Contents

Chapter 2

Electrical Fundamentals

Building on the work of numerous earlier contributors (most recently Roger Taylor, K9ALD's section on DC Circuits and Resistance), this chapter has been updated by Ward Silver, NØAX. Look for the section "Radio Mathematics," added at the end of this chapter, for a list of online math resources and sections on some of the mathematical techniques used in radio and electronics. Glossaries follow the sections on dc and ac theory, as well.

2.1 Introduction to Electricity

The *atom* is the primary building block of matter and is made up of a *nucleus*, containing *protons* and *neutrons*, surrounded by *electrons*. Protons have a positive electrical charge, electrons a negative charge, and neutrons have no electrical charge. An *element* (or *chemical element*) is a type of atom that has a specific number of protons, the element's *atomic number*. Each different element, such as iron, oxygen, silicon, or bromine has a distinct chemical and physical identity determined primarily by the number of protons. A *molecule* is two or more atoms bonded together and acting as a single particle.

Unless modified by chemical, mechanical, or electrical processes, all atoms are electrically neutral because they have the same number of electrons as protons. If an atom loses electrons, it has more protons than electrons and thus has a net positive charge. If an atom gains electrons, it has more electrons than protons and a net negative charge. Atoms or molecules with a positive or negative charge are called *ions*. Electrons not bound to any atom, or *free electrons*, can also be considered as ions because they have a negative charge.

2.1.1 Electric Charge, Voltage and Current

Any piece of matter that has a net positive or negative electrical charge is said to be *electrically charged*. An electrical force exists between electrically charged particles, pushing charges of the same type apart (like charges repel each other) and pulling opposite charges together (opposite charges attract). This is the *electromotive force* (or EMF), also referred to as *voltage* or *potential*. The higher the EMF's voltage, the stronger is its force on an electrical charge.

Under most conditions, the number of positive and negative charges in any volume of space is very close to balanced and so the region has no net charge. When there are extra positive ions in one region and extra negative ions (or electrons) in another region, the resulting EMF attracts the charges toward each other. The direction of the force, from the positive region to the negative region, is called its *polarity*. Because EMF results from an imbalance of charge between two regions, its voltage is always measured between two points, with positive voltage defined as being in the direction from the positively-charged to the negatively-charged region.

If there is no path by which electric charge can move in response to an EMF (called a *conducting path*), the charges cannot move together and so remain separated. If a conducting path is available, then the electrons or ions will flow along the path, neutralizing the net imbalance of charge. The movement of electrical charge is called *electric current*. Materials through which current flows easily are called *conductors*. Most metals, such as copper or aluminum are good conductors. Materials in which it is difficult for current to flow are *insulators*. *Semiconductors*, such as silicon or germanium, are materials with much poorer conductivity than metals. Semiconductors can be chemically altered to acquire properties that make them useful in solid-state devices such as diodes, transistors and integrated circuits.

Voltage differences can be created in a variety of ways. For example, chemical ions can be physically separated to form a battery. The resulting charge imbalance creates a voltage difference at the battery terminals so that if a conductor is connected to both terminals at once, electrons flow between the terminals and gradually eliminate the charge imbalance, discharging the battery's stored energy. Mechanical means such as friction (static electric-

ity, lightning) and moving conductors in a magnetic field (generators) can also produce voltages. Devices or systems that produce voltage are called *voltage sources*.

2.1.2 Electronic and Conventional Current

Electrons move in the direction of positive voltage — this is called *electronic current*. *Conventional current* takes the other point of view — of positive charges moving in the direction of negative voltage. Conventional current was the original model for electricity and results from an arbitrary decision made by Benjamin Franklin in the 18th century when the nature of electricity and atoms was still unknown. It can be imagined as electrons flowing "backward" and is completely equivalent to electronic current.

Conventional current is used in nearly all electronic literature and is the standard used in this book. The direction of conventional current direction establishes the polarity for most electronics calculations and circuit diagrams. The arrows in the drawing symbols for transistors point in the direction of conventional current, for example.

2.1.3 Units of Measurement

To measure electrical quantities, certain definitions have been adopted. Charge is measured in *coulombs* (C) and represented by *q* in equations. One coulomb is equal to 6.25 × 10^{18} electrons (or protons). Current, the flow of charge, is measured in *amperes* (A) and represented by *i* or I in equations. One ampere represents one coulomb of charge flowing past a point in one second and so amperes can also be expressed as coulombs per second. Electromotive force is measured in *volts* (V) and represented by *e*, E, *v,* or V in equations. One volt is defined as the electromotive force between two points required to cause one ampere of current to do one *joule* (measure of energy) of work in flowing between the points. Voltage can also be expressed as joules per coulomb.

2.1.4 Series and Parallel Circuits

A *circuit* is any conducting path through which current can flow between two points of that have different voltages. An *open circuit* is a circuit in which a desired conducting path is interrupted, such as by a broken wire or a switch. A *short circuit* is a circuit in which a conducting path allows current to flow directly between the two points at different voltages.

The two fundamental types of circuits are shown in **Fig 2.1**. Part A shows a *series circuit* in which there is only one current path. The current in this circuit flows from the voltage source's positive terminal (the symbol for a battery is shown with its voltage polarity shown as + and –) in the direction shown by the arrow through three *resistors* (electronic components discussed below) and back to the battery's negative terminal. Current is the same at every point in a series circuit.

Part B shows a *parallel circuit* in which there are multiple paths for the current to take. One terminal of both resistors is connected to the battery's positive terminal. The other terminal of both resistors is connected to the battery's negative terminal. Current flowing out of the battery's positive terminal divides into smaller currents that flow through the individual resistors and then recombine at the battery's negative terminal. All of the components in a parallel circuit experience the same voltage. All circuits are made up of series and parallel combinations of components and sources of voltage and current.

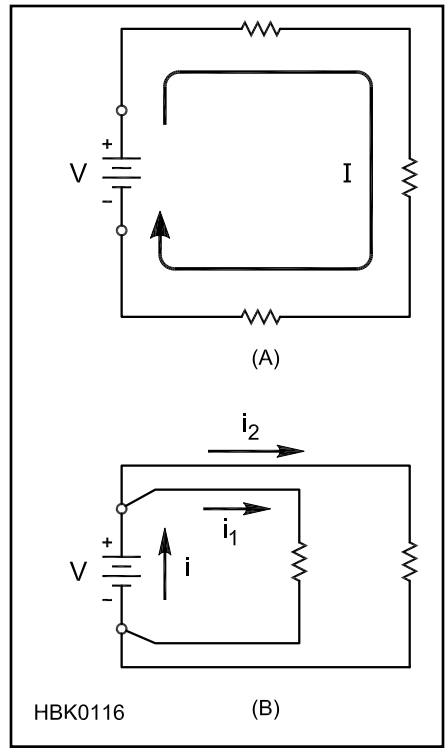


Fig 2.1 — A series circuit (A) has the same current through all components. Parallel circuits (B) apply the same voltage to all components.

Schematic Diagrams

The drawing in Fig 2.1 is a *schematic diagram*. Schematics are used to show the electrical connections in a circuit without requiring a drawing of the actual components or wires, called a *pictorial diagram*. Pictorials are fine for very simple circuits like these, but quickly become too detailed and complex for everyday circuits. Schematics use lines and dots to represent the conducting paths and connections between them. Individual electrical devices and electronic components are represented by *schematic symbols* such as the resistors shown here. A set of the most common schematic symbols is provided in the **Component Data and References** chapter. You will find additional information on reading and drawing schematic diagrams in the ARRL Web site Technology section at **www.arrl.org/circuit-construction**.

2.1.5 Glossary — DC and Basic Electricity

Alternating current (ac) — A flow of charged particles through a conductor, first in one direction, then in the other direction.

Ampere — A measure of flow of charged particles per unit of time. One ampere (A) represents one coulomb of charge flowing past a point in one second.

Atom — The smallest particle of matter that makes up a distinct chemical element. Atoms consist of protons and neutrons in the central region called the nucleus, with electrons surrounding the nucleus.

Circuit — Conducting path between two points of different voltage. In a *series circuit*, there is only one current path. In a *parallel circuit*, there are multiple current paths.

Conductance (G) — The reciprocal of resistance, measured in siemens (S).

Conductor — Material in which electrons or ions can move easily.

Conventional Current — Current defined as the flow of positive charges in the direction of positive to negative voltage. Conventional current flows in the opposite direction of electronic current, the flow of negative charges (electrons) from negative to positive voltage.

Coulomb — A unit of measure of a quantity of electrically charged particles. One coulomb (C) is equal to 6.25 × 10^{18} electrons.

Current (I) — The movement of electrical charge, measured in amperes and represented by *i* in equations.

Direct current (dc) — A flow of charged particles through a conductor in one direction only.

Electronic current — see **Conventional Current**

EMF — Electromotive Force is the term used to define the force of attraction or repulsion between electrically-charged regions. Also called *voltage* or *potential*.

Energy — Capability of doing work. It is usually measured in electrical terms as the number of watts of power consumed during a specific period of time, such as watt-seconds or kilowatt-hours.

Insulator — Material in which it is difficult for electrons or ions to move.

Ion — Atom or molecule with a positive or negative electrical charge.

Joule — Measure of a quantity of energy. One joule is defined as one newton (a measure of force) acting over a distance of one meter.

Ohm — Unit of resistance. One ohm is defined as the resistance that will allow one ampere of current when one volt of EMF is impressed across the resistance.

Polarity — The direction of EMF or voltage, from positive to negative.

Potential — See **EMF**.

Power — Power is the rate at which work is done. One watt of power is equal to one volt of EMF causing a current of one ampere through a resistor.

Resistance (R) — Opposition to current by conversion into other forms of energy, such as heat, measured in ohms (Ω).

Volt, voltage — See **EMF**.

Voltage source — Device or system that creates a voltage difference at its terminals.

2.2 Resistance and Conductance

2.2.1 Resistance

Any conductor connected to points at different voltages will allow current to pass between the points. No conductor is perfect or lossless, however, at least not at normal temperatures. The moving electrons collide with the atoms making up the conductor and lose some of their energy by causing the atoms to vibrate, which is observed externally as heat. The property of energy loss due to interactions between moving charges and the atoms of the conductor is called *resistance*. The amount of resistance to current is measured in *ohms* (Ω) and is represented by *r* or R in equations.

Suppose we have two conductors of the same size and shape, but of different materials. Because all materials have different internal structures, the amount of energy lost by current flowing through the material is also different. The material's ability to impede current flow is its *resistivity*. Numerically, the resistivity of a material is given by the resistance, in ohms, of a cube of the material measuring one centimeter on each edge. The symbol for resistivity is the Greek letter rho, ρ.

The longer a conductor's physical path, the higher the resistance of that conductor. For direct current and low-frequency alternating currents (up to a few thousand hertz) the conductor's resistance is inversely proportional to the cross-sectional area of the conductor. Given two conductors of the same material and having the same length, but differing in cross-sectional area, the one with the larger area (for example, a thicker wire or sheet) will have the lower resistance.

One of the best conductors is copper, and it is frequently convenient to compare the resistance of a material under consideration with that of a copper conductor of the same size and shape. **Table 2.1** gives the ratio of the resistivity of various conductors to the resistivity of copper.

2.2.2 Conductance

The reciprocal of resistance (1/R) is *conductance*. It is usually represented by the symbol G. A circuit having high conductance has low resistance, and vice versa. In radio work, the term is used chiefly in connection with electron-tube and field-effect transistor characteristics. The units of conductance are siemens (S). A resistance of 1 Ω has a conductance of 1 S, a resistance of 1000 Ω has a conductance of 0.001 S, and so on. A unit frequently used in regards to vacuum tubes and the field-effect transistor is the μS or one millionth of a siemens. It is the conductance of a 1-MΩ resistance. Siemens have replaced the obsolete unit *mho* (abbreviated as an upside-down Ω symbol).

2.2.3 Ohm's Law

The amount of current that will flow through a conductor when a given EMF is applied will vary with the resistance of the conductor. The lower the resistance, the greater the current for a given EMF. One ohm is defined as the amount of resistance that allows one ampere of current to flow between two points that have a potential difference of one volt. This proportional relationship is known as *Ohm's Law*:

$$R = \frac{E}{I} \qquad (1)$$

where

R = resistance in ohms,
E = potential or EMF in volts and
I = current in amperes.

Transposing the equation gives the other common expressions of Ohm's Law as:

Table 2.1
Relative Resistivity of Metals

Material	*Resistivity Compared to Copper*
Aluminum (pure)	1.60
Brass	3.7-4.90
Cadmium	4.40
Chromium	8.10
Copper (hard-drawn)	1.03
Copper (annealed)	1.00
Gold	1.40
Iron (pure)	5.68
Lead	12.80
Nickel	5.10
Phosphor bronze	2.8-5.40
Silver	0.94
Steel	7.6-12.70
Tin	6.70
Zinc	3.40

The Origin of Unit Names

Many units of measure carry names that honor scientists who made important discoveries in or advanced the state of scientific knowledge of electrical and radio phenomena. For example, Georg Ohm (1787-1854) discovered the relationship between current, voltage and resistance that now bears his name as Ohm's Law and as the unit of resistance, the ohm. The following table lists the most common electrical units and the scientists for whom they are named. You can find more information on these and other notable scientists in encyclopedia entries on the units that bear their names.

Electrical Units and Their Namesakes

Unit	*Measures*	*Named for*
Ampere	Current	Andree Ampere 1775 -1836
Coulomb	Charge	Charles Coulomb 1736-1806
Farad	Capacitance	Michael Faraday 1791-1867
Henry	Inductance	Joseph Henry 1797-1878
Hertz	Frequency	Heinrich Hertz 1857-1894
Ohm	Resistance	Georg Simon Ohm 1787-1854
Watt	Power	James Watt 1736-1819
Volt	Voltage	Alessandro Volta 1745-1827

$$E = I \times R \quad (2)$$

and

$$I = \frac{E}{R} \quad (3)$$

All three forms of the equation are used often in electronics and radio work. You must remember that the quantities are in volts, ohms and amperes; other units cannot be used in the equations without first being converted. For example, if the current is in milliamperes you must first change it to the equivalent fraction of an ampere before substituting the value into the equations.

The following examples illustrate the use of Ohm's Law in the simple circuit of **Fig 2.2**. If 150 V is applied to a circuit and the current is measured as 2.5 A, what is the resistance of the circuit? In this case R is the unknown, so we will use equation 1:

$$R = \frac{E}{I} = \frac{150\ V}{2.5\ A} = 60\ \Omega$$

No conversion of units was necessary because the voltage and current were given in volts and amperes.

If the current through a 20,000-Ω resistance is 150 mA, what is the voltage? To find voltage, use equation 2. Convert the current from milliamperes to amperes by dividing by 1000 mA / A (or multiplying by 10^{-3} A / mA) so that 150 mA becomes 0.150 A. (Notice the conversion factor of 1000 does not limit the number of significant figures in the calculated answer.)

$$I = \frac{150\ mA}{1000\ \frac{mA}{A}} = 0.150\ A$$

Then:

$$E = 0.150\ A \times 20000\ \Omega = 3000\ V$$

In a final example, how much current will flow if 250 V is applied to a 5000-Ω resistor? Since I is unknown,

$$I = \frac{E}{R} = \frac{250\ V}{5000\ \Omega} = 0.05\ A$$

It is more convenient to express the current in mA, and 0.05 A × 1000 mA / A = 50 mA.

It is important to note that Ohm's Law applies in any portion of a circuit as well as to the circuit as a whole. No matter how many resistors are connected together or how they are connected together, the relationship between the resistor's value, the voltage across the resistor, and the current through the resistor still follows Ohm's Law.

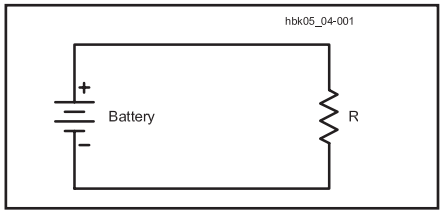


Fig 2.2 — A simple circuit consisting of a battery and a resistor.

Ohm's Law Timesaver

This simple diagram presents the mathematical equations relating voltage, current, and resistance. Cover the unkown quantity (E,I, or R) and the remaining symbols are shown as in the equation. For example, covering I shows E over R, as they would be written in the equation I=E/R.

When the current is small enough to be expressed in milliamperes, calculations are simplified if the resistance is expressed in kilohms rather than in ohms. With voltage in volts, if resistance in kilohms is substituted directly in Ohm's Law, the current will be milliamperes. Expressed as an equation: V = mA × kΩ.

2.2.4 Resistance of Wires

The problem of determining the resistance of a round wire of given diameter and length — or the converse, finding a suitable size and length of wire to provide a desired amount of resistance — can easily be solved with the help of the copper wire table given in the chapter on **Component Data and References**. This table gives the resistance, in ohms per 1000 ft, of each standard wire size. For example, suppose you need a resistance of 3.5 Ω, and some #28 AWG wire is on hand. The wire table shows that #28 AWG wire has a resistance of 65.31 Ω / 1000 ft. Since the desired resistance is 3.5 Ω, the required length of wire is:

$$\text{Length} = \frac{R_{DESIRED}}{\frac{R_{WIRE}}{1000\ ft}} = \frac{3.5\ \Omega}{\frac{66.17\ \Omega}{1000\ ft}} \quad (4)$$

$$= \frac{3.5\ \Omega \times 1000\ ft}{65.31\ \Omega} = 55\ ft$$

As another example, suppose that the resistance of wire in a radio's power cable must not exceed 0.05 Ω and that the length of wire required for making the connections totals 14 ft. Then:

$$\frac{R_{WIRE}}{1000\ ft} < \frac{R_{MAXIMUM}}{\text{Length}} = \frac{0.05\ \Omega}{14.0\ ft} \quad (5)$$

$$= 3.57 \times 10^{-3}\ \frac{\Omega}{ft} \times \frac{1000\ ft}{1000\ ft}$$

$$\frac{R_{WIRE}}{1000\ ft} < \frac{3.57\ \Omega}{1000\ ft}$$

Find the value of R_{WIRE} / 1000 ft that is less than the calculated value. The wire table shows that #15 AWG is the smallest size having a resistance less than this value. (The resistance of #15 AWG wire is given as 3.1810 Ω / 1000 ft.) Select any wire size larger than this for the connections in your circuit, to ensure that the total wire resistance will be less than 0.05 Ω.

When the wire in question is not made of copper, the resistance values in the wire table should be multiplied by the ratios shown in Table 2.1 to obtain the resulting resistance. If the wire in the first example were made from nickel instead of copper, the length required for 3.5 Ω would be:

$$\text{Length} = \frac{R_{DESIRED}}{\frac{R_{WIRE}}{1000\ ft}} \quad (6)$$

$$= \frac{3.5\ \Omega}{\frac{65.31\ \Omega}{1000\ ft} \times 5.1} = \frac{3.5\ \Omega \times 1000\ ft}{66.17\ \Omega \times 5.1}$$

$$\text{Length} = \frac{3500\ ft}{337.5} = 10.5\ ft$$

Component Tolerance

Resistors are manufactured with a specific *nominal* value of resistance. This is the value printed on the body of the resistor or marked with stripes of colored paint. The *actual* value of resistance varies from the nominal value because of random variations in the manufacturing process. The maximum allowable amount of variation is called the *tolerance* and it is expressed in percent. For example, a 1000-Ω resistor with a tolerance of 5% could have any value of resistance between 95% and 105% of 1000 Ω; 950 to 1050 Ω. In most circuits, this small variation doesn't have much effect, but it is important to be aware of tolerance and choose the correct value (10%, 5%, 1%, or even tighter tolerance values are available for *precision components*) of tolerance for the circuit to operate properly, no matter what the actual value of resistance. All components have this same nominal-to-actual value relationship.

2.2.5 Temperature Effects

The resistance of a conductor changes with its temperature. The resistance of practically

every metallic conductor increases with increasing temperature. Carbon, however, acts in the opposite way; its resistance decreases when its temperature rises. It is seldom necessary to consider temperature in making resistance calculations for amateur work. The temperature effect is important when it is necessary to maintain a constant resistance under all conditions, however. Special materials that have little or no change in resistance over a wide temperature range are used in that case.

2.2.6 Resistors

A package of material exhibiting a certain amount of resistance and made into a single unit is called a *resistor*. (See the **Component Data and References** chapter for information on resistor value marking conventions.) The size and construction of resistors having the same value of resistance in ohms may vary considerably based on how much power they are intended to dissipate, how much voltage is expected to be applied to them, and so forth (see **Fig 2.3**).

TYPES OF RESISTORS

Resistors are made in several different ways: carbon composition, metal oxide, carbon film, metal film and wirewound. In some circuits, the resistor value may be critical. In this case, precision resistors are used. These are typically wirewound or carbon-film devices whose values are carefully controlled during manufacture. In addition, special material or construction techniques may be used to provide temperature compensation, so the value does not change (or changes in a precise manner) as the resistor temperature changes.

Carbon composition resistors are simply small cylinders of carbon mixed with various binding agents to produce any desired resistance. The most common sizes of "carbon comp" resistors are ½- and ¼-W resistors. They are moderately stable from 0 to 60 °C (their resistance increases above and below this temperature range). They can absorb short overloads better than film-type resistors, but they are relatively noisy, and have relatively wide tolerances. Because carbon composition resistors tend to be affected by humidity and other environmental factors and because they are difficult to manufacture in surface-mount packages, they have largely been replaced by film-type resistors.

Metal-oxide resistors are similar to carbon composition resistors in that the resistance is supplied by a cylinder of metal oxide. Metal-oxide resistors have replaced carbon composition resistors in higher power applications because they are more stable and can operate at higher temperatures.

Wirewound resistors are made from wire, which is cut to the proper length and wound on a coil form (usually ceramic). They are capable of handling high power; their values are very stable, and they are manufactured to close tolerances.

Metal-film resistors are made by depositing a thin film of aluminum, tungsten or other metal on an insulating substrate. Their resistances are controlled by careful adjustments of the width, length and depth of the film. As a result, they have very tight tolerances. They are used extensively in surface-mount technology. As might be expected, their power handling capability is somewhat limited. They also produce very little electrical noise.

Carbon-film resistors use a film of carbon mixed with other materials instead of metal. They are not quite as stable as other film resistors and have wider tolerances than metal-film resistors, but they are still as good as (or better than) carbon composition resistors.

THERMAL CONSIDERATIONS FOR RESISTORS

Current through a resistance causes the conductor to become heated; the higher the resistance and the larger the current, the greater the amount of heat developed. Resistors intended for carrying large currents must be physically large so the heat can be radiated quickly to the surrounding air or some type of heat sinking material. If the resistor does not dissipate the heat quickly, it may get hot enough to melt or burn.

The amount of heat a resistor can safely dissipate depends on the material, surface area and design. Typical resistors used in amateur electronics (⅛ to 2-W resistors) dissipate heat primarily through the surface area of the case, with some heat also being carried away through the connecting leads. Wirewound resistors are usually used for higher power levels. Some have finned cases for better convection cooling and/or metal cases for better conductive cooling.

The major departure of resistors from ideal behavior at low-frequencies is their *temperature coefficient* (TC). (See the chapter on **RF Techniques** for a discussion of the behavior of resistors at high frequencies.) The resistivity of most materials changes with temperature, and typical TC values for resistor materials are given in **Table 2.2**. TC values are usually expressed in parts-per-million (PPM) for each degree (centigrade) change from some nominal temperature, usually room temperature (77 °F or 27 °C). A positive TC indicates an increase in resistance with increasing temperature while a negative TC indicates a decreasing resistance. For example, if a 1000-Ω resistor with a TC of +300 PPM/°C is heated to 50 °C, the *change* in resistance

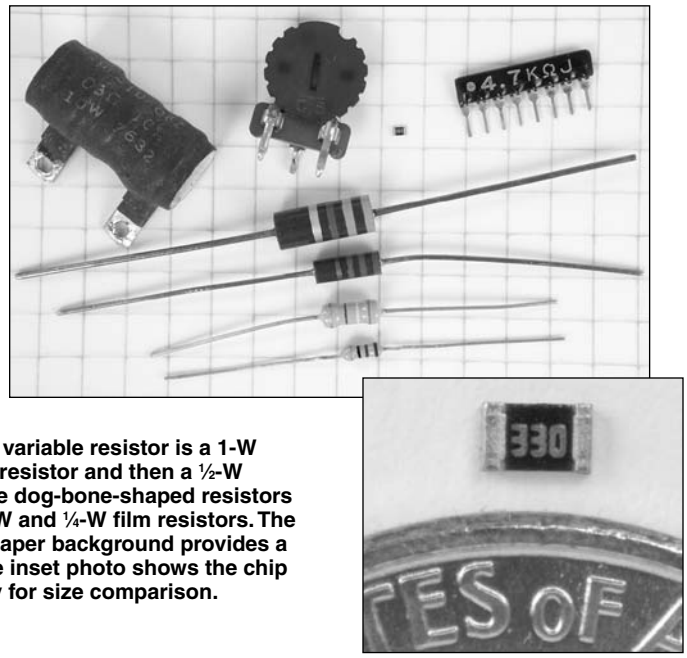


Fig 2.3 — Examples of various resistors. At the top left is a small 10-W wirewound resistor. A single in-line package (SIP) of resistors is at the top right. At the top center is a small PC-board-mount variable resistor. A tiny surface-mount (chip) resistor is also shown at the top. Below the variable resistor is a 1-W carbon composition resistor and then a ½-W composition unit. The dog-bone-shaped resistors at the bottom are ½-W and ¼-W film resistors. The ¼-inch-ruled graph paper background provides a size comparison. The inset photo shows the chip resistor with a penny for size comparison.

Table 2.2

Temperature Coefficients for Various Resistor Compositions

1 PPM = 1 part per million = 0.0001%

Type	*TC (PPM/°C)*
Wire wound	±(30 - 50)
Metal Film	±(100 - 200)
Carbon Film	+350 to –800
Carbon composition	±800

is 300 × (50 – 27) = 6900 PPM, yielding a new resistance of

$$1000\left(1+\frac{6900}{1000000}\right)=1006.9\,\Omega$$

Carbon-film resistors are unique among the major resistor families because they alone have a negative temperature coefficient. They are often used to "offset" the thermal effects of the other components.

If the temperature increase is small (less than 30-40 °C), the resistance change with temperature is nondestructive — the resistor will return to normal when the temperature returns to its nominal value. Resistors that get too hot to touch, however, may be permanently damaged even if they appear normal. For this reason, be conservative when specifying power ratings for resistors. It's common to specify a resistor rated at 200% to 400% of the expected dissipation.

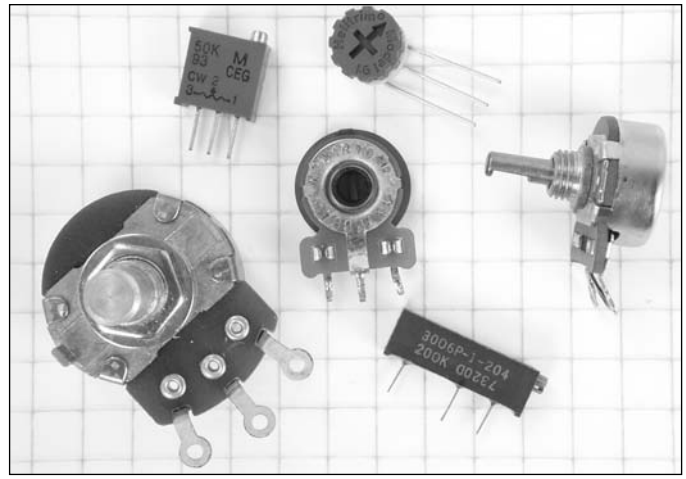


Fig 2.4 — This photo shows examples of different styles of potentiometers. The ¼-inch-ruled graph paper background provides a size comparison.

POTENTIOMETERS

Potentiometer (pronounced po-ten-tchee-AH-meh-tur) is a formal name for a variable resistor and the common name for these components is "pots." A typical potentiometer consists of a circular pattern of resistive material, usually a carbon compound similar to that used in carbon composition resistors, with a wiper contact on a shaft moving across the material. For higher power applications, the resistive material may be wire, wound around a core, like a wirewound resistor.

As the wiper moves along the material, more resistance is introduced between the wiper and one of the fixed contacts on the material. A potentiometer may be used to control current, voltage or resistance in a circuit. **Fig 2.4** shows several different types of potentiometers. **Fig 2.5** shows the schematic symbol for a potentiometer and how changing the position of the shaft changes the resistance between its three terminals. The figure shows a *panel pot*, designed to be mounted on an equipment panel and adjusted by an operator. The small rectangular *trimmer* potentiometers in Fig 2.5 are adjusted with a screwdriver and have wire terminals.

Typical specifications for a potentiometer include element resistance, power dissipation, voltage and current ratings, number of turns (or degrees) the shaft can rotate, type and size of shaft, mounting arrangements and resistance "taper."

Not all potentiometers have a *linear* taper. That is, the change in resistance may not be the same for a given number of degrees of shaft rotation along different portions of the resistive material. These are called *nonlinear tapers*

A typical use of a potentiometer with a nonlinear taper is as a volume control in an audio amplifier. Since the human ear has a logarithmic response to sound, a volume control must change the amplifier output much more near one end of the potentiometer than the other (for a given amount of rotation) so that the "perceived" change in volume is about the same for a similar change in the control's position. This is commonly called an *audio taper* or *log taper* as the change in resistance per degree of rotation attempts to match the response of the human ear. Tapers can be designed to match almost any desired control function for a given application. Linear and audio tapers are the most common tapers.

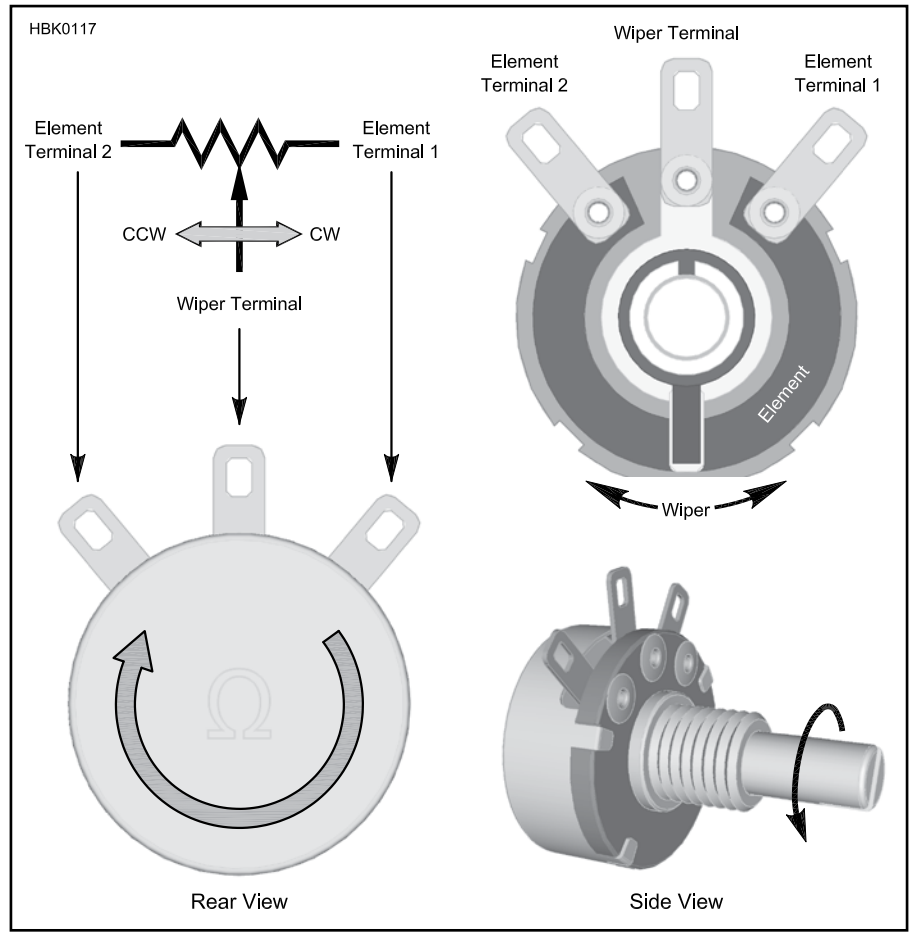


Fig 2.5 — Typical potentiometer construction and schematic symbol. Rotation on the shaft moves the wiper along the element, changing the resistance between the wiper terminal and the element terminals. Moving the wiper closer to an element terminal reduces the resistance between them.

2.3 Basic Circuit Principles

Circuits are composed of *nodes* and *branches*. A node is any point in the circuit at which current can divide between conducting paths. For example, in the parallel circuit of **Fig 2.6**, the node is represented by the schematic dot. A branch is any unique conducting path between nodes. A series of branches that make a complete current path, such as the series circuit of Fig 2.1A, is called a *loop*.

Very few actual circuits are as simple as those shown in Fig 2.1. However, all circuits, no matter how complex, are constructed of combinations of these series and parallel circuits. We will now use these simple circuits of resistors and batteries to illustrate two fundamental rules for voltage and current, known as *Kirchoff's Laws.*

2.3.1 Kirchhoff's Current Law

Kirchhoff's Current Law (KCL) states, *"The sum of all currents flowing into a node and all currents flowing out of a node is equal to zero."* KCL is stated mathematically as:

$$(I_{in1} + I_{in2} + \cdots) - (I_{out1} + I_{out2} + \cdots) = 0 \qquad (7A)$$

The dots indicate that as many currents as necessary may be added.

Another way of stating KCL is that the sum of all currents flowing into a node must balance the sum of all currents flowing out of a node as shown in equation 7B:

$$(I_{in1} + I_{in2} + \cdots) = (I_{out1} + I_{out2} + \cdots) \qquad (7B)$$

Equations 7A and 7B are mathematically equivalent.

KCL is illustrated by the following example. Suppose three resistors (5.0 kΩ, 20.0 kΩ and 8.0 kΩ) are connected in parallel as shown in Fig 2.6. The same voltage, 250 V, is applied to all three resistors. The current through R1 is I1, I2 is the current through R2 and I3 is the current through R3.

The current in each can be found from Ohm's Law, as shown below. For convenience, we can use resistance in kΩ, which gives current in milliamperes.

$$I1 = \frac{E}{R1} = \frac{250\ V}{5.0\ k\Omega} = 50.0\ mA$$

$$I2 = \frac{E}{R2} = \frac{250\ V}{20.0\ k\Omega} = 12.5\ mA$$

$$I3 = \frac{E}{R3} = \frac{250\ V}{8.0\ k\Omega} = 31.2\ mA$$

Notice that the branch currents are inversely proportional to the resistances. The 20,000-Ω resistor has a value four times larger than the 5000-Ω resistor, and has a current one-quarter as large. If a resistor has a value twice as large as another, it will have half as much current through it when they are connected in parallel.

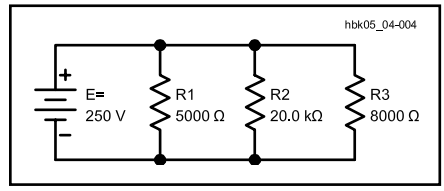


Fig 2.6 — An example of resistors in parallel.

Using the balancing form of KCL (equation 7B) the current that must be supplied by the battery is therefore:

$$I_{BATT} = I1 + I2 + I3$$

$$I_{BATT} = 50.0\ mA + 12.5\ mA + 31.2\ mA$$

$$I_{BATT} = 93.7\ mA$$

2.3.2 Resistors in Parallel

In a circuit made up of resistances in parallel, the resistors can be represented as a single *equivalent* resistance that has the same value as the parallel combination of resistors. In a parallel circuit, the equivalent resistance is less than that of the lowest resistance value present. This is because the total current is always greater than the current in any individual resistor. The formula for finding the equivalent resistance of resistances in parallel is:

$$R_{EQUIV} = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3} + \frac{1}{R4} \ldots} \qquad (8)$$

where the dots indicate that any number of parallel resistors can be combined by the same method. In the example of the previous section, the equivalent resistance is:

$$R = \frac{1}{\frac{1}{5000\ \Omega} + \frac{1}{20\ k\Omega} + \frac{1}{8000\ \Omega}} = 2.67\ k\Omega$$

The notation "//" (two slashes) is frequently used to indicate "in parallel with." Using that notation, the preceding example would be given as "5000 Ω // 20 kΩ // 8000 Ω."

For only two resistances in parallel (a very common case) the formula can be reduced to the much simpler (and easier to remember):

$$R_{EQUIV} = \frac{R1 \times R2}{R1 + R2} \qquad (9)$$

Example: If a 500-Ω resistor is connected in parallel with one of 1200 Ω, what is the total resistance?

$$R = \frac{R1 \times R2}{R1 + R2} = \frac{500\ \Omega \times 1200\ \Omega}{500\ \Omega + 1200\ \Omega}$$

$$R = \frac{600000\ \Omega^2}{1700\ \Omega} = 353\ \Omega$$

Any number of parallel resistors can be combined two at a time by using equation 9 until all have been combined into a single equivalent. This is a bit easier than using equation 8 to do the conversion in a single step.

CURRENT DIVIDERS

Resistors connected in parallel form a circuit called a *resistive current divider*. For any number of resistors connected in parallel (R1, R2, R3, ... R4), the current through one of the resistors, R_n, is equal to the sum of all resistor currents multiplied by the ratio of the sum of all resistors *except* R_n to the sum of all resistors.

For example, in a circuit with three parallel resistors; R1, R2, and R3, the current through R2 is equal to:

$$I_2 = I \frac{R1 + R3}{R1 + R2 + R3}$$

where I is the total of the individual currents flowing through all resistors. If I = 100 mA, R1 = 100 Ω, R2 = 50 Ω, and R3 = 200 Ω:

$$100\ mA \frac{100 + 200}{100 + 50 + 200} = 85.7\ mA$$

2.3.3 Kirchhoff's Voltage Law

Kirchhoff's Voltage Law (KVL) states, *"The sum of the voltages around a closed current loop is zero."* Where KCL is somewhat intuitive, KVL is not as easy to visualize. In the circuit of **Fig 2.7**, KVL requires that the battery's voltage must be balanced exactly by the voltages that appear across the three resistors in the circuit. If it were not, the "extra" voltage would create an infinite current with no limiting resistance, just as KCL prevents charge from "building up" at a circuit node.

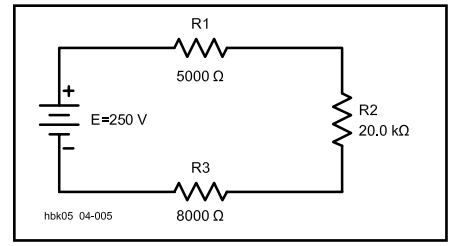


Fig 2.7 — An example of resistors in series.

KVL is stated mathematically as:

$$E_1 + E_2 + E_3 + \cdots = 0 \qquad (10)$$

where each E represents a voltage encountered by current as it flows around the circuit loop.

This is best illustrated with an example. Although the current is the same in all three of the resistances in the example of Fig 2.7, the total voltage divides between them, just as current divides between resistors connected in parallel. The voltage appearing across each resistor (the *voltage drop*) can be found from Ohm's Law. (Voltage across a resistance is often referred to as a "drop" or "I-R drop" because the value of the voltage "drops" by the amount $E = I \times R$.)

For the purpose of KVL, it is common to assume that if current flows *into* the more positive terminal of a component the voltage is treated as positive in the KVL equation. If the current flows *out* of a positive terminal, the voltage is treated as negative in the KVL. Positive voltages represent components that consume or "sink" power, such as resistors. Negative voltages represent components that produce or "source" power, such as batteries. This allows the KVL equation to be written in a balancing form, as well:

$$(E_{source1} + E_{source2} + \cdots) =$$
$$(E_{sink1} + E_{sink2} + \cdots)$$

All of the voltages are treated as positive in this form, with the power sources (current flowing *out* of the more positive terminal) on one side and the power sinks (current flowing *into* the more positive terminal) on the other side.

Note that it doesn't matter what a component terminal's *absolute* voltage is with respect to ground, only which terminal of the component is more positive than the other. If one side of a resistor is at +1000 V and the other at +998 V, current flowing into the first terminal and out of the second experiences a +2 V voltage drop. Similarly, current supplied by a 9 V battery with its positive terminal at –100 V and its negative terminal at –108.5 V still counts for KVL as an 8.5 V power source. Also note that current can flow *into* a battery's positive terminal, such as during recharging, making the battery a power sink, just like a resistor.

Here's an example showing how KVL works: In Fig 2.7, if the voltage across R1 is E1, that across R2 is E2 and that across R3 is E3, then:

$$-250 + I \times R1 + I \times R2 + I \times R3 = 0$$

This equation can be simplified to:

$$-250 + I\,(R1 + R2 + R3) =$$
$$-250 + I\,(33000\ \Omega) = 0$$

Solving for I gives I = 250 / 33000 = 0.00758 A = 7.58 mA. This allows us to calculate the value of the voltage across each resistor:

$$E1 = I \times R1 = 0.00758\ A \times 5000\ \Omega = 37.9\ V$$

$$E2 = I \times R2 = 0.00758\ A \times 20000\ \Omega = 152\ V$$

$$E3 = I \times R3 = 0.00758\ A \times 8000\ \Omega = 60.6\ V$$

Verifying that the sum of E1, E2, and E3 does indeed equal the battery voltage of 250 V ignoring rounding errors.

$$E_{TOTAL} = E1 + E2 + E3$$

$$E_{TOTAL} = 37.9\ V + 152\ V + 60.6\ V$$

$$E_{TOTAL} = 250\ V$$

2.3.4 Resistors in Series

The previous example illustrated that in a circuit with a number of resistances connected in series, the equivalent resistance of the circuit is the sum of the individual resistances. If these are numbered R1, R2, R3 and so on, then:

$$R_{EQUIV} = R1 + R2 + R3 + R4... \qquad (11)$$

Example: Suppose that three resistors are connected to a source of voltage as shown in Fig 2.7. The voltage is 250 V, R1 is 5.0 kΩ, R2 is 20.0 kΩ and R3 is 8.0 kΩ. The total resistance is then

$$R_{EQUIV} = R1 + R2 + R3$$

$$R_{EQUIV} = 5.0\ k\Omega + 20.0\ k\Omega + 8.0\ k\Omega$$

$$R_{EQUIV} = 33.0\ k\Omega$$

The current in the circuit is then

$$I = \frac{E}{R} = \frac{250\ V}{33.0\ k\Omega} = 7.58\ mA$$

VOLTAGE DIVIDERS

Notice that the voltage drop across each resistor in the KVL example is directly proportional to the resistance. The value of the 20,000 Ω resistor is four times larger than the 5000 Ω resistor, and the voltage drop across the 20,000 Ω resistor is four times larger. A resistor that has a value twice as large as another will have twice the voltage drop across it when they are connected in series.

Resistors in series without any other connections form a *resistive voltage divider*. (Other types of components can form voltage dividers, too.) The voltage across any specific resistor in the divider, R_n, is equal to the voltage across the entire string of resistors multiplied by the ratio of R_n to the sum of all resistors in the string.

For example, in the circuit of Fig 2.7, the voltage across the 5000 Ω resistor is:

$$E1 = 250\,\frac{5000}{5000 + 20000 + 8000} = 37.9\ V$$

This is a more convenient method than calculating the current through the resistor and using Ohm's Law.

Voltage dividers can be used as a source of voltage. As long as the device connected to the output of the divider has a much higher resistance than the resistors in the divider, there will be little effect on the divider output voltage. For example, for a voltage divider with a voltage of E = 15 V and two resistors of R1 = 5 kΩ and R2 = 10 kΩ, the voltage across R2 will be 10 V measured on a high-impedance voltmeter because the measurement draws very little current from the divider. However, if the measuring device or load across R2 draws significant current, it will increase the amount of current drawn through the divider and change the output voltage.

A good rule of thumb is that the load across a voltage divider's output should be at least ten times the value of the highest resistor in the divider to stay reasonably close to the required voltage. As the load resistance gets closer to the value of the divider resistors, the current drawn by the load affects the voltage division and causes changes from the desired value.

Potentiometers (variable resistors described previously) are often used as adjustable voltage dividers and this is how they got their name. Potential is an older name for voltage and a "potential-meter" is a device that can "meter" or adjust potential, thus potentiometer.

2.3.5 Conductances in Series and Parallel

Since conductance is the reciprocal of resistance, G = 1/R, the formulas for combining resistors in series and in parallel can be converted to use conductance by substituting 1/G for R. Substituting 1/G into equation 11 shows that conductances in series are combined this way:

$$G = \frac{1}{\frac{1}{G1} + \frac{1}{G2} + \frac{1}{G3} + \frac{1}{G4}...}$$

and two conductances in series may be combined in a manner similar to equation 9:

$$G_{EQUIV} = \frac{G1 \times G2}{G1 + G2}$$

Substituting 1/G into equation 8 shows that conductances in parallel are combined this way:

$$G_{TOTAL} = G1 + G2 + G3 + G4...$$

This also shows that when faced with a large number of parallel resistances, converting

them to conductances may make the math a little easier to deal with.

2.3.6 Equivalent Circuits

A circuit may have resistances both in parallel and in series, as shown in **Fig 2.8A**. In order to analyze the behavior of such a circuit, *equivalent circuits* are created and combined by using the equations for combining resistors in series and resistors in parallel. Each separate combination of resistors, series or parallel, can be reduced to a single equivalent resistor. The resulting combinations can be reduced still further until only a single resistor remains.

The method for analyzing the circuit of Fig 2.8A is as follows: Combine R2 and R3 to create the equivalent single resistor, R_{EQ} whose value is equal to R2 and R3 in parallel.

$$R_{EQ} = \frac{R2 \times R3}{R2 + R3} = \frac{20000\ \Omega \times 8000\ \Omega}{20000\ \Omega + 8000\ \Omega}$$

$$= \frac{1.60 \times 10^8\ \Omega^2}{28000\ \Omega} = 5710\ \Omega = 5.71\ k\Omega$$

This resistance in series with R1 then forms a simple series circuit, as shown in Fig 2.8B. These two resistances can then be combined into a single equivalent resistance, R_{TOTAL}, for the entire circuit:

$$R_{TOTAL} = R1 + R_{EQ} = 5.0\ k\Omega + 5.71\ k\Omega$$

$$R_{TOTAL} = 10.71\ k\Omega$$

The battery current is then:

$$I = \frac{E}{R} = \frac{250\ V}{10.71\ k\Omega} = 23.3\ mA$$

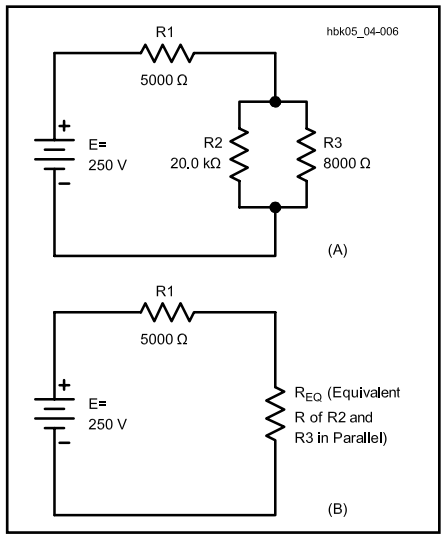


Fig 2.8 — At A, an example of resistors in series-parallel. The equivalent circuit is shown at B.

The voltage drops across R1 and R_{EQ} are:

$$E1 = I \times R1 = 23.3\ mA \times 5.0\ k\Omega = 117\ V$$

$$E2 = I \times R_{EQ} = 23.3\ mA \times 5.71\ k\Omega = 133\ V$$

These two voltage drops total 250 V, as described by Kirchhoff's Voltage Law. E2 appears across both R2 and R3 so,

$$I2 = \frac{E2}{R2} = \frac{133\ V}{20.0\ k\Omega} = 6.65\ mA$$

$$I3 = \frac{E3}{R3} = \frac{133\ V}{8.0\ k\Omega} = 16.6\ mA$$

where

I2 = current through R2 and
I3 = current through R3.

The sum of I2 and I3 is equal to 23.3 mA, conforming to Kirchhoff's Current Law.

2.3.7 Voltage and Current Sources

In designing circuits and describing the behavior of electronic components, it is often useful to use *ideal sources*. The two most common types of ideal sources are the *voltage source* and the *current source*, symbols for which are shown in **Fig 2.9**. These sources are considered ideal because no matter what circuit is connected to their terminals, they continue to supply the specified amount of voltage or current. Practical voltage and current sources can approximate the behavior of an ideal source over certain ranges, but are limited in the amount of power they can supply and so under excessive load, their output will drop.

Voltage sources are defined as having zero *internal impedance*, where impedance is a more general form of resistance as described in the sections of this chapter dealing with alternating current. A short circuit across an ideal voltage source would result in the source providing an infinite amount of current. Practical voltage sources have non-zero internal impedance and this also limits the amount of power they can supply. For example, placing a short circuit across the terminals of a practical voltage source such as 1.5 V dry-cell battery may produce a current of several amperes, but the battery's internal impedance acts to limit the amount of current produced in accordance with Ohm's Law — as if the resistor in Fig 2.2 was inside of or internal to the battery.

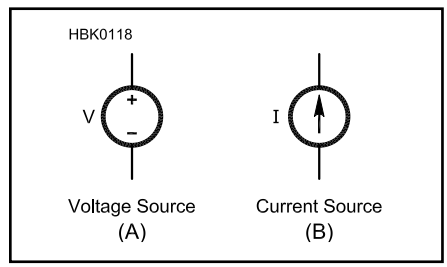


Fig 2.9 — Voltage sources (A) and current sources (B) are examples of ideal energy sources.

Current sources are defined to have infinite internal impedance. This means that no matter what is connected to the terminals of an ideal current source, it will supply the same amount of current. An open circuit across the terminal of an ideal current source will result in the source generating an infinite voltage at its terminals. Practical current sources will raise their voltage until the internal power supply limits are reached and then reduce output current.

2.3.8 Thevenin's Theorem and Thevenin Equivalents

Thevenin's Voltage Theorem (usually just referred to as "Thevenin's Theorem") is a useful tool for simplifying electrical circuits or *networks* (the formal name for circuits) by allowing circuit designers to replace a circuit with a simpler equivalent circuit. Thevenin's Theorem states, "*Any two-terminal network made up of resistors and voltage or current sources can be replaced by an equivalent network made up of a single voltage source and a series resistor.*"

Thevenin's Theorem can be readily applied to the circuit of Fig 2.8A, to find the current through R3. In this example, illustrated in **Fig 2.10**, the circuit is redrawn to show R1 and R2 forming a voltage divider, with R3 as the load (Fig 2.10A). The current drawn by the load (R3) is simply the voltage across R3, divided by its resistance. Unfortunately, the value of R2 affects the voltage across R3, just as the presence of R3 affects the voltage appearing across R2. Some means of separating the two is needed; hence the *Thevenin-equivalent circuit* is constructed, replacing everything connected to terminals A and B with a single voltage source (the *Thevenin-equivalent voltage,* E_{THEV}) and series resistor (the *Thevenin-equivalent resistance,* R_{TH}).

The first step of creating the Thevenin-equivalent of the circuit is to determine its *open-circuit voltage,* measured when there is no load current drawn from either terminal A or B. Without a load connected between A and B, the total current through the circuit is (from Ohm's Law):

$$I = \frac{E}{R1 + R2} \qquad (12)$$

and the voltage between terminals A and B (E_{AB}) is:

$$E_{AB} = I \times R2 \qquad (13)$$

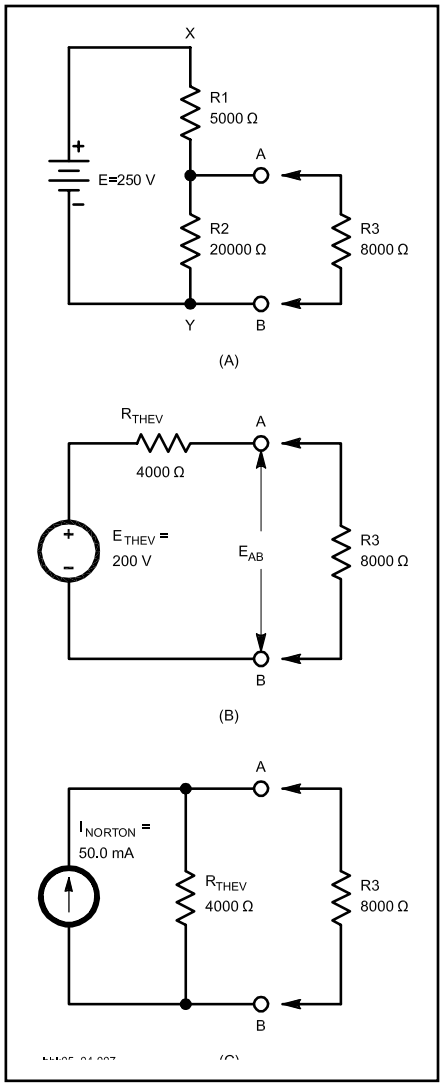


Fig 2.10 — Equivalent circuits for the circuit shown in Fig 2.8. A shows the circuit to be replaced by an equivalent circuit from the perspective of the resistor (R3 load). B shows the Thevenin-equivalent circuit, with a resistor and a voltage source in series. C shows the Norton-equivalent circuit, with a resistor and current source in parallel.

By substituting equation 12 into equation 13, we have an expression for E_{AB} in which all values are known:

$$E_{AB} = \frac{R2}{R1 + R2} \times E \qquad (14)$$

Using the values in our example, this becomes:

$$E_{AB} = \frac{20.0\ k\Omega}{25.0\ k\Omega} \times 250\ V = 200\ V$$

when nothing is connected to terminals A or B. E_{THEV} is equal to E_{AB} with no current drawn.

The equivalent resistance between terminals A and B is R_{THEV}. R_{THEV} is calculated as the equivalent circuit at terminals A and B with all sources, voltage or current, replaced by their internal impedances. The ideal voltage source, by definition, has zero internal resistance and is replaced by a short circuit. The ideal current source has infinite internal impedance and is replaced by an open circuit.

Assuming the battery to be a close approximation of an ideal source, replace it with a short circuit between points X and Y in the circuit of Fig 2.10A. R1 and R2 are then effectively placed in parallel, as viewed from terminals A and B. R_{THEV} is then:

$$R_{THEV} = \frac{R1 \times R2}{R1 + R2} \qquad (15)$$

$$R_{THEV} = \frac{5000\ \Omega \times 20000\ \Omega}{5000\ \Omega + 20000\ \Omega}$$

$$R_{THEV} = \frac{1.0 \times 10^8\ \Omega^2}{25000\ \Omega} = 4000\ \Omega$$

This gives the Thevenin-equivalent circuit as shown in Fig 2.10B. The circuits of Figs 2.10A and 2.10B are completely equivalent from the perspective of R3, so the circuit becomes a simple series circuit.

Once R3 is connected to terminals A and B, there will be current through R_{THEV}, causing a voltage drop across R_{THEV} and reducing E_{AB}. The current through R3 is equal to

$$I3 = \frac{E_{THEV}}{R_{TOTAL}} = \frac{E_{THEV}}{R_{THEV} + R3} \qquad (16)$$

Substituting the values from our example:

$$I3 = \frac{200\ V}{4000\ \Omega + 8000\ \Omega} = 16.7\ mA$$

This agrees with the value calculated earlier.

The Thevenin-equivalent circuit of an ideal voltage source in series with a resistance is a good model for a real voltage source with non-zero internal resistance. Using this more realistic model, the maximum current that a real voltage source can deliver is seen to be

$$I_{sc} = \frac{E_{THEV}}{R_{THEV}}$$

and the maximum output voltage is $V_{oc} = E_{THEV}$.

Sinusoidal voltage or current sources can be modeled in much the same way, keeping in mind that the internal impedance, Z_{THEV}, for such a source may not be purely resistive, but may have a reactive component that varies with frequency.

2.3.9 Norton's Theorem and Norton Equivalents

Norton's Theorem is another method of creating an equivalent circuit. Norton's Theorem states, "*Any two-terminal network made up of resistors and current or voltage sources can be replaced by an equivalent network made up of a single current source and a parallel resistor.*" Norton's Theorem is to current sources what Thevenin's Theorem is to voltage sources. In fact, the Thevenin-resistance calculated previously is also the *Norton-equivalent resistance*.

The circuit just analyzed by means of Thevenin's Theorem can be analyzed just as easily by Norton's Theorem. The equivalent Norton circuit is shown in Fig 2.10C. The short circuit current of the equivalent circuit's current source, I_{NORTON}, is the current through terminals A and B with the load (R3) replaced by a short circuit. In the case of the voltage divider shown in Fig 2.10A, the short circuit completely bypasses R2 and the current is:

$$I_{AB} = \frac{E}{R1} \qquad (17)$$

Substituting the values from our example, we have:

$$I_{AB} = \frac{E}{R1} = \frac{250\ V}{5000\ \Omega} = 50.0\ mA$$

The resulting Norton-equivalent circuit consists of a 50.0-mA current source placed in parallel with a 4000-Ω resistor. When R3 is connected to terminals A and B, one-third of the supply current flows through R3 and the remainder through R_{THEV}. This gives a current through R3 of 16.7 mA, again agreeing with previous conclusions.

A Norton-equivalent circuit can be transformed into a Thevenin-equivalent circuit and vice versa. The equivalent resistor, R_{THEV}, is the same in both cases; it is placed in series with the voltage source in the case of a Thevenin-equivalent circuit and in parallel with the current source in the case of a Norton-equivalent circuit. The voltage for the Thevenin-equivalent source is equal to the open-circuit voltage appearing across the resistor in the Norton-equivalent circuit. The current for a Norton-equivalent source is equal to the short circuit current provided by the Thevenin source. A Norton-equivalent circuit is a good model for a real current source that has a less-than infinite internal impedance.

2.4 Power and Energy

Regardless of how voltage is generated, energy must be supplied if current is drawn from the voltage source. The energy supplied may be in the form of chemical energy or mechanical energy. This energy is measured in joules (J). One joule is defined from classical physics as the amount of energy or *work* done when a force of one newton (a measure of force) is applied to an object that is moved one meter in the direction of the force.

Power is another important concept and measures the rate at which energy is generated or used. One *watt* (W) of power is defined as the generation (or use) of one joule of energy (or work) per second.

One watt is also defined as one volt of EMF causing one ampere of current to flow through a resistance. Thus,

$$P = I \times E \tag{18}$$

where
P = power in watts
I = current in amperes
E = EMF in volts.

(This discussion pertains only to direct current in resistive circuits. See the **AC Theory and Reactance** section of this chapter for a discussion about power in ac circuits, including reactive circuits.)

Common fractional and multiple units for power are the milliwatt (mW, one thousandth of a watt) and the kilowatt (kW, 1000 W).

Example: The plate voltage on a transmitting vacuum tube is 2000 V and the plate current is 350 mA. (The current must be changed to amperes before substitution in the formula, and so is 0.350 A.) Then:

$$P = I \times E = 2000\ V \times 0.350\ A = 700\ W$$

Power may be expressed in *horsepower* (hp) instead of watts, using the following conversion factor:

$$1\ \text{horsepower} = 746\ W$$

This conversion factor is especially useful if you are working with a system that converts electrical energy into mechanical energy, and vice versa, since mechanical power is often expressed in horsepower in the US. In metric countries, mechanical power is usually expressed in watts. All countries use the metric power unit of watts in electrical systems, however. The value 746 W/hp assumes lossless conversion between mechanical and electrical power; practical efficiency is taken up shortly.

2.4.1 Energy

When you buy electricity from a power company, you pay for electrical energy, not power. What you pay for is the work that the electrical energy does for you, not the rate at which that work is done. Like energy, work is equal to power multiplied by time. The common unit for measuring electrical energy is the *watt-hour* (Wh), which means that a power of one watt has been used for one hour. That is:

$$Wh = P\,t \tag{19}$$

where
Wh = energy in watt-hours
P = power in watts
t = time in hours.

Actually, the watt-hour is a fairly small energy unit, so the power company bills you for *kilowatt-hours* (kWh) of energy used. Another energy unit that is sometimes useful is the *watt-second* (Ws), which is equal to joules.

It is important to realize, both for calculation purposes and for efficient use of power resources, a small amount of power used for a long time can eventually result in a power bill that is just as large as if a large amount of power had been used for a very short time.

A common use of energy units in radio is in specifying the energy content of a battery. Battery energy is rated in *ampere-hours* (Ah) or *milliampere-hours* (mAh). While the multiplication of amperes and hours does not result in units of energy, the calculation assumes the result is multiplied by a specified (and constant) battery voltage. For example, a rechargeable NiMH battery rated to store 2000 mAh of energy is assumed to supply that energy at a terminal voltage of 1.5 V. Thus, after converting 2000 mA to 2 A, the actual energy stored is:

$$\text{Energy} = 1.5\ V \times 2\ A \times 1\ \text{hour} = 3\ Wh$$

Another common energy unit associated with batteries is *energy density*, with units of Ah per unit of volume or weight.

One practical application of energy units is to estimate how long a radio (such as a hand-held unit) will operate from a certain battery. For example, suppose a fully charged battery stores 900 mAh of energy and that the radio draws 30 mA on receive. A simple calculation indicates that the radio will be able receive 900 mAh / 30 mA = 30 hours with this battery, assuming 100% efficiency. You shouldn't expect to get the full 900 mAh out of the battery because the battery's voltage will drop as it is discharged, usually causing the equipment it powers to shut down before the last fraction of charge is used. Any time spent transmitting will also reduce the time the battery will last. The **Power Supplies** chapter includes additional information about batteries and their charge/discharge cycles.

2.4.2 Generalized Definition of Resistance

Electrical energy is not always turned into heat. The energy used in running a motor, for example, is converted to mechanical motion. The energy supplied to a radio transmitter is largely converted into radio waves. Energy applied to a loudspeaker is changed into sound waves. In each case, the energy is converted to other forms and can be completely accounted for. None of the energy just disappears! These are examples of the Law of Conservation of Energy. When a device converts energy from one form to another, we often say it *dissipates* the energy, or power. (Power is energy divided by time.) Of course the device doesn't really "use up" the energy, or make it disappear, it just converts it to another form. Proper operation of electrical devices often requires that the power be supplied at a specific ratio of voltage to current. These features are characteristics of resistance, so it can be said that any device that "dissipates power" has a definite value of resistance.

This concept of resistance as something that absorbs power at a definite voltage-to-current ratio is very useful; it permits substituting a simple resistance for the load or power-consuming part of the device receiving power, often with considerable simplification of calculations. Of course, every electrical device has some resistance of its own in the more narrow sense, so a part of the energy supplied to it is converted to heat in that resistance even though the major part of the energy may be converted to another form.

2.4.3 Efficiency

In devices such as motors and transmitters, the objective is to convert the supplied energy (or power) into some form other than heat. In such cases, power converted to heat is considered to be a loss because it is not useful power. The efficiency of a device is the useful power output (in its converted form) divided by the power input to the device. In a transmitter, for example, the objective is to convert power from a dc source into ac power at some radio frequency. The ratio of the RF power output to the dc input is the *efficiency* (*Eff* or η) of the transmitter. That is:

$$\text{Eff} = \frac{P_O}{P_I} \tag{20}$$

where
Eff = efficiency (as a decimal)
P_O = power output (W)

P_I = power input (W).

Example: If the dc input to the transmitter is 100 W, and the RF power output is 60 W, the efficiency is:

$$\text{Eff} = \frac{P_O}{P_I} = \frac{60\ \text{W}}{100\ \text{W}} = 0.6$$

Efficiency is usually expressed as a percentage — that is, it expresses what percent of the input power will be available as useful output. To calculate percent efficiency, multiply the value from equation 20 by 100%. The efficiency in the example above is 60%.

Suppose a mobile transmitter has an RF power output of 100 W with 52% efficiency at 13.8 V. The vehicle's alternator system charges the battery at a rate of 5.0 A at this voltage. Assuming an alternator efficiency of 68%, how much horsepower must the engine produce to operate the transmitter and charge the battery? Solution: To charge the battery, the alternator must produce 13.8 V × 5.0 A = 69 W. The transmitter dc input power is 100 W / 0.52 = 190 W. Therefore, the total electrical power required from the alternator is 190 + 69 = 259 W. The engine load then is:

$$P_I = \frac{P_O}{\text{Eff}} = \frac{259\ \text{W}}{0.68} = 381\ \text{W}$$

We can convert this to horsepower using the conversion factor given earlier to convert between horsepower and watts:

$$\frac{381\ \text{W}}{746\ \text{W/hp}} = 0.51\ \text{horsepower (hp)}$$

2.4.4 Ohm's Law and Power Formulas

Electrical power in a resistance is turned into heat. The greater the power, the more rapidly the heat is generated. By substituting the Ohm's Law equivalent for E and I, the following formulas are obtained for power:

$$P = \frac{E^2}{R} \qquad (21)$$

and

$$P = I^2 \times R \qquad (22)$$

These formulas are useful in power calculations when the resistance and either the current or voltage (but not both) are known.

Example: How much power will be converted to heat in a 4000-Ω resistor if the potential applied to it is 200 V? From equation 21,

$$P = \frac{E^2}{R} = \frac{40000\ \text{V}^2}{4000\ \Omega} = 10.0\ \text{W}$$

As another example, suppose a current of 20 mA flows through a 300-Ω resistor. Then:

$$P = I^2 \times R = 0.020^2\ \text{A}^2 \times 300\ \Omega$$

$$P = 0.00040\ \text{A}^2 \times 300\ \Omega$$

$$P = 0.12\ \text{W}$$

Note that the current was changed from milliamperes to amperes before substitution in the formula.

Resistors for radio work are made in many sizes, the smallest being rated to safely operate at power levels of about 1/16 W. The largest resistors commonly used in amateur equipment are rated at about 100 W. Large resistors, such as those used in dummy-load antennas, are often cooled with oil to increase their power-handling capability.

Ohm's Law and Power Circle

During the first semester of my *Electrical Power Technology* program, one of the first challenges issued by our dedicated instructor — Roger Crerie — to his new freshman students was to identify and develop 12 equations or formulas that could be used to determine voltage, current, resistance and power. Ohm's Law is expressed as R = E / I and it provided three of these equation forms while the basic equation relating power to current and voltage (P = I × E) accounted for another three. With six known equations, it was just a matter of applying mathematical substitution for his students to develop the remaining six. Together, these 12 equations compose the *circle* or *wheel* of voltage (E), current (I), resistance (R) and power (P) shown in **Fig 2.A1**. Just as Roger's previous students had learned at the Worcester Industrial Technical Institute (Worcester, Massachusetts), our Class of '82 now held the basic electrical formulas needed to proceed in our studies or professions. As can be seen in Fig 2.A1, we can determine any one of these four electrical quantities by knowing the value of any two others. You may want to keep this page bookmarked for your reference. You'll probably be using many of these formulas as the years go by — this has certainly been my experience. — *Dana G. Reed, W1LC*

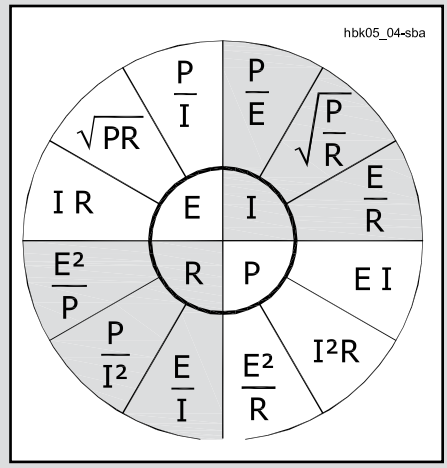


Figure 2.A1 — Electrical formulas.

2.5 Circuit Control Components

2.5.1 Switches

Switches are used to allow or interrupt a current flowing in a particular circuit. Most switches are mechanical devices, although the same effect may be achieved with solid-state devices.

Switches come in many different forms and a wide variety of ratings. The most important ratings are the *voltage-handling* and *current-handling* capabilities. The voltage rating usually includes both the *breakdown voltage rating* and the *interrupt voltage rating*. The breakdown rating is the maximum voltage that the switch can withstand when it is open before the voltage will arc between the switch's terminals. The interrupt voltage rating is the maximum amount of voltage that the switch can interrupt without arcing. Normally, the interrupt voltage rating is the lower value, and therefore the one given for (and printed on) the switch.

Switches typically found in the home are usually rated for 125 V ac and 15 to 20 A. Switches in cars are usually rated for 12 V dc and several amperes. The breakdown voltage rating of a switch primarily depends on the insulating material surrounding the contacts and the separation between the contacts. Plastic or phenolic material normally provides both structural support and insulation. Ceramic material may be used to provide better insulation, particularly in rotary (wafer) switches.

A switch's current rating includes both the *current-carrying capacity* and the *interrupt capability*. The current-carrying capacity of the switch depends on the contact material

and size, and on the pressure exerted to keep the contacts closed. It is primarily determined from the allowable contact temperature rise. On larger ac switches and most dc switches, the interrupt capability is usually lower than the current carrying value.

Most power switches are rated for alternating current use. Because ac current goes through zero twice in each cycle, switches can successfully interrupt much higher alternating currents than direct currents without arcing. A switch that has a 10-A ac current rating may arc and damage the contacts if used to turn off more than an ampere or two of dc.

Switches are normally designated by the number of *poles* (circuits controlled) and *throws* or *positions* (circuit path choices). The simplest switch is the on-off switch, which is a single-pole, single-throw (SPST) switch as shown in **Fig 2.11A**. The off position does not direct the current to another circuit. The next step would be to change the current path to another path. This would be a single-pole, double-throw (SPDT) switch as shown in Fig 2.11B. Adding an off position would give a single-pole, double-throw, center-off switch as shown in Fig 2.11C.

Several such switches can be "ganged" to or actuated by the same mechanical activator to provide double-pole, triple-pole or even more, separate control paths all activated at once. Switches can be activated in a variety of ways. The most common methods include lever or toggle, push-button and rotary switches. Samples of these are shown in **Fig 2.12**. Most switches stay in position once set, but some are spring-loaded so they only stay in the desired position while held there. These are called *momentary* switches.

Rotary/wafer switches can provide very complex switching patterns. Several poles (separate circuits) can be included on each wafer. Many wafers may be stacked on the same shaft. Not only may many different circuits be controlled at once, but by wiring different poles/positions on different wafers together, a high degree of circuit switching logic can be developed. Such switches can select different paths as they are turned and can also "short" together successive contacts to connect numbers of components or paths.

Rotary switches can also be designed to either break one contact before making another (*break-before-make*), or to short two contacts together before disconnecting the first one (*make-before-break*) to eliminate arcing or perform certain logic functions. The two types of switches are generally not interchangeable and may cause damage if inappropriately substituted for one another during circuit construction or repair. When buying rotary switches from a surplus or flea-market vendor, check to be sure the type of switch is correct.

Microswitches are designed to be actuated by the operation of machine components, opening or closing of a door, or some other mechanical movement. Instead of a handle or button-type actuator that would be used by a human, microswitches have levers or buttons more suitable for being actuated as part of an enclosure or machine.

In choosing a switch for a particular task, consideration should be given to function, voltage and current ratings, ease of use, availability and cost. If a switch is to be operated frequently, a better-quality switch is usually less costly over the long run. If signal noise or contact corrosion is a potential problem (usually in low-current signal applications), it is best to get gold-plated contacts. Gold does not oxidize or corrode, thus providing surer contact, which can be particularly important at very low signal levels. Gold plating will not hold up under high-current-interrupt applications, however.

2.5.2 Fuses and Circuit Breakers

Fuses self-destruct to protect circuit wiring or equipment. The fuse *element* that melts or *blows* is a carefully shaped piece of soft metal, usually mounted in a cartridge of some kind. The element is designed to safely carry a given amount of current and to melt at a current value that is a certain percentage above the rated value.

The most important fuse rating is the *nominal current rating* that it will safely carry for an indefinite period without blowing. A fuse's melting current depends on the type of material, the shape of the element and the

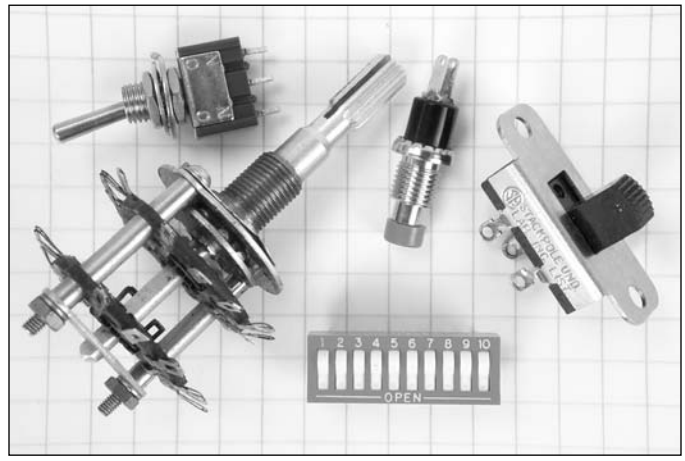


Fig 2.12 — This photo shows examples of various styles of switches. The ¼-inch-ruled graph paper background provides for size comparison.

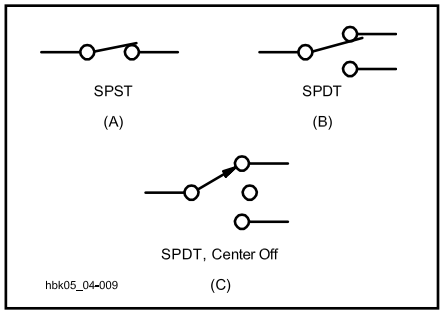


Fig 2.11 — Schematic diagrams of various types of switches. A is an SPST, B is an SPDT, and C is an SPDT switch with a center-off position.

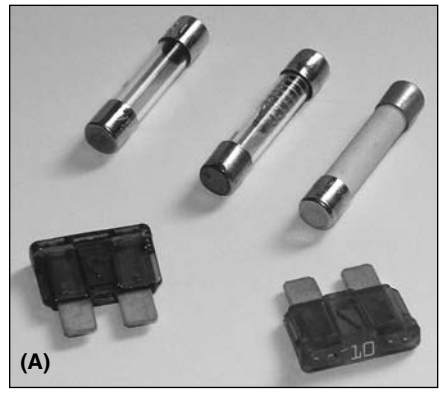


Fig 2.13 — These photos show examples of various styles of fuses. Cartridge-type fuses (A, top) can use glass or ceramic construction. The center fuse is a slow-blow type. Automotive blade-type fuses (A, bottom) are common for low-voltage dc use. A typical home circuit breaker for ac wiring is shown at B.

heat dissipation capability of the cartridge and holder, among other factors.

Next most important are the timing characteristics, or how quickly the fuse element blows under a given current overload. Some fuses (*slow-blow*) are designed to carry an overload for a short period of time. They typically are used in motor-starting and power-supply circuits that have a large inrush current when first started. Other fuses are designed to blow very quickly to protect delicate instruments and solid-state circuits.

A fuse also has a voltage rating, both a value in volts and whether it is expected to be used in ac or dc circuits. The voltage rating is the amount of voltage an open fuse can withstand without arcing. While you should never substitute a fuse with a higher current rating than the one it replaces, you may use a fuse with a higher voltage rating.

Fig 2.13A shows typical cartridge-style cylindrical fuses likely to be encountered in ac-powered radio and test equipment. Automotive style fuses, shown in the lower half of Fig 2.13A, have become widely used in low-voltage dc power wiring of amateur stations. These are called "blade" fuses. Rated for vehicle-level voltages, automotive blade fuses should never be used in ac line-powered circuits.

Circuit breakers perform the same function as fuses — they open a circuit and interrupt current flow when an overload occurs. Instead of a melting element, circuit breakers use spring-loaded magnetic mechanisms to open a switch when excessive current is present. Once the overload has been corrected, the circuit-breaker can be reset. Circuit breakers are generally used by amateurs in home ac wiring (a typical ac circuit breaker is shown in Fig 2.13B) and in dc power supplies.

A replacement fuse or circuit breaker should have the same current rating and the same characteristics as the fuse it replaces. Never substitute a fuse with a larger current rating. You may cause permanent damage (maybe even a fire) to wiring or circuit elements by allowing larger currents to flow when there is an internal problem in equipment. (Additional discussion of fuses and circuit breakers is provided in the chapter on **Safety**.)

Fuses blow and circuit breakers open for several reasons. The most obvious reason is that a problem develops in the circuit, causing too much current to flow. In this case, the circuit problem needs to be fixed. A fuse can fail from being cycled on and off near its current rating. The repeated thermal stress causes metal fatigue and eventually the fuse blows. A fuse can also blow because of a momentary power surge, or even by rapidly turning equipment with a large inrush current on and off several times. In these cases it is only necessary to replace the fuse with the same type and value.

Panel-mount fuse holders should be wired with the hot lead of an ac power circuit (the black wire of an ac power cord) connected to the end terminal, and the ring terminal is connected to the power switch or circuit inside the chassis. This removes voltage from the fuse as it is removed from the fuse holder. This also locates the line connection at the far end of the fuse holder where it is not easily accessible.

2.5.3 Relays And Solenoids

Relays are switches controlled by an electrical signal. *Electromechanical relays* consist of a electromagnetic *coil* and a moving *armature* attracted by the coil's magnetic field when energized by current flowing in the coil. Movement of the armature pushes the switch contacts together or apart. Many sets of contacts can be connected to the same armature, allowing many circuits to be controlled by a single signal. In this manner, the signal voltage that energizes the coil can control circuits carrying large voltages and/or currents.

Relays have two positions or *states* — *energized* and *de-energized*. Sets of contacts called *normally-closed* (NC) are closed when the relay is de-energized and open when it is energized. *Normally-open* contact sets are closed when the relay is energized.

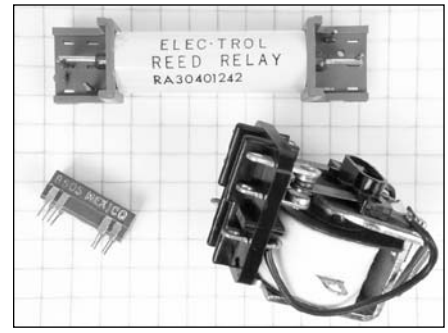
Like switches, relay contacts have breakdown voltage, interrupting, and current-carrying ratings. These are not the same as the voltage and current requirements for energizing the relay's coil. Relay contacts (and housings) may be designed for ac, dc or RF signals. The most common control voltages for relays used in amateur equipment are 12 V dc or 120 V ac. Relays with 6, 24, and 28 V dc, and 24 V ac coils are also common. **Fig 2.14** shows some typical relays found in amateur equipment.

A relay's p*ull-in voltage* is the minimum voltage at which the coil is guaranteed to cause the armature to move and change the relay's state. *Hold-in voltage* is the minimum voltage at which the relay is guaranteed to hold the armature in the energized position after the relay is activated. A relay's pull-in voltage is higher than its hold-in voltage due to magnetic hysteresis of the coil (see the section on magnetic materials below). *Current-sensing relays* activate when the current through the coil exceeds a specific value, regardless of the voltage applied to the coil. They are used when the control signal is a current rather than a voltage.

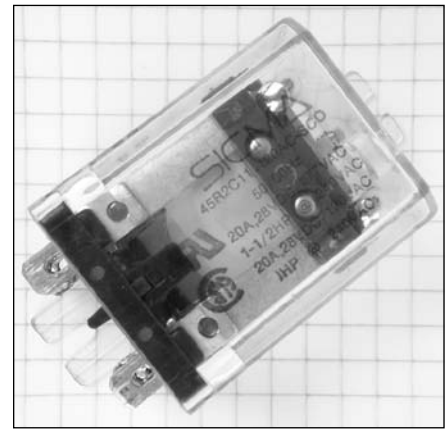
Latching relays have two coils; each moves the armature to a different position where it remains until the other coil is energized. These relays are often used in portable and low-power equipment so that the contact configuration can be maintained without the need to continuously supply power to the relay.

Reed relays have no armature. The contacts are attached to magnetic strips or "reeds" in a glass or plastic tube, surrounded by a coil. The reeds move together or apart when current is applied to the coil, opening or closing contacts. Reed relays can open and close very quickly and are often used in transmit-receive switching circuits.

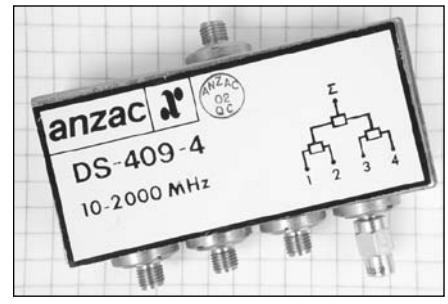
Solid-state relays (SSR) use transistors in-



(A)



(B)



(C)

Fig 2.14 — These photos show examples of various styles and sizes of relays. Photo A shows a large reed relay, and a small reed relay in a package the size of a DIP IC. The contacts and coil can clearly be seen in the open-frame relay. Photo B shows a relay inside a plastic case. Photo C shows a four-position coaxial relay-switch combination with SMA connectors. The ¼-inch-ruled graph paper background provides a size comparison.

stead of mechanical contacts and electronic circuits instead of magnetic coils. They are designed as substitutes for electromechanical relays in power and control circuits and are not used in low-level ac or dc circuits.

Coaxial relays have an armature and contacts designed to handle RF signals. The signal path in coaxial relays maintains a specific characteristic impedance for use in RF systems. Coaxial connectors are used for the RF circuits. Coaxial relays are typically used to control antenna system configurations or to switch a transceiver between a linear amplifier and an antenna.

A *solenoid* is very similar to a relay, except that instead of the moving armature actuating switch contacts, the solenoid moves a lever or rod to actuate some mechanical device. Solenoids are not commonly used in radio equipment, but may be encountered in related systems or devices.

2.6 AC Theory and Waveforms

2.6.1 AC in Circuits

A circuit is a complete conductive path for current to flow from a source, through a load and back to the source. If the source permits the current to flow in only one direction, the current is *dc* or *direct current*. If the source permits the current to change direction, the current is *ac* or *alternating current*. **Fig 2.15** illustrates the two types of circuits. Circuit A shows the source as a battery, a typical dc source. Circuit B shows the more abstract voltage source symbol to indicate ac. In an ac circuit, both the current and the voltage reverse direction. For nearly all ac signals in electronics and radio, the reversal is *periodic*, meaning that the change in direction occurs on a regular basis. The rate of reversal may range from a few times per second to many billions per second.

Graphs of current or voltage, such as Fig 2.15, begin with a horizontal axis that represents time. The vertical axis represents the amplitude of the current or the voltage, whichever is graphed. Distance above the zero line indicates larger positive amplitude; distance below the zero line means larger negative amplitude. Positive and negative only designate the opposing directions in which current may flow in an alternating current circuit or the opposing *polarities* of an ac voltage.

If the current and voltage never change direction, then from one perspective, we have a dc circuit, even if the level of dc constantly changes. **Fig 2.16** shows a current that is always positive with respect to 0. It varies periodically in amplitude, however. Whatever the shape of the variations, the current can be called *pulsating dc*. If the current periodically reaches 0, it can be called *intermittent dc*. From another perspective, we may look at intermittent and pulsating dc as a combination of an ac and a dc current. Special circuits can separate the two currents into ac and dc *components* for separate analysis or use. There are circuits that combine ac and dc currents and voltages, as well.

2.6.2 AC Waveforms

A *waveform* is the pattern of amplitudes reached by an ac voltage or current as measured over time. The combination of ac and dc voltages and currents results in *complex waveforms*. **Fig 2.17** shows two ac waveforms fairly close in frequency and their combination. **Fig 2.18** shows two ac waveforms dissimilar in both frequency and wavelength, along with the resultant combined waveform. Note the similarities (and the differences) between the resultant waveform in Fig 2.18 and the combined ac-dc waveform in Fig 2.16.

Alternating currents may take on many useful waveforms. **Fig 2.19** shows a few that are commonly used in electronic and radio circuits. The *sine wave* is both mathematically and practically the foundation of all other forms of ac; the other forms can usually be reduced to (and even constructed from) a particular collection of sine waves. The *square wave* is vital to digital electronics. The *triangular* and *ramp waves* — the latter sometimes called a *sawtooth* waveform — are especially useful in timing circuits. The individual repeating patterns that make up these *periodic waveforms* are called *cycles*. Waveforms that do not consist of repetitive patterns are called *aperiodic* or *irregular* waveforms. Speech is

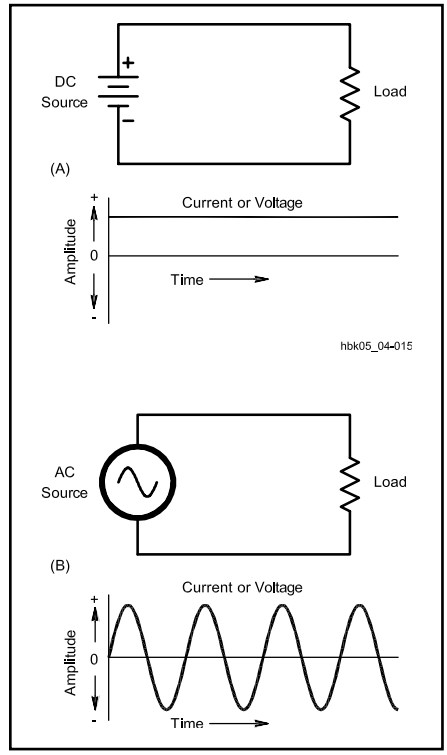


Fig 2.15 — Basic circuits for direct and alternating currents. With each circuit is a graph of the current, constant for the dc circuit, but periodically changing direction in the ac circuit.

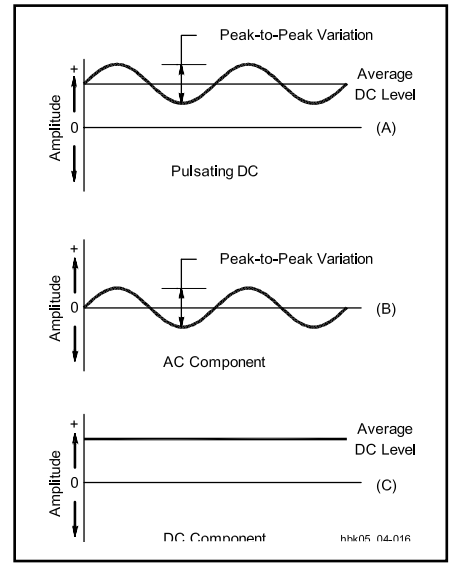


Fig 2.16 — A pulsating dc current (A) and its resolution into an ac component (B) and a dc component (C).

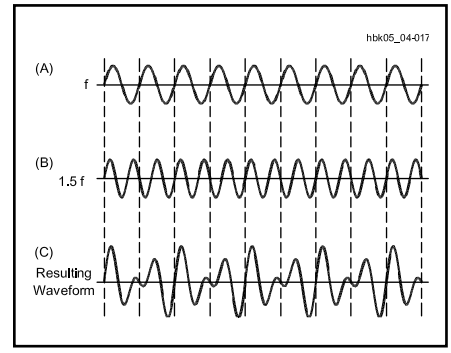


Fig 2.17 — Two ac waveforms of similar frequencies (f1 = 1.5 f2) and amplitudes form a composite wave. Note the points where the positive peaks of the two waves combine to create high composite peaks at a frequency that is the difference between f1 and f2. The beat note frequency is 1.5f – f = 0.5f and is visible in the drawing.

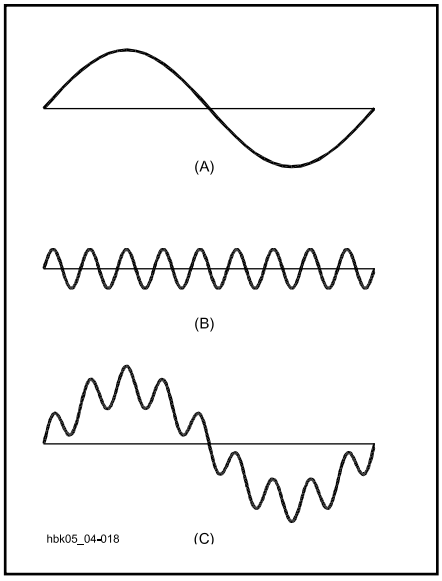


Fig 2.18 — Two ac waveforms of widely different frequencies and amplitudes form a composite wave in which one wave appears to ride upon the other.

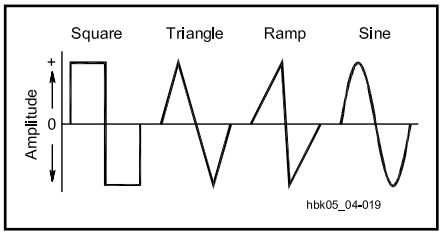


Fig 2.19 — Some common ac waveforms: square, triangle, ramp and sine.

a good example of an irregular waveform, as is noise or static.

There are numerous ways to generate alternating currents: with an ac power generator (an *alternator*), with a *transducer* (for example, a microphone) or with an electronic circuit (for example, an RF oscillator). The basis of the sine wave is circular or cyclical motion, which underlies the most usual methods of generating alternating current. The circular motion of the ac generator may be physical or mechanical, as in an alternator. Currents in the resonant circuit of an oscillator may also produce sine waves as repetitive electrical cycles without mechanical motion.

SINE WAVES AND CYCLICAL MOTION

The relationship between circular motion and the sine wave is illustrated in **Fig 2.20** as the correlation between the amplitude of an ac current (or voltage) and the relative positions of a point making circular rotations through one complete revolution of 360º. Following the height of the point above the horizontal axis, its height is zero at point 1. It rises to a maximum at a position 90º from position 1, which is position 3. At position 4, 180º from position 1, the point's height falls back to zero. Then the point's height begins to increase again, but below the horizontal axis, where it reaches a maximum at position 5.

An alternative way of visualizing the relationship between circular motion and sine waves is to use a yardstick and a light source, such as a video projector. Hold the yardstick horizontally between the light source and the wall so that it points directly at the light source. This is defined as the 0° position and the length of the yardstick's shadow is zero. Now rotate the yardstick, keeping the end farthest from the projector stationary, until the yardstick is vertical. This is the 90° position and the shadow's length is a maximum in the direction defined as positive. Continue to rotate the yardstick until it reaches horizontal again. This is the 180° position and the shadow once again has zero length. Keep rotating the yardstick in the same direction until it is vertical and the shadow is at maximum, but now below the stationary end of the yardstick. This is the 270° position and the direction of the shadow is defined to be negative. Another 90° of rotation brings the pointer back to its original position at 0° and the shadow's length to zero. If the length of the shadow is plotted against the amount of rotation of the yardstick, the result is a sine wave.

The corresponding rise and fall of a *sinusoidal* or sine wave current (or voltage) along a linear time line produces the curve accompanying the circle in Fig 2.20. The curve is sinusoidal because its amplitude varies as the sine of the angle made by the circular movement with respect to the zero position. This is often referred to as the *sine function*, written sin(angle). The sine of 90° is 1, and 90° is also the position of maximum current (along with 270°, but with the opposite polarity). The sine of 45° (point 2) is 0.707, and the value of current at point 2, the 45° position of rotation, is 0.707 times the maximum current. Both ac current and voltage sine waves vary in the same way.

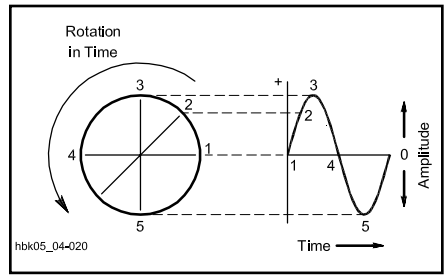


Fig 2.20 — The relationship of circular motion and the resultant graph of ac current or voltage. The curve is sinusoidal, a sine wave.

FREQUENCY AND PERIOD

With a continuously rotating generator, alternating current or voltage will pass through many equal cycles over time. This is a *periodic waveform*, composed of many 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 the 19th century radio-phenomena pioneer, Heinrich Hertz).

The length of any cycle in units of time is the *period* of the cycle, as measured between equivalent points on succeeding cycles. Mathematically, the period is the inverse of frequency. That is,

$$\text{Frequency (f) in Hz} = \frac{1}{\text{Period (T) in seconds}} \tag{23}$$

and

$$\text{Period (T) in seconds} = \frac{1}{\text{Frequency (f) in Hz}} \tag{24}$$

Example: What is the period of a 400-hertz ac current?

$$T = \frac{1}{f} = \frac{1}{400\ \text{Hz}} = 0.00250\ \text{s} = 2.5\ \text{ms}$$

The frequency of alternating currents used in radio circuits varies from a few hertz, or cycles per second, to thousands of millions of hertz. Likewise, the period of alternating currents amateurs use ranges from significant fractions of a second down to nanoseconds or smaller. In order to express compactly the units of frequency, time and almost everything else in electronics, a standard system of prefixes is used. In magnitudes of 1000 or 10^3, frequency is measurable in hertz, in kilohertz (1000 hertz or kHz), in megahertz (1 million hertz or MHz), gigahertz (1 billion hertz or GHz — pronounced "gee-gah" with a hard 'g' as in 'golf') and even in terahertz (1 trillion hertz or THz). For units smaller than unity, as in the measurement of period, the basic unit can be milliseconds (1 thousandth of a second or ms), microseconds (1 millionth of a second or μs), nanoseconds (1 billionth of a second or ns) and picoseconds (1 trillionth of a second or ps).

It is common for complex ac signals to contain a series of sine waves with frequencies related by integer multiples of some lowest or *fundamental* frequency. Sine waves with the higher frequency are called harmonics and are

said to be *harmonically related* to the fundamental. 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.

The uses of ac in radio circuits are many and varied. Most can be cataloged by reference to ac frequency ranges used in circuits. For example, the frequency of ac power used in the home, office and factory is ordinarily 60 Hz in the United States and Canada. In Great Britain and much of Europe, ac power is 50 Hz. Sonic and ultrasonic applications of ac run from about 20 Hz (audio) up to several MHz.

Radio circuits include both power- and sonic-frequency-range applications. Radio communication and other electronics work, however, require ac circuits capable of operation with frequencies up to the GHz range. Some of the applications include signal sources for transmitters (and for circuits inside receivers); industrial induction heating; diathermy; microwaves for cooking, radar and communication; remote control of appliances, lighting, model planes and boats and other equipment; and radio direction finding and guidance.

PHASE

When drawing the sine-wave curve of an ac voltage or current, the horizontal axis represents time. This type of drawing in which the amplitude of a waveform is shown compared to time is called the *time domain*. Events to the right take place later; events to the left occur earlier. 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 counting degrees in a sinusoidal waveform such as ac is the *zero point* at which the voltage or current begins the positive *half cycle*. The essential elements of an ac cycle appear in **Fig 2.21**.

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.

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.*

Fig 2.22 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 Fig 2.17, creates a complex waveform with a *beat frquency* that is the difference between the two sine waves.

Fig 2.22 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 **Fig 2.23**. 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*.

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 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.

Phase also has other meanings. It is important to distinguish between *polarity*, meaning the sense in which positive and negative voltage or current relate to circuit operation, and phase, which 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.

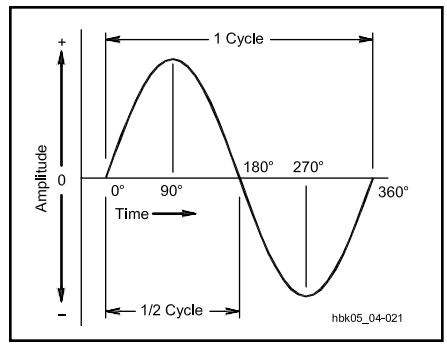


Fig 2.21 — An ac cycle is divided into 360° that are used as a measure of time or phase.

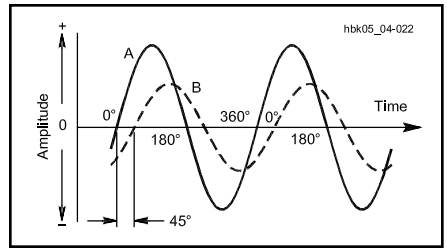


Fig 2.22 — When two waves of the same frequency start their cycles at slightly different times, the time difference or phase difference is measured in degrees. In this drawing, wave B starts 45° (one-eighth cycle) later than wave A, and so lags 45° behind A.

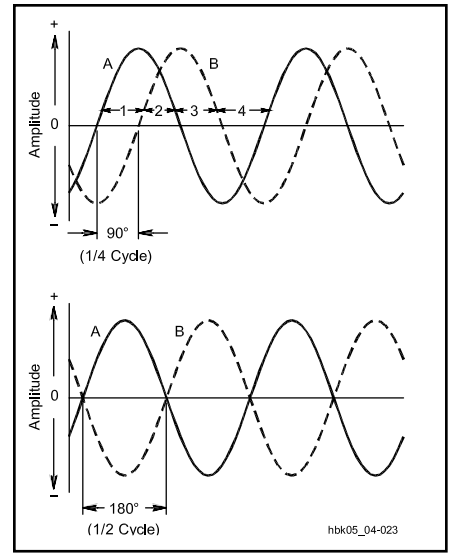


Fig 2.23 — Two important special cases of phase difference: In the upper drawing, the phase difference between A and B is 90°; in the lower drawing, the phase difference is 180°.

Degrees and Radians

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 equations and computations. Radians are used for angular measurements when the *angular frequency* (ω) is being used. There are 2π radians in a circle, just as there are 360°, so one radian = $360/2\pi \approx 57.3°$. Angular frequency is measured in radians/second (not revolutions/second or cycles/second) so $\omega = 2\pi f$. In this *Handbook*, unless it is specifically noted otherwise, the convention will be to use degrees in all calculations that require an angle.

2.6.3 Electromagnetic Energy

Alternating currents are often loosely classified as audio frequency (AF) and radio frequency (RF). Although these designations are handy, they actually represent something other than electrical energy: They designate special forms of energy that we find useful.

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.

RF energy occurs at frequencies for which it is practical to generate and detect *electromagnetic* or *RF 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 2.3** provides a brief glimpse at the total spectrum of electromagnetic energy. The *radio spectrum* is generally considered to begin around 3 kHz and end at infrared light.

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.

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. Electromagnetic-energy waves have a length uniquely associated with each possible frequency. The *wavelength* (λ) is the speed of propagation divided by the frequency (f) in hertz.

$$f\,(\text{Hz}) = \frac{3.0 \times 10^8 \left(\frac{\text{m}}{\text{s}}\right)}{\lambda\,(\text{m})} \qquad (25)$$

and

$$\lambda\,(\text{m}) = \frac{3.0 \times 10^8 \left(\frac{\text{m}}{\text{s}}\right)}{f\,(\text{Hz})} \qquad (26)$$

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

$$f\,(\text{Hz}) = \frac{3.0 \times 10^8 \left(\frac{\text{m}}{\text{s}}\right)}{\lambda\,(\text{m})}$$

$$= \frac{3.0 \times 10^8 \left(\frac{\text{m}}{\text{s}}\right)}{80.0\ \text{m}}$$

$$= 3.75 \times 10^6\ \text{Hz}$$

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 ºC and 344 m/s at 20 ºC.

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

$$f\,(\text{kHz}) = \frac{3.0 \times 10^5 \left(\frac{\text{km}}{\text{s}}\right)}{\lambda\,(\text{m})} \qquad (27)$$

and

$$\lambda\,(\text{m}) = \frac{3.0 \times 10^5 \left(\frac{\text{km}}{\text{s}}\right)}{f\,(\text{kHz})} \qquad (28)$$

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

$$f\,(\text{MHz}) = \frac{300 \left(\frac{\text{Mm}}{\text{s}}\right)}{\lambda\,(\text{m})} \qquad (29)$$

and

$$\lambda\,(\text{m}) = \frac{300 \left(\frac{\text{Mm}}{\text{s}}\right)}{f\,(\text{MHz})} \qquad (30)$$

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.

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.

Within the part of the electromagnetic-energy spectrum of most interest to radio applications, frequencies have been classified into groups and given names. **Table 2.4** 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

Table 2.3

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

Note: The range of radio frequencies can also be written as 3 kHz to 300 GHz

Table 2.4

Classification of the Radio Frequency Spectrum

Abbreviation	*Classification*	*Frequency Range*			
VLF	Very low frequencies	3	to	30	kHz
LF	Low frequencies	30	to	300	kHz
MF	Medium frequencies	300	to	3000	kHz
HF	High frequencies	3	to	30	MHz
VHF	Very high frequencies	30	to	300	MHz
UHF	Ultrahigh frequencies	300	to	3000	MHz
SHF	Superhigh frequencies	3	to	30	GHz
EHF	Extremely high frequencies	30	to	300	GHz

3 to 30 MHz, all exhibit *skip* or 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.

2.6.4 Measuring AC Voltage, Current and Power

Measuring the voltage or current in a dc circuit is straightforward, as **Fig 2.24A** demonstrates. Since the current flows in only one direction, the voltage and current have constant values until the resistor values are changed.

Fig 2.24B 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.

INSTANTANEOUS VOLTAGE AND CURRENT

By far, the most common waveform associated with ac of any frequency is the sine wave. Unless otherwise noted, it is safe to assume that measurements of ac voltage or current are of a sinusoidal waveform. **Fig 2.25** 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 Fig 2.25, 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,

$$E_{inst} = E_{max} \sin (ft) \quad (31)$$

assuming all sine calculations are done in degrees. (See the sidebar "Degrees and Radians".) If the sine calculation is done in radians, substitute $2\pi ft$ for ft in equation 31.

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

$$E_{inst} = E_{max} \sin \theta \quad (32)$$

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 Fig 2.25, if the maximum voltage value is 120 V and point D's phase is 60.0°?

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

$$= 120 \times 0.866 = 104 \text{ V}$$

PEAK AND PEAK-TO-PEAK VOLTAGE AND CURRENT

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 Fig 2.25, 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

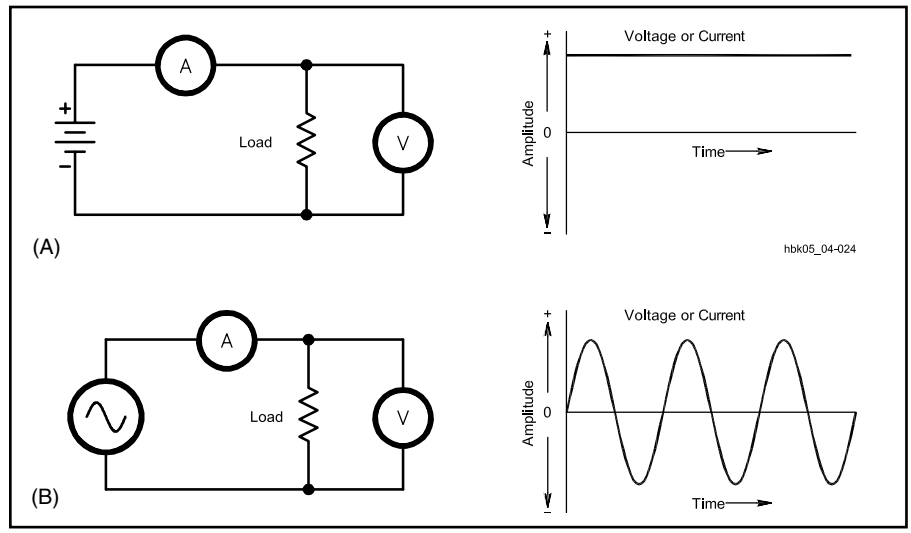


Fig 2.24 — Voltage and current measurements in dc and ac circuits.

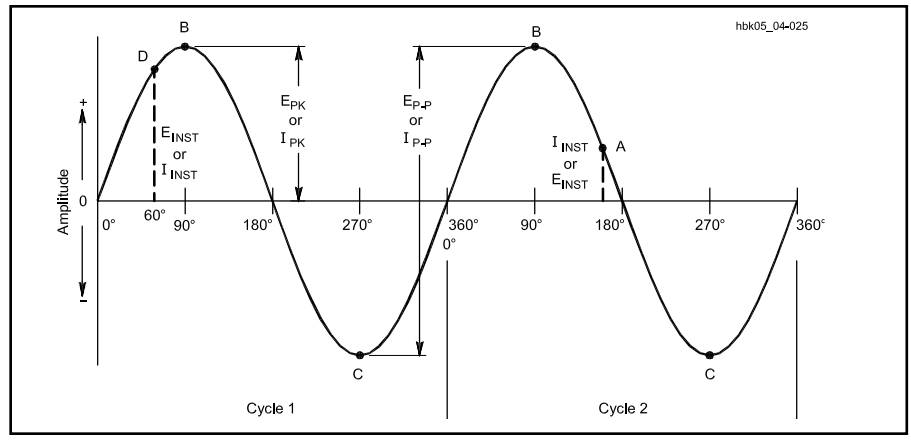


Fig 2.25 — Two cycles of a sine wave to illustrate instantaneous, peak, and peak-to-peak ac voltage and current values.

product of the peak voltage and the peak current, or

$$P_{pk} = E_{pk} \times I_{pk} \quad (33)$$

The span from points B to C in Fig 2.25 represents the largest difference in value of the sine wave. Designated the *peak-to-peak* value (indicated by a "P-P" subscript), this span is equal to twice the peak value of the waveform. Thus, peak-to-peak voltage is:

$$E_{P-P} = 2\,E_{pk} \quad (34)$$

Amplifying devices often specify their input limits in terms of peak-to-peak voltages. Operational amplifiers, which have almost unlimited gain, often require input-level limiting to prevent the output signals from distorting if they exceed the peak-to-peak output rating of the devices.

RMS VALUES OF VOLTAGES AND CURRENTS

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. 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. This procedure produces the RMS value of voltage or current for all waveforms, sinusoidal or not. (The waveform is assumed to be periodic.)

For the remainder of this discussion, remember that we are assuming that 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 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.

$$P_{ave} = \frac{P_{pk}}{2} \quad (35)$$

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:

$$E_{RMS} = \frac{E_{pk}}{\sqrt{2}} = \frac{E_{pk}}{1.414} = E_{pk} \times 0.707 \quad (36)$$

$$I_{RMS} = \frac{I_{pk}}{\sqrt{2}} = \frac{I_{pk}}{1.414} = I_{pk} \times 0.707 \quad (37)$$

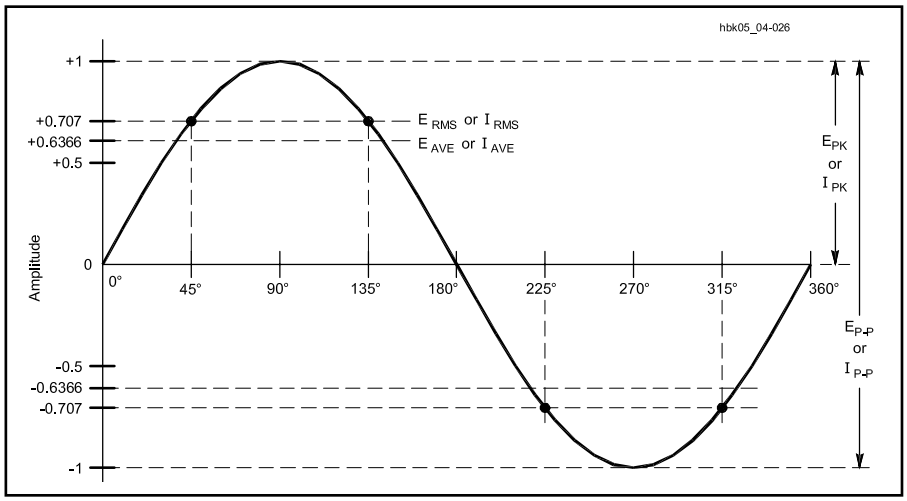


Fig 2.26 — The relationships between RMS, average, peak, and peak-to-peak values of ac voltage and current.

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

$$E_{pk} = E_{RMS} \times 1.414 \quad (38)$$

$$I_{pk} = I_{RMS} \times 1.414 \quad (39)$$

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?

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

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

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 **Fig 2.26**. (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 stress components.

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

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

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 Supplies** chapter for more information on specifying components in this application.)

AVERAGE VALUES OF AC VOLTAGE AND CURRENT

Certain kinds of circuits respond to the *average* voltage or current (not power) of an ac waveform. Among these circuits are 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

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.

of a sine wave are $2/\pi$ (or 0.6366) times the peak value.

$$E_{ave} = 0.6366\, E_{pk} \quad (40)$$

$$I_{ave} = 0.6366\, I_{pk} \quad (41)$$

For convenience, **Table 2.5** summarizes the relationships between all of the common ac values. All of these relationships apply only to sine waves.

COMPLEX WAVEFORMS AND PEAK-ENVELOPE VALUES

Complex waveforms, as shown earlier in Fig 2.18, differ from sine waves. The amplitude of 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.

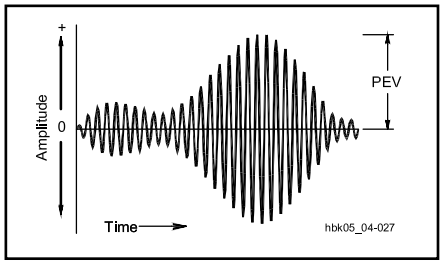


Fig 2.27 — The peak envelope voltage (PEV) for a composite waveform.

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 **Fig 2.27**. 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.

$$PEP = \frac{(PEV \times 0.707)^2}{R} \quad (42)$$

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

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

2.6.5 – Glossary — AC Theory and Reactance

Admittance (Y) — The reciprocal of impedance, measured in siemens (S).

Capacitance (C) — The ability to store electrical energy in an electrostatic field, measured in farads (F). A device with capacitance is a capacitor.

Flux density (B) — The number of magnetic-force lines per unit area, measured in gauss.

Frequency (f) — The rate of change of an ac voltage or current, measured in cycles per second, or hertz (Hz).

Fundamental — The lowest frequency in a series of sine waves whose frequencies have an integer relationship.

Harmonic — A sine wave whose frequency is an integer multiple of a fundamental frequency.

Impedance (Z) — The complex combination of resistance and reactance, measured in ohms (Ω).

Inductance (L) — The ability to store electrical energy in a magnetic field, measured in henrys (H). A device, such as a coil of wire, with inductance is an inductor.

Peak (voltage or current) — The maximum value relative to zero that an ac voltage or current attains during any cycle.

Peak-to-peak (voltage or current) — The value of the total swing of an ac voltage

Table 2.5
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.

or current from its peak negative value to its peak positive value, ordinarily twice the value of the peak voltage or current.

Period (T) — The duration of one ac voltage or current cycle, measured in seconds (s).

Permeability (μ) — The ratio of the magnetic flux density of an iron, ferrite, or similar core in an electromagnet compared to the magnetic flux density of an air core, when the current through the electromagnet is held constant.

Power (P) — The rate of electrical-energy use, measured in watts (W).

Q (quality factor) — The ratio of energy stored in a reactive component (capacitor or inductor) to the energy dissipated, equal to the reactance divided by the resistance.

Reactance (X) — Opposition to alternating current by storage in an electrical field (by a capacitor) or in a magnetic field (by an inductor), measured in ohms (Ω).

Resonance — Ordinarily, the condition in an ac circuit containing both capacitive and inductive reactance in which the reactances are equal.

RMS (voltage or current) — Literally, "root mean square," the square root of the average of the squares of the instantaneous values for one cycle of a waveform. A dc voltage or current that will produce the same heating effect as the waveform. For a sine wave, the RMS value is equal to 0.707 times the peak value of ac voltage or current.

Susceptance (B) — The reciprocal of reactance, measured in siemens (S).

Time constant (τ) — The time required for the voltage in an RC circuit or the current in an RL circuit to rise from zero to approximately 63.2% of its maximum value or to fall from its maximum value 63.2% toward zero.

Toroid — Literally, any donut-shaped solid; most commonly referring to ferrite or powdered-iron cores supporting inductors and transformers.

Transducer — Any device that converts one form of energy to another; for example an antenna, which converts electrical energy to electromagnetic energy or a speaker, which converts electrical energy to sonic energy.

Transformer — A device consisting of at least two coupled inductors capable of transferring energy through mutual inductance.

2.7 Capacitance and Capacitors

It is possible to build up and hold an electrical charge in an *electrostatic field*. This phenomenon is called *capacitance*, and the devices that exhibit capacitance are called *capacitors*. (Old articles and texts use the obsolete term *condenser*.) **Fig 2.28** shows several schematic symbols for capacitors. Part A shows a fixed capacitor; one that has a single value of capacitance. Part B shows the symbol for variable capacitors; these are adjustable over a range of values. If the capacitor is of a type that is *polarized*, meaning that dc voltages must be applied with a specific polarity, the straight line in the symbol should be connected to the most positive voltage, while the curved line goes to the more negative voltage, which is often ground. For clarity, the positive terminal of a polarized capacitor symbol is usually marked with a + symbol. The symbol for *non-polarized* capacitors may be two straight lines or the + symbol may be omitted. When in doubt, consult the capacitor's specifications or the circuits parts list.

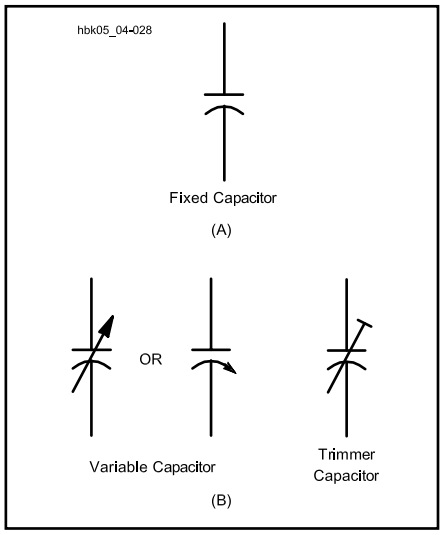


Fig 2.28 — Schematic symbol for a fixed capacitor is shown at A. The symbols for a variable capacitor are shown at B.

2.7.1 Electrostatic Fields and Energy

An *electrostatic field* is created wherever a voltage exists between two points, such as two opposite electric charges or regions that contain different amounts of charge. The field causes electric charges (such as electrons or ions) in the field to feel a force in the direction of the field. If the charges are not free to move, as in an insulator, they store the field's energy as *potential energy*, just as a weight held in place by a surface stores gravitational energy. If the charges are free to move, the field's stored energy is converted to *kinetic energy* of motion just as if the weight is released to fall in a gravitational field.

The field is represented by *lines of force* that show the direction of the force felt by the electric charge. Each electric charge is surrounded by an electric field. The lines of force of the field begin on the charge and extend away from charge into space. The lines of force can terminate on another charge (such as lines of force between a proton and an electron) or they can extend to infinity.

The strength of the electrostatic field is measured in *volts per meter* (V/m). Stronger fields cause the moving charges to accelerate more strongly (just as stronger gravity causes weights to fall faster) and stores more energy in fixed charges. The stronger the field in V/m, the more force an electric charge in the field will feel. The strength of the electric field diminishes with the square of the distance from its source, the electric charge.

2.7.2 The Capacitor

Suppose two flat metal plates are placed close to each other (but not touching) and are connected to a battery through a switch, as illustrated in **Fig 2.29A**. At the instant the switch is closed, electrons are attracted from the upper plate to the positive terminal of the battery, while the same quantity is repelled from the negative battery terminal and pushed into the lower plate. This imbalance of charge creates a voltage between the plates. Eventually, enough electrons move into one plate and out of the other to make the voltage between the plates the same as the battery voltage.

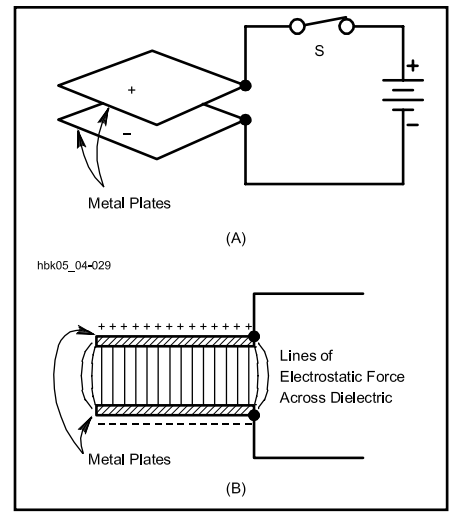


Fig 2.29 — A simple capacitor showing the basic charging arrangement at A, and the retention of the charge due to the electrostatic field at B.

At this point, the voltage between the plates opposes further movement of electrons and no further current flow occurs.

If the switch is opened after the plates have been charged in this way, the top plate is left with a deficiency of electrons and the bottom plate with an excess. Since there is no current path between the two, the plates remain *charged* despite the fact that the battery no longer is connected to a source of voltage. As illustrated in Fig 2.29B, the separated charges create an electrostatic field between the plates. The electrostatic field contains the energy that was expended by the battery in causing the electrons to flow off of or onto the plates. These two plates create a *capacitor*, a device that has the property of storing electrical energy in an electric field, a property called *capacitance*.

The amount of electric charge that is held on the capacitor plates is proportional to the applied voltage and to the capacitance of the capacitor:

$$Q = CV \tag{43}$$

where

Q = charge in coulombs,
C = capacitance in farads (F), and
V = electrical potential in volts. (The symbol E is also commonly used instead of V in this and the following equation.)

The energy stored in a capacitor is also a function of voltage and capacitance:

$$W = \frac{V^2 C}{2} \tag{44}$$

where W = energy in joules (J) or watt-seconds.

If a wire is simultaneously touched to the two plates (short circuiting them), the voltage between the plates causes the excess electrons on the bottom plate to flow through the wire to the upper plate, restoring electrical neutrality. The plates are then *discharged*.

Fig 2.30 illustrates the voltage and current in the circuit, first, at the moment the switch is closed to charge the capacitor and, second, at the moment the shorting switch is closed to discharge the capacitor. Note that the periods of charge and discharge are very short, but that they are not zero. This finite charging and discharging time can be controlled and that will prove useful in the creation of timing circuits.

During the time the electrons are moving — that is, while the capacitor is being charged or discharged — a current flows in the circuit even though the circuit apparently is broken by the gap between the capacitor plates. The current flows only during the time of charge and discharge, however, and this time is usually very short. There is no continuous flow of direct current through a capacitor.

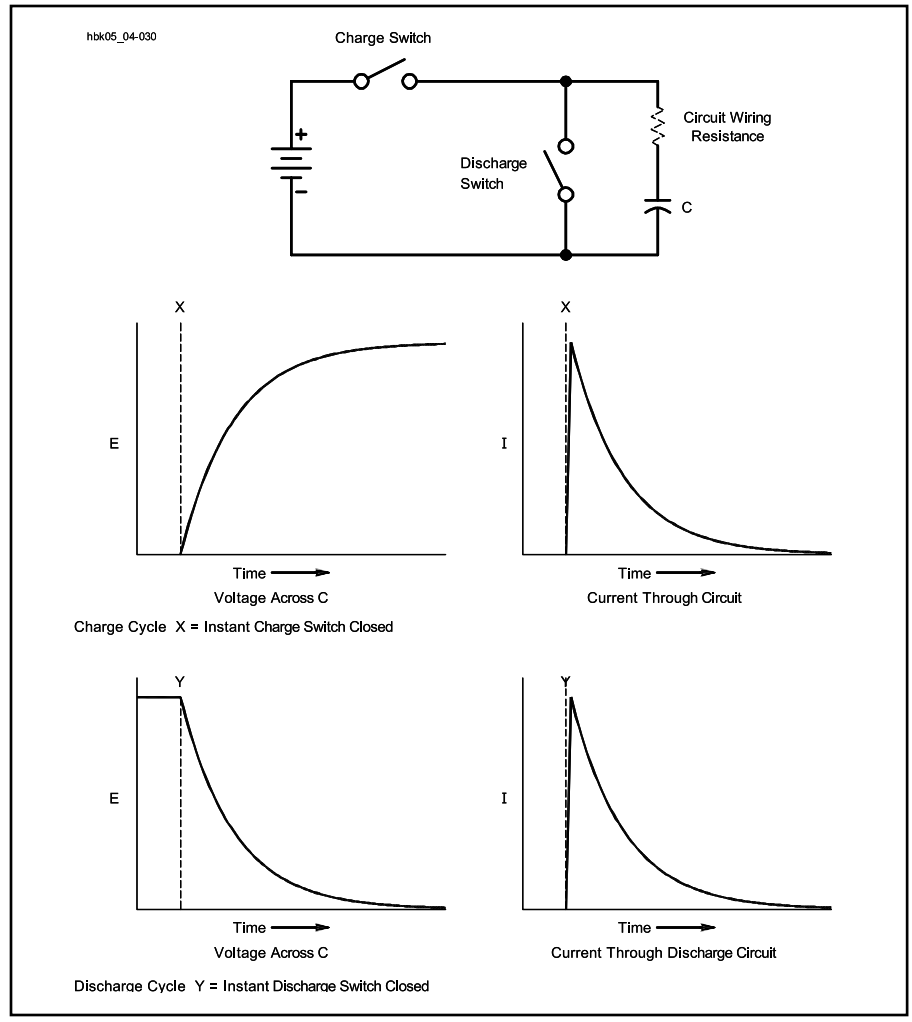


Fig 2.30 — The flow of current during the charge and discharge of a capacitor. The charging graphs assume that the charge switch is closed and the discharge switch is open. The discharging graphs assume just the opposite.

Although dc cannot pass through a capacitor, alternating current can. At the same time one plate is charged positively by the positive excursion of the alternating current, the other plate is being charged negatively at the same rate. (Remember that conventional current is shown as the flow of positive charge, equal to and opposite the actual flow of electrons.) The reverse process occurs during the second half of the cycle as the changing polarity of the applied voltage causes the flow of charge to change direction, as well. The continual flow into and out of the capacitor caused by ac voltage appears as an ac current, although with a phase difference between the voltage and current flow as described below.

UNITS OF CAPACITANCE

The basic unit of capacitance, the ability to store electrical energy in an electrostatic field, is the *farad*. This unit is generally too large for practical radio circuits, although capacitors of several farads in value are used in place of small batteries or as a power supply filter for automotive electronics. Capacitance encountered in radio and electronic circuits is usually measured in microfarads (abbreviated μF), nanofarads (abbreviated nF) or picofarads (pF). The microfarad is one millionth of a farad (10^{-6} F), the nanofarad is one thousandth of a microfarad (10^{-9} F) and the picofarad is one millionth of a microfarad (10^{-12} F). Old articles and texts use the obsolete term micromicrofarad (μμF) in place of picofarad.

CAPACITOR CONSTRUCTION

An ideal capacitor is a pair of parallel metal plates separated by an insulating or *dielectric* layer, ideally a vacuum. The capacitance of a vacuum-dielectric capacitor is given by

$$C = \frac{A\,\varepsilon_r\,\varepsilon_0}{d} \tag{45}$$

where

C = capacitance, in farads
A = area of plates, in cm^2
d = spacing of the plates in cm

ε_r = dielectric constant of the insulating material

ε_0 = permittivity of free space, 8.85 × 10^{-14} F/cm.

The actual capacitance of such a parallel-plate capacitor is somewhat higher due to *end effect* caused by the electric field that exists just outside the edges of the plates.

The *larger* the plate area and the *smaller* the spacing between the plates, the *greater* the amount of energy that can be stored for a given voltage, and the *greater* the capacitance. The more general name for the capacitor's plates is *electrodes*. However, amateur radio literature generally refers to a capacitor's electrodes as plates and that is the convention in this text.

The amount of capacitance also depends on the material used as insulating material between the plates; capacitance is smallest with air or a vacuum as the insulator. Substituting other insulating materials for air may greatly increase the capacitance.

The ratio of the capacitance with a material other than a vacuum or air between the plates to the capacitance of the same capacitor with air insulation is called the *dielectric constant*, or K, of that particular insulating material. The dielectric constants of a number of materials commonly used as dielectrics in capacitors are given in **Table 2.6**. For example, if polystyrene is substituted for air in a capacitor, the capacitance will be 2.6 times greater.

In practice, capacitors often have more than two plates, with alternating plates being connected in parallel to form two sets, as shown in **Fig 2.31**. This practice makes it possible to obtain a fairly large capacitance in a small space, since several plates of smaller individual area can be stacked to form the equivalent of a single large plate of the same total area. Also, all plates except the two on the ends of the stack are exposed to plates of the other group on both sides, and so are twice as effective in increasing the capacitance.

The formula for calculating capacitance from these physical properties is:

$$C = \frac{0.2248\ K\ A\ (n-1)}{d} \qquad (46)$$

where

C = capacitance in pF,

K = dielectric constant of material between plates,

A = area of one side of one plate in square inches,

d = separation of plate surfaces in inches, and

n = number of plates.

If the area (A) is in square centimeters and the separation (d) is in centimeters, then the formula for capacitance becomes

$$C = \frac{0.0885\ K\ A\ (n-1)}{d} \qquad (47)$$

If the plates in one group do not have the same area as the plates in the other, use the area of the smaller plates.

Example: What is the capacitance of two copper plates, each 1.50 square inches in area, separated by a distance of 0.00500 inch, if the dielectric is air?

$$C = \frac{0.2248\ K\ A\ (n-1)}{d}$$

$$C = \frac{0.2248 \times 1 \times 1.50\ (2-1)}{0.00500}$$

$$C = 67.4\ \text{pF}$$

What is the capacitance if the dielectric is mineral oil? (See Table 2.6 for the appropriate dielectric constant.)

$$C = \frac{0.2248 \times 2.23 \times 1.50\ (2-1)}{0.00500}$$

$$C = 150.3\ \text{pF}$$

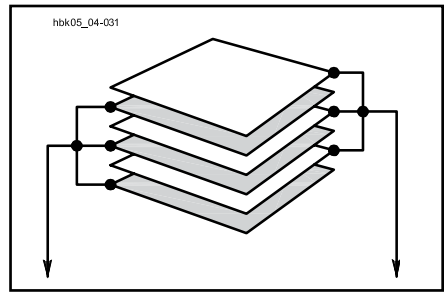


Fig 2.31 — A multiple-plate capacitor. Alternate plates are connected to each other, increasing the total area available for storing charge.

2.7.3 Capacitors in Series and Parallel

When a number of capacitors are connected in parallel, as in **Fig 2.32A**, the total capacitance of the group is equal to the sum of the individual capacitances:

$$C_{total} = C1 + C2 + C3 + C4 + ... + C_n \qquad (48)$$

When two or more capacitors are con-

Table 2.6

Relative Dielectric Constants of Common Capacitor Dielectric Materials

Material	*Dielectric Constant (k)*	*(O)rganic or (I)norganic*
Vacuum	1 (by definition)	I
Air	1.0006	I
Ruby mica	6.5 - 8.7	I
Glass (flint)	10	I
Barium titanate (class I)	5 - 450	I
Barium titanate (class II)	200 - 12000	I
Kraft paper	≈ 2.6	O
Mineral Oil	≈ 2.23	O
Castor Oil	≈ 4.7	O
Halowax	≈ 5.2	O
Chlorinated diphenyl	≈ 5.3	O
Polyisobutylene	≈ 2.2	O
Polytetrafluoroethylene	≈ 2.1	O
Polyethylene terephthalate	≈ 3	O
Polystyrene	≈ 2.6	O
Polycarbonate	≈ 3.1	O
Aluminum oxide	≈ 8.4	I
Tantalum pentoxide	≈ 28	I
Niobium oxide	≈ 40	I
Titanium dioxide	≈ 80	I

(Adapted from: Charles A. Harper, *Handbook of Components for Electronics*, p 8-7.)

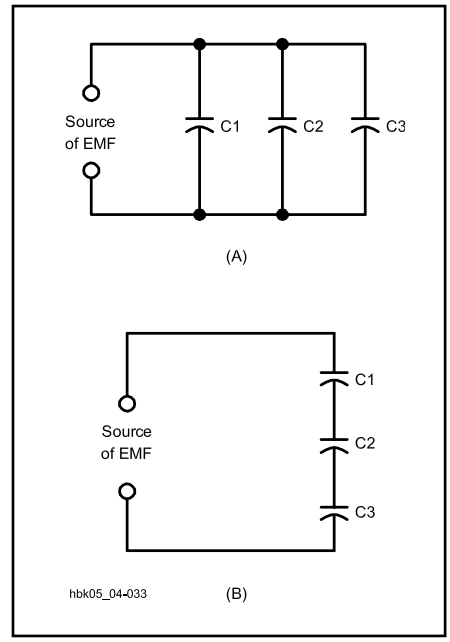


Fig 2.32 — Capacitors in parallel are shown at A, and in series at B.

nected in series, as in Fig 2.32B, the total capacitance is less than that of the smallest capacitor in the group. The rule for finding the capacitance of a number of series-connected capacitors is the same as that for finding the resistance of a number of parallel-connected resistors.

$$C_{total} = \frac{1}{\frac{1}{C1}+\frac{1}{C2}+\frac{1}{C3}+...+\frac{1}{C_n}} \quad (49)$$

For only two capacitors in series, the formula becomes:

$$C_{total} = \frac{C1 \times C2}{C1 + C2} \quad (50)$$

The same units must be used throughout; that is, all capacitances must be expressed in μF, nF or pF, etc. Different units cannot be combined in the same equation.

Capacitors are often connected in parallel to obtain a larger total capacitance than is available in one unit. The voltage rating of capacitors connected in parallel is the lowest voltage rating of any of the capacitors.

When capacitors are connected in series, the applied voltage is divided between them according to Kirchhoff's Voltage Law. The situation is much the same as when resistors are in series and there is a voltage drop across each. The voltage that appears across each series-connected capacitor is inversely proportional to its capacitance, as compared with the capacitance of the whole group. (This assumes ideal capacitors.)

Example: Three capacitors having capacitances of 1, 2 and 4 μF, respectively, are connected in series as shown in **Fig 2.33**. The voltage across the entire series is 2000 V. What is the total capacitance? (Since this is a calculation using theoretical values to illustrate a technique, we will not follow the rules of significant figures for the calculations.)

$$C_{total} = \frac{1}{\frac{1}{C1}+\frac{1}{C2}+\frac{1}{C3}}$$

$$= \frac{1}{\frac{1}{1\,\mu F}+\frac{1}{2\,\mu F}+\frac{1}{4\,\mu F}}$$

$$= \frac{1}{\frac{7}{4\,\mu F}} = \frac{4\,\mu F}{7} = 0.5714\,\mu F$$

The voltage across each capacitor is proportional to the total capacitance divided by the capacitance of the capacitor in question. So the voltage across C1 is:

$$E1 = \frac{0.5714\,\mu F}{1\,\mu F} \times 2000\ V = 1143\ V$$

Similarly, the voltages across C2 and C3 are:

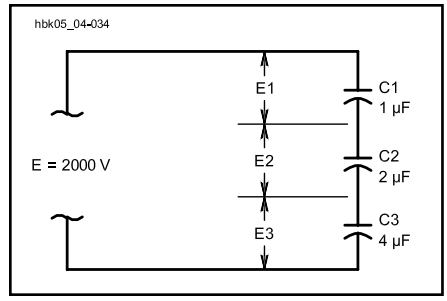


Fig 2.33 — An example of capacitors connected in series. The text shows how to find the voltage drops, E1 through E3.

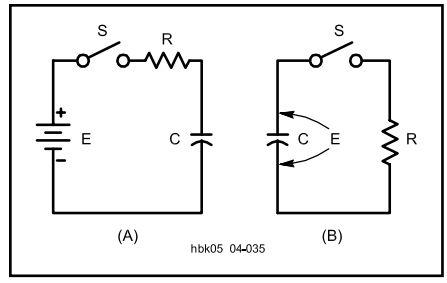


Fig 2.34 — An RC circuit. The series resistance delays the process of charging (A) and discharging (B) when the switch, S, is closed.

$$E2 = \frac{0.5714\,\mu F}{2\,\mu F} \times 2000\ V = 571\ V$$

and

$$E3 = \frac{0.5714\,\mu F}{4\,\mu F} \times 2000\ V = 286\ V$$

The sum of these three voltages equals 2000 V, the applied voltage.

Capacitors may be connected in series to enable the group to withstand a larger voltage than any individual capacitor is rated to withstand. The trade-off is a decrease in the total capacitance. As shown by the previous example, the applied voltage does not divide equally between the capacitors except when all the capacitances are precisely the same. Use care to ensure that the voltage rating of any capacitor in the group is not exceeded. If you use capacitors in series to withstand a higher voltage, you should also connect an "equalizing resistor" across each capacitor as described in the **Power Supplies** chapter.

2.7.4 RC Time Constant

Connecting a dc voltage source directly to the terminals of a capacitor charges the capacitor to the full source voltage almost instantaneously. Any resistance added to the circuit (as R in **Fig 2.34A)** limits the current, lengthening the time required for the voltage between the capacitor plates to build up to the source-voltage value. During this charging period, the current flowing from the source into the capacitor gradually decreases from its initial value. The increasing voltage stored in the capacitor's electric field offers increasing opposition to the steady source voltage.

While it is being charged, the voltage between the capacitor terminals is an exponential function of time, and is given by:

$$V(t) = E\left(1 - e^{-\frac{t}{RC}}\right) \quad (51)$$

where

V(t) = capacitor voltage at time t,
E = power source potential in volts,
t = time in seconds after initiation of charging current,
e = natural logarithmic base = 2.718,
R = circuit resistance in ohms, and
C = capacitance in farads.

(References that explain exponential equations, e, and other mathematical topics are found in the "Radio Math" section of this chapter.)

If t = RC, the above equation becomes:

$$V(RC) = E\,(1 - e^{-1}) \approx 0.632\ E \quad (52)$$

The product of R in ohms times C in farads is called the *time constant* (also called the *RC time constant*) of the circuit and is the time in seconds required to charge the capacitor to 63.2% of the applied voltage. (The lower-case Greek letter tau, τ, is often used to represent the time constant in electronics circuits.) After two time constants (t = 2τ) the capacitor charges another 63.2% of the difference between the capacitor voltage at one time constant and the applied voltage, for a total charge of 86.5%. After three time constants the capacitor reaches 95% of the applied voltage, and so on, as illustrated in the curve of **Fig 2.35A**. After five time constants, a capacitor is considered fully charged, having reached 99.24% of the applied voltage. Theoretically, the charging process is never really finished, but eventually the charging current drops to an immeasurably small value and the voltage is effectively constant.

If a charged capacitor is discharged through a resistor, as in Fig 2.34B, the same time constant applies to the decay of the capacitor voltage. A direct short circuit applied between the capacitor terminals would discharge the capacitor almost instantly. The resistor, R, limits the current, so a capacitor discharging through a resistance exhibits the same time-constant characteristics (calculated in the same way as above) as a charging capacitor. The voltage, as a function of time while the capacitor is being discharged, is given by:

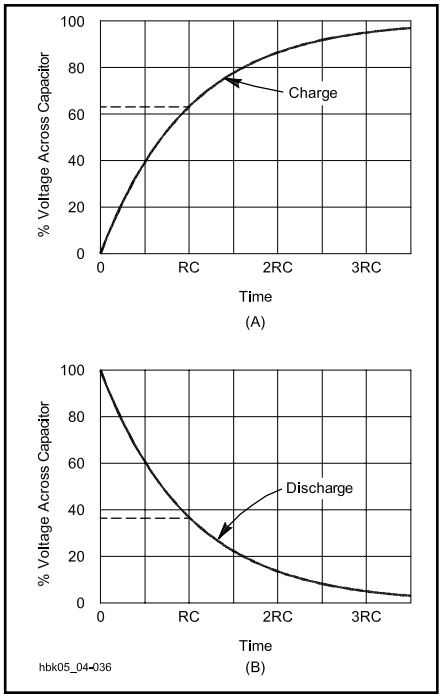


Fig 2.35 — At A, the curve shows how the voltage across a capacitor rises, with time, when charged through a resistor. The curve at B shows the way in which the voltage decreases across a capacitor when discharging through the same resistance. For practical purposes, a capacitor may be considered charged or discharged after five RC periods.

$$V(t) = E\left(e^{-\frac{t}{RC}}\right) \qquad (53)$$

where t = time in seconds after initiation of discharge and E is the fully-charged capacitor voltage prior to beginning discharge.

Again, by letting t = RC, the time constant of a discharging capacitor represents a decrease in the voltage across the capacitor of about 63.2%. After five time-constants, the capacitor is considered fully discharged, since the voltage has dropped to less than 1% of the full-charge voltage. **Fig 2.35B** is a graph of the discharging capacitor voltage in terms of time constants..

Time constant calculations have many uses in radio work. The following examples are all derived from practical-circuit applications.

Example 1: A 100-µF capacitor in a high-voltage power supply is shunted by a 100-kΩ resistor. What is the minimum time before the capacitor may be considered fully discharged? Since full discharge is approximately five RC periods,

$$t = 5 \times RC = 5 \times 100 \times 10^{3}\ \Omega \times 100 \times 10^{-6}\ F$$

$$= 50000 \times 10^{-3}$$

$$t = 50.0\ s$$

Note: Although waiting almost a minute for the capacitor to discharge seems safe in this high-voltage circuit, never rely solely on capacitor-discharging resistors (often called *bleeder resistors*). Be certain the power source is removed and the capacitors are totally discharged before touching any circuit components. (See the **Power Supplies** chapter for more information on bleeder resistors.)

Example 2: Smooth CW keying without clicks requires approximately 5 ms (0.005 s) of delay in both the rising and falling edges of the waveform, relative to full charging and discharging of a capacitor in the circuit. What typical values might a builder choose for an RC delay circuit in a keyed voltage line? Since full charge and discharge require 5 RC periods,

$$RC = \frac{t}{5} = \frac{0.005\ s}{5} = 0.001\ s$$

Any combination of resistor and capacitor whose values, when multiplied together, equal 0.001 would do the job. A typical capacitor might be 0.05 µF. In that case, the necessary resistor would be:

$$R = \frac{0.001\ s}{0.05 \times 10^{-6}\ F}$$

$$= 0.02 \times 10^{6}\ \Omega = 20000\ \Omega = 20\ k\Omega$$

In practice, a builder would use the calculated value as a starting point. The final value would be selected by monitoring the waveform on an oscilloscope.

Example 3: Many modern integrated circuit (IC) devices use RC circuits to control their timing. To match their internal circuitry, they may use a specified threshold voltage as the trigger level. For example, a certain IC uses a trigger level of 0.667 of the supply voltage. What value of capacitor and resistor would be required for a 4.5-second timing period?

First we will solve equation 51 for the time constant, RC. The threshold voltage is 0.667 times the supply voltage, so we use this value for V(t).

$$V(t) = E\left(1 - e^{-\frac{t}{RC}}\right)$$

$$0.667\ E = E\left(1 - e^{-\frac{t}{RC}}\right)$$

$$e^{-\frac{t}{RC}} = 1 - 0.667$$

$$\ln\left(e^{-\frac{t}{RC}}\right) = \ln(0.333)$$

$$-\frac{t}{RC} = -1.10$$

We want to find a capacitor and resistor combination that will produce a 4.5 s timing period, so we substitute that value for t.

$$RC = \frac{4.5\ s}{1.10} = 4.1\ s$$

If we select a value of 10 µF, we can solve for R.

$$R = \frac{4.1\ s}{10 \times 10^{-6}\ F} = 0.41 \times 10^{6}\ \Omega = 410\ k\Omega$$

A 1% tolerance resistor and capacitor will give good results. You could also use a variable resistor and an accurate method of measuring time to set the circuit to a 4.5 s period.

As the examples suggest, RC circuits have numerous applications in electronics. The number of applications is growing steadily, especially with the introduction of integrated circuits controlled by part or all of a capacitor charge or discharge cycle.

2.7.5 Alternating Current in Capacitance

Whereas 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 limit 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 **Fig 2.36** 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

RC Timesaver

When calculating time constants, it is handy to remember that if R is in units of MΩ and C is in units of µF, the result of R × C will be in seconds. Expressed as an equation: MΩ × µF = seconds

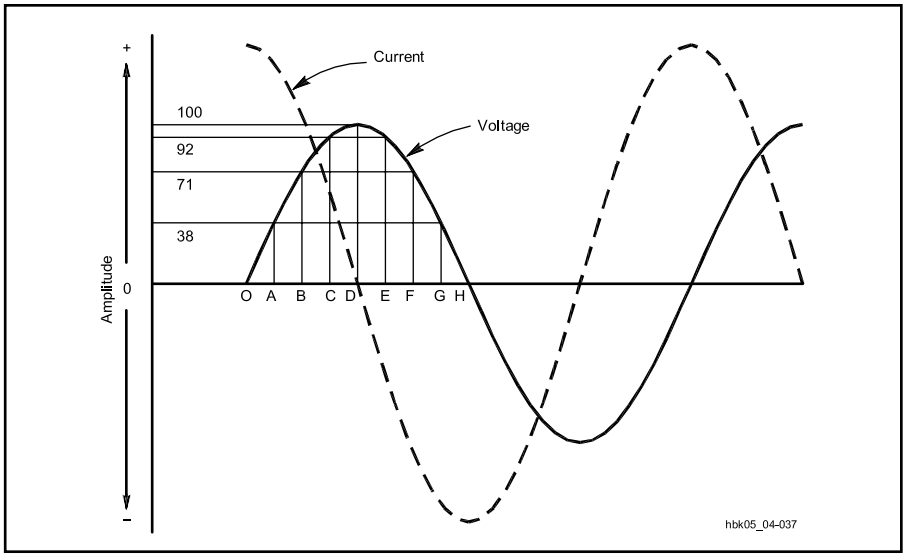


Fig 2.36 — Voltage and current phase relationships when an alternating current is applied to a capacitor.

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 Fig 2.36, 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 Fig 2.36, the current starts its cycle 90° before the voltage, 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°.

2.7.6 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 reactance of a capacitor at a given frequency is:

$$X_C = \frac{1}{2\pi f C} \qquad (54)$$

where

X_C = capacitive reactance in ohms,
f = frequency in hertz,
C = capacitance in farads
$\pi = 3.1416$

Note: In many references and texts, angular frequency $\omega = 2\pi f$ is used and equation 54 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?

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

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

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

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

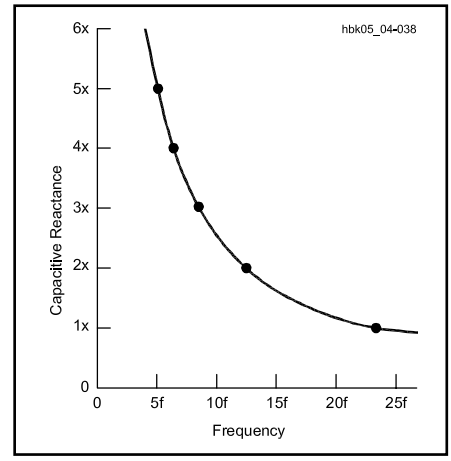


Fig 2.37 — A graph showing the general relationship of reactance to frequency for a fixed value of capacitance.

Capacitive Reactance Timesaver

The fundamental units for frequency and capacitance (hertz and farads) are too cumbersome for practical use in radio circuits. If the capacitance is specified in microfarads (μF) and the frequency is in megahertz (MHz), however, the reactance calculated from equation 54 is in units of ohms (Ω).

Table 2.7
Typical Temperature Coefficients and Leakage Resistances for Various Capacitor Constructions

Type	*TC @ 20°C (PPM/°C)*	*DC Leakage Resistance (Ω)*
Ceramic Disc	±300(NP0)	> 10 M
	+150/–1500(GP)	> 10 M
Mica	–20 to +100	> 100,000 M
Polyester	±500	> 10 M
Tantalum Electrolytic	±1500	> 10 MΩ
Small Al Electrolytic(≈ 100 μF)	–20,000	500 k - 1 M
Large Al Electrolytic(≈ 10 mF)	–100,000	10 k
Vacuum (glass)	+100	≈ ∞
Vacuum (ceramic)	+50	≈ ∞

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. **Fig 2.37** 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 losses, susceptance is simply the reciprocal of reactance. Hence,

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

where

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

2.7.7 Characteristics of Capacitors

The ideal capacitor does not conduct any current at dc, dissipates none of the energy stored in it, has the same value at all temperatures, and operates with any amount of voltage applied across it or ac current flowing through it. Practical capacitors deviate considerably from that ideal and have imperfections that must be considered when selecting capacitors and designing circuits that use capacitors. (The characteristics of capacitors at high frequencies is discussed in the **RF Techniques** chapter.)

LEAKAGE RESISTANCE

If we use anything other than a vacuum for the insulating layer, even air, two imperfections are created. Because there are atoms between the plates, some electrons will be available to create a current between the plates when a dc voltage is applied. The magnitude of this *leakage current* will depend on the insulator quality, and the current is usually very small. Leakage current can be modeled by a resistance R_L in parallel with the capacitance (in an ideal capacitor, R_L is infinite). **Table 2.7** shows typical dc leakage resistances for different dielectric materials. Leakage also generally increases with increasing temperature.

CAPACITOR LOSSES

When an ac current flows through the capacitor (even at low frequencies), capacitors dissipate some of the energy stored in the dielectric due to the electromagnetic properties of dielectric materials. This loss can be thought of as a resistance in series with the capacitor and it is often specified in the manufacturer's data for the capacitor as *effective (or equivalent) series resistance (ESR).*

Loss can also be specified as the capacitor's *loss angle,* θ. (Some literature uses δ for loss angle.) Loss angle is the angle between X_C (the reactance of the capacitor without any loss) and the impedance of the capacitor (impedance is discussed later in this chapter) made up of the combination of ESR and X_C. Increasing loss increases loss angle. The loss angle is usually quite small, and is zero for an ideal capacitor.

Dissipation Factor (DF) or *loss tangent* = tan θ = ESR / X_C and is the ratio of loss resistance to reactance. The loss angle of a given capacitor is relatively constant over frequency, meaning that ESR = (tan θ) / 2πfC goes down as frequency goes up. For this reason, ESR must be specified at a given frequency.

TOLERANCE AND TEMPERATURE COEFFICIENT

As with resistors, capacitor values vary in production, and most capacitors have a tolerance rating either printed on them or listed on a data sheet. Typical capacitor tolerances and the labeling of tolerance are described in the chapter on **Component Data and References**.

Because the materials that make up a capacitor exhibit mechanical changes with temperature, capacitance also varies with temperature. This change in capacitance with temperature is the capacitor's *temperature coefficient* or *tempco (TC)*. The lower a capacitor's TC, the less its value changes with temperature.

TC is important to consider when constructing a circuit that will carry high power levels, operate in an environment far from room temperature, or must operate consistently at different temperatures. Typical temperature coefficients for several capacitor types are given in Table 2.7. (Capacitor temperature coefficient behaviors are listed in the **Component Data and References** chapter.)

VOLTAGE RATINGS AND BREAKDOWN

When voltage is applied to the plates of a capacitor, force is exerted on the atoms and molecules of the dielectric by the electrostatic field between the plates. If the voltage is high enough, the atoms of the dielectric will ionize (one or more of the electrons will be pulled away from the atom), causing a large dc current to flow discharging the capacitor. This is *dielectric breakdown,* and it is generally destructive to the capacitor because it creates punctures or defects in solid dielectrics that provide permanent low-resistance current paths between the plates. (*Self-healing* dielectrics have the ability to seal off this type of damage.) With most gas dielectrics such as air, once the voltage is removed, the arc ceases and the capacitor is ready for use again.

The *breakdown voltage* of a dielectric depends on the chemical composition and thickness of the dielectric. Breakdown voltage is not directly proportional to the thickness;

doubling the thickness does not quite double the breakdown voltage. A thick dielectric must be used to withstand high voltages. Since capacitance is inversely proportional to dielectric thickness (plate spacing) for a given plate area, a high-voltage capacitor must have more plate area than a low-voltage one of the same capacitance. High-voltage, high-capacitance capacitors are therefore physically large.

Dielectric strength is specified in terms of a *dielectric withstanding voltage* (DWV), given in volts per mil (0.001 inch) at a specified temperature. Taking into account the design temperature range of a capacitor and a safety margin, manufacturers specify *dc working voltage* (dcwv) to express the maximum safe limits of dc voltage across a capacitor to prevent dielectric breakdown.

For use with ac voltages, the peak value of ac voltage should not exceed the dc working voltage, unless otherwise specified in component ratings. In other words, the RMS value of sine-wave ac waveforms should be 0.707 times the dcwv value, or lower. With many types of capacitors, further derating is required as the operating frequency increases. An additional safety margin is good practice.

Dielectric breakdown in a gas or air dielectric capacitor occurs as a spark or arc between the plates. Spark voltages are generally given with the units *kilovolts per centimeter*. For air, the spark voltage or V_s may range from more than 120 kV/cm for gaps as narrow as 0.006 cm down to 28 kV/cm for gaps as wide as 10 cm. In addition, a large number of variables enter into the actual breakdown voltage in a real situation. Among the variables are the plate shape, the gap distance, the air pressure or density, the voltage, impurities in the air (or any other dielectric material) and the nature of the external circuit (with air, for instance, the humidity affects conduction on the surface of the capacitor plate).

Dielectric breakdown occurs at a lower voltage between pointed or sharp-edged surfaces than between rounded and polished surfaces. Consequently, the breakdown voltage between metal plates of any given spacing in air can be increased by buffing the edges of the plates. If the plates are damaged so they are no longer smooth and polished, they may have to be polished or the capacitor replaced.

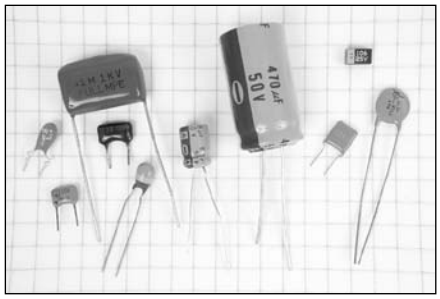
2.7.8 Capacitor Types and Uses

Quite a variety of capacitors are used in radio circuits, differing considerably in physical size, construction and capacitance. Some of the different types are shown in **Fig 2.38** and many other types and packages are available. (See the **Component Data and References** chapter for illustrations of capacitor types and labeling conventions.)

The dielectric determines many properties of the capacitor, although the construction of the plates strongly affects the capacitor's ac performance and some dc parameters. Various materials are used for different reasons such as working voltage and current, availability, cost, and desired capacitance range.

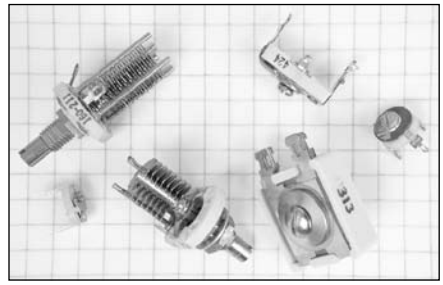
Fixed capacitors having a single, nonadjustable value of capacitance can also be made with metal plates and with air as the dielectric, but are usually constructed from strips of metal foil with a thin, solid or liquid dielectric sandwiched between, so that a relatively large capacitance can be obtained in a small package. Solid dielectrics commonly used in fixed capacitors are plastic films, mica, paper and special ceramics. Two typical types of fixed capacitor construction are shown in **Fig 2.39**.

For capacitors with wire leads, there are two

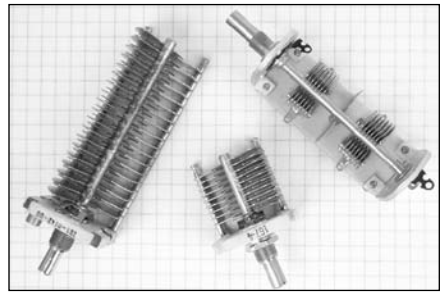


(A)

(B)



(C)



(D)

(E)

Fig 2.38 — Fixed-value capacitors are shown in parts A and B. Aluminum electrolytic capacitors are pictured near the center of photo A. The small tear-drop units to the left of center are tantalum electrolytic capacitors. The rectangular units are silvered-mica, polystyrene film and monolithic ceramic. At the right edge is a disc-ceramic capacitor and near the top right corner is a surface-mount capacitor. B shows a large "computer-grade" electrolytic. These have very low equivalent series resistance (ESR) and are often used as filter capacitors in switch-mode power supplies, and in series-strings for high-voltage supplies of RF power amplifiers. Parts C and D show a variety of variable capacitors, including air variable capacitors and mica compression units. Part E shows a vacuum variable capacitor such as is sometimes used in high-power amplifier circuits. The ¼-inch-ruled graph paper backgrounds provide size comparisons.

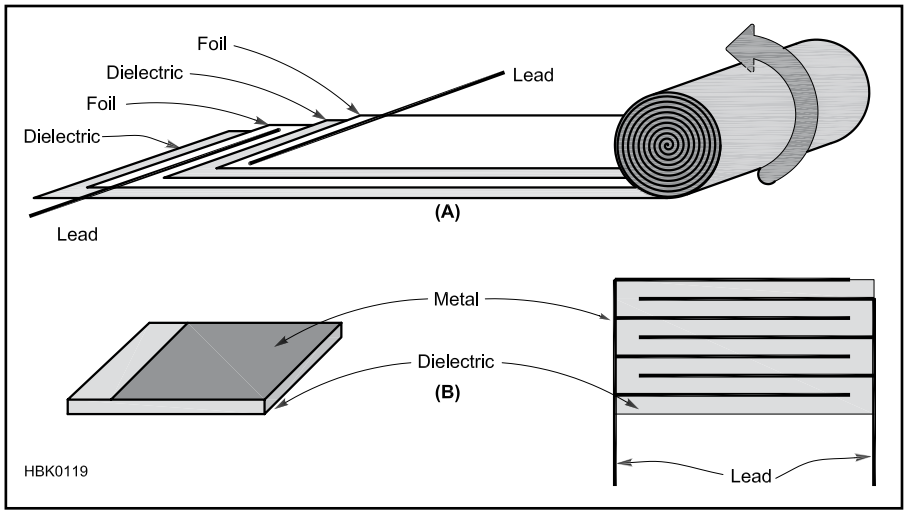


Fig 2.39 — Two common types of capacitor construction. A shows the roll method for film capacitors with axial leads. B shows the alternating layer method for ceramic capacitors. Axial leads are shown in A and radial leads in B.

basic types of lead orientation; *axial* (shown in Fig 2.39A) in which the leads are aligned with the long axis of the capacitor body and *radial* (shown in Fig 2.39B) in which the leads are at right angles to the capacitor's length or width.

Vacuum. Both fixed and variable vacuum capacitors are available. They are rated by their maximum working voltages (3 to 60 kV) and currents. Losses are specified as negligible for most applications. The high working voltage and low losses make vacuum capacitors widely used in transmitting applications. Vacuum capacitors are also unaffected by humidity, moisture, contamination, or dust, unlike air-dielectric capacitors discussed next. This allows them to be used in environments for which air-dielectric capacitors would be unsuitable.

Air. Since $K \approx 1$ for air, air-dielectric capacitors are large when compared to those of the same value using other dielectrics. Their capacitance is very stable over a wide temperature range, leakage losses are low, and therefore a high Q can be obtained. They also can withstand high voltages. Values range from a few tens to hundreds of pF.

For these reasons (and ease of construction) most variable capacitors in tuning circuits are *air-variable* capacitors made with one set of plates movable with respect to the other set to vary the area of overlap and thus the capacitance. A *transmitting-variable* capacitor has heavy plates far enough apart to withstand the high voltages and currents encountered in a transmitter. (Air variable capacitors with more closely-spaced plates are often referred to as *receiving-variables.*)

Plastic film. Capacitors with plastic film (such as polystyrene, polyethylene or Mylar) dielectrics are widely used in bypassing and coupling applications up to several megahertz. They have high leakage resistances (even at high temperatures) and low TCs. Values range from tens of pF to 1 μF. Plastic-film variable capacitors are also available.

Most film capacitors are not polarized; however, the body of the capacitor is usually marked with a color band at one end. The band indicates the terminal that is connected to the outermost plate of the capacitor. This terminal should be connected to the side of the circuit at the lower potential as a safety precaution.

Mica. The capacitance of mica capacitors is very stable with respect to time, temperature and electrical stress. Leakage and losses are very low and they are often used in transmitting equipment. Values range from 1 pF to 0.1 μF. High working voltages are possible, but they must be derated severely as operating frequency increases.

Silver-mica capacitors are made by depositing a thin layer of silver on the mica dielectric. This makes the value even more stable, but it presents the possibility of silver migration through the dielectric. The migration problem worsens with increased dc voltage, temperature and humidity. Avoid using silver-mica capacitors under such conditions. Silver-mica capacitors are often used in RF circuits requiring stable capacitor values, such as oscillators and filters.

Ceramic. Ceramic capacitors are available with values from 1 pF to 1 μF and with voltage ratings up to 1 kV. *Monolithic ceramic* capacitors are constructed from a stack of thin ceramic layers with a metal coating on one side. The layer is then compressed with alternating metal coatings connected together to form the capacitor's plates. The high dielectric constant makes these capacitors physically small for their capacitance, but their value is not as stable and their dielectric properties vary with temperature, applied voltage and operating frequency. They also exhibit piezoelectric behavior. Use them only in coupling and bypass roles. *Disc ceramic* capacitors are made similarly to monolithic ceramic capacitors but with a lower dielectric constant so they are larger and tend to have higher voltage ratings. Ceramic capacitors are useful into the VHF and UHF ranges.

Transmitting ceramic capacitors are made, like transmitting air-variables, with heavy plates and high-voltage ratings. They are relatively large, but very stable and have nearly as low losses as mica capacitors at HF.

Electrolytic. Electrolytic capacitors are constructed with plates made of aluminum-foil strips and a semi-liquid conducting chemical compound between them. They are sometimes called *aluminum electrolytics*. The actual dielectric is a very thin film of insulating material that forms on one set of plates through electrochemical action when a dc voltage is applied to the capacitor. The capacitance of an electrolytic capacitor is very large compared to capacitors having other dielectrics, because the dielectric film is so thin — much thinner than is practical with a solid dielectric. Electrolytic capacitors are available with values from approximately 1 μF to 1 F and with voltage ratings up to hundreds of volts.

Electrolytic capacitors are popular because they provide high capacitance values in small packages at a reasonable cost. Leakage resistance is comparatively low and they are polarized — there is a definite positive and negative plate, due to the chemical reaction that creates the dielectric. Internal inductance restricts aluminum-foil electrolytics to low-frequency applications such as power-supply filtering and bypassing in audio circuits. To maintain the dielectric film, electrolytic capacitors should not be used if the applied dc potential will be well below the capacitor working voltage.

A cautionary note is warranted regarding electrolytic capacitors found in older equipment, both vacuum tube and solid-state. The chemical paste in electrolytics dries out when the component is heated and with age, causing high losses and reduced capacitance. The dielectric film also disappears when the capacitor is not used for long periods. It is possible to "reform" the dielectric by applying a low voltage to an old or unused capacitor and gradually increasing the voltage. However, old electrolytics rarely perform as well as new units. To avoid expensive failures and circuit damage, it is recommended that electrolytic capacitors in old equipment be replaced if they have not been in regular use for more than ten years.

Tantalum. Related to the electrolytic capacitor, *tantalum* capacitors substitute a *slug* of extremely porous tantalum (a rare-earth metallic element) for the aluminum-foil strips as one plate. As in the electrolytic

capacitor, the dielectric is an oxide film that forms on the surface of the tantalum. The slug is immersed in a liquid compound contained in a metal can that serves as the other plate. Tantalum capacitors are commonly used with values from 0.1 to several hundred μF and voltage ratings of less than 100 V. Tantalum capacitors are smaller, lighter and more stable, with less leakage and inductance than their aluminum-foil electrolytic counterparts but their cost is higher.

Paper. Paper capacitors are generally not used in new designs and are largely encountered in older equipment; capacitances from 500 pF to 50 μF are available. High working voltages are possible, but paper-dielectric capacitors have low leakage resistances and tolerances are no better than 10 to 20%.

Trimming capacitors. Small-value variable capacitors are often referred to as *trimmers* because they are used for fine-tuning or frequency adjustments, called *trimming*. Trimmers have dielectrics of Teflon, air, or ceramic and generally have values of less than 100 pF. *Compression trimmers* have higher values of up to 1000 pF and are constructed with mica dielectrics.

Oil-filled. Oil-filled capacitors use special high-strength dielectric oils to achieve voltage ratings of several kV. Values of up to 100 μF are commonly used in high-voltage applications such as high-voltage power supplies and energy storage. (See the chapter on **Power Supplies** for additional information about the use of oil-filled and electrolytic capacitors.)

2.8 Inductance and Inductors

A second way to store electrical energy is in a *magnetic field*. This phenomenon is called *inductance*, and the devices that exhibit inductance are called *inductors*. Inductance is derived from some basic underlying magnetic properties.

2.8.1 Magnetic Fields and Magnetic Energy Storage

MAGNETIC FLUX

As an electric field surrounds an electric charge, magnetic fields surround *magnets*. You are probably familiar with metallic bar, disc, or horseshoe-shaped magnets. **Fig 2.40** shows a bar magnet, but particles of matter as small as an atom can also be magnets.

Fig 2.40 also shows the magnet surrounded by lines of force called *magnetic flux*, representing a *magnetic field*. (More accurately, a *magnetostatic field*, since the field is not changing.) Similar to those of an electric field, magnetic lines of force (or *flux lines*) show the direction in which a magnet would feel a force in the field.

There is no "magnetic charge" comparable to positive and negative electric charges. All magnets and magnetic fields have a polarity, represented as *poles*, and every magnet — from atoms to bar magnets — possesses both a *north* and *south pole*. The size of the source of the magnetism makes no difference. The north pole of a magnet is defined as the one attracted to the Earth's north magnetic pole. (Confusingly, this definition means the Earth's North Magnetic Pole is magnetically a south pole!) Like conventional current, the direction of magnetic lines of force was assigned arbitrarily by early scientists as pointing *from* the magnet's north pole *to* the south pole.

An electric field is *open* — that is, its lines of force have one end on an electric charge and can extend to infinity. A magnetic field is *closed* because all magnetic lines of force form a loop passing through a magnet's north and south poles.

Magnetic fields exist around two types of materials; *permanent magnets* and *electromagnets*. Permanent magnets consist of *ferromagnetic* and *ferrimagnetic* materials whose atoms are or can be aligned so as to produce a magnetic field. Ferro- or ferrimagnetic materials are strongly attracted to magnets. They can be *magnetized*, meaning to be made magnetic, by the application of a magnetic field. Lodestone, magnetite, and ferrites are examples of ferrimagnetic materials. Iron, nickel, cobalt, Alnico alloys and other materials are ferromagnetic. Magnetic materials with high *retentivity* form permanent magnets because they retain their magnetic properties for long periods. Other materials, such as soft iron, yield temporary magnets that lose their magnetic properties rapidly.

Paramagnetic substances are very weakly attracted to a magnet and include materials such as platinum, aluminum, and oxygen. *Diamagnetic* substances, such as copper, carbon, and water, are weakly repelled by a magnet.

The second type of magnet is an electrical conductor with a current flowing through it. As shown in **Fig 2.41**, moving electrons are surrounded by a closed magnetic field, illustrated as the circular lines of force around the wire lying in planes perpendicular to the current's motion. The magnetic needle of a compass placed near a wire carrying direct current will be deflected as its poles respond to the forces created by the magnetic field around the wire.

If the wire is coiled into a *solenoid* as shown in **Fig 2.42**, the magnetic field greatly intensifies. This occurs as the magnetic fields from each successive turn in the coil add together

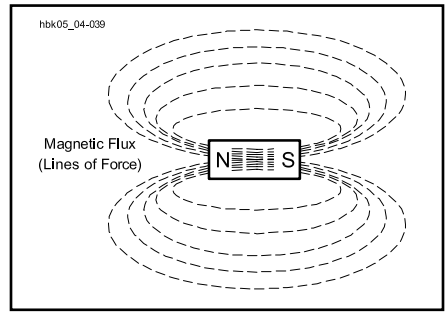


Fig 2.40 — The magnetic field and poles of a permanent magnet. The magnetic field direction is from the north to the south pole.

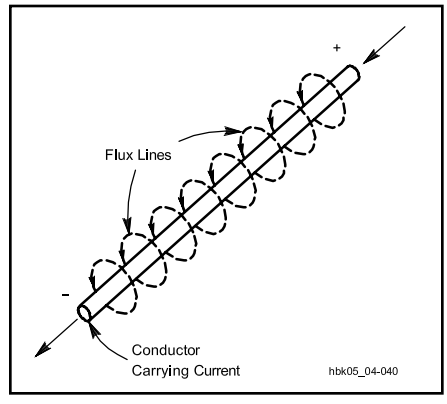


Fig 2.41 — The magnetic field around a conductor carrying an electrical current. If the thumb of your right hand points in the direction of the conventional current (plus to minus), your fingers curl in the direction of the magnetic field around the wire.

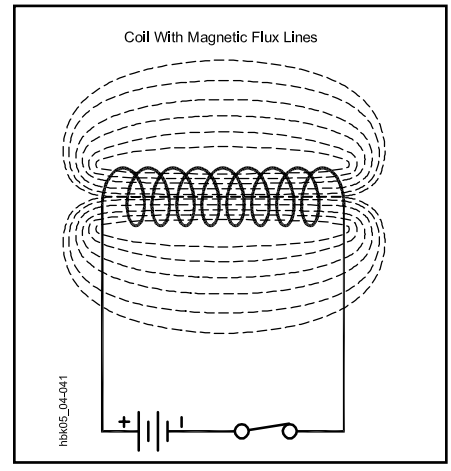


Fig 2.42 — Cross section of an inductor showing its flux lines and overall magnetic field.

because the current in each turn is flowing in the same direction.

Note that the resulting *electromagnet* has magnetic properties identical in principle to those of a permanent magnet, including poles and lines of force or flux. The strength of the magnetic field depends on several factors: the number and shape of turns of the coil, the magnetic properties of the materials surrounding the coil (both inside and out), the length of the coil and the amplitude of the current.

Magnetic fields and electric current have a special two-way relationship: voltage causing an electrical current (moving charges) in a conductor will produce a magnetic field and a moving magnetic field will create an electrical field (voltage) that produces current in a conductor. This is the principle behind motors and generators, converting mechanical energy into electrical energy and vice-versa.

MAGNETIC FLUX

Magnetic flux is measured in the SI unit (International System of Units) of the weber, which is a volt-second (Wb = Vs). In the *centimeter-gram-second* (*cgs*) metric system units, magnetic flux is measured in maxwells (1 Mx = 10^{-8} Wb). The volt-second is used because of the relationship described in the previous paragraph: 1 volt of electromotive force will be created in a loop of wire in which magnetic flux through the loop changes at the rate of 1 weber per second. The relationship between current and magnetic fields is one of motion and change.

Magnetic field intensity, known as *flux density*, decreases with the square of the distance from the source, either a magnet or current. Flux density (*B*) is represented in gauss (G), where one gauss is equivalent to one line of force (1 Mx) per square centimeter of area measured perpendicularly to the direction of the field (G = Mx / cm^2). The gauss is a *cgs* unit. In SI units, flux density is represented by the tesla (T), which is one weber per square meter (T = Wb/m^2 and 1T = 10,000 G).

Magnetomotive Force and Field Strength

The magnetizing or *magnetomotive force* ($\Im$) that produces a flux or total magnetic field is measured in gilberts (Gb). Magnetomotive force is analogous to electromotive force in that it produces the magnetic field. The SI unit of magnetomotive force is the ampere-turn, abbreviated A, just like the ampere. (1 Gb = 0.79577 A)

$$\Im = \frac{10\,N\,I}{4\pi} \qquad (55)$$

where

$\Im$ = magnetomotive strength in gilberts,
N = number of turns in the coil creating the field,
I = dc current in amperes in the coil, and
π = 3.1416.

MAGNETIC FIELD STRENGTH

The magnetic field strength (*H*) measured in oersteds (Oe) produced by any particular magnetomotive force (measured in gilberts) is given by:

$$H = \frac{\Im}{\ell} = \frac{10\,N\,I}{4\,\pi\,\ell} \qquad (56)$$

where

H = magnetic field strength in oersteds, and
ℓ = mean magnetic path length in centimeters.

The *mean magnetic path length* is the average length of the lines of magnetic flux. If the inductor is wound on a closed core as shown in the next section, *l* is approximately the average of the inner and outer circumferences of the core. The SI unit of magnetic field strength is the ampere-turn per meter. (1 Oe = 79.58 A/m) The *cgs* units gilbert and oersted are the units most commonly used for radio circuits.

The Right-hand Rule

How do you remember which way the magnetic field around a current is pointing? Luckily, there is a simple method, called the *right-hand rule*. Make your right hand into a fist, then extend your thumb, as in **Fig 2.A2**. If your thumb is pointing in the direction of conventional current flow, then your fingers curl in the same direction as the magnetic field. (If you are dealing with electronic current, use your left hand, instead!)

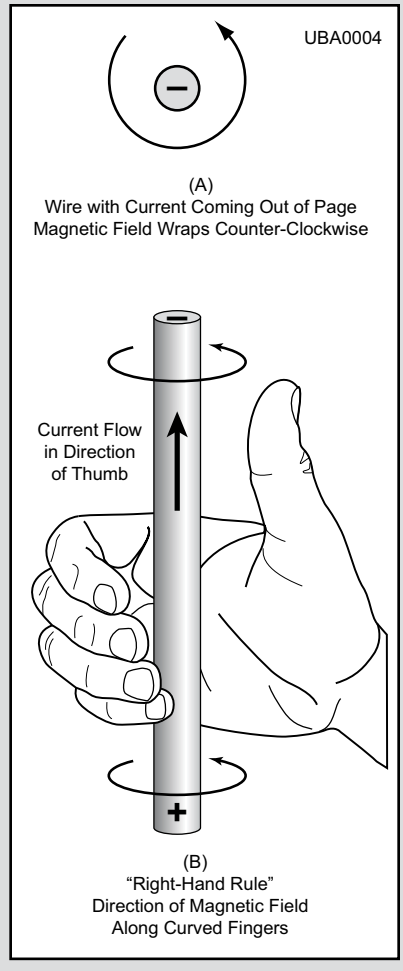


Fig 2.A2 — Use the right-hand rule to determine magnetic field direction from the direction of current flow.

2.8.2 Magnetic Core Properties

PERMEABILITY

The nature of the material within the coil of an electromagnet, where the lines of force are most concentrated, has the greatest effect upon the magnetic field established by the coil. All core materials are compared relatively to air. The ratio of flux density produced by a given material compared to the flux density produced by an air core is the *permeability* (μ) of the material. Air and non-magnetic materials have a permeability of one.

Suppose the coil in **Fig 2.43** is wound on an iron core having a cross-sectional area of 2 square inches. When a certain current is sent through the coil, it is found that there are 80,000 lines of force in the core. Since the area is 2 square inches, the magnetic flux density is 40,000 lines per square inch. Now suppose that the iron core is removed and the same current is maintained in the coil. Also suppose the flux density without the iron core is found to be 50 lines per square inch. The ratio of these flux densities, iron core to air, is 40,000 / 50 or 800. This ratio is the core's permeability.

Permeabilities as high as 10^6 have been attained. The three most common types of materials used in magnetic cores are these:

A. stacks of thin steel laminations (for

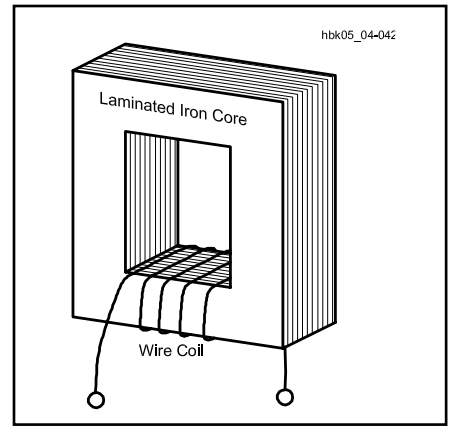


Fig 2.43 — A coil of wire wound around a laminated iron core.

Table 2.8
Properties of Some High-Permeability Materials

Material	*Approximate Percent Composition*					*Maximum Permeability*
	Fe	Ni	Co	Mo	Other	
Iron	99.91	—	—	—	—	5000
Purified Iron	99.95	—	—	—	—	180000
4% silicon-iron	96	—	—	—	4 Si	7000
45 Permalloy	54.7	45	—	—	0.3 Mn	25000
Hipernik	50	50	—	—	—	70000
78 Permalloy	21.2	78.5	—	—	0.3 Mn	100000
4-79 Permalloy	16.7	79	—	—	0.3 Mn	100000
Supermalloy	15.7	79	—	5	0.3 Mn	800000
Permendur	49.7	—	50	—	0.3 Mn	5000
2V Permendur	49	—	49	—	2 V	4500
Hiperco	64	—	34	—	2 Cr	10000
2-81 Permalloy*	17	81	—	2	—	130
Carbonyl iron*	99.9	—	—	—	—	132
Ferroxcube III**	($MnFe_2O_4$ + $ZnFe_2O_4$)		1500			

Note: all materials in sheet form except * (insulated powder) and ** (sintered powder).
(Reference: L. Ridenour, ed., *Modern Physics for the Engineer*, p 119.)

power and audio applications, see the discussion on eddy currents below);

B. various ferrite compounds (for cores shaped as rods, toroids, beads and numerous other forms); and

C. powdered iron (shaped as slugs, toroids and other forms for RF inductors).

The permeability of silicon-steel power-transformer cores approaches 5000 in high-quality units. Powdered-iron cores used in RF tuned circuits range in permeability from 3 to about 35, while ferrites of nickel-zinc and manganese-zinc range from 20 to 15000. Not all materials have permeabilities higher than air. Brass has a permeability of less than one. A brass core inserted into a coil will decrease the magnetic field compared to an air core.

Table 2.8 lists some common magnetic materials, their composition and their permeabilities. Core materials are often frequency sensitive, exhibiting excessive losses outside the frequency band of intended use. (Ferrite materials are discussed separately in a later section of the chapter on **RF Techniques**.)

As a measure of the ease with which a magnetic field may be established in a material as compared with air, permeability (μ) corresponds roughly to electrical conductivity. Higher permeability means that it is easier to establish a magnetic field in the material. Permeability is given as:

$$\mu = \frac{B}{H} \qquad (57)$$

where

B is the flux density in gauss, and

H is the magnetic field strength in oersteds.

RELUCTANCE

That a force (the magnetomotive force) is required to produce a given magnetic field strength implies that there is some opposition to be overcome. This opposition to the creation of a magnetic field is called *reluctance*. Reluctance ($\Re$) is the reciprocal of permeability and corresponds roughly to resistance in an electrical circuit. Carrying the electrical resistance analogy a bit further, the magnetic equivalent of Ohm's Law relates reluctance, magnetomotive force, and flux density: $\Re = \Im / B$.

SATURATION

Unlike electrical conductivity, which is independent of other electrical parameters, the permeability of a magnetic material varies with the flux density. At low flux densities (or with an air core), increasing the current through the coil will cause a proportionate increase in flux. This occurs because the current passing through the coil forces the atoms of the iron (or other material) to line up, just like many small compass needles. The magnetic field that results from the atomic alignment is *much* larger than that produced by the current with no core. As more and more atoms align, the magnetic flux density also increases.

At very high flux densities, increasing the current beyond a certain point may cause no appreciable change in the flux because all of the atoms are aligned. At this point, the core is said to be *saturated*. Saturation causes a rapid decrease in permeability, because it decreases the ratio of flux lines to those obtainable with the same current using an air core. **Fig 2.44** displays a typical permeability curve, showing the region of saturation. The saturation point varies with the makeup of different magnetic materials. Air and other nonmagnetic materials do not saturate.

HYSTERESIS

Retentivity in magnetic core materials is caused by atoms retaining their alignment from an applied magnetizing force. Retentivity is desirable if the goal is to create a permanent magnet. In an electronic circuit, however, the changes caused by retentivity cause the properties of the core material to depend on the history of how the magnetizing force was applied.

Fig 2.45 illustrates the change of flux density (B) with a changing magnetizing force (H). From starting point A, with no flux in the core, the flux reaches point B at the maximum magnetizing force. As the force decreases, so too does the flux, but it does not reach zero simultaneously with the force at point D. As

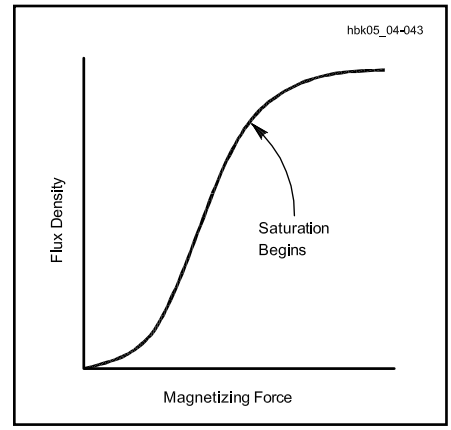


Fig 2.44 — A typical permeability curve for a magnetic core, showing the point where saturation begins.

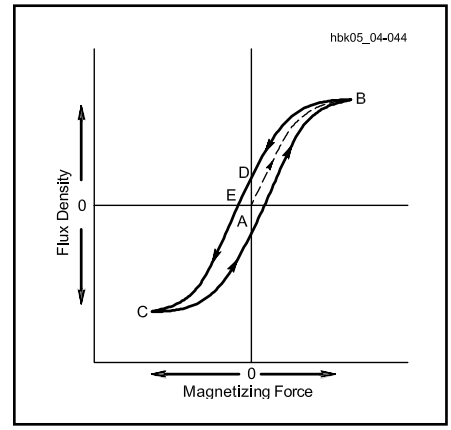


Fig 2.45 — A typical hysteresis curve for a magnetic core, showing the additional energy needed to overcome residual flux.

the force continues in the opposite direction, it brings the flux density to point C. As the force decreases to zero, the flux once more lags behind. This occurs because some of the atoms in core retain their alignment, even after the external magnetizing force is removed. This creates *residual flux* that is present even with no applied magnetizing force. This is the property of *hysteresis*.

In effect, a *coercive force* is necessary to reverse or overcome the residual magnetism retained by the core material. If a circuit carries a large ac current (that is, equal to or larger than saturation), the path shown in Fig 2.45 will be retraced with every cycle and the reversing force each time. The result is a power loss to the magnetic circuit, which appears as heat in the core material. Air cores are immune to hysteresis effects and losses.

2.8.3 Inductance and Direct Current

In an electrical circuit, any element whose operation is based on the transfer of energy into and out of magnetic fields is called an *inductor* for reasons to be explained shortly. **Fig 2.46** shows schematic-diagram symbols and photographs of a few representative inductors. The photograph shows an air-core inductor, a slug-tuned (variable-core) inductor with a nonmagnetic core and an inductor with a magnetic (iron) core. Inductors are often called *coils* because of their construction.

As explained above, when current flows through any conductor — even a straight wire — a magnetic field is created. The transfer of energy to the magnetic field represents work performed by the source of the voltage. Power is required for doing work, and since power is equal to current multiplied by voltage, there must be a voltage drop across the inductor while energy is being stored in the field. This voltage drop, exclusive of any voltage drop caused by resistance in the conductor, is the result of an opposing voltage created in the conductor while the magnetic field is building up to its final value. Once the field becomes constant, the *induced voltage* or *back-voltage* disappears, because no further energy is being stored. Back voltage is analogous to the opposition to current flow in a capacitor from the increasing capacitor voltage.

The induced voltage opposes the voltage of the source, preventing the current from rising rapidly when voltage is applied. **Fig 2.47A** illustrates the situation of energizing an inductor or magnetic circuit, showing the relative amplitudes of induced voltage and the delayed rise in current to its full value.

The amplitude of the induced voltage is proportional to the rate at which the current changes (and consequently, the rate at which the magnetic field changes) and to a constant associated with the inductor itself, *inductance* (*L*). (*Self-inductance* is sometimes used to distinguish between mutual inductance as described below.) The basic unit of inductance is the *henry* (abbreviated H).

$$V = L\frac{\Delta I}{\Delta t} \qquad (58)$$

where

V is the induced voltage in volts,
L is the inductance in henries, and
$\Delta I/\Delta t$ is the rate of change of the current in amperes per second.

An inductance of 1 H generates an induced voltage of one volt when the inducing current is varying at a rate of one ampere per second.

The energy stored in the magnetic field of an inductor is given by the formula:

$$W = \frac{I^2 L}{2} \qquad (59)$$

where

W = energy in joules,
I = current in amperes, and
L = inductance in henrys.

This formula corresponds to the energy-storage formula for capacitors: energy storage is a function of current squared. Inductance is proportional to the amount of energy stored in an inductor's magnetic field for a given amount of current. The magnetic field strength, H, is proportional to the number of turns in the inductor's winding, N, (see equation 56) and for a given amount of current, to the value of μ for the core. Thus, inductance is directly proportional to both N and μ.

The polarity of the induced voltage is always such as to oppose any change in the circuit current. (This is why the term "back" is used, as in back-voltage or *back-EMF* for this reason.) This means that when the current in the circuit is increasing, work is being done against the induced voltage by storing energy in the magnetic field. Likewise, if the current in the circuit tends to decrease, the stored energy of the field returns to the circuit, and adds to the energy being supplied by the voltage source. The net effect of storing and releasing energy is that inductors oppose changes in current just as capacitors oppose changes in voltage. This phenomenon tends to keep the current flowing even though the applied voltage may be decreasing or be removed entirely. Fig 2.47B illustrates the decreasing but continuing flow of current caused by the induced voltage after the source voltage is removed from the circuit.

Inductance depends on the physical configuration of the inductor. All conductors, even straight wires, have inductance. Coiling a conductor increases its inductance. In

Rate of Change

The symbol Δ represents change in the following variable, so that ΔI represents "change in current" and Δt "change in time". A rate of change per unit of time is often expressed in this manner. When the amount of time over which the change is measured becomes very small, the letter *d* replaces Δ in both the numerator and denominator to indicate infinitesimal changes. This notation is used in the derivation and presentation of the functions that describe the behavior of electric circuits.

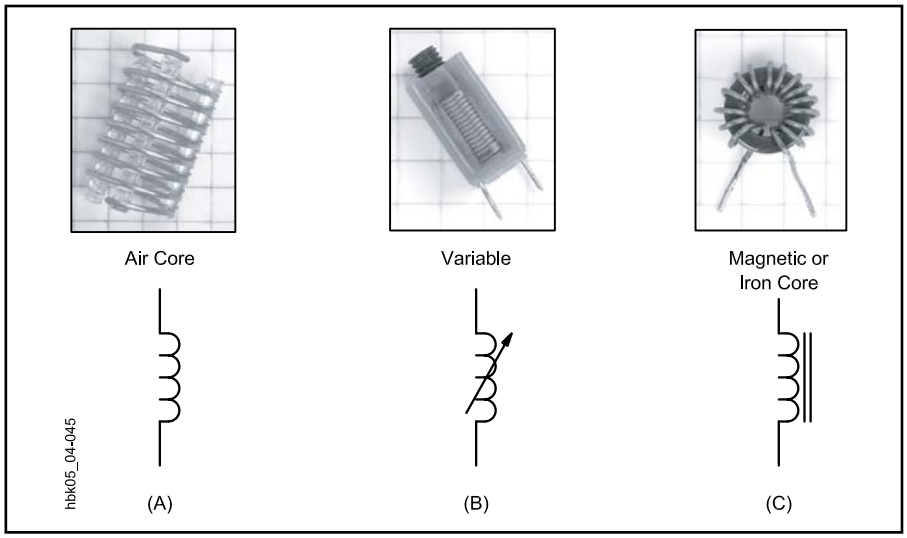


Fig 2.46 — Photos and schematic symbols for representative inductors. A, an air-core inductor; B, a variable inductor with a nonmagnetic slug and C, an inductor with a toroidal magnetic core. The ¼-inch-ruled graph paper background provides a size comparison.

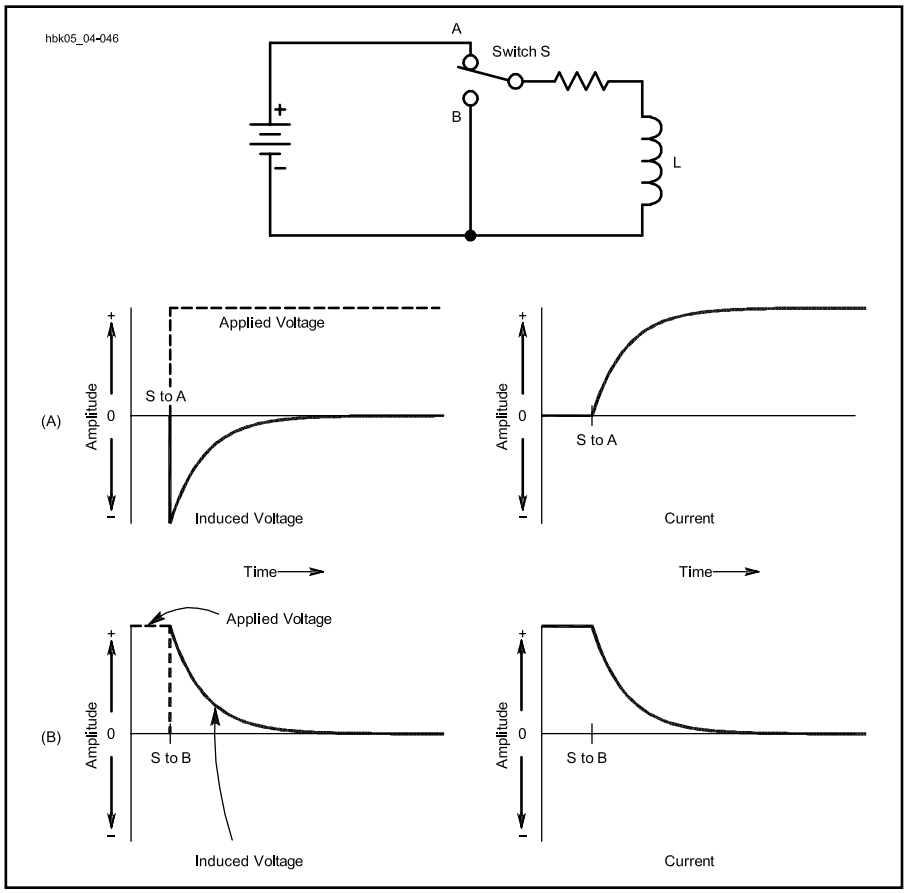


Fig 2.47 — Inductive circuit showing the generation of induced voltage and the rise of current when voltage is applied to an inductor at A, and the decay of current as the coil shorted at B.

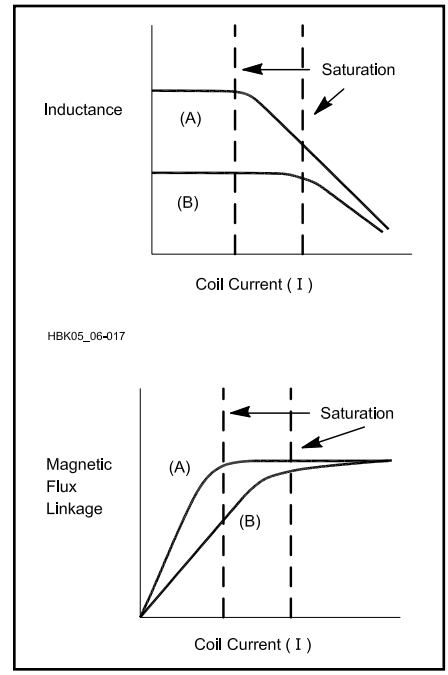


Fig 2.48 — Magnetic flux linkage and inductance plotted versus coil current for (A) a typical iron-core inductor. As the flux linkage Nϕ in the coil saturates, the inductance begins to decrease since inductance = flux linkage / current. The curves marked B show the effect of adding an air gap to the core. The current-handling capability has increased, but at the expense of reduced inductance.

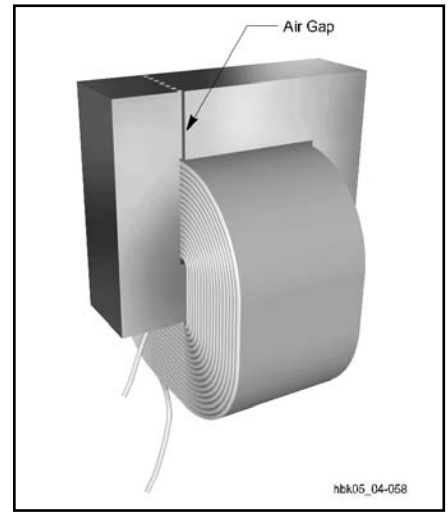


Fig 2.49 — Typical construction of a magnetic-core inductor. The air gap greatly reduces core saturation at the expense of reducing inductance. The insulating laminations between the core layers help to minimize eddy currents, as well.

effect, the growing (or shrinking) magnetic field of each turn produces magnetic lines of force that — in their expansion (or contraction) — intercept the other turns of the coil, inducing a voltage in every other turn. (Recall the two-way relationship between a changing magnetic field and the voltage it creates in a conductor.) The mutuality of the effect, called *magnetic flux linkage* (ψ), multiplies the ability of the coiled conductor to store magnetic energy.

A coil of many turns will have more inductance than one of few turns, if both coils are otherwise physically similar. Furthermore, if an inductor is placed around a magnetic core, its inductance will increase in proportion to the permeability of that core, if the circuit current is below the point at which the core saturates.

In various aspects of radio work, inductors may take values ranging from a fraction of a nanohenry (nH) through millihenrys (mH) up to about 20 H.

EFFECTS OF SATURATION

An important concept for using inductors is that as long as the coil current remains below saturation, the inductance of the coil is essentially constant. **Fig 2.48** shows graphs of magnetic flux linkage (ψ) and inductance (L) vs. current (I) for a typical iron-core inductor both saturated and non-saturated. These quantities are related by the equation

$$\psi = N\phi = LI \qquad (60)$$

where

ψ = the flux linkage
N = number of turns,
ϕ = flux density in webers
L = inductance in henrys, and
I = current in amperes.

In the lower graph, a line drawn from any point on the curve to the (0,0) point will show the effective inductance, L = Nϕ / I, at that current. These results are plotted on the upper graph.

Note that below saturation, the inductance is constant because both ψ and I are increasing at a steady rate. Once the saturation current is reached, the inductance decreases because ψ does not increase anymore (except for the tiny additional magnetic field the current itself provides).

One common method of increasing the saturation current level is to cut a small air gap in the core (see **Fig 2.49**). This gap forces the flux lines to travel through air for a short distance, reducing the permeability of the core. Since the saturation flux linkage of the core is unchanged, this method works by requiring a higher current to achieve saturation. The price that is paid is a reduced inductance below

saturation. The curves in Fig 2.48B show the result of an air gap added to that inductor.

Manufacturer's data sheets for magnetic cores usually specify the saturation flux density. Saturation flux density (ϕ) in gauss can be calculated for ac and dc currents from the following equations:

$$\phi_{ac} = \frac{3.49\ V}{fNA}$$

$$\phi_{dc} = \frac{NIA_L}{10A}$$

where

V = RMS ac voltage
f = frequency, in MHz
N = number of turns
A = equivalent area of the magnetic path in square inches (from the data sheet)
I = dc current, in amperes, and
A_L = inductance index (also from the data sheet).

2.8.4 Mutual Inductance and Magnetic Coupling

MUTUAL INDUCTANCE

When two inductors are arranged with their axes aligned as shown in **Fig 2.50**, current flowing in through inductor 1 creates a magnetic field that intercepts inductor 2. Consequently, a voltage will be induced in inductor 2 whenever the field strength of inductor 1 is changing. This induced voltage is similar to the voltage of self-induction, but since it appears in the second inductor because of current flowing in the first, it is a mutual effect and results from the *mutual inductance* between the two inductors.

When all the flux set up by one coil intercepts all the turns of the other coil, the mutual inductance has its maximum possible value. If only a small part of the flux set up by one coil intercepts the turns of the other, the mutual inductance is relatively small. Two inductors having mutual inductance are said to be *coupled*.

The ratio of actual mutual inductance to the maximum possible value that could theoretically be obtained with two given inductors is called the *coefficient of coupling* between the inductors. It is expressed as a percentage or as a value between 0 and 1. Inductors that have nearly the maximum possible mutual inductance (coefficient = 1 or 100%) are said to be closely, or tightly, coupled. If the mutual inductance is relatively small the inductors are said to be loosely coupled. The degree of coupling depends upon the physical spacing between the inductors and how they are placed with respect to each other. Maximum coupling exists when they have a common or parallel axis and are as close together as possible (for example, one wound over the other). The coupling is least when the inductors are far apart or are placed so their axes are at right angles.

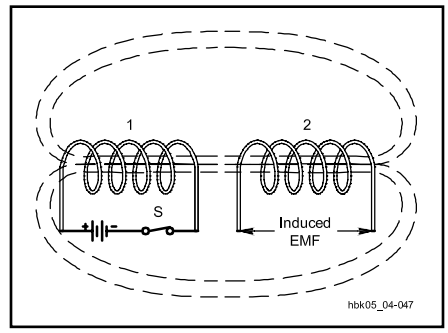


Fig 2.50 — Mutual inductance: When S is closed, current flows through coil number 1, setting up a magnetic field that induces a voltage in the turns of coil number 2.

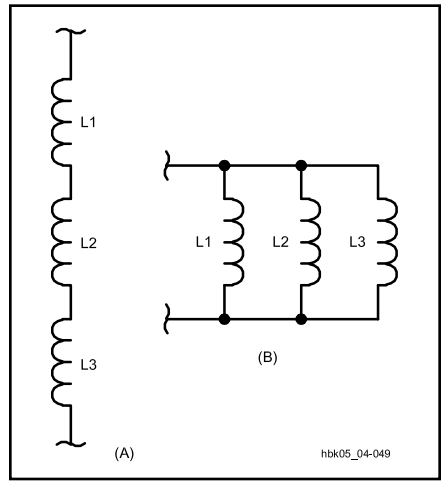


Fig 2.51 — Part A shows inductances in series, and Part B shows inductances in parallel.

The maximum possible coefficient of coupling is closely approached when the two inductors are wound on a closed iron core. The coefficient with air-core inductors may run as high as 0.6 or 0.7 if one inductor is wound over the other, but will be much less if the two inductors are separated. Although unity coupling is suggested by Fig 2.50, such coupling is possible only when the inductors are wound on a closed magnetic core.

Coupling between inductors can be minimized by using separate closed magnetic cores for each. Since an inductor's magnetic field is contained almost entirely in a closed core, two inductors with separate closed cores, such as the toroidal inductor in Fig 2.46 C, can be placed close together in almost any relative orientation without coupling.

UNWANTED COUPLING

The inductance of a short length of straight wire is small, but it may not be negligible. (In free-space, round wire has an inductance on the order of 1 μH/m, but this is affected by wire diameter and the total circuit's physical configuration.) Appreciable voltage may be induced in even a few inches of wire carrying ac by changing magnetic fields with a frequency on the order of 100 MHz or higher. At much lower frequencies or at dc, the inductance of the same wire might be ignored because the induced voltage would be very small.

There are many phenomena, both natural and man-made, that create sufficiently strong or rapidly-changing magnetic fields to induce voltages in conductors. Many of them create brief but intense pulses of energy called *transients* or "spikes." The magnetic fields from these transients intercept wires leading into and out of — and wires wholly within — electronic equipment, inducing unwanted voltages by mutual coupling.

Lightning is a powerful natural source of magnetically-coupled transients. Strong transients can also be generated by sudden changes in current in nearby circuits or wiring. High-speed digital signals and pulses can also induce voltages in adjacent conductors.

2.8.5 Inductances in Series and Parallel

When two or more inductors are connected in series (**Fig 2.51A**), the total inductance is equal to the sum of the individual inductances, provided that the inductors are sufficiently separated so that there is no coupling between them (see the preceding section). That is:

$$L_{total} = L1 + L2 + L3 ... + L_n \qquad (61)$$

If inductors are connected in parallel (Fig 2.51B), again assuming no mutual coupling, the total inductance is given by:

$$L_{total} = \frac{1}{\frac{1}{L1} + \frac{1}{L2} + \frac{1}{L3} + ... + \frac{1}{L_n}} \qquad (62)$$

For only two inductors in parallel, the formula becomes:

$$L_{total} = \frac{L1 \times L2}{L1 + L2} \qquad (63)$$

Thus, the rules for combining inductances in series and parallel are the same as those for resistances, assuming there is no coupling between the inductors. When there is coupling between the inductors, the formulas given above will not yield correct results.

2.8.6 RL Time Constant

As with capacitors, the time dependence of inductor current is a significant property. A comparable situation to an RC circuit exists when resistance and inductance are connected in series. In **Fig 2.52**, first consider the case in which R is zero. Closing S1 sends a current

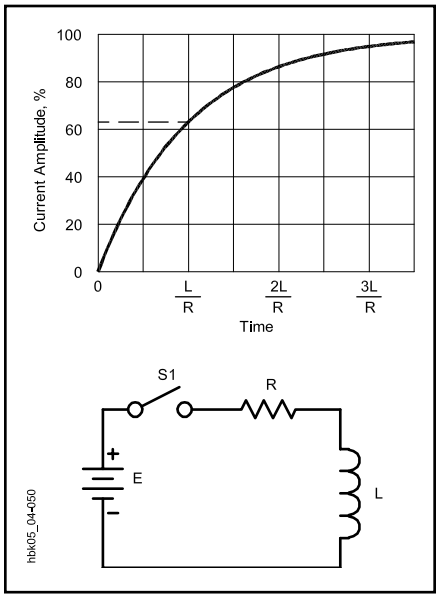


Fig 2.52 — Time constant of an RL circuit being energized.

through the circuit. The instantaneous transition from no current to a finite value, however small, represents a rapid change in current, and an opposing voltage is induced in L. The value of the opposing voltage is almost equal to the applied voltage, so the resulting initial current is very small.

The opposing voltage is created by change in the inductor current and would cease to exist if the current did not continue to increase. With no resistance in the circuit, the current would increase forever, always growing just fast enough to keep the self-induced opposing voltage just below the applied voltage.

When resistance in the circuit limits the current, the opposing voltage induced in L must only equal the difference between E and the drop across R, because that is the voltage actually applied to L. This difference becomes smaller as the current approaches its final value, limited by Ohm's Law to I = E/R. Theoretically, the opposing voltage never quite disappears, and so the current never quite reaches the Ohm's Law limit. In practical terms, the difference eventually becomes insignificant, just as described above for capacitors charging to an applied voltage through a resistor.

The inductor current at any time after the switch in Fig 2.52 has been closed, can be found from:

$$I(t) = \frac{E}{R}\left(1 - e^{-\frac{tR}{L}}\right) \qquad (64)$$

where

I(t) = current in amperes at time t,
E = power source potential in volts,
t = time in seconds after application of voltage,
e = natural logarithmic base = 2.718,
R = circuit resistance in ohms, and
L= inductance in henrys.

(References that explain exponential equations, e, and other mathematical topics are found in the "Radio Math" section of this chapter.) The term E/R in this equation represents the dc value of I, or the value of I(t) when t becomes very large; this is the *steady-state value* of I. If t = L/R, the above equation becomes:

$$V(L/R) = \frac{E}{R}(1 - e^{-1}) \approx 0.632\,\frac{E}{R} \qquad (65)$$

The time in seconds required for the current to build up to 63.2% of the maximum value is called the *time constant* (also the *RL time constant*), and is equal to L/R, where L is in henrys and R is in ohms. (Time constants are also discussed in the section on RC circuits above.) After each time interval equal to this constant, current increases by an additional 63.2% closer to the final value of E/R. This behavior is graphed in Fig 2.52. As is the case with capacitors, after five time constants the current is considered to have reached its maximum value. As with capacitors, we often use the lower- case Greek tau (τ) to represent the time constant.

Example: If a circuit has an inductor of 5.0 mH in series with a resistor of 10 Ω, how long will it take for the current in the circuit to reach full value after power is applied? Since achieving maximum current takes approximately five time constants,

$$t = \frac{5\,L}{R} = \frac{5 \times 5.0 \times 10^{-3}\ \text{H}}{10\ \Omega}$$

$$= 2.5 \times 10^{-3}\ \text{seconds} = 2.5\ \text{ms}$$

Note that if the inductance is increased to 5.0 H, the required time increases by a factor of 1000 to 2.5 seconds. Since the circuit resistance didn't change, the final current is the same for both cases in this example. Increasing inductance increases the time required to reach full current.

Zero resistance would prevent the circuit from ever achieving full current. All practical inductors have some resistance in the wire making up the inductor.

An inductor cannot be discharged in the simple circuit of Fig 2.52 because the magnetic field ceases to exist or "collapses" as soon as the current ceases. Opening S1 does not leave the inductor charged in the way that a capacitor would remain charged. Energy storage in a capacitor depends on the separated charges staying in place. Energy storage in an inductor depends on the charges continuing to move as current.

The energy stored in the inductor's magnetic field attempts to return instantly to the circuit when S1 is opened. The rapidly changing (collapsing) field in the inductor causes a very large voltage to be induced across the

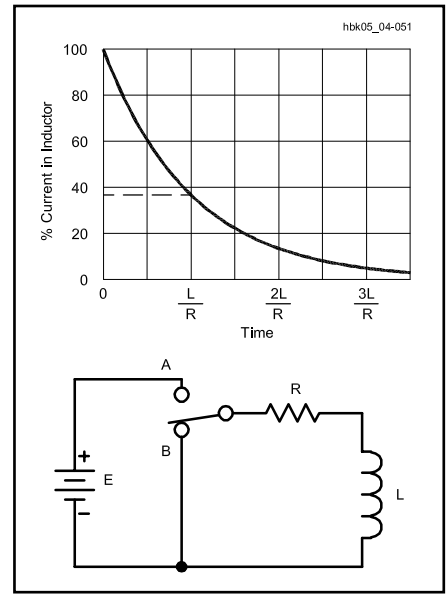


Fig 2.53 — Time constant of an RL circuit being de-energized. This is a theoretical model only, since a mechanical switch cannot change state instantaneously.

inductor. Because the change in current is now in the opposite direction, the induced voltage also reverses polarity. This induced voltage (called *inductive kick-back*) is usually many times larger than the originally applied voltage, because the induced voltage is proportional to the rate at which the field changes.

The common result of opening the switch in such a circuit is that a spark or arc forms at the switch contacts during the instant the switch opens. When the inductance is large and the current in the circuit is high, large amounts of energy are released in a very short time. It is not at all unusual for the switch contacts to burn or melt under such circumstances.

The spark or arc at the opened switch can be reduced or suppressed by connecting a suitable capacitor and resistor in series across the contacts to absorb the energy non-destructively. Such an RC combination is called a *snubber network*. The current rating for a switch may be significantly reduced if it is used in an inductive circuit.

Transistor switches connected to and controlling inductors, such as relays and solenoids, also require protection from the high kick-back voltages. In most cases, a small power diode connected across the relay coil so that it does not conduct current when the inductor is energized (called a *kick-back diode*) will protect the transistor.

If the excitation is removed without breaking the circuit, as shown in **Fig 2.53**, the current will decay according to the formula:

$$I(t) = \frac{E}{R}\left(e^{-\frac{t\,R}{L}}\right) \qquad (66)$$

where t = time in seconds after removal of the source voltage.

After one time constant the current will decay by 63.2% of its steady-state value. (It will decay to 36.8% of the steady-state value.) The graph in Fig 2.53 shows the current-decay waveform to be identical to the voltage-discharge waveform of a capacitor. Be careful about applying the terms *charge* and *discharge* to an inductive circuit, however. These terms refer to energy storage in an electric field. An inductor stores energy in a magnetic field and the usual method of referring to the process is *energize* and *de-energize* (although it is not always followed).

2.8.7 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 **Fig 2.54**, 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 Fig 2.54 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 Fig 2.54 with Fig 2.36. 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.

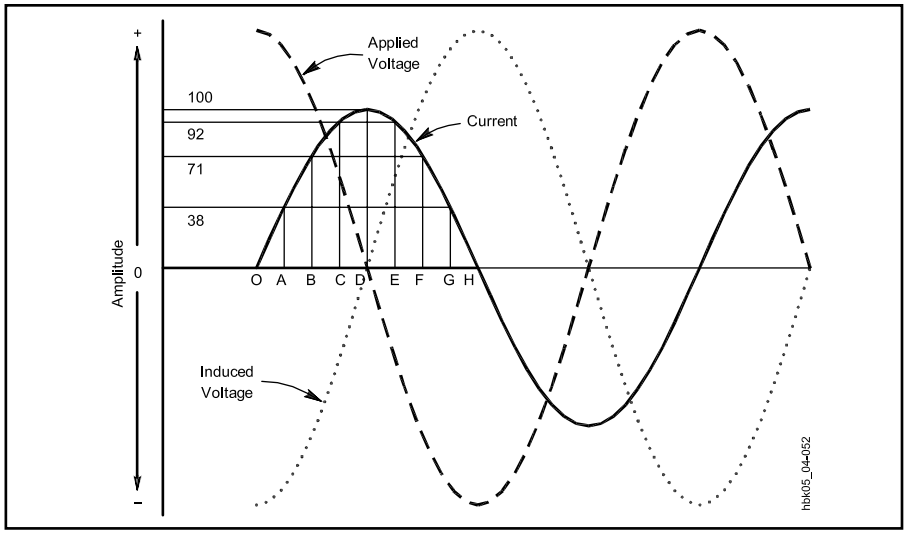


Fig 2.54 — Phase relationships between voltage and current when an alternating current is applied to an inductance.

EDDY CURRENT

Since magnetic core material is usually conductive, the changing magnetic field produced by an ac current in an inductor also induces a voltage in the core. This voltage causes a current to flow in the core. This *eddy current* (so-named because it moves in a closed path, similarly to eddy currents in water) serves no useful purpose and results in energy being dissipated as heat from the core's resistance. Eddy currents are a particular problem in inductors with iron cores. Cores made of thin strips of magnetic material, called *laminations*, are used to reduce eddy currents. (See also the section on Practical Inductors elsewhere in the chapter.)

2.8.8 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 inductive reactance is:

$$X_L = 2 \pi f L \qquad (67)$$

where

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

Example: What is the reactance of an inductor having an inductance of 8.00 H at a frequency of 120 Hz?

$$X_L = 2 \pi f L$$
$$= 6.2832 \times 120 \text{ Hz} \times 8.0 \text{ H}$$
$$= 6030 \Omega$$

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

$$X_L = 2 \pi f L$$
$$= 6.2832 \times 14.0 \text{ MHz} \times 15.0 \mu\text{H}$$
$$= 1320 \Omega$$

The resistance of the wire used to wind 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?

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 calculated from equation 66 is in units of ohms (Ω). The same is true for the combination of mH and kHz.

$$X_L = 2 \pi f L$$
$$= 6.2832 \times 7.0 \text{ MHz} \times 15.0 \mu\text{H}$$
$$= 660 \Omega$$

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 reciprocal of reactance.

$$B = \frac{1}{X_L} \quad (68)$$

where
X_L = reactance, and
B = susceptance.

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

2.9 Working with Reactance

2.9.1 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:

$$E = I X \quad (69)$$

$$I = \frac{E}{X} \quad (70)$$

$$X = \frac{E}{I} \quad (71)$$

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:

$$X_C = \frac{1}{2 \pi f C}$$
$$= \frac{1}{2 \times 3.1416 \times 7.15 \times 10^6 \text{ Hz} \times 200 \times 10^{-12} \text{ F}}$$
$$= \frac{10^6 \Omega}{8980} = 111 \Omega$$

Next, use equation 69:

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

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

$$X_L = 2 \pi f L$$
$$= 2 \times 3.1416 \times 120 \text{ Hz} \times 8.0 \text{ H}$$
$$= 6030 \Omega$$
$$I = E / X_L = 420 / 6030 = 69.6 \text{ mA}$$

Fig 2.55 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 "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 **Fig 2.56**. 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.

2.9.2 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

$$X_{total} = X1 + X2 + X3 ... + X_n \quad (72)$$

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:

$$X_L = 2 \pi f L$$
$$= 2 \times 3.1416 \times 3.8 \times 10^6 \text{ Hz} \times 4 \times 10^{-6} \text{ H}$$
$$= 96 \Omega$$
$$X_{total} = X1 + X2 = 96 \Omega + 96 \Omega = 192 \Omega$$

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

$$L_{total} = L1 + L2 = 4.0 \mu\text{H} + 4.0 \mu\text{H} = 8.0 \mu\text{H}$$
$$X_{total} = 2 \pi f L$$
$$= 2 \times 3.1416 \times 3.8 \times 10^6 \text{ Hz} \times 8.0 \times 10^{-6} \text{ H}$$
$$= 191 \Omega$$

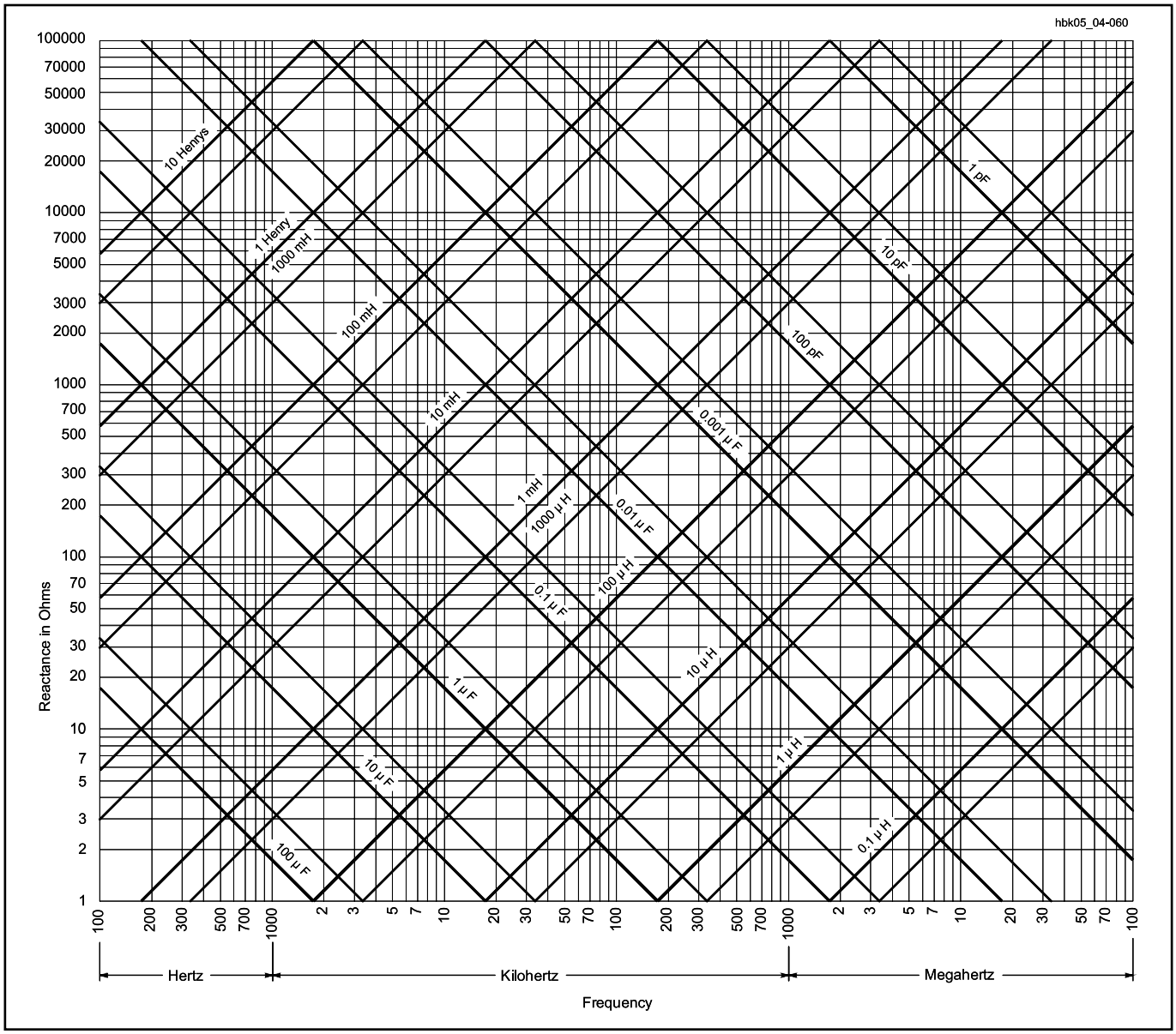


Fig 2.55 — Inductive and capacitive reactance vs frequency. Heavy lines represent multiples of 10, intermediate lines multiples of 5. For example, the light line between 10 µH and 100 µH represents 50 µH; the light line between 0.1 µF and 1 µF represents 0.5 µF, and so on. Other values can be extrapolated 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.)

(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?

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

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

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

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

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

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

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

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

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

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

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

(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

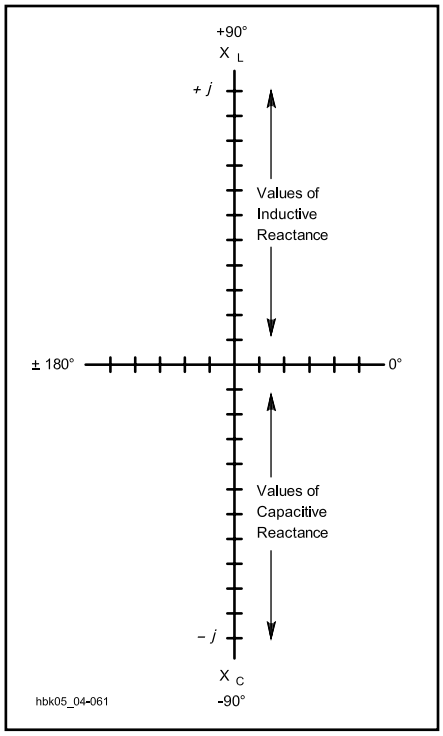


Fig 2.56 — The conventional method of plotting reactances on the vertical axis of a graph, using the upward or "plus" direction for inductive reactance and the downward or "minus" direction for capacitive reactance. The horizontal axis will be used for resistance in later examples.

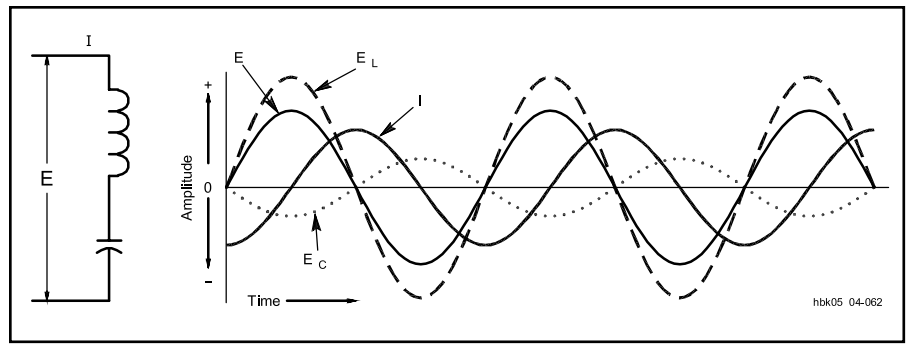


Fig 2.57 — A series circuit containing both inductive and capacitive components, together with representative voltage and current relationships.

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:

$$X_{total} = \frac{1}{\frac{1}{X1} + \frac{1}{X2} + \frac{1}{X3} + ... + \frac{1}{X_n}} \quad (73)$$

or, for exactly two reactances in parallel

$$X_{total} = \frac{X1 \times X2}{X1 + X2} \quad (74)$$

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.

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

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

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

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

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. **Fig 2.57** 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 Fig 2.57 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 as in equation 72. 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).

Another common method of adding the unlike reactances is to use the absolute values of the reactances and subtract capacitive reactance from inductive reactance:

$$X_{total} = X_L - X_C \quad (75)$$

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 Fig 2.57 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, reactance is inductive. Had the result been a negative number, the 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 **Fig 2.58**. 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 Fig 2.58.

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

$$X_{total} = \frac{E}{I_L - I_C} \qquad (76)$$

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 for 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 equation 74 can be used, carrying the positive and negative signs:

$$X_{total} = \frac{X_L \times (-X_C)}{X_L - X_C} = \frac{-X_L \times X_C}{X_L - X_C} \qquad (77)$$

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 Fig 2.58 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, 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.

2.9.3 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.

When a series circuit like the one shown in Fig 2.57 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 **Fig 2.59**, 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 Fig 2.58, the current I_L and I_C are equal and cancel to zero. Since the reactance is inversely

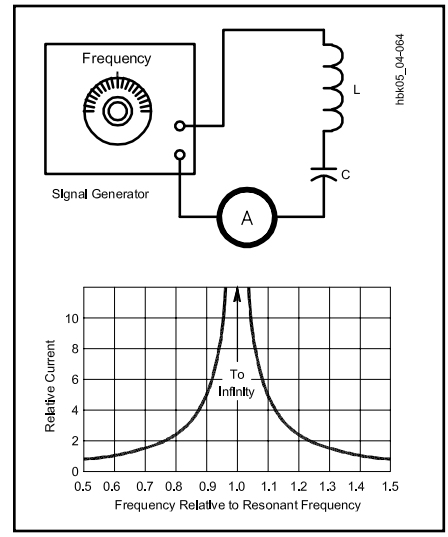


Fig 2.59 — The relative generator current with a fixed voltage in a series circuit containing inductive and capacitive reactances as the frequency approaches and departs from resonance.

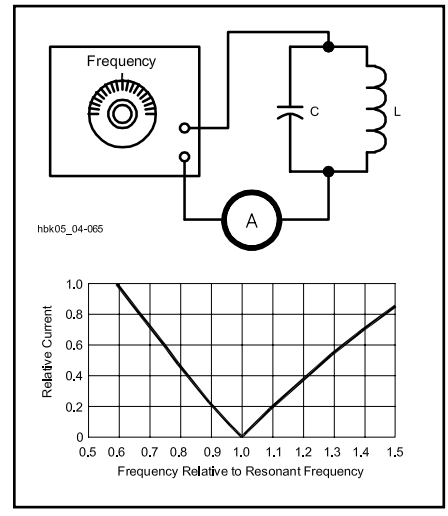


Fig 2.60 — The relative generator current with a fixed voltage in a parallel circuit containing inductive and capacitive reactances as the frequency approaches and departs from resonance. (The circulating current through the parallel inductor and capacitor is a maximum at resonance.)

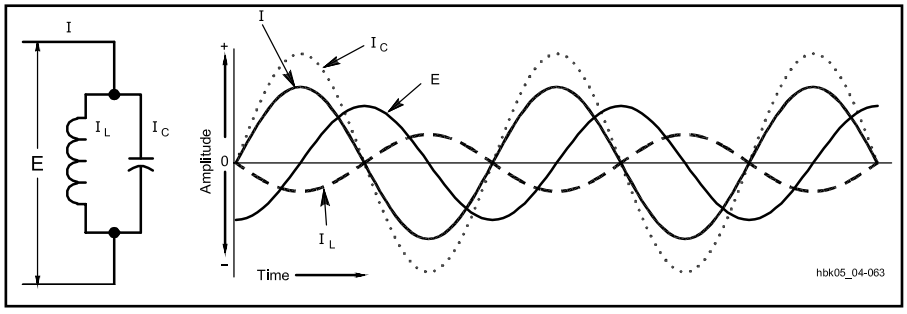


Fig 2.58 — A parallel circuit containing both inductive and capacitive components, together with representative voltage and current relationships.

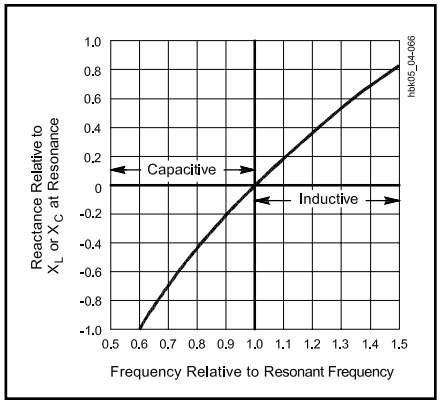


Fig 2.61 — The transition from capacitive to inductive reactance in a series-resonant circuit as the frequency passes resonance.

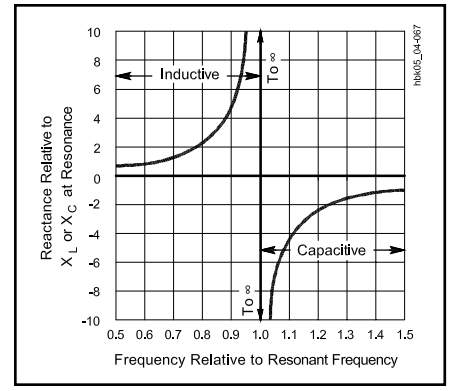


Fig 2.62 — The transition from inductive to capacitive reactance in a parallel-resonant circuit as the frequency passes resonance.

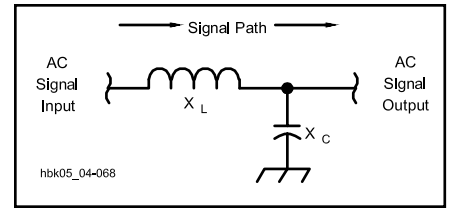


Fig 2.63 — A signal path with a series inductor and a shunt capacitor. The circuit presents different reactances to an ac signal and to its harmonics.

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. **Fig 2.60** shows the theoretical current curve near and at resonance for a purely reactive parallel-resonant circuit. Note that in both Fig 2.59 and Fig 2.60, the departure of current from the resonance value is close to, but not quite, symmetrical above and below the resonant frequency.

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. **Fig 2.61** 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 **Fig 2.62**. 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.

2.9.4 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.

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 **Fig 2.63**. 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. See the **RF and AF Filters** chapter for detailed information on filter theory and construction.

2.10 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 **Fig 2.64**. 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 Fig 2.56 in the preceding section, reactance is graphed on the vertical (Y) axis to record the phase difference between the voltage and the current. **Fig 2.65** adds

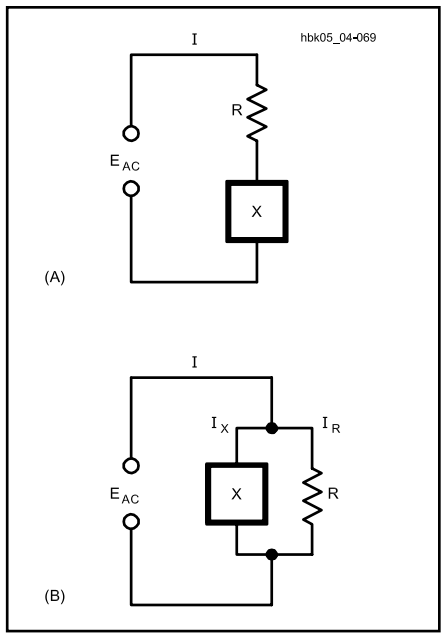


Fig 2.64 — Series and parallel circuits containing resistance and reactance.

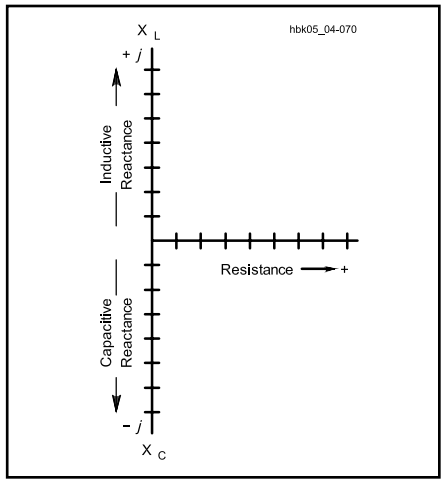


Fig 2.65 — The conventional method of charting impedances on a graph, using the vertical axis for reactance (the upward or "plus" direction for inductive reactance and the downward or "minus" direction for capacitive reactance), and using the horizontal axis for resistance.

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.

2.10.1 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 like "Z = R ± X" can be misleading, because they suggest simple addition. As a result, impedance is often expressed "Z = R ± *j*X."

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 Math" section at the end of this chapter.) 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 actual function of *j* is to indicate that calculating impedance from resistance and reactance requires *vector addition*.

A *vector* is a value with both magnitude and direction, such as velocity; "10 meters/sec to the north". Impedance also has a "direction" as described below. In vector addition, the result of combining two values with a 90° phase difference is a quantity different from the simple *algebraic addition* of the two values. The result will have a phase difference intermediate between 0° and 90°.

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 **Fig 2.66**, 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

$$Z = R + jX \qquad (78)$$

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.

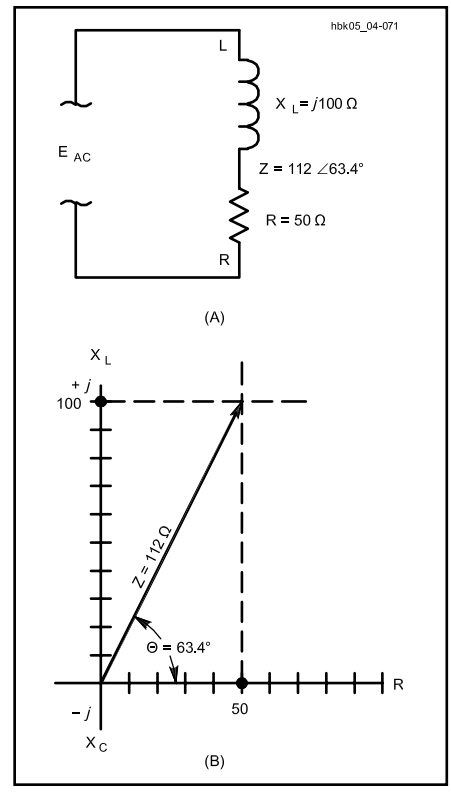


Fig 2.66 — A series circuit consisting of an inductive reactance of 100 Ω and a resistance of 50 Ω. At B, the graph plots the resistance, reactance, and impedance.

POLAR FORM OF IMPEDANCE AND PHASE ANGLE

As the graph in Fig 2.66 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, |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

$$Z = |Z| \angle\theta \qquad (79)$$

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:

$$|Z| = \sqrt{R^2 + X^2} \qquad (80)$$

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

$$\tan \theta = \frac{X}{R} \quad (81)$$

where

X = the reactance, and
R = the resistance.

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

$$\theta = \arctan \frac{X}{R} \quad (82)$$

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 Fig 2.66,

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

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 $|Z| \angle \theta$ 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. **Fig 2.67** 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 $Z = |12.0\ \Omega| \angle -42.0°$.

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:

$$X = |Z| \times \sin \theta \text{ (ohms)} \quad (83)$$

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:

$$R = |Z| \times \cos \theta \text{ (ohms)} \quad (84)$$

Then from our example:

$$X = 12.0\ \Omega \times \sin(-42°)$$
$$= 12.0\ \Omega \times -0.669 = -8.03\ \Omega$$

$$R = 12.0\ \Omega \times \cos(-42°)$$
$$= 12.0\ \Omega \times 0.743 = 8.92\ \Omega$$

Since X is a negative value, it is plotted on the lower vertical axis, as shown in Fig 2.67, 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 section on "Radio Math" at the end of the chapter 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, equation 80 can be expressed in either of two forms, depending on whether the reactance is inductive (equation 85) or capacitive (equation 86):

$$|Z| = \sqrt{R^2 + (2\ \pi\ f\ L)^2} \quad (85)$$

$$|Z| = \sqrt{R^2 + \left(\frac{1}{2\ \pi\ f\ C}\right)^2} \quad (86)$$

Example: What is the impedance of a circuit like Fig 2.66 with a resistance of 100 Ω and a 7.00-μH inductor operating at a frequency of 7.00 MHz? Using equation 85,

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

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

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

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

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

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 (equation 82):

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

$$= \arctan(3.08) = 72.0°$$

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

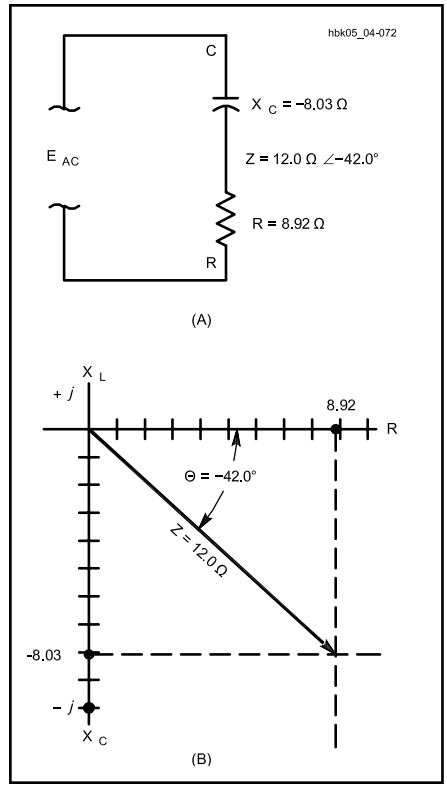


Fig 2.67 — A series circuit consisting of a capacitive reactance and a resistance: the impedance is given as 12.0 Ω at a phase angle θ of –42 degrees. At B, the graph plots the resistance, reactance, and impedance.

2.10.2 Calculating Z From R and X in Parallel Circuits

In a parallel circuit containing reactance and resistance, such as shown in **Fig 2.68**, calculation of the resultant impedance from the values of R and X does not proceed by

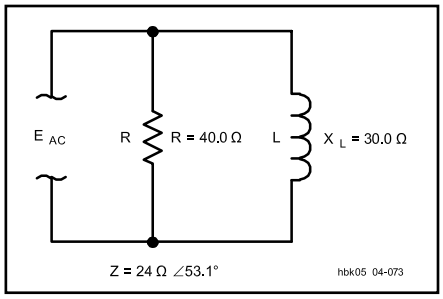


Fig 2.68 — A parallel circuit containing an inductive reactance of 30.0 Ω and a resistor of 40.0 Ω. No graph is given, since parallel impedances can not be manipulated graphically in the simple way of series impedances.

direct combination as for series circuits. The general formula for parallel circuits is:

$$|Z| = \frac{RX}{\sqrt{R^2 + X^2}} \qquad (87)$$

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

$$\theta = \arctan\left(\frac{R}{X}\right) \qquad (88)$$

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: An inductor with a reactance of 30.0 Ω is in parallel with a resistor of 40.0 Ω. What is the resulting impedance and phase angle?

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

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

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

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

$$\theta = \arctan\left(1.33\right) = 53.1°$$

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.)

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

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

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

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

$$\theta = \arctan\left(-0.750\right) = -36.9°$$

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

2.10.3 Admittance

Just as the inverse of resistance is conductance (*G*) and the inverse of reactance is susceptance (*B*), so, too, impedance has an inverse: admittance (*Y*), measured in siemens (S). Thus,

$$Y = \frac{1}{Z} \qquad (89)$$

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,

$$Y = \frac{I}{E} \qquad (90)$$

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

$$Y = G \pm jB = |Y|\angle\theta \qquad (91)$$

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

One handy use for admittance is in simplifying parallel circuit impedance calculations. Similarly to the rules stated previously for combining conductance, the admittance of a parallel combination of reactance and resistance is the vector addition of susceptance and conductance. In other words, for parallel circuits:

$$|Y| = \sqrt{G^2 + B^2} \qquad (92)$$

where

$|Y|$ = magnitude of the admittance in siemens,
G = conductance or 1 / R in siemens, and
B = susceptance or 1 / X in siemens.

Example: An inductor with a reactance of 30.0 Ω is in parallel with a resistor of 40.0 Ω. What is the resulting impedance and phase angle? The susceptance is 1 / 30.0 Ω = 0.0333 S and the conductance is 1 / 40.0 Ω = 0.0250 S.

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

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

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

The phase angle in terms of conductance and susceptance is:

$$\theta = \arctan\left(\frac{B}{G}\right) \qquad (93)$$

In this example,

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

Again, since the reactive component is inductive, the phase angle is positive. For a capacitively reactive parallel circuit, the phase angle would have been negative. Compare these results with the example calculation of the impedance for the same circuit earlier in the section.

Conversion between resistance, reactance and impedance and conductance, susceptance 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.

2.10.4 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.

Fig 2.69 illustrates a more difficult case in which a circuit contains two different reactive elements in series, along with a 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 Fig 2.69B demonstrates, the resultant reactance is the

difference between the two series reactances.

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 **Fig 2.70** 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.

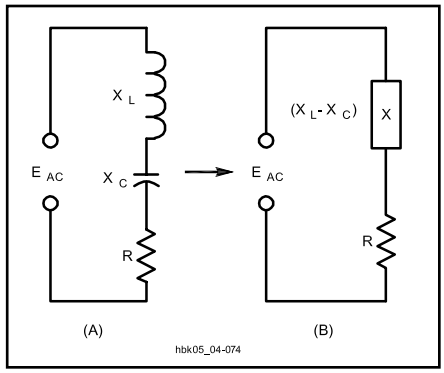


Fig 2.69 — A series impedance containing mixed capacitive and inductive reactances can be reduced to a single reactance plus resistance by combining the reactances algebraically.

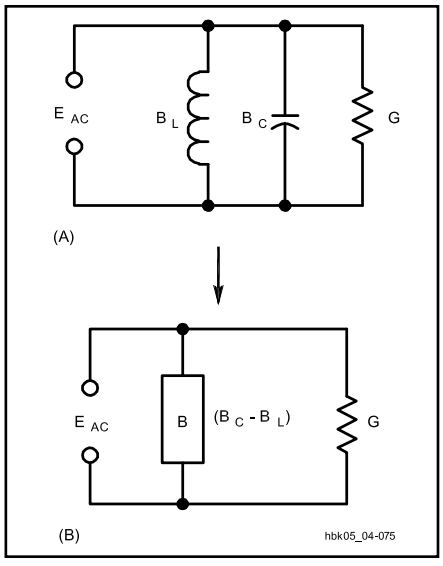


Fig 2.70 — A parallel impedance containing mixed capacitive and inductive reactances can be reduced to a single reactance plus resistance using formulas shown earlier in the chapter. By converting reactances to susceptances, as shown in A, you can combine the susceptances algebraically into a single susceptance, as shown in B.

2.10.5 Equivalent Series and Parallel Circuits

The two circuits shown in Fig 2.64 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.

A series RX circuit can be converted into its parallel equivalent by means of the formulas:

$$R_P = \frac{R_S^2 + X_S^2}{R_S} \qquad (94)$$

$$X_P = \frac{R_S^2 + X_S^2}{X_S} \qquad (95)$$

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:

$$R_S = \frac{R_P\, X_P^2}{R_P^2 + X_P^2} \qquad (96)$$

and

$$X_S = \frac{R_P^2\, X_P}{R_P^2 + X_P^2} \qquad (97)$$

Example: Let the series circuit in Fig 2.64 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?

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

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

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

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

In the parallel circuit of Fig 2.64, a capacitive reactance of 100 Ω and a resistance of 100 Ω to be equivalent would the series circuit.

2.10.6 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:

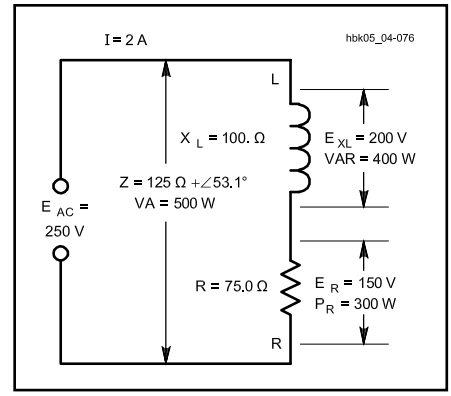


Fig 2.71 — A series circuit consisting of an inductive reactance of 100 Ω and a resistance of 75.0 Ω. Also shown is the applied voltage, voltage drops across the circuit elements, and the current.

$$E = I\,Z \qquad (98)$$

$$I = \frac{E}{Z} \qquad (99)$$

$$Z = \frac{E}{I} \qquad (100)$$

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). **Fig 2.71** 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:

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

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

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,

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

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

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

2.10.7 Reactive Power and Power Factor

Although purely reactive circuits, whether simple or complex, show a measurable ac voltage and current, we cannot simply multiply the two together to arrive at power. Power is the rate at which energy is consumed by a circuit, and purely reactive circuits do not consume energy. 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 as the field collapses 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.

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:

$$VAR = I \times E \qquad (101)$$

$$VAR = I^2 \times X \qquad (102)$$

$$VAR = \frac{E^2}{X} \qquad (103)$$

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 Fig 2.71, 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).

$$PF = \frac{P_{consumed}}{P_{apparent}} = \frac{R}{Z} \qquad (104)$$

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 Fig 2.71 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.

$$\cdots = \cdots\cdots\cdots \times \qquad (105)$$

An equivalent definition of power factor is:

$$PF = \cos\theta \qquad (106)$$

where θ is the phase angle of the circuit impedance.

Since the phase angle in the example equals:

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

$$\theta = \arctan\left(1.33\right) = 53.1°$$

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."

2.11 Quality Factor (Q) of Components

Components that store energy, like capacitors and inductors, may be compared in terms of *quality factor* or *Q factor*, abbreviated *Q*. The Q of any such component is the ratio of its ability to store energy to the sum total of all energy losses within the component. In practical terms, this ratio reduces to the formula:

$$Q = \frac{X}{R} \qquad (107)$$

where

Q = quality factor (no units),

X = X_L (inductive reactance) for inductors and X_C (capacitive reactance) for capacitors (in ohms), and

R = the sum of all resistances associated with the energy losses in the component (in ohms).

The Q of capacitors is ordinarily high. Good quality ceramic capacitors and mica capacitors may have Q values of 1200 or more. Small ceramic trimmer capacitors may have Q values too small to ignore in some applica-

tions. Microwave capacitors can have poor Q values; 10 or less at 10 GHz and higher frequencies.

Inductors are subject to many types of electrical energy losses, however, such as wire resistance, core losses and skin effect. All electrical conductors have some resistance through which electrical energy is lost as heat. Moreover, inductor wire must be sized to handle the anticipated current through the inductor. 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 a conductive material, such as iron, ferrite, or brass, the core will introduce additional losses of energy. 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 a resistor in series with the inductor (as in Figs 2.52 and 2.53), although there is no separate component represented by the resistor symbol. As a result of inherent energy losses, inductor Q rarely, if ever, 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.

AC Component Summary

	Resistor	*Capacitor*	*Inductor*
Basic Unit	ohm (Ω)	farad (F)	henry (H)
Units Commonly Used		microfarads (μF)	millihenrys (mH)
		picofarads (pF)	microhenrys (μH)
Time constant	(None)	RC	L/R
Voltage-Current Phase	In phase	Current leads voltage	Voltage leads current
		Voltage lags 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

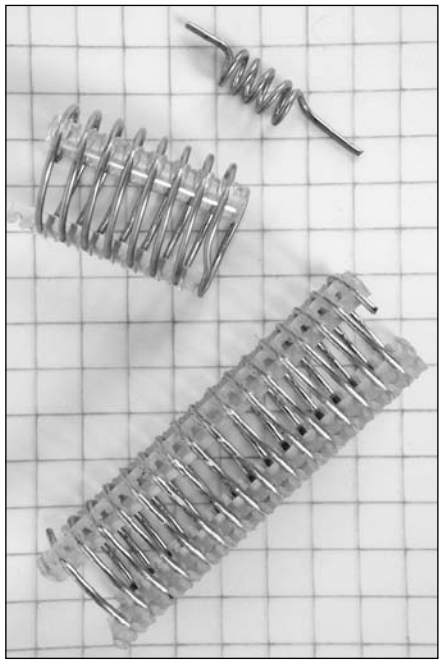
2.12 Practical Inductors

Various facets of radio circuits make use of inductors ranging from the tiny up to the massive. Small values of inductance, such as those inductors in **Fig 2.72A**, serve mostly in RF circuits. They may be self-supporting, air-core or air-wound inductors or the winding may be supported by nonmagnetic strips or a form. Phenolic, certain plastics and ceramics are the most common *coil forms* for air-core inductors. These inductors range in value from a few hundred μH for medium- and high-frequency circuits down to tenths of a μH at VHF and UHF.

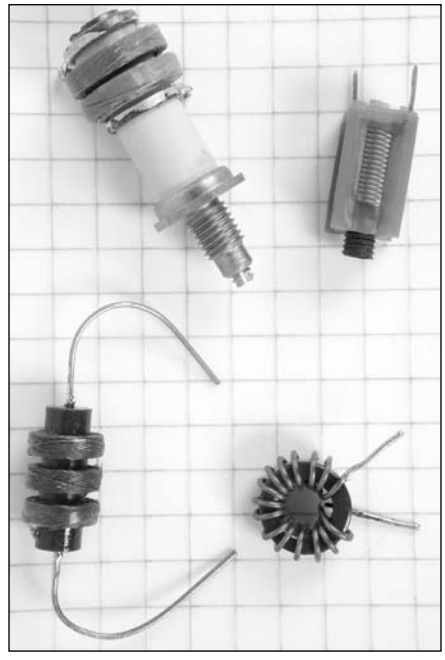
The most common inductor in small-signal RF circuits is the *encapsulated inductor*. These components look a lot like carbon-composition or film resistors and are often marked with colored paint stripes to indicate value. (The chapter **Component Data and References** contains information on inductor color codes and marking schemes.) These inductors have values from less than 1 μH to a few mH. They cannot handle much current without saturating or over-heating.

It is possible to make solenoid inductors variable by inserting a moveable *slug* in the center of the inductor. (Slug-tuned inductors normally have a ceramic, plastic or phenolic insulating form between the conductive slug and the inductor winding.) If the slug material is magnetic, such as powdered iron, the inductance increases as the slug is moved into the center of the inductor. If the slug is brass or some other nonmagnetic material, inserting the slug will reduce the inductor's inductance.

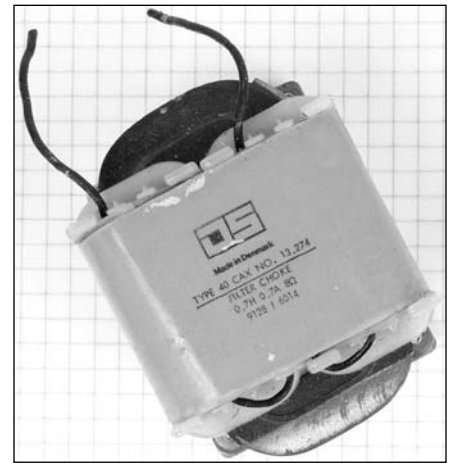
An alternative to air-core inductors for



(A)



(B)



(C)

Fig 2.72 — Part A shows small-value air-wound inductors. Part B shows some inductors with values in the range of a few millihenrys and C shows a large inductor such as might be used in audio circuits or as power-supply chokes. The ¼-inch-ruled graph paper background provides a size comparison.

RF work are *toroidal* inductors (or *toroids*) wound on powdered-iron or ferrite cores. The availability of many types and sizes of powdered-iron cores has made these inductors popular for low-power fixed-value service. The toroidal shape concentrates the inductor's field nearly completely inside the inductor, eliminating the need in many cases for other forms of shielding to limit the interaction of the inductor's magnetic field with the fields of other inductors. (Ferrite core materials are discussed here and in the chapter on **RF Techniques**.)

Fig 2.72B shows samples of inductors in the millihenry (mH) range. Among these inductors are multi-section RF chokes designed to block RF currents from passing beyond them to other parts of circuits. Low-frequency radio work may also use inductors in this range of values, sometimes wound with *litz wire*. Litz wire is a special version of stranded wire, with each strand insulated from the others, and is used to minimize losses associated with skin effect.

For audio filters, toroidal inductors with values up to 100 mH are useful. Resembling powdered-iron-core RF toroids, these inductors are wound on ferrite or molybdenum-permalloy cores having much higher permeabilities.

Audio and power-supply inductors appear in Fig 2.72C. Lower values of these iron-core inductors, in the range of a few henrys, are useful as audio-frequency chokes. Larger values up to about 20 H may be found in power supplies, as choke filters, to suppress 120-Hz ripple. Although some of these inductors are open frame, most have iron covers to confine the powerful magnetic fields they produce.

Although builders and experimenters rarely construct their own capacitors, inductor fabrication is common. In fact, it is often necessary, since commercially available units may be unavailable or expensive. Even if available, they may consist of inductor stock to be trimmed to the required value. Core materials and wire for winding both solenoid and toroidal inductors are readily available. The following information includes fundamental formulas and design examples for calculating practical inductors, along with additional data on the theoretical limits in the use of some materials.

2.12.1 Air-Core Inductors

Many circuits require air-core inductors using just one layer of wire. The approximate inductance of a single-layer air-core inductor may be calculated from the simplified formula:

$$L\,(\mu H) = \frac{d^2\,n^2}{18d + 40\ell} \qquad (108)$$

where

L = inductance in microhenrys,
d = inductor diameter in inches (from wire center to wire center),
ℓ = inductor length in inches, and
n = number of turns.

The notation is illustrated in **Fig 2.73**. This formula is a close approximation for inductors having a length equal to or greater than 0.4d. (Note: Inductance varies as the square of the turns. If the number of turns is doubled, the inductance is quadrupled. This relationship is inherent in the equation, but is often overlooked. For example, to double the inductance, add additional turns equal to 1.4 times the original number of turns, or 40% more turns.)

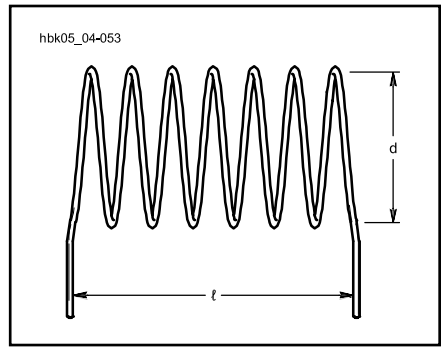


Fig 2.73 — Coil dimensions used in the inductance formula for air-core inductors.

Example: What is the inductance of an inductor if the inductor has 48 turns wound at 32 turns per inch and a diameter of ¾ inch? In this case, d = 0.75, ℓ = 48/32 = 1.5 and n = 48.

$$L\,(\mu H) = \frac{0.75^2 \times 48^2}{(18 \times 0.75) + (40 \times 1.5)}$$

$$= \frac{1300}{74} = 18\ \mu H$$

To calculate the number of turns of a single-layer inductor for a required value of inductance, the formula becomes:

$$n = \frac{\sqrt{L\,(18\,d + 40\,\ell)}}{d} \qquad (109)$$

Example: Suppose an inductance of 10.0 µH is required. The form on which the inductor is to be wound has a diameter of one inch and is long enough to accommodate an inductor of 1¼ inches. Then d = 1.00 inch, ℓ= 1.25 inches and L = 10.0. Substituting:

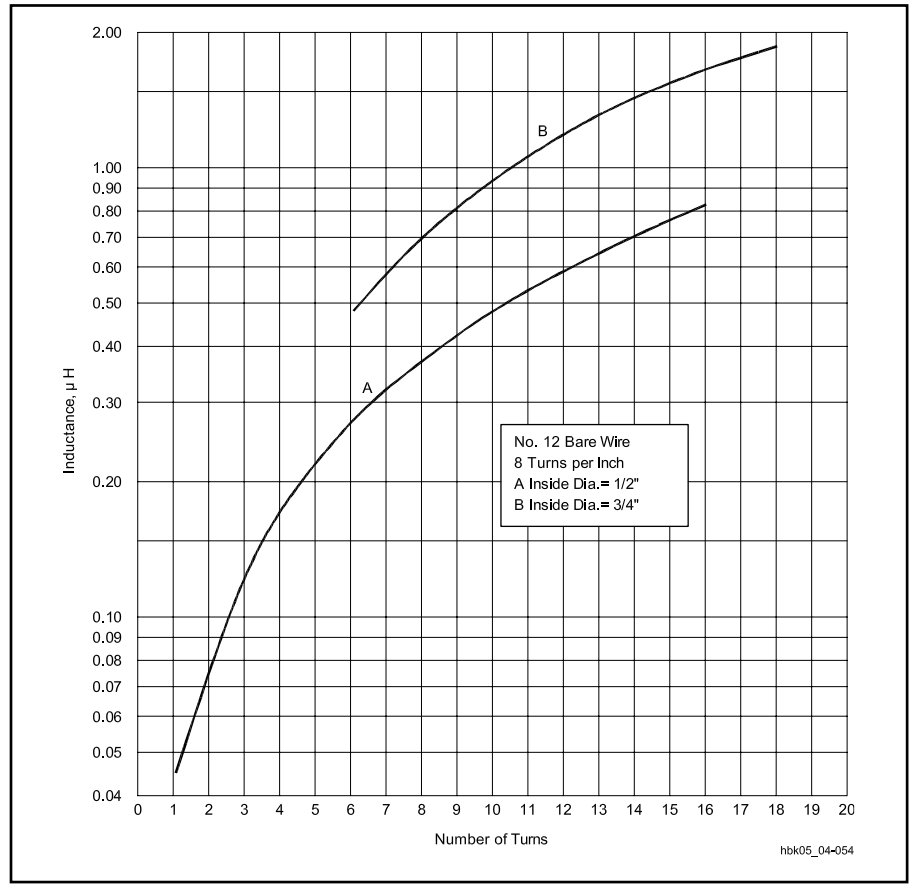


Fig 2.74 — Measured inductance of coils wound with #12 bare wire, eight turns to the inch. The values include half-inch leads.

$$n = \frac{\sqrt{10.0\,[\,(18 \times 1.0) + (40 \times 1.25)}}{1}$$
$$= \sqrt{680} = 26.1 \text{ turns}$$

A 26-turn inductor would be close enough in practical work. Since the inductor will be 1.25 inches long, the number of turns per inch will be 26.1 / 1.25 = 20.9. Consulting the wire table in the **Component Data and References** chapter, we find that #17 AWG enameled wire (or anything smaller) can be used. The proper inductance is obtained by winding the required number of turns on the form and then adjusting the spacing between the turns to make a uniformly spaced inductor 1.25 inches long.

Most inductance formulas lose accuracy when applied to small inductors (such as are used in VHF work and in low-pass filters built for reducing harmonic interference to televisions) because the conductor thickness is no longer negligible in comparison with the size of the inductor. **Fig 2.74** shows the measured inductance of VHF inductors and may be used as a basis for circuit design. Two curves are given; curve A is for inductors wound to an inside diameter of ½ inch; curve B is for inductors of ¾ -inch inside diameter. In both curves, the wire size is #12 AWG and the winding pitch is eight turns to the inch (⅛-inch turn spacing). The inductance values include leads ½-inch long.

Machine-wound inductors with the preset diameters and turns per inch are available in many radio stores, under the trade names of B&W Miniductor and Airdux. Information on using such coil stock is provided in the **Component Data and References** chapter to simplify the process of designing high-quality inductors for most HF applications.

Forming a wire into a solenoid increases its inductance, but also introduces distributed capacitance. Since each turn is at a slightly different ac potential, each pair of turns effectively forms a capacitor in parallel with part of the inductor. (See the chapter on **RF Techniques** for information on the effects of these and other factors that affect the behavior of the "ideal" inductors discussed in this chapter.)

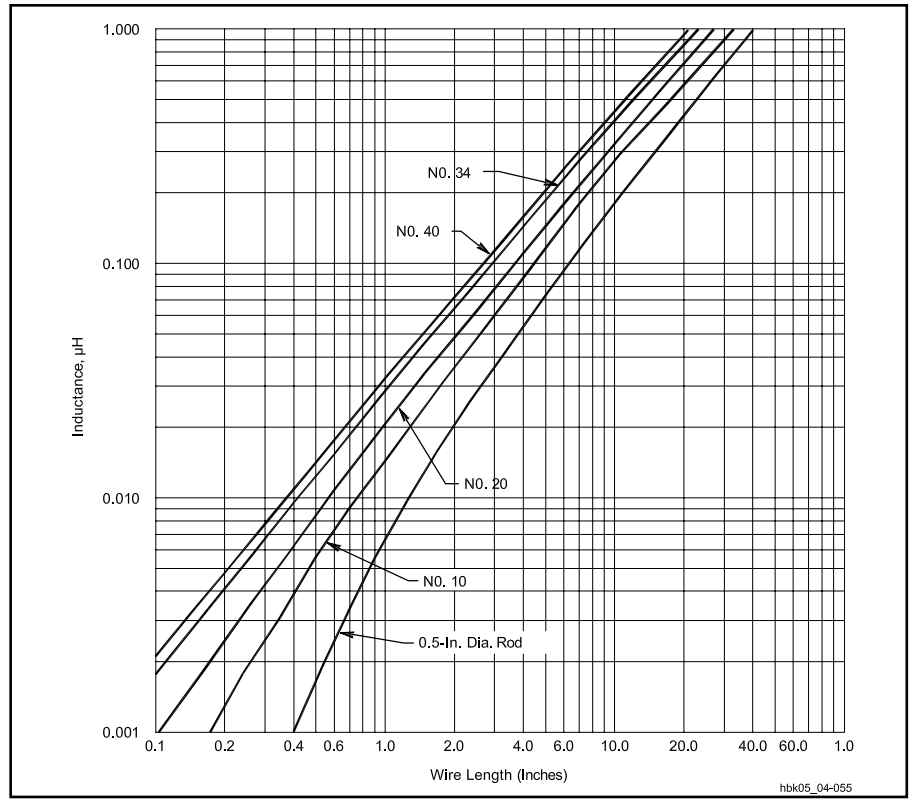


Fig 2.75 — Inductance of various conductor sizes as straight wires.

Moreover, the Q of air-core inductors is, in part, a function of the inductor shape, specifically its ratio of length to diameter. Q tends to be highest when these dimensions are nearly equal. With wire properly sized to the current carried by the inductor, and with high-caliber construction, air-core inductors can achieve Q above 200.

For a large collection of formulas useful in constructing air-core inductors of many configurations, see the "Circuit Elements" section in Terman's *Radio Engineers' Handbook*.

2.12.2 Straight-Wire Inductance

At low frequencies the inductance of a straight, round, nonmagnetic wire in free space is given by:

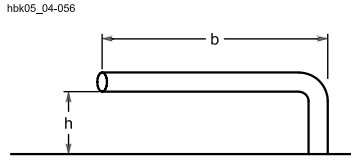
$$L = 0.00508\, b \left\{ \left[\ln\left(\frac{2\,b}{a}\right) \right] - 0.75 \right\} \quad (110)$$

where

L = inductance in µH,
a = wire radius in inches,
b = wire length in inches, and
ln = natural logarithm = 2.303 × common logarithm (base 10).

If the dimensions are expressed in millimeters instead of inches, the equation may still be used, except replace the 0.00508 value with 0.0002.

Skin effect reduces the inductance at VHF and above. As the frequency approaches infinity, the 0.75 constant within the brackets approaches unity. As a practical matter, skin



$$L = 0.0117b\left\{\log_{10}\left[\frac{2h}{a}\left(\frac{b+\sqrt{b^2+a^2}}{b+\sqrt{b^2+4h^2}}\right)\right]\right\} + 0.00508\left(\sqrt{b^2+4h^2}-\sqrt{b^2+a^2}+\frac{b}{4}-2h+a\right)$$

where
L=inductance in µH
a=wire radius in inches
b=wire length parallel to ground plane in inches
h= wire height above ground plane in inches

Fig 2.76 — Equation for determining the inductance of a wire parallel to a ground plane, with one end grounded. If the dimensions are in millimeters, the numerical coefficients become 0.0004605 for the first term and 0.0002 for the second term.

effect will not reduce the inductance by more than a few percent.

Example: What is the inductance of a wire that is 0.1575 inch in diameter and 3.9370 inches long? For the calculations, a = 0.0787 inch (radius) and b = 3.9370 inch.

$$L = 0.00508\, b \left\{ \left[\ln\left(\frac{2\,b}{a} \right) \right] - 0.75 \right\}$$

$$= 0.00508\,(3.9370) \times \left\{ \left[\ln\left(\frac{2 \times 3.9370}{0.0787} \right) \right] - 0.75 \right\}$$

$$= 0.020 \times [\ln(100) - 0.75]$$

$$= 0.020 \times (4.60 - 0.75)$$

$$= 0.020 \times 3.85 = 0.077\ \mu H$$

Fig 2.75 is a graph of the inductance for wires of various radii as a function of length.

A VHF or UHF tank circuit can be fabricated from a wire parallel to a ground plane, with one end grounded. A formula for the inductance of such an arrangement is given in **Fig 2.76**.

Example: What is the inductance of a wire 3.9370 inches long and 0.0787 inch in radius, suspended 1.5748 inch above a ground plane? (The inductance is measured between the free end and the ground plane, and the formula includes the inductance of the 1.5748-inch grounding link.) To demonstrate the use of the formula in Fig 2.76, begin by evaluating these quantities:

$$b + \sqrt{b^2 + a^2}$$

$$= 3.9370 + \sqrt{3.9370^2 + 0.0787^2}$$

$$= 3.9370 + 3.94 = 7.88$$

$$b + \sqrt{b^2 + 4\,(h^2)}$$

$$= 3.9370 + \sqrt{3.9370^2 + 4\,(1.5748^2)}$$

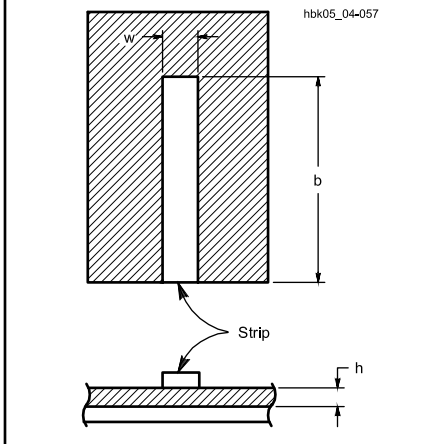
$$= 3.9370 + \sqrt{15.50 + 4\,(2.480)}$$

$$= 3.9370 + \sqrt{15.50 + 9.920}$$

$$= 3.9370 + 5.0418 = 8.9788$$

$$\frac{2\,h}{a} = \frac{2 \times 1.5748}{0.0787} = 40.0$$

$$\frac{b}{a} = \frac{3.9370}{4} = 0.98425$$



$$L = 0.00508b\left(\ln\frac{2b}{w+h} + 0.5 + 0.2235\frac{w+h}{b} \right)$$

where
L=inductance in microhenrys
b=length in inches
w=width in inches
h=thickness in inches

Fig 2.77 — Equation for determining the inductance of a flat strip inductor.

Substituting these values into the formula yields:

$$L = 0.0117 \times 3.9370 \left\{ \log_{10}\left[40.0 \times \left(\frac{7.88}{8.9788} \right) \right] \right\}$$

$$+ 0.00508 \times (5.0418 - 3.94 + 0.98425 - 3.1496 + 0.0787)$$

$$L = 0.0662\ \mu H$$

Another conductor configuration that is frequently used is a flat strip over a ground plane. This arrangement has lower skin-effect loss at high frequencies than round wire because it has a higher surface-area to volume ratio. The inductance of such a strip can be found from the formula in **Fig 2.77**.

2.12.3 Iron-Core Inductors

If the permeability of an iron core in an inductor is 800, then the inductance of any given air-wound inductor is increased 800 times by inserting the iron core. The inductance will be proportional to the magnetic flux through the inductor, other things being equal. The inductance of an iron-core inductor is highly dependent on the current flowing in the inductor, in contrast to an air-core inductor, where the inductance is independent of current because air does not saturate.

Iron-core inductors such as the one sketched in Fig 2.49 are used chiefly in power-supply equipment. They usually have direct current flowing through the winding, and any variation in inductance with current is usually undesirable. Inductance variations may be overcome by keeping the flux density below the saturation point of the iron. Opening the core so there is a small air gap will achieve this goal, as discussed in the earlier section on inductors. The reluctance or magnetic resistance introduced by such a gap is very large compared with that of the iron, even though the gap is only a small fraction of an inch. Therefore, the gap — rather than the iron — controls the flux density. Air gaps in iron cores reduce the inductance, but they hold the value practically constant regardless of the current magnitude.

When alternating current flows through an inductor wound on an iron core, a voltage is induced. Since iron is a conductor, eddy currents also flow in the core as discussed earlier. Eddy currents represent lost power because they flow through the resistance of the iron and generate heat. Losses caused by eddy currents can be reduced by laminating the core (cutting the core into thin strips). These strips or laminations are then insulated from each other by painting them with some insulating material such as varnish or shellac. Eddy current losses add to hysteresis losses, which are also significant in iron-core inductors.

Eddy-current and hysteresis losses in iron increase rapidly as the frequency of the alternating current increases. For this reason, ordinary iron cores can be used only at power-line and audio frequencies — up to approximately 15000 Hz. Even then, a very good grade of iron or steel is necessary for the core to perform well at the higher audio frequencies. Laminated iron cores become completely useless at radio frequencies because of eddy current and hysteresis losses.

2.12.4 Slug-Tuned Inductors

For RF work, the losses in iron cores can be reduced to a more useful level by grinding the iron into a powder and then mixing it with a binder of insulating material in such a way that the individual iron particles are insulated from each other. Using this approach, cores can be made that function satisfactorily even into the VHF range. Because a large part of the magnetic path is through a nonmagnetic material (the binder), the permeability of the powdered iron core is low compared with the values for solid iron cores used at power-line frequencies.

The slug is usually shaped in the form of a cylinder that fits inside the insulating form on which the inductor is wound. Despite the fact that the major portion of the magnetic path for the flux is in air, the slug is quite effective in increasing the inductor inductance. By pushing (or screwing) the slug in and out

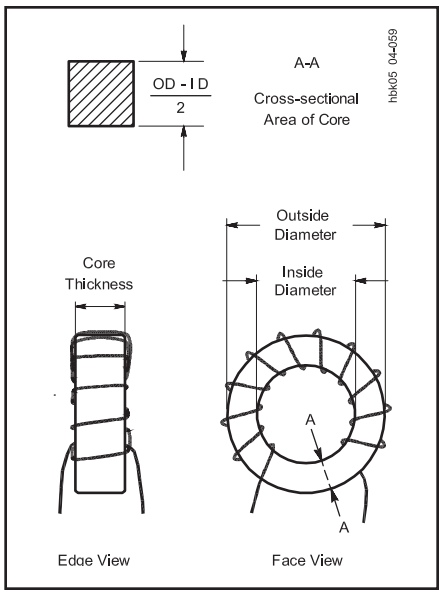


Fig 2.78 — A typical toroidal inductor wound on a powdered-iron or ferrite core. Some key physical dimensions are noted. Equally important are the core material, its permeability, its intended range of operating frequencies, and its A_L value. This is an 11-turn toroid.

of the inductor, the inductance can be varied over a considerable range.

2.12.5 Powdered-Iron Toroidal Inductors

For fixed-value inductors intended for use at HF and VHF, the powdered-iron toroidal core has become the standard in low- and medium-power circuits. **Fig 2.78** shows the general outlines of a toroidal inductor on a magnetic core.

Manufacturers offer a wide variety of core materials, or *mixes*, to create conductor cores that will perform over a desired frequency range with a reasonable permeability. Permeabilities for powdered-iron cores fall in the range of 3 to 35 for various mixes. In addition, core sizes are available in the range of 0.125-inch outside diameter (OD) up to 1.06-inch OD, with larger sizes to 5-inch OD available in certain mixes. The range of sizes permits the builder to construct single-layer inductors for almost any value using wire sized to meet the circuit current demands.

The use of powdered iron in a binder reduces core losses usually associated with iron, while the permeability of the core permits a reduction in the wire length and associated resistance in forming an inductor of a given inductance. Therefore, powdered-iron-core toroidal inductors can achieve Q well above 100, often approaching or exceeding 200 within the frequency range specified for a given core. Moreover, these inductors are considered *self-shielding* since most of the magnetic flux is within the core, a fact that simplifies circuit design and construction.

Each powdered-iron core has an *inductance factor* or *index* A_L determined by the manufacturer. Amidon Corp. is the most common supplier of cores for amateurs (see the Component Data and References chapter) and specifies A_L in μH per 100 turns-squared. The following calculations are based on the Amidon specification. Other manufacturers specify A_L in μH per turn-squared.

To calculate the inductance of a powdered-iron toroidal inductor using the Amidon convention for A_L when the number of turns and the core material are known:

$$L = \frac{A_L \times N^2}{10000} \qquad (111)$$

where

L = the inductance in μH,
A_L = the inductance index in μH per 100 turns-squared, and
N = the number of turns.

The builder must then ensure that the core is capable of holding the calculated number of turns of wire of the required wire size.

Example: What is the inductance of a 60-turn inductor on a core with an A_L of 55? This A_L value was selected from manufacturer's information about a 0.8-inch OD core with an initial permeability of 10. This particular core is intended for use in the range of 2 to 30 MHz. See the **Component Data and References** chapter for more detailed data on the range of available cores.

$$L = \frac{A_L \times N^2}{10000} = \frac{55 \times 60^2}{10000}$$

$$= \frac{198000}{10000} = 19.8\ \mu H$$

To calculate the number of turns needed for a particular inductance, use the formula:

$$N = 100\sqrt{\frac{L}{A_L}} \qquad (112)$$

Example: How many turns are needed for a 12.0-μH inductor if the A_L for the selected core is 49?

$$N = 100\sqrt{\frac{L}{A_L}} = 100\sqrt{\frac{12.0}{49}}$$

$$N = 100\sqrt{0.245} = 100 \times 0.495 = 49.5 \text{ turns}$$

If the value is critical, experimenting with 49-turn and 50-turn inductors is in order, especially since core characteristics may vary slightly from batch to batch. Count turns by each pass of the wire through the center of the core. (A straight wire through a toroidal core amounts to a one-turn inductor.) Fine adjustment of the inductance may be possible by spreading or compressing inductor turns.

The power-handling ability of toroidal cores depends on many variables, which include the cross-sectional area through the core, the core material, the numbers of turns in the inductor, the applied voltage and the operating frequency. Although powdered-iron cores can withstand dc flux densities up to 5000 gauss without saturating, ac flux densities from sine waves above certain limits can overheat cores. Manufacturers provide guideline limits for ac flux densities to avoid overheating. The limits range from 150 gauss at 1 MHz to 30 gauss at 28 MHz, although the curve is not linear. To calculate the maximum anticipated flux density for a particular inductor, use the formula:

$$B_{max} = \frac{E_{RMS} \times 10^8}{4.44 \times A_e \times N \times f} \qquad (113)$$

where

B_{max} = the maximum flux density in gauss,
E_{RMS} = the voltage across the inductor,
A_e = the cross-sectional area of the core in square centimeters,
N = the number of turns in the inductor, and
f = the operating frequency in Hz.

Example: What is the maximum ac flux density for an inductor of 15 turns if the frequency is 7.0 MHz, the RMS voltage is 25 V and the cross-sectional area of the core is 0.133 cm^2?

$$B_{max} = \frac{E_{RMS} \times 10^8}{4.44 \times A_e \times N \times f}$$

$$= \frac{25 \times 10^8}{4.44 \times 0.133 \times 15 \times 7.0 \times 10^6}$$

$$= \frac{25 \times 10^8}{62 \times 10^6} = 40 \text{ gauss}$$

Since the recommended limit for cores operated at 7 MHz is 57 gauss, this inductor is well within guidelines.

2.12.6 Ferrite Toroidal Inductors

Although nearly identical in general appearance to powdered-iron cores, ferrite cores differ in a number of important characteristics. Composed of nickel-zinc ferrites for lower permeability ranges and of manganese-zinc ferrites for higher permeabilities, these cores span a permeability range from 20 to above 10000. Nickel-zinc cores with permeabilities from 20 to 800 are useful in high-Q applications, but function more commonly in amateur applications as RF chokes. They are also useful in wide-band transformers, discussed in the chapter **RF Techniques**.

Ferrite cores are often unpainted, unlike powdered-iron toroids. Ferrite toroids and rods often have sharp edges, while powdered-

iron toroids usually have rounded edges.

Because of their higher permeabilities, the A_L values for ferrite cores are higher than for powdered-iron cores. Amidon Corp. is the most common supplier of cores for amateurs (see the Component Data and References chapter) and specifies A_L in mH per 1000 turns-squared. Other manufacturers specify A_L in nH per turns-squared.

To calculate the inductance of a ferrite toroidal inductor using the Amidon convention for A_L when the number of turns and the core material are known:

$$L = \frac{A_L \times N^2}{1000000} \qquad (114)$$

where

L = the inductance in mH,

A_L = the inductance index in mH per 1000 turns-squared, and

N = the number of turns.

The builder must then ensure that the core is capable of holding the calculated number of turns of wire of the required wire size.

Example: What is the inductance of a 60-turn inductor on a core with an A_L of 523? (See the chapter **Component Data and References** for more detailed data on the range of available cores.)

$$L = \frac{A_L \times N^2}{1000000} = \frac{523 \times 60^2}{1000000}$$

$$= \frac{1.88 \times 10^6}{1 \times 10^6} = 1.88 \text{ mH}$$

To calculate the number of turns needed for a particular inductance, use the formula:

$$N = 1000\sqrt{\frac{L}{A_L}} \qquad (115)$$

Example: How many turns are needed for a 1.2-mH inductor if the A_L for the selected core is 150?

$$N = 1000\sqrt{\frac{L}{A_L}} = 1000\sqrt{\frac{1.2}{150}}$$

$$N = 1000\sqrt{0.008} = 1000 \times 0.089 = 89 \text{ turns}$$

For inductors carrying both dc and ac currents, the upper saturation limit for most ferrites is a flux density of 2000 gauss, with power calculations identical to those used for powdered- iron cores. More detailed information is available on specific cores and manufacturers in the **Component Data and References** chapter.

2.13 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. (Special considerations apply to parallel-resonant circuits; they will be covered in their own section.) 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:

$$f = \frac{1}{2\pi\sqrt{L C}} \qquad (116)$$

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:

$$f = \frac{10^3}{2\pi\sqrt{L C}} \qquad (117)$$

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:

$$f^2 = \frac{1}{4\pi^2 L C} \qquad (118)$$

where f, L and C are in basic units. For HF work in terms of MHz, μH and pF, the basic relationship rearranges to these handy formulas:

$$L = \frac{25330}{f^2 C} \qquad (119)$$

$$C = \frac{25330}{f^2 L} \qquad (120)$$

where

f = frequency in MHz,

L = inductance in μH, and

C = capacitance in pF

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?

$$C = \frac{25330}{f^2 L} = \frac{25330}{(21.1^2 \times 2.0)}$$

$$= \frac{25330}{890} = 28.5 \text{ pF}$$

Fig 2.55 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

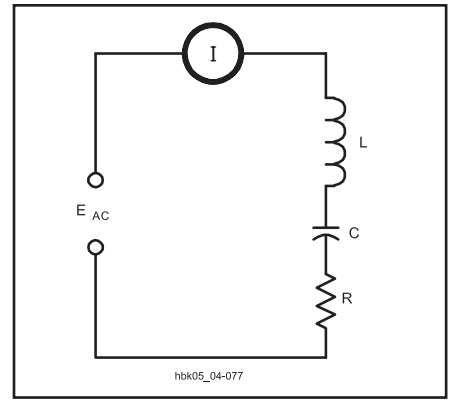


Fig 2.79 — A series circuit containing L, C, and R is resonant at the applied frequency when the reactance of C is equal to the reactance of L. The I in the circle is the schematic symbol for an ammeter.

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.

2.13.1 Series-Resonant Circuits

Fig 2.79 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 losses in the inductor at HF; resistive losses in the capacitor are low enough at those frequencies 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:

$$I = \frac{E}{Z} = \frac{E}{\sqrt{R^2 + \left[2\pi f L - \frac{1}{(2\pi f C}\right]^2}} \qquad (121)$$

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.

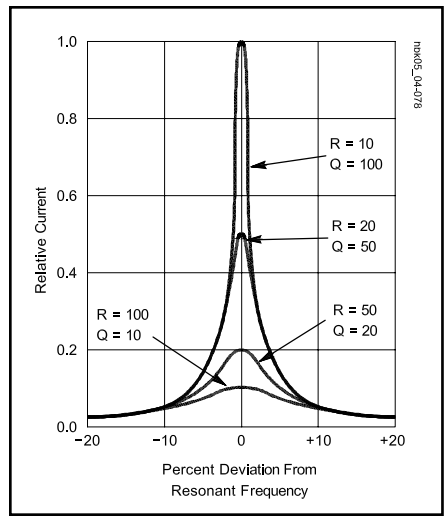


Fig 2.80 — Relative current in series-resonant circuits with various values of series resistance and Q. (An arbitrary maximum value of 1.0 represents current at resonance.) The reactance at resonance for all curves is 1000 Ω. Note that the current is hardly affected by the resistance in the circuit at frequencies more than 10% away from the resonant frequency.

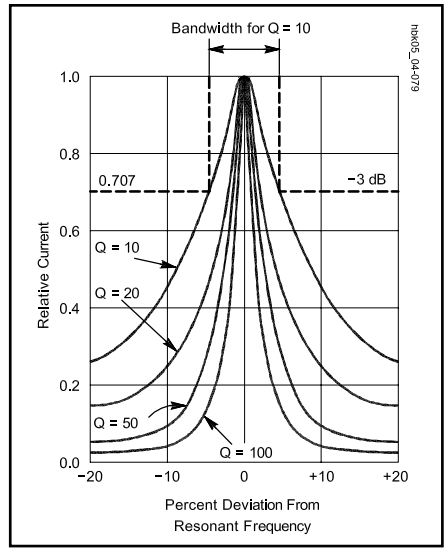


Fig 2.81 — Relative current in series-resonant circuits having different values of Q_U. The current at resonance is normalized to the same level for all curves in order to show the rate of change of decrease in current for each value of Q_U. The half-power points are shown to indicate relative bandwidth of the response for each curve. The bandwidth is indicated for a circuit with a Q_U of 10.

UNLOADED Q

As noted in earlier sections of this chapter, Q is the ratio of reactance or stored energy to 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 resistive losses of the coil 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 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

Fig 2.81 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 ampli-

tude decreases to 0.707 (or $1/\sqrt{2}$) 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 Fig 2.81.

For Q values of 10 or greater, the curves shown in Fig 2.81 are approximately symmetrical. On this assumption, bandwidth (BW) can be easily calculated:

$$BW = \frac{f}{Q_U} \qquad (122)$$

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

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

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

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:

$$Q_U = \frac{f}{BW} \qquad (123)$$

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

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

Table 2.9 provides some simple formulas for estimating the maximum current and phase angle for various bandwidths, if both f and Q_U are known.

Table 2.9
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:

$$E_X = Q_U\ E_{AC} \qquad (124)$$

where

E_X = the voltage across the reactive component,
Q_U = the circuit unloaded Q, and
E_{AC} = the applied voltage in Fig 2.79.

(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, is not exactly like the wire resistance losses in a coil. 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) must be converted to an equivalent series resistance (R_S) before adding it to the inductor's resistance.

$$R_S = \frac{{X_C}^2}{R_P} = \frac{1}{R_P \times (2\,\pi\,f\,C)^2} \qquad (125)$$

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

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

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

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

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

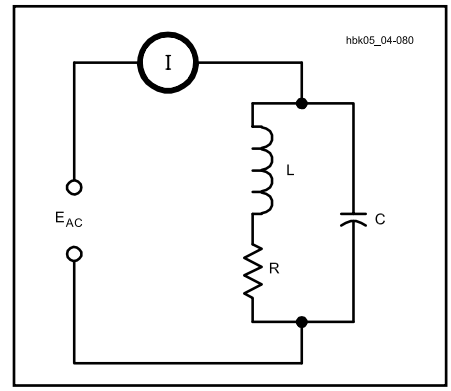


Fig 2.82 — A typical parallel-resonant circuit, with the resistance shown in series with the inductive leg of the circuit. Below a Q_U of 10, resonance definitions may lead to three separate frequencies which converge at higher Q_U levels. See text.

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, the combined losses in the capacitor and the inductor can seriously reduce circuit Q, without special component- and circuit-construction techniques.

2.13.2 Parallel-Resonant Circuits

Although series-resonant circuits are common, the vast majority of resonant circuits used in radio work are *parallel-resonant circuits*. **Fig 2.82** represents a typical HF parallel-resonant circuit. As is the case for series-resonant circuits, the inductor is the chief source of resistive losses, 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

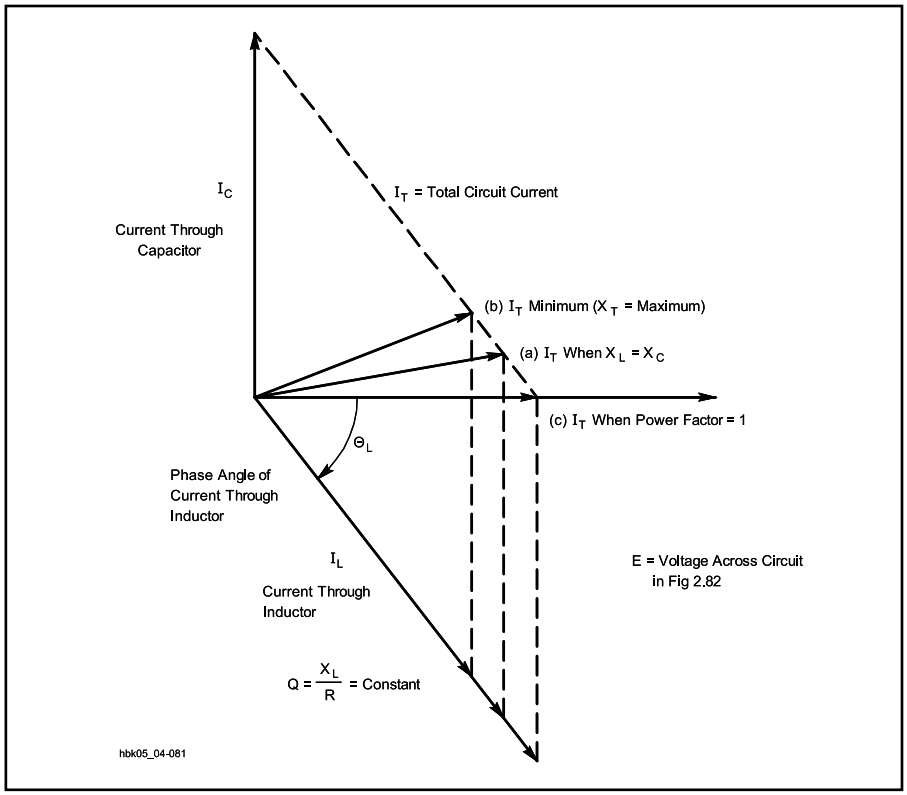


Fig 2.83 — Resonant conditions for a low- Q_U parallel circuit. Resonance may be defined as (a) $X_L = X_C$ (b) minimum current flow and maximum impedance or (c) voltage and current in phase with each other. With the circuit of Fig 2.82 and a Q_U of less than 10, these three definitions may represent three distinct frequencies.

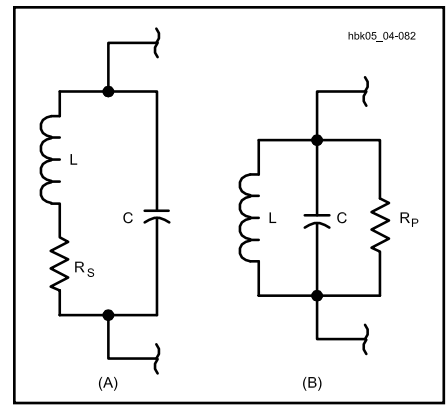


Fig 2.84 — Series and parallel equivalents when both circuits are resonant. The series resistance, R_S in A, is replaced by the parallel resistance, R_p in B, and vice versa. $R_P = X_L^2 / R_S$.

parallel-resonant circuits, respectively.)

Because the conditions in the two legs of the parallel circuit in Fig 2.82 are not the same — the resistance is in only one of the legs — all of the conditions by which series resonance is determined do not occur simultaneously in a parallel-resonant circuit. **Fig 2.83** 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 its minimum possible value and 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 inductor's new value of 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 Fig 2.83 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 **Fig 2.84**. 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:

$$R_P = \frac{X_L^2}{R_S} = \frac{(2\pi f L)^2}{R_S} = Q_U X_L \qquad (126)$$

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?

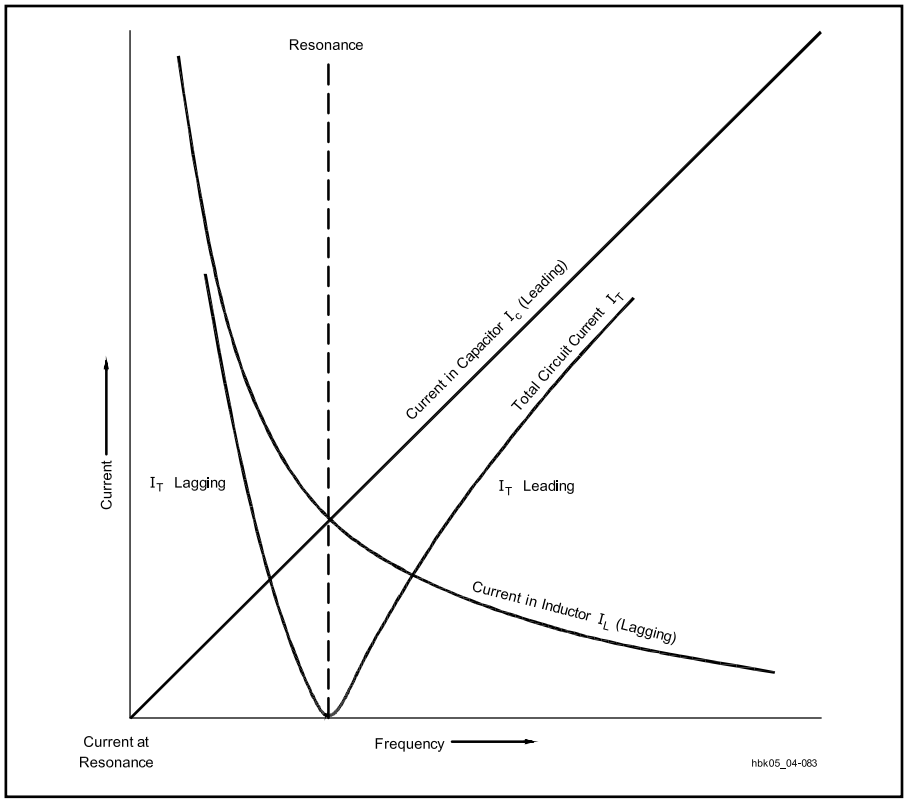


Fig 2.85 — The currents in a parallel-resonant circuit as the frequency moves through resonance. Below resonance, the current lags the voltage; above resonance the current leads the voltage. The base line represents the current level at resonance, which depends on the impedance of the circuit at that frequency.

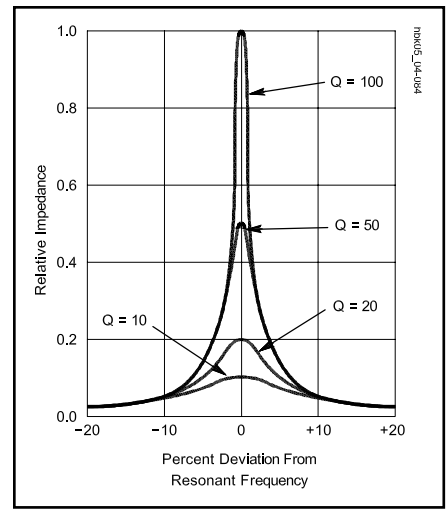


Fig 2.86 — Relative impedance of parallel-resonant circuits with different values of Q_U. The curves are similar to the series-resonant circuit current level curves of Fig 2.80. The effect of Q_U on impedance is most pronounced within 10% of the resonance frequency.

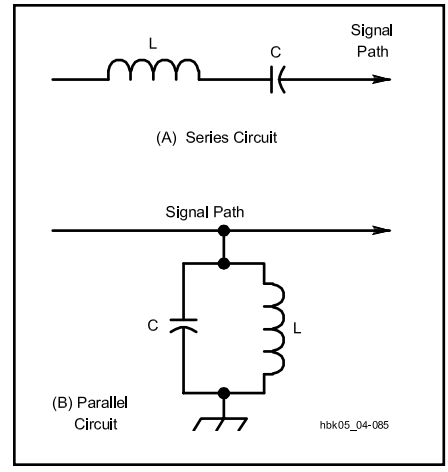


Fig 2.87 — Series- and parallel-resonant circuits configured to perform the same theoretical task: passing signals in a narrow band of frequencies along the signal path. A real design example would consider many other factors.

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

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

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, equation 126 can be rewritten as:

$$Z = \frac{X_L{}^2}{R_S} = \frac{(2\,\pi\,f\,L)^2}{R_S} = Q_U X_L \qquad (127)$$

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. **Fig 2.85** 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:

$$Z = \frac{Z_C\,Z_L}{Z_S} \qquad (128)$$

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 denominator of equation 121.

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,

Table 2.10

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.

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. **Fig 2.86** 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, equations 122 and 123 apply equally to parallel-tuned circuits. That is, $BW = f / Q_U$ and $Q_U = f / BW$, where the resonant frequency and the bandwidth are in the same units. As a handy reminder, **Table 2.10** 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. **Fig 2.87** illustrates this general principle by showing a series-resonant circuit 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.

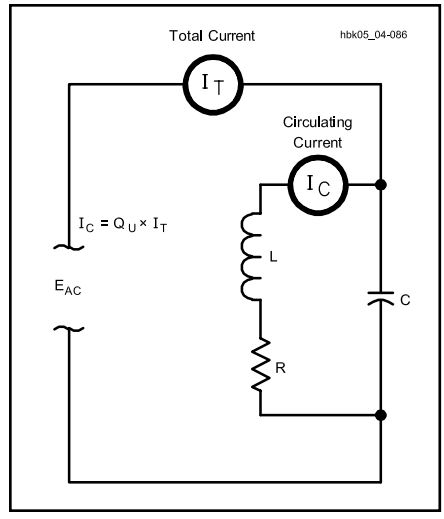


Fig 2.88 — A parallel-resonant circuit redrawn to illustrate both the total current and the circulating current.

Fig 2.88 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:

$$I_C = Q_U I_T \qquad (129)$$

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

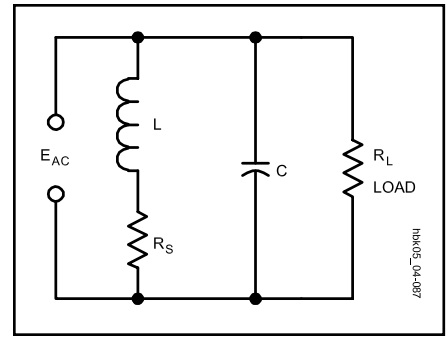


Fig 2.89 — A loaded parallel-resonant circuit, showing both the inductor-loss resistance and the load, R_L. If smaller than the inductor resistance, R_L will control the loaded Q of the circuit (Q_L).

(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 **Fig 2.89**, 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:

$$Q_L = \frac{R_L}{X} \qquad (130)$$

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:

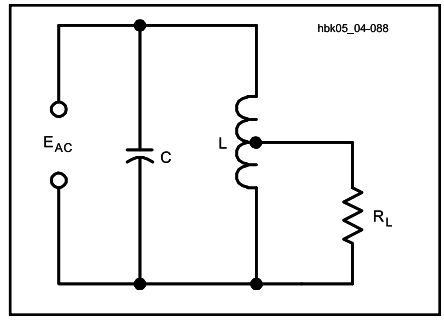


Fig 2.90 — A parallel-resonant circuit with a tapped inductor to effect an impedance match. Although the impedance presented to E_{AC} is very high, the impedance at the connection of the load, R_L, is lower.

$$BW = \frac{f}{Q_U} = \frac{14.0\ MHz}{70.0} = 0.200\ MHz$$
$$= 200\ kHz$$

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

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

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:

$$Z = Q_U\ X_L = 35.0 \times 350\ \Omega$$
$$= 12250\ \Omega = 12.25\ k\Omega$$

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.

Fig 2.90 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 Fig 2.84A, 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:

$$Z_R = \frac{X^2}{R_L} \qquad (131)$$

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 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.

2.14 Transformers

When the ac source current flows through every turn of an inductor, the generation of a counter-voltage and the storage of energy during each half cycle is said to be by virtue of self-inductance. If another inductor — not connected to the source of the original current — is positioned so the magnetic field of the first inductor intercepts the turns of the second inductor, coupling the two inductors and creating mutual inductance as described earlier, a voltage will be induced and current will flow in the second inductor. A load such as a resistor may be connected across the second inductor to consume the energy transferred magnetically from the first inductor.

Fig 2.91 illustrates a pair of coupled inductors, showing an ac energy source connected to one, called the *primary inductor*, and a load connected to the other, called the *secondary inductor*. If the inductors are wound tightly on a magnetic core so that nearly all or magnetic flux from the first inductor intersects with the turns of the second inductor, the pair is said to be *tightly coupled*. Inductors not sharing a common core and separated by a distance would be *loosely coupled*.

The signal source for the primary inductor may be household ac power lines, audio or other waveforms at low frequencies, or RF currents. The load may be a device needing power, a speaker converting electrical energy into sonic energy, an antenna using RF energy for communications or a particular circuit set up to process a signal from a preceding circuit. The uses of magnetically coupled energy in electronics are innumerable.

Mutual inductance (M) between inductors is measured in henrys. Two inductors have a mutual inductance of 1 H under the following conditions: as the primary inductor current changes at a rate of 1 A/s, the voltage across the secondary inductor is 1 V. The level of mutual inductance varies with many factors: the size and shape of the inductors, their relative positions and distance from each other, and the permeability of the inductor core material and of the space between them.

If the self-inductance values of two inductors are known (self-inductance is used in this section to distinguish it from the mutual inductance), it is possible to derive the mutual inductance by way of a simple experiment schematically represented in **Fig 2.92**. Without altering the physical setting or position of two inductors, measure the total coupled inductance, L_C, of the series-connected inductors with their windings complementing each other and again with their windings opposing each other. Since, for the two inductors, $L_C = L1 + L2 + 2M$, in the complementary case, and $L_O = L1 + L2 - 2M$ for the opposing case,

$$M = \frac{L_C - L_O}{4} \qquad (132)$$

The ratio of magnetic flux set up by the secondary inductor to the flux set up by the primary inductor is a measure of the extent to which two inductors are coupled, compared to the maximum possible coupling between them. This ratio is the *coefficient of coupling* (k) and is always less than 1. If k were to equal 1, the two inductors would have the maximum possible mutual coupling. Thus:

$$M = k\sqrt{L1\,L2} \qquad (133)$$

where

M = mutual inductance in henrys,
L1 and L2 = individual coupled inductors, each in henrys, and
k = the coefficient of coupling.

Using the experiment above, it is possible to solve equation 133 for k with reasonable accuracy.

Any two inductors having mutual inductance comprise a *transformer* having a *primary winding* or inductor and a *secondary winding* or inductor. The word "winding" is generally dropped so that transformers are said to have "primaries" and "secondaries." **Fig 2.93** provides a pictorial representation of a typical iron-core transformer, along with the schematic symbols for both iron-core and air-core transformers. Conventionally, the term *transformer* is most commonly applied to coupled inductors having a magnetic core material, while coupled air-wound inductors are not called by that name. They

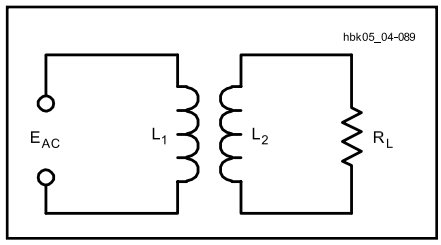


Fig 2.91 — A basic transformer: two inductors — one connected to an ac energy source, the other to a load — with coupled magnetic fields.

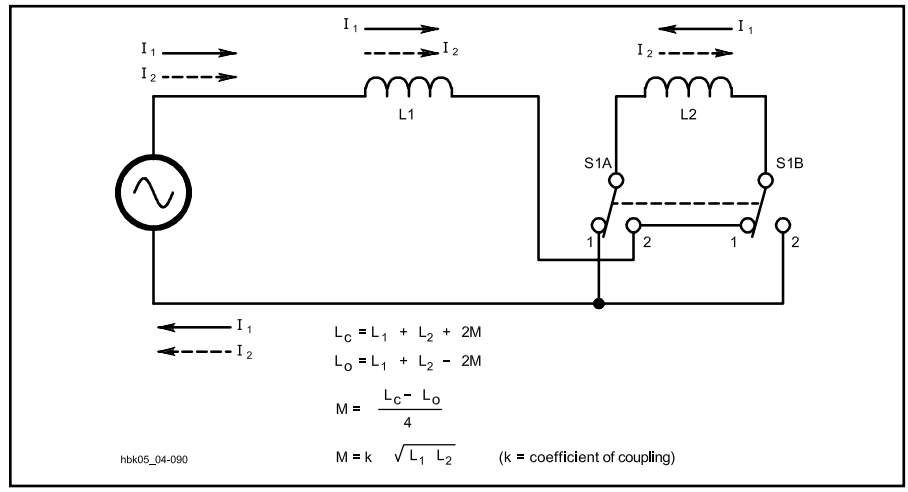


Fig 2.92 — An experimental setup for determining mutual inductance. Measure the inductance with the switch in each position and use the formula in the text to determine the mutual inductance.

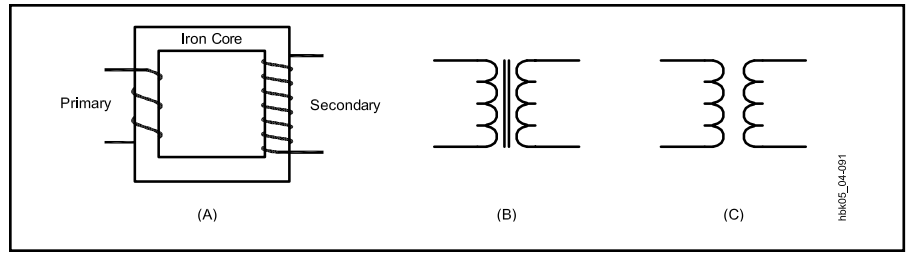


Fig 2.93 — A transformer. A is a pictorial diagram. Power is transferred from the primary coil to the secondary by means of the magnetic field. B is a schematic diagram of an iron-core transformer, and C is an air-core transformer.

are still transformers, however.

We normally think of transformers as ac devices, since mutual inductance only occurs when magnetic fields are changing. A transformer connected to a dc source will exhibit mutual inductance only at the instants of closing and opening the primary circuit, or on the rising and falling edges of dc pulses, because only then does the primary winding have a changing field. There are three principle uses of transformers: to physically isolate the primary circuit from the secondary circuit, to transform voltages and currents from one level to another, and to transform circuit impedances from one level to another. These functions are not mutually exclusive and have many variations.

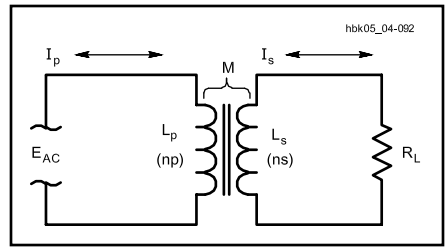


Fig 2.94 — The conditions for transformer action: two coils that exhibit mutual inductance, an ac power source, and a load. The magnetic field set up by the energy in the primary circuit transfers energy to the secondary for use by the load, resulting in a secondary voltage and current.

2.14.1 Basic Transformer Principles

The primary and secondary windings of a transformer may be wound on a core of magnetic material. The permeability of the magnetic material increases the inductance of the windings so a relatively small number of turns may be used to induce a given voltage value with a small current. A closed core having a continuous magnetic path, such as that shown in Fig 2.93, also tends to ensure that practically all of the field set up by the current in the primary winding will intercept or "cut" the turns of the secondary winding.

For power transformers and impedance-matching transformers used at audio frequencies, cores made of soft iron strips or sheets called *laminations* are most common and generally very efficient. At higher frequencies, ferrite or powdered-iron cores are more frequently used. This section deals with basic transformer operation at audio and power frequencies. RF transformer operation is discussed in the chapter on **RF Techniques**.

The following principles presume a coefficient of coupling (k) of 1, that is, a perfect transformer. The value k = 1 indicates that all the turns of both windings link with all the magnetic flux lines, so that the voltage induced per turn is the same with both windings. This condition makes the induced voltage independent of the inductance of the primary and secondary inductors. Iron-core transformers for low frequencies closely approach this ideal condition. **Fig 2.94** illustrates the conditions for transformer action.

VOLTAGE RATIO

For a given varying magnetic field, the voltage induced in an inductor within the field is proportional to the number of turns in the inductor. When the two windings of a transformer are in the same field (which is the case when both are wound on the same closed core), it follows that the induced voltages will be proportional to the number of turns in each winding. In the primary, the induced voltage practically equals, and opposes, the applied voltage, as described earlier. Hence:

$$E_S = E_P\left(\frac{N_S}{N_P}\right) \qquad (134)$$

where

E_S = secondary voltage,
E_P = primary applied voltage,
N_S = number of turns on secondary, and
N_P = number of turns on primary.

Example: A transformer has a primary with 400 turns and a secondary with 2800 turns, and a voltage of 120 V is applied to the primary. What voltage appears across the secondary winding?

$$E_S = 120\text{ V}\left(\frac{2800}{400}\right) = 120\text{ V} \times 7 = 840\text{ V}$$

(Notice that the number of turns is taken as a known value rather than a measured quantity, so they do not limit the significant figures in the calculation.) Also, if 840 V is applied to the 2800-turn winding (which then becomes the primary), the output voltage from the 400-turn winding will be 120 V.

Either winding of a transformer can be used as the primary, provided the winding has enough turns (enough inductance) to induce a voltage equal to the applied voltage without requiring an excessive current. The windings must also have insulation with a voltage rating sufficient for the voltages applied or created. Transformers are called *step-up* or *step-down* transformers depending on whether the secondary voltage is higher or lower than the primary voltage, respectively.

CURRENT OR AMPERE-TURNS RATIO

The current in the primary when no current is taken from the secondary is called the *magnetizing current* of the transformer. An ideal transformer, with no internal losses, would consume no power, since the current through the primary inductor would be 90° out of phase with the voltage. In any properly designed transformer, the power consumed by the transformer when the secondary is open (not delivering power) is only the amount necessary to overcome the losses in the iron core and in the resistance of the wire with which the primary is wound.

When power is transferred from the secondary winding to a load, the secondary current creates a magnetic field that opposes the field established by the primary current. For the induced voltage in the primary to equal the applied voltage, the original magnetizing field must be maintained. Therefore, enough additional current must flow in the primary to create a field exactly equal and opposite to the field set up by the secondary current, leaving the original magnetizing field.

In practical transformer calculations it may be assumed that the entire primary current is caused by the secondary load. This is justifiable because the magnetizing current should be very small in comparison with the primary load current at rated power output.

If the magnetic fields set up by the primary and secondary currents are to be equal, the number of ampere-turns must be equal in each winding. (See the previous discussion of magnetic fields and magnetic flux density.) Thus, primary current multiplied by the primary turns must equal the secondary current multiplied by the secondary turns.

$$I_P N_P = I_S N_S$$

and

$$I_P = I_S\left(\frac{N_S}{N_P}\right) \qquad (135)$$

where

I_P = primary current,
I_S = secondary current,
N_P = number of turns in the primary winding, and
N_S = number of turns in the secondary winding.

Example: Suppose the secondary of the transformer in the previous example is delivering a current of 0.20 A to a load. What will be the primary current?

$$I_P = 0.20\text{ A}\left(\frac{2800}{400}\right) = 0.20\text{ A} \times 7 = 1.4\text{ A}$$

Although the secondary voltage is higher than the primary voltage, the secondary current is lower than the primary current, and by the same ratio. The secondary current in an ideal transformer is 180° out of phase with the primary current, since the field in the secondary just offsets the field in the primary. The phase relationship between the currents in the windings holds true no matter what the phase

difference between the current and the voltage of the secondary. In fact, the phase difference, if any, between voltage and current in the secondary winding will be *reflected* back to the primary as an identical phase difference.

Power Ratio

A transformer cannot create power; it can only transfer it and change the voltage and current ratios. Hence, the power taken from the secondary cannot exceed that taken by the primary from the applied voltage source. There is always some power loss in the resistance of the windings and in the iron core, so in all practical cases the power taken from the source will exceed that taken from the secondary.

$$P_O = \eta P_I \qquad (136)$$

where

P_O = power output from secondary,
P_I = power input to primary, and
η = efficiency factor.

The efficiency, η, is always less than 1 and is commonly expressed as a percentage: if η is 0.65, for instance, the efficiency is 65%.

Example: A transformer has an efficiency of 85.0% at its full-load output of 150 W. What is the power input to the primary at full secondary load?

$$P_I = \frac{P_O}{\eta} = \frac{150\ \text{W}}{0.850} = 176\ \text{W}$$

A transformer is usually designed to have the highest efficiency at the power output for which it is rated. The efficiency decreases with either lower or higher outputs. On the other hand, the losses in the transformer are relatively small at low output but increase as more power is taken. The amount of power that the transformer can handle is determined by its own losses, because these losses heat the wire and core. There is a limit to the temperature rise that can be tolerated, because too high a temperature can either melt the wire or cause the insulation to break down. A transformer can be operated at reduced output, even though the efficiency is low, because the actual loss will be low under such conditions. The full-load efficiency of small power transformers such as are used in radio receivers and transmitters usually lies between about 60 and 90%, depending on the size and design.

IMPEDANCE RATIO

In an ideal transformer — one without losses or leakage inductance (see Transformer Losses) — the primary power, $P_P = E_P I_P$, and secondary power, $P_S = E_S I_S$, are equal. The relationships between primary and secondary voltage and current are also known (equations 134 and 135). Since impedance is the ratio of voltage to current, Z = E/I, the impedances represented in each winding are related as follows:

$$Z_P = Z_S \left(\frac{N_P}{N_S} \right)^2 \qquad (137)$$

where

Z_P = impedance at the primary terminals from the power source,
Z_S = impedance of load connected to secondary, and
N_P/N_S = turns ratio, primary to secondary.

A load of any given impedance connected to the transformer secondary will thus be transformed to a different value at the primary terminals. The impedance transformation is proportional to the square of the primary-to-secondary turns ratio. (Take care to use the primary-to-secondary turns ratio, since the secondary-to-primary ratio is more commonly used to determine the voltage transformation ratio.)

The term *looking into* is often used to mean the conditions observed from an external perspective at the terminals specified. For example, "impedance looking into" the transformer primary means the impedance measured externally to the transformer at the terminals of the primary winding.

Example: A transformer has a primary-to-secondary turns ratio of 0.6 (the primary has six-tenths as many turns as the secondary) and a load of 3000 Ω is connected to the secondary. What is the impedance looking into the primary of the transformer?

$$Z_P = 3000\ \Omega \times (0.6)^2 = 3000\ \Omega \times 0.36$$

$$Z_P = 1080\ \Omega$$

By choosing the proper turns ratio, the impedance of a fixed load can be transformed to any desired value, within practical limits. If transformer losses can be neglected, the transformed (reflected) impedance has the same phase angle as the actual load impedance. Thus, if the load is a pure resistance, the load presented by the primary to the power source will also be a pure resistance. If the load impedance is complex, that is, if the load current and voltage are out of phase with each other, then the primary voltage and current will have the same phase angle.

Many devices or circuits require a specific value of load resistance (or impedance) for optimum operation. The impedance of the actual load that is to dissipate the power may be quite different from the impedance of the source device or circuit, so an *impedance-matching transformer* is used to change the actual load into an impedance of the desired value. The turns ratio required is:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}} \qquad (138)$$

where

N_P / N_S = required turns ratio, primary to secondary,
Z_P = primary impedance required, and
Z_S = impedance of load connected to secondary.

Example: A transistor audio amplifier requires a load of 150 Ω for optimum performance, and is to be connected to a loudspeaker having an impedance of 4.0 Ω. What primary-to-secondary turns ratio is required in the coupling transformer?

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}} = \frac{N_P}{N_S} \sqrt{\frac{150\ \Omega}{4.0\ \Omega}} = \sqrt{38} = 6.2$$

The primary therefore must have 6.2 times as many turns as the secondary.

These relationships may be used in practical circuits even though they are based on an ideal transformer. Aside from the normal design requirements of reasonably low internal losses and low leakage reactance, the only other requirement is that the primary have enough inductance to operate with low magnetizing current at the voltage applied to the primary.

The primary terminal impedance of an iron-core transformer is determined wholly by the load connected to the secondary and by the turns ratio. If the characteristics of the transformer have an appreciable effect on the impedance presented to the power source, the transformer is either poorly designed or is not suited to the voltage and frequency at which it is being used. Most transformers will operate quite well at voltages from slightly above to well below the design figure.

TRANSFORMER LOSSES

In practice, none of the formulas given so far provides truly exact results, although they afford reasonable approximations. Transformers in reality are not simply two coupled inductors, but a network of resistances and reactances, most of which appear in **Fig 2.95**. Since only the terminals numbered 1 through 4 are accessible to the user, transformer ratings and specifications take into account the additional losses created by these complexities.

In a practical transformer not all of the magnetic flux is common to both windings, although in well-designed transformers the amount of flux that cuts one winding and not the other is only a small percentage of the total flux. This *leakage flux* causes a voltage by self-induction in the winding creating the flux. The effect is the same as if a small *leakage inductance* existed independently of the main windings. Leakage inductance acts in exactly the same way as inductance inserted in series with the winding. It has, therefore, a certain reactance, depending on the amount

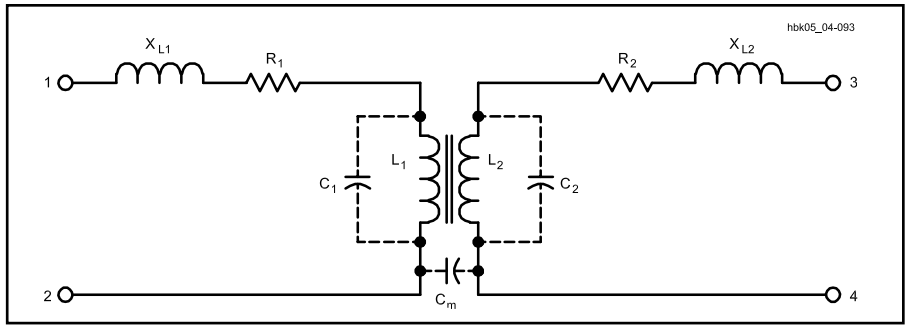


Fig 2.95 — A transformer as a network of resistances, inductances and capacitances. Only L1 and L2 contribute to the transfer of energy.

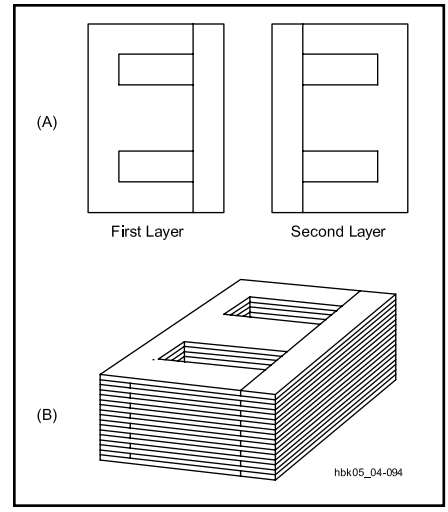


Fig 2.96 — A typical transformer iron core. The E and I pieces alternate direction in successive layers to improve the magnetic path while attenuating eddy currents in the core.

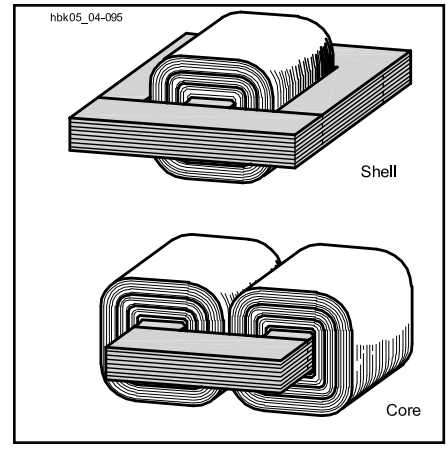


Fig 2.97 — Two common transformer constructions: shell and core.

of leakage inductance and the frequency. This reactance is called *leakage reactance* and is shown as X_{L1} and X_{L2} in Fig 2.95.

Current flowing through the leakage reactance causes a voltage drop. This voltage drop increases with increasing current; hence, it increases as more power is taken from the secondary. The resistances of the transformer windings, R1 and R2, also cause voltage drops when there is current. Although these voltage drops are not in phase with those caused by leakage reactance, together they result in a lower secondary voltage under load than is indicated by the transformer turns ratio. Thus, in a practical transformer, the greater the secondary current, the smaller the secondary terminal voltage becomes.

At ac line frequencies (50 or 60 Hz), the voltage at the secondary, with a reasonably well-designed iron-core transformer, should not drop more than about 10% from open-circuit conditions to full load. The voltage drop may be considerably more than this in a transformer operating at audio frequencies, because the leakage reactance increases with frequency.

In addition to wire resistances and leakage reactances, certain unwanted or "stray" capacitances occur in transformers. The wire forming the separate turns of the windings acts as the plates of a small capacitor, creating a capacitance between turns and between the windings. This *distributed capacitance* appears in Fig 2.95 as C1, C2, and C_M. Moreover, transformer windings can exhibit capacitance relative to nearby metal, for example, the chassis, the shield and even the core. When current flows through a winding, each turn has a slightly different voltage than its adjacent turns. This voltage causes a small current to flow in these *interwinding* and *winding-to-winding capacitances*.

Although stray capacitances are of little concern with power and audio transformers, they become important as the frequency increases. In transformers for RF use, the stray capacitance can resonate with either the leakage reactance or, at lower frequencies, with the winding reactances, L1 or L2, especially under very light or zero loads. In the frequency region around resonance, transformers no longer exhibit the properties formulated above or the impedance properties to be described below.

Iron-core transformers also experience losses within the core itself. *Hysteresis losses* include the energy required to overcome the retentivity of the core's magnetic material. Circulating currents through the core's resistance are *eddy currents*, which form part of the total core losses. These losses, which add to the required magnetizing current, are equivalent to adding a resistance in parallel with L1 in Fig 2.95.

CORE CONSTRUCTION

Audio and power transformers usually employ silicon steel as the core material. With permeabilities of 5000 or greater, these cores saturate at flux densities approaching 10^5 lines of flux per square inch of cross section. The cores consist of thin insulated laminations to break up potential eddy current paths.

Each core layer consists of an "E" and an "I" piece butted together, as represented in **Fig 2.96**. The butt point leaves a small gap. Each layer is reversed from the adjacent layers so that each gap is next to a continuous magnetic path so that the effect of the gaps is minimized. This is different from the air-gapped inductor of Fig 2.49 in which the air gap is maintained for all layers of laminations.

Two core shapes are in common use, as shown in **Fig 2.97**. In the shell type, both windings are placed on the inner leg, while in the core type the primary and secondary windings may be placed on separate legs, if desired. This is sometimes done when it is necessary to minimize capacitance between the primary and secondary, or when one of the windings must operate at very high voltage.

The number of turns required in the primary for a given applied voltage is determined by the size, shape and type of core material used, as well as the frequency. The number of turns required is inversely proportional to the cross-sectional area of the core. As a rough indication, windings of small power transformers frequently have about six to eight turns per volt on a core of 1-square-inch cross section and have a magnetic path 10 or 12 inches in length. A longer path or smaller cross section requires more turns per volt, and vice versa.

In most transformers the windings are wound in layers, with a thin sheet of treated-paper insulation between each layer. Thicker insulation is used between adjacent windings and between the first winding and the core.

SHIELDING

Because magnetic lines of force are continuous loops, shielding requires a complete path for the lines of force of the leakage flux. The high-permeability of iron cores tends to concentrate the field, but additional shielding is often needed. As depicted in **Fig 2.98**,

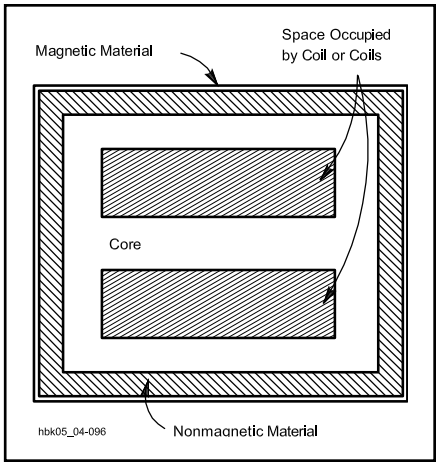


Fig 2.98 — A shielded transformer cross-section: the core plus an outer shield of magnetic material contain nearly all of the magnetic field.

enclosing the transformer in a good magnetic material can restrict virtually all of the magnetic field in the outer case. The nonmagnetic material between the case and the core creates a region of high reluctance, attenuating the field before it reaches the case.

2.14.2 Autotransformers

The transformer principle can be used with only one winding instead of two, as shown in **Fig 2.99A**. The principles that relate voltage, current and impedance to the turns ratio also apply equally well. A one-winding transformer is called an *auto-transformer*. The current in the common section (A) of the winding is the difference between the line (primary) and the load (secondary) currents, since these currents are out of phase. Hence, if the line and load currents are nearly equal, the common section of the winding may be wound with comparatively small wire. The line and load currents will be equal only when the primary (line) and secondary (load) voltages are not very different.

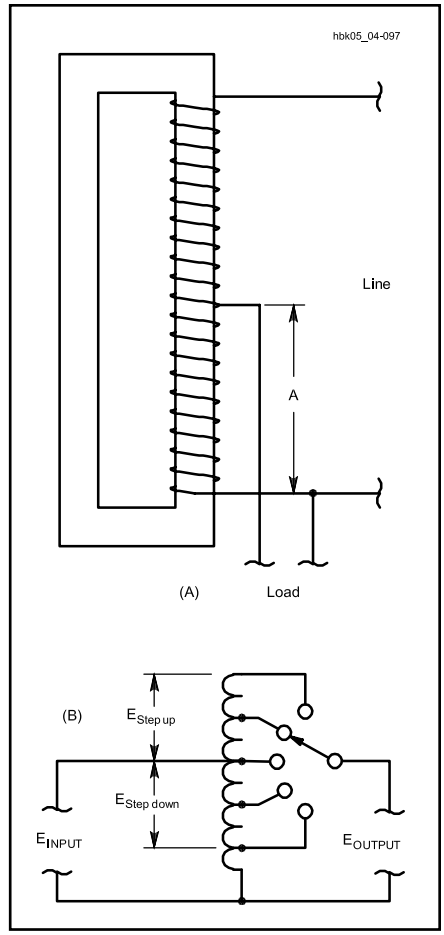


Fig 2.99 — The autotransformer is based on the transformer, but uses only one winding. The pictorial diagram at A shows the typical construction of an autotransformer. The schematic diagram at B demonstrates the use of an autotransformer to step up or step down ac voltage, usually to compensate for excessive or deficient line voltage.

Autotransformers are used chiefly for boosting or reducing the power-line voltage by relatively small amounts. Fig 2.99B illustrates the principle schematically with a switched, stepped autotransformer. Continuously variable autotransformers are commercially available under a variety of trade names; Variac and Powerstat are typical examples.

Technically, tapped air-core inductors, such as the one in the network in Fig 2.90 at the close of the discussion of resonant circuits, are also autotransformers. The voltage from the tap to the bottom of the winding is less than the voltage across the entire winding. Likewise, the impedance of the tapped part of the winding is less than the impedance of the entire winding. Because in this case, leakage reactances are great and the coefficient of coupling is quite low, the relationships that are true in a perfect transformer grow quite unreliable in predicting the exact values. For this reason, tapped inductors are rarely referred to as transformers. The stepped-down situation in Fig 2.90 is better approximated — at or close to resonance — by the formula

$$R_P = \frac{R_L X_{COM}^2}{X_L} \qquad (139)$$

where

R_P = tuned-circuit parallel-resonant impedance,

R_L = load resistance tapped across part of the winding,

X_{COM} = reactance of the portion of the winding common to both the resonant circuit and the load tap, and

X_L = reactance of the entire winding.

The result is approximate and applies only to circuits with a Q of 10 or greater.

2.15 Heat Management

While not strictly an electrical fundamental, managing the heat generated by electronic circuits is important in nearly all types of radio equipment. Thus, the topic is included in this chapter. Information on the devices and circuits discussed in this section may be found in other chapters.

Any actual energized circuit consumes electric power because any such circuit contains components that convert electricity into other forms of energy. This dissipated power appears in many forms. For example, a loudspeaker converts electrical energy into sound, the motion of air molecules. An antenna (or a light bulb) converts electricity into electromagnetic radiation. Charging a battery converts electrical energy into chemical energy (which is then converted back to electrical energy upon discharge). But the most common transformation by far is the conversion, through some form of resistance, of electricity into heat.

Sometimes the power lost to heat serves a useful purpose — toasters and hair dryers come to mind. But most of the time, this heat represents a power loss that is to be minimized wherever possible or at least taken into account. Since all real circuits contain resistance, even those circuits (such as a loudspeaker) whose primary purpose is to convert electricity to some *other* form of energy also convert some part of their input power to heat. Often, such losses are negligible, but sometimes they are not.

If unintended heat generation becomes significant, the involved components will get warm. Problems arise when the temperature increase affects circuit operation by either

- causing the component to fail, by explosion, melting, or other catastrophic event, or, more subtly,
- causing a slight change in the properties of the component, such as through a temperature coefficient (TC).

In the first case, we can design conservatively, ensuring that components are rated to safely handle two, three or more times the maximum power we expect them to dissipate.

In the second case, we can specify components with low TCs, or we can design the circuit to minimize the effect of any one component. Occasionally we even exploit temperature effects (for example, using a resistor, capacitor or diode as a temperature sensor). Let's look more closely at the two main categories of thermal effects.

Not surprisingly, heat dissipation (more correctly, the efficient removal of generated heat) becomes important in medium- to high-power circuits: power supplies, transmitting circuits and so on. While these are not the only examples where elevated temperatures and related failures are of concern, the techniques we will discuss here are applicable to all circuits.

2.15.1 Thermal Resistance

The transfer of heat energy, and thus the change in temperature, between two ends of a block of material is governed by the following heat flow equation and illustrated in **Fig 2.100**):

$$P = \frac{kA}{L}\Delta T = \frac{\Delta T}{\theta} \qquad (140)$$

where

P = power (in the form of heat) conducted between the two points

k = *thermal conductivity*, measured in W/(m °C), of the material between the two points, which may be steel, silicon, copper, PC board material and so on

L = length of the block

A = area of the block, and

ΔT = *difference* in temperature between the two points.

Thermal conductivities of various common materials at room temperature are given in **Table 2.11**.

The form of equation 140 is the same as the variation of Ohm's Law relating current flow to the ratio of a difference in potential to resistance; I = E/R. In this case, what's flowing is heat (P), the difference in potential is a temperature difference (ΔT), and what's resisting the flow of heat is the *thermal resistance*;

$$\theta = \frac{L}{kA} \qquad (141)$$

with units of °C/W. (The units of resistance are equivalent to V/A.) The analogy is so apt that the same principles and methods apply to heat flow problems as circuit problems. The following correspondences hold:

• Thermal conductivity W/(m °C) ↔ Electrical conductivity (S/m).

• Thermal resistance (°C/W) ↔ Electrical resistance (Ω).

• Thermal current (heat flow) (W) ↔ Electrical current (A).

Table 2.11

Thermal Conductivities of Various Materials

Gases at 0 °C, Others at 25 °C; from *Physics,* by Halliday and Resnick, 3rd Ed.

Material	*k in units of W/m°C*
Aluminum	200
Brass	110
Copper	390
Lead	35
Silver	410
Steel	46
Silicon	150
Air	0.024
Glass	0.8
Wood	0.08

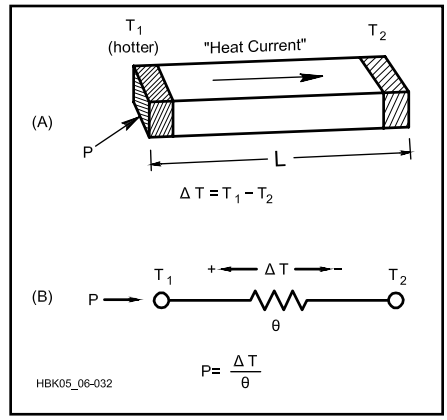


Fig 2.100 — Physical and "circuit" models for the heat-flow equation.

• Thermal potential (T) ↔ Electrical potential (V).

• Heat source ↔ Power source.

For example, calculate the temperature of a 2-inch (0.05 m) long piece of #12 copper wire at the end that is being heated by a 25 W (input power) soldering iron, and whose other end is clamped to a large metal vise (assumed to be an infinite heat sink), if the ambient temperature is 25 °C (77 °F).

First, calculate the thermal resistance of the copper wire (diameter of #12 wire is 2.052 mm, cross-sectional area is 3.31×10^{-6} m^2)

$$\theta = \frac{L}{k\,A} = \frac{(0.05\ \text{m})}{(390\ \text{W}/(\text{m}\ ^\circ\text{C})) \times (3.31 \times 10^{-6}\ \text{m}^2)}$$

$$= 38.7\ ^\circ\text{C/W}$$

Then, rearranging the heat flow equation above yields (after assuming the heat energy actually transferred to the wire is around 10 W)

$$\Delta T = P\,\theta = (10\ \text{W}) \times (38.7\ ^\circ\text{C}/\text{W}) = 387\ ^\circ\text{C}$$

So the wire temperature at the hot end is 25 C + ΔT = 412 °C (or 774 °F). If this sounds a little high, remember that this is for the steady state condition, where you've been holding the iron to the wire for a long time.

From this example, you can see that things can get very hot even with application of moderate power levels. For this reason, circuits that generate sufficient heat to alter, not necessarily damage, the components must employ some method of cooling, either active or passive. Passive methods include heat sinks or careful component layout for good ventilation. Active methods include forced air (fans) or some sort of liquid cooling (in some high-power transmitters).

2.15.2 Heat Sink Selection and Use

The purpose of a heat sink is to provide a high-power component with a large surface area through which to dissipate heat. To use the models above, it provides a low thermal-resistance path to a cooler temperature, thus allowing the hot component to conduct a large "thermal current" away from itself.

Power supplies probably represent one of the most common high-power circuits amateurs are likely to encounter. Everyone has certainly noticed that power supplies get warm or even hot if not ventilated properly. Performing the thermal design for a properly cooled power supply is a very well-defined process and is a good illustration of heat-flow concepts.

This material was originally prepared by ARRL Technical Advisor Dick Jansson, KD1K, during the design of a 28-V, 10-A power supply. (The chapter **Analog Basics** discusses semiconductor diodes, rectifiers, and transistors. The **Power Supplies** chapter has more information on power supply design.) An outline of the design procedure shows the logic applied:

1. Determine the expected power dissipation (P_{in}).

2. Identify the requirements for the dissipating elements (maximum component temperature).

3. Estimate heat-sink requirements.

4. Rework the electronic device (if necessary) to meet the thermal requirements.

5. Select the heat exchanger (from heat sink data sheets).

The first step is to estimate the filtered, unregulated supply voltage under full load. Since the transformer secondary output is 32 V ac (RMS) and feeds a full-wave bridge rectifier, let's estimate 40 V as the filtered dc output at a 10-A load.

The next step is to determine our critical components and estimate their power dissipations. In a regulated power supply, the pass transistors are responsible for nearly all the power lost to heat. Under full load, and allowing for some small voltage drops in the

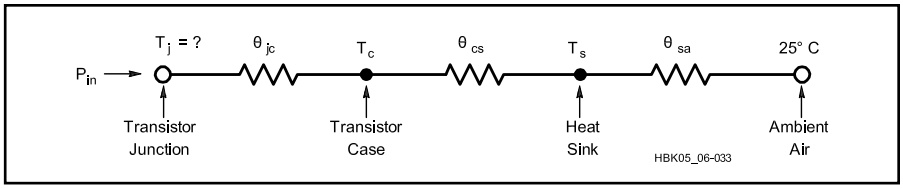


Fig 2.101 — Resistive model of thermal conduction in a power transistor and associated heat sink. See text for calculations.

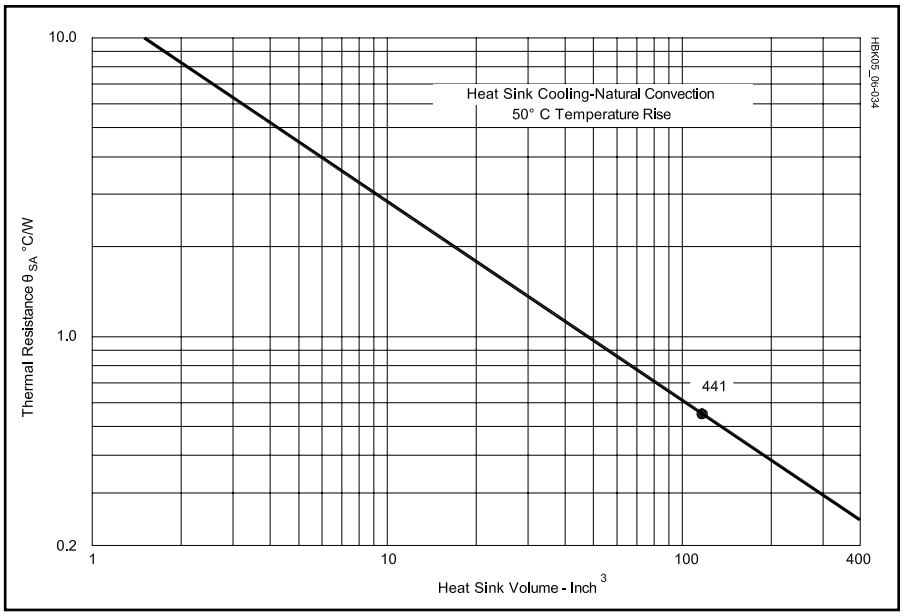


Fig 2.102 — Thermal resistance vs heat-sink volume for natural convection cooling and 50 °C temperature rise. The graph is based on engineering data from Wakefield Thermal Solutions, Inc.

power-transistor emitter circuitry, the output of the series pass transistors is about 29 V for a delivered 28 V under a 10-A load. With an unregulated input voltage of 40 V, the total energy heat dissipated in the pass transistors is (40 V – 29 V) × 10 A = 110 W. The heat sink for this power supply must be able to handle that amount of dissipation and still keep the transistor junctions below the specified safe operating temperature limits. It is a good rule of thumb to select a transistor that has a maximum power dissipation of twice the desired output power.

Now, consider the ratings of the pass transistors to be used. This supply calls for 2N3055s as pass transistors. The data sheet shows that a 2N3055 is rated for 15-A service and 115-W dissipation. But the design uses *four* in parallel. Why? Here we must look past the big, bold type at the top of the data sheet to such subtle characteristics as the junction-to-case thermal resistance, q_{jc}, and the maximum allowable junction temperature, T_j.

The 2N3055 data sheet shows θ_{jc} = 1.52 °C/W, and a maximum allowable case (and junction) temperature of 220 °C. While it seems that one 2N3055 could barely, on paper at least, handle the electrical requirements — at what temperature would it operate?

To answer that, we must model the entire "thermal circuit" of operation, starting with the transistor junction on one end and ending at some point with the ambient air. A reasonable model is shown in **Fig 2.101**. The ambient air is considered here as an infinite heat sink; that is, its temperature is assumed to be a constant 25 °C (77 °F). θ_{jc} is the thermal resistance from the transistor junction to its case. θ_{cs} is the resistance of the mounting interface between the transistor case and the heat sink. θ_{sa} is the thermal resistance between the heat sink and the ambient air. In this "circuit," the generation of heat (the "thermal current source") occurs in the transistor at P_{in}.

Proper mounting of most TO-3 package power transistors such as the 2N3055 requires that they have an electrical insulator between the transistor case and the heat sink. However, this electrical insulator must at the same time exhibit a low thermal resistance. To achieve a quality mounting, use thin polyimid or mica formed washers and a suitable thermal compound to exclude air from the interstitial space. "Thermal greases" are commonly available for this function. Any silicone grease may be used, but filled silicone oils made specifically for this purpose are better.

Using such techniques, a conservatively high value for θ_{cs} is 0.50 °C/W. Lower values are possible, but the techniques needed to achieve them are expensive and not generally available to the average amateur. Furthermore, this value of θ_{cs} is already much lower than θ_{jc}, which cannot be lowered without going to a somewhat more exotic pass transistor.

Finally, we need an estimate of θ_{sa}. **Fig 2.102** shows the relationship of heat-sink volume to thermal resistance for natural-convection cooling. This relationship presumes the use of suitably spaced fins (0.35 inch or greater) and provides a "rough order-of-magnitude" value for sizing a heat sink. For a first calculation, let's assume a heat sink of roughly 6 × 4 × 2 inch (48 cubic inches). From Fig 2.102, this yields a θ_{sa} of about 1 °C/W.

Returning to Fig 2.101, we can now calculate the approximate temperature increase of a single 2N3055:

$$\delta T = P\,\theta_{total}$$

$$= 110\text{ W} \times (1.52\text{ °C/W} + 0.5\text{ °C/W} + 1.0\text{ °C/W})$$

$$= 332\text{ °C}$$

Given the ambient temperature of 25 °C, this puts the junction temperature T_j of the 2N3055 at 25 + 332 = 357 °C! This is clearly too high, so let's work backward from the air end and calculate just how many transistors we need to handle the heat.

First, putting more 2N3055s in parallel means that we will have the thermal model illustrated in **Fig 2.103**, with several identical θ_{jc} and θ_{cs} in parallel, all funneled through the same θ_{as} (we have one heat sink).

Keeping in mind the physical size of the project, we could comfortably fit a heat sink of approximately 120 cubic inches (6 × 5 × 4 inches), well within the range of commercially available heat sinks. Furthermore, this application can use a heat sink where only "wire access" to the transistor connections is required. This allows the selection of a more efficient design. In contrast, RF designs require the transistor mounting surface to be completely exposed so that the PC board can be mounted close to the transistors to minimize parasitics. Looking at Fig 2.102, we see that a 120-cubic-inch heat sink yields a θ_{sa} of 0.55 °C/W. This means that the temperature of the heat sink when dissipating 110 W will be 25 °C + (110 W × 0.55 °C/W) = 85.5 °C.

Industrial experience has shown that silicon transistors suffer substantial failure when junctions are operated at highly elevated temperatures. Most commercial and military specifications will usually not permit design junction temperatures to exceed 125 °C. To arrive at a safe figure for our maximum

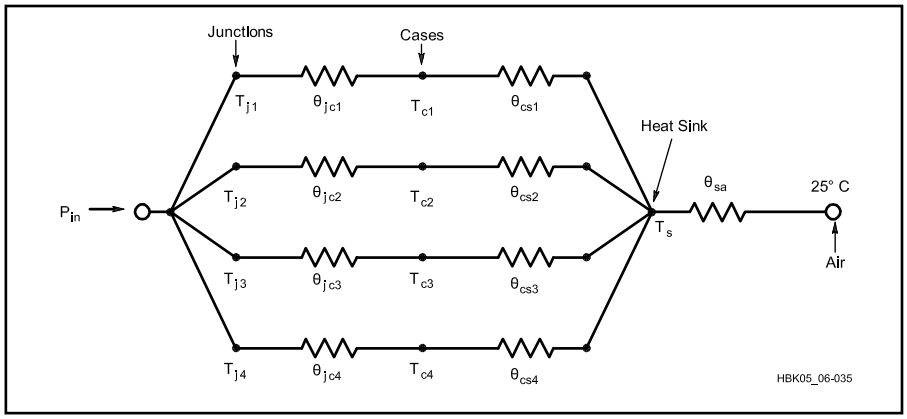


Fig 2.103 — Thermal model for multiple power transistors mounted on a common heat sink.

allowed T_j, we must consider the intended use of the power supply. If we are using it in a 100% duty-cycle transmitting application such as RTTY or FM, the circuit will be dissipating 110 W continuously. For a lighter duty-cycle load such as CW or SSB, the "key-down" temperature can be slightly higher as long as the average is less than 125 °C. In this intermittent type of service, a good conservative figure to use is T_j = 150 °C.

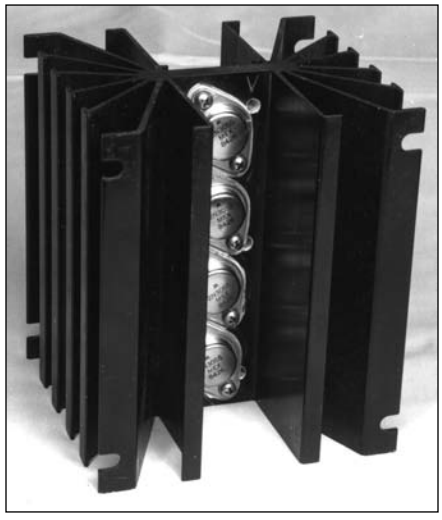
Given this scenario, the temperature rise across each transistor can be 150 – 85.5 = 64.5 °C. Now, referencing Fig 2.103, remembering the total θ for each 2N3055 is 1.52 + 0.5 = 2.02 °C/W, we can calculate the maximum power each 2N3055 can safely dissipate:

$$P = \frac{\delta T}{\theta} = \frac{64.5\ °C}{2.02\ °C/W} = 31.9\ W$$

Thus, for 110 W full load, we need four 2N3055s to meet the thermal requirements of the design. Now comes the big question: What is the "right" heat sink to use? We have already established its requirements: it must be capable of dissipating 110 W, and have a θ_{sa} of 0.55 °C/W (see above).

A quick consultation with several manufacturer's catalogs reveals that Wakefield Thermal Solutions, Inc. model nos. 441 and 435 heat sinks meet the needs of this application. A Thermalloy model no. 6441 is suitable as well. Data published in the catalogs of these manufacturers show that in natural-convection service, the expected temperature rise for 100 W dissipation would be just under 60 °C, an almost perfect fit for this application. Moreover, the no. 441 heat sink can easily mount four TO-3-style 2N3055 transistors as shown in **Fig 2.104**. Remember: heat sinks should be mounted with the fins and transistor mounting area vertical to promote convection cooling.

The design procedure just described is applicable to any circuit where heat buildup is a potential problem. By using the thermal-resistance model, we can easily calculate whether or not an external means of cooling is necessary, and if so, how to choose it. Aside from heat sinks, forced air cooling (fans) is another common method. In commercial transceivers, heat sinks with forced-air cooling are common.



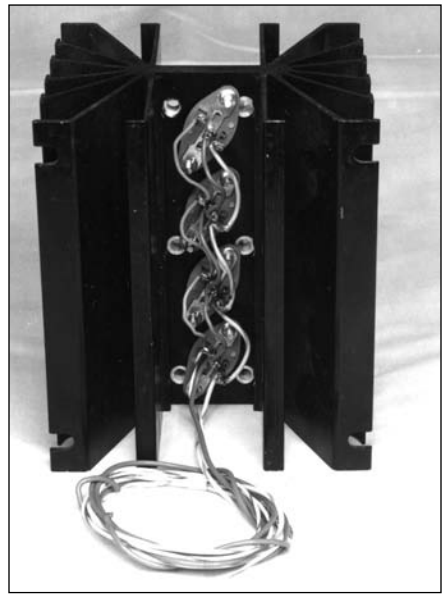


Fig 2.104 — A Wakefield 441 heat sink with four 2N3055 transistors mounted.

2.15.3 Transistor Derating

Maximum ratings for power transistors are usually based on a case temperature of 25 °C. These ratings will decrease with increasing operating temperature. Manufacturer's data sheets usually specify a *derating* figure or curve that indicates how the maximum ratings change per degree rise in temperature. If such information is not available (or even if it is!), it is a good rule of thumb to select a power transistor with a maximum power dissipation of at least twice the desired output power.

2.15.4 Rectifiers

Diodes are physically quite small, and they operate at high current densities. As a result their heat-handling capabilities are somewhat limited. Normally, this is not a problem in high-voltage, low-current supplies in which rectifiers in axial-lead DO-type packages are used. (See the **Component Data and References** chapter for information on device packages.) The use of high-current (2 A or greater) rectifiers at or near their maximum ratings, however, requires some form of heat sinking. The average power dissipated by a rectifier is

$$P = I_{AVG} \times V_F \qquad (142)$$

where

I_{AVG} is the average current, and
V_F is the forward voltage drop.

Average current must account for the conduction duty cycle and the forward voltage drop must be determined at the average current level.

Rectifiers intended for such high-current applications are available in a variety of packages suitable for mounting to flat surfaces. Frequently, mounting the rectifier on the main chassis (directly, or with thin mica insulating washers) will suffice. If the diode is insulated from the chassis, thin layers of thermal compound or thermal insulating washers should be used to ensure good heat conduction. Large, high-current rectifiers often require special heat sinks to maintain a safe operating temperature. Forced-air cooling is sometimes used as a further aid.

2.15.5 RF Heating

RF current often causes component heating problems where the same level of dc current may not. An example is the tank circuit of

an RF oscillator. If several small capacitors are connected in parallel to achieve a desired capacitance, skin effect will be reduced and the total surface area available for heat dissipation will be increased, thus significantly reducing the RF heating effects as compared to a single large capacitor. This technique can be applied to any similar situation; the general idea is to divide the heating among as many components as possible.

2.15.6 Forced-Air and Water Cooling

In Amateur Radio today, forced-air cooling is most commonly found in vacuum-tube circuits or in power supplies built in small enclosures, such as those in solid-state transceivers or computers. Fans or blowers are commonly specified in cubic feet per minute (CFM). While the nomenclature and specifications differ from those used for heat sinks, the idea remains the same: to offer a low thermal resistance between the inside of the enclosure and the (ambient) exterior.

For forced air cooling, we basically use the "one resistor" thermal model of Fig 2.100. The important quantity to be determined is heat generation, P_{in}. For a power supply, this can be easily estimated as the difference between the input power, measured at the transformer primary, and the output power at full load. For variable-voltage supplies, the worst-case output condition is minimum voltage with maximum current. A discussion of forced-air cooling for vacuum tube equipment appears in the **RF Power Amplifiers** chapter.

Dust build-up is a common problem for forced-air cooling systems, even with powerful blowers and fans. If air intake grills and vents are not kept clean and free of lint and debris, air flow can be significantly reduced, leading to excessive equipment temperature and premature failure. Cleaning of air passageways should be included in regular equipment maintenance for good performance and maximum equipment life.

Water cooling systems are much less common in amateur equipment, used primarily for high duty cycle operating, such as RTTY, and at frequencies where the efficiency of the amplifier is relatively low, such as UHF and microwaves. In these situations, water cooling is used because water can absorb and transfer more than 3000 times as much heat as the same volume of air!

The main disadvantage of water cooling is that it requires pumps, hoses, and reservoirs whereas a fan or blower is all that is required for forced-air cooling. For high-voltage circuits, using water cooling also requires special insulation techniques and materials to allow water to circulate in close contact with the heat source while remaining electrically isolated.

Nevertheless, the technique can be effective. The increased availability of inexpensive materials designed for home sprinkler and other low-pressure water distribution systems make water-cooling less difficult to implement. It is recommended that the interested reader review articles and projects in the amateur literature to observe the successful implementation of water cooling systems.

2.15.7 Heat Pipe Cooling

A heat pipe is a device containing a *working fluid* in a reservoir where heat is absorbed and a channel to a second reservoir where heat is dissipated. Heat pipes work by *evaporative cooling*. The heat-absorbing reservoir is placed in thermal contact with the heat source which transfers heat to the working fluid, usually a liquid substance with a boiling point just above room temperature. The working fluid vaporizes and the resulting vapor pressure pushes the hot vapor through the channel to the cooling reservoir.

In the cooling reservoir, the working fluid gives up its heat of vaporization, returning to the fluid state. The cooled fluid then flows back through the channel to the heat-absorbing reservoir where the process is repeated.

Heat pipes require no fans or pumps — movement of the working fluid is driven entirely from the temperature difference between the two reservoirs. The higher the temperature difference between the absorbing and dissipating reservoirs, the more effective the heat pump becomes, up to the limit of the dissipating reservoir to dissipate heat.

The principle application for heat pipes is for space operations. Heat pipe applications in terrestrial applications are limited due to the gravity gradient sensitivity of heat pipes. At present, the only amateur equipment making use of heat pipes are computers and certain amplifier modules. Nevertheless, as more general-purpose products become available, this technique will become more common.

2.15.8 Thermoelectric Cooling

Thermoelectric cooling makes use of the *Peltier effect* to create heat flow across the junction of two different types of materials. This process is related to the *thermoelectric effect* by which thermocouples generate voltages based on the temperature of a similar junction. A *thermoelectric cooler* or *TEC* (also known as a *Peltier cooler*) requires only a source of dc power to cause one side of the device to cool and the other side to warm. TECs are available with different sizes and power ratings for different applications.

TECs are not available with sufficient heat transfer capabilities that they can be used in high-power applications, such as RF amplifiers. However, they can be useful in lowering the temperature of sensitive receiver circuits, such as preamplifiers used at UHF and microwave frequencies, or imaging devices, such as charge-coupled devices (CCDs). Satellites use TECs as *radiative coolers* that dissipate heat directly as thermal or infrared radiation. TECs are also found in some computing equipment where they are used to remove heat from microprocessors and other large integrated circuits.

2.15.9 Temperature Compensation

Aside from catastrophic failure, temperature changes may also adversely affect circuits if the temperature coefficient (TC) of one or more components is too large. If the resultant change is not too critical, adequate temperature stability can often be achieved simply by using higher-precision components with low TCs (such as NP0/C0G capacitors or metal-film resistors). For applications where this is impractical or impossible (such as many solid-state circuits), we can minimize temperature sensitivity by *compensation* or *matching* — using temperature coefficients to our advantage.

Compensation is accomplished in one of two ways. If we wish to keep a certain circuit quantity constant, we can interconnect pairs of components that have equal but opposite TCs. For example, a resistor with a negative TC can be placed in series with a positive TC resistor to keep the total resistance constant. Conversely, if the important point is to keep the *difference* between two quantities constant, we can use components with the *same* TC so that the pair "tracks." That is, they both change by the same amount with temperature.

An example of this is a Zener reference circuit. Since a diode is strongly affected by operating temperature, circuits that use diodes or transistors to generate stable reference voltages must use some form of temperature compensation. Since, for a constant current, a reverse-biased PN junction has a negative voltage TC while a forward-biased junction has a positive voltage TC, a good way to temperature-compensate a Zener reference diode is to place one or more forward-biased diodes in series with it.

2.15.10 Thermistors

Thermistors can be used to control temperature and improve circuit behavior or protect against excessive temperatures, hot or cold. Circuit temperature variations can affect gain, distortion or control functions like receiver AGC or transmitter ALC. Thermistors can be used in circuits that compensate for temperature changes, such as the simplified AGC system shown in **Fig 2.105** where temperature-sensitive resistors ($R_{th1} - R_{th3}$) are used to counteract the thermal changes of other resistors.

In this discussion, William Sabin, WØIYH describes how a simple temperature control circuit can be used to protect a power transistor, such as a MOSFET in an RF power amplifier, or produce a temperature-stable heater to regulate the temperature of a VFO to reduce oscillator drift.

A *thermistor* is a small bit of intrinsic (no N or P doping) metal-oxide semiconductor compound material between two wire leads. As temperature increases, the number of liberated hole/electron pairs increases exponentially, causing the resistance to decrease exponentially. You can see this in the resistance equation:

$$R(T) = R(T0)e^{-\beta(1/T0-1/T)} \qquad (143)$$

where T is some temperature in Kelvins and T0 is a reference temperature, usually 298 K (25°C), at which the manufacturer specifies R(T0).

The constant β is experimentally determined by measuring resistance at various temperatures and finding the value of β that best agrees with the measurements. A simple way to get an approximate value of β is to make two measurements, one at room temperature, say 25 °C (298 K) and one at 100 °C (373 K) in boiling water. Suppose the resistances are 10 kΩ and 938 Ω. Eq 1 is solved for β

$$\beta = \frac{\ln\left(\frac{R(T)}{R(T0)}\right)}{\frac{1}{T}-\frac{1}{T0}} = \frac{\ln\left(\frac{938}{1000}\right)}{\frac{1}{373}-\frac{1}{298}} = 3507 \qquad (144)$$

With the behavior of the thermistor known — either by equation or calibration table — its change in resistance can be used to create an electronic circuit whose behavior is controlled by temperature in a known fashion. Such a circuit can be used for controlling temperature or detecting specific temperatures.

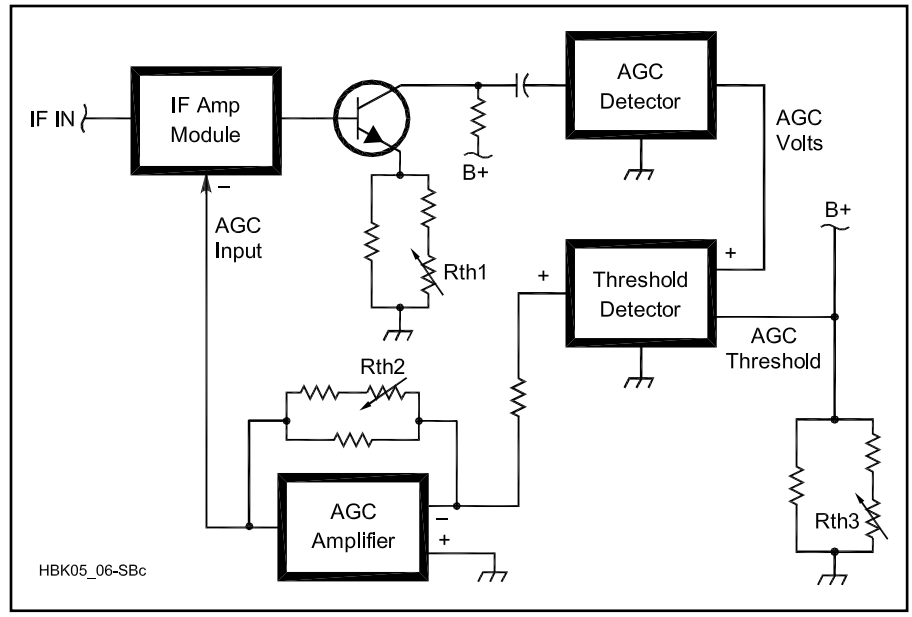


Fig 2.105 — Temperature compensation of IF amplifier and AGC circuitry. R_{th1} – R_{th3} are temperature-sensitive resistors that counteract thermal drift in gain and thresholds.

THERMISTOR-BASED TEMPERATURE CONTROLLER

For many electronics applications, the exact value of temperature is not as important as the ability to maintain that temperature. The following project shows how to use a thermistor to create a circuit that acts as a temperature-controlled switch. The circuit is then used to detect over-temperature conditions and to act as a simple on/off temperature controller.

The circuit of **Fig 2.106** uses a thermistor (R1) to detect a case temperature of about 93 °C, with an on/off operating range of about 0.3°C. A red LED warns of an over-temperature condition that requires attention. The LM339 comparator toggles when R_t = R5. Resistors R3, R4, R5 and the thermistor form a Wheatstone bridge. R3 and R4 are

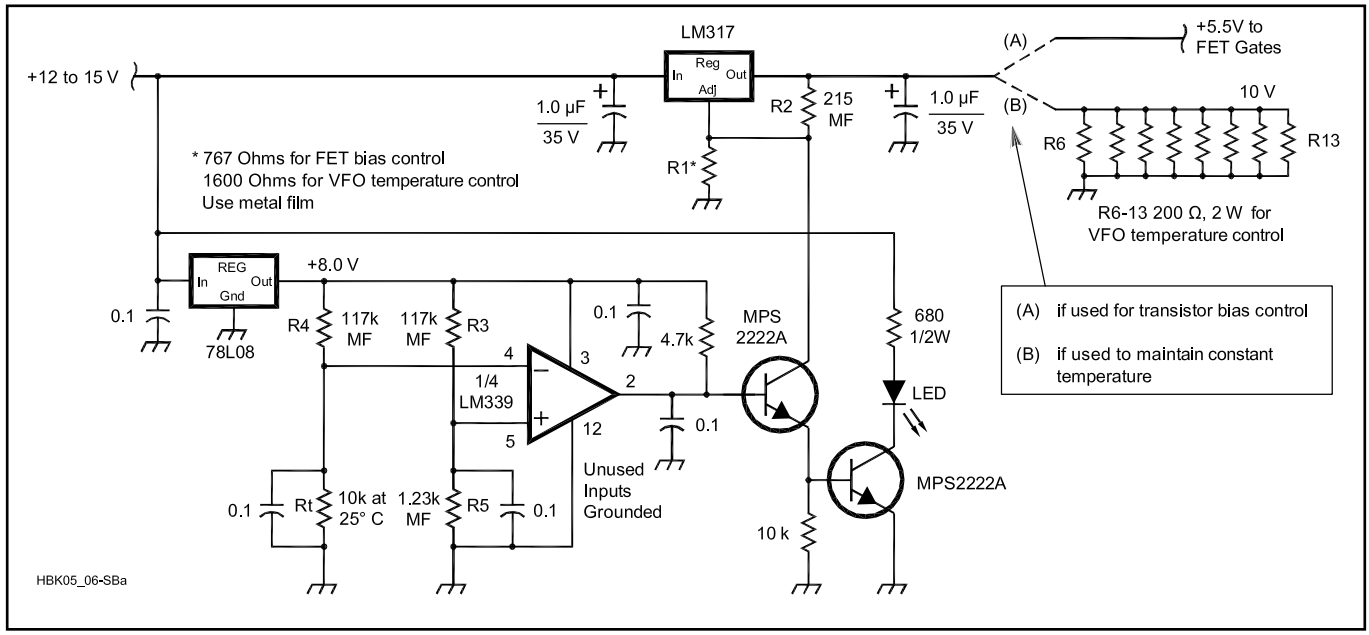


Fig 2.106 — Temperature controller for a power transistor (option A) or VFO (option B).

chosen to make the inputs to the LM339 about 0.5 V at the desired temperature. This greatly reduces self-heating of the thermistor, which could cause a substantial error in the circuit behavior.

The RadioShack Precision Thermistor (RS part number 271-110) used in this circuit is rated at 10 kΩ ±1% at 25 °C. It comes with a calibration chart from –50 °C to +110 °C that you can use to get an approximate resistance at any temperature in that range. For many temperature protection applications you don't have to know the exact temperature. You won't need an exact calibration, but you can verify that the thermistor is working properly by measuring its resistance at 20 °C (68 °F) and in boiling water (≈ 100 °C).

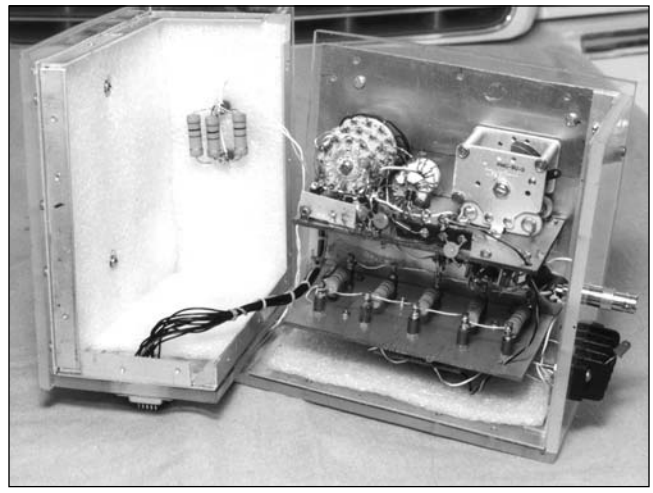


Fig 2.107 — Temperature-controlled VFO cabinet. ¼-inch plexiglass on the outside and ¼-inch thick insulating foam panels on the inside are used to improve temperature stability.

MOSFET POWER TRANSISTOR PROTECTOR

It is very desirable to compensate the temperature sensitivity of power transistors. In bipolar (BJT) transistors, thermal runaway occurs because the dc current gain increases as the transistors get hotter. In MOSFET transistors, thermal runaway is less likely to occur but with excessive drain dissipation or inadequate cooling, the junction temperature may increase until its maximum allowable value is exceeded.

Suppose you want to protect a power amplifier that uses a pair of MRF150 MOSFETs. You can place a thermistor on the heat sink close to the transistors so that the bias adjustment tracks the flange temperature. Another good idea is to attach a thermistor to the ceramic case of the FET with a small drop of epoxy. That way it will respond more quickly to a sudden temperature increase, possibly saving the transistor from destruction. Use the following procedure to get the desired temperature control:

The MRF150 MOSFET has a maximum allowed junction temperature of 200 °C. The thermal resistance θ_{JC} from junction to case is 0.6 °C per watt. The maximum expected dissipation of the FET in normal operation is 110 W. Select a case temperature of 93 °C. This makes the junction temperature $93\,°C + (0.6\,°C/W) \times (110\,W) = 159\,°C$, which is a safe 41 °C below the maximum allowed temperature.

The FET has a rating of 300 W maximum dissipation at a case temperature of 25 °C, derated at 1.71 W per °C. At 93 °C case temperature the maximum allowed dissipation is $300\,W - 1.71\,W/°C \times (93\,°C - 25\,°C) = 184\,W$. The safety margin at that temperature is 184 – 110 = 74 W.

A very simple way to determine the correct value of R5 is to put the thermistor in 93 °C water (let it stabilize) and adjust R5 so that the circuit toggles. At 93°C, the measured value of the thermistor should be about 1230 Ω.

To use the circuit of Fig 2.106 to control the FET bias voltage, set R1 to be 767 Ω. This value, along with R2, adjusts the LM317 voltage regulator output to 5.5V for the FET gate bias. When the thermistor heats to the set point, the LM339 comparator toggles on. This brings the FET gate voltage to a low level and completely turns off the FETs until the temperature drops about 0.3°C. Use metal film resistors throughout the circuit.

TEMPERATURE CONTROL FOR VFOS

The same circuit can be used to control the temperature of a VFO (variable frequency oscillator) with a few simple changes. Select R1 to be 1600 Ω (metal film) to adjust the LM317 for 10 V. Add resistors R6 through R13 as a heater element. Place the VFO in a thermally insulated enclosure such as shown in **Fig 2.107**. The eight 200-Ω, 2-W, metal-oxide resistors at the output of the LM317 supply about 4 W to maintain a temperature of about 33°C inside the enclosure. The resistors are placed so that heat is distributed uniformly. Five are placed near the bottom and three near the top. The thermistor is mounted in the center of the box, close to the tuned circuit and in physical contact with the oscillator ground-plane surface, using a small drop of epoxy.

Use a massive and well-insulated enclosure to slow the rate of temperature change. In one project, over a 0.1 °C range, the frequency at 5.0 MHz varied up and down ±20 Hz or less, with a period of about five minutes. Superimposed is a very slow drift of average frequency that is due to settling down of components, including possibly the thermistor. These gradual changes became negligible after a few days of "burn-in." One problem that is virtually eliminated by the constant temperature is a small but significant "retrace" effect of the cores and capacitors, and perhaps also the thermistor, where a substantial temperature transient of some kind may take from minutes to hours to recover the previous L and C values.

For a much more detailed discussion of this material, see Sabin W.E., WØIYH "Thermistors in Homebrew Projects," *QEX* Nov/Dec 2000. This article is included (with all *QT*, *QEX* and *NCJ* articles from 2000) on the *2000 ARRL Periodicals CD-ROM*. (Order No. 8209.)

2.16 Radio Mathematics

Understanding radio and electronics beyond an intuitive or verbal level requires the use of some mathematics. None of the math is more advanced than trigonometry or advanced algebra, but if you don't use math regularly, it's quite easy to forget what you learned during your education. In fact, depending on your age and education, you may not have encountered some of these topics at all.

It is well beyond the scope of this book to be a math textbook, but this section touches some of topics that most need explanation for amateurs. For introductory-level tutorials and explanations of advanced topics, a list of on-line, no-cost tutorials is presented in the sidebar, "Online Math Resources". You can browse these tutorials whenever you need them in support of the information in this reference text.

Online Math Resources

The following Web links are a compilation of on-line resources organized by topic. Other resources are available online at **www.arrl.org/tech-prep-resource-library**. Look for the Math Tutorials section.

Many of the tutorials listed below are part of the *Interactive Mathematics* Web site (**www.intmath.com**), a free, online system of tutorials. The system begins with basic number concepts and progresses all the way through introductory calculus. The lessons referenced here are those of most use to a student of radio electronics.

Basic Numbers & Formulas

Order of Operations — **www.intmath.com/Numbers/3_Order-of-operations.php**
Powers, Roots, and Radicals — **www.intmath.com/Numbers/4_Powers-roots-radicals.php**
Introduction to Scientific Notation — **www.ieer.org/clssroom/scinote.html**
Scientific Notation — **www.intmath.com/Numbers/6_Scientific-notation.php**
Ratios and Proportions — **www.intmath.com/Numbers/7_Ratio-proportion.php**
Geometric Formulas — **www.equationsheet.com/sheets/Equations-4.html**
Tables of Conversion Factors — **en.wikipedia.org/wiki/Conversion_of_units**

Metric System

Metric System Overview — **en.wikipedia.org/wiki/Metric_system**
Metric System Tutorial — **www.arrl.org/FandES/tbp/radio-lab/RLH%20Unit%203%20Lesson%20%233.1.doc**

Conversion of Units

Metric-English — **www.m-w.com/mw/table/metricsy.htm**
Metric-English — **en.wikipedia.org/wiki/Metric_yardstick**
Conversion Factors — **oakroadsystems.com/math/convert.htm**

Fractions

Equivalent Fractions — **www.intmath.com/Factoring-fractions/5_Equivalent-fractions.php**
Multiplication and Division— **www.intmath.com/Factoring-fractions/6_Multiplication-division-fractions.php**
Adding and Subtracting — **www.intmath.com/Factoring-fractions/7_Addition-subtraction-fractions.php**
Equations Involving Fractions — **www.intmath.com/Factoring-fractions/8_Equations-involving-fractions.php**
Basic Algebra — **www.intmath.com/Basic-algebra/Basic-algebra-intro.php**

Graphs

Basic Graphs — **www.intmath.com/Functions-and-graphs/Functions-graphs-intro.php**
Polar Coordinates — **www.intmath.com/Plane-analytic-geometry/7_Polar-coordinates.php**
Exponents & Radicals — **www.intmath.com/Exponents-radicals/Exponent-radical.php**
Exponential & Logarithmic Functions — **www.intmath.com/Exponential-logarithmic-functions/Exponential-log-functionsintro.php**

Trigonometry

Basic Trig Functions — **www.intmath.com/Trigonometric-functions/Trig-functions-intro.php**
Graphs of Trig Functions — **www.intmath.com/Trigonometric-graphs/Trigo-graph-intro.php**

Miscellaneous

Complex Numbers — **www.intmath.com/Complex-numbers/imaginary-numbers-intro.php**

2.16.1 Working With Decibels

The *decibel* (*dB*) is a way of expressing a ratio *logarithmically*, meaning as a power of some *base* number, such as 10 or the base for *natural logarithms*, the number e ≈ 2.71828. A decibel, using the metric prefix "deci" (d) for one-tenth, is one-tenth of a *Bel* (B), a unit used in acoustics representing a ratio of 10, and named for the telephone's inventor, Alexander Graham Bell. As it turns out, the Bel was too large a ratio for common use and the deci-Bel, or decibel, became the standard unit.

In radio, the logarithmic ratio is used, but it compares signal strengths — power, voltage, or current. The decibel is more useful than a *linear* ratio because it can represent a wider scale of ratios. The numeric values encountered in radio span a very wide range and so the dB is more suited to discuss large ratios. For example, a typical receiver encounters signals that have powers differing by a factor of 100,000,000,000,000 (10^{14} or 100 trillion!). Expressed in dB, that range is 140, which is a lot easier to work with than the preceding number, even in scientific notation!

The formula for computing the decibel equivalent of a power ratio is

$$dB = 10 \log (\text{power ratio}) = 10 \log (P1/P2) \quad (145)$$

or if voltage is used

$$dB = 20 \log (\text{voltage ratio}) = 20 \log (V1/V2) \quad (146)$$

Positive values of dB mean the ratio is greater than 1 and negative values of dB indicate a ratio of less than 1. For example, if an amplifier turns a 5 W signal into a 25 W signal, that's a gain of 10 log (25 / 5) = 10 log (5) = 7 dB. On the other hand, if by adjusting a receiver's volume control the audio output signal voltage is reduced from 2 V to 0.1 V, that's a loss: 20 log (0.1 / 2) = 20 log (0.05) = –26 dB.

If you are comparing a measured power or voltage (P_M or V_M) to some reference power (P_{REF} or V_{REF}) the formulas are:

$$dB = 10 \log (P_M / P_{REF})$$

$$dB = 20 \log (V_M / V_{REF})$$

There are several commonly-used reference powers and voltages, such as 1 V or 1 mW. When a dB value uses one of them as

the references, dB is followed with a letter. Here are the most common:

- dBV means dB with respect to 1 V (V_{REF} = 1 V)
- dBμV means dB with respect to 1 μV (V_{REF} = 1 μV)
- dBm means dB with respect to 1 mW (P_{REF} = 1 mW)

If you are given a ratio in dB and asked to calculate the power or voltage ratio, use the following formulas:

$$\text{Power ratio} = \text{antilog}\ (\text{dB} / 10) \quad (147)$$

$$\text{Voltage ratio} = \text{antilog}\ (\text{dB} / 20) \quad (148)$$

Example: A power ratio of –9 dB = antilog –(9 / 10) = antilog (–0.9) = ⅛ = 0.125

Example: A voltage ratio of 32 dB = antilog (32 / 20) = antilog (1.6) = 40

Antilog is also written as $\log^{-1}$ and may be labeled that way on calculators.

CONVERTING BETWEEN DB AND PERCENTAGE

You may also have to convert back and forth between decibels and percentages. Here are the required formulas:

$$\text{dB} = 10 \log\ (\text{percentage power} / 100\%) \quad (149)$$

$$\text{dB} = 20 \log\ (\text{percentage voltage} / 100\%) \quad (150)$$

$$\text{Percentage power} = 100\% \times \text{antilog}\ (\text{dB} / 10) \quad (151)$$

$$\text{Percentage voltage} = 100\% \times \text{antilog}\ (\text{dB} / 20) \quad (152)$$

Example: A power ratio of 20% = 10 log (20% / 100%) = 10 log (0.2) = –7 dB

Example: A voltage ratio of 150% = 20 log (150% / 100%) = 20 log (1.5) = 3.52 dB

Example: –1 dB represents a percentage power = 100% × antilog (–1 / 10) = 79%

Example: 4 dB represents a percentage voltage = 100% × antilog (4 / 20) = 158%

2.16.2 Rectangular and Polar Coordinates

Graphs are drawings of what equations describe with symbols — they're both saying the same thing. Graphs are used to present a visual representation of what an equation expresses. The way in which mathematical quantities are positioned on the graph is called the *coordinate system*. *Coordinate* is another name for the numeric scales that divide the graph into regular units. The location of every point on the graph is described by a pair of *coordinates*.

The two most common coordinate systems used in radio are the *rectangular-coordinate* system shown in **Fig 2.108** (sometimes called *Cartesian coordinates*) and the *polar-coordinate* system shown in **Fig 2.109**.

The line that runs horizontally through the center of a rectangular coordinate graph is called the *X axis*. The line that runs vertically through the center of the graph is normally called the *Y axis*. Every point on a rectangular coordinate graph has two coordinates that identify its location, X and Y, also written as (X,Y). Every different pair of coordinate values describes a different point on the graph. The point at which the two axes cross — and where the numeric values on both axes are zero — is called the *origin*, written as (0,0).

In Fig 2.108, the point with coordinates (3,5) is located 3 units from the origin along

Computer Software and Calculators

Every version of the *Windows* operating system comes with a calculator program located in the Accessories program group and a number of free calculators are available on-line. Enter "online" and "calculator" into the search window of an Internet search engine for a list. If you can express your calculation as a mathematical expression, such as sin(45) or 10log(17.5), it can be entered directly into the search window at **www.google.com** and the Google calculator will attempt to solve for the answer.

Microsoft *Excel* (and similar spreadsheet programs) also make excellent calculators. If you are unfamiliar with the use of spreadsheets, here are some online tutorials and help programs:

homepage.cs.uri.edu/tutorials/csc101/pc/excel97/excel.html
www.baycongroup.com/el0.htm
www.bcschools.net/staff/ExcelHelp.htm

For assistance in converting units of measurement, the Web site **www.onlineconversion.com** is very useful. The Google online unit converter can also be used by entering the required conversion, such as "12 gauss in tesla", into the Google search window.

Decibels In Your Head

Every time you increase the power by a factor of 2 times, you have a 3-dB increase of power. Every 4 times increase of power is a 6-dB increase of power. When you increase the power by 10 times, you have a power increase of 10 dB. You can also use these same values for a decrease in power. Cut the power in half for a 3-dB loss of power. Reduce the power to ¼ the original value for a 6-dB loss in power. If you reduce the power to ⅒ of the original value you will have a 10-dB loss. The power-loss values are often written as negative values: –3, –6 or –10 dB. The following tables show these common decibel values and ratios.

Common dB Values and Power Ratios

P2/P1	*dB*
0.1	–10
0.25	–6
0.5	–3
1	0
2	3
4	6
10	10

Common dB Values and Voltage Ratios

V2/V1	*dB*
0.1	–20
0.25	–12
0.5	–6
0.707	–3
1.0	0
1.414	3
2	6
4	12
10	20

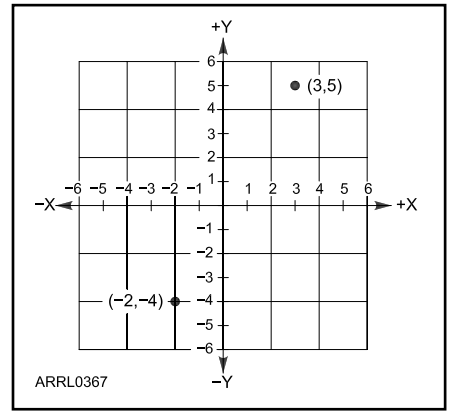


Fig 2.108 — Rectangular-coordinate graphs use a pair of axes at right angles to each other; each calibrate in numeric units. Any point on the resulting grid can be expressed in terms of its horizontal (X) and vertical (Y) values, called coordinates.

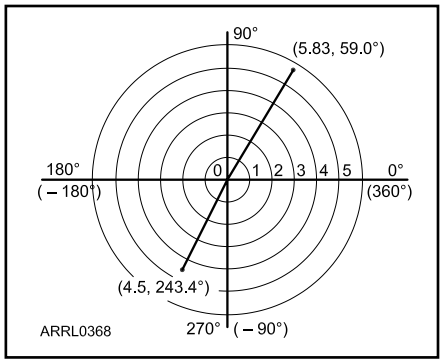


Fig 2.109 — Polar-coordinate graphs use a radius from the origin and an angle from the 0° axis to specify the location of a point. Thus, the location of any point can be specified in terms of a radius and an angle.

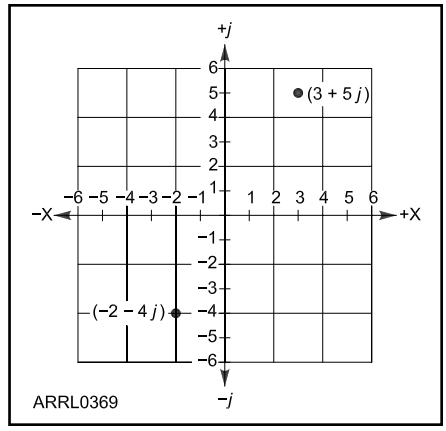


Fig 2.110 — The Y axis of a complex-coordinate graph represents the imaginary portion of complex numbers. This graph shows the same numbers as in Figure 4-1, graphed as complex numbers.

the X axis and 5 units from the origin along the Y axis. Another point at (–2,–4) is found 2 units to the left of the origin along the X axis and 4 units below the origin along the Y axis. Do not confuse the X coordinate with the X representing reactance! Reactance is usually plotted on the Y-axis, creating occasional confusion.

In the polar-coordinate system, points on the graph are described by a pair of numeric values called *polar coordinates*. In this case, a length, or *radius*, is measured from the origin, and an *angle* is measured counterclockwise from the 0° line as shown in Fig 2.109. The symbol *r* is used for the radius and θ for the angle. A number in polar coordinates is written r∠θ. So the two points described in the last paragraph could also be written as (5.83, ∠59.0°) and (4.5, ∠243.4°). Unlike geographic maps, 0° is always to the *right* along the X-axis and 90°at the top along the Y-axis!

Angles are specified counterclockwise from the 0° axis. A negative value in front of an angle specifies an angle measured clockwise from the 0° axis. (Angles can also be given with reference to some arbitrary line, as well.) For example, –270° is equivalent to 90°, –90° is equivalent to 270°, 0° and –360° are equivalent, and +180° and –180° are equivalent. With an angle measured clockwise from the 0° axis, the polar coordinates of the second point in Fig 2.109 would be (4.5, ∠–116.6°).

In some calculations the angle will be specified in *radians* (1 radian = 360 / 2π degrees = 57.3°), but you may assume that all angles are in degrees in this book. Radians are used when *angular frequency* (ω) is used instead of frequency (*f*). Angular frequency is measured in units of 2π radians/sec and so ω = 2πf.

In electronics, it's common to use both the rectangular and polar-coordinate systems when dealing with impedance problems. The examples in the next few sections of this book should help you become familiar with these coordinate systems and the techniques for changing between them.

2.16.3 Complex Numbers

Mathematical equations that describe phases and angles of electrical quantities use the symbol *j* to represent the square root of minus one ($\sqrt{-1}$). (Mathematicians use *i* for the same purpose, but *i* is used to represent current in electronics.) The number *j* and any real number multiplied by *j* are called *imaginary numbers* because no real number is the square root of minus one. For example, 2*j*, 0.1*j*, 7*j*/4, and 457.6*j* are all imaginary numbers. Imaginary numbers are used in electronics to represent reactance, such as *j*13 Ω of inductive reactance or –*j*25 Ω of capacitive reactance.

Real and imaginary numbers can be combined by using addition or subtraction. Adding a real number to an imaginary number creates a hybrid called a *complex number*, such as 1 + *j* or 6 – 7*j*. These numbers come in very handy in radio, describing impedances, relationships between voltage and current, and many other phenomena.

If the complex number is broken up into its real and imaginary parts, those two numbers can also be used as coordinates on a graph using *complex coordinates*. This is a special type of rectangular-coordinate graph that is also referred to as the *complex plane*. By convention, the X axis coordinates represent the real number portion of the complex number and Y axis coordinates the imaginary portion. For example, the complex number 6 – 7*j* would have the same location as the point (6,–7) on a rectangular-coordinate graph. **Fig 2.110** shows the same points as Fig 2.108,

The number *j* has a number of interesting properties:

$1/j = -j$

$j^2 = -1$

$j^3 = -j$

$j^4 = 1$

Multiplication by *j* can also represent a phase shift or rotation of +90°. In polar coordinates $j = 1 \angle 90°$.

WORKING WITH COMPLEX NUMBERS

Complex numbers representing electrical quantities can be expressed in either rectangular form (a + *j*b) or polar form (r∠θ) as described above. Adding complex numbers is easiest in rectangular form:

$$(a + jb) + (c + jd) = (a+c) + j(b+d) \quad (153)$$

Multiplying and dividing complex numbers is easiest in polar form:

$$a\angle\theta_1 \times b\angle\theta_2 = (a \times b)\angle(\theta_1 + \theta_2) \quad (154)$$

and

$$\frac{a\angle\theta_1}{b\angle\theta_2} = \left(\frac{a}{b}\right)\angle(\theta_1 - \theta_2) \quad (155)$$

Converting from one form to another is useful in some kinds of calculations. For example, to calculate the value of two complex impedances in parallel you use the formula

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2}$$

To calculate the numerator (Z_1Z_2) you would write the impedances in polar form. To calculate the numerator ($Z_1 + Z_2$) you would write the impedances in rectangular form. So you need to be able to convert back and forth from one form to the other. Here is the procedure:

To convert from rectangular (a+*j*b) to polar form (r ∠ θ):

$$r = \sqrt{a^2 + b^2} \text{ and } \theta = \tan^{-1}(b / a) \quad (156)$$

To convert from polar to rectangular form:

$$a = r\cos\theta \text{ and } b = r\sin\theta \quad (157)$$

Many calculators have polar-rectangular

conversion functions built-in and they are worth learning how to use. Be sure that your calculator is set to the correct units for angles, radians or degrees.

Example: Convert $3 \angle 60°$ to rectangular form:

$$a = 3 \cos 60° = 3\,(0.5) = 1.5$$

$$b = 3 \sin 60° = 3\,(0.866) = 26$$

$$3\angle 60° = 1.5 + j2.6$$

Example: Convert $0.8 + j0.6$ to polar form:

$$r = \sqrt{0.8^2 + 0.6^2} = 1$$

$$\theta = \tan^{-1}\left(0.6 / 0.8\right) = 36.8^{\circ}$$

$$0.8 + j0.6 = 1\angle 36.8°$$

2.16.5 Accuracy and Significant Figures

The calculations you have encountered in this chapter and will find throughout this handbook follow the rules for accuracy of calculations. Accuracy is represented by the number of *significant digits* in a number. That is, the number of digits that carry numeric value information beyond order of magnitude. For example, the numbers 0.123, 1.23, 12.3, 123, and 1230 all have three significant digits.

The result of a calculation can only be as accurate as its *least* accurate measurement or known value. This is important because it is rare for measurements to be more accurate than a few percent. This limits the number of useful significant digits to two or three. Here's another example; what is the current through a 12-Ω resistor if 4.6 V is applied? Ohm's Law says I in amperes = 4.6 / 12 = 0.3833333... but because our most accurate numeric information only has two significant digits (12 and 4.6), strictly applying the significant digits calculation rule limits our answer to 0.38. One extra digit is often included, 0.383 in this case, to act as a guide in rounding off the answer.

Quite often a calculator is used, and the result of a calculation fills the numeric window. Just because the calculator shows 9 digits after the decimal point, this does not mean that is a more correct or even useful answer.

2.17 References and Bibliography

BOOKS

Alexander and Sadiku, *Fundamentals of Electric Circuits* (McGraw-Hill)

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

Glover, T., *Pocket Ref* (Sequoia Publishing)

Horowitz and Hill, *The Art of Electronics* (Cambridge University Press)

Kaiser, C., *The Resistor Handbook* (Saddleman Press, 1998)

Kaiser, C., *The Capacitor Handbook* (Saddleman Press, 1995)

Kaiser, C., *The Inductor Handbook* (Saddleman Press, 1996)

Kaiser, C., *The Diode Handbook* (Saddleman Press, 1999)

Kaiser, C., *The Transistor Handbook* (Saddleman Press, 1999)

Kaplan, S., *Wiley Electrical and Electronics Dictionary* (Wiley Press)

Orr, W., *Radio Handbook* (Newnes)

Terman, F., *Radio Engineer's Handbook* (McGraw-Hill)

MAGAZINE AND JOURNAL ARTICLES

ARRL Lab Staff, "Lab Notes — Capacitor Basics," *QST*, Jan 1997, pp 85-86

B. Bergeron, NU1N, "Under the Hood II: Resistors," *QST,* Nov 1993, pp 41-44

B. Bergeron, NU1N, "Under the Hood III: Capacitors," *QST,* Jan 1994, pp 45-48

B. Bergeron, NU1N, "Under the Hood IV: Inductors," *QST,* Mar 1994, pp 37-40

B. Bergeron, NU1N, "Under the Hood: Lamps, Indicators, and Displays," *QST,* Sep 1994, pp 34-37

J. McClanahan, W4JBM, "Understanding and Testing Capacitor ESR," *QST*, Sep 2003, pp 30-32

W. Silver, NØAX, "Experiment #24 — Heat Management," *QST*, Jan 2005, pp 64-65

W. Silver, NØAX, "Experiment #62 — About Resistors," *QST*, Mar 2008, pp 66-67

W. Silver, NØAX, "Experiment #63 — About Capacitors," *QST*, Apr 2008, pp 70-71

W. Silver, NØAX, "Experiment #74 — Resonant Circuits," *QST*, Mar 2009, pp 72-73

J. Smith, K8ZOA, "Carbon Composition, Carbon Film and Metal Oxide Film Resistors," *QEX*, Mar/Apr 2008, pp 46-57