Contents

Chapter 2
Electrical Fundamentals

Collecting material on fundamental concepts from previous editions, this chapter summarizes the basic ideas of electricity and electronics. It covers the physical quantities, elementary circuits, basic components, and the laws that govern their behavior. These are the foundations on which all of electronics is constructed. Glossaries are included for each group of topics, as well.

Since many of the basic ideas are expressed or defined in terms of mathematics, the collection of tutorials "Radio Mathematics" is available online in the Handbook's reference web page at **www.arrl.org/arrl-handbook-reference**. Tutorials on scientific notation and using a scientific calculator are also provided.

Chapter 2 — Online Content

Articles
- Hands On Radio: Kirchoff's Laws by Ward Silver, NØAX
- Hands On Radio: Laying Down the Laws by Ward Silver, NØAX
- Hands On Radio: Putting the Laws to Work by Ward Silver, NØAX
- Hands On Radio: Thevenin Equivalents by Ward Silver, NØAX

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). Moving charges in a magnetic field also generates an electrical force. This is the *electromotive force* (or EMF), the source of energy that causes charged particles to move. *Voltage* is the general term for the strength of the electromotive force or the difference in electrical potential between two points. Voltage and EMF are often used interchangeably in radio. A good diagram showing the relationship between EMF and voltage is available at **hyperphysics.phy-astr.gsu.edu/hbase/electric/elevol.html#c2**.

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 an imbalance of charge between two regions generates an EMF, 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 electricity, 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 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

Measurement of electrical quantities is made in several standard units. 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 (or through a specific area) in one second so 1 A = 1 C/s. Electromotive force (EMF) is measured in *volts* (V) and represented by *e*, *E*, *v*, or *V* in equations. One volt is defined as the EMF required for one ampere of current to do one joule (J, a measure of energy) of work and 1 V = 1 J/C.

2.1.4 Series and Parallel Circuits

A *circuit* is any conducting path through which current can flow between two points 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 **Figure 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 as + and –) in the direction shown by the arrow through three *resistors* (electronic components discussed later in this chapter) and back to the battery's negative terminal. Current is

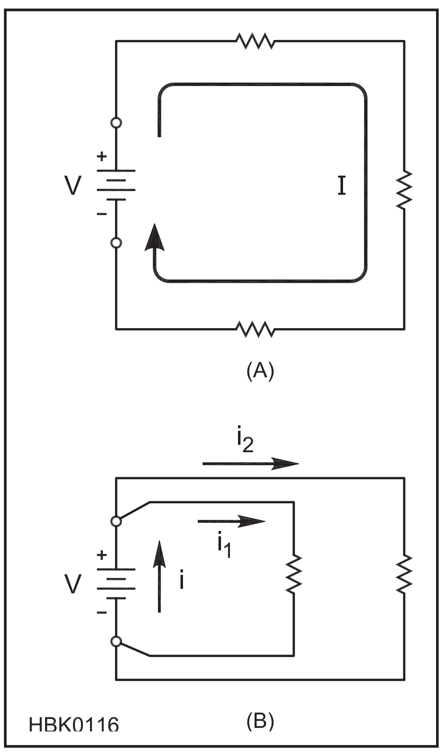


Figure 2.1 — A series circuit (A) has the same current through all components. Parallel circuits (B) apply the same voltage to all components.

When Is E a V and V an E?

Beginners in electronics are often confused about the interchange of *V* and *E* to refer to voltage in a circuit. When should each be used? Unfortunately, there is no universal convention. *E* or *e* usually refers to force created by an electric or magnetic field, or a battery. *E* is also commonly used in the equation for Ohm's Law: *I* = *E*/*R*. *V* or *v* is used when describing a measured difference in voltage between two points in a circuit or the terminal voltage of a power supply or battery. Capital *V* is always used as an abbreviation for volts.

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, what they are used to measure, 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*	*Physical Quantities*	*Named for*
Ampere (A)	Current	Coulombs per second	Andree Ampere 1775 – 1836
Coulomb (C)	Charge	Quantity of charge	Charles Coulomb 1736 – 1806
Farad (F)	Capacitance	Coulombs per volt	Michael Faraday 1791 – 1867
Henry (H)	Inductance	Volts per amp per second	Joseph Henry 1797 – 1878
Hertz (Hz)	Frequency	Cycles per second	Heinrich Hertz 1857 – 1894
Ohm (Ω)	Resistance	Volts per amp	Georg Simon Ohm 1787 – 1854
Watt (W)	Power	Joules per second	James Watt 1736 – 1819
Volt (V)	Voltage	Joules per coulomb	Alessandro Volta 1745 – 11827

Schematic Diagrams

The drawing in Figure 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** provided at the end of this chapter. You will find additional information on reading and drawing schematic diagrams in the ARRL website Technology section at **www.arrl.org/circuit-construction**.

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.

2.1.5 Direct and Alternating Current

A circuit is a complete conductive path for current to flow from a source, through a load and back to the source. *Direct current* or *dc* flows in only one direction. *Alternating current* or *ac* changes direction. **Figure 2.2** illustrates the two types of circuits. Circuit A shows the source as a battery, a typical dc source. Circuit B shows a voltage source symbol to indicate ac such as from a generator or household power outlet. 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 billion times per second.

Graphs of current or voltage, such as Figure 2.2, 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 we have a dc circuit, even if the level of dc constantly changes. **Figure 2.3A** 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*.

We can also look at intermittent and pulsating dc as a combination of an ac and a dc current (Figures 2.3B and 2.3C). 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.

2.1.6 Glossary — 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.

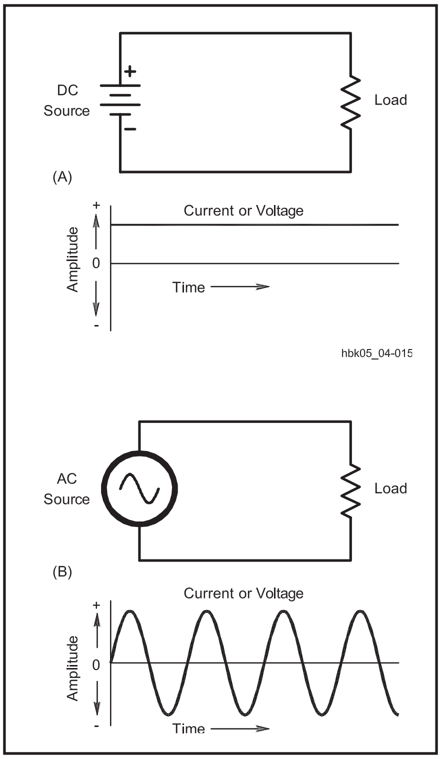


Figure 2.2 — Basic circuits for direct and alternating currents. With each circuit is a graph of the current, constant for the dc circuit, but periodically changing direction in the ac circuit.

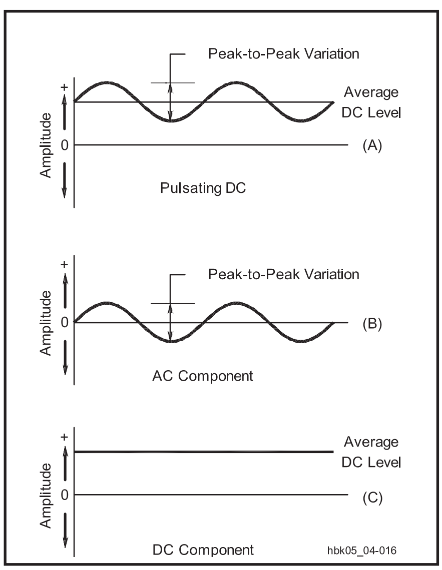


Figure 2.3 — A pulsating dc current (A) and its resolution into an ac component (B) and a dc component (C).

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.

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* or *I* in equations.

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

Electronic current — see **Conventional Current**

Electromotive force (EMF) — The source of energy that creates a force between charged particles or regions. The term used to define the force of attraction or repulsion between electrically charged regions. Also see *voltage*.

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 through 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.

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

Potential — See **voltage**

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.

Voltage — The general term for the difference in electrical potential energy between two points. Measured in volts or joules/coulomb.

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

2.2 Resistance and Conductance

2.2.1 Resistance and Resistors

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.

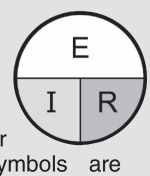
One of the best conductors is copper, and it is frequently convenient to compare the resistance of a material under consideration with that of a copper conductor of the same size and shape. **Table 2.1** gives the ratio of the resistivity of various conductors to the resistivity of copper.

A package of material exhibiting a certain amount of resistance and made into a single unit or component is called a *resistor*. There are many types of resistors described in the **Circuits and Components** chapter, each suited to different applications and power levels. Next to the transistors built into microprocessors by the billion, resistors are the most common electronic component of all.

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

Ohm's Law Timesaver



This simple diagram presents the mathematical equations relating voltage, current, and resistance. Cover the unknown 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.

2.2.2 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 regard 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 voltage is applied will vary with the resistance of the conductor. The lower the resistance, the greater the current for a given EMF. One ohm (Ω) 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 = E / I

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

Rearranging the equation gives the other common forms of Ohm's Law as:

E = I × R

and

I = E / R

All three forms of the equation are used often in electronics and radio. 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 **Figure 2.4**. 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:

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

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 E = I × R. Convert the current from milliamperes to amperes by dividing by 1,000 mA / A (or multiplying by 10^{-3} A / mA) so that 150 mA becomes 0.150 A. (Notice the conversion factor of 1,000 does not limit the number of significant figures in the calculated answer.)

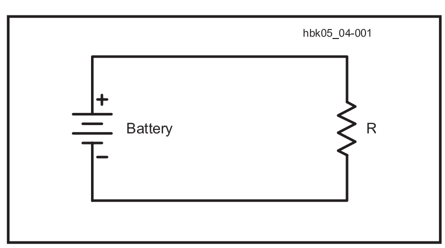


Figure 2.4 — A simple circuit consisting of a battery and a resistor.

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

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 5,000-Ω resistor? Since I is unknown,

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

This value of current is more conveniently stated 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.

2.2.4 Glossary — Conductance and Resistance

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

Ohm (Ω) — 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.

Ohm's law — The expression that describes resistance (R) as the proportional relationship between voltage (E) and current (I); R = E / I. Named for Georg Ohm, who first described the relationship.

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

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 **Figure 2.5A**, 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 Figure 2.1A, is called a *loop*.

Very few actual circuits are as simple as those shown in Figure 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:

$$(I_{in1} + I_{in2} + ...) - (I_{out1} + I_{out2} + ...) = 0$$

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

An equivalent 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:

$$(I_{in1} + I_{in2} + ...) = (I_{out1} + I_{out2} + ...)$$

KCL is illustrated by the following example. Suppose three resistors (R1 = 5.0 kΩ, R2 = 20.0 kΩ, and R3 = 8.0 kΩ) are connected in parallel as shown in Figure 2.5A. 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.

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

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

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

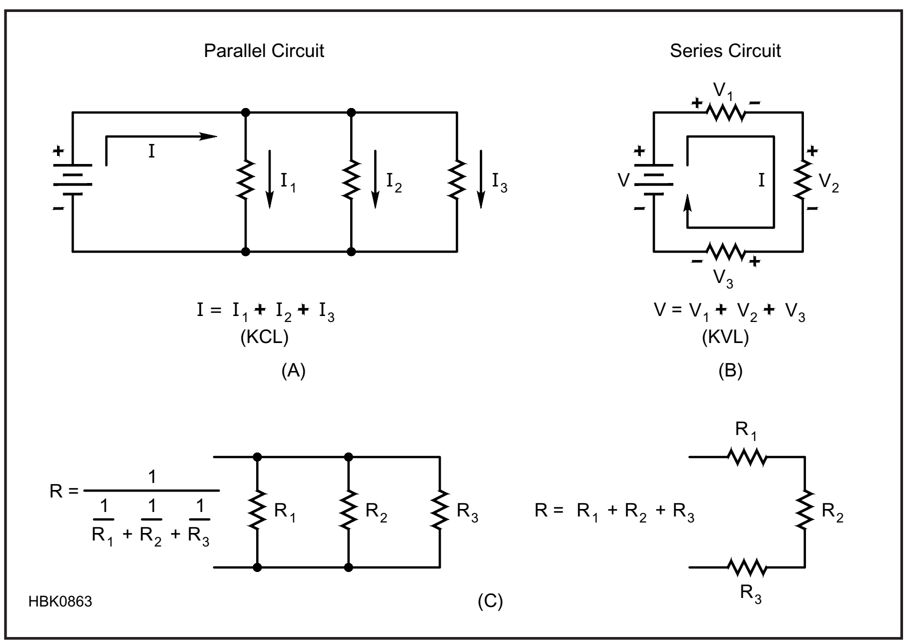


Figure 2.5 — An example of resistors in parallel (A) and series (B). In series circuits, the current is the same in all components, and voltages are summed. In parallel circuits, voltage across all components is the same and the sum of currents into and out of circuit junctions must be equal. Part C shows how to calculate equivalent values for series and parallel combinations.

Notice that the branch currents are inversely proportional to the resistances. The 20-kΩ resistor has a value four times larger than the 5-kΩ 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.

Using the balancing form of KCL the current that must be supplied by the battery is:

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

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

$I_{Batt} = 93.7\ mA$

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:

$$R_{EQUIV} = \frac{1}{\frac{1}{R1} + \frac{1}{R2} + \frac{1}{R3} + \frac{1}{R4} ...}$$

where the dots indicate that any number of parallel resistors can be combined by the same method. The equation is often referred to as the "reciprocal of reciprocals." Figure 2.5C shows the general rule on a schematic.

In the example of the previous section, the equivalent resistance is:

$$R = \frac{1}{\frac{1}{5\ k\Omega} + \frac{1}{20\ k\Omega} + \frac{1}{8\ k\Omega}} = 2.67\ k\Omega$$

The notation "//" (two slashes) is frequently used to indicate "in parallel with." Using that notation, the preceding example would be given as "5.0 kΩ // 20 kΩ // 8.0 kΩ."

If all the resistors in parallel have the same value, divide the resistor value by the number of resistors, N, to get the parallel resistance. For example, for five 50-Ω resistors in parallel, the equivalent resistance is R = 50 / N = 50 / 5 = 10 Ω.

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

$$R_{EQUIV} = \frac{R1 \times R2}{R1 + R2}$$

Example: If a 500-Ω resistor is connected in parallel with a 1200-Ω resistor, 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 this equation until all have been combined into a single equivalent. This is a bit easier than using the general "reciprocal of reciprocals" equation to do the conversion in a single step.

Another useful equation finds the value of a parallel resistor, R_{PAR}, that when connected across a known resistor, R_1, creates a desired equivalent resistance, R_{DES}:

$$R_{PAR} = \frac{R_1 R_{DES}}{R_1 - R_{DES}}$$

For example, what value of resistor connected across a 10 kΩ resistor results in a 3 kΩ equivalent resistance?

$$R_{PAR} = \frac{10\ k\Omega \times 3\ k\Omega}{10\ k\Omega - 3\ k\Omega} = 4.28\ k\Omega$$

From the 5% series of resistance values, a 4.3 kΩ value is the closest available value.

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 Figure 2.5B, 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. KVL is stated mathematically as:

$E_1 + E_2 + E_3 + ... = 0$

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 previous example, 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 × 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:

$(E_{source1} + E_{source2} + ...) = (E_{sin\ k1} + E_{sin\ k2} + ...)$

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 Figure 2.5B, 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:

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

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

$E3 = I \times R3 = 0.00758\ A \times 000\ \Omega = 60.6\ V$

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:

$$R_{EQUIV} = R1 + R2 + R3 + R4 \ldots$$

Figure 2.5C shows the general rule on a schematic.

Example: Suppose that three resistors are connected to a source of voltage as shown in Figure 2.5B. 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

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

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. Conductances in series are thus combined similarly to resistors in parallel:

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

and two conductances in series may be combined in a manner similar to two parallel resistors:

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

Conductances in parallel are combined similarly to resistances in series:

$$G_{TOTAL} = G1 + G2 + G3 + G4 \ldots$$

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 **Figure 2.6A**. In

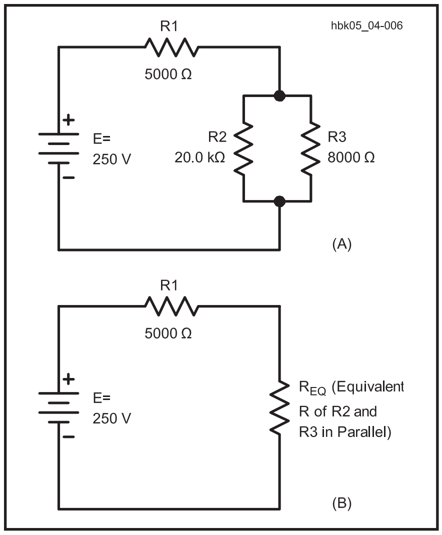


Figure 2.6 — At A, an example of resistors in series-parallel. The equivalent circuit is shown at B.

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 simplest process begins with combining any two of the resistors into a single equivalent resistance using the formulas for series or parallel resistances. Then combine the resulting equivalent resistance with any single remaining resistor into a new equivalent resistance. Repeat the process of combining the equivalent resistance with a single resistor until all resistances have been combined into a single equivalent resistance. For example, to find the equivalent resistance for the circuit in Figure 2.5A: 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 Figure 2.5B. These two resistances can then be combined into a single equivalent resistance, R_{TOTAL}, for the entire circuit:

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

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

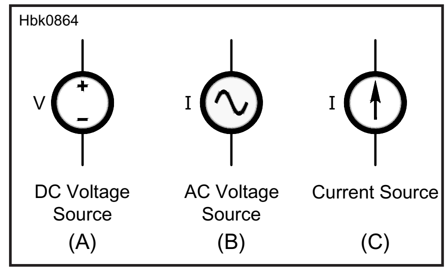


Figure 2.7— Voltage sources for dc (A) and ac (B) and current sources (C) are examples of ideal energy sources.

The battery current is then:

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

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 **Figure 2.7**. 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 Figure 2.2 were inside of or internal to the battery.

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 Figure 2.6A, to find the current through R3. In this example, illustrated in **Figure 2.8**, the circuit is redrawn to show R1 and R2 forming a voltage divider, with R3 as the load (Figure 2.8A). 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):

$$I = \frac{E}{R1 + R2}$$

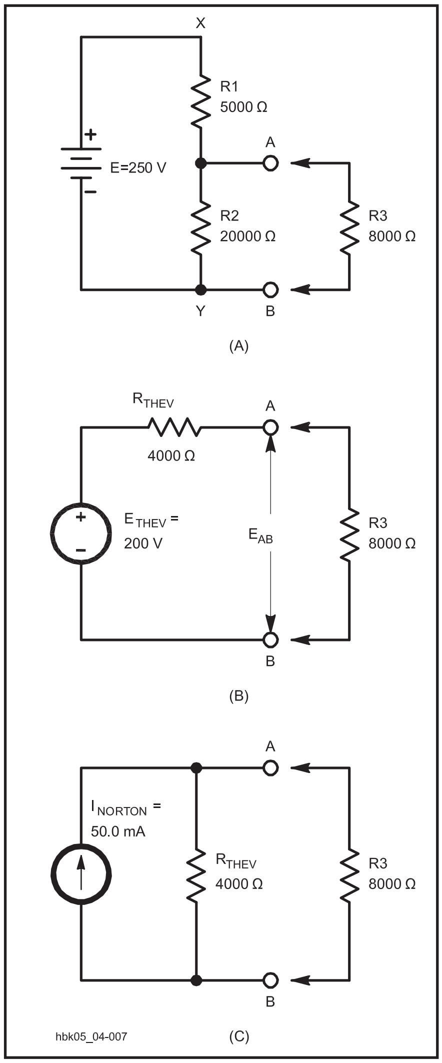


Figure 2.8 — Equivalent circuits for the circuit shown in Figure 2.6. A shows the circuit to be replaced by an equivalent circuit from the perspective of the resistor (R3 load). B shows the Thevenin-equivalent circuit, with a resistor and a voltage source in series. C shows the Norton-equivalent circuit, with a resistor

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

$$E_{AB} = I \times R2$$

By substituting the equation for current for I, we have an expression for E_{AB} in which all values are known:

$$E_{AB} = \frac{R2}{R1 + R2} \times E$$

Using the values in our example, this becomes:

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

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 Figure 2.8A. R1 and R2 are then effectively placed in parallel, as viewed from terminals A and B. R_{THEV} is then:

$$R_{THEV} = \frac{R1 \times R2}{R1 + R2}$$

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

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

This gives the Thevenin-equivalent circuit as shown in Figure 2.8B. The circuits of Figures 2.8A and 2.8B 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

$$I3 = \frac{E_{THEV}}{R_{TOTAL}} = \frac{E_{THEV}}{R_{THEV} + R3}$$

Substituting the values from our example:

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

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 Figure 2.8C. 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 Figure 2.8A, the short circuit completely bypasses R2 and the current is:

$$I_{AB} = \frac{E}{R1}$$

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,

$$P = I \times E$$

where

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

(This discussion pertains only to direct current in resistive circuits. See the **Radio Fundamentals** 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 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 U.S. 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:

$$Wh = P \times t$$

where

Wh = energy in watt-hours,
P = power in watts, and
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 equivalent to joules.

It is important to realize, both for calculation purposes and for efficient use of power resources, that 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:

Energy = 1.5 V × 2 A × 1 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 Sources** 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:

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

where

Eff = efficiency (as a value or fraction between 0 and 1),

P_O = power output (W), and

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 efficiency ratio 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:

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

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 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 **Figure 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 Figure 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*

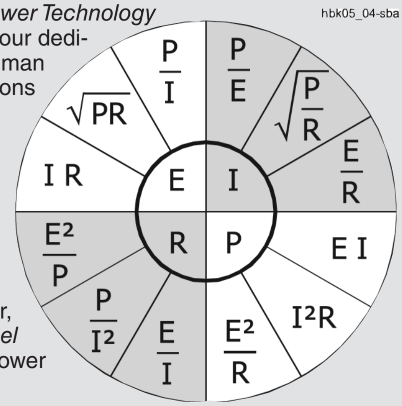


Figure 2.A1 — Electrical formulas

2.4.4 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:

$P = E^2 / R$

and

$P = I^2 \times R$

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 dissipated by (converted to heat in) a 4000-Ω resistor if the potential applied to it is 200 V?

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

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

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

$P = 0.00040\ A^2 \times 300\ \Omega$

$P = 0.12\ W$

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.

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 **Figure 2.9A**. 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 Figure 2.9B. Adding an off position would give a single-pole, double-throw, center-off (ON-OFF-ON) switch as shown in Figure 2.9C.

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 Figure 2.9D. Most switches stay in position once set, but some are spring-loaded so they

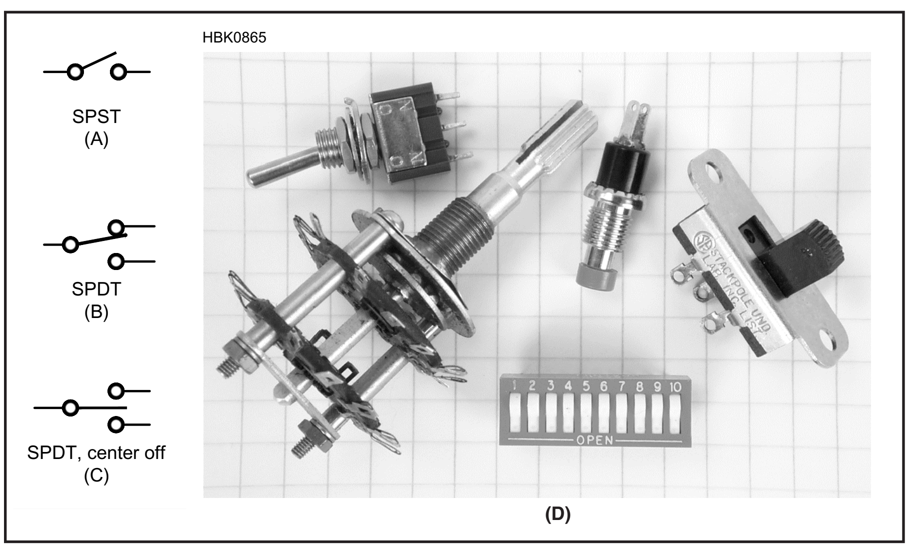


Figure 2.9 — Schematic diagrams of various types of switches. A is an SPST, B is an SPDT, and C is an SPDT switch with a center-off position. The photo (D) shows examples of various styles of switches. The ¼-inch-ruled graph paper background provides for size comparison.

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

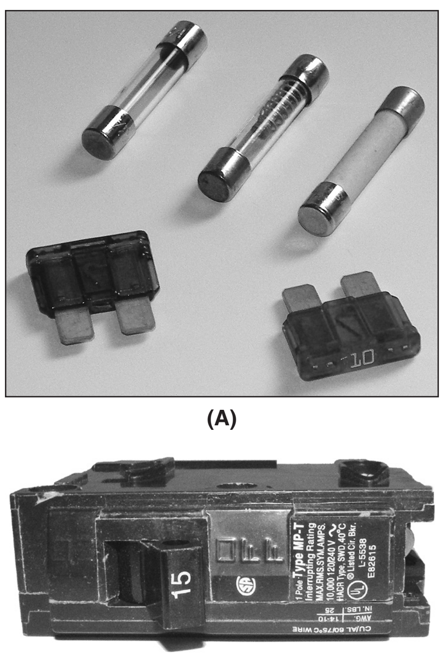
Fuses self-destruct to protect circuit wiring or equipment. The fuse *element* that melts or *blows* is a carefully shaped piece of soft metal, usually mounted in a cartridge of some kind. The element is designed to safely carry a given amount of current and to melt at a current value that is a certain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

Figure 2.10A 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 Figure 2.10A, 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.



(A)

(B)

Figure 2.10 — 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.

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 Figure 2.10B) 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**.)

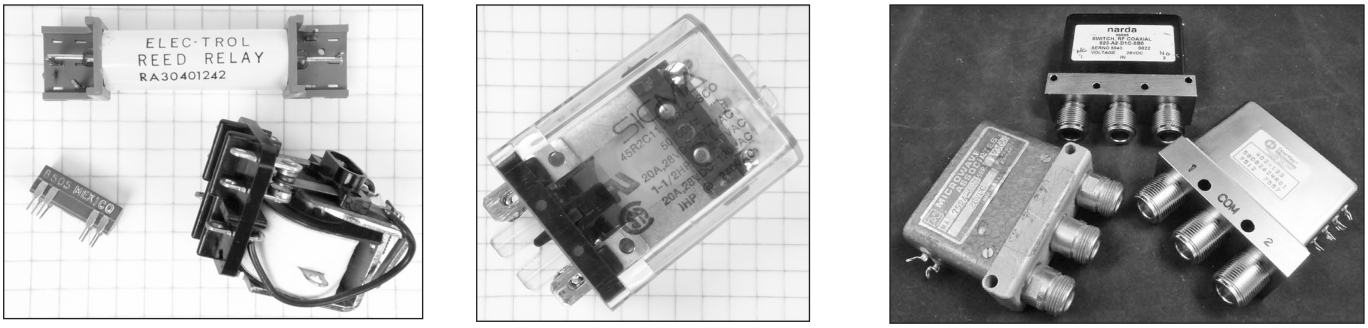
Fuses blow and circuit breakers open for several reasons. The most obvious reason is that a problem develops in the circuit, causing too much current to flow. In this case, the circuit problem needs to be fixed. A fuse can fail from being cycled on and off near its current rating. The repeated thermal 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* (NO) contact sets



(A) (B) (C)

Figure 2.11 — 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 an integrated circuit. The contacts and coil can clearly be seen in the open-frame relay. Photo B shows a relay inside a plastic case. The background grid for Photos A and B is ¼ inch to provide a size comparison. Photo C shows three surplus SPDT coaxial relays with type N connectors used for RF switching at VHF and higher frequencies.

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. **Figure 2.11** shows some typical relays found in amateur equipment.

Relays with dc coils store energy in the coil when energized. When voltage to the coil is removed, the stored energy creates a large voltage transient that can damage the relay driver circuit. Techniques for managing the coil's stored energy are presented in "Using the Transistor as a Switch" in the **Circuits and Components** chapter.

A relay's *pull-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 later in this chapter). *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 supply power to the relay continuously.

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 instead 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 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*.) **Figure 2.12** shows schematic symbols for capacitors: a fixed capacitor with a single value of capacitance (Figure 2.12A) and variable capacitors adjustable over a range of values (Figure 2.12B). 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. 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.

2.6.1 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.

2.6.2 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 **Figure 2.13A**. 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. 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 plates, they remain charged despite the fact that they are no longer connected to the battery, which is the source of the voltage. In Figure 2.13B, 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 more general name for the capacitor's plates is *electrodes*. However, amateur literature generally refers to a capacitor's electrodes as *plates,* and that is the convention in this text.

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:

$$Q = CV$$

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:

$$U = \frac{V^2 C}{2}$$

where U = 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*.

Figure 2.14 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, which 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

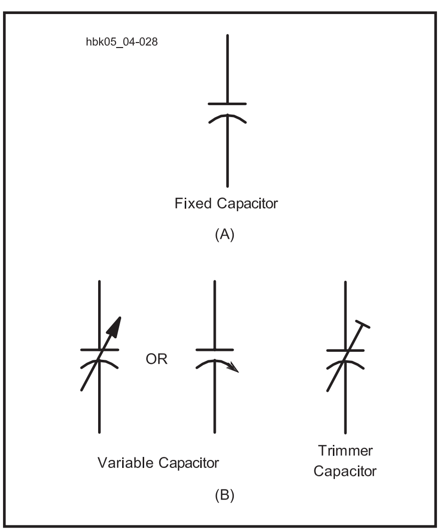


Figure 2.12 — Schematic symbol for a fixed capacitor is shown at A. The symbols for a variable capacitor are shown at B.

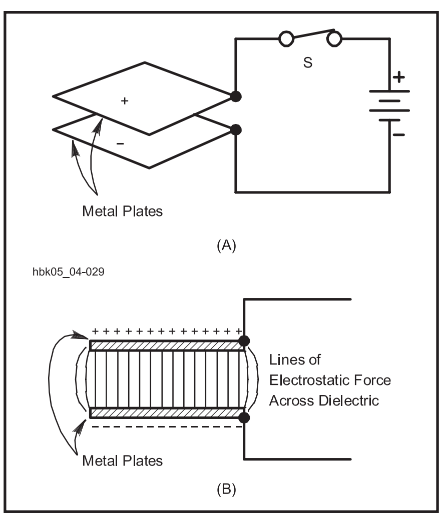


Figure 2.13 — A simple capacitor showing the basic charging arrangement at A, and the retention of the charge due to the electrostatic field at B.

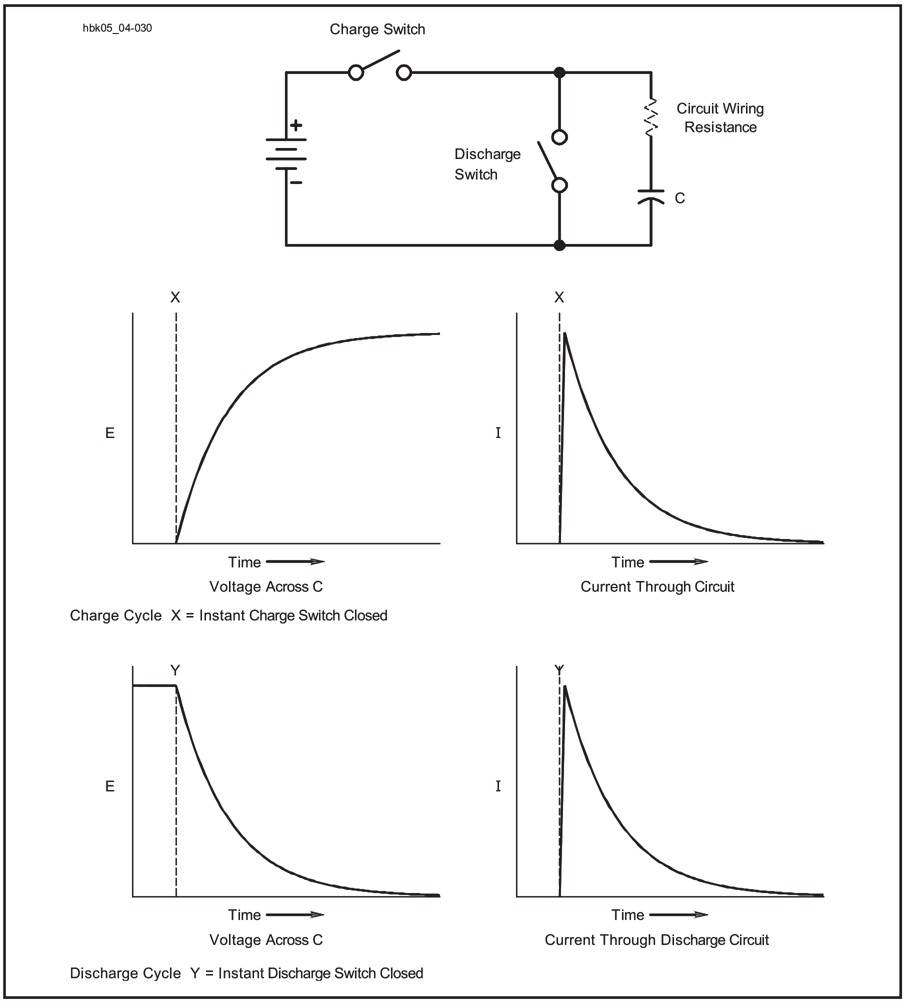


Figure 2.14 — The flow of current during the charge and discharge of a capacitor. The charging graphs assume that the charge switch is closed and the discharge switch is open. The discharging graphs assume just the opposite.

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.

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 (mmF or µµF) in place of picofarad.

CAPACITOR CONSTRUCTION

An idealized 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

$$C = \frac{A\ \varepsilon_r\ \varepsilon_0}{d}$$

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, and
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 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* (ε_r 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 2.2**. 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 **Figure 2.15**. 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:

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

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.

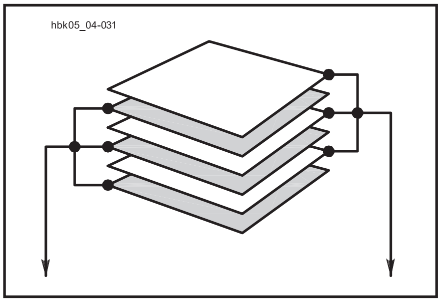


Figure 2.15 — A multiple-plate capacitor. Alternate plates are connected to each other, increasing the total area available for storing charge.

Table 2.2
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.)

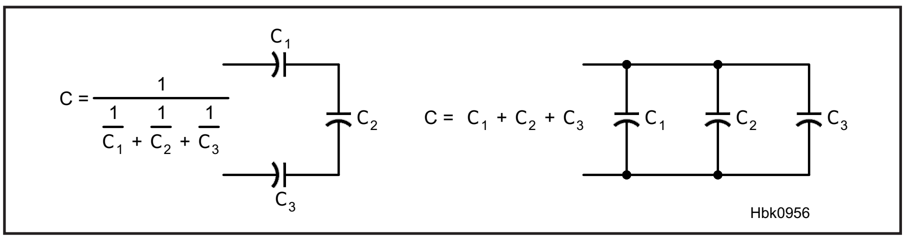


Figure 2.16 — Capacitors in series and parallel.

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

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

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?

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

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

$$C = 67.4\ pF$$

What is the capacitance if the dielectric is mineral oil? (See Table 2.2 for the appropriate dielectric constant.)

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

$$C = 150.3\ pF$$

2.6.3 Capacitors in Series and Parallel

When a number of capacitors are connected in parallel, as in the right side of **Figure 2.16**, the total capacitance of the group is equal to the sum of the individual capacitances:

$$C_{total} = C1 + C2 + C3 + C4 \ldots + C_n$$

When two or more capacitors are connected in series, as in the left side of Figure 2.16, 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.

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

For only two capacitors in series, the formula becomes:

$$C_{total} = \frac{C1 \times C2}{C1 + C2}$$

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.)

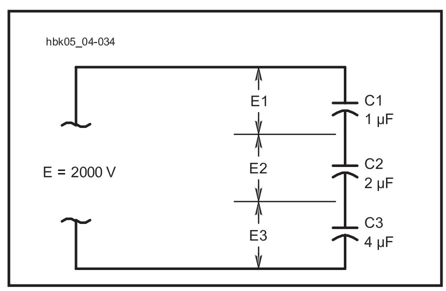


Figure 2.17 — An example of capacitors connected in series. The text shows how to find the voltage drops, E1 through E3.

Example: Three capacitors having capacitances of 1, 2, and 4 µF, respectively, are connected in series as in **Figure 2.17**. 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:

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

and

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

The sum of these three voltages equals 2,000 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 Sources** chapter.

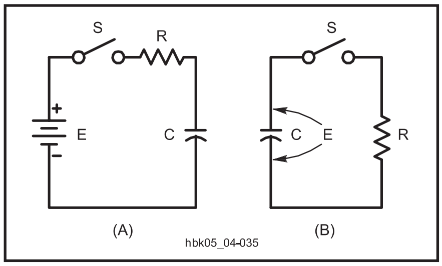


Figure 2.18 — An RC circuit. The series resistance delays the process of charging (A) and discharging (B) when the switch, S, is closed.

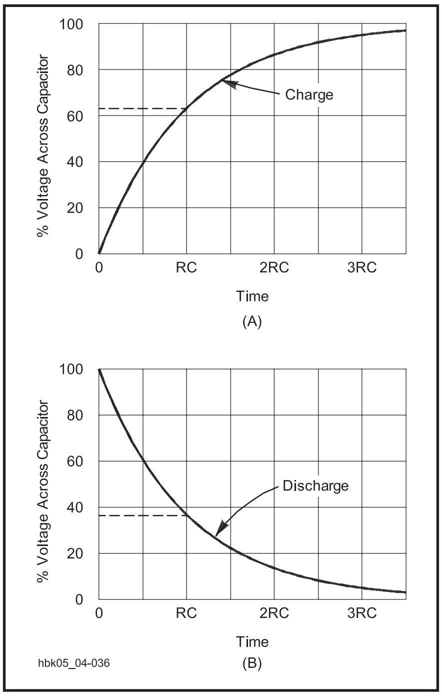


Figure 2.19 — At A, the curve shows how the voltage across a capacitor rises, with time, when charged through a resistor. The curve at B shows the way in which the voltage decreases across a capacitor when discharging through the same resistance. For practical purposes, a capacitor may be considered charged or discharged after five RC periods.

2.6.4 RC Time Constant

An Excel spreadsheet by Lou Ernst, WA2GKH, illustrates charging and discharging time constants. The file name is Capacitor_TC_Calc_Version_2.1r.xls, and it is available in this book's online content, found at the website, **www.arrl.org/arrl-handbook-reference**.

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 **Figure 2.18A)** 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:

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

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 Mathematics" article in this book's online content.)

By letting t = RC, the above equation becomes:

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

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 **Figure 2.19A**. 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 Figure 2.18B, 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:

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

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%

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

of the full-charge voltage. Figure 2.19B 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,

$t = 5 \times RC = 5 \times 100 \times 10^3\ \Omega \times 100 \times 10^{-6}\ F = 50000 \times 10^{-3} = 50\ s$

Caution: 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 Sources** chapter for more information on bleeder resistors.)

Example 2: Smooth CW keying without clicks requires both the rising and falling edges of the waveform to take approximately 5 ms (0.005 s). If an RC delay circuit in a keyed voltage line is used to set the rise and fall time, what values of R and C should be used? 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: The popular 555 timer IC activates its output pin when the trigger input reaches 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 the charging equation 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.

2.7 Inductance and Inductors

A complete treatment of magnetism and magnetic properties of materials can be found in *Electricity and Magnetism* by Edward Purcell (see **References and Bibliography**). This section is limited to a discussion of inductance, inductors, and related materials.

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.

2.7.1 Magnetic Fields and Magnetic Energy Storage

As an electric field surrounds an electric charge, magnetic fields surround *magnets*. You are probably familiar with metallic bar, disc, or horseshoe-shaped magnets. **Figure 2.20** shows a bar magnet, but particles of matter as small as an atom can also be magnets.

Figure 2.20 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.

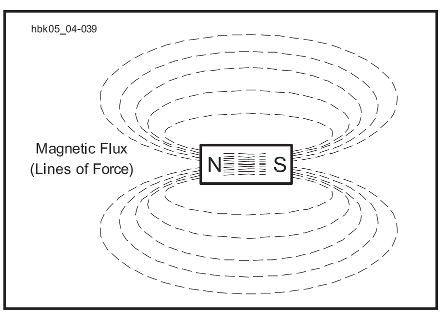


Figure 2.20 — The magnetic field and poles of a permanent magnet. The magnetic field direction is from the north to the south pole.

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 **Figure 2.21**, 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 **Figure 2.22**, the magnetic field greatly intensifies. This occurs as the magnetic fields from each successive turn in the coil add together 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 connection: 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. **Table 2.3** shows the similarities: magnetic quantities and circuits have analogues in electrical quantities and circuits. This illustrates the deep relationship between electricity and magnetism, two sides of the same coin.

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 the figure below. 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!)

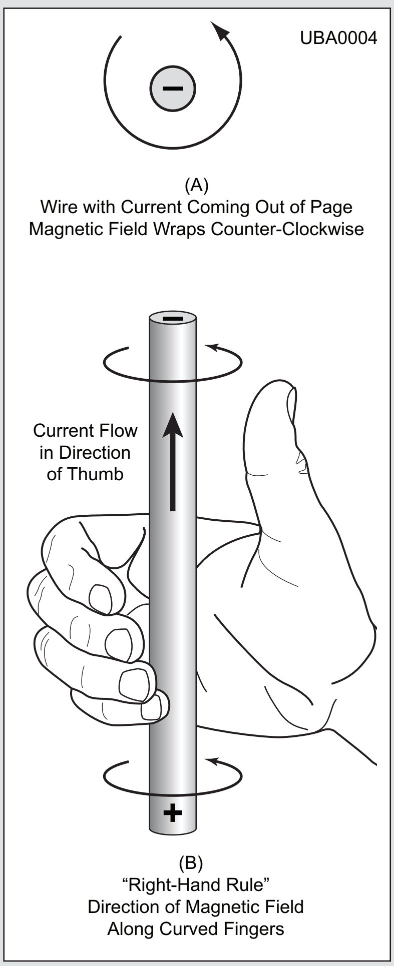


Figure 2.A2 — Use the right-hand rule to determine magnetic field direction from the direction of current flow.

MAGNETIC FLUX

Magnetic flux is measured in the SI unit (International System of Units) of the weber,

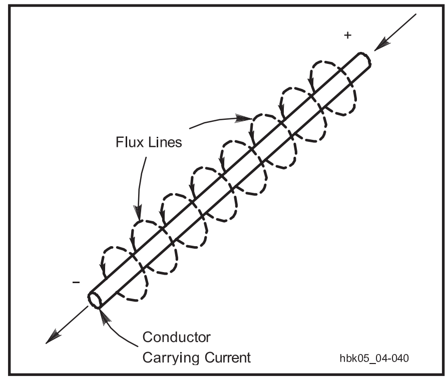


Figure 2.21 — The magnetic field around a conductor carrying an electrical current. If the thumb of your right hand points in the direction of the conventional current (plus to minus), your fingers curl in the direction of the magnetic field around the wire.

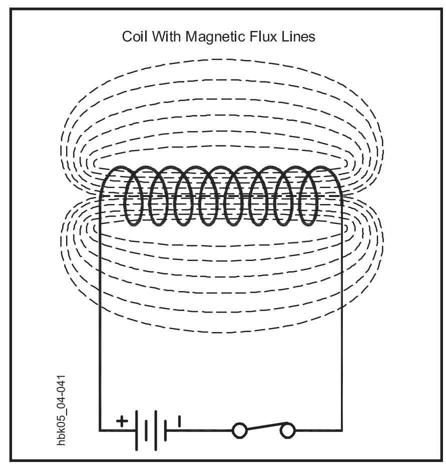


Figure 2.22 — Cross section of an inductor showing its flux lines and overall magnetic field.

Table 2.3

Magnetic Quantities

Value	*Symbol*	*MKS*	*cgs*
Magnetic Flux	lines	Weber, Wb = V-s	Maxwell, Mx = 10^{-8} Wb
Magnetic Flux Density	B	Tesla, T = Wb/m^2	Gauss, G = Mx/ cm^2 T = 10,000 G
Magnetomotive Force	[T]	Amp-turn = A	Gilbert, Gb = 0.79577 A
Magnetic Field Strength	H	A / meter	Oersted, Oe = Gb/cm = 79.58 A/m

Magnetic Circuit Analogies

Electric Circuit	*Magnetic Circuit*
Voltage drop V	Hl magnetovoltage drop
Voltage source V	nI magnetomotive force
Current I	psi = BA magnetic flux

Note – Magnetic circuit analogies as described by Shen and Kong, *Applied Electromagnetism* which is a volt-second (Wb = V-s). 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 Earth's magnetic field at the surface is approximately one gauss. 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)

$$\Im = \frac{10\, N\, I}{4\pi}$$

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.

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

$$H = \frac{\Im}{\ell} = \frac{10\, N\, I}{4\, \pi\, \ell}$$

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)

2.7.2 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 **Figure 2.23** 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.

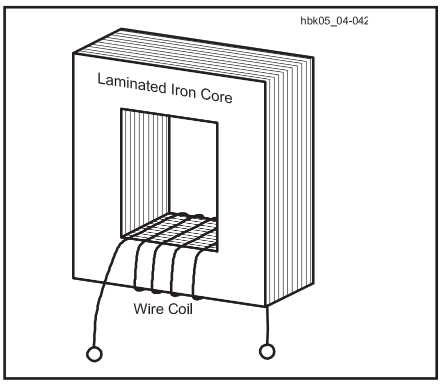


Figure 2.23 — A coil of wire wound around a laminated iron core.

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

A) stacks of thin steel laminations (for 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 5,000 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 15,000. 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 2.4 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:

Table 2.4
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	—	—	—	—	180,000
4% silicon-iron	96	—	—	—	4 Si	7000
45 Permalloy	54.7	45	—	—	0.3 Mn	25,000
Hipernik	50	50	—	—	—	70,000
78 Permalloy	21.2	78.5	—	—	0.3 Mn	100,000
4-79 Permalloy	16.7	79	—	—	0.3 Mn	100,000
Supermalloy	15.7	79	—	5	0.3 Mn	800,000
Permendur	49.7	—	50	—	0.3 Mn	5000
2V Permendur	49	—	49	—	2 V	4500
Hiperco	64	—	34	—	2 Cr	10,000
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.)

$$\mu = \frac{B}{H}$$

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 inversely proportional to 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. There are two related properties. *Reluctivity* is the reciprocal of permeability, and *permeance* is the reciprocal of reluctance.

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.

Figure 2.24 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 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 the 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 Figure 2.24 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.

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.

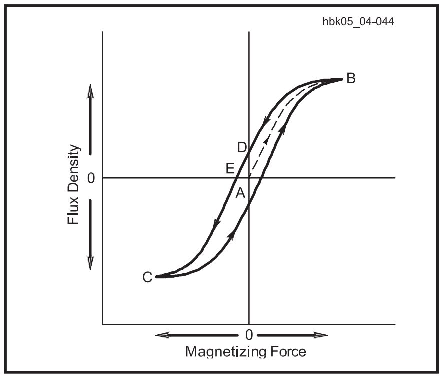


Figure 2.24 — A typical hysteresis curve for a magnetic core, showing the additional energy needed to overcome residual flux.

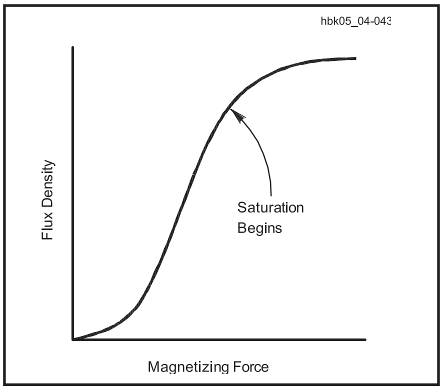


Figure 2.25 — A typical permeability curve for a magnetic core, showing the point where saturation begins.

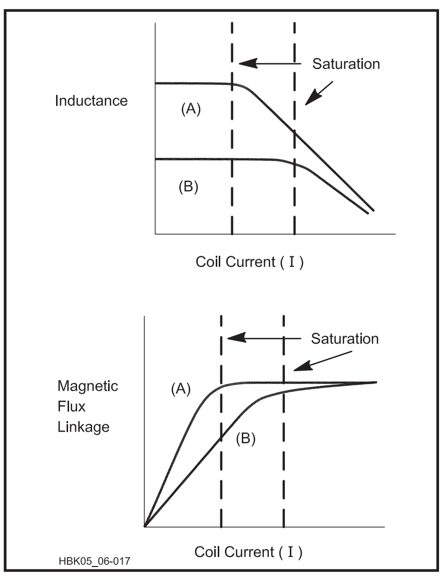


Figure 2.26 — Magnetic flux linkage and inductance plotted versus coil current for (A) a typical iron-core inductor. As the flux linkage Nϕ in the coil saturates, the inductance begins to decrease since inductance = flux linkage / current. The curves marked B show the effect of adding an air gap to the core. The current-handling capability has increased, but at the expense of reduced inductance.

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. **Figure 2.25** 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.

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. **Figure 2.26** 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

$$\psi = N\phi = LI$$

where

ψ = the flux linkage,

N = number of turns,

ϕ = flux 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 **Figure 2.27**). 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

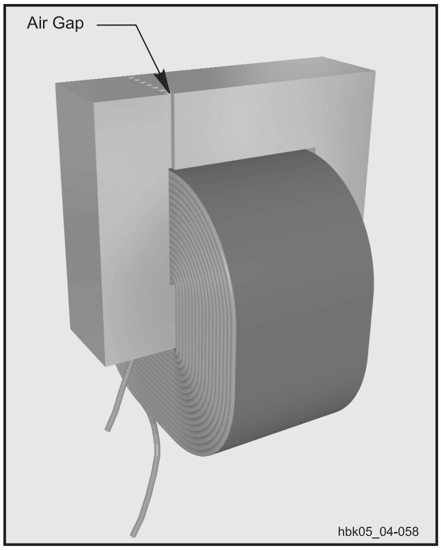


Figure 2.27 — Typical construction of a magnetic-core inductor. The air gap greatly reduces core saturation at the expense of reducing inductance. The insulating laminations between the core layers help to minimize eddy currents, as well.

works by requiring a higher current to achieve saturation. The price that is paid is a reduced inductance below saturation. The curves in Figure 2.26B 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:

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

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

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).

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 in the **Circuits and Components** chapter.)

2.7.3 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. **Figure 2.28** 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.

> **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.

The induced voltage opposes the voltage of the source, preventing the current from rising rapidly when voltage is applied. **Figure 2.29A** 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, distinguishing it from *mutual inductance*, as described below.) The basic unit of inductance is the *henry* (abbreviated H).

$$V = -L\frac{\Delta I}{\Delta t}$$

where

V is the induced voltage in volts,
L is the inductance in henries, and
ΔI/Δt is the rate of change of the current in amperes per second.

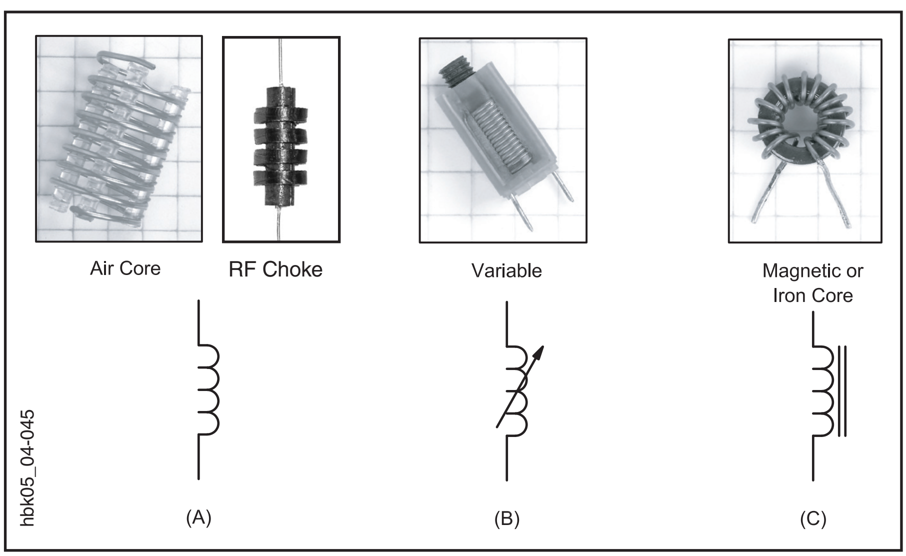


Figure 2.28 — Photos and schematic symbols for representative inductors. A, an air core inductor and a pie-wound RF choke; B, a variable inductor with a nonmagnetic slug, and C, an inductor with a toroidal magnetic core. The ¼-inch-ruled graph paper background provides a size comparison.

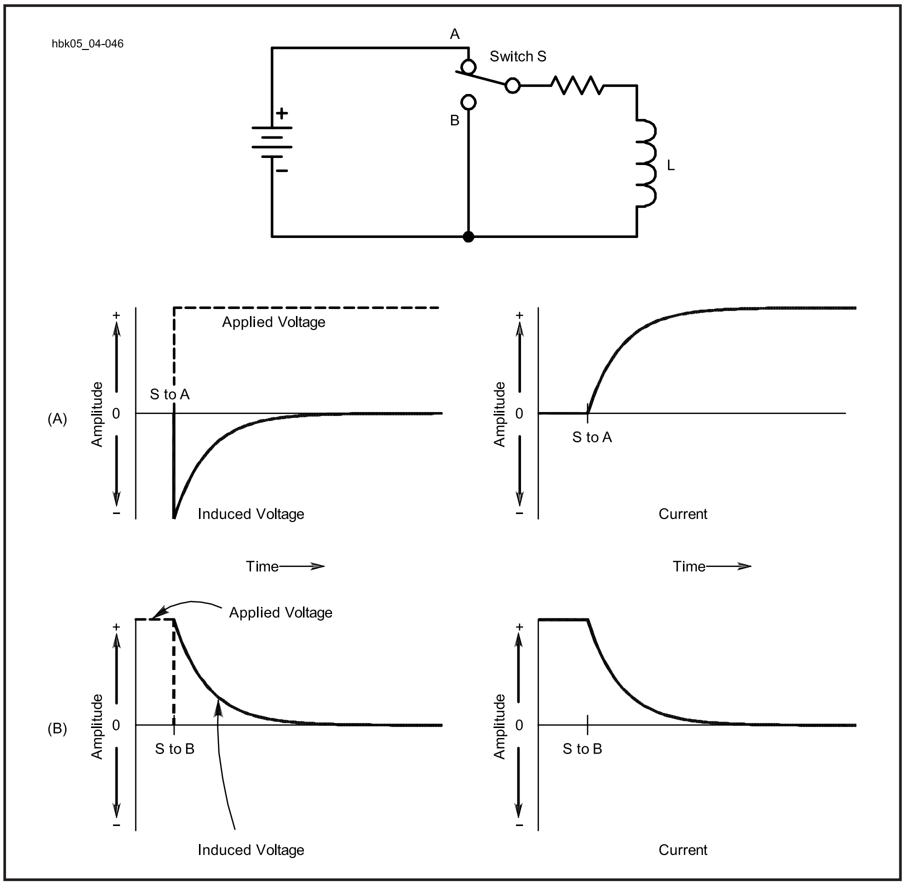


Figure 2.29 — Inductive circuit showing the generation of induced voltage and the rise of current when voltage is applied to an inductor at A, and the decay of current as the coil shorted at B.

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 minus sign (–) indicates that the induced voltage has a polarity opposing the change in current.

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

$$U = \frac{I^2 L}{2}$$

where

U = 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 the equation for magnetic field strength given previously) 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. Figure 2.29B 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 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.

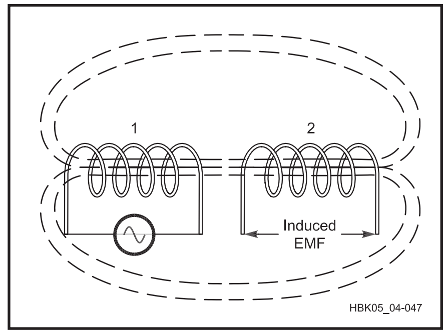


Figure 2.30 — Mutual inductance: When S is closed, current flows through coil number 1, setting up a magnetic field that induces a voltage in the turns of coil number 2.

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.

2.7.4 Mutual Inductance and Magnetic Coupling

When two inductors are arranged with their axes aligned as shown in **Figure 2.30**, 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.

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 Figure 2.30, 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 Figure 2.28C, 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.

2.7.5 Inductances in Series and Parallel

When two or more inductors are connected in series (left side of **Figure 2.31**), the total inductance is equal to the sum of the individual inductances, provided that the inductors are sufficiently separated that there is no coupling between them (see the preceding section):

$$L_{total} = L1 + L2 + L3 \ldots + L_n$$

If inductors are connected in parallel (right side of Figure 2.31), again assuming no mutual coupling, the total inductance is given by:

$$L_{total} = \frac{1}{\frac{1}{L1} + \frac{1}{L2} + \frac{1}{L3} + \ldots + \frac{1}{L_n}}$$

For only two inductors in parallel, the formula becomes:

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

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.

2.7.6 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 **Figure 2.32**, first consider the case in which R is zero. Closing S1 sends a current 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 Figure 2.32 has been closed can be found from:

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

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.

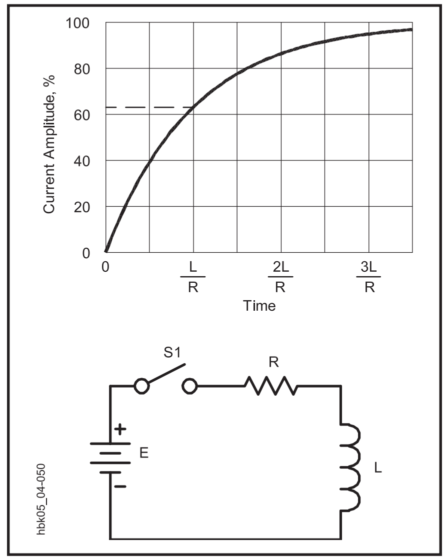


Figure 2.32 — Time constant of an RL circuit being energized.

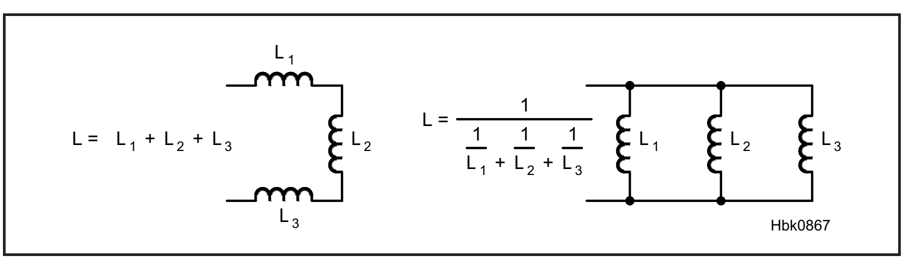


Figure 2.31 — Inductors in series and parallel.

(References that explain exponential equations, e, and other mathematical topics are found in the "Radio Mathematics" article in this book's online information.) 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:

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

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, the current increases by an additional 63.2% closer to the final value of E/R. This behavior is graphed in Figure 2.32. As is the case with capacitors, after five time constants the current is considered to have reached its maximum value. As with capacitors, we 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,

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

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

Note that if the inductance is increased to 5.0 H, the required time increases by a factor of 1,000 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 Figure 2.32 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 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 **Figure 2.33**, the current will decay according to the formula:

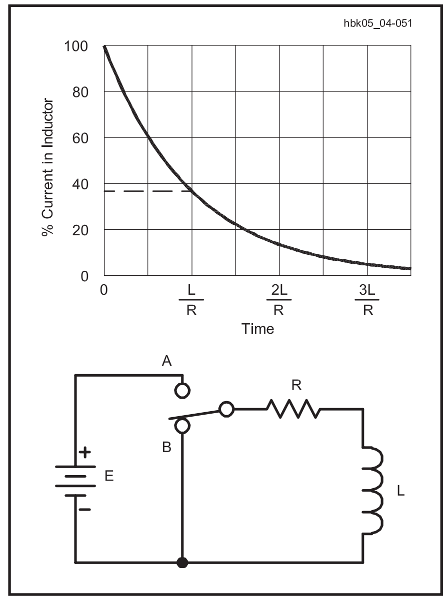


Figure 2.33 — Time constant of an RL circuit being de-energized. This is a theoretical model only, since a mechanical switch cannot change state instantaneously.

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

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 Figure 2.33 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 Semiconductor Devices

2.8.1 Introduction to Semiconductors

In a conductor, such as a metal, some of the outer, or *valence*, electrons of each atom are free to move about between atoms. These *free electrons* are the constituents of electrical current. In a good conductor, the concentration of these free electrons is very high, on the order of 10^{22} electrons/cm^3. In an insulator, nearly all the electrons are tightly held by their atoms and the concentration of free electrons is very small — on the order of 10 electrons/cm^3.

Between the classes of materials considered to be conductors and insulators is a class of elements called *semiconductors*, materials with conductivity much poorer than metals and much better than insulators. (In electronics, "semiconductor" means a device made from semiconductor elements that have been chemically manipulated as described below, leading to interesting properties that create useful applications.)

Semiconductor atoms (silicon, Si, is the most widely used) share their valence electrons in a chemical bond that holds adjacent atoms together, forming a three-dimensional *lattice* that gives the material its physical characteristics. A lattice of pure semiconductor material (one type of atom or molecule) can form a crystal, in which the lattice structure and orientation is preserved throughout the material. *Monocrystalline* or "single-crystal" is the type of material used in electronic semiconductor devices. *Polycrystalline* material is made of up many smaller crystals with their own individual lattice orientations.

MAJORITY AND MINORITY CARRIERS

Crystals of pure semiconductor material are called *intrinsic* semiconductors. When energy, generally in the form of heat, is added to a semiconductor crystal lattice, some electrons are liberated from their bonds and move freely throughout the lattice. The bond that loses an electron is then unbalanced and the space that the electron came from is referred to as a *hole*. In these materials the number of free electrons is equal to the number of holes.

Electrons from adjacent bonds can leave their positions to fill the holes, thus leaving behind a hole in their old location. As a consequence of the electron moving, two opposite movements can be said to occur: negatively charged electrons move from bond to bond in one direction and positively charged holes move from bond to bond in the opposite direction. Both of these movements represent forms of electrical current, but this is very different from the current in a conductor. While a conductor has free electrons that flow independently from the bonds of the crystalline lattice, the current in a pure semiconductor is constrained to move from bond to bond.

MATERIAL DOPING

Impurities can be added to intrinsic semiconductors (by a process called *doping*) to enhance the formation of electrons or holes and thus improve conductivity. These materials are *extrinsic* semiconductors. Since the additional electrons and holes can move, their movement is current and they are called *carriers*. The type of carrier that predominates in the material is called the *majority carrier*. In N-type material the majority carriers are electrons and in P-type material, holes.

There are two types of impurities that can be added: a *donor impurity* with five valence electrons donates free electrons to the crystalline structure; this is called an *N-type* impurity, for the negative charge of the majority carriers. Some examples of donor impurities are antimony (Sb), phosphorus (P), and arsenic (As). N-type extrinsic semiconductors have more electrons and fewer holes than intrinsic semiconductors. *Acceptor impurities* with three valence electrons accept free electrons from the lattice, adding holes to the overall structure. These are called P-type impurities, for the positive charge of the majority carriers; some examples are boron (B), gallium (Ga), and indium (In).

It is important to note that even though N-type and P-type material have different numbers of holes and free electrons than intrinsic material, they are still electrically neutral. When an electron leaves an atom, the positively charged atom that remains in place in the crystal lattice electrically balances the roaming free electron. Similarly, an atom gaining an electron acquires a negative charge that balances the positively charged atom it left. At no time does the material acquire a net electrical charge, positive or negative.

Compound semiconductor material can be formed by combining equal amounts of N-type and P-type impurity materials. Some examples of this include gallium-arsenide (GaAs), gallium-phosphate (GaP), and indium-phosphide (InP). To make an N-type compound semiconductor, a slightly higher amount of N-type material is used in the mixture. A P-type compound semiconductor has a little more P-type material in the mixture.

Impurities are introduced into intrinsic semiconductors by diffusion, the same physical process that lets you smell cookies baking from several rooms away. (Molecules diffuse through air much faster than through solids.) Rates of diffusion are proportional to temperature, so semiconductors are doped with impurities at high temperature to save time. Once the doped semiconductor material is cooled, the rate of diffusion of the impurities is so low that they are essentially immobile for many years to come. If an electronic device made from a structure of N- and P-type materials is raised to a high temperature, such as by excessive current, the impurities can again migrate and the internal structure of the device may be destroyed. The maximum operating temperature for semiconductor devices is specified at a level low enough to limit additional impurity diffusion.

The conductivity of an extrinsic semiconductor depends on the charge density (in other words, the concentration of free electrons in N-type, and holes in P-type, semiconductor material). As the energy in the semiconductor increases, the charge density also increases. This is the basis of how all semiconductor devices operate: the major difference is the way in which the energy level is increased. Variations include: The *transistor*, where conductivity is altered by injecting current into the device via a wire; the *thermistor*, where the level of heat in the device is detected by its conductivity, and the *photoconductor*, where light energy that is absorbed by the semiconductor material increases the conductivity.

2.8.2 The PN Semiconductor Junction

If a piece of N-type semiconductor material is placed against a piece of P-type semiconductor material, the location at which they join is called a *PN junction*. The junction has characteristics that make it possible to develop diodes and transistors. The action of the junction is best described by a diode operating as a rectifier.

Initially, when the two types of semiconductor material are placed in contact, each type of material will have only its majority carriers: P-type will have only holes and N-type will have only free electrons. The presence of the positive charges (holes) in the P-type material attracts free electrons from the N-type material immediately across the junction. The opposite is true in the N-type material.

These attractions lead to diffusion of some of the majority carriers across the junction, which combine with and neutralize the majority carriers immediately on the other side (a process called *recombination*). As distance from the junction increases, the attraction quickly becomes too small to cause the carriers to move. The region close to the junction is then *depleted* of carriers, and so is named the *depletion region* (also the *space-charge region* or the *transition region*). The width of the depletion region is very small, on the order of 0.5 µm.

If the N-type material (the *cathode*) is placed at a more negative voltage than the P-type material (the *anode*), current will pass through the junction because electrons are attracted from the lower potential to the higher potential and holes are attracted in the opposite direction. This *forward bias* forces the majority carriers toward the junction where recombination occurs with the opposite type of majority carrier. The source of voltage supplies replacement electrons to the N-type material and removes electrons from the P-type material so that the majority carriers are continually replenished. Thus, the net effect is a *forward current* flowing through the semiconductor, across the PN junction. The *forward resistance* of a diode conducting current is typically very low and varies with the amount of forward current.

When the polarity is reversed, majority-carriers are attracted away from the junction, not toward it. Very little current flows across the PN junction — called *reverse leakage current* — in this case. Allowing only unidirectional current flow is what allows a semiconductor diode to act as rectifier.

2.8.3 Junction Semiconductors

Semiconductor devices that operate using the principles of a PN junction are called *junction semiconductors*. These devices can have one or several junctions. The properties of junction semiconductors can be tightly controlled by the characteristics of the materials used and the size and shape of the junctions.

SEMICONDUCTOR DIODES

Diodes are commonly made of silicon and occasionally germanium. Although they act similarly, they have slightly different characteristics. The *junction threshold voltage*, or *junction barrier voltage,* is the forward bias voltage (V_F) at which current begins to pass through the device. This voltage is different for the two kinds of diodes. In the diode response curve of **Figure 2.34**, V_F corresponds to the voltage at which the positive portion of the curve begins to rise sharply from the x-axis. Most silicon diodes have a junction threshold voltage of about 0.7 V, while the voltage for germanium diodes is typically 0.3 V. Reverse leakage current is much lower for silicon diodes than for germanium diodes.

The characteristic curve for a semiconductor diode junction is given by the following equation (slightly simplified) called the *Fundamental Diode Equation* because it describes the behavior of all semiconductor PN junctions.

$$I = I_S\left(e^{\frac{V}{\eta V_t}} - 1\right)$$

where

I = diode current,

V = diode voltage,

I_s = reverse-bias saturation current,

V_t = kT/q, the thermal equivalent of voltage (about 25 mV at room temperature), and

η = emission coefficient

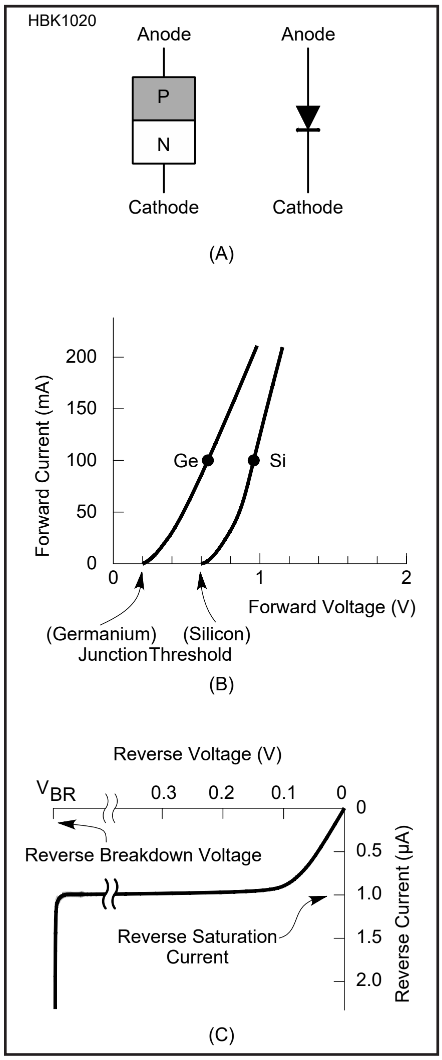


Figure 2.34 — Semiconductor diode (PN junction) structure (A). Characteristic curve (B) shows forward-biased (anode positive with respect to cathode) response for germanium (Ge) and silicon (Si) devices. Each curve breaks away from the X-axis at its junction threshold voltage. The slope of each curve is its forward resistance. (C) Reverse-biased response. Very small reverse current increases until it reaches the reverse saturation current (I_0). The reverse current increases suddenly and drastically when the reverse voltage reaches the reverse breakdown voltage, V_{BR}.

The value of I_s varies with the type of semiconductor material, with the value of 10^{-12} used for silicon. η also varies from 1 to 2 with the type of material and method of fabrication. (η is close to 1 for silicon at normal current values, increasing to 2 at high currents.) This curve is shown in **Figure 2.35B**.

The obvious differences between Figure 2.35A and B are that the semiconductor diode has a finite *turn-on* voltage — it requires a small but nonzero forward bias voltage before it begins conducting. Furthermore, once conducting, the diode voltage continues to increase very slowly with increasing current, unlike a true short circuit. Finally, when the applied voltage is negative, the reverse current is not exactly zero but very small (microamperes). The reverse current flow rapidly reaches a level that varies little with the reverse bias voltage. This is the *reverse-bias saturation current*, I_s.

For bias (dc) circuit calculations, a useful model for the diode that takes these two effects into account is shown by the artificial I-V curve in Figure 2.35C. This model neglects the negligible reverse bias current I_s.

When converted into an equivalent circuit, the model in Figure 2.35C yields the circuit in Figure 2.35D. The ideal voltage source V_a represents the turn-on voltage and R_f represents the effective resistance caused by the small increase in diode voltage as the diode current increases. The turn-on voltage is material-dependent: approximately 0.3 V

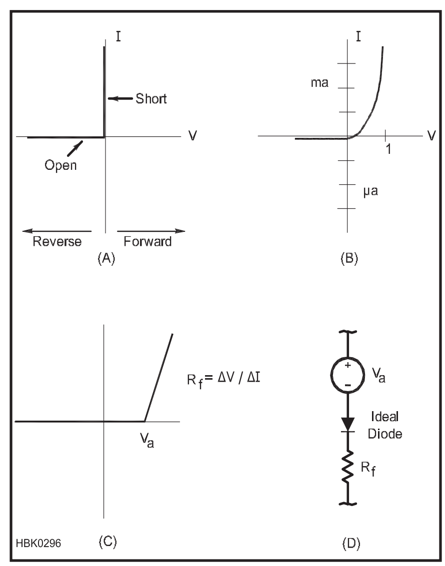


Figure 2.35 — Circuit models for rectifying switches (diodes). A: I-V curve of the ideal rectifier. B: I-V curve of a typical semiconductor diode showing the typical small leakage current in the reverse direction. Note the different scales for forward and reverse current. C shows a simplified diode I-V curve for dc-circuit calculations (at a much larger scale than B). D is an equivalent circuit for C.

for germanium diodes and 0.7 for silicon. R_f is typically on the order of 10 Ω, but it can vary according to the specific component. R_f can often be completely neglected in comparison to the other resistances in the circuit. This very common simplification leaves only a pure *voltage drop* for the diode model.

THYRISTORS

Thyristors are semiconductors made with four or more alternating layers of P- and N-type semiconductor material. In a four-layer thyristor, when the anode is at a higher potential than the cathode, the first and third junctions are forward biased and the center junction reverse biased. In this state, there is little current, just as in the reverse-biased diode. The different types of thyristor have different ways in which they turn on to conduct current and in how they turn off to interrupt current flow.

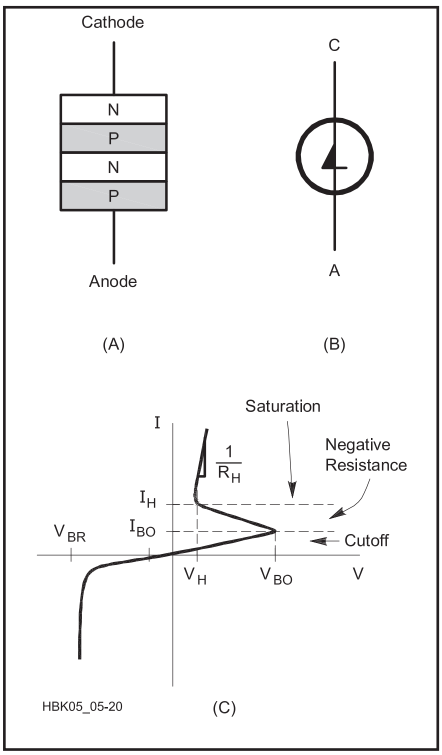


Figure 2.36 — PNPN diode. (A) Alternating layers of P-type and N-type semiconductor. (B) Schematic symbol with cathode (C) and anode (A) leads. (C) I-V curve. Reverse-biased response is the same as normal PN junction diodes. Forward biased response acts as a hysteresis switch. Resistance is very high until the bias voltage reaches V_{BO} (where the center junction breaks over) and exceeds the cutoff current, I_{BO}. The device exhibits a negative resistance when the current increases as the bias voltage decreases until a voltage of V_H and saturation current of I_H is reached. After this, the resistance is very low, with large increases in current for small voltage increases.

PNPN Diode

The simplest thyristor is a PNPN (usually pronounced like *pinpin*) diode with three junctions (see **Figure 2.36**). As the forward bias voltage is increased, the current through the device increases slowly until the *breakover* (or *firing)* voltage, V_{BO}, is reached and the flow of current abruptly increases. The PNPN diode is often considered to be a switch that is off below V_{BO} and on above it.

Bilateral Diode Switch (Diac)

A semiconductor device similar to two PNPN diodes facing in opposite directions and attached in parallel is the *bilateral diode switch* or *diac*. This device has the characteristic curve of the PNPN diode for both positive and negative bias voltages. Its construction, schematic symbol and characteristic curve are shown in **Figure 2.37**.

Silicon Controlled Rectifier (SCR)

Another device with four alternate layers of P-type and N-type semiconductor is the *silicon controlled rectifier (SCR)*. (Some sources refer to an SCR as a thyristor, as well.) In addition to the connections to the outer two layers, two other terminals can be brought out for the inner two layers. The connection to the P-type material near the cathode is called the *cathode gate* and the N-type material near the anode is called the *anode gate*. In nearly all commercially available SCRs, only the cathode gate is connected (**Figure 2.38**).

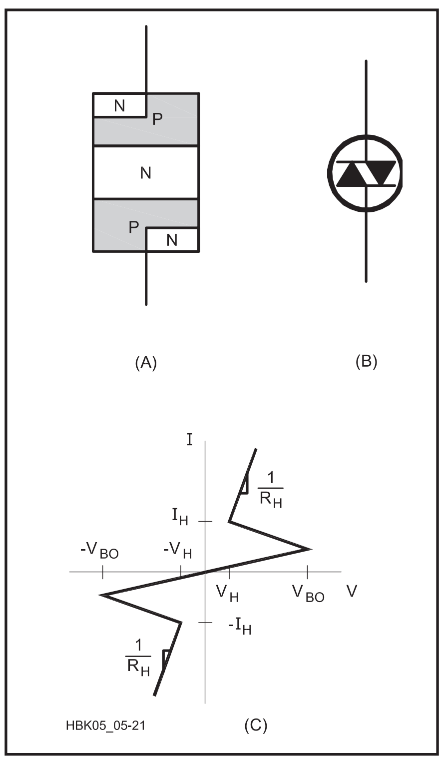


Figure 2.37 — Bilateral switch. (A) Alternating layers of P-type and N-type semiconductor. (B) Schematic symbol. (C) I-V curve. The right-hand side of the curve is identical to the PNPN diode response in Figure 2.36. The device responds identically for both forward and reverse bias so the left-hand side of the curve is symmetrical with the right-hand side.

Like the PNPN diode switch, the SCR is used to abruptly start conducting when the voltage exceeds a given level. By biasing the gate terminal appropriately, the breakover voltage can be adjusted.

Triac

A five-layered semiconductor whose operation is similar to a bidirectional SCR is the *triac* (**Figure 2.39**). This is also similar to a bidirectional diode switch with a bias control gate. The gate terminal of the triac can control both positive and negative breakover voltages, and the devices can pass both polarities of voltage.

Thyristor Applications

The SCR is highly efficient and is used in power control applications. SCRs are available that can handle currents of greater than 100 A and voltage differentials of greater

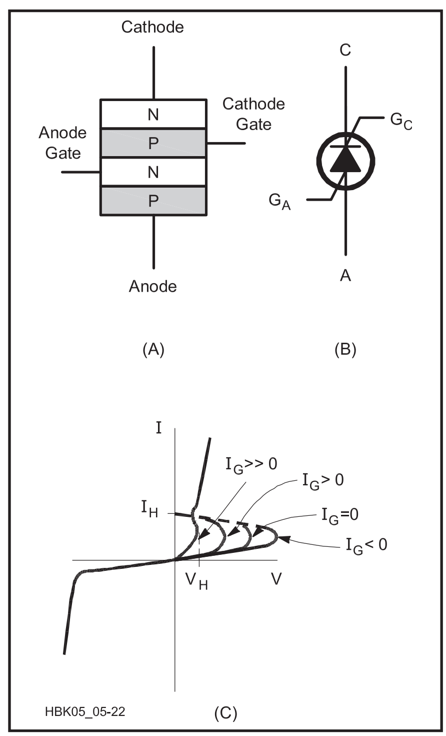


Figure 2.38 — SCR. (A) Alternating layers of P-type and N-type semiconductor. This is similar to a PNPN diode with gate terminals attached to the interior layers. (B) Schematic symbol with anode (A), cathode (C), anode gate (G_A) and cathode gate (G_C). Many devices are constructed without G_A. (C) I-V curve with different responses for various gate currents. I_G = 0 has a similar response to the PNPN diode.

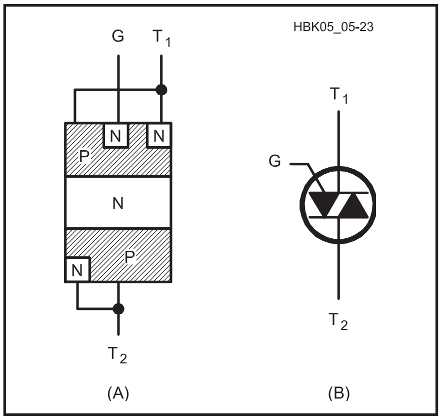


Figure 2.39 — Triac. (A) Alternating layers of P-type and N-type semiconductor. This behaves as two SCR devices facing in opposite directions with the anode of one connected to the cathode of the other and the cathode gates connected together. (B) Schematic symbol.

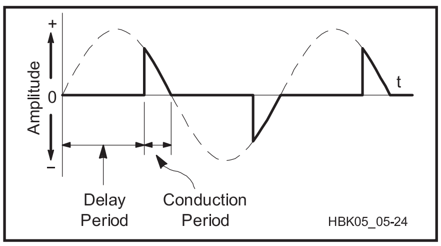


Figure 2.40 — Triac operation on sine wave. The dashed line is the original sine wave and the solid line is the portion that conducts through the triac. The relative delay and conduction period times are controlled by the amount or timing of gate current, I_G. The response of an SCR is the same as this for positive voltages (above the X-axis) and with no conduction for negative voltages.

than 1000 V, yet can be switched with gate currents of less than 50 mA. Because of their high current-handling capability, SCRs are used as "crowbars" in power supply circuits, to short the output to ground and blow a fuse when an overvoltage condition exists.

SCRs and triacs are often used to control ac power sources. A sine wave with a given RMS value can be switched on and off at preset points during the cycle to decrease the RMS voltage. When conduction is delayed until after the peak (as **Figure 2.40** shows) the peak-to-peak voltage is reduced. If conduction starts before the peak, the RMS voltage is reduced, but the peak-to-peak value remains the same. This method is used to operate light dimmers and 240 V ac to 120 V ac converters. The sharp switching transients created when these devices turn on are common sources of RF interference. (See the chapter on **RF Interference** for information on dealing with interference from thyristors.)

BIPOLAR TRANSISTOR

A *bipolar junction transistor (BJT*) is created by three alternating layers or regions of N- and P-type material as shown in **Figure 2.41**. The designation NPN or PNP describes the order of the regions as shown in the figure.

Physically, we can think of the transistor as two PN junctions back-to-back, such as two diodes connected at their *anodes* (the positive terminal) for an NPN transistor or two diodes connected at their *cathodes* (the negative terminal) for a PNP transistor. The connection point is the base of the transistor. (You can't actually make a transistor this way — this is a representation for illustration only.)

A transistor conducts when the base-emitter junction is forward biased and the base-collector is reverse biased. Under these conditions, the emitter region emits majority carriers into the base region, where they become minority carriers because the materials of the emitter and base regions have opposite polarity. The excess minority carriers in the base are then attracted across the very thin base to the base-collector junction, where they are collected and are once again considered majority carriers before they can flow to the base terminal.

The flow of majority carriers from emitter to collector can be modified by the application of a bias current to the base terminal. If the bias current causes majority carriers to be injected into the base material (electrons flowing into an N-type base or out of a P-type base), the emitter-collector current increases. In this way, a transistor allows a small base current to control a much larger collector current.

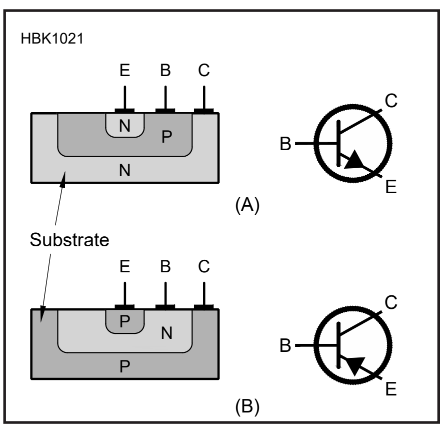


Figure 2.41 — Bipolar junction transistors (BJT) are usually constructed as three regions of N- or P-type material with a substrate at the bottom, the base region embedded into the substrate, and the emitter region embedded in the base. The order of the layers determines whether the device is NPN (A) or PNP (B). There are three electrodes: collector (C), base (B), and emitter (E). In the schematic symbol for an PNP device, the arrow points toward the base and away from the base for a NPN device.

As in a semiconductor diode, the forward biased base-emitter junction has a threshold voltage (V_{BE}) that must be exceeded before the emitter current increases. As the base-emitter current continues to increase, the point is reached at which further increases in base-emitter current cause no additional change in collector current. This is the condition of *saturation*. Conversely, when base-emitter current is reduced to the point at which collector current ceases to flow, that is the situation of *cutoff*.

2.8.4 Field-Effect Transistors (FET)

The *field-effect transistor (FET)* controls the current between two points but does so differently than the bipolar transistor. The FET operates by the effects of an electric field on the flow of electrons through a single type of semiconductor material. This is why the FET is sometimes called a *unipolar* transistor. Unlike bipolar semiconductors that can be arranged in many configurations to provide diodes, transistors, photoelectric devices, temperature sensitive devices, and so on, the field effect technique is usually only used to make transistors, although FETs are also available as special-purpose diodes, for use as constant current sources.

FET devices are constructed on a *substrate* of doped semiconductor material. The channel is formed within the substrate and has the opposite polarity (a P-channel FET has N-type substrate). Most FETs are constructed with silicon.

Within the FET, current moves in a *channel* as shown in **Figure 2.42**. The channel is made

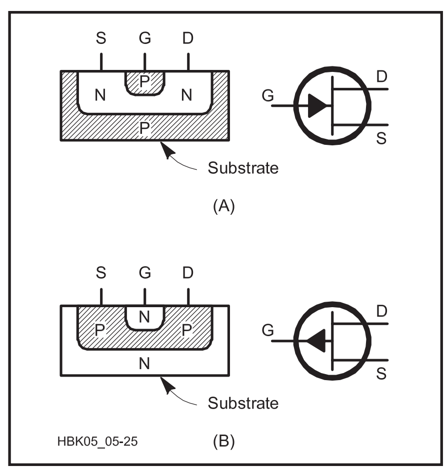


Figure 2.42 — JFET devices with terminals labeled: source (S), gate (G) and drain (D). (A) Pictorial of N-type channel embedded in P-type substrate and schematic symbol. (B) P-channel embedded in N-type substrate and schematic symbol.

of either N-type or P-type semiconductor material; an FET is specified as either an N-channel or P-channel device. Current flows from the *source* terminal (where majority carriers are injected) to the *drain* terminal (where majority carriers are removed). A *gate* terminal generates an electric field that controls the current in the channel.

In N-channel devices, the drain potential must be higher than that of the source ($V_{DS} > 0$) for electrons (the majority carriers) to flow in the channel. In P-channel devices, the flow of holes requires that $V_{DS} < 0$. The polarity of the electric field that controls current in the channel is determined by the majority carriers of the channel, ordinarily positive for P-channel FETs and negative for N-channel FETs.

Variations of FET technology are based on different ways of generating the electric field. In all of these, however, electrons at the gate are used only for their charge in order to create an electric field around the channel. There is a minimal flow of electrons through the gate. This leads to a very high dc input resistance in devices that use FETs for their input circuitry. There may be quite a bit of capacitance between the gate and the other FET terminals, however, causing the input impedance to be quite low at high frequencies.

The current through an FET only has to pass through a single type of semiconductor material. Depending on the type of material and the construction of the FET, drain-source resistance when the FET is conducting ($r_{DS(ON)}$) may be anywhere from a few hundred ohms to much less than an ohm. The output impedance of devices made with FETs is generally quite low. If a gate bias voltage is added to operate the transistor near cutoff, the circuit output impedance may be much higher.

In order to achieve a higher gain-bandwidth product, other materials have been used. Gallium-arsenide (GaAs) has *electron mobility* and *drift velocity* (both are measures of how easily electrons are able to move through the crystal lattice) far higher than the standard doped silicon. Amplifiers designed with GaAsFET devices operate at much higher frequencies and with a lower noise factor at VHF and UHF than those made with silicon FETs (although silicon FETs have improved dramatically in recent years).

JFET

One of two basic types of FET, the *junction FET (JFET)* gate material is made of the opposite polarity semiconductor to the channel material (for a P-channel FET the gate is made of N-type semiconductor material). The gate-channel junction is similar to a diode's PN junction with the gate material in direct contact with the channel. JFETs are used with the junction reverse-biased, since any current in the gate is undesirable. The reverse bias of the junction creates an electric field that "pinches" the channel. Since the magnitude of the electric field is proportional to the reverse-bias voltage, the current in the channel is reduced for higher reverse gate bias voltages. When current in the channel is completely halted by the electric field, this is called *pinch-off,* which is analogous to cutoff in a bipolar transistor. The channel in a JFET is at its maximum conductivity when the gate and source voltages are equal ($V_{GS} = 0$).

Because the gate-channel junction in a JFET is similar to a bipolar junction diode, this junction must never be forward biased; otherwise large currents will pass through the gate and into the channel. For an N-channel JFET, the gate must always be at a lower potential than the source ($V_{GS} < 0$). The prohibited condition is for $V_{GS} > 0$. For P-channel JFETs these conditions are reversed (in normal operation $V_{GS} > 0$ and the prohibited condition is for $V_{GS} < 0$).

MOSFET

Placing an insulating layer between the gate and the channel allows for a wider range of control (gate) voltages and further decreases the gate current (and thus increases the device input resistance). The insulator is typically made of an oxide (such as silicon dioxide, SiO_2). This type of device is called a *metal-oxide-semiconductor FET (MOSFET)* or *insulated-gate FET (IGFET).*

The substrate is often connected to the source internally. The insulated gate is on the opposite side of the channel from the substrate (see **Figure 2.43**). The bias voltage on the gate terminal either attracts or repels the majority carriers of the substrate across its PN-junction with the channel. This narrows (*depletes*) or widens (*enhances*) the channel, respectively, as V_{GS} changes polarity. For example, in the N-channel enhancement-mode MOSFET, positive gate voltages with respect to the substrate and the source ($V_{GS} > 0$) repel holes from the channel into the substrate, thereby widening the channel and

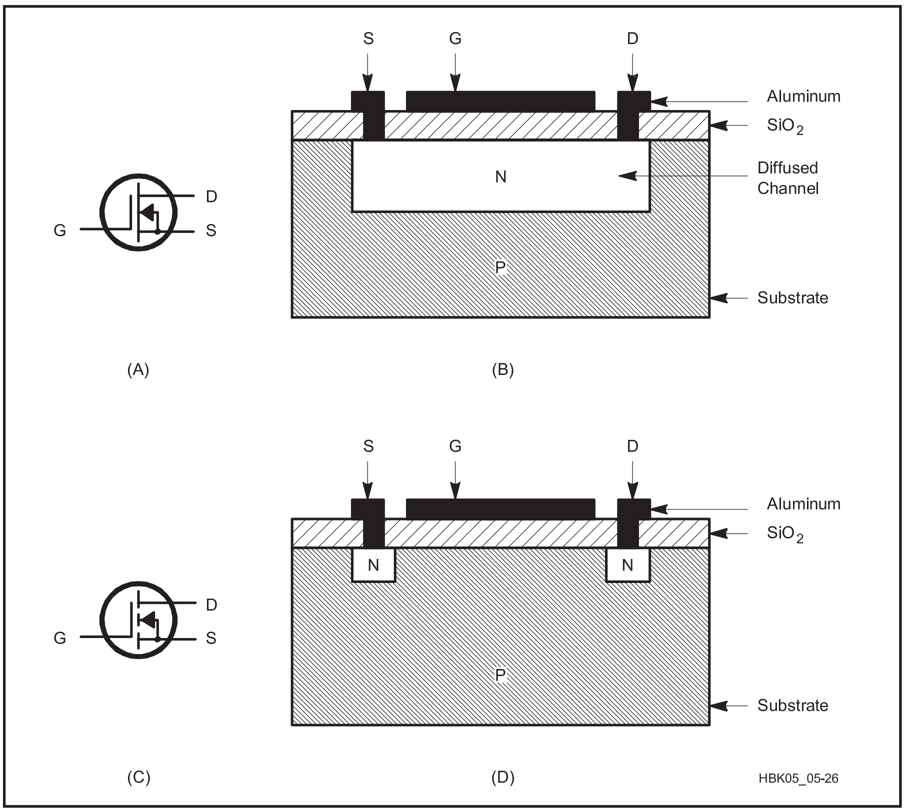


Figure 2.43 — MOSFET devices with terminals labeled: source (S), gate (G) and drain (D). N-channel devices are pictured. P-channel devices have the arrows reversed in the schematic symbols and the opposite type semiconductor material for each of the layers. (A) N-channel depletion mode device schematic symbol and (B) pictorial of P-type substrate, diffused N-type channel, SiO_2 insulating layer, and aluminum gate region and source and drain connections. The substrate is connected to the source internally. A negative gate potential narrows the channel. (C) N-channel enhancement mode device schematic and (D) pictorial of P-type substrate, N-type source and drain wells, SiO_2 insulating layer, and aluminum gate region and source and drain connections. Positive gate potential forms a channel between the two N-type wells by repelling the P-carriers away from the channel region in the substrate.

decreasing channel resistance. Conversely, $V_{GS} < 0$ causes holes to be attracted from the substrate, narrowing the channel and increasing the channel resistance. Once again, the polarities discussed in this example are reversed for P-channel devices. The common abbreviation for an N-channel MOSFET is *NMOS*, and for a P-channel MOSFET, *PMOS*.

Because of the insulating layer next to the gate, input resistance of a MOSFET is usually greater than 10^{12} Ω (a million megohms). Since MOSFETs can both deplete the channel, like the JFET, and also enhance it, the construction of MOSFET devices differs based on the channel size in the quiescent state, $V_{GS} = 0$.

A *depletion mode* device (also called a *normally-on MOSFET*) has a channel in the quiescent state that gets smaller as a reverse bias is applied; this device conducts current with no bias applied (see Figure 2.43A and B). An *enhancement mode* device (also called a *normally-off MOSFET*) is built without a channel and does not conduct current when $V_{GS} = 0$; increasing forward bias forms a temporary channel that conducts current (see Figure 2.43C and D).

Complementary Metal Oxide Semiconductors (CMOS)

Power dissipation in a circuit can be reduced to very small levels (on the order of a few nanowatts) by using MOSFET devices in complementary pairs (CMOS). Each amplifier is constructed of a series circuit of MOSFET devices, as in **Figure 2.44**. The gates are tied together for the input signal, as are the drains for the output signal. In saturation and cutoff, only one of the devices conducts. The current drawn by the circuit under no load is equal to the OFF leakage current of either device and the voltage drop across the pair is equal to V_{DD}, so the steady-state power used by the circuit is always equal to $V_{DD} \times I_{D(OFF)}$. Power is only consumed during the switching process, so for ac signals, power consumption is proportional to frequency.

CMOS circuitry could be built with discrete components, but the number of extra parts and the need for the complementary components to be matched has made that an unusual design technique. The low power consumption and ease of fabrication has made CMOS the most common of all IC technologies. Although CMOS is most commonly used in digital integrated circuitry, its low power consumption has also been put to work by manufacturers of analog as well as digital ICs.

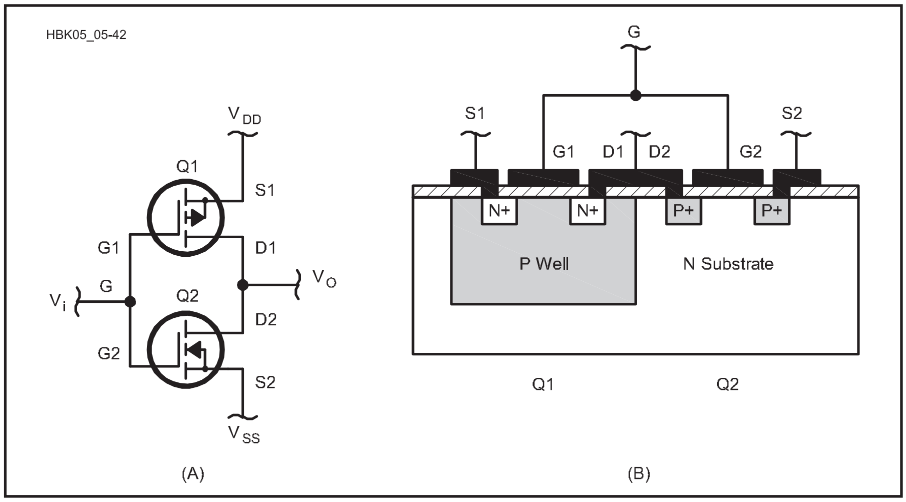


Figure 2.44 — Complementary metal oxide semiconductor (CMOS). (A) CMOS device is made from a pair of enhancement mode MOS transistors. The upper is a P-channel device, and the lower is an N-channel device. When one transistor is biased on, the other is biased off; therefore, there is minimal current from V_{DD} to ground. (B) Implementation of a CMOS pair as an integrated circuit.

2.9 References and Bibliography

BOOKS AND ARTICLES

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

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

Glover, T., *Pocket Ref* 4th ed. (Sequoia Publishing, 2010).

Horowitz and Hill, *The Art of Electronics* 3rd ed. (Cambridge University Press, 2015).

Kaplan, S., *Wiley Electrical and Electronics Dictionary* (Wiley Press, 2004).

Orr, W., *Radio Handbook* 23rd ed. (Newnes, 1997).

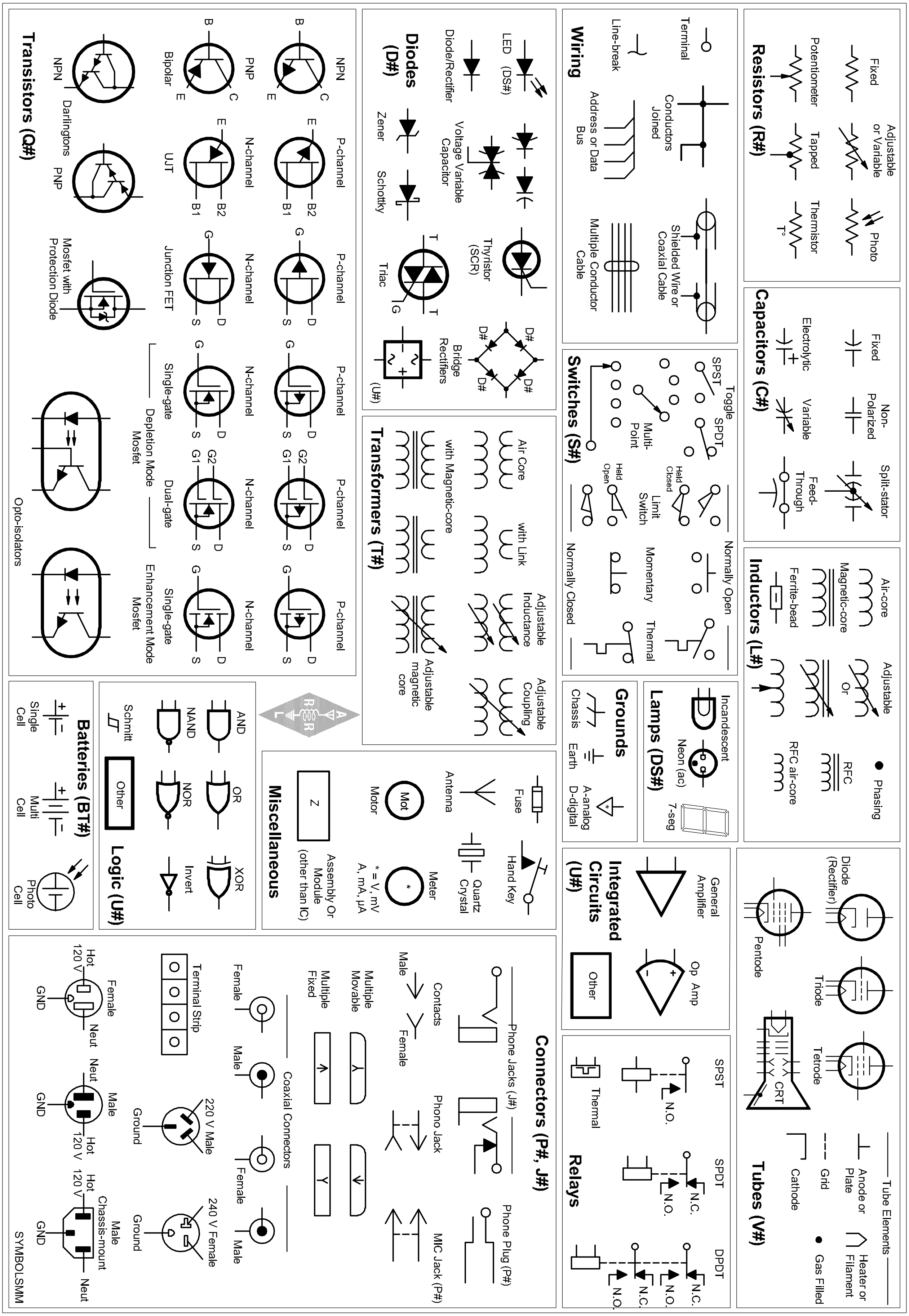
Purcell, E., *Electricity and Magnetism*, 3rd ed. (Cambridge University Press, 2013).

Silver, W., NØAX, "Experiment #117 — Laying Down the Laws," *QST*, Oct 2012, pp. 60-61.

Silver, W., NØAX, "Experiment #118 — The Laws at Work," *QST*, Nov 2012, pp. 70-71.

Silver, W., NØAX, "Experiment #138 — E vs V," *QST*, Jul 2014, pp. 59–60.

Terman, F., *Radio Engineers' Handbook* (McGraw-Hill, 1943).



Common Schematic Symbols Used in Circuit Diagrams
Resistors (R#)
Fixed
Adjustable or Variable
Photo
Potentiometer
Tapped
Thermistor
T°
Capacitors (C#)
Fixed
Non-Polarized
Split-stator
Electrolytic
Variable
Feed-Through
Inductors (L#)
Air-core
Adjustable
Or
Phasing
Magnetic-core
RFC
Ferrite-bead
RFC air-core
Tubes (V#)
Diode (Rectifier)
Pentode
Triode
Tetrode
CRT
Tube Elements
Anode or Plate
Heater or Filament
Grid
Gas Filled
Cathode
Wiring
Terminal
Conductors Joined
Shielded Wire or Coaxial Cable
Line-break
Address or Data Bus
Multiple Conductor Cable
Switches (S#)
Toggle
SPST
SPDT
Multi-Point
Held Closed
Held Open
Limit Switch
Normally Open
Momentary
Normally Closed
Thermal
Lamps (DS#)
Incandescent
Neon (ac)
7-seg
Grounds
Chassis
Earth
A-analog
D-digital
Integrated Circuits (U#)
General Amplifier
Op Amp
Other
Relays
SPST
N.O.
SPDT
N.C.
DPDT
Thermal
Diodes (D#)
LED (DS#)
Diode/Rectifier
Voltage Variable Capacitor
Zener
Schottky
Thyristor (SCR)
Triac
T
G
Bridge Rectifiers
D#
(U#)
Transformers (T#)
Air Core
with Link
Adjustable Inductance
Adjustable Coupling
with Magnetic-core
Adjustable magnetic core
Miscellaneous
Fuse
Hand Key
Antenna
Quartz Crystal
Motor
Mot
Meter
* = V, mV A, mA, μA
Assembly Or Module (other than IC)
Z
Connectors (P#, J#)
Phone Jacks (J#)
Phone Plug (P#)
Contacts
Male
Female
Phono Jack
MIC Jack (P#)
Multiple Movable
Multiple Fixed
Coaxial Connectors
Terminal Strip
220 V Male
240 V Female
Ground
Hot 120 V
Neut
GND
Chassis-mount
SYMBOLSMM
Transistors (Q#)
NPN
PNP
Bipolar
B
C
E
P-channel
N-channel
UJT
B1
B2
Junction FET
G
D
S
G1
G2
Single-gate
Dual-gate
Depletion Mode Mosfet
Enhancement Mode Mosfet
Darlingtons
Mosfet with Protection Diode
Opto-isolators
Logic (U#)
AND
OR
XOR
NAND
NOR
Invert
Schmitt
Other
Batteries (BT#)
Single Cell
Multi Cell
Photo Cell