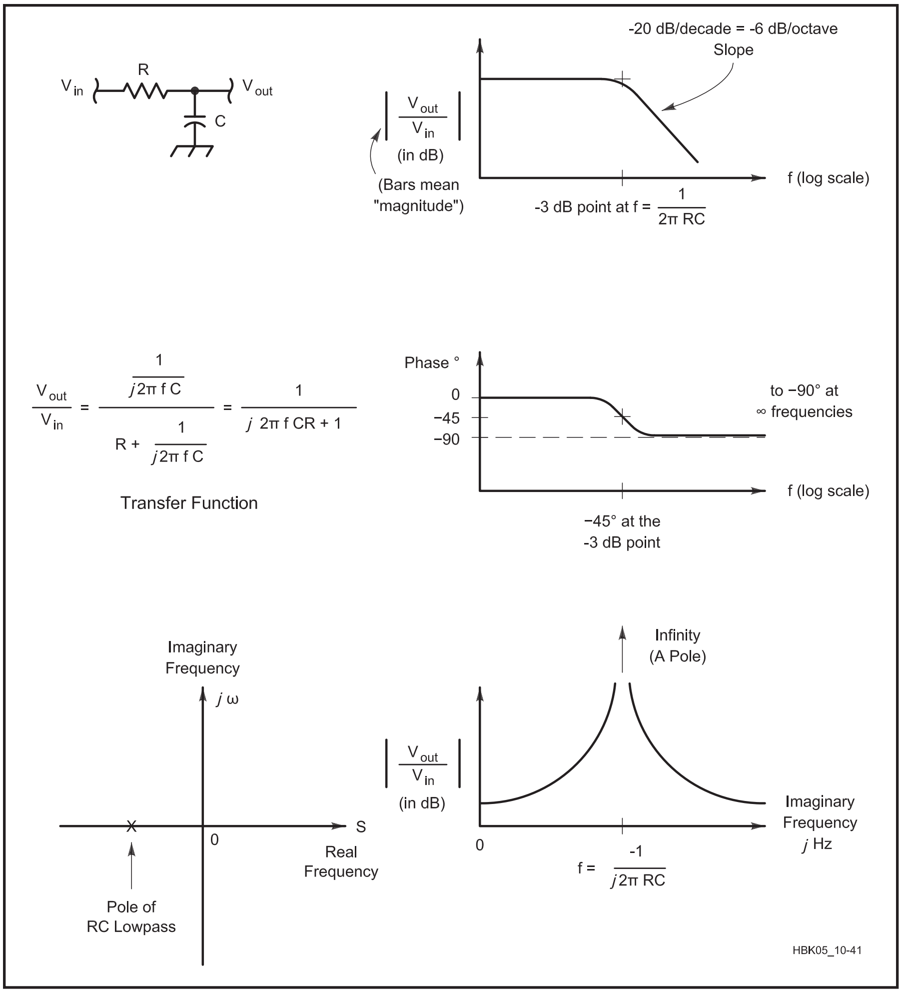

Figure 3.51 — The transfer function for a circuit describes both the magnitude and phase response of a circuit. The RC circuit shown at the upper left has a pole at f = 2πRC, the filter's –3 dB or cutoff frequency, at which the phase response is a 45° lagging phase shift. Poles cause an infinite response on the imaginary frequency axis.

POLES AND ZEROES

Frequencies that cause the transfer function to become infinite are called *poles*. This is shown at the bottom right of Figure 3.51 in the graph of the circuit's amplitude response for imaginary frequencies shown on the horizontal axis. (The pole causes the graph to extend up "as a pole under a tent," thus the name.) Similarly, circuits can have *zeroes* which occur at imaginary frequencies that cause the transfer function to be zero, a less imaginative name, but quite descriptive.

A circuit can also have poles and zeroes at frequencies of zero and infinity. For example, the circuit in Figure 3.51 has a zero at infinity because the capacitor's reactance is zero at infinity and the transfer function is zero, as well. If the resistor and capacitor were exchanged, so that the capacitor was in series with the output, then at zero frequency (dc), the output would be zero because the capacitor's reactance was infinite, creating a zero.

Complex circuits can have multiple poles or multiple zeroes at the same frequency. Poles and zeroes can also occur at frequencies that are combinations of real and imaginary numbers. The poles and zeroes of a circuit form a pattern in the complex plane that corresponds to certain types of circuit behavior. (The relationships between the pole-zero pattern and circuit behavior is beyond the scope of this book, but are covered in textbooks on circuit theory.)

What is a Pole?

Poles cause a specific change in the circuit's amplitude and phase response for real-world frequencies, even though we can't experience imaginary frequencies directly. A pole is associated with a bend in a magnitude response plot that changes the slope of the response downward with increasing frequency by 6 dB per octave (20 dB per decade; an octave is a 2:1 frequency ratio, a decade is a 10:1 frequency ratio).

There are four ways to identify the existence and frequency of a pole as shown in Figure 3.51:

1) For a downward bend in the magnitude versus frequency plot, the pole is at the –3 dB frequency for a single pole. If the bend causes a change in slope of more than 6 dB/octave, there must be multiple poles at this frequency.

2) A 90° lagging change on a phase versus frequency plot, where the lag increases with frequency. The pole is at the point of 45° added lag on the S-shaped transition. Multiple poles will add their phase lags, as above.

3) On a circuit diagram, a single pole looks like a simple RC low-pass filter. The pole is at the –3 dB frequency (f = 1/2πRC Hz). Any other circuit with the same response has a pole at the same frequency.

4) In an equation for the transfer function of a circuit, a pole is a theoretical value of frequency that would result in infinite gain. This is clearly impossible, but as the value of frequency will either be absolute zero, or will have an imaginary component, it is impossible to make an actual real-world signal at a pole frequency.

For example, comparing the amplitude responses at top and bottom of Figure 3.51 shows that the frequency of the pole is equal to the circuit's –3 dB cutoff frequency (1/2πfC) multiplied by *j*, which is also the frequency at which the circuit causes a –45° (lagging) phase shift from input to output.

What Is a Zero?

A zero is the complement of a pole. In math, it is a frequency at which the transfer function equation of a circuit is zero. This is not impossible in the real world (unlike the pole), so zeroes can be found at real-number frequencies as well as complex-number frequencies.

Each zero is associated with an *upward* bend of 6 dB per octave in a magnitude response. Similarly to a pole, the frequency of the zero is at the +3 dB point. Each zero is associated with a transition on a phase-versus-frequency plot that reduces the lag by 90°. The zero is at the 45° leading phase point. Multiple zeroes add their phase shifts just as poles do.

In a circuit, a zero creates gain that increases with frequency forever above the zero frequency. This requires active circuitry that would inevitably run out of gain at some frequency, which implies one or more poles up there. In real-world circuits, zeroes are usually not found by themselves, making the magnitude response go up, but rather paired with a pole of a different frequency, resulting in the magnitude response having a slope between two frequencies but flat above and below them.

Real-world circuit zeroes are only found accompanied by a greater or equal number of poles. Consider a classic RC high-pass filter, such as if the resistor and capacitor in Figure 3.51 were exchanged. The response of such a circuit increases at 6 dB per octave from 0 Hz (so there must be a zero at 0 Hz) and then levels off at 1/2πRC Hz. This leveling off is due to the presence of a pole adding its 6 dB-per-octave roll-off to cancel the 6 dB-per-octave roll-up of the zero. The transfer function for such as circuit would equal zero at zero frequency and infinity at the imaginary pole frequency.

FEEDBACK AND OSCILLATION

The *stability* of an amplifier refers to its ability to provide gain to a signal without tending to oscillate. For example, an amplifier just on the verge of oscillating is not generally considered to be "stable." If the output of an amplifier is fed back to the input, the feedback can affect the amplifier stability. If the amplified output is added to the input, the output of the sum will be larger. This larger output, in turn, is also fed back. As this process continues, the amplifier output will continue to rise until the amplifier cannot go any higher (clamps). Such *positive feedback* increases the amplifier gain, and is called *regeneration*. (The chapter on **Oscillators and Synthesizers** includes a discussion of positive feedback.)

Most practical amplifiers have some intrinsic and unavoidable feedback either as part of the circuit or in the amplifying device(s) itself. To improve the stability of an amplifier, *negative feedback* can be added to counteract any unwanted positive feedback. Negative feedback is often combined with a phase-shift *compensation* network to improve the amplifier stability.

Although negative feedback reduces amplifier or stage gain, the advantages of *stable* gain, freedom from unwanted oscillations and the reduction of distortion are often key design objectives and advantages of using negative feedback.

The design of feedback networks depends on the desired result. For amplifiers, which should not oscillate, the feedback network is customized to give the desired frequency response without loss of stability. For oscillators, the feedback network is designed to create a steady oscillation at the desired frequency.

SUMMING

In a linear system, nature does most of the work for us when it comes to adding signals; placing two signals together naturally causes them to add according to the principle of superposition. When processing signals, we would like to control the summing operation so the signals do not distort or combine in a nonlinear way. If two signals come from separate stages and they are connected together directly, the circuitry of the stages may interact, causing distortion of either or both signals.

Summing amplifiers generally use a resistor in series with each stage, so the resistors connect to the common input of the following stage. This provides some *isolation* between the output circuits of each stage. **Figure 3.52** illustrates the resistors connecting to a summing amplifier. Ideally, any time we wanted to combine signals (for example, combining an audio signal with a sub-audible tone in a 2 meter FM transmitter prior to modulating the RF signal) we could use a summing amplifier.

FILTERING

A *filter* is a common linear stage in radio equipment. Filters are characterized by their ability to selectively attenuate certain frequencies in the filter's *stop band*, while passing or amplifying other frequencies in the *passband*. If the filter's passband extends to or near dc, it is a *low-pass* filter, and if to infinity (or at least very high frequencies for the circuitry involved), it is a *high-pass* filter. Filters that pass a range of frequencies are *band-pass* filters. *All-pass* filters are designed to affect only the phase of a signal without changing the signal amplitude. The range of frequencies between a band-pass circuit's low-pass and high-pass regions is its *mid-band*.

Figure 3.50A is the amplitude response for a typical band-pass audio filter. It shows that the input signal is passed to the output with no loss (0 dB) between 200 Hz and 5 kHz. This is the filter's *mid-band response*. Above and below those frequencies the response of the filter begins to drop. By 20 Hz and 20 kHz, the amplitude response has been reduced to one-half of (–3 dB) the mid-band response. These points are called the circuit's *cutoff* or *corner* or *half-power frequencies*. The range between the cutoff frequencies is the filter's passband. Outside the filter's passband, the amplitude response drops to 1/200th (–23 dB) of mid-band response at 1 Hz and only 1/1000th (–30 dB) at 500 kHz. The steepness of the change in response with frequency is the filter's *roll-off* and it is usually specified in dB/octave (an octave is a doubling or halving of frequency) or dB/decade (a decade is a change to 10 times or 1/10th frequency).

Figure 3.50B represents the phase response of a different filter — the simple RC low-pass filter shown at the upper right. As frequency increases, the reactance of the capacitor becomes smaller, causing most of the input signal to appear across the fixed-value resistor instead. At low frequencies, the capacitor has little effect on phase shift. As the signal frequency rises, however, there is more and more phase shift until at the cutoff frequency, there is 45° of lagging phase shift, plotted as a negative number. Phase shift then gradually approaches 90°.

Practical analog (both passive and active) and digital filters are discussed in the chapter **Analog and Digital Filtering**. Filters at RF may also be created by using transmission lines as described in the **Transmission Lines** chapter. All practical amplifiers are in effect either low-pass filters or band-pass filters, because their magnitude response decreases as the frequency increases beyond their ability to amplify signals.

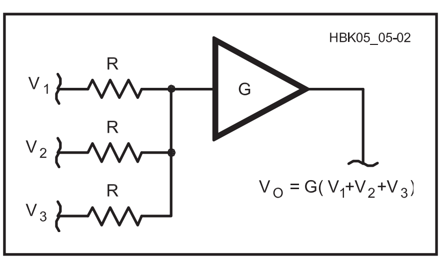

Figure 3.52 — Summing amplifier. The output voltage is equal to the sum of the input voltages times the amplifier gain, G. As long as the resistance values, R, are equal and the amplifier input impedance is much higher, the actual value of R does not affect the output signal.

3.9.4 Nonlinear Operations

All signal processing doesn't have to be linear. Any time that we treat various signal levels differently, the operation is called *nonlinear*. This is not to say that all signals must be treated the same for a circuit to be linear. High-frequency signals are attenuated in a low-pass filter while low-frequency signals are not, yet the filter can be linear. The distinction is that the amount of attenuation at different frequencies is always the same, regardless of the amplitude of the signals passing through the filter.

What if we do not want to treat all voltage levels the same way? This is commonly desired in analog signal processing for clipping, rectification, compression, modulation and switching.

CLIPPING AND RECTIFICATION

Clipping is the process of limiting the range of signal voltages passing through a circuit (in other words, *clipping* those voltages outside the desired range from the signals). There are a number of reasons why we would like to do this. As shown in Figure 3.49, clipping is the process of limiting the positive and negative peaks of a signal. (Clipping is also called *clamping*.)

Clipping might be used to prevent a large audio signal from causing excessive deviation in an FM transmitter that would interfere with communications on adjacent channels.

Clipping circuits are also used to protect sensitive inputs from excessive voltages. Clipping distorts the signal, changing it so that the original signal waveform is lost.

Another kind of clipping results in *rectification*. A *rectifier* circuit clips off all voltages of one polarity (positive or negative) and passes only voltages of the other polarity, thus changing ac to pulsating dc (see the **Power Sources** chapter). Another use of rectification is in a *peak detection* circuit that measures the peak value of a waveform. Only one polarity of the ac voltage needs to be measured and so a rectifier clips the unwanted polarity.

LIMITING

Another type of clipping occurs when an amplifier is intentionally operated with so much gain that the input signals result in an output that is clipped at the limits of its power supply voltages (or some other designated voltages). The amplifier is said to be driven into *limiting* and an amplifier designed for this behavior is called a *limiter*. Limiters are used in FM receivers to amplify the signal until all amplitude variations in the signal are removed and the only characteristic of the original signal that remains is the frequency.

LOGARITHMIC AMPLIFICATION

It is sometimes desirable to amplify a signal logarithmically, which means amplifying low levels more than high levels. This type of amplification is often called *signal compression*. Speech compression is sometimes used in audio amplifiers that feed modulators. The voice signal is compressed into a small range of amplitudes, allowing more voice energy to be transmitted without overmodulation (see the **Modulation** and the **Transmitting** chapters).

3.9.5 System Design Functions

Many kinds of electronic equipment are developed by combining basic analog signal processing circuits, often treating them as independent functional blocks. This section describes several topics associated with building systems from multiple blocks. Because analog circuits often interface with digital circuits or include digital elements in a hybrid circuit, some topics associated with digital systems are included. Many similar functions are implemented as part of Digital Signal Processing (DSP) systems and have similar behaviors and concerns. Although not all basic electronic functions are discussed here, the concepts associated with combining them can be applied generally.

Since our main concern is the effect that circuitry has on a signal, we often describe the circuit by its actions rather than by its specific components. A *black box* is a circuit that can be described entirely by the behavior of its interfaces with other blocks and circuitry. When circuits are combined in such as way as to perform sequential operations on a signal, the individual circuits are called *stages*.

The most general way of referring to an analog circuit is as a *network*. Two basic properties of analog networks are of principal concern: the effect that the network has on a signal and the interaction that the network has with the circuitry surrounding it. Interfaces between the network and the rest of the network are called *ports*.

Many analog circuits are analyzed as *two-port networks* with an input and an output port. The signal is fed into the input port, is modified inside the network and then exits from the output port. (See the chapter on **RF Techniques** for more information on two-port networks.)

TRANSFER FUNCTIONS

The specific way in which the analog circuit modifies the signal can be described mathematically as a transfer function. The mathematical operation that combines a signal with a transfer function is pictured symbolically in **Figure 3.53**. The transfer function, h(t) or h(f), describes the circuit's modification of the input signal in the time domain where all values are functions of time, such as a(t) or b(t), or in the frequency domain where all values are functions of frequency, such as a(f) or b(f). The mathematical operation by which h(t) operates on a(t) is called *convolution* and is represented as a dot, as in $a(t) \bullet h(t) = b(t)$. In the frequency domain, the transfer function multiplies the input, as in $a(f) \times h(f) = b(f)$.

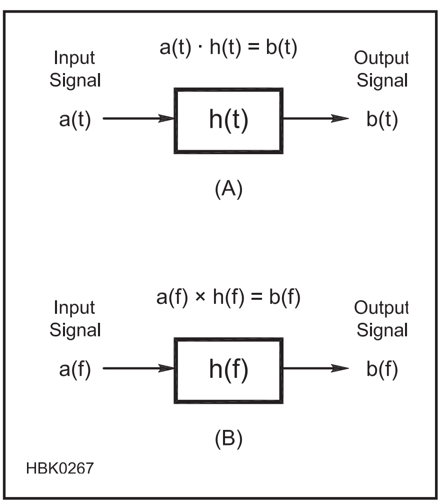

Figure 3.53 — Linear function blocks and transfer functions. The transfer function can be expressed in the time domain (A) or in the frequency domain (B). The transfer function describes how the input signal a(t) or a(f) is transformed into the output signal b(t) or b(f).

While it is not necessary to understand transfer functions mathematically to work with analog circuits, it is useful to realize that they describe how a signal interacts with other signals in an electronic system. In general, the output signal of an analog system depends not only on the input signal at the same time, but also on past values of the input signal. This is a very important concept and is the basis of such essential functions as analog filtering.

CASCADING STAGES

If an analog circuit can be described with a transfer function, a combination of analog circuits can also be described similarly. This description of the combined circuits depends upon the relationship between the transfer functions of the parts and that of the combined circuits. In many cases this relationship allows us to predict the behavior of large and complex circuits from what we know about the parts of which they are made. This aids in the design and analysis of analog circuits.

When two analog circuits are cascaded (the output signal of one stage becomes the input signal to the next stage) their transfer functions are combined. The mechanism of the combination depends on the interaction between the stages. The ideal case is the functions of the stages are completely independent. In other words, when the action of a stage is unchanged, regardless of the characteristics of any stages connected to its input or output.

Just as the signal entering the first stage is modified by the action of the first transfer function, the ideal cascading of analog circuits results in changes produced only by the individual transfer functions. For any number of stages that are cascaded, the combination of their transfer functions results in a new transfer function. The signal that enters the circuit is changed by the composite transfer function to produce the signal that exits in the cascaded circuits.

While each stage in a series may use feedback within itself, feedback around more than one stage may create a function — and resultant performance — different from any of the included stages. Examples include oscillation or negative feedback.

AMPLIFIER FREQUENCY RESPONSE

At higher frequencies a typical amplifier acts as a low-pass filter, decreasing amplification with increasing frequency. Signals within a range of frequencies are amplified consistently but outside that range the amplification changes. At high gains many amplifiers work properly only over a small range of frequencies. The combination of gain and frequency response is often expressed as a *gain-bandwidth product*. For many amplifiers, gain times bandwidth is approximately constant. As gain increases, bandwidth decreases, and vice versa.

Performance at lower frequencies depends on whether the amplifier is *dc- or ac-coupled*. Coupling refers to the transfer of signals between circuits. A dc-coupled amplifier amplifies signals at all frequencies down to dc. An ac-coupled amplifier acts as a high-pass filter, decreasing amplification as the frequency decreases toward dc. Ac-coupled circuits usually use capacitors to allow ac signals to flow between stages while blocking the dc bias voltages of the circuit.

INTERSTAGE LOADING AND IMPEDANCE MATCHING

Every two-port network can be further defined by its input and output impedance. The input impedance is the opposition to current, as a function of frequency, seen when looking into the input port of the network. Likewise, the output impedance is similarly defined when looking back into a network through its output port.

If the transfer function of a stage changes when it is cascaded with another stage, we say that the second stage has *loaded* the first stage. This often occurs when an appreciable amount of current passes from one stage to the next. Interstage loading is related to the relative output impedance of a stage and the input impedance of the stage that is cascaded after it.

In some applications, the goal is to transfer a maximum amount of power from the output of the stage to a load connected to the output. In this case, the output impedance of the stage is *matched* or transformed to that of the load (or vice versa). This allows the stage to operate at its optimum voltage and current levels. In an RF amplifier, the impedance at the input of the transmission line feeding an antenna is transformed by means of a matching network to produce the resistance the amplifier needs in order to efficiently produce RF power.

In contrast, it is the goal of most analog signal processing circuitry to modify a signal rather than to deliver large amounts of power. Thus, an impedance-matched condition may not be required. Instead, current between stages can be minimized by using mismatched impedances. Ideally, if the output impedance of a network is very low and the input impedance of the following stage is very high, very little current will pass between the stages, and interstage loading will be negligible.

NOISE

Generally we are only interested in specific man-made signals. Nature allows many signals to combine, however, so the desired signal becomes combined with many other unwanted signals, both man-made and naturally occurring. The broadest definition of noise is any signal that is not the one in which we are interested. One of the goals of signal processing is to separate desired signals from noise. (See the **RF Techniques** chapter for a more complete discussion on noise, including calculation and use of noise factor and noise figure.)

One form of noise that occurs naturally and must be dealt with in low-level processing circuits is called *thermal noise*, or *Johnson noise*. Thermal noise is produced by random motion of free electrons in conductors and semiconductors. This motion increases as temperature increases, hence the name. This kind of noise is present at all frequencies and is proportional to temperature. Naturally occurring noise can be reduced either by decreasing the circuit's bandwidth or by reducing the temperature in the system. Thermal noise voltage and current vary with the circuit impedance and follow Ohm's Law. Low-noise-amplifier-design techniques are based on these relationships.

Analog signal processing stages are characterized in part by the noise they add to a signal. A distinction is made between enhancing existing noise (such as amplifying it) and adding new noise. The noise added by analog signal processing is commonly quantified by the *noise factor, f*. Noise factor is the ratio of the total output noise power (thermal noise plus noise added by the stage) to the amplifier input noise power when the termination is at the standard temperature of 290 K (17 °C). When the noise factor is expressed in dB, we often call it *noise figure, NF*.

In a system of many cascaded signal processing stages, such as a communications receiver, each stage contributes to the total noise of the system. The noise factor of the first stage dominates the noise factor of the entire system because noise added at the first stage is then multiplied by each following stage. Noise added by later stages is not multiplied to the same degree and so is a smaller contribution to the overall noise at the output. Designers try to optimize system noise factor by using a first stage with a minimum possible noise factor and maximum possible gain. (Caution: A circuit that overloads is often as useless as one that generates too much noise.)

BUFFERING

It is often necessary to isolate the stages of an analog circuit. This isolation reduces the loading, coupling and feedback between stages. It is often necessary to connect circuits that operate at different impedance levels between stages. An intervening stage, a type of amplifier called a *buffer*, is often used for this purpose. If signal level is sufficient, an attenuation can also serve as a buffer at the expense of some signal loss.

Buffers can have high values of amplification but this is unusual. A buffer used for impedance transformation generally has a low or unity gain. In some circuits, notably power amplifiers, the desired goal is to deliver a maximum amount of power to the output device (such as a speaker or an antenna). Matching the amplifier output impedance to the output-device impedance provides maximum power transfer. A buffer amplifier may be just the circuit for this type of application. Such amplifier circuits must be carefully designed to avoid distortion. Combinations of buffer stages can also be effective at isolating the stages from each other and making impedance transformations, as well.

TRANSITION TIME

The transition between the binary 0 and binary 1 states of a digital signal or circuit does not occur instantly. There is a *transition time* between states. This transition time is a result of the time it takes to charge or discharge the stray capacitance in wires and other components because voltage cannot change instantaneously across a capacitor. Stray inductance in the wires also has an effect because the current through an inductor can't change instantaneously. The rate at which the digital circuit's output transistors can change state may also be a factor.

Distributed inductances and capacitances in wires or PC-board traces may cause rise and fall times to increase as the pulse moves away from the source. Ringing and reflections may occur due to transmission line effects as discussed in the **Transmission Lines** chapter.

The transition from a 0 to a 1 state is called the *rise time*, and is usually specified as the

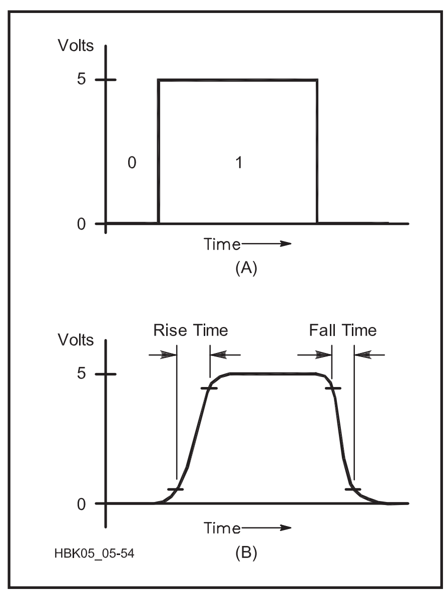

Figure 3.54 — (A) An ideal digital pulse and (B) a typical actual pulse, showing the gradual transition between states.

time for the pulse to rise from 10% of its final value to 90% of its final value. Similarly, the transition from a 1 to a 0 state is called the *fall time,* with a similar 10% to 90% definition. Note that these times need not be the same. **Figure 3.54A** shows an ideal signal, or *pulse*, with zero-time switching. Figure 3.54B shows a typical pulse, as it changes between states in a smooth curve.

The faster the rise or fall time, the wider the bandwidth of signals associated with the transition. These signals can be radiated, causing noise and interfering signals in receivers or sensitive circuits. The general rule of thumb for the bandwidth of digital signals is:

$$BW(GHz) = \frac{0.35}{RT(ns)} \text{ and}$$

$$RT(ns) = \frac{0.35}{BW(GHz)}$$

Rise and fall times for digital integrated circuits vary with the logic family used and the location in a circuit. Typical values of transition time are in the range of microseconds (4000-series CMOS with high-impedance loads) to sub-nanosecond range for current digital logic technology.

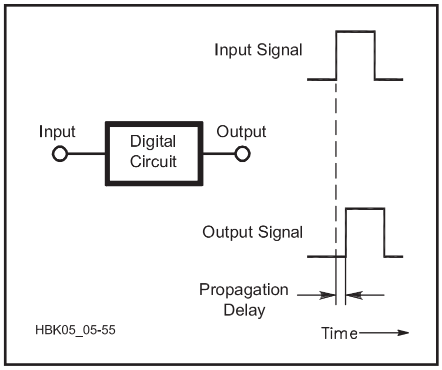

Figure 3.55 — Propagation delay in a digital circuit.

PROPAGATION DELAY

Rise and fall times only describe a relationship within a pulse. For a circuit, a pulse input into the circuit must propagate through the circuit; in other words it must pass through each component in the circuit until eventually it arrives at the circuit output. The time delay between providing an input to a circuit and seeing a response at the output is the *propagation delay* and is illustrated by **Figure 3.55**.

For modern switching logic, typical propagation delay values are in the sub-nanosecond range. (It is useful to remember that the propagation delay along a wire or printed-circuit-board trace adds about 1.0 to 1.5 ns per inch.) Propagation delay is the result of cumulative transition times as well as transistor switching delays, reactive element charging times and the time for signals to travel through wires. In complex circuits, different propagation delays through different paths can cause problems when pulses must arrive somewhere at exactly the same time.

The effect of these delays on digital devices can be seen by looking at the speed of the digital pulses. Most digital devices and all PCs use *clock pulses*. If two pulses are supposed to arrive at a logic circuit at the same time, or very close to the same time, the path length for the two signals cannot be any different than two to three inches. This can be a very significant design problem for high-speed logic designs.

3.10 Electromagnetic Waves

Audio or *sonic* energy is the energy imparted by the mechanical movement of a medium, which can be air, metal, water or even the human body. Sound that humans can hear normally requires the movement of air between 20 Hz and 20 kHz, although the human ear loses its ability to detect the extremes of this range as we age. Some animals, such as elephants, can detect air vibrations well below 20 Hz, while others, such as dogs and cats, can detect air vibrations well above 20 kHz.

Electrical circuits do not directly produce air vibrations. Sound production requires a *transducer*, a device to transform one form of energy into another form of energy; in this case electrical energy into sonic energy. The speaker and the microphone are the most common audio transducers. There are numerous ultrasonic transducers for various applications.

Radio frequency energy exists at frequencies for which it is practical to generate and detect waves that exist independently of the movement of electrical charge, such as a radio signal. Like sonic energy, a transducer — an antenna — is required to convert the electrical energy in a circuit to electromagnetic waves. In a physical circuit, such as a wire, electromagnetic energy exists as both electromagnetic waves and the physical movement of electrical charge.

Electromagnetic waves have been generated and detected in many forms with frequencies from below 1 Hz to above 10^{12} GHz, including at the higher frequencies infrared, visible, and ultraviolet light, and a number of energy forms of greatest interest to physicists and astronomers. **Table 3.5** provides a brief glimpse at the total spectrum of electromagnetic energy. The *radio spectrum* is generally

Table 3.5
Key Regions of the Electromagnetic Energy Spectrum

Region Name	*Frequency Range*		
Radio frequencies*	3.0×10^{3} Hz	to	3.0×10^{11} Hz
Infrared	3.0×10^{11} Hz	to	4.3×10^{14} Hz
Visible light	4.3×10^{14} Hz	to	7.5×10^{14} Hz
Ultraviolet	7.5×10^{14} Hz	to	6.0×10^{16} Hz
X-rays	6.0×10^{16} Hz	to	3.0×10^{19} Hz
Gamma rays	3.0×10^{19} Hz	to	5.0×10^{20} Hz
Cosmic rays	5.0×10^{20} Hz	to	8.0×10^{21} Hz

Range Name	*Abbreviation*	*Frequency Range*
Very Low Frequency	VLF	3 kHz - 30 kHz
Low Frequency	LF	30 kHz - 300 kHz
Medium Frequency	MF	300 kHz - 3 MHz
High Frequency	HF	3 MHz - 30 MHz
Very High Frequency	VHF	30 MHz - 300 MHz
Ultra High Frequency	UHF	300 MHz - 3 GHz
Super High Frequency	SHF	3 GHz - 30 GHz
Extremely High Frequency	EHF	30 GHz - 300 GHz

*Note: The range of radio frequencies can also be written as 3 kHz to 300 GHz considered to begin around 3 kHz and end at infrared light.

Within the part of the electromagnetic energy spectrum of most interest to radio amateurs, frequencies have been classified into groups and given names. Table 3.5 provides a reference list of these classifications. To a significant degree, the frequencies within each group exhibit similar properties, both in circuits and as RF waves. For example, HF or high frequency waves, with frequencies from 3 to 30 MHz, all exhibit ionospheric refraction that permits regular long-range radio communications. This property also applies occasionally both to MF (medium frequency) and to VHF (very high frequency) waves, as well.

Despite the close relationship between electromagnetic energy and waves, it remains important to distinguish the two. To a circuit producing or amplifying a 15-kHz alternating current, the ultimate transformation and use of the electrical energy may make no difference to the circuit's operation. By choosing the right transducer, one can produce either a sonic wave or an electromagnetic wave — or both. Such is a common problem of video monitors and switching power supplies; forces created by the ac currents cause electronic parts both to vibrate audibly *and* to radiate electromagnetic energy.

3.10.1 Electric and Magnetic Fields

Electrical and magnetic energy are invisible — you can't detect them with any of your senses. All you can do is observe their effects such as when a resistor gets hot, a motor spins, or an electromagnet picks up iron or steel. The energy exists as a *field* — a region of space in which energy is stored and through which electrical and magnetic forces act. (For serious inquiries as to the nature of fields, see **en.wikipedia.org/wiki/Electric_field** and **en.wikipedia.org/wiki/Magnetic_field**.)

You are already quite familiar with fields in the form of gravity. You are being pulled down toward the Earth as you read this because you are in the Earth's *gravitational field*. Because your body has mass it interacts with the gravitational field in such a way that the Earth attracts you. (You have your own gravitational field, too, but many orders of magnitude smaller than that of the Earth.) Think of a bathroom scale as a "gravitational voltmeter" that instead of reading "volts," reads "pounds." The heavier something is, the stronger the Earth is attracting it. Weight is the same as force. (Metric scales provide readings in kilograms, a unit of mass. To do so, the scales assume a standard strength for gravity in order to convert weight [a force] to an equivalent mass in kilograms.)

This field makes you do work, such as when you climb stairs. Work has a precise definition when it comes to fields: *Work* equals force times distance moved in the direction of the field's force. For example, let's say you pick up a mass — a stone that weighs 1 pound — and lift it to a shelf 10 feet above where it previously lay. How much work did you do? You moved a weight of 1 pound a distance of 10 feet against the attraction of the field, so you have done 10 foot-pounds of work. (It doesn't count if you move the stone sideways instead of vertically.)

What did that work accomplish? You stored gravitational energy in the stone equal to the amount of work that you performed. This stored energy is called *potential energy*, whether gravitational, electrical or magnetic. You could store the same amount of gravitational energy by lifting a 10-pound stone 1 foot or by lifting a stone that weighs 1/10th of a pound 100 feet. If you drop the stone (or it falls off the shelf), the same amount of potential energy is converted back to *kinetic energy* as the stone moves toward the Earth in the gravitational field.

In electronics we are interested in two types of fields: *electric fields* and *magnetic fields*. Electric fields can be detected as voltage differences between two points. The electric field's analog to gravitational mass is electric charge. Every electric charge has its own electric field, just as every mass has its own gravitational field. The more charged a body is, the "heavier" it is in terms of an electric field. Just as a body with mass feels a force to move in a gravitational field, so does an electric charge in an electric field. Electrical energy is stored by moving electrical charges apart so that there is a voltage between them. If the field does not change with time, it is called an *electrostatic field*.

Magnetic energy is detected by its effects on moving electrical charges or current. Magnetic energy is stored through the motion of electric charge (current) creating a magnetic field. Magnetic fields that don't change with time, such as from a stationary permanent magnet, are called *magnetostatic fields*.

The potential energy is released by allowing the charges to move in the field. For example, electric energy is released when a current flows from a charged-up capacitor. Magnetic energy stored by current flowing in an inductor is released when the current is allowed to change, such as a relay's armature does when the coil is de-energized.

3.10.2 Electromagnetic Fields and Waves

An *electromagnetic field* is created when the potential energy stored in an electric field or magnetic field changes. The changing electric and magnetic fields create *electromagnetic waves* (what we call *radio waves*) that propagate through space carrying both electric and magnetic energy. The electric and magnetic fields in the wave vary with time in a sinusoidal pattern. The potential energy is shared between the electric and magnetic fields making up the electromagnetic field.

The *field strength* of an electromagnetic wave can be measured either by the electric field (volts/meter) or the magnetic field (amps/meter). Usually the wave's field strength is stated only in volts/meter since that is easier to measure than the magnetic field. If we multiply the electric and magnetic field strengths, we have power per unit area:

$$\frac{E}{m} \times \frac{H}{m} = \frac{watts}{m^2}$$

MAXWELL'S EQUATIONS

The basic theory of electromagnetic fields was established by James Maxwell in 1860-1864. The behavior of the fields are described by the four equations known today as Maxwell's equations. (This form of the equations was actually produced by Oliver Heaviside in his work prior to 1890.) The existence of the electromagnetic waves predicted by Maxwell in 1864 was demonstrated by Heinrich Hertz in 1886. For more information about the experiment, see the References and Bibliography listing at the end of this chapter for G.S. Smith's article analyzing Hertz's experiments.

A discussion of Maxwell's equations and their application to electromagnetic simulation is presented in the **Electronic Design Automation** chapter. A more complete treatment is provided by Bob Zavrel, W7SX, in his book *Antenna Physics: An Introduction* (see the References and Bibliography section of this chapter). A summary treatment of the equations in a pair of Hand-On Radio columns by Ward Silver, NØAX, is included in the online content. These include an explanation of the vector calculus concepts of gradient, divergence, and curl, as well as illustrating how waves are created by moving electric charge.

3.10.3 Electromagnetic Wave Propagation

All electromagnetic energy has one thing in common: it travels, or *propagates,* at the speed of light, abbreviated *c*. This speed is approximately 300,000,000 (or 3×10^8) meters per second in a vacuum, termed *free space*.

In general, the speed at which electromagnetic waves travel or *propagate* depends on the permittivity and permeability of the medium through which they travel.

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

The speed of light is highest in the vacuum of free space and only slightly lower in air. In materials such as glass or plastic, however, velocity can be quite a bit lower. For example, in polyethylene (commonly used as a center insulator in coaxial cable), the *velocity of propagation* is about two-thirds (67%) of that in free space.

Electromagnetic waves have a wavelength uniquely associated with each possible frequency. (See **Figure 3.56**) The *wavelength* (λ) is the speed of propagation, c, divided by the frequency (f) in hertz.

$$f\,(\text{Hz}) = \frac{3.0\times10^{8}\left(\frac{\text{m}}{\text{s}}\right)}{\lambda\,(\text{m})}$$

and

$$\lambda\,(\text{m}) = \frac{3.0\times10^{8}\left(\frac{\text{m}}{\text{s}}\right)}{f\,(\text{Hz})}$$

Example: What is the frequency of an RF wave with wavelength of 80 meters?

$$f\,(\text{Hz}) = \frac{3.0\times10^{8}\left(\frac{\text{m}}{\text{s}}\right)}{\lambda\,(\text{m})}$$

$$= \frac{3.0\times10^{8}\left(\frac{\text{m}}{\text{s}}\right)}{80.0\ \text{m}}$$

$$= 3.75\times10^{6}\ \text{Hz}$$

This is 3.750 MHz or 3750 kHz, a frequency in the middle of the ham band known as "80 meters."

A similar equation is used to calculate the wavelength of a sound wave in air, substituting the speed of sound instead of the speed of light in the numerator. The speed of propagation of the mechanical movement of air that we call sound varies considerably with air temperature and altitude. The speed of sound at sea level is about 331 m/s at 0 °C and 344 m/s at 20 °C.

To calculate the frequency of an electromagnetic wave directly in kilohertz, change the speed constant to 300,000 (3×10^5) km/s.

$$f\,(\text{kHz}) = \frac{3.0\times10^{5}\left(\frac{\text{km}}{\text{s}}\right)}{\lambda\,(\text{m})}$$

and

$$\lambda\,(\text{m}) = \frac{3.0\times10^{5}\left(\frac{\text{km}}{\text{s}}\right)}{f\,(\text{kHz})}$$

For frequencies in megahertz, change the speed constant to 300 (3×10^2) Mm/s.

$$f\,(\text{MHz}) = \frac{300\left(\frac{\text{Mm}}{\text{s}}\right)}{\lambda\,(\text{m})}$$

and

$$\lambda\,(\text{m}) = \frac{300\left(\frac{\text{Mm}}{\text{s}}\right)}{f\,(\text{MHz})}$$

Stated as it is usually remembered and used, "wavelength in meters equals 300 divided by frequency in megahertz." Assuming the proper units for the speed of light constant simplify the equation.

$$\lambda\,(\text{in m}) = \frac{300}{f\,(\text{in MHz})}$$

$$\text{and } f\,(\text{in MHz}) = \frac{300}{\lambda\,(\text{in m})}$$

Example: What is the wavelength of an RF wave whose frequency is 4.0 MHz?

$$\lambda(\text{m}) = \frac{300}{f\,(\text{MHz})} = \frac{300}{4.0} = 75\ \text{m}$$

At higher frequencies, circuit elements with lengths that are a significant fraction of a wavelength can act like transducers. This property can be useful, but it can also cause problems for circuit operations. Therefore, wavelength calculations are of some importance in designing ac circuits for those frequencies.

3.10.4 Electromagnetic Wave Structure

The waves move through space independently of any component or conductor. The electric and magnetic fields of the wave are oriented at right angles to each other as shown by **Figure 3.57**. The direction of the right angle between the electric and magnetic fields determines the direction the wave travels, as illustrated by Figure 3.50. The term "lines of force" in the figure means the direction in which a force would be felt by an electron (from the electric field) or by a magnet (from the magnetic field).

An important note about electromagnetic waves: The electric and magnetic fields making up the wave are not just perpendicular electric and magnetic fields that simply happen to be in the same place at the same time! The fields are *coupled*; that is they are both aspects of the same entity — the electromagnetic wave. The fields cannot be separated although the energy in the wave can be detected as either electric or magnetic force. The electromagnetic wave is created as a single entity by the motion of electrons, such as in a transmitting antenna.

WAVEFRONTS

To an observer staying in one place, such as a fixed station's receiving antenna, the electric and magnetic fields of the wave appear to oscillate as the wave passes. That is, the fields create forces on electrons in the antenna that increase and decrease in the sine wave pattern. Some of the energy in the propagating wave is transferred to the electrons as the forces from the changing fields cause them to move. This creates a sine wave current in the antenna with a frequency determined by the rate at which the field strength changes in the passing wave.

If the observer is moving along with the wave at the same speed, however, the strength

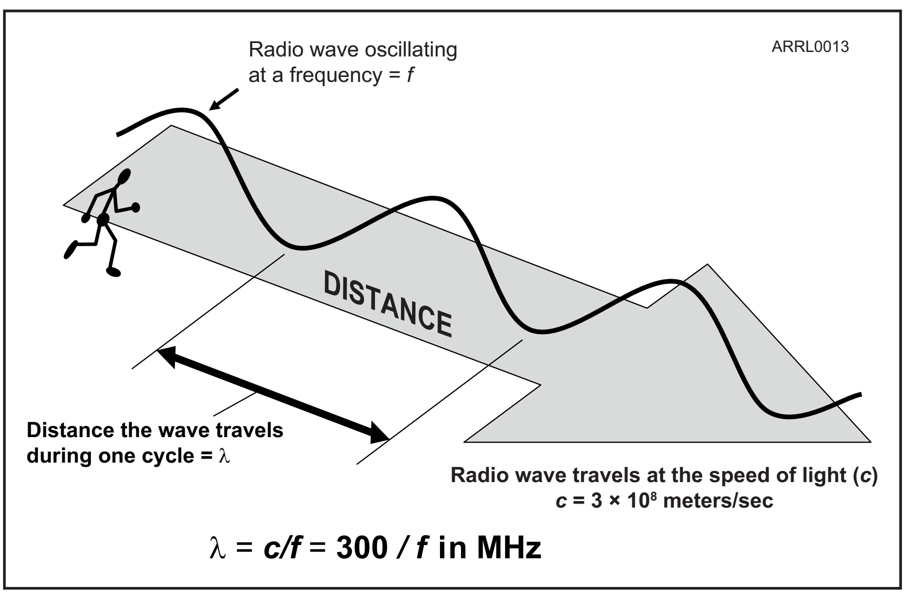

Figure 3.56 — As a radio wave travels, it oscillates at the frequency of the signal. The distance covered by the wave during the time it takes for one complete cycle is its wavelength.

of the fields will not change. To that observer, the electric and magnetic field strengths are fixed, as in a photograph. This is a *wavefront* of the electromagnetic wave — a flat surface or plane moving through space on which the electric and magnetic fields have a constant value as illustrated in Figure 3.57.

Just as an ac voltage is made up of an infinite sequence of instantaneous voltages, each slightly larger or smaller than the next, an infinite number of wavefronts make up an electromagnetic wave, one behind another like a deck of cards. The direction of the wave is the direction in which the wavefronts move. The fields on each successive wavefront have a slightly different strength, so as they pass a fixed location the detected field strength changes as well. The result is that the fixed observer "sees" fields with strengths varying as a sine wave.

Figure 3.58 is a drawing of what would happen if we could suddenly freeze all of the wavefronts in the wave and take measurements of the electric and magnetic field strengths in each. In this example, the electric field is oriented vertically and the magnetic field horizontally. (Each of the vertical lines in the electric field can be thought of as representing an individual wavefront.) **Figure 3.59** illustrates the right-angle relationship of the *E* and *H* fields, and the direction of their motion.

All of the wavefronts are moving in the direction indicated — the whole set of them moves together at the same speed. As the wave — the set of wavefronts — moves past the receive antenna, the varying field strengths of the different wavefronts are perceived as a continuously changing wave. What we call a "wave" is really this entire group of wavefronts moving through space.

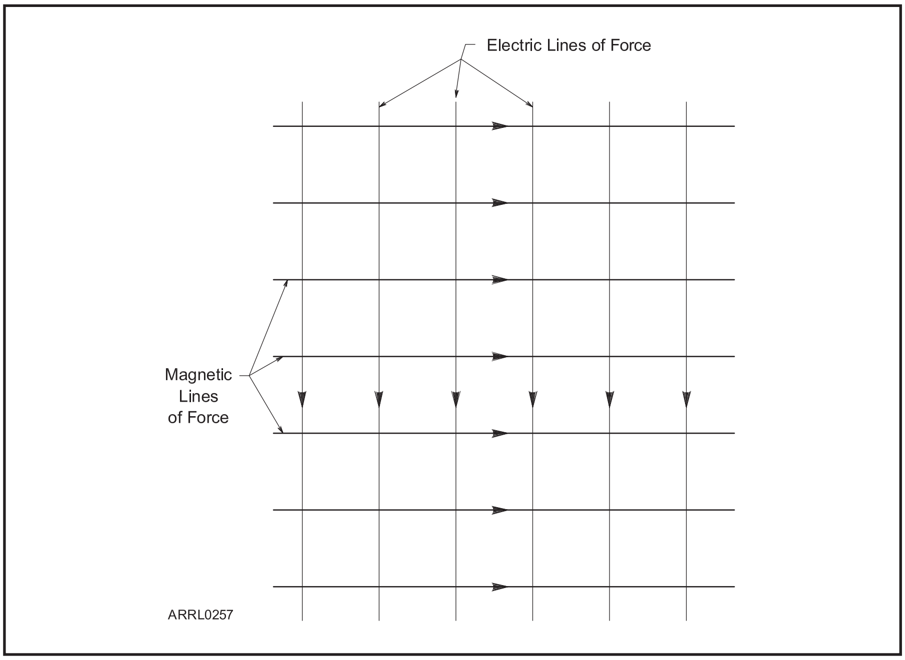

Figure 3.57 — Representation of electric and magnetic lines of force in an electromagnetic wavefront. Arrows indicate the instantaneous directions of the fields for a wavefront in a wave traveling toward you, out of the page. Reversing the direction of either of the fields would also reverse the direction of the wave.

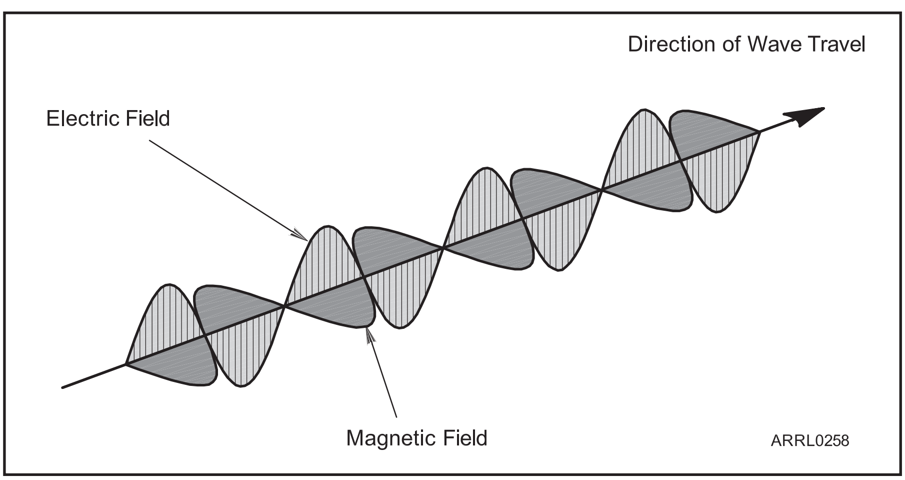

Figure 3.58 — Representation of the magnetic and electric field strengths of a vertically polarized electromagnetic wave. In the diagram, the electric field is oriented vertically and the magnetic field horizontally.

POLARIZATION

The orientation of the pair of fields in an electromagnetic wave can have any orientation with respect to the surface of the Earth, but the electric and magnetic fields will always be at right angles to each other. The orientation of the wave's electric field determines the *polarization* of the wave. If the electric field's lines of force are parallel to the surface of the Earth (meaning those of the magnetic field are perpendicular to the Earth), the wave is *horizontally polarized*. Conversely, if the magnetic field's lines of force are parallel to the surface of the Earth (and those of the electric field are perpendicular to the Earth), the wave is *vertically polarized*. Knowing the polarization of the wave allows the receiving antenna to be oriented so that the passing wave will exert the maximum force on the electrons in the antenna, maximizing received signal strength.

For the most part, the wave's polarization is determined by the type of transmitting antenna and its orientation. For example, a Yagi antenna with its elements parallel to the

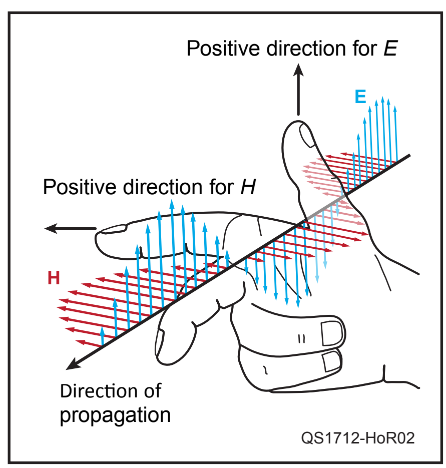

Figure 3.59 — The right-hand rule shows how to determine direction of propagation of an electromagnetic wave. Point your thumb in the positive direction for the *E* field, your Index finger in the positive direction for *H*, and your middle finger will point in the direction the wave is traveling.

Earth's surface transmits a horizontally polarized wave. On the other hand, an amateur mobile whip antenna, mounted vertically on an automobile, radiates a vertically polarized wave. If a vertically polarized antenna is used to receive a horizontally polarized radio wave (or vice versa), received signal strength can be reduced by more than 20 dB as compared to using an antenna with the same polarization as the wave. This is called *cross-polarization*.

It is also possible to generate electromagnetic waves in which the orientation of successive wavefronts rotates around the direction of travel — both the electric and magnetic fields. This is called *circular polarization*. Imagine the wave of Figure 3.51 being twisted so at one point the direction of the electric field is horizontal and a bit further along the wave it is vertical. As the twisted, circularly polarized wave passes the receiving antenna, the polarization of its fields will appear to rotate. The rate at which the polarization changes and the direction of the rotation — *right-handed* or *left-handed* — is determined by the construction of the transmitting antenna. Note that the electric and magnetic fields rotate together so the right-angle between them remains fixed. Polarization that does not rotate is called *linear polarization* or *plane polarization*. Horizontal and vertical polarization are examples of linear polarization.

To best receive a circularly polarized wave, the structure of the receiving antenna should match that of the transmitting antenna. It is particularly helpful to use circular polarization in satellite communication, where polarization tends to shift with the orientation of the satellite and the path of its signal through the atmosphere. Circular polarization is usable with linearly polarized antennas at one end of the signal's path. There will be some small loss in this case, however.

IMPEDANCE OF FREE SPACE

Maxwell's equations provide the relationships (direction and ratio) between the magnetic and electric fields associated with an electromagnetic wave. Far from an antenna or other distorting surfaces, the wave is treated as a *plane* wave in which the wavefronts are infinite, flat planes. In a plane wave, the ratio of the two fields' amplitudes remains constant and the two fields are always at right angles. (There are exceptions but they are not discussed here.)

Since the ratio of the two fields is constant in free space, that gives rise to the idea of an intrinsic impedance. If the impedance is zero or infinite, the magnetic and electric fields would have to be infinite or zero and there could be no electromagnetic radiation.

The impedance can be derived from the permittivity (ε_0) and permeability (μ_0) of free space:

$$\varepsilon_0 = \frac{\text{F(farads)}}{\text{meter}} \text{ and } \mu_0 = \frac{\text{H(henries)}}{\text{meter}}$$

Taking the ratio of the two, noting that farads have units of joules/volt^2 and henries have units of joules/ampere^2:

$$\sqrt{\frac{\mu_0}{\varepsilon_0}} = \sqrt{\frac{H}{F}} = \sqrt{\frac{J/I^2}{J/V^2}} = \sqrt{\frac{V^2}{I^2}} = \frac{V}{I} \approx 377\Omega$$

It is interesting to remember that the speed of light is:

$$c = \frac{1}{\sqrt{\varepsilon_0 \mu_0}}$$

This links the fundamental electrical and magnetic constants to both velocity of electromagnetic waves and an impedance describing how energy is distributed between the electric and magnetic fields.

3.11 References and Bibliography

BOOKS AND ARTICLES

Alexander and Sadiku, *Fundamentals of Electric Circuits* 7th ed (McGraw-Hill, 2021).

Banzhaf, W., WB1ANE, *Understanding Basic Electronics*, 2nd ed (ARRL, 2010).

Horowitz and Hill, *The Art of Electronics* 3rd ed (Cambridge University Press, 2015).

Kaplan, S., *Wiley Electrical and Electronics Dictionary* (Wiley Press, 2004).

Millman and Grabel, *Microelectronics: Digital and Analog Circuits and Systems* (McGraw-Hill, 1988).

Ohira, T. "What in the World is Q?," *IEEE Microwave Magazine*, Jun 2016, pp 42–49.

Orr, W., *Radio Handbook* 23rd ed (Newnes, 1997).

Rautio, J., AJ3K, "The Long Road to Maxwell's Equations," IEEE Spectrum, **spectrum.ieee.org/tech-history/dawn-of-electronics/the-long-road-to-maxwells-equations**

Silver, W., NØAX, "Experiment #74 — Resonant Circuits," *QST*, Mar 2009, pp 72-73.

Silver, W., NØAX, "Experiment #178 — Maxwell's Equations — Grad, Div, and Curl," *QST*, Nov 2017, pp 77-78.

Silver, W., NØAX, "Experiment #179 — Maxwell's Equations — The Wave Emerges," *QST*, Dec 2017, pp 59-60.

Smith, GS, "Analysis of Hertz's Experimentum Crucis on Electromagnetic Waves," IEEE Antennas and Propagation Magazine, Oct 2016, pp. 96-108.

Terman, F., *Radio Engineers' Handbook* (McGraw-Hill, 1943).

Zavrel, R., W7SX, *Antenna Physics: An Introduction*, 2nd ed (ARRL, 2020).

WEBSITES

Oliver Heaviside — **en.wikipedia.org/wiki/Oliver_Heaviside**

History of Maxwell's Equations — **en.wikipedia.org/wiki/History_of_Maxwell%27s_equations**

Rohde & Schwarz, "dB or not dB? Everything you ever wanted to know about decibels but were afraid to ask…," Application Note 1MA98, **www.rohde-schwarz.com/us/applications/db-or-not-db-educational-note_230850-15534.html**