Contents

Chapter 4
Circuits and Components

This chapter begins by covering the various aspects of dealing with real components — resistors, capacitors, inductors, and transformers. (A sheet of standard schematic symbols is available in the online content.) All of these have special features and behaviors you must take into account, especially at RF. We then cover common semiconductors you'll encounter in radio and their important characteristics.

The discussion then proceeds to building-block circuits, beginning with various types of amplifiers constructed with bipolar and field-effect transistors. Building-block circuits including op amps and miscellaneous analog ICs are also covered. Finally, a section on Heat Management discusses how to deal with heat in electronic devices.

Chapter 4 — Online Content

Articles

- Common Schematic Symbols
- Copper Wire Tables
- Digital Electronics Tutorial
- Hands On Radio: Basic Operational Amplifiers by Ward Silver, NØAX
- Hands On Radio: Field Effect Transistors by Ward Silver, NØAX
- Hands On Radio: Load Lines by Ward Silver, NØAX
- Hands On Radio: The Common Base Amplifier by Ward Silver, NØAX
- Hands On Radio: The Common Emitter Amplifier by Ward Silver, NØAX
- Hands On Radio: The Emitter-Follower Amplifier by Ward Silver, NØAX
- Large Signal Transistor Operation

Tools and Data

- *LTSpice* Simulation Files
- Frequency Response Spreadsheet

4.1 EIA and Industry Standards

The American National Standards Institute (ANSI), the Electronic Industries Alliance (EIA), and the Electronic Components Association (ECA) establish the US standards for most electronic components, connectors, wire and cables. These standards establish component sizes, wattages, "standard values," tolerances and other performance characteristics. A branch of the EIA sets the standards for Mil-spec (standard military specification) and special electronic components used by defense and government agencies. The Joint Electron Devices Engineering Council (JEDEC), another branch of the EIA, develops the standards for the semiconductor industry. The EIA cooperates with other standards agencies such as the International Electrotechnical Commission (IEC), a worldwide standards agency. You can often find published EIA standards in the engineering library of a college or university.

And finally, the International Organization of Standardization (ISO), headquartered in Geneva, Switzerland, sets the global standards for nearly everything from paper sizes to photographic film speeds. ANSI is the US representative to the ISO.

These organizations, or their acronyms, are familiar to most of us. They are much more than a label on a component. EIA and other industry standards are what mark components for identification, establish the "preferred standard values" and ensure their reliable performance from one unit to the next, regardless of their source. Standards require that a 1.2 kΩ 5% resistor from Ohmite Corp. has the same performance as a 1.2 kΩ 5% resistor from Vishay, or a 2N3904 has the same performance characteristics and physical packaging whether from ON Semi or NTE.

By selecting components manufactured under these industry standards, building a project from the *Handbook* or other source will ensure nearly identical performance to the original design.

There are many sources you can consult for detailed component data but the best source of component information and data sheets is the internet. Most manufacturers maintain extensive websites with information and data on their products. Often, the quickest route to detailed product information is to search for "data sheet" and the part number. Distributors such as Digi-Key and Mouser include links to useful information in their online catalogs as well. Some manufacturers publish parts catalogs which also good sources of component data and application notes and bulletins.

The ARRL Technical Information Service on the ARRL website (**www.arrl.org/technical-information-service**) provides technical assistance to members and nonmembers, including information about components and useful references. The TIS includes links to detailed, commonly needed information in many technical areas. Questions may also be submitted via email (**tis@arrl.org**); fax (860-594-0259); or mail (TIS, ARRL, 225 Main St, Newington, CT 06111).

4.1.1 Component Packaging

Through-hole or *leaded* components have wire leads intended to be inserted into holes in printed-circuit boards or used in point-to-point wiring. Passive leaded components come in two styles: *axial* and *radial*. Axial leads are aligned in opposite directions along a common axis, usually the largest dimension of the component. Radial leads leave the component body in the same direction and are usually arranged radially about the center of the component body. Many types of leaded passive components (R, L, and C) are available in both styles. As an example, see the Capacitors section for drawings of typical axial and radial lead packages for electrolytic capacitors.

"SMT" is used throughout this book to refer to components, printed-circuit boards, or assembly techniques that involve surface-mount technology. SMT components are often referred to by the abbreviations "SMD" and "SMC," but all three abbreviations are considered to be effectively equivalent.

Many different types of electronic components, both active and passive, are now available in surface-mount packages. Each package is identified by a code, such as 1802 or SOT. Resistors in SMT packages are referred to by package code and not by power dissipation, as through-hole resistors are. SMT package outlines and PCB pad placement for a variety of standard package sizes are provided in the **Construction Techniques** chapter's section on PCB Layout.

The very small size of these components leaves little space for marking with conventional codes, so brief alphanumeric codes are used to convey the most information in the smallest possible space. You will need a magnifying glass to read the markings on the bodies of SMT components.

In many cases, vendors will deliver SMT components packaged in tape from master reels and the components will not be marked. This is often the case with SMT resistors and small capacitors. However, the tape will be marked or the components are delivered in a plastic bag with a label. Take care to keep the components separated and labeled or you'll have to measure their values one by one!

4.2 Practical Resistors

4.2.1 Resistance of Wires

The problem of determining the resistance of a round wire of given diameter and length — or the converse, finding a suitable size and length of wire to provide a desired amount of resistance — can easily be solved with the help of the copper wire tables in the online content which give the resistance, in ohms per 1,000 feet, of each standard wire size. For example, suppose you need a resistance of 3.5 Ω, and some #28 AWG wire is on hand. The wire table shows that #28 AWG wire has a resistance of 63.31 Ω / 1,000 ft. Since the desired resistance is 3.5 Ω, the required length of wire is:

$$\text{Length} = \frac{R_{DESIRED}}{\dfrac{R_{WIRE}}{1000\text{ ft}}} = \frac{3.5\,\Omega}{\dfrac{63.31\,\Omega}{1000\text{ ft}}}$$

$$= \frac{3.5\,\Omega \times 1000\text{ ft}}{63.31\,\Omega} = 53.6\text{ ft}$$

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

$$\frac{R_{WIRE}}{1000\text{ ft}} < \frac{R_{MAXIMUM}}{\text{Length}} = \frac{0.05\,\Omega}{14.0\text{ ft}}$$

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

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

Find the value of R_{WIRE} / 1,000 ft that is less than the calculated value. The wire table shows that #15 AWG is the smallest size having a resistance less than this value. (The resistance of #15 AWG wire is given as 3.1810 Ω / 1,000 feet.) Select any wire size larger than this for the connections in your circuit to ensure that the total wire resistance will be less than 0.05 Ω.

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

$$\text{Length} = \frac{R_{DESIRED}}{\dfrac{R_{WIRE}}{1000\text{ ft}}}$$

$$= \frac{3.5\,\Omega}{\dfrac{66.17\,\Omega}{1000\text{ ft}} \times 5.1} = \frac{3.5\,\Omega \times 1000\text{ ft}}{66.17\,\Omega \times 5.1}$$

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

Component Tolerance and Temperature Coefficient

Electronic components such as resistors, capacitors, and inductors are manufactured having a *nominal* value — the value with which they are labeled. The component's *actual* value is what is measured with a suitable measuring instrument. The actual value of resistance varies from the nominal value because of random variations in the manufacturing process. If the nominal value is given as text characters, an "R" in the value (for example "4R7") stands for radix and is read as a decimal point, thus "4.7."

The maximum allowable amount of variation is the component's *tolerance*, and it is expressed in percent. For example, a 1,000-Ω resistor with a tolerance of 5% could have any value of resistance between 95% and 105% of 1,000 Ω; 950 to 1,050 Ω. In most circuits, this small variation doesn't have much effect, but it is important to be aware of tolerance and choose the correct value (10%, 5%, 1%, or even tighter tolerance values are available for *precision components*) of tolerance for the circuit to operate properly, no matter what the actual value of resistance. All components have this same nominal-to-actual value relationship.

The *temperature coefficient* or *tempco* of a component describes its change in value with temperature. Tempco may be expressed as a change in unit value per degree (ohms per degree Celsius) or as a relative change per degree (parts per million per degree). Except for temperature sensing components that may use Fahrenheit or Kelvin, Celsius is almost always used for the temperature scale. Temperature coefficients may not be linear, such as those for capacitors, thermistors, or quartz crystals. In such cases, tempco is specified by an identifier such as Z5U or C0G and an equation or graph of the change with temperature is provided by the manufacturer.

TEMPERATURE EFFECTS ON RESISTORS

Current through a resistance causes the conductor to become heated; the higher the resistance and the larger the current, the greater the amount of heat developed. The power to be dissipated is $P = I^2R$ or $P = V^2/R$ using RMS values for the current through the resistor, I, or voltage across the resistor, V.

Resistors intended for carrying large currents must be physically large so the heat can be radiated quickly to the surrounding air or some type of heat sinking material. If the resistor does not dissipate the heat quickly, it may get hot enough to melt or burn.

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

4.2.2 Component Resistors

TYPES OF RESISTORS

The size and construction of resistors having the same value of resistance in ohms may vary considerably based on how much power they are intended to dissipate, how much voltage is expected to be applied to them, and so forth (see **Figure 4.1** and **Table 4.1**).

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

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

Metal-film resistors are made by depositing a thin film of aluminum, tungsten, or other metal on an insulating substrate. Their resistances are controlled by careful adjustments of the width, length and depth of the film. As a result, they have very tight tolerances. They are used extensively in surface-mount technology. As might be expected, their power handling capability is somewhat limited. They also produce very little electrical noise. For new designs metal-film and carbon-film (see next entry) resistors should be used for their improved characteristics and lower cost compared to the older carbon composition resistors. Metal-film resistors have lower residual inductance and are often preferred at VHF.

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

Metal-oxide resistors are similar to carbon composition resistors in that the resistance is supplied by a cylinder of metal oxide. Metal-oxide resistors have replaced carbon composition resistors in higher power applications because they are more stable and can operate at higher temperatures. These resistors are available in leaded styles and in transistor-style TO-220 packages for use at high power and RF.

Wirewound resistors are made from wire that is wound on a coil form (usually ceramic). They are capable of handling high power, their values are very stable, and they are manufactured to close tolerances. The wound-wire construction creates inductance so these

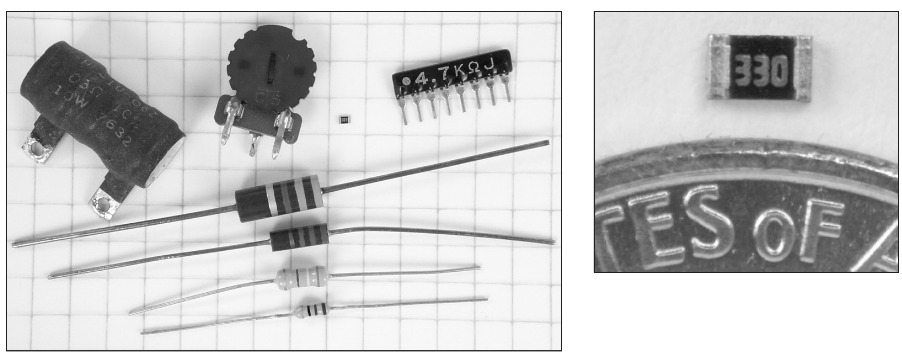
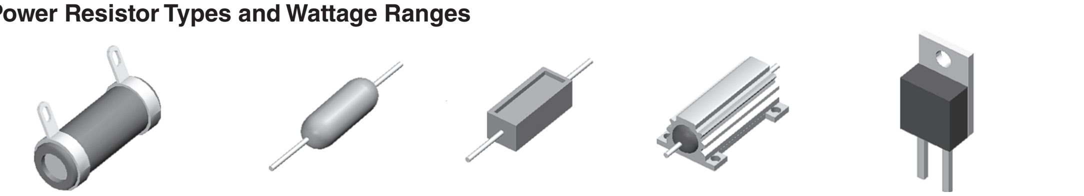


Figure 4.1 — Examples of various resistor types. At the top left is a small 10-W wirewound resistor. A single in-line package (SIP) of resistors is at the top right. At the top center is a small PC-board-mount variable resistor. A tiny surface-mount (chip) resistor is also shown at the top. Below the variable resistor is a 1-W carbon composition resistor and then a ½-W composition unit. The dog-bone-shaped resistors at the bottom are ½-W and ¼-W film resistors. The ¼-inch-ruled graph paper background provides a size comparison. The photo on the right shows the chip resistor with a penny for size comparison. The drawing in Table 4.1 shows several different types of wirewound power resistors.

Table 4.1
Power Resistor Types and Wattage Ranges

HBK0463



Wire-wound, ceramic core, 10-300 W*	Wire-wound, axial, 3-10 W*	Metal-oxide, 5-25 W**	Wire-wound, aluminum housing, 3-50 W*	Thick-film resistors, 15-100 W**

*Wire-wound resistors are inductive, though seldom noted as such on the data sheets, and are not recommended for RF.

**Thick-film and metal-oxide power resistors are low inductance or noninductive.

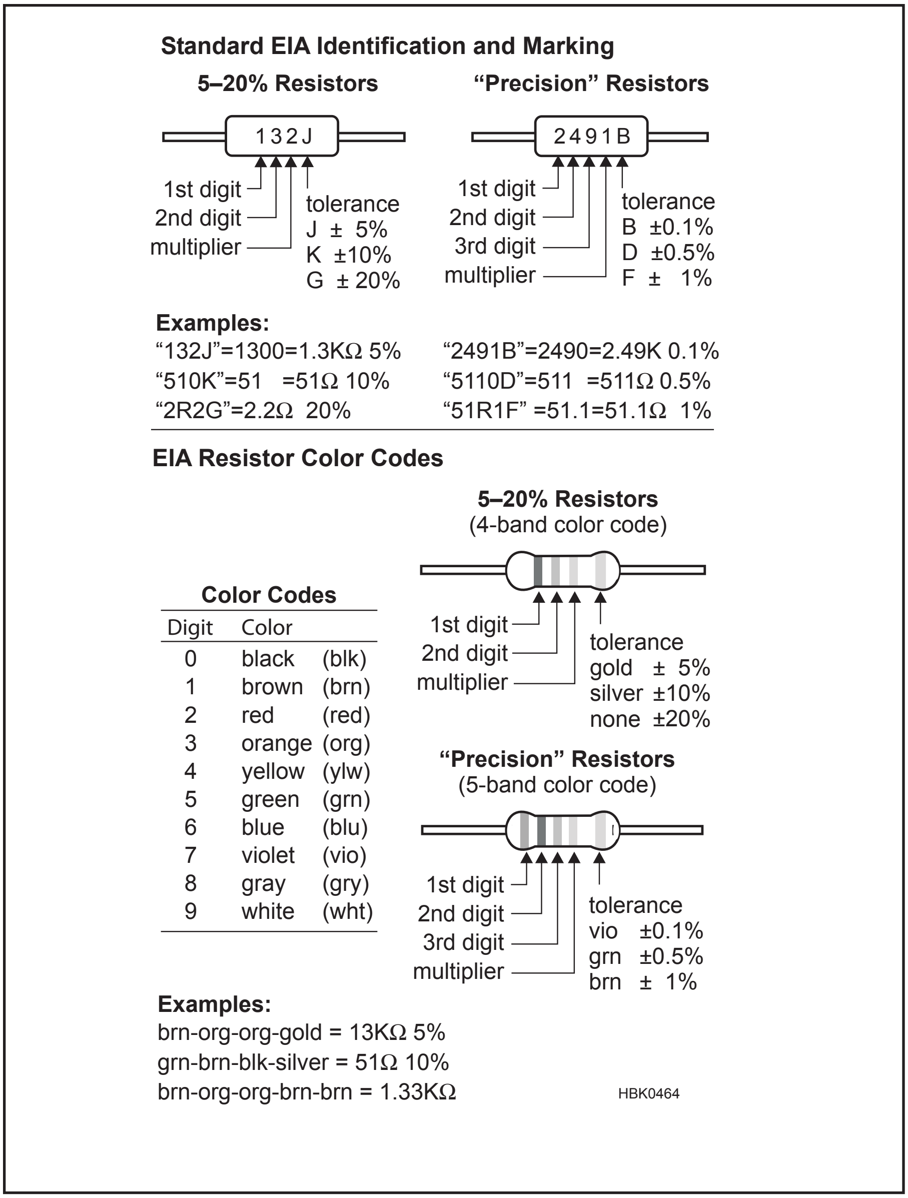


Figure 4.2 — Resistor value identification.

resistors are not suitable for ac circuits above a few kHz. They have power ratings from 3 to 300 W.

Thick-film resistors are also available at high power ratings (up to 100 W) in a transistor-like TO-220 package. Because of their packaging, they can be mounted to heat sinks and printed-circuit boards. These resistors also have low inductance for use in switch-mode circuits and high-frequency applications.

RESISTOR IDENTIFICATION

Resistors are identified by the IEC 60062:2016 standard as shown in **Figure 4.2**. The IEC numerical code is used worldwide in industry. The nominal resistance, expressed in ohms, is identified by three digits for 2% (and greater) tolerance devices. The first two digits represent the significant figures; the last digit specifies the multiplier as the exponent of 10. (The multiplier is simply the number of zeros following the significant numerals.) For values less than 100 Ω, the letter R is substituted for one of the significant digits and represents a decimal point. An alphabetic character indicates the tolerance as follows:

D	±0.5%
F	±1.0%
G	±2.0%
J	±5.0%

For example, a resistor marked with "122J" would be a 1,200 Ω, or a 1.2 kΩ 5% resistor. If the tolerance of the unit is narrower than ±2%, the code used is a four-digit code where the first three digits are the significant figures and the last is the multiplier. A resistor containing four digits, such as "1211F," would be a 1,210 Ω, or a 1.21 kΩ 1% precision resistor.

The letter R is used in the same way to represent a decimal point. For example, 22R0 indicates a 22-Ω unit.

Here are some additional examples of resistor value markings:

Code	*Value*
101	10 and 1 zero = 100 Ω
224	22 and 4 zeros = 220,000 Ω
1R0	1 point zero = 1 Ω
22R	22.0 and no zeros = 22 Ω
R10	0 point 1 zero = 0.1 Ω

The resistor color code, used only with through-hole components, assigns colors to the numerals one through nine and zero, as shown in **Table 4.2**, to represent the significant numerals, the multiplier and the tolerance. The color code is often memorized with a mnemonic such as "Big boys race our young girls, but Violet generally wins" to represent the colors black (0), brown (1), red (2), orange (3), yellow (4), green (5), blue (6), violet (7), gray (8) and white (9). You will no doubt discover other versions of this memory aid made popular over the years.

For example, a resistor with color bands black (1), red (2), red (2) and gold would be a 1,200 Ω, or 1.2 kΩ 5% resistor, with the gold band signifying 5% tolerance.

The resistor color code should be memorized as it is also used for identifying capacitors and inductors. It is also handy to use when connecting multiconductor or ribbon cables.

Resistors are also identified by an "E" series classification, such as E12 or E48. The number following the letter E signifies the number of logarithmic steps per decade. The more steps per decade, the more the choices of resistor values and the tighter the tolerances can be. For example, in the E12 series, there are twelve resistor values between 1 kΩ and 10 kΩ with 10% tolerance; E48 provides 48 values between 1 kΩ and 10 kΩ at 1% tolerance. This system is often used with online circuit calculators to indicate the resistor accuracy and tolerance desired. The standard resistor values of the E12 (±10%), E24 (±5%), E48 (±2%) and E96 (±1%) series are listed in **Table 4.3**.

Resistors used in military electronics (Mil-spec) use the type identifiers listed in **Table 4.4**. In addition, Mil-spec resistors with paint-stripe value bands have an extra band indicating the reliability level to which they are certified.

Table 4.2
Resistor Color Codes

Color	*Significant Figure*	*Decimal Multiplier*	*Tolerance (%)*
Black	0	1	
Brown	1	10	1
Red	2	100	2
Orange	3	1,000	
Yellow	4	10,000	
Green	5	100,000	0.5
Blue	6	1,000,000	0.25
Violet	7	10,000,000	0.1
Gray	8	100,000,000	0.05
White	9	1,000,000,000	
Gold		0.1	5
Silver		0.01	10
No color			20

Table 4.3
EIA Standard Resistor Values

±10% (E12)	*±5% (E24)*	*±2% (E48)*		*±1% (E96)*			
100	100	100	316	100	178	316	562
120	110	105	332	102	182	323	576
150	120	110	348	105	187	332	590
180	130	115	365	107	191	340	604
220	150	121	383	110	196	348	619
270	160	127	402	113	200	357	634
330	180	133	422	115	205	365	649
390	200	140	442	118	210	374	665
470	220	147	464	121	215	383	681
560	240	154	487	124	221	392	698
680	270	162	511	127	226	402	715
820	300	169	536	130	232	412	732
	330	178	562	133	237	422	750
	360	187	590	137	243	432	768
	390	196	619	140	249	442	787
	430	205	649	143	255	453	806
	470	215	681	147	261	464	825
	510	226	715	150	267	475	845
	560	237	750	154	274	487	866
	620	249	787	158	280	499	887
	680	261	825	162	287	511	909
	750	274	866	165	294	523	931
	820	287	909	169	301	536	953
	910	301	953	174	309	549	976

Use Table 4.3 values for each decade.
Example: 133 = 13.3 Ω, 133 Ω, 1.33 kΩ, 13.3 Ω, 133 kΩ, 1.33 MΩ

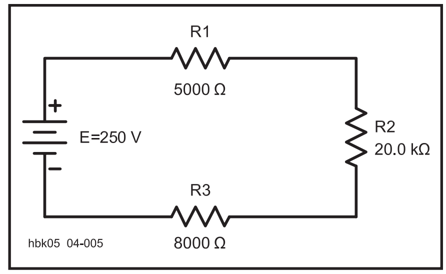


Figure 4.3 — An example of resistors in series.

Surface-mount (SMT) resistors are labeled with an alphanumeric code. There are several identification conventions, including the three-digit and four-digit value-and-exponent and an EIA-96 labeling standard described at **www.hobby-hour.com/electronics/smdcalc.php**.

4.2.3 Effect of Temperature on Resistors

The major departure of resistors from ideal behavior at low frequencies is their *temperature coefficient* (TC or TCR for resistors). (See the chapter on **RF Techniques** for a discussion of the behavior of resistors at high frequencies.) The resistivity of most materials changes with temperature, and typical TC values for resistor materials are given in **Table 4.5**. TC values are usually expressed in parts-per-million (ppm) for each degree (centigrade) change from some nominal temperature, usually room temperature (77 °F or 27 °C). A positive TC indicates an increase in resistance with increasing temperature, while a negative TC indicates a decreasing resistance. For example, if a 1,000-Ω resistor with a TC of +300 ppm/°C is heated to 50 °C,

Table 4.4
Mil-Spec Resistors

Wattage	*Metal Film Types*	*Fixed Film Types*	*Composition Types*	
1⁄10 W	RN50			
1⁄8 W	RN55	RL05	RLR05	RCR05
1⁄4 W	RN60	RL07	RLR07	RCR07
1⁄2 W	RN65	RL20	RLR20	RCR20
1 W	RN75	RL32	RLR32	RCR32
2 W	RN80	RL42	RLR62	RCR42

Tolerance Codes

B	±0.1%
C	±0.25%
D	±0.5%
F	±1%
G	±2%
J	±5%
K	±10%

Examples:
RN60D-2202F = 22 kΩ 1%
RL07S-471J = 470 Ω ±5%
RLR07C-471J = 470 Ω ±5%

Note: The RN Mil-Spec was discontinued in 1996. It is still used by some manufacturers such as Vishay-Dale.

Table 4.5
Temperature Coefficients for Various Resistor Compositions
1 PPM = 1 part per million = 0.0001%

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

the *change* in resistance is 300 × (50 – 27) = 6,900 ppm, yielding a new resistance of

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

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

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

4.2.4 Voltage Dividers

According to Kirchoff's Voltage Law (KVL), the voltage drop across each resistor in a series circuit is directly proportional to the resistance. When connected in series, a resistor that has a value twice as large as another will have twice the voltage drop across it.

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

For example, in the circuit of **Figure 4.3**, the voltage across the 5,000 Ω resistor is:

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

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

Voltage dividers can be used as a source of voltage. As long as the device connected to the output of the divider has a much higher

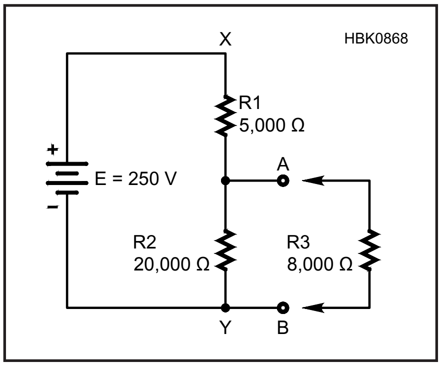


Figure 4.4 — A voltage divider showing both unloaded (R3 not connected) and loaded (R3 connected) conditions.

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

The following equations show how to calculate the voltage produced by a voltage divider. Using the circuit in **Figure 4.4**, the *unloaded* output voltage (with R3 not connected) is:

$$V_{OUT}=V_{IN}\left(\frac{R2}{R1+R2}\right)$$

If R3 is connected, the *loaded* output voltage is:

$$V_{OUT}=V_{IN}\left(\frac{R2//R3}{R1+R2//R3}\right)$$

where // indicates "in parallel with."

A good rule of thumb to keep the loaded output voltage within about 10% of the unloaded voltage is for the load resistance to be at least 10 times higher than the output resistor of the divider. As the load resistance approaches the value of the output resistor, the additional current through the load causes additional voltage drop across the divider's input resistor.

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

ATTENUATORS

A special type of voltage divider is used to reduce signal power levels while maintaining a constant resistance through the divider. The two primary types of divider circuits are the Pi-network and T-network shown at the bottom of **Table 4.6** and **Table 4.7**. These circuits are symmetric: the input and output resistor values are equal. The values for R1 and R2 in the tables are chosen to present a 50-Ω resistance at both the input and output. See the **Test Equipment and Measurements** chapter for a discussion of switched-step attenuators and how to use them.

4.2.5 Current Dividers

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

$$I_n = I_{TOT}\, R_{EQ}/(R_n + R_{EQ})$$

For example, in a circuit with three parallel resistors R1, R2, and R3, the current through

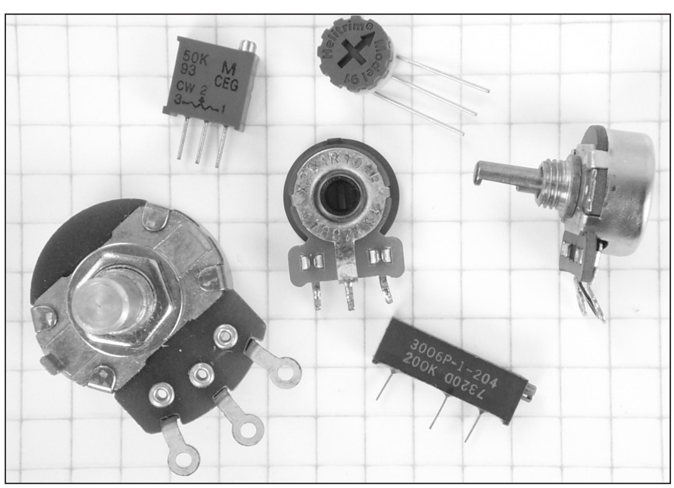
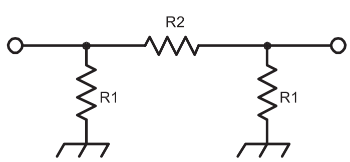


Figure 4.5 — This photo shows examples of different styles of potentiometers. The ¼-inch-ruled graph paper background provides a size comparison.

Table 4.6
Pi-Network Resistive Attenuators (50 Ω)

dB Atten.	*R1 (Ohms)*	*R2 (Ohms)*
1.0	870	5.77
2.0	436	11.6
3.0	292	17.6
4.0	221	23.8
5.0	178	30.4
6.0	150	37.4
7.0	131	44.8
8.0	116	52.8
9.0	105	61.6
10.0	96.2	71.2
11.0	89.2	81.7
12.0	83.5	93.2
13.0	78.8	106
14.0	74.9	120
15.0	71.6	136
16.0	68.8	154
17.0	66.4	173
18.0	64.4	195
19.0	62.6	220
20.0	61.1	248
21.0	59.8	278
22.0	58.6	313
23.0	57.6	352
24.0	56.7	395
25.0	56.0	443
30.0	53.2	790
35.0	51.8	1405
40.0	51.0	2500
45.0	50.5	4446
50.0	50.3	7906
55.0	50.2	14,058
60.0	50.1	25,000

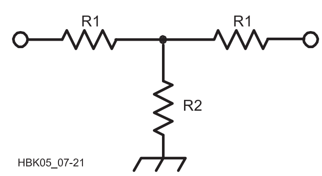
An RF Step Attenuator project is shown in the **Test Equipment and Measurements** chapter of this *Handbook*, and a Low Power Step Attenuator PC board is available from FAR Circuits at **www.farcircuits.net/test2.htm**.



HBK05_07-20

Table 4.7
T-Network Resistive Attenuators (50 Ω)

dB Atten.	*R1 (Ohms)*	*R2 (Ohms)*
1.0	2.88	433
2.0	5.73	215
3.0	8.55	142
4.0	11.3	105
5.0	14.0	82.2
6.0	16.6	66.9
7.0	19.1	55.8
8.0	21.5	47.3
9.0	23.8	40.6
10.0	26.0	35.1
11.0	28.0	30.6
12.0	30.0	26.8
13.0	31.7	23.5
14.0	33.3	20.8
15.0	35.0	18.4
16.0	36.3	16.2
17.0	37.6	14.4
18.0	38.8	12.8
19.0	40.0	11.4
20.0	41.0	10.0
21.0	41.8	9.0
22.0	42.6	8.0
23.0	43.4	7.1
24.0	44.0	6.3
25.0	44.7	5.6
30.0	47.0	3.2
35.0	48.2	1.8
40.0	49.0	1.0
45.0	49.4	0.56
50.0	49.7	0.32
55.0	49.8	0.18
60.0	49.9	0.10



R2 is equal to:

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

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

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

4.2.6 Potentiometers

Potentiometer (pronounced po-ten-tchee-AH-meh-tur) is a formal name for a variable resistor and the common name for these components is "pots." A typical potentiometer consists of a circular *element* of resistive material which can be a carbon compound similar to that used in carbon composition resistors or a conductive plastic. One contact is made to each end of the element. A variable-position *wiper* makes contact with the element at different positions. As the wiper moves along the material, more resistance is introduced between the wiper and one of the element's contacts. The wiper is turned by a shaft to move across the material.

For higher power applications, the element may be wire wound around a core, like a wirewound resistor. Like the wirewound resistor, this type of pot is not suitable for high-frequency applications due to inductance.

A potentiometer may be used to control current, voltage, or resistance in a circuit. **Figure 4.5** shows several different types of potentiometers. **Figure 4.6** shows the schematic symbol for a potentiometer and how changing the position of the shaft changes the resistance between its three terminals. The figure shows a *panel pot*, designed to be mounted on an equipment panel and adjusted by an operator. The small rectangular *trimmer* potentiometers in Figure 4.6 are adjusted with a screwdriver and have wire terminals.

Typical specifications for a potentiometer include element resistance, power dissipation, voltage and current ratings, number of turns (or degrees) the shaft can rotate, type and size of shaft, mounting arrangements, and resistance *taper*. Taper describes how the resistance of the element changes with position along it.

Pots with a *linear taper* have equal change in resistance with position along the element. That is, the change in resistance is the same for a given number of degrees of shaft rotation

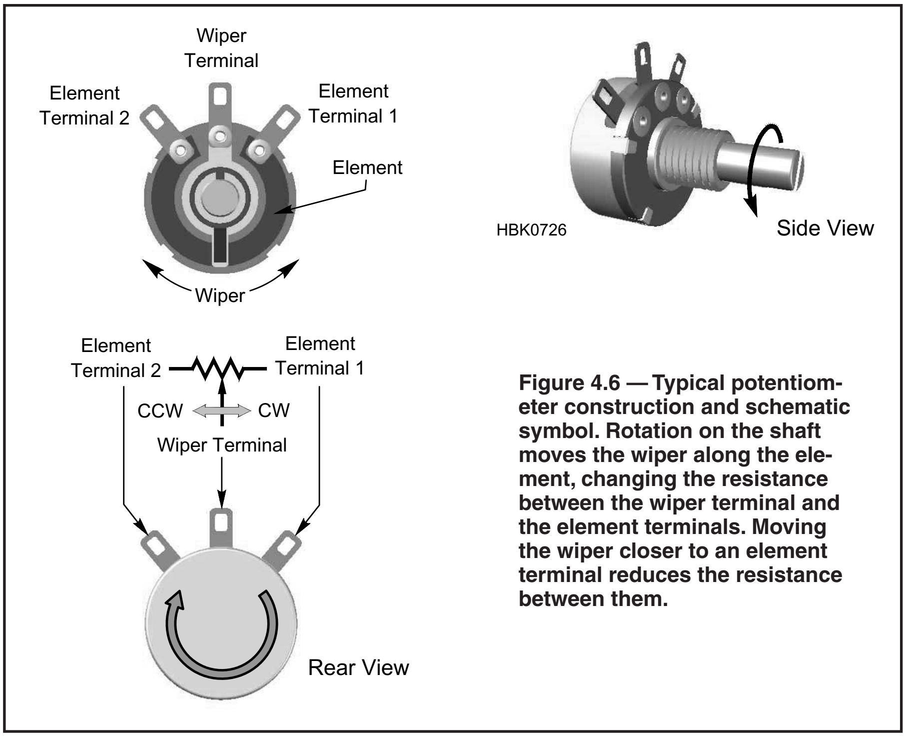


Figure 4.6 — Typical potentiometer construction and schematic symbol. Rotation on the shaft moves the wiper along the element, changing the resistance between the wiper terminal and the element terminals. Moving the wiper closer to an element terminal reduces the resistance between them.

anywhere along different portions of the resistive material.

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

4.3 Practical Capacitors

Like resistors, capacitors are available in EIA standard series of values, E6 and E12, shown in **Table 4.8**. Most capacitors have a tolerance of 5% or greater. High-value capacitors used for filtering may have asymmetric tolerances, such as –5% and +10%, since the primary concern is for a guaranteed minimum value of capacitance.

SMT capacitors are generally film, ceramic, or tantalum electrolytics. Although the IEC scheme is the standard method of labeling capacitor value, you may encounter a two-character alphanumeric code (see **Table 4.9**) consisting of a letter indicating the significant digits and a number indicating the multiplier. The code represents the capacitance in picofarads. For example, a chip capacitor marked "A4" would have a capacitance of 10,000 pF, or 0.01 µF. A unit marked "N1" would be a 33-pF capacitor. If there is sufficient space on the device package, a tolerance code may be included.

Table 4.8
EIA Standard Capacitor Values

±20% Capacitors (E6)

pF	*pF*	*pF*	*µF*	*µF*	*µF*	*µF*
1.0	10	100	0.001	0.01	0.1	1
1.5	15	150	0.0015	0.015	0.15	1.5
2.2	22	220	0.0022	0.022	0.22	2.2
3.3	33	330	0.0033	0.033	0.33	3.3
4.7	47	470	0.0047	0.047	0.47	4.7
6.8	68	680	0.0068	0.068	0.68	6.8

±10%, ±5% Capacitors (E12)

pF	*pF*	*pF*	*µF*	*µF*	*µF*	*µF*
1.0	10	100	0.001	0.01	0.1	1
1.2	12	120	0.0012	0.012	0.12	
1.5	15	150	0.0015	0.015	0.15	
1.8	18	180	0.0018	0.018	0.18	
2.2	22	220	0.0022	0.022	0.22	2.2
2.7	27	270	0.0027	0.027	0.27	
3.3	33	330	0.0033	0.033	0.33	3.3
3.9	39	390	0.0039	0.039	0.39	
4.7	47	470	0.0047	0.047	0.47	4.7
5.6	56	560	0.0056	0.056	0.56	
6.8	68	680	0.0068	0.068	0.68	
8.2	82	820	0.0082	0.082	0.82	

4.3.1 Capacitor Types and Characteristics

Capacitors exhibit the largest variety of electronic components. So many varieties and types are available that selecting the proper capacitor for a particular application can be overwhelming. Capacitors are generally grouped by the type of dielectric; ceramic, mica, film, polymer, and electrolytic are the most common. Some of the different types are shown in **Figure 4.7** and many other types and packages are available. (See the **RF Techniques** chapter for information about capacitor characteristics at RF.) The type of capacitor can often be determined by its appearance, as shown in **Figures 4.8** and **4.9**.

The type of dielectric determines many properties of the capacitor, although the construction of the electrodes or plates strongly affects the capacitor's ac performance and

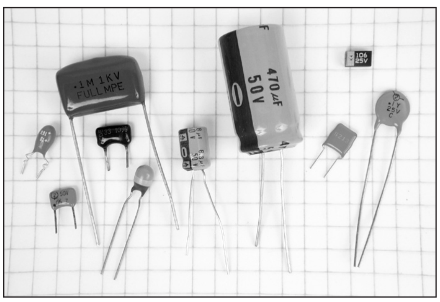
Table 4.9
SMT Capacitor Two-Character Labeling

Significant Figure Codes

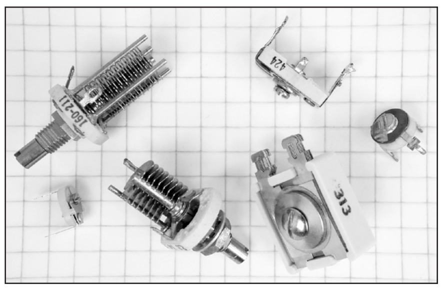
Character	*Significant Figures*	*Character*	*Significant Figures*
A	1.0	T	5.1
B	1.1	U	5.6
C	1.2	V	6.2
D	1.3	W	6.8
E	1.5	X	7.5
F	1.6	Y	8.2
G	1.8	Z	9.1
H	2.0	a	2.5
J	2.2	b	3.5
K	2.4	d	4.0
L	2.7	e	4.5
M	3.0	f	5.0
N	3.3	m	6.0
P	3.6	n	7.0
Q	3.9	t	8.0
R	4.3	y	9.0
S	4.7		

Multiplier Codes

Numeric Character	*Decimal Multiplier*
0	1
1	10
2	100
3	1,000
4	10,000
5	100,000
6	1,000,000
7	10,000,000
8	100,000,000
9	0.1



(A)



(C)

(B)

(D)

(E)

Figure 4.7 — Fixed-value capacitors are shown in parts A and B. Aluminum electrolytic capacitors are pictured near the center of photo A. The small tear-drop units to the left of center are tantalum electrolytic capacitors. The rectangular units are silvered-mica, polystyrene film and monolithic ceramic. At the right edge is a disc-ceramic capacitor and near the top right corner is a surface-mount capacitor. B shows a large "computer-grade" electrolytic. These have very low equivalent series resistance (ESR) and are often used as filter capacitors in switch-mode power supplies, and in series-strings for high-voltage supplies of RF power amplifiers. Parts C and D show a variety of variable capacitors, including air variable capacitors and mica compression units. Part E shows a vacuum variable capacitor such as is sometimes used in high-power amplifier circuits. The ¼-inch-ruled graph paper backgrounds provide size comparisons.

some dc parameters. Various materials are used for different reasons such as working voltage and current, availability, cost, and desired capacitance range.

Disc* (or *disk) ceramic capacitors consist of two metal plates separated by a ceramic dielectric that establishes the desired capacitance. Due to their low cost, they are the most common capacitor type. The main disadvantage is their sensitivity to temperature changes (that is, a high temperature coefficient).

Monolithic ceramic capacitors are made by sandwiching layers of metal electrodes and ceramic layers to form the desired capacitance as shown in **Figure 4.10**. "Monolithics" are physically smaller than disc ceramics for the same value of capacitance and cost, but exhibit the same high temperature coefficients. They are made from a stack of metal plating on ceramic layers that are pressed together and coated with an insulating jacket.

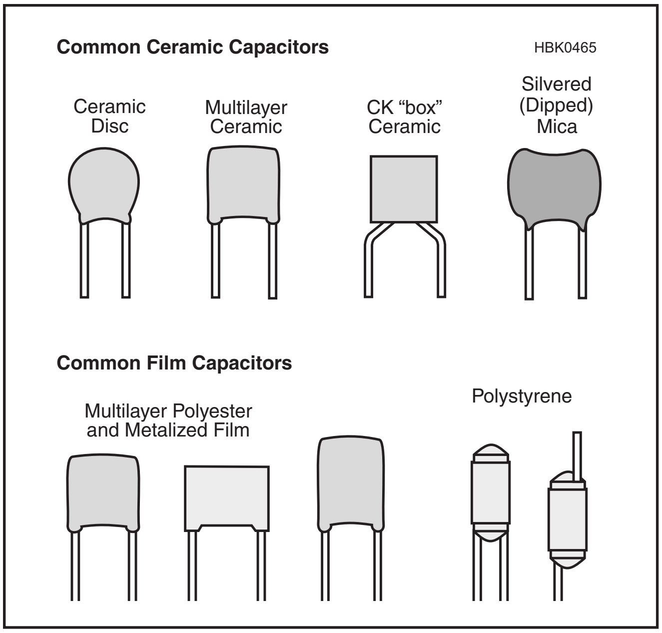


Figure 4.8 — Common capacitor types and package styles.

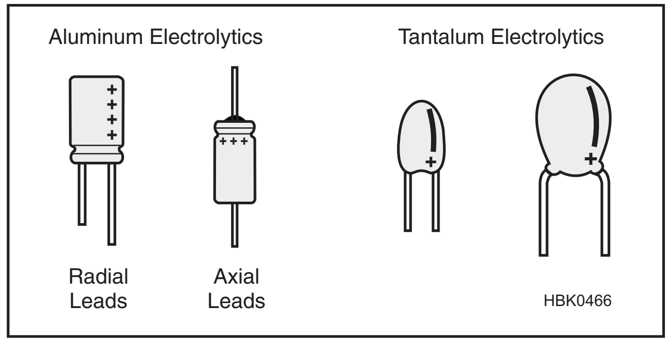


Figure 4.9 — Aluminum and tantalum electrolytic capacitors.

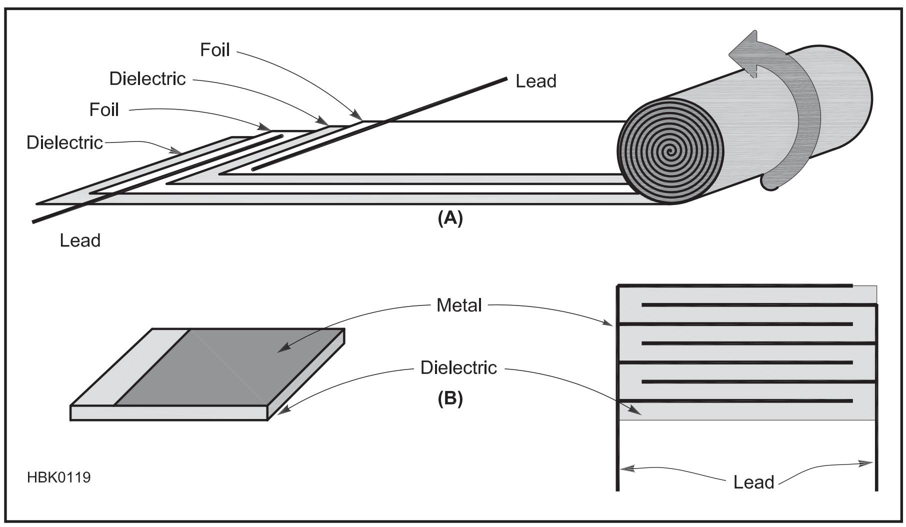


Figure 4.10 — Two common types of capacitor construction. A shows the roll method for film capacitors with axial leads. B shows the alternating layer method for ceramic capacitors with radial leads.

Plastic film capacitors use layers of metal and a plastic such as polystyrene, polypropylene, or polyester (Mylar) to make a wide range of capacitances. Values range from tens of pF to 1 µF. They have high leakage resistances (even at high temperatures) and low TCs. Smaller sizes can be used up to several megahertz. Film capacitors are often made using roll construction as in Figure 4.10. This construction has relatively high inductance and the capacitor may not be suitable for RF applications. Plastic-film variable capacitors are also available.

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

Paper capacitors are generally not used in new designs and are largely encountered in older equipment; capacitances from 500 pF to 50 µF are available. High working voltages are possible, but paper-dielectric capacitors have low leakage resistances and tolerances are no better than 10 to 20%. Paper capacitors are not available as new stock (except possibly for specialty restoration applications) and should not be used in new equipment.

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

Mica capacitors are very stable with respect to time, temperature, and electrical stress. Leakage and losses are very low and they are often used in transmitting equipment. Values range from 1 pF to 0.1 µF. High working voltages are possible, but they must be derated severely as operating frequency increases. Note that many WWII-era mica capacitors are still available on the used and surplus market. Given the age of these components (now approaching 80 years) either use newer components or test them before using them in a high-power or high-voltage circuit.

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

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

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

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

Air-variable capacitors have one set of plates that is movable with respect to the other set to vary the area of overlap and thus the capacitance. A *transmitting-variable* capacitor has heavy plates far enough apart to withstand the high voltages and currents encountered in a transmitter. (Air variable capacitors with more closely spaced plates are often referred to as *receiving-variables.*)

Trimming capacitors. Small-value variable capacitors are often referred to as *trimmers* because they are used for fine-tuning or frequency adjustments, called *trimming*. Trimmers have dielectrics of Teflon, plastic film, air, or ceramic and generally have values of less than 100 pF. *Compression trimmers* constructed with mica dielectrics have higher values of up to 1,000 pF and are no longer manufactured but widely available as surplus or NOS (new old stock).

Table 4.10
Ceramic Temperature Characteristics

Common EIA Types:

EIA Class	*EIA Code*	*Characteristics*	*Temp. Range**
1	C0G	0 ± 30 ppm/°C	–55 °C to +125 °C
2	Y5P	±10%	–30 °C to + 85 °C
2	X7R	±15%	–55 °C to +125 °C
2	Y5U	±20%	–10 °C to + 85 °C
2	Z5U	±20%	+10 °C to + 85 °C
2	Z5V	+80%, –20%	–30 °C to + 85 °C
3	Y5V	+80%, –20%	–10 °C to + 85 °C

Common Industry Types:

EIA Class	*EIA Code*	*Characteristics*	*Temp. Range**
1	NP0	0 ± 30 ppm/°C	–55 °C to +125 °C
2	CK05	±10%	–55 °C to +125 °C

*Temp. range for which characteristics are specified and may vary slightly between different manufacturers

Temperature Coefficient Codes

Minimum Temperature	*Maximum Temperature*	*Maximum capacitance change over temp range*
X –55 °C	2 +45°C	A ±1.0%
Y –30 °C	4 +65°C	B ±1.5%
Z +10 °C	5 +85°C	C ±2.2%
6 +105 °C		D ±3.3%
7 +125 °C		E ±4.7%
		F ±7.5%
		P ±10%
		R ±15%
		S ±22%
		T –33%, +22%
		U –56%, +22%
		V –82%, +22%

TEMPERATURE COEFFICIENT

Capacitors are particularly sensitive to temperature changes because the physical dimensions of the capacitor determine its value. Standard temperature coefficient codes are shown in **Table 4.10**. Each code is made up of one character from each column in the table. For example, a capacitor marked Z5U is suitable for use between +10 and +85 °C, with a maximum change in capacitance of –56% or +22%.

Capacitors with highly predictable temperature coefficients of capacitance are sometimes used in circuits whose performance must remain stable with temperature. If an application called for a temperature coefficient of –750 ppm/°C (N750), a capacitor marked U2J would be suitable. The older industry code for these ratings is being replaced with the EIA code shown in Table 4.10. NP0 (that is, N-P-zero) means "negative, positive, zero." It is a characteristic often specified for RF circuits requiring temperature stability, such as VFOs. A capacitor of the proper value marked C0G is a suitable replacement for an NP0 unit.

LEAKAGE RESISTANCE

If we use anything other than a vacuum for the insulating layer, even air, two imperfections are created. Because there are atoms between the plates, some electrons will be available to create a current between the plates when a dc voltage is applied. The magnitude of this *leakage current* will depend on the insulator quality, and the current is usually very small. Leakage current can be modeled by a resistance R_L in parallel with the capacitance (in an ideal capacitor, R_L is infinite). **Table 4.11** shows typical dc leakage resis-

Table 4.11
Typical Temperature Coefficients and Leakage Resistances for Various Capacitor Constructions

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

tances for different dielectric materials. Leakage also generally increases with increasing temperature.

CAPACITOR LOSSES

When an ac current flows through the capacitor (even at low frequencies), capacitors dissipate some of the energy stored in the dielectric due to the electromagnetic properties of dielectric materials. This loss can be thought of as a resistance in series with the capacitor and it is often specified in the manufacturer's data for the capacitor as *effective (or equivalent) series resistance* (*ESR*) which is specified in ohms. (Capacitors that are constructed using roll construction as in Figure 4.10 also have significant inductance specified as *equivalent series inductance (ESL)* which is specified in micro- or nano-henries.)

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

Dissipation Factor (DF) or *loss tangent* is the ratio of loss resistance to reactance.

$$DF = \tan\theta = ESR / X_C$$

The loss angle of a given capacitor is relatively constant over frequency, meaning that $ESR = (\tan\theta) / 2\pi fC$ goes down as frequency goes up. For this reason, ESR must be specified at a given frequency.

4.3.2 Electrolytic Capacitors

Aluminum electrolytic capacitors use aluminum foil "wetted" with a chemical agent and formed into layers to increase the effective area, and therefore the capacitance. Aluminum electrolytics provide high capacitance in small packages at low cost. Most varieties are polarized, that is, voltage should only be applied in one "direction." Polarized capacitors have a negative (–) and positive (+) lead. EIA standard values for aluminum electrolytics are given in **Table 4.12**.

Table 4.12
Aluminum Electrolytic Capacitors
EIA ±20%Standard Values

μF	*μF*	*μF*	*μF*	*μF*
0.1	1.0	10	100	1,000
0.22	2.2	22	220	2200
0.33	3.3	33	330	3300
0.47	4.7	47	470	4700
0.68	6.8	68	680	6800
0.82	8.2	82	820	8200

Very old electrolytic capacitors should be used with care or, preferably, replaced. The wet dielectric agent can dry out during prolonged periods of non-use, causing the internal capacitor plates to form a short circuit when energized. Applying low voltage and gradually increasing it over a period of time may restore the capacitor to operation, but if the dielectric agent has dried out, the capacitor will have lost some or most of its value and will likely be lossy and prone to failure.

Tantalum electrolytic capacitors consist of an extremely porous tantalum (a rare-earth metallic element) pentoxide powder mixed with a wet or dry electrolyte, then formed into a pellet or slug for a large effective area. Tantalums also provide high capacitance values in very small packages. Tantalums tend to be more expensive than aluminum electrolytic capacitors. Like the aluminum electrolytic capacitor, tantalum capacitors are also polarized for which care should be exercised. Some varieties of tantalums can literally explode or burst open if voltage is applied with reverse polarity or the voltage rating is exceeded. Tantalum electrolytics are used almost exclusively as high-value SMT components due to their small sizes. Capacitance values up to 1,000 μF at 4 V are available with body sizes about a quarter-inch square.

Identifying the polarity markings of aluminum and tantalum electrolytics (shown in Figure 4.9) can be confusing. Most tantalum electrolytics are marked with a solid band indicating the positive lead. Aluminum electrolytics are available with bands or symbols marking *either* the negative or positive lead. The positive lead of axial-lead electrolytic capacitors is usually manufactured to be longer than the negative lead and often enters the capacitor through fiber or plastic insulating material while the negative lead is connected directly to the metallic case of the capacitor. Misidentifying the polarity of capacitors is a common error during assembly or repair.

Supercapacitors are a special type of electrolytic capacitor with very high capacitance (greater than 1 F) and a low working voltage (a few volts, typically). Supercapacitors have one electrode that is a porous material with an extremely high ratio of surface area to volume and is immersed in electrolyte fluid or gel. The porous electrode is contained in a metal can that forms the other electrode. The dielectric forms on the porous material's surface, similarly to foil-type electrolytics. "Supercaps" are used as short-term power sources and as filter capacitors for power connections with large, sudden, short-term current demands.

4.3.3 Capacitor Voltage Ratings

Capacitors are also rated by their maximum operating voltage. The importance of selecting a capacitor with the proper voltage rating is often overlooked. Exceeding the voltage rating, even momentarily, can cause excessive heating, a permanent shift of the capacitance value, a short circuit, or outright destruction. As a result, the voltage rating should be at least 25% higher than the working voltage across the capacitor; many designers use 50 – 100%.

Following the 25% guideline, filter capacitors for a 12-V system should have at least a 15-V rating (12 V × 1.25). However, 12-V systems such as 12-V power supplies and automotive 12-V electrical systems actually operate near 13.8 V and in the case of automotive systems, as high as 15 V. In such cases, capacitors rated for 15 V would be an insufficient margin of safety; 20 to 25-V capacitors should be used in such cases.

In large signal ac circuits, the maximum voltage rating of the capacitor should be based on the peak-to-peak voltages present. For example, the output of a 5-W QRP transmitter is 16 V_{RMS}, or about 45 V_{P-P}. Capacitors exposed to the 5 W RF power, such as in the output low-pass filter, should be rated well above 50 V for the 25% rule. A 100 W transmitter produces RF voltages of about 200 V_{P-P}. Remember, too, that ac line voltage is given as RMS, with peak-to-peak voltage 2.83 times higher: 120 V_{RMS} = 339 V_{P-P}.

Capacitors that are to be connected to primary ac circuits (directly to the ac line) for filtering or coupling *must* be rated for ac line use. These capacitors are listed as Class X and Class Y with several levels of line voltage ratings. (See **www.allaboutcircuits.com/technical-articles/safety-capacitor-class-x-and-class-y-capacitors**). They are designed to have failure modes that minimize fire and other hazards in case of failure.

Applying peak-to-peak voltages approaching the maximum voltage rating will cause excessive heating of the capacitor. This, in turn, will cause a permanent shift in the capacitance value. This could be undesirable in the output low pass filter example cited above in trying to maintain the proper impedance match between transmitter and antenna.

Exceeding the maximum voltage rating can also cause a breakdown of the dielectric material in the capacitor. The voltage can jump between the plates causing momentary or permanent electrical shorts between the capacitor plates.

In electrolytic and tantalum capacitors, exceeding the voltage rating can produce extreme heating of the oil or wetting agent used as the dielectric material. The expanding gases can cause the capacitor to burst or explode.

These over-voltage problems are easily avoided by selecting a capacitor with a voltage rating 25 – 50% above the normal peak-to-peak operating voltage. **Table 4.13** lists standard working voltages for common capacitor types.

DIELECTRIC BREAKDOWN

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

The *breakdown voltage* of a dielectric depends on the chemical composition and thickness of the dielectric. Breakdown voltage is not directly proportional to the thickness; doubling the thickness does not quite double the breakdown voltage. A thick dielectric must be used to withstand high voltages. Since capacitance is inversely proportional to dielectric thickness (plate spacing) for a given plate area, a high-voltage capacitor must have more plate area than a low-voltage one of the same capacitance. High-voltage, high-capacitance capacitors are therefore physically large. *Dielectric strength* is specified in terms of a *dielectric withstanding voltage* (DWV), given in volts per mil (0.001 inch) at a specified temperature.

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

Dielectric breakdown occurs at a lower voltage between pointed or sharp-edged surfaces than between rounded and polished surfaces. This can be a problem when constructing or repairing high-voltage power supplies such as are used in RF power amplifiers. Consequently, the breakdown voltage between metal plates of any given spacing in air can be increased by buffing the edges of the plates. If the plates are damaged so they are no longer smooth, they may have to be polished or the capacitor replaced.

Table 4.13
Capacitor Standard Working Voltages

Ceramic	*Polyester*	*Electrolytic*	*Tantalum*
		6.3 V	6.3 V
		10 V	10 V
16 V		16 V	16 V
			20 V
25 V		25 V	25 V
		35 V	35 V
50 V	50 V	50 V	50 V
		63 V	63 V
100 V	100 V	100 V	
	150 V	150 V	
200 V	200 V		
	250 V	250 V	

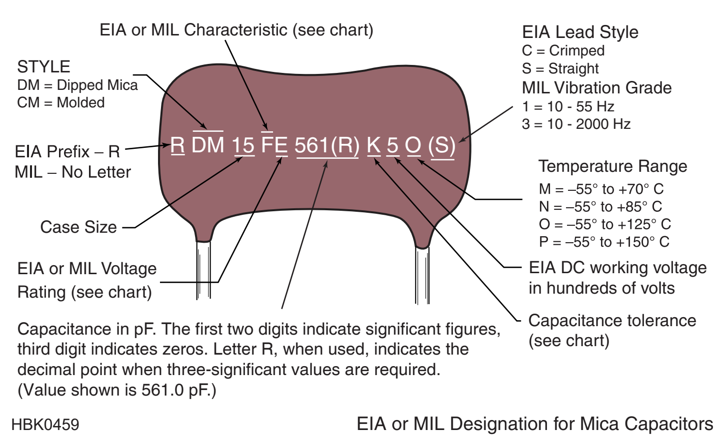
4.3.4 Capacitor Identification

Capacitors are identified by the IEC numerical code as shown in **Figure 4.11**. Color coding schemes used prior to 2000 are becoming rare, used only by a few non-US manufacturers. Some thru-hole "gum drop" tantalum capacitors also still use the color codes of **Figure 4.12**. Electrolytic and tantalum capacitors are often labeled with capacitance and working voltage in µF and V as in Figure 4.12C.

The IEC capacitor identification scheme is shown in **Figure 4.13**. Similar to the resistor IEC code, numerals are used to indicate the significant numerals and the multiplier, followed by an alphabetic character to indicate the tolerance. The multiplier is simply the number of zeros following the significant numerals. For example, a capacitor marked with "122K" would be a 1,200 pF 10% capacitor. The use of R to denote a decimal point in a value can be confusing if pF or µF are not specified. Generally, an inspection of the capacitor will determine which is correct, but a capacitance meter may be required. Additional digits and codes may be encountered

Table 4.14
European Marking Standards for Capacitors

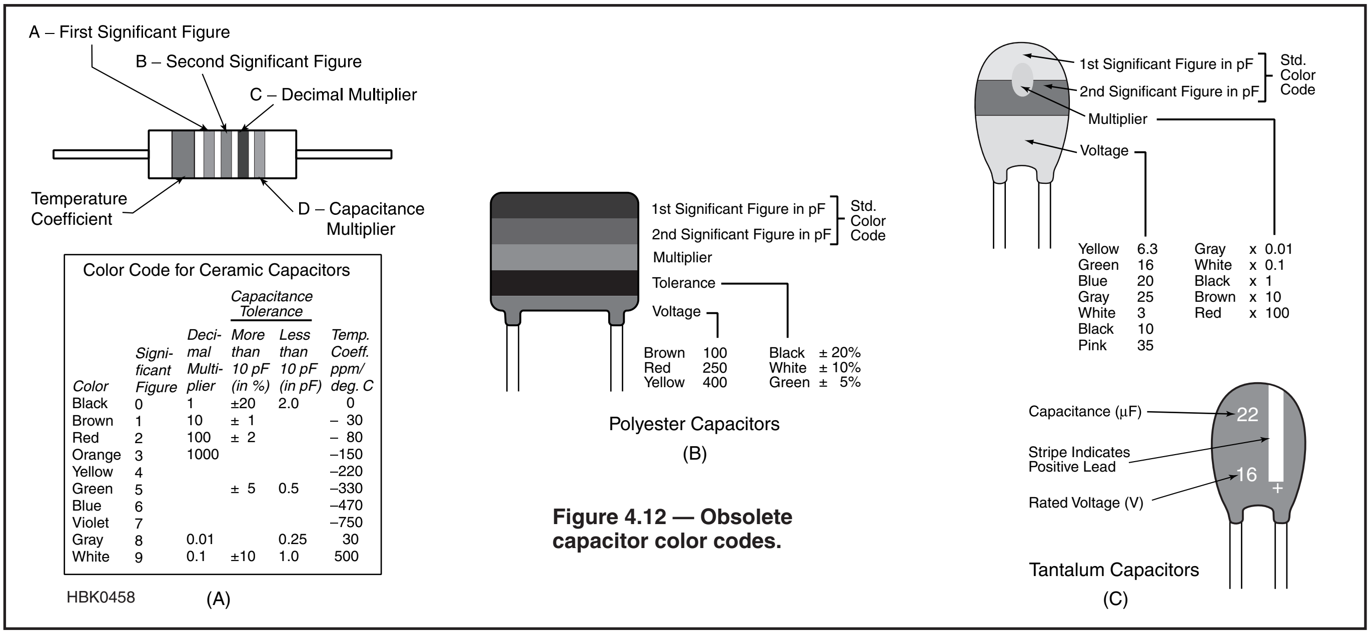
Marking	*Value*
1p	1 pF
2p2	2.2 pF
10p	10 pF
100p	100 pF
1n	1 nF (= 0.001 µF)
2n2	2.2 nF (= 0.0022 µF)
10n	10 nF (= 0.01 µF)
100n	100 nF (= 0.1 µF)
1u	1 µF
5u6	5.6 µF
10u	10 µF
100u	100 µF



Letter Designator	*"Characteristic" Max Capacitance Drift*	*"Characteristic" Max Range of Temp Coeff (ppm / deg. C)*	*MIL Voltage Rating (V)*	*Capactitance Tolerance (Percent)*
A	-	-	100	-
B	Not Specified	Not Specified	250	-
C	±(0.5% + 0.1 pF)	±200	300	-
D	±(0.3% + 0.1 pF)	±100	500	-
E	±(0.1% + 0.1 pF)	–20 to +100	600	-
F	±(0.05% + 0.1 pF)	0 to +70	1000	±1
G	-	-	1200	±2
H	-	-	1500	-
J	-	-	2000	±5
K	-	-	2500	±10
L	-	-	3000	-
M	-	-	4000	±20

MIL voltage ratings for other letter designators: N=5000 V, P=6000 V, Q=8000 V, R=10,000 V, S-12,000 V, T=15,000 V, U=20,000 V, V=25,000 V, W=30,000 V, X=35,000 V.

Figure 4.11 — Complete capacitor labeling scheme.



Color Code for Ceramic Capacitors

Color	Significant Figure	Decimal Multiplier	Capacitance Tolerance: More than 10 pF (in %)	Capacitance Tolerance: Less than 10 pF (in pF)	Temp. Coeff. ppm/ deg. C
Black	0	1	±20	2.0	0
Brown	1	10	± 1		– 30
Red	2	100	± 2		– 80
Orange	3	1000			–150
Yellow	4				–220
Green	5		± 5	0.5	–330
Blue	6				–470
Violet	7				–750
Gray	8	0.01		0.25	30
White	9	0.1	±10	1.0	500

Figure 4.12 — Obsolete capacitor color codes.

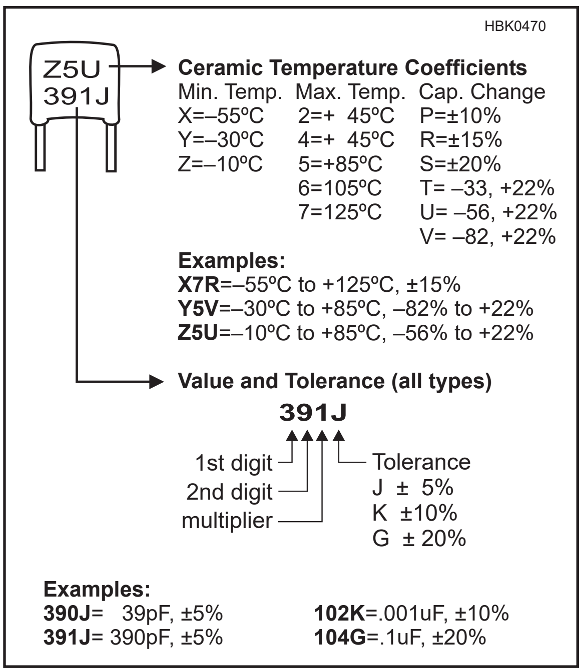


Figure 4.13 — Abbreviated IEC capacitor identification. This method is used on SMT capacitors. An "R" in the numeric field stands for "radix" and represents a decimal point so the "4R7" indicates "4.7" for example.

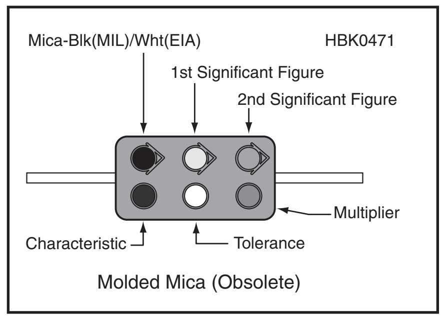


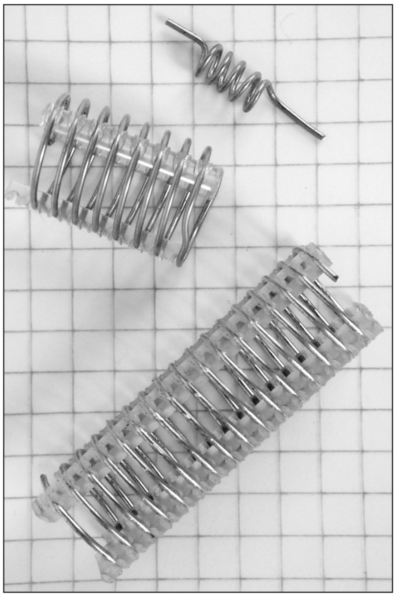
Figure 4.14 — Obsolete JAN "postage stamp" capacitor labeling.

as shown in Figure 4.11.

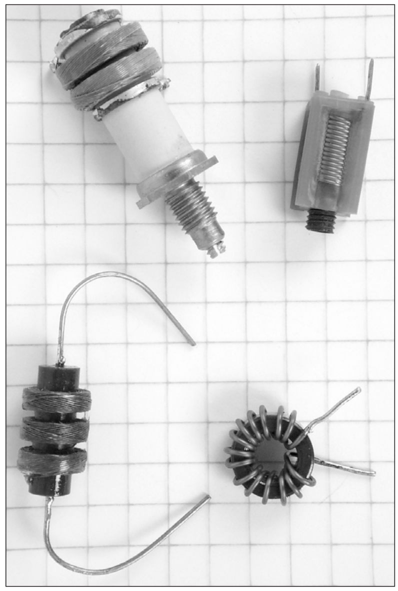
Military-surplus equipment using the obsolete "postage stamp" capacitors is still encountered in amateur radio. These capacitors used the colored dot method of value identification shown in **Figure 4.14**.

European manufacturers often use nanofarads or nF, such that 10 nF, or simply 10N, indicates 10 nanofarads. This is equivalent to 10,000 pF or 0.01 µF. This notational scheme, shown in **Table 4.14**, is more commonly found on schematic diagrams than actual part markings.

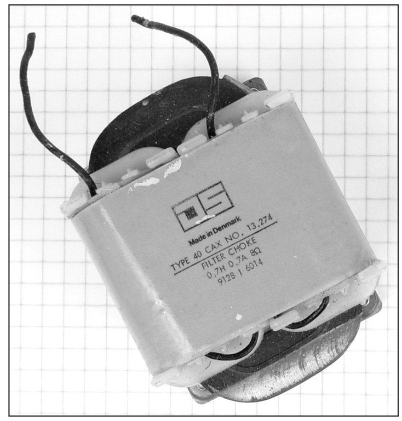
4.4 Practical Inductors



(A)



(B)



(C)

Figure 4.15 — Part A shows small-value air-wound inductors. Part B shows some inductors with values in the range of a few millihenrys and C shows a large inductor such as might be used in audio circuits or as power-supply chokes. The ¼-inch-ruled graph paper background provides a size comparison.

4.4.1 Component Inductors

Various facets of radio circuits make use of inductors ranging from the tiny up to the massive. Small values of inductance, such as those inductors in **Figure 4.15A**, serve mostly in RF circuits. They may be self-supporting, air-core or air-wound inductors, or the winding may be supported by nonmagnetic strips or a form. Phenolic, certain plastics and ceramics are the most common *coil forms* for air-core inductors. These inductors range in value from a few hundred µH for medium- and high-frequency circuits down to tenths of a µH at VHF and UHF. **Table 4.15** is a list of the EIA standard inductor values.

The most common inductor in small-signal RF circuits is the *encapsulated inductor* or *RF choke*. These components look a lot like carbon-composition or film resistors and are often marked with colored paint stripes to indicate value. (See **Figure 4.16**) The body of these components is often green to identify them as inductors and not resistors. Measure the resistance of the component with an ohmmeter if there is any doubt as to the identity of the component. Pay attention to the limit current-carrying abilities of these inductors. These inductors have values from less than 1 µH to a few mH.

It is possible to make solenoid inductors variable by inserting a moveable *slug* in the center of the inductor. (Slug-tuned inductors normally have a ceramic, plastic, or phenolic insulating form between the conductive slug and the inductor winding.) If the slug material is magnetic, such as powdered iron, the induc-

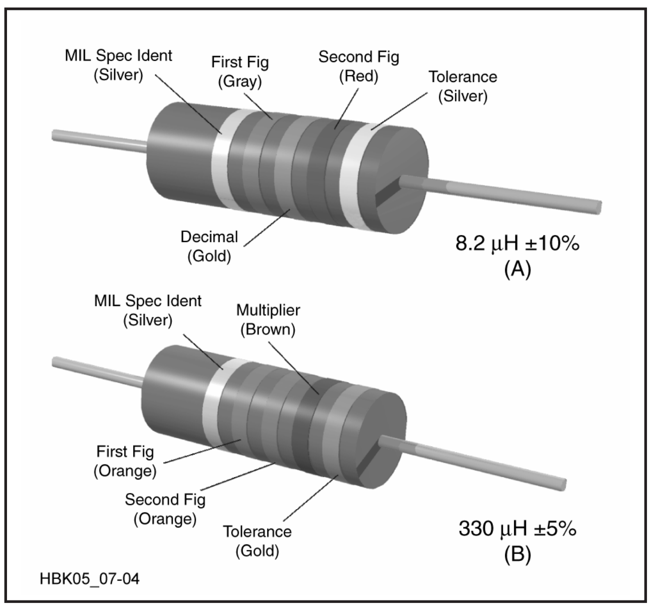


Figure 4.16 — Color coding for cylindrical encapsulated RF chokes. At A, an example of the coding for an 8.2-µH choke is given. At B, the color bands for a 330-µH inductor are illustrated. The color code is given in Table 4.2.

Table 4.15
EIA Standard Inductor Values

µH	*µH*	*µH*	*mH*	*mH*	*mH*
1.0	10	100	1.0	10	100
1.2	12	120	1.2	12	120
1.5	15	150	1.5	15	150
2.2	22	220	2.2	22	220
2.7	27	270	2.7	27	270
3.3	33	330	3.3	33	330
3.9	39	390	3.9	39	390
4.7	47	470	4.7	47	470
5.6	56	560	5.6	56	560
6.8	68	680	6.8	68	680
8.2	82	820	8.2	82	820

tance increases as the slug is moved into the center of the inductor. If the slug is brass or some other nonmagnetic material, inserting the slug will reduce the inductor's inductance.

Alternatives to air-core inductors for RF work are *toroidal* inductors (or *toroids*) wound on powdered-iron or ferrite cores. The availability of many types and sizes of powdered-iron cores has made these inductors popular for low-power fixed-value service. The toroidal shape concentrates the inductor's field nearly completely inside the inductor, eliminating the need in many cases for other forms of shielding to limit the interaction of the inductor's magnetic field with the fields of other inductors. (Ferrite core materials are discussed here and in the chapter on **RF Techniques**.)

Figure 4.15B shows samples of inductors in the millihenry (mH) range. Among these inductors are multi-section RF chokes designed to block RF currents from passing beyond them to other parts of circuits. Low-frequency radio work may also use inductors in this range of values, sometimes wound with *Litz wire*. Litz wire is a special version of stranded wire, with each strand insulated from the others, and is used to minimize losses associated with skin effect.

For audio filters, toroidal inductors with values up to 100 mH are useful. Resembling powdered-iron-core RF toroids, these inductors are wound on ferrite or molybdenum-permalloy cores having much higher permeabilities.

Audio and power-supply inductors appear in Figure 4.15C. Lower values of these iron-core inductors, in the range of a few henrys, are useful as audio-frequency chokes. Larger values up to about 20 H may be found in power supplies, as choke filters, to suppress 120-Hz ripple. Although some of these inductors are open frame, most have iron covers to confine the powerful magnetic fields they produce.

Although builders and experimenters rarely construct their own capacitors, inductor fabrication for low- and high-power RF circuits is common. In fact, it is often necessary, since commercially available units may be unavailable or expensive. Even if available, they may consist of inductor stock to be trimmed to the required value. Core materials and wire for winding both solenoid and toroidal inductors are readily available. The following information includes fundamental formulas and design examples for calculating practical inductors, along with additional data on the theoretical limits in the use of some materials.

4.4.2 Air-Core Inductors

Many circuits require air-core inductors using just one layer of wire. The approximate inductance of a single-layer air-core inductor may be calculated from the simplified formula:

$$L\ (\mu H) = \frac{d^2\ n^2}{18d + 40\ell}$$

where

L = inductance in microhenrys,

d = inductor diameter in inches (from wire center to wire center),

ℓ = inductor length in inches, and

n = number of turns.

If dimensions are given in cm, the equation is:

$$L(\mu H) = \frac{d^2 n^2}{6.4(7.2d + 15.8\ell)}$$

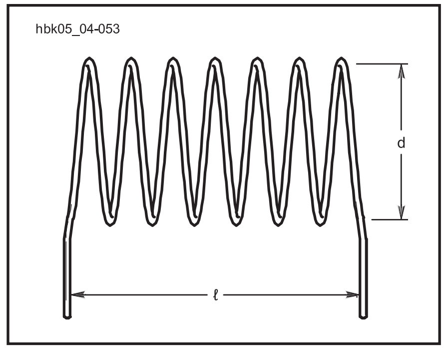


Figure 4.17 — Coil dimensions used in the inductance formula for air-core inductors.

The notation is illustrated in **Figure 4.17**. This formula is a close approximation for inductors having a length equal to or greater than 0.4d. (Note: Inductance varies as the square of the turns. If the number of turns is doubled, the inductance is quadrupled. This relationship is inherent in the equation but is often overlooked. For example, to double the inductance, add additional turns equal to 1.4 times the original number of turns, or 40% more turns.)

Example: What is the inductance of an inductor if the inductor has 48 turns wound at 32 turns per inch and a diameter of ¾ inch? In this case, d = 0.75, ℓ = 48/32 = 1.5 and n = 48.

$$L\ (\mu H) = \frac{0.75^2 \times 48^2}{(18 \times 0.75) + (40 \times 1.5)}$$

$$= \frac{1300}{74} = 18\ \mu H$$

To calculate the number of turns of a single-layer inductor for a required value of inductance, the formula becomes:

$$n = \frac{\sqrt{L(18d + 40\ell)}}{d}\ \text{(inches)}$$

$$n = \frac{\sqrt{L(7.2d + 15.8\ell)}}{d}\ \text{(cm)}$$

Example: Suppose an inductance of 10.0 µH is required. The form on which the inductor is to be wound has a diameter of one inch and is long enough to accommodate an inductor of 1¼ inches. Then d = 1.00 inch, ℓ= 1.25 inches and L = 10.0. Substituting:

$$n = \frac{\sqrt{10.0\,[\,(18 \times 1.0) + (40 \times 1.25)}}{1}$$

$$= \sqrt{680} = 26.1 \text{ turns}$$

A 26-turn inductor would be close enough in practical work. Since the inductor will be 1.25 inches long, the number of turns per inch will be 26.1 / 1.25 = 20.9. Consulting the wire table, we find that #17 AWG enameled wire (or anything smaller) can be used. The proper inductance is obtained by winding the required number of turns on the form and then adjusting the spacing between the turns to make a uniformly spaced inductor 1.25 inches long.

Most inductance formulas lose accuracy when applied to small inductors (such as those used in VHF work and in low-pass filters built for reducing harmonic interference to televisions) because the conductor thickness is no longer negligible in comparison with the size of the inductor. **Figure 4.18** shows the measured inductance of VHF inductors and may be used as a basis for circuit design. Two curves are given: curve A is for inductors wound to an inside diameter of ½ inch; curve B is for inductors of ¾ -inch inside diameter. In both curves, the wire size is #12 AWG and the winding pitch is eight turns to the inch (⅛-inch turn spacing). The inductance values include leads ½-inch long.

Machine-wound inductors with the preset diameters and turns per inch were sold as Miniductor and Airdux coils. These are no longer manufactured but are commonly available as surplus and NOS (new old stock). Charts and tables of the inductance of these coils is provided in the online content.

Forming a wire into a solenoid increases its inductance but also introduces distributed capacitance. Since each turn is at a slightly different ac potential, each pair of turns effectively forms a capacitor in parallel with part of the inductor. (See the chapter on **RF Techniques** for information on the effects of these and other factors that affect the behavior of the "ideal" inductors discussed in this chapter.)

Moreover, the Q of air-core inductors is, in part, a function of the inductor shape, specifically its ratio of length to diameter. Q tends to be highest when these dimensions are nearly equal. With wire properly sized to the current carried by the inductor, and with high-caliber construction, air-core inductors can achieve Q above 200.

For a large collection of formulas useful in

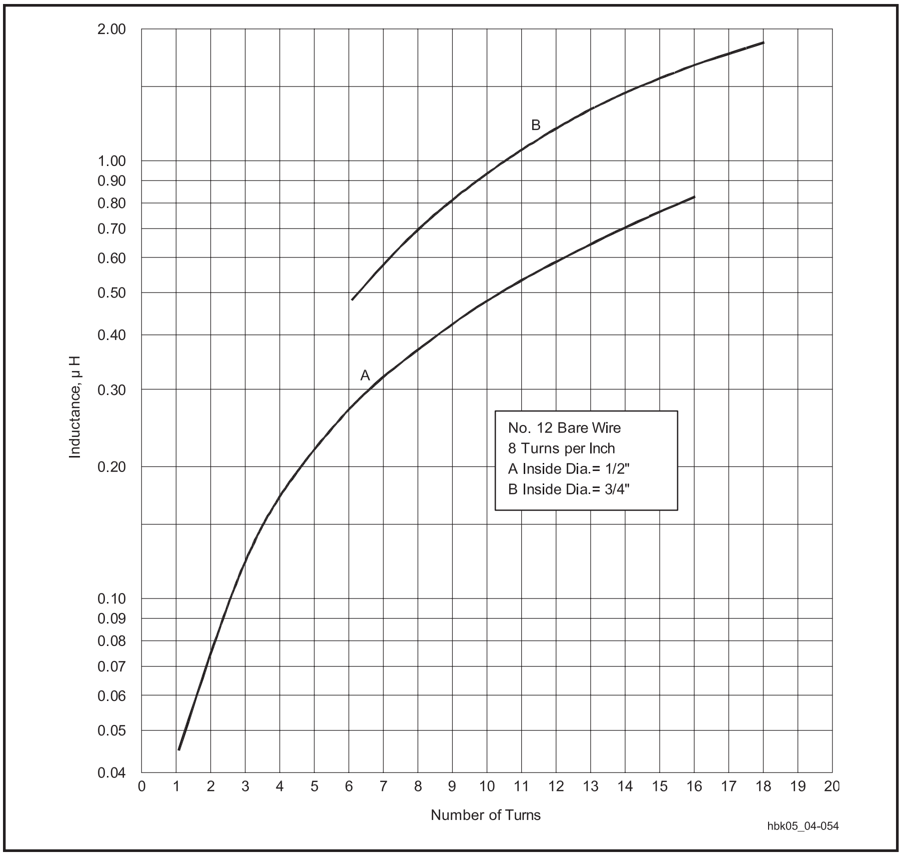


Figure 4.18 — Measured inductance of coils wound with #12 bare wire, eight turns to the inch. The values include half-inch leads.

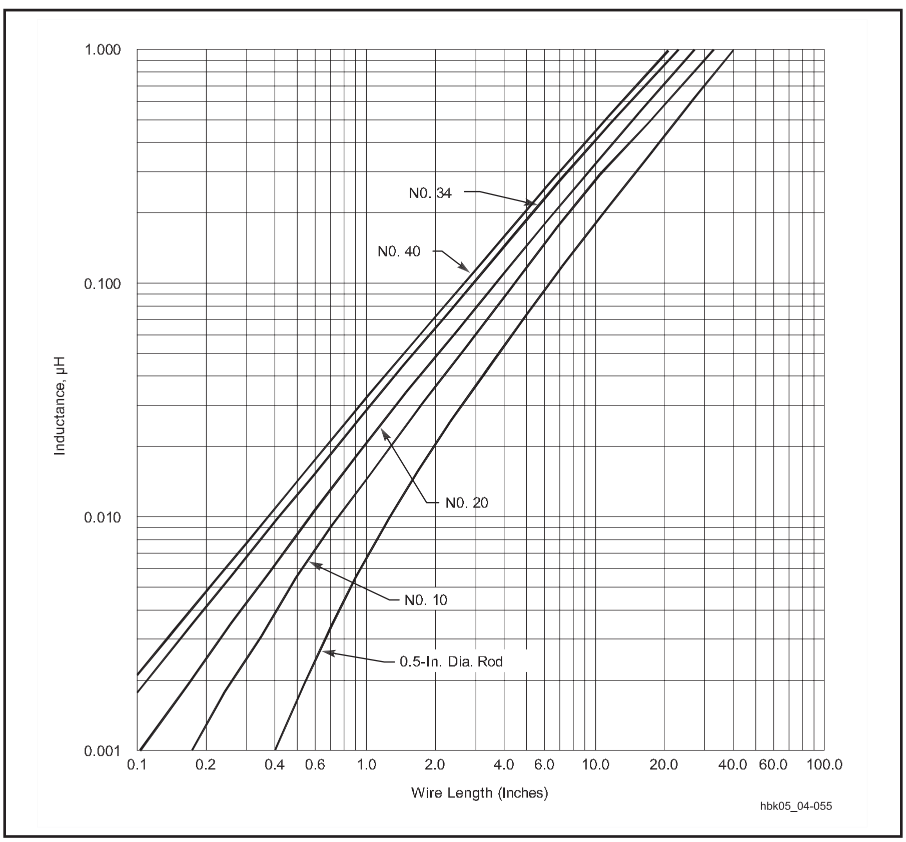


Figure 4.19 — Inductance of various conductor sizes as straight wires.

constructing air-core inductors of many configurations, see the "Circuit Elements" section in Terman's *Radio Engineers' Handbook* (listed in the References section).

4.4.3 Straight-Wire Inductance

At low frequencies the inductance of a straight, round, nonmagnetic wire in free space is given by:

$$L = 0.00508\, b \left\{ \left[\ln\left(\frac{2\,b}{a}\right) \right] - 0.75 \right\}$$

where

L = inductance in µH,
a = wire radius in inches,
b = wire length in inches, and
ln = natural logarithm = 2.303 × common logarithm (base 10).

If the dimensions are expressed in millimeters instead of inches, the equation may still be used, except replace the 0.00508 value with 0.0002.

Skin effect reduces the inductance at VHF and above. As the frequency approaches infinity, the 0.75 constant within the brackets approaches unity. As a practical matter, skin effect will not reduce the inductance by more than a few percent.

Example: What is the inductance of a wire that is 0.1575 inch in diameter and 3.9370 inches long? For the calculations, a = 0.0787 inch (radius) and b = 3.9370 inch.

$$L = 0.00508\, b \left\{ \left[\ln\left(\frac{2\,b}{a}\right) \right] - 0.75 \right\}$$

$$= 0.00508\,(3.9370) \times \left\{ \left[\ln\left(\frac{2 \times 3.9370}{0.0787}\right) \right] - 0.75 \right\}$$

$$= 0.20 \times [\ln(100) - 0.75)]$$

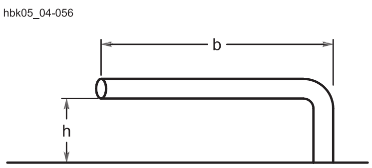
$$= 0.20 \times (4.60 - 0.75)$$

$$= 0.20 \times 3.85 = 0.077\ \mu H$$

Figure 4.19 is a graph of the inductance for wires of various radii as a function of length.

A VHF or UHF tank circuit can be fabricated from a wire parallel to a ground plane, with one end grounded. A formula for the inductance of such an arrangement is given in **Figure 4.20**.

Example: What is the inductance of a wire 3.9370 inches long and 0.0787 inch in radius, suspended 1.5748 inch above a ground plane? (The inductance is measured between the free end and the ground plane, and the formula includes the inductance of the 1.5748-inch



$$L = 0.0117b\left\{\text{Log}_{10}\left[\frac{2h}{a}\left(\frac{b+\sqrt{b^2+a^2}}{b+\sqrt{b^2+4h^2}}\right)\right]\right\} + 0.00508\left(\sqrt{b^2+4h^2} - \sqrt{b^2+a^2} + \frac{b}{4} - 2h + a\right)$$

where
L = inductance in μH
a = wire radius in inches
b = wire length parallel to ground plane in inches
h = wire height above ground plane in inches

Figure 4.20 — Equation for determining the inductance of a wire parallel to a ground plane, with one end grounded. If the dimensions are in millimeters, the numerical coefficients become 0.0004605 for the first term and 0.0002 for the second term.

grounding link.) To demonstrate the use of the formula in Figure 4.20, begin by evaluating these quantities:

$$b + \sqrt{b^2 + a^2}$$

$$= 3.9370 + \sqrt{3.9370^2 + 0.0787^2}$$

$$= 3.9370 + 3.94 = 7.88$$

$$b + \sqrt{b^2 + 4\,(h^2)}$$

$$= 3.9370 + \sqrt{3.9370^2 + 4\,(1.5748^2)}$$

$$= 3.9370 + \sqrt{15.50 + 4\,(2.480)}$$

$$= 3.9370 + \sqrt{15.50 + 9.920}$$

$$= 3.9370 + 5.0418 = 8.9788$$

$$\frac{2\,h}{a} = \frac{2 \times 1.5748}{0.0787} = 40.0$$

$$\frac{b}{a} = \frac{3.9370}{4} = 0.098425$$

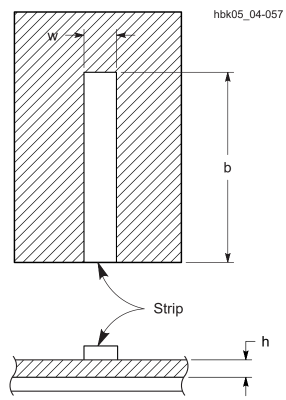
Substituting these values into the formula yields:

$$L = 0.0117 \times 3.9370\left\{\log_{10}\left[40.0 \times \left(\frac{7.88}{8.9788}\right)\right]\right\}$$

$$+ 0.00508 \times (5.0418 - 3.94 + 0.98425 - 3.1496 + 0.0787)$$

$$L = 0.0662\ \mu H$$

Another conductor configuration that is frequently used is a flat strip over a ground plane. This arrangement has lower skin-effect loss at high frequencies than round wire because it has a higher surface-area to volume ratio. The inductance of such a strip can be found from the formula in **Figure 4.21**.



$$L = 0.00508\ b\left(\ln\frac{2b}{w+h} + 0.5 + 0.2235\,\frac{w+h}{b}\right)$$

where
L = inductance in μH
b = length in inches
w = width in inches
h = thickness in inches

Figure 4.21 — Equation for determining the inductance of a flat strip inductor.

4.4.4 Iron-Core Inductors

If the permeability of an iron core in an inductor is 800, then the inductance of any given air-wound inductor is increased 800 times by inserting the iron core. The inductance will be proportional to the magnetic flux through the inductor, other things being equal. The inductance of an iron-core inductor is highly dependent on the current flowing in the inductor, in contrast to an air-core inductor, where the inductance is independent of current because air does not saturate.

Iron-core inductors are used chiefly in power-supply equipment. They usually have direct current flowing through the winding, and any variation in inductance with current is usually undesirable. Inductance variations may be overcome by keeping the flux density below the saturation point of the iron. Opening the core so there is a small air gap will achieve this goal, as discussed in the earlier section on inductors. The reluctance or magnetic resistance introduced by such a gap is very large compared with that of the iron, even though the gap is only a small fraction of an inch. Therefore, the gap — rather than the iron — controls the flux density. Air gaps in iron cores reduce the inductance, but they hold the value practically constant regardless of the current magnitude.

When alternating current flows through an inductor wound on an iron core, a voltage is induced. Since iron is a conductor, eddy currents also flow in the core as discussed earlier. Eddy currents represent lost power because they flow through the resistance of the iron and generate heat. Losses caused by eddy currents can be reduced by laminating the core (cutting the core into thin strips). These strips or laminations are then insulated from each other by painting them with some insulating material such as varnish or shellac. Eddy current losses add to hysteresis losses, which are also significant in iron-core inductors.

Eddy-current and hysteresis losses in iron increase rapidly as the frequency of the alternating current increases. For this reason, ordinary iron cores can be used only at power-line and audio frequencies — up to approximately 15,000 Hz. Even then, a very good grade of iron or steel is necessary for the core to perform well at the higher audio frequencies. Laminated iron cores become completely useless at radio frequencies because of eddy current and hysteresis losses.

4.4.5 Slug-Tuned Inductors

For RF work, the losses in iron cores can be reduced to a more useful level by grinding the iron into a powder and then mixing it with a binder of insulating material in such a way that the individual iron particles are insulated from each other. Using this approach, cores can be made that function satisfactorily even into the VHF range. Because a large part of the magnetic path is through a nonmagnetic material (the binder), the permeability of the powdered iron core is low compared with the values for solid iron cores used at power-line frequencies.

The slug is usually shaped in the form of a cylinder that fits inside the insulating form on which the inductor is wound. Despite the fact that the major portion of the magnetic path for the flux is in air, the slug is quite effective in increasing the inductor inductance. By pushing (or screwing) the slug in and out of the inductor, the inductance can be varied over a considerable range.

4.4.6 Powdered-Iron Toroidal Inductors

For fixed-value inductors intended for use at HF and VHF, the powdered-iron toroidal core has become the standard in low- and medium-power circuits. **Figure 4.22** shows the general outlines of a toroidal inductor on a magnetic core.

Manufacturers offer a wide variety of core materials, or *mixes*, to create conductor cores that will perform over a desired frequency range with a reasonable permeability. Permeabilities for powdered-iron cores fall in the range of 3 to 35 for various mixes. In addition, core sizes are available in the range of 0.125-inch outside diameter (OD) up to 1.06-inch OD, with larger sizes to 5-inch OD available in certain mixes. The range of sizes permits the builder to construct single-layer inductors for almost any value using wire sized to meet the circuit current demands.

The use of powdered iron in a binder reduces core losses usually associated with iron, while the permeability of the core permits a reduction in the wire length and associated resistance in forming an inductor of a given inductance. Therefore, powdered-iron-core toroidal inductors can achieve Q well above 100, often approaching or exceeding 200 within the frequency range specified for a given core. Moreover, these inductors are considered *self-shielding* since most of the magnetic flux is within the core, a fact that simplifies circuit design and construction.

Table 4.16 lists the properties of common powdered-iron cores. Formulas are given for calculating the number of required turns based on a given inductance and for calculating the inductance given a specific number of turns. Most powdered-iron toroid cores that amateurs use are manufactured by Micrometals (**www.micrometals.com**). Paint is used to identify the material used in the core. The Micrometals color code is part of Table 4.16. **Table 4.17** gives the physical dimensions of powdered-iron toroids.

Each powdered-iron core has an *inductance factor* or *index* A_L determined by the manufacturer. Amidon specifies A_L in µH per 100 turns-squared and other manufacturers specify A_L in µH or nH per turns-squared. Check the manufacturer's website or product information for the correct units for A_L and method of calculating the inductance or

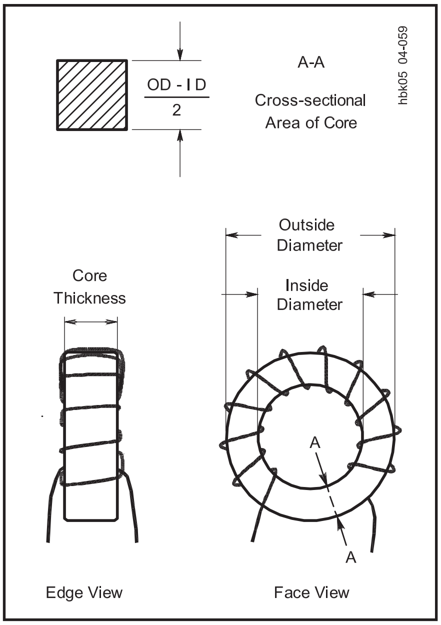


Figure 4.22 — A typical toroidal inductor wound on a powdered-iron or ferrite core. Some key physical dimensions are noted. Equally important are the core material, its permeability, its intended range of operating frequencies, and its A_L value. This is an 11-turn toroid.

desired number of turns.

$$L = \frac{A_L \times N^2}{10000} \text{ (Amidon) or}$$

$$L = A_L \times N^2 \text{ (Other manufacturers)}$$

where

L = the inductance in µH,
A_L = the inductance index (see text above)
N = the number of turns.

The builder must then ensure that the core is capable of holding the calculated number of turns of wire of the required wire size.

Example: What is the inductance of a 60-turn inductor on a core with an A_L of 55 µH/100-turns²?

$$L = \frac{A_L \times N^2}{10000} = \frac{55 \times 60^2}{10000}$$

$$= \frac{198000}{10000} = 19.8 \ \mu\text{H}$$

Example: What is the inductance of a 20-turn inductor on a core with an A_L of 0.3 µH/turns²?

$$L = A_L \times N^2 = 0.3 \times 20^2$$
$$= 0.3 \times 400 = 120 \ \mu\text{H}$$

To calculate the number of turns needed for a particular inductance, use the formula:

$$N = 100\sqrt{\frac{L}{A_L}} \text{ (Amidon) or}$$

$$N = \sqrt{\frac{L}{A_L}} \text{ (Other manufacturers)}$$

Example: How many turns are needed for a 12.0-µH inductor if the A_L for the selected core is 49 µH/100-turns²?

$$N = 100\sqrt{\frac{L}{A_L}} = 100\sqrt{\frac{12.0}{49}}$$

$$= 100\sqrt{0.245} = 100 \times 0.495 = 49.5 \text{ turns}$$

Example: How many turns are needed for a 300-nH inductor if the A_L for the selected core is 12 nH/turns²?

$$N = \sqrt{\frac{L}{A_L}} = \sqrt{\frac{300}{12}} = \sqrt{25} = 5 \text{ turns}$$

Count turns by each pass of the wire through the center of the core. (*A straight wire through a toroidal core counts as a one-turn inductor.*) Fine adjustment of the inductance may be possible by spreading or compressing inductor turns.

If the value is critical, experiment by starting with an extra turn or two, then measure the inductance or test the circuit. Core characteristics may vary slightly from batch to batch and winding style has a small effect on inductance, as well.

The power-handling ability of toroidal cores depends on many variables, which include the cross-sectional area through the core, the core material, the numbers of turns in the inductor, the applied voltage and the operating frequency. Although powdered-iron cores can withstand dc flux densities up to 5,000 gauss without saturating, ac flux densities from sine waves above certain limits can overheat cores. Manufacturers provide guideline limits for ac flux densities to avoid overheating. The limits range from 150 gauss at 1 MHz to 30 gauss at 28 MHz, although the curve is not linear. To calculate the maximum anticipated flux density for a particular inductor, use the formula:

$$B_{max} = \frac{E_{RMS} \times 10^8}{4.44 \times A_e \times N \times f}$$

where

B_{max} = the maximum flux density in gauss,
E_{RMS} = the voltage across the inductor,
A_e = the cross-sectional area of the core in square centimeters,
N = the number of turns in the inductor, and
f = the operating frequency in Hz.

Example: What is the maximum ac flux density for an inductor of 15 turns if the frequency is 7.0 MHz, the RMS voltage is 25 V

Table 4.16
Powdered-Iron Toroidal Cores: Magnetic Properties

There are differing conventions for referring to the type of core material: #, mix and type are all used. For example, all of the following designate the same material: #12, Mix 12, 12-Mix, Type 12 and 12-Type.

Inductance and Turns Formula

The turns required for a given inductance or inductance for a given number of turns can be calculated from:

$$N = 100\sqrt{\frac{L}{A_L}} \qquad L = A_L\left(\frac{N^2}{10{,}000}\right) \qquad \text{(Amidon cores)}$$

$$N = \sqrt{\frac{L}{A_L}} \qquad L = A_L \times N^2 \qquad \text{(Non-Amidon cores)}$$

where N = number of turns; L = desired inductance; A_L = inductance index (µH per 100 turns-squared for Amidon cores; nH or µH per turns-squared for non-Amidon cores; see core data sheet)

Amidon Associates literature gives the value of A_L as inductance per 100 turns but the correct units are inductance per 100 turns-squared. The units of inductance are generally in nH but may also be mH. Make sure you understand which units apply and use the A_L value and formula provided by the manufacturer of the core to calculate number of turns or inductance.

Toroid diameter is indicated by the number following "T" — T-200 is 2.00 in. dia; T-68 is 0.68 in. diameter, etc.

A_L Values — ***Mix***

Size	*26**	*3*	*15*	*1*	*2*	*7*	*6*	*10*	*12*	*17*	*0*
T-12	na	60	50	48	20	18	17	12	7.5	7.5	3.0
T-16	145	61	55	44	22	na	19	13	8.0	8.0	3.0
T-20	180	76	65	52	27	24	22	16	10.0	10.0	3.5
T-25	235	100	85	70	34	29	27	19	12.0	12.0	4.5
T-30	325	140	93	85	43	37	36	25	16.0	16.0	6.0
T-37	275	120	90	80	40	32	30	25	15.0	15.0	4.9
T-44	360	180	160	105	52	46	42	33	18.5	18.5	6.5
T-50	320	175	135	100	49	43	40	31	18.0	18.0	6.4
T-68	420	195	180	115	57	52	47	32	21.0	21.0	7.5
T-80	450	180	170	115	55	50	45	32	22.0	22.0	8.5
T-94	590	248	200	160	84	na	70	58	32.0	na	10.6
T-106	900	450	345	325	135	133	116	na	na	na	19.0
T-130	785	350	250	200	110	103	96	na	na	na	15.0
T-157	870	420	360	320	140	na	115	na	na	na	na
T-184	1640	720	na	500	240	na	195	na	na	na	na
T-200	895	425	na	250	120	105	100	na	na	na	na

*Mix-26 is similar to the older Mix-41, but can provide an extended frequency range.

Magnetic Properties Iron Powder Cores

Mix	*Color*	*Material*	µ	*Temp stability (ppm/°C)*	*f (MHz)*	*Notes*
26	Yellow/white	Hydrogen reduced	75	825	dc - 1	Used for EMI filters and dc chokes
3	Gray	Carbonyl HP	35	370	0.05 - 0.50	Excellent stability, good Q for lower frequencies
15	Red/white	Carbonyl GS6	25	190	0.10 - 2	Excellent stability, good Q
1	Blue	Carbonyl C	20	280	0.50 - 5	Similar to Mix-3, but better stability
2	Red	Carbonyl E	10	95	2 - 30	High Q material
7	White	Carbonyl TH	9	30	3 - 35	Similar to Mix-2 and Mix-6, but better temperature stability
6	Yellow	Carbonyl SF	8	35	10 - 50	Very good Q and temperature stability for 20-50 MHz
10	Black	Powdered iron W	6	150	30 - 100	Good Q and stability for 40 - 100 MHz
12	Green/white	Synthetic oxide	4	170	50 - 200	Good Q, moderate temperature stability
17	Blue/yellow	Carbonyl	4	50	40 - 180	Similar to Mix-12, better temperature stability, Q drops about 10% above 50 MHz, 20% above 100 MHz
0	Tan	phenolic	1	0	100 - 300	Inductance may vary greatly with winding technique

Courtesy of Amidon Assoc and Micrometals
Note: Color codes hold only for cores manufactured by Micrometals, which makes the cores sold by most amateur radio distributors.

and the cross-sectional area of the core is 0.133 cm²?

$$B_{max} = \frac{E_{RMS} \times 10^8}{4.44 \times A_e \times N \times f}$$

$$= \frac{25 \times 10^8}{4.44 \times 0.133 \times 15 \times 7.0 \times 10^6}$$

$$= \frac{25 \times 10^8}{62 \times 10^6} = 40 \text{ gauss}$$

Since the recommended limit for cores operated at 7 MHz is 57 gauss, this inductor is well within guidelines.

Table 4.17

Powdered-Iron Toroidal Cores: Dimensions

Toroid diameter is indicated by the number following "T" — T-200 is 2.00 in. dia; T-68 is 0.68 in. diameter, etc.
See Table 4.16 for a core sizing guide.

Red E Cores—500 kHz to 30 MHz (μ = 10)

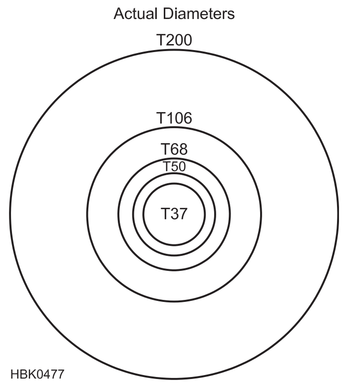
No.	*OD (in)*	*ID (in)*	*H (in)*
T-200-2	2.00	1.25	0.55
T-94-2	0.94	0.56	0.31
T-80-2	0.80	0.50	0.25
T-68-2	0.68	0.37	0.19
T-50-2	0.50	0.30	0.19
T-37-2	0.37	0.21	0.12
T-25-2	0.25	0.12	0.09
T-12-2	0.125	0.06	0.05

Black W Cores—30 MHz to 200 MHz (μ=6)

No.	*OD (In)*	*ID (In)*	*H (In)*
T-50-10	0.50	0.30	0.19
T-37-10	0.37	0.21	0.12
T-25-10	0.25	0.12	0.09
T-12-10	0.125	0.06	0.05

Yellow SF Cores—10 MHz to 90 MHz (μ=8)

No.	*OD (In)*	*ID (In)*	*H (In)*
T-94-6	0.94	0.56	0.31
T-80-6	0.80	0.50	0.25
T-68-6	0.68	0.37	0.19
T-50-6	0.50	0.30	0.19
T-26-6	0.25	0.12	0.09
T-12-6	0.125	0.06	0.05



Number of Turns vs Wire Size and Core Size

Approximate maximum number of turns—single layer wound—enameled wire.

Wire Size	*T-200*	*T-130*	*T-106*	*T-94*	*T-80*	*T-68*	*T-50*	*T-37*	*T-25*	*T-12*
10	33	20	12	12	10	6	4	1		
12	43	25	16	16	14	9	6	3		
14	54	32	21	21	18	13	8	5	1	
16	69	41	28	28	24	17	13	7	2	
18	88	53	37	37	32	23	18	10	4	1
20	111	67	47	47	41	29	23	14	6	1
22	140	86	60	60	53	38	30	19	9	2
24	177	109	77	77	67	49	39	25	13	4
26	223	137	97	97	85	63	50	33	17	7
28	281	173	123	123	108	80	64	42	23	9
30	355	217	154	154	136	101	81	54	29	13
32	439	272	194	194	171	127	103	68	38	17
34	557	346	247	247	218	162	132	88	49	23
36	683	424	304	304	268	199	162	108	62	30
38	875	544	389	389	344	256	209	140	80	39
40	1103	687	492	492	434	324	264	178	102	51

Actual number of turns may differ from above figures according to winding techniques, especially when using the larger size wires. Chart prepared by Michel J. Gordon, Jr, WB9FHC. Courtesy of Amidon Assoc.

4.4.7 Ferrite Toroidal Inductors

Although nearly identical in general appearance to powdered-iron cores, ferrite cores differ in a number of important characteristics. Composed of nickel-zinc ferrites for lower permeability ranges and of manganese-zinc ferrites for higher permeabilities, these cores span a permeability range from 20 to above 1,0000. Nickel-zinc cores with permeabilities from 20 to 800 are useful in high-Q applications, but function more commonly in amateur applications as RF chokes. They are also useful in wide-band transformers, discussed in the **RF Techniques** chapter. An excellent design resource for ferrite-based components is the Fair-Rite Materials Corp on-line catalog at **www.fair-rite.com**. The Fair-Rite website's "Technical Resources" section also has free papers on the use of ferrites for EMI suppression and broadband transformers. The following list presents the general characteristics (material and composition and intended application) of Fair-Rite's ferrite materials:

- Type 31 (MnZn) — EMI suppression applications from 1 MHz up to 500 MHz.
- Type 43 (NiZn) — Suppression of conducted EMI, inductors and HF common-mode chokes from 20 MHz to 250 MHz.
- Type 44 (NiZn) — EMI suppression from 30 MHz to 500 MHz.
- Type 61 (NiZn) — Inductors up to 25 MHz and EMI suppression above 200 MHz.
- Type 67 (NiZn) — Broadband transformers, antennas and high-Q inductors up to 50 MHz.
- Type 73 (MnZn) — Suppression of conducted EMI below 50 MHz.
- Type 75 (MnZn) — Broadband and pulse transformers.

See **Table 4.18** for information about the dimensions and magnetic properties of ferrite cores. Identification can often be difficult because ferrite cores are often unpainted, unlike powdered-iron toroids. Ferrite toroids and rods often have sharp edges, while powdered-iron toroids usually have rounded edges.

Table 4.18
Ferrite Toroids: Magnetic Properties and Dimensions

There are differing conventions for referring to the type of ferrite material: #, mix and type are all used. For example, all of the following designate the same ferrite material: #43, Mix 43, 43-Mix, Type 43, and 43-Type.

Fair-Rite Corporation (**www.fair-rite.com**) and Amidon (**www.amidoncorp.com**) ferrite toroids can be cross-referenced as follows:

For Amidon toroids, “FT-XXX-YY” indicates a ferrite toroid, with XXX as the OD in hundredths of an inch and YY the mix. For example, an FT-23-43 core has an OD of 0.23 inch and is made of type 43 material. Additional letters (usually “C”) are added to indicate special coatings or different thicknesses.

For Fair-Rite toroids, digits 1 and 2 of the part number indicate product type (59 indicates a part for inductive uses), digits 3 and 4 indicate the material type, digits 5 through 9 indicate core size, and the final digit indicates coating (1 for Paralene and 2 for thermo-set). For example, Fair-Rite part number 5943,000101 is equivalent to the Amidon FT-23-43 core.

Ferrite Toroids: A_L Chart (mH per 1,000 turns-squared)
Toroid diameter is specified as the outside diameter of the core. See Table 4.16 for a core sizing guide.

Core Size (in)	*63/67-Mix* $\mu = 40$	*61-Mix* $\mu = 125$	*43-Mix* $\mu = 850$	*77 (72)-Mix* $\mu = 2{,}000$	*J (75)-Mix* $\mu = 5{,}000$
0.23	7.9	24.8	188	396	980
0.37	19.7	55.3	420	884	2196
0.50	22.0	68.0	523	1100	2715
0.82	22.4	73.3	557	1170	NA
1.14	25.4	79.3	603	1270	3170
1.40	45	140	885	2400	5500
2.40	55	170	1075	2950	6850

31-Mix is an EMI suppression material and not recommended for inductive use.

Inductance and Turns Formula
The turns required for a given inductance or inductance for a given number of turns can be calculated from:

$$N = 100\sqrt{\frac{L}{A_L}} \qquad L = A_L\left(\frac{N^2}{10{,}000}\right) \quad \text{(Amidon cores)}$$

$$N = \sqrt{\frac{L}{A_L}} \qquad L = A_L \times N^2 \quad \text{(Non-Amidon cores)}$$

where N = number turns; L = desired inductance; A_L = inductance index. Amidon specifies A_L as mH per 1,000 turns-squared. See non-Amidon manufacturer’s core data sheet to determine appropriate units for L and A_L, usually nH and nH/turns-squared. Make sure you understand which units apply and use the A_L value appropriate for your core.

Ferrite Magnetic Properties

Property	*Unit*	*63/67-Mix*	*61-Mix*	*43-Mix*	*77 (72)-Mix*	*J (75)-Mix*	*31-Mix*
Initial perm.	(μ_i)	40	125	850	2,000	5,000	1500
Max. perm.		125	450	3,000	6000	8000	Not spec.
Saturation flux density @ 10 oe	gauss	1850	2350	2750	4600	3900	3400
Residual flux density	gauss	750	1,200	1,200	1150	1250	2500
Curie temp.	°C	450	350	130	200	140	>130
Vol. resistivity	ohm/cm	1×10^8	1×10^8	1×10^5	1×10^2	5×10^2	3×10^3
Resonant circuit frequency	MHz	15-25	0.2-10	0.01-1	0.001-1	0.001-1	*
Specific gravity		4.7	4.7	4.5	4.8	4.8	4.7
Loss factor	$\frac{1}{\mu_i Q}$	110×10^{-6} @25 MHz	32×10^{-6} @2.5 MHz	120×10^{-6} @1 MHz	4.5×10^{-6} @0.1 MHz	15×10^{-6} @0.1 MHz	20×10^{-6} @0.1 MHz
Coercive force	Oe	2.40	1.60	0.30	0.22	0.16	0.35
Temp. Coef. of initial perm.	%/°C (20°-70°)	0.10	0.15	1.0	0.60	0.90	1.6

*31-Mix is an EMI suppression material and not recommended for inductive uses.

Ferrite Toroids—Physical Properties
All physical dimensions in inches.

OD (in)	*ID (in)*	*Height (in)*	A_e	ℓ_e	V_e
0.230	0.120	0.060	0.00330	0.529	0.00174
0.375	0.187	0.125	0.01175	0.846	0.00994
0.500	0.281	0.188	0.02060	1.190	0.02450
0.825	0.520	0.250	0.03810	2.070	0.07890
1.142	0.750	0.295	0.05810	2.920	0.16950
1.400	0.900	0.500	0.12245	3.504	0.42700
2.400	1.400	0.500	0.24490	5.709	1.39080

Different height cores may be available for each core size.
A_e — Effective magnetic cross-sectional area $(in)^2$
ℓ_e — Effective magnetic path length (inches)
V_e — Effective magnetic volume $(in)^3$
To convert from $(in)^2$ to $(cm)^2$, divide by 0.155
To convert from $(in)^3$ to $(cm)^3$, divide by 0.0610
Courtesy of Amidon Assoc. and Fair-Rite Corp.

Because of their higher permeabilities, the A_L values for ferrite cores are higher than for powdered-iron cores. Amidon Corp. is the most common supplier of cores for amateurs and specifies A_L in mH per 1,000 turns-squared. Other manufacturers specify A_L in nH per turns-squared.

To calculate the inductance of a ferrite toroidal inductor when the number of turns and the core material are known:

$$L = \frac{A_L \times N^2}{1000000} \text{ (Amidon)}$$

where

L = the inductance in mH,

A_L = the inductance index in mH per 1,000 turns-squared, and

N = the number of turns.

$$L = A_L \times N^2 \text{ (Other manufacturers)}$$

where

L = the inductance in nH,

A_L = the inductance index in nH per turns-squared, and

N = the number of turns.

Calculations are performed similarly to the examples given for powdered-iron cores in the previous section. The builder must then ensure that the core is capable of holding the calculated number of turns of wire of the required wire size.

For inductors carrying both dc and ac currents, the upper saturation limit for most ferrites is a flux density of 2,000 gauss, with power calculations identical to those used for powdered-iron cores.

4.5 Transformers

When the ac source current flows through every turn of an inductor, the generation of a counter-voltage and the storage of energy during each half cycle is said to be by virtue of self-inductance. If another inductor — not connected to the source of the original current — is positioned so the magnetic field of the first inductor intercepts the turns of the second inductor, coupling the two inductors and creating mutual inductance as described earlier, a voltage will be induced and current will flow in the second inductor. A load such as a resistor may be connected across the second inductor to consume the energy transferred magnetically from the first inductor.

Figure 4.23 illustrates a pair of coupled inductors, showing an ac energy source connected to one, called the *primary inductor*, and a load connected to the other, called the *secondary inductor*. If the inductors are wound tightly on a magnetic core so that nearly all magnetic flux from the first inductor intersects with the turns of the second inductor, the pair is said to be *tightly coupled*. Inductors not sharing a common core and separated by a distance would be *loosely coupled*.

The signal source for the primary inductor may be household ac power lines, audio or other waveforms at low frequencies, or RF currents. The load may be a device needing power, a speaker converting electrical energy into sonic energy, an antenna using RF energy for communications or a particular circuit set up to process a signal from a preceding circuit. The uses of magnetically coupled energy in electronics are innumerable.

Mutual inductance (M) between inductors is measured in henrys. Two inductors have a mutual inductance of 1 H under the following conditions: as the primary inductor current changes at a rate of 1 A/s, the voltage across the secondary inductor is 1 V. The level of mutual inductance varies with many factors: the size and shape of the inductors, their relative positions and distance from each other, and the permeability of the inductor core material and of the space between them.

If the self-inductance values of two inductors are known (self-inductance is used in this section to distinguish it from the mutual inductance), it is possible to derive the mutual inductance by way of a simple experiment schematically represented in **Figure 4.24**. Without altering the physical setting or position of two inductors, measure the total coupled inductance, L_C, of the series-connected inductors with their windings complementing each other and again with their windings opposing each other. Since, for the two inductors, $L_C = L1 + L2 + 2M$, in the complementary case, and $L_O = L1 + L2 - 2M$ for the opposing case,

$$M = \frac{L_C - L_O}{4}$$

The ratio of magnetic flux set up by the secondary inductor to the flux set up by the primary inductor is a measure of the extent to which two inductors are coupled, compared to the maximum possible coupling between them. This ratio is the *coefficient of coupling*

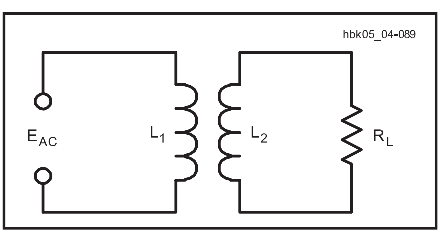


Figure 4.23 — A basic transformer: two inductors — one connected to an ac energy source, the other to a load — with coupled magnetic fields.

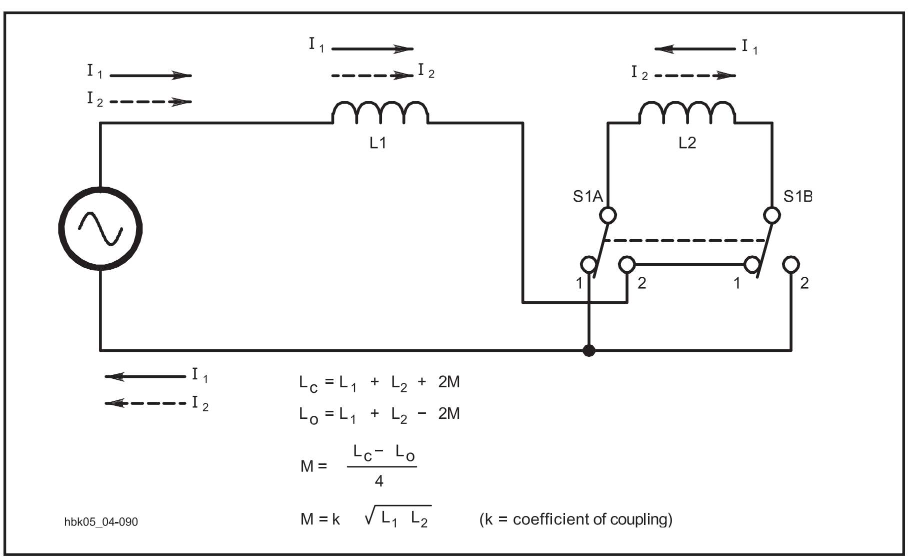


Figure 4.24 — An experimental setup for determining mutual inductance. Measure the inductance with the switch in each position and use the formula in the text to determine the mutual inductance.

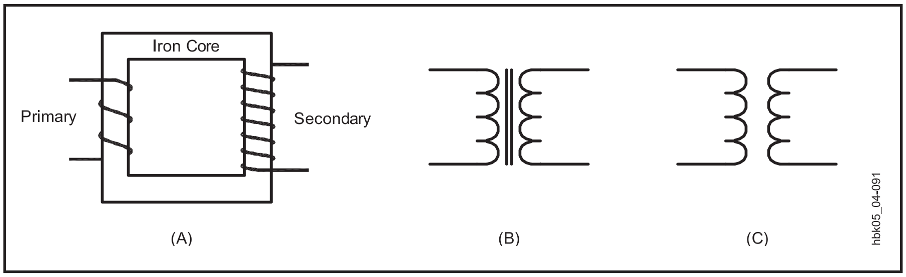


Figure 4.25 — A transformer. A is a pictorial diagram. Power is transferred from the primary coil to the secondary by means of the magnetic field. B is a schematic diagram of an iron-core transformer, and C is an air-core transformer.

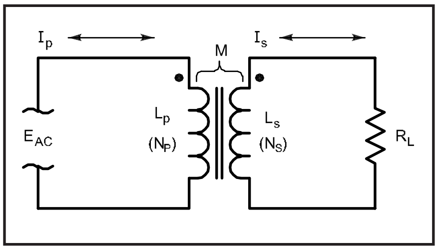


Figure 4.26 — The conditions for transformer action: two coils that exhibit mutual inductance, an ac power source, and a load. The magnetic field set up by the energy in the primary circuit transfers energy to the secondary for use by the load, resulting in a secondary voltage and current. The phasing dots at the top of each winding indicate the points at which ac voltages are in-phase.

(k) and is always less than 1. If k were to equal 1, the two inductors would have the maximum possible mutual coupling. Thus:

$$M = k\sqrt{L1\,L2}$$

where
- M = mutual inductance in henrys,
- L1 and L2 = individual coupled inductors, each in henrys, and
- k = the coefficient of coupling.

Using the experiment above, it is possible to solve for k with reasonable accuracy.

Any two inductors having mutual inductance comprise a *transformer* having a *primary winding* or inductor and a *secondary winding* or inductor. The word "winding" is generally dropped so that transformers are said to have "primaries" and "secondaries." **Figure 4.25** provides a pictorial representation of a typical iron-core transformer, along with the schematic symbols for both iron-core and air-core transformers. Conventionally, the term *transformer* is most commonly applied to coupled inductors having a magnetic core material, while coupled air-wound inductors are not called by that name. They are still transformers, however.

We normally think of transformers as ac devices, since mutual inductance only occurs when magnetic fields are changing. A transformer connected to a dc source will exhibit mutual inductance only at the instants of closing and opening the primary circuit, or on the rising and falling edges of dc pulses, because only then does the primary winding have a changing field. There are three principle uses of transformers: to physically isolate the primary circuit from the secondary circuit, to transform voltages and currents from one level to another, and to transform circuit impedances from one level to another. These functions are not mutually exclusive and have many variations.

4.5.1 Basic Transformer Principles

The primary and secondary windings of a transformer may be wound on a core of magnetic material. The permeability of the magnetic material increases the inductance of the windings so a relatively small number of turns may be used to induce a given voltage value with a small current. A closed core having a continuous magnetic path, such as that shown in Figure 4.25, also tends to ensure that practically all of the field set up by the current in the primary winding will intercept or "cut" the turns of the secondary winding.

For power transformers and impedance-matching transformers used at audio frequencies, cores made of soft iron strips or sheets called *laminations* are most common and generally very efficient. At higher frequencies, ferrite or powdered-iron cores are more frequently used. This section deals with basic transformer operation at audio and power frequencies. RF transformer operation is discussed in the chapter on **RF Techniques**.

The following principles presume a coefficient of coupling (k) of 1, that is, a perfect transformer. The value k = 1 indicates that all the turns of both windings link with all the magnetic flux lines, so that the voltage induced per turn is the same with both windings. This condition makes the induced voltage independent of the inductance of the primary and secondary inductors. Iron-core transformers for low frequencies closely approach this ideal condition. **Figure 4.26** illustrates the conditions for transformer action.

The *phasing dots* at the top of each winding in Figure 4.26 are polarity markings that indicate phase relationships in transformer schematic diagrams. If the dots are placed at the same ends of the primary and secondary windings as in the figure, the polarity of the instantaneous voltage across the primary winding will be the same as that across the secondary winding. That means the phase shift between the primary and secondary winding will be zero and the directions of the secondary current (I_S) and primary current (I_P) will be the same. If the dots are at the opposite ends of the windings, the primary and secondary voltages will be out of phase and the primary and secondary currents will be in the opposite directions.

Winding polarity is important because it is determined by the direction in which the windings are wrapped on the transformer core. This also determines the direction of magnetic flux in the core and the polarity of the resulting induced voltage and current. This relationship is illustrated with diagrams at **circuitdigest.com/article/understanding-dot-convention-in-transformers**. Dot convention is also used in switchmode power conversion circuits (see the **Power Sources** chapter) to show how inductors and transformers store and release magnetic energy.

VOLTAGE RATIO

For a given varying magnetic field, the voltage induced in an inductor within the field is proportional to the number of turns in the inductor. When the two windings of a transformer are in the same field (which is the case when both are wound on the same closed core), it follows that the induced voltages will be proportional to the number of turns in each winding. In the primary, the induced voltage practically equals, and opposes, the applied voltage, as described earlier. Hence:

$$E_S = E_P\left(\frac{N_S}{N_P}\right)$$

where
- E_S = secondary voltage,
- E_P = primary applied voltage,
- N_S = number of turns on secondary, and
- N_P = number of turns on primary.

Example: A transformer has a primary with 400 turns and a secondary with 2,800 turns, and a voltage of 120 V is applied to the primary. What voltage appears across the secondary winding?

$$E_S = 120\text{ V}\left(\frac{2800}{400}\right) = 120\text{ V} \times 7 = 840\text{ V}$$

(Notice that the number of turns is taken as a known value rather than a measured quantity, so they do not limit the significant figures in the calculation.) Also, if 840 V is applied to the 2,800-turn winding (which then becomes the primary), the output voltage from the 400-turn winding will be 120 V.

Either winding of a transformer can be used as the primary, provided the winding has enough turns (enough inductance) to induce a voltage equal to the applied voltage without requiring an excessive current. The windings must also have insulation with a voltage rating sufficient for the voltages applied or created. Transformers are called *step-up* or *step-down* transformers depending on whether the secondary voltage is higher or lower than the primary voltage, respectively.

CURRENT OR AMPERE-TURNS RATIO

The current in the primary when no current is taken from the secondary is called the *magnetizing current* of the transformer. An ideal transformer, with no internal losses, would consume no power, since the current through the primary inductor would be 90° out of phase with the voltage. In any properly designed transformer, the power consumed by the transformer when the secondary is open (not delivering power) is only the amount necessary to overcome the losses in the iron core and in the resistance of the wire with which the primary is wound.

When power is transferred from the secondary winding to a load, the secondary current creates a magnetic field that opposes the field established by the primary current. For the induced voltage in the primary to equal the applied voltage, the original magnetizing field must be maintained. Therefore, enough additional current must flow in the primary to create a field exactly equal and opposite to the field set up by the secondary current, leaving the original magnetizing field.

In practical transformer calculations it may be assumed that the entire primary current is caused by the secondary load. This is justifiable because the magnetizing current should be very small in comparison with the primary load current at rated power output.

If the magnetic fields set up by the primary and secondary currents are to be equal, the number of ampere-turns must be equal in each winding. (See the previous discussion of magnetic fields and magnetic flux density.) Thus, primary current multiplied by the primary turns must equal the secondary current multiplied by the secondary turns.

$$I_P N_P = I_S N_S$$

and

$$I_P = I_S \left(\frac{N_S}{N_P} \right)$$

where

I_P = primary current,
I_S = secondary current,
N_P = number of turns in the primary winding, and
N_S = number of turns in the secondary winding.

Example: Suppose the secondary of the transformer in the previous example is delivering a current of 0.20 A to a load. What will be the primary current?

$$I_P = 0.20 \text{ A} \left(\frac{2800}{400} \right) = 0.20 \text{ A} \times 7 = 1.4 \text{ A}$$

Although the secondary voltage is higher than the primary voltage, the secondary current is lower than the primary current, and by the same ratio. The secondary current in an ideal transformer is 180° out of phase with the primary current, since the field in the secondary just offsets the field in the primary. The phase relationship between the currents in the windings holds true no matter what the phase difference between the current and the voltage of the secondary. In fact, the phase difference, if any, between voltage and current in the secondary winding will be *reflected* back to the primary as an identical phase difference.

POWER RATIO

A transformer cannot create power; it can only transfer it and change the voltage and current ratios. Hence, the power taken from the secondary cannot exceed that taken by the primary from the applied voltage source. There is always some power loss in the resistance of the windings and in the iron core, so in all practical cases the power taken from the source will exceed that taken from the secondary.

$$P_O = \eta \, P_I$$

where

P_O = power output from secondary,
P_I = power input to primary, and
η = efficiency.

The efficiency, η, is always less than 1 and is commonly expressed as a percentage: if η is 0.65, for instance, the efficiency is 65%.

Example: A transformer has an efficiency of 85.0% at its full-load output of 150 W. What is the power input to the primary at full secondary load?

$$P_I = \frac{P_O}{\eta} = \frac{150 \text{ W}}{0.850} = 176 \text{ W}$$

A transformer is usually designed to have the highest efficiency at the power output for which it is rated. The efficiency decreases with either lower or higher outputs. On the other hand, the losses in the transformer are relatively small at low output but increase as more power is taken. The amount of power that the transformer can handle is determined by its own losses, because these losses heat the wire and core. There is a limit to the temperature rise that can be tolerated, because too high a temperature can either melt the wire or cause the insulation to break down. A transformer can be operated at reduced output, even though the efficiency is low, because the actual loss will be low under such conditions. The full-load efficiency of small power transformers such as are used in radio receivers and transmitters usually lies between about 60 and 90%, depending on the size and design.

IMPEDANCE RATIO

In an ideal transformer — one without losses or leakage inductance (see Transformer Losses) — the primary power, $P_P = E_P I_P$, and secondary power, $P_S = E_S I_S$, are equal. The relationships between primary and secondary voltage and current are also known. Since impedance is the ratio of voltage to current, Z = E/I, the impedances represented in each winding are related as follows:

$$Z_P = Z_S \left(\frac{N_P}{N_S} \right)^2$$

where

Z_P = impedance at the primary terminals from the power source,
Z_S = impedance of load connected to secondary, and
N_P/N_S = turns ratio, primary to secondary.

The transformer is converting input power at one ratio of voltage to current (i.e., impedance) to output power at a different ratio of voltage to current (i.e., a different impedance).

A load of any given impedance connected to the transformer secondary will thus be transformed to a different value at the primary terminals. The impedance transformation is proportional to the square of the primary-to-secondary turns ratio. (Take care to use the primary-to-secondary turns ratio, since the secondary-to-primary ratio is more commonly used to determine the voltage transformation ratio.)

The term *looking into* is often used to mean the conditions observed from an external perspective at the terminals specified. For example, "impedance looking into" the transformer primary means the impedance measured externally to the transformer at the terminals of the primary winding.

Example: A transformer has a primary-to-secondary turns ratio of 0.6 (the primary has six-tenths as many turns as the secondary) and a load of 3,000 Ω is connected to the secondary. What is the impedance looking into the primary of the transformer?

$Z_P = 3,000\ \Omega \times (0.6)^2 = 3,000\ \Omega \times 0.36$

$Z_P = 1,080$ W

By choosing the proper turns ratio, the impedance of a fixed load can be transformed to any desired value, within practical limits. If transformer losses can be neglected, the transformed (reflected) impedance has the same phase angle as the actual load impedance. Thus, if the load is a pure resistance, the load presented by the primary to the power source will also be a pure resistance. If the load impedance is complex, that is, if the load current and voltage are out of phase with each other, then the primary voltage and current will have the same phase angle.

Many devices or circuits require a specific value of load resistance (or impedance) for optimum operation. The impedance of the actual load that is to dissipate the power may be quite different from the impedance of the source device or circuit, so an *impedance-matching transformer* is used to change the actual load into an impedance of the desired value. The turns ratio required is:

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}}$$

where

N_P / N_S = required turns ratio, primary to secondary,

Z_P = primary impedance required, and

Z_S = impedance of load connected to secondary.

Example: A transistor audio amplifier requires a load of 150 Ω for optimum performance, and is to be connected to a loudspeaker having an impedance of 4.0 Ω. What primary-to-secondary turns ratio is required in the coupling transformer?

$$\frac{N_P}{N_S} = \sqrt{\frac{Z_P}{Z_S}} = \sqrt{\frac{150\ \Omega}{4.0\ \Omega}} = \sqrt{38} = 6.2$$

The primary therefore must have 6.2 times as many turns as the secondary.

These relationships may be used in practical circuits even though they are based on an ideal transformer. Aside from the normal design requirements of reasonably low internal losses and low leakage reactance, the only other requirement is that the primary has enough inductance to operate with low magnetizing current at the voltage applied to the primary.

The primary terminal impedance of an iron-core transformer is determined wholly by the load connected to the secondary and by the turns ratio. If the characteristics of the transformer have an appreciable effect on the impedance presented to the power source, the transformer is either poorly designed or is not suited to the voltage and frequency at which it is being used. Most transformers will operate quite well at voltages from slightly above to well below the design figure.

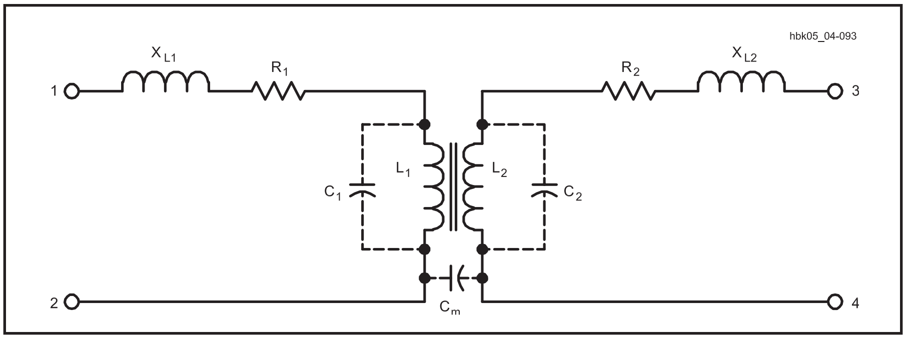


Figure 4.27 — A transformer as a network of resistances, inductances and capacitances. Only L1 and L2 contribute to the transfer of energy.

TRANSFORMER LOSSES

In practice, none of the formulas given so far provides truly exact results, although they afford reasonable approximations. Transformers in reality are not simply two coupled inductors, but a network of resistances and reactances, most of which appear in **Figure 4.27**. Since only the terminals numbered 1 through 4 are accessible to the user, transformer ratings and specifications take into account the additional losses created by these complexities.

In a practical transformer not all of the magnetic flux is common to both windings, although in well-designed transformers the amount of flux that cuts one winding and not the other is only a small percentage of the total flux. This *leakage flux* causes a voltage by self-induction in the winding creating the flux. The effect is the same as if a small *leakage inductance* existed independently of the main windings. Leakage inductance acts in exactly the same way as inductance inserted in series with the winding. It has, therefore, a certain reactance, depending on the amount of leakage inductance and the frequency. This reactance is called *leakage reactance* and is shown as X_{L1} and X_{L2} in Figure 4.27.

Current flowing through the leakage reactance causes a voltage drop. This voltage drop increases with increasing current (or frequency); hence, it increases as more power is taken from the secondary. The resistances of the transformer windings, R1 and R2, also cause voltage drops when there is current. Although these voltage drops are not in phase with those caused by leakage reactance, together they result in a lower secondary voltage under load than is indicated by the transformer turns ratio. Thus, in a practical transformer, the greater the

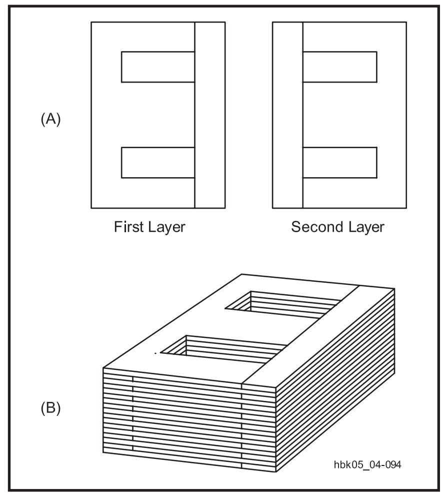


Figure 4.28 — A typical transformer iron core. The E and I pieces alternate direction in successive layers to improve the magnetic path while attenuating eddy currents in the core.

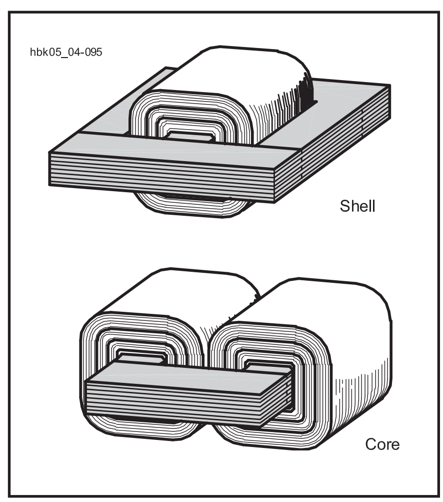


Figure 4.29 — Two common transformer constructions: shell and core.

secondary current, the smaller the secondary terminal voltage becomes.

At ac line frequencies (50 or 60 Hz), the voltage at the secondary, with a reasonably well-designed iron-core transformer, should not drop more than about 10% from open-circuit conditions to full load. The voltage drop may be considerably more than this in a transformer operating at audio frequencies, because the leakage reactance increases with frequency.

In addition to wire resistances and leakage reactances, certain unwanted or "stray" capacitances occur in transformers. The wire forming the separate turns of the windings acts as the plates of a small capacitor, creating a capacitance between turns and between the windings. This *distributed capacitance* appears in Figure 4.27 as C1, C2, and C_M. Moreover, transformer windings can exhibit capacitance relative to nearby metal, for example, the chassis, the shield and even the core. When current flows through a winding, each turn has a slightly different voltage than its adjacent turns. This voltage causes a small current to flow in these *interwinding* and *winding-to-winding capacitances*.

Although stray capacitances are of little concern with power and audio transformers, they become important as the frequency increases. In transformers for RF use, the stray capacitance can resonate with either the leakage reactance or, at lower frequencies, with the winding reactances, L1 or L2, especially under very light or zero loads. In the frequency region around resonance, transformers no longer exhibit the properties formulated above or the impedance properties to be described below.

Iron-core transformers also experience losses within the core itself. *Hysteresis losses* include the energy required to overcome the retentivity of the core's magnetic material. Circulating currents through the core's resistance are *eddy currents*, which form part of the total core losses. These losses, which add to the required magnetizing current, are equivalent to adding a resistance in parallel with L1 in Figure 4.27.

CORE CONSTRUCTION

Audio and power transformers usually employ silicon steel as the core material. With permeabilities of 5,000 or greater, these cores saturate at flux densities approaching 10^5 (Mx) per square inch of cross section. The cores consist of thin insulated laminations to break up potential eddy current paths.

Each core layer consists of an "E" and an "I" piece butted together, as represented in **Figure 4.28**. The butt point leaves a small gap. Each layer is reversed from the adjacent layers so that each gap is next to a continuous magnetic path, minimizing the effect of the gaps. This is different from an air-gapped inductor in which the air gap is maintained for all layers of laminations.

Two core shapes are in common use, as shown in **Figure 4.29**. In the shell type, both windings are placed on the inner leg, while in the core type the primary and secondary windings may be placed on separate legs, if desired. This is sometimes done when it is necessary to minimize capacitance between the primary and secondary, or when one of the windings must operate at very high voltage.

The number of turns required in the primary for a given applied voltage is determined by the size, shape, and type of core material used, as well as the frequency. The number of turns required is inversely proportional to the cross-sectional area of the core. As a rough indication, windings of small power transformers frequently have about six to eight turns per volt on a core of 1-square-inch cross-section and have a magnetic path 10 or 12 inches in length. A longer path or smaller cross-section requires more turns per volt, and vice versa.

In most transformers the windings are wound in layers, with a thin sheet of treated-paper insulation between each layer. Thicker insulation is used between adjacent windings and between the first winding and the core.

SHIELDING

Because magnetic lines of force are continuous loops, shielding requires a complete path for the lines of force of the leakage flux. The high permeability of iron cores tends to concentrate the field, but additional shielding is often needed. As depicted in **Figure 4.30**, enclosing the transformer in a good magnetic material can restrict virtually all of the magnetic field in the outer case. The nonmagnetic material between the case and the core creates a region of high reluctance, attenuating the field before it reaches the case.

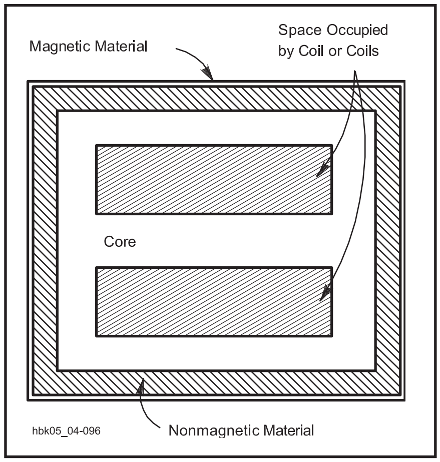


Figure 4.30 — A shielded transformer cross-section: the core plus an outer shield of magnetic material contain nearly all of the magnetic field.

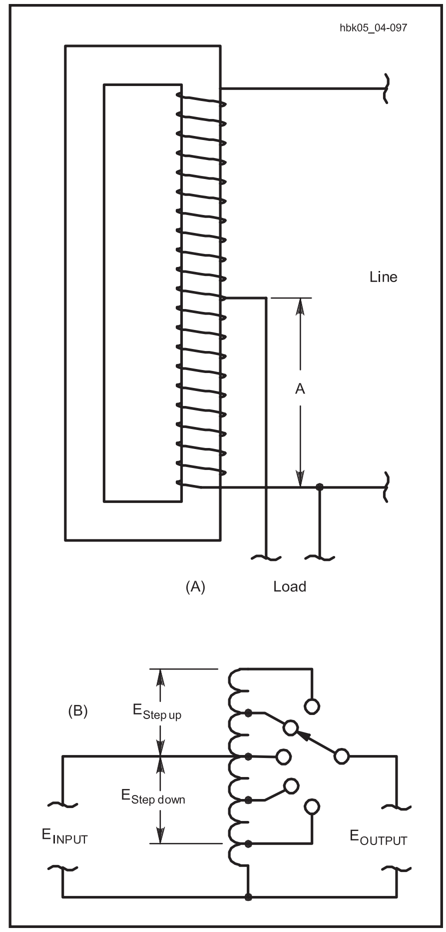


Figure 4.31 — The autotransformer is based on the transformer but uses only one winding. The pictorial diagram at A shows the typical construction of an autotransformer. The schematic diagram at B demonstrates the use of an autotransformer to step up or step down ac voltage, usually to compensate for excessive or deficient line voltage.

4.5.2 Transformer Identification

Many transformers, including power transformers, IF transformers, and audio transformers, are made to be installed on PC boards, and have terminals designed for that purpose. Some transformers are manufactured with wire leads that are color-coded to identify each connection. When colored wire leads are present, the color codes in **Tables 4.19**, **4.20**, and **4.22** usually apply. In addition, many miniature IF transformers are tuned with slugs, color-coded to signify their application. **Table 4.21** lists application versus slug color.

4.5.3 Autotransformers

The transformer principle can be used with only one winding instead of two, as shown in **Figure 4.31A**. The principles that relate volt-

Table 4.19
Power-Transformer Wiring Color Codes

Non-tapped primary leads:	Black
Tapped primary leads:	Common: Black Tap: Black/yellow striped Finish: Black/red striped
High-voltage plate winding:	Red
Center tap:	Red/yellow striped
Rectifier filament winding:	Yellow
Center tap:	Yellow/blue striped
Filament winding 1:	Green
Center tap:	Green/yellow striped
Filament winding 2:	Brown
Center tap:	Brown/yellow striped
Filament winding 3:	Slate
Center tap:	Slate/yellow striped

Table 4.22
Audio Transformer Wiring Color Codes

Plate lead of primary	Blue
B+ lead (plain or center-tapped)	Red
Plate (start) lead on center-tapped primaries	Brown (or blue if polarity is not important)
Grid (finish) lead to secondary	Green
Grid return (plain or center tapped)	Black
Grid (start) lead on center tapped secondaries	Yellow (or green if polarity not important)

Note: These markings also apply to line-to-grid and tube-to-line transformers.

Table 4.20
IF Transformer Wiring Color Codes

Plate lead:	Blue
B+ lead:	Red
Grid (or diode) lead:	Green
Grid (or diode) return:	Black

Note: If the secondary of the IF transformer is center-tapped, the second diode plate lead is green-and-black striped, and black is used for the center-tap lead.

Table 4.21
IF Transformer Slug Color Codes

Frequency	*Application*	*Slug color*
455 kHz	1st IF	Yellow
	2nd IF	White
	3rd IF	Black
	Osc tuning	Red
10.7 MHz	1st IF	Green
	2nd or 3rd IF	Orange, Brown or Black

age, current and impedance to the turns ratio also apply equally well. A one-winding transformer is called an *autotransformer*. The current in the common section (A) of the winding is the difference between the line (primary) and the load (secondary) currents, since these currents are out of phase. Hence, if the line and load currents are nearly equal, the common section of the winding may be wound with comparatively small wire. The line and load currents will be equal only when the primary (line) and secondary (load) voltages are not very different.

Autotransformers are used chiefly for boosting or reducing the power-line voltage by relatively small amounts. Figure 4.31B illustrates the principle schematically with a switched, stepped autotransformer. Continuously variable autotransformers are commercially available under a variety of trade names; Variac and Powerstat are typical examples.

Technically, tapped air-core inductors, such as the one in the network in Figure 3.48 at the close of the discussion of resonant circuits in the chapter on **Radio Fundamentals**, are also autotransformers. The voltage from the tap to the bottom of the winding is less than the voltage across the entire winding. Likewise, the impedance of the tapped part of the winding is less than the impedance of the entire winding. Because in this case, leakage reactances are great and the coefficient of coupling is quite low, the relationships that are true in a perfect transformer grow quite unreliable in predicting the exact values. For this reason, tapped inductors are rarely referred to as transformers. The stepped-down situation in Figure 3.48 is better approximated — at or close to resonance — by the formula

$$R_P = \frac{R_L\ X_{COM}^{\ 2}}{X_L}$$

where

R_P = tuned-circuit parallel-resonant impedance,

R_L = load resistance tapped across part of the winding,

X_{COM} = reactance of the portion of the winding common to both the resonant circuit and the load tap, and

X_L = reactance of the entire winding.

The result is approximate and applies only to circuits with a Q of 10 or greater.

4.6 Practical Semiconductors

There are several different kinds of semiconductors used to build analog electronic circuits, including bipolar diodes, thyristors, transistors, field-effect transistors, and integrated circuits. (Vacuum tubes are discussed in the chapter on **RF Power Amplifiers**, their primary application in amateur radio.) Manufacturer websites are often a rich source of information on applying semiconductors, both in general and about the specific devices they offer.

Most semiconductors are labeled with industry standard part numbers, such as 1N4148 or 2N3904, and possibly a date or batch code. You will also encounter numerous manufacturer-specific part numbers and the so-called "house numbers" (marked with codes used by an equipment manufacturer instead of the standard part numbers). In such cases, it is often possible to find the standard equivalent or a suitable replacement by using one of the semiconductor cross-reference directories available from various replacement parts distributors. If you look up the house number and find the recommended replacement part, you can often find other standard parts that are replaced by that same part.

4.6.1 Device Characteristics

CHARACTERISTIC CURVES

Analog devices are described most completely with their *characteristic curves*. The characteristic curve is a plot of the interrelationships between two or three variables. The vertical (y) axis parameter is the output, or result of the device being operated with an input parameter on the horizontal (x) axis. Often the output is the result of two input values. The first input parameter is represented along the X-axis and the second input parameter by several curves, each for a different value. Characteristic curves are used to graphically describe a device's operation in both its linear and nonlinear regions.

Figure 4.32A shows the characteristic curve for a semiconductor diode with the Y-axis showing the forward current, I_F, flowing through the diode and the X-axis showing forward voltage, V_F, across the diode. This curve shows the relationship between current and voltage in the diode when it is conducting current. Characteristic curves showing voltage and current in two-terminal devices such as diodes are often called *I-V curves*. Characteristic curves may include all four quadrants of operation in which both axes include positive and negative values. It is also common for different scales to be used in the different quadrants, so inspect the legend for the curves carefully.

The parameters plotted in a characteristic curve depend on how the device will be used so that the applicable design values can be obtained from the characteristic curve. The slope of the curve is often important because it relates changes in output to changes in input. To determine the slope of the curve, two closely spaced points along that portion of the curve are selected, each defined by its location along the x and y axes. If the two points are defined by (x_1,y_1) and (x_2,y_2), the slope, m, of the curve (which can be a gain, a resistance or a conductance, for example) is calculated as:

$$m = \frac{\Delta y}{\Delta x} = \frac{y_1 - y_2}{x_1 - x_2}$$

It is important to pick points that are close together or the slope will not reflect the actual behavior of the device. A device whose characteristic curve is not a straight line will not have a linear response to inputs because the slope changes with the value of the input parameter.

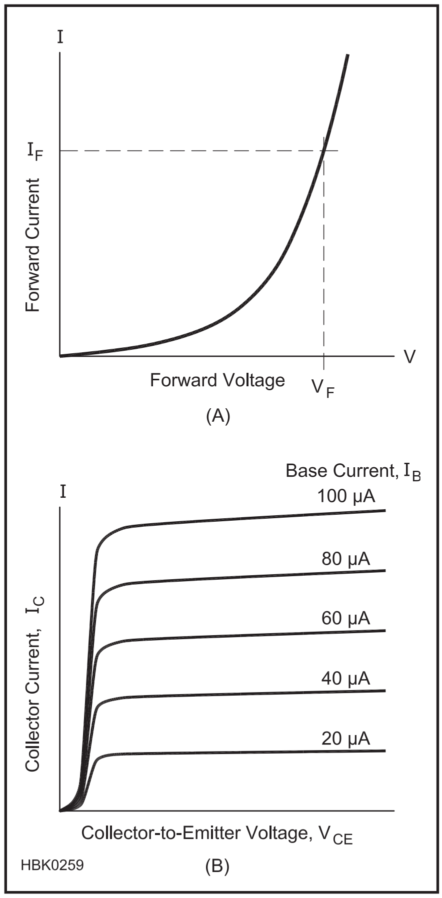


Figure 4.32 — Characteristic curves. A forward voltage vs forward current characteristic curve for a semiconductor diode is shown at (A). (B) shows a set of characteristic curves for a bipolar transistor in which the collector current vs collector-to-emitter voltage curve is plotted for five different values of base current.

For a device in which three parameters interact, such as a transistor, sets of characteristic curves can be drawn. Figure 4.32B shows a set of characteristic curves for a bipolar transistor where collector current, I_C, is shown on the y axis and collector-to-emitter voltage, V_{CE}, is shown on the x axis. Because the amount of collector current also depends on base current, I_B, the curve is repeated several times for different values of I_B. From this set of curves, an amplifier circuit using this transistor can be designed to have specific values of gain.

BIASING

The operation of an analog signal-processing device is greatly affected by which portion of the characteristic curve is used to do the processing. The device's *bias point* is its set of operating parameters when no input signal is applied. The bias point is also known as the *quiescent point* or *Q-point*. By changing the bias point, the circuit designer can affect the relationship between the input and output signal. The bias point can also be considered as a dc offset of the input signal. Devices that perform analog signal processing require appropriate input signal biasing.

As an example, consider the characteristic curve shown in **Figure 4.33**. (The exact types of device and circuit are unimportant.) The characteristic curve shows the relationship between an input voltage and an output current. Increasing input voltage results in an increase in output current so that an input signal is reproduced at the output. The characteristic curve is linear in the middle but is quite nonlinear in its upper and lower regions.

In the circuit described by the figure, bias points are established by adding one of the three voltages, V_1, V_2 or V_3 to the input signal. Bias voltage V_1 results in an output current of I_1 when no input signal is present. This is shown as Bias Point 1 on the characteristic curve. When an input signal is applied, the input voltage varies around V_1 and the output current varies around I_1 as shown. If the dc value of the output current is subtracted, a reproduction of the input signal is the result.

If Bias Point 2 is chosen, we can see that the input voltage is reproduced as a changing output current with the same shape. In this case, the device is operating linearly. If either Bias Point 1 or Bias Point 3 is chosen, however, the shape of the output signal is distorted because the characteristic curve of the device is nonlinear in this region. Either the increasing portion of the input signal results in more variation than the decreasing portion (Bias

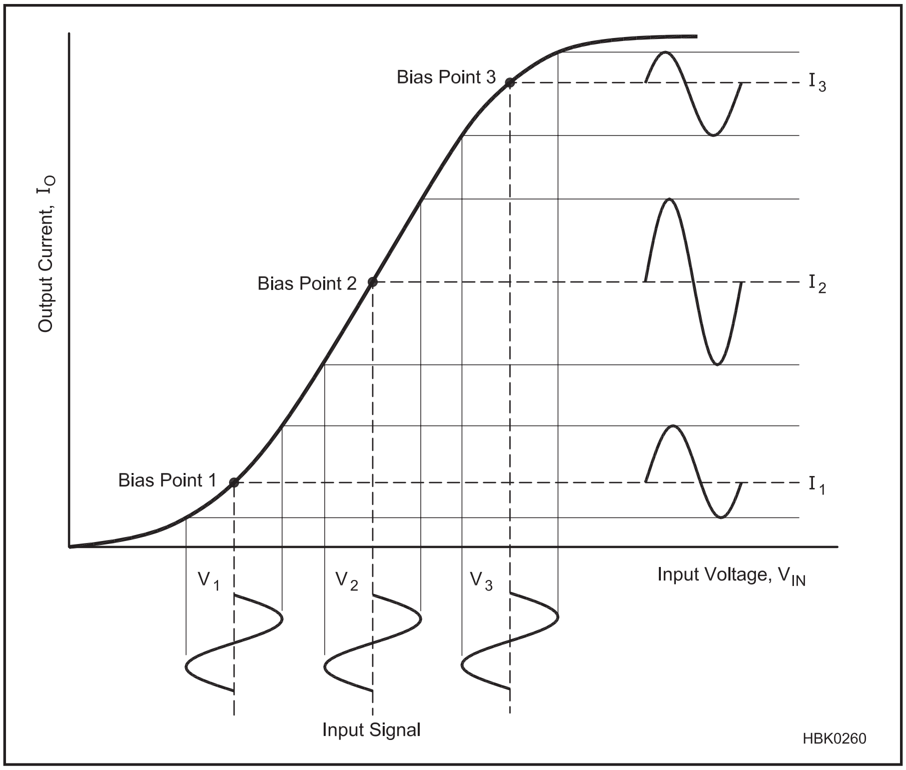


Figure 4.33 — Effect of biasing. An input signal may be reproduced linearly or nonlinearly depending on the choice of bias points.

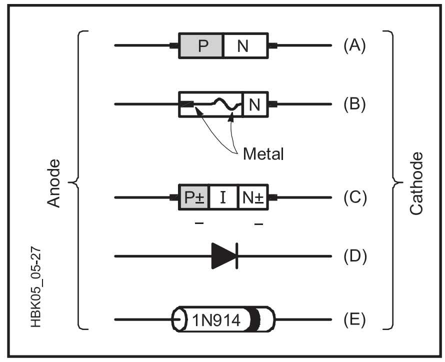


Figure 4.34 — Practical semiconductor diodes. All devices are aligned with anode on the left and cathode on the right. (A) Standard PN junction diode. (B) Point-contact or "cat's whisker" diode. (C) PIN diode formed with heavily doped P-type (P^+), undoped (intrinsic) and heavily doped N-type (N^+) semiconductor material. (D) Diode schematic symbol. (E) Diode package with marking stripe on the cathode end.

Point 1) or vice versa (Bias Point 3). Proper biasing is crucial to ensure that a device operates linearly.

MANUFACTURER'S DATA SHEETS

Manufacturer's data sheets list device characteristics, along with the specifics of the part type (polarity, semiconductor type), identity of the pins and leads (*pinouts*), and the typical use (such as small signal, RF, switching or power amplifier). The pin identification is important because, although common package pinouts are normally used, there are exceptions. Manufacturers may differ slightly in the values reported, but certain basic parameters are listed. Different batches of the same devices are rarely identical, so manufacturers specify the guaranteed limits for the parameters of their device. There are usually three values listed in the data sheet for each parameter: guaranteed minimum value, the guaranteed maximum value, and/or the typical value.

Another section of the data sheet lists ABSOLUTE MAXIMUM RATINGS, beyond which device damage may result. For example, the parameters listed in the ABSOLUTE MAXIMUM RATINGS section for a solid-state device are typically voltages, continuous currents, total device power dissipation (P_D) and operating- and storage-temperature ranges.

Rather than plotting the characteristic curves for each device, the manufacturer often selects key operating parameters that describe the device operation for the configurations and parameter ranges that are most commonly used. For example, a bipolar transistor data sheet might include an OPERATING PARAMETERS section. Parameters are listed in an OFF CHARACTERISTICS subsection and an ON CHARACTERISTICS subsection that describe the conduction properties of the device for dc voltages. The SMALL-SIGNAL CHARACTERISTICS section might contain a minimum Gain-Bandwidth Product (f_T or GBW), maximum output capacitance, maximum input capacitance, and the range of the transfer parameters applicable to a given device. Finally, the SWITCHING CHARACTERISTICS section might list absolute maximum ratings for Delay Time (t_d), Rise Time (t_r), Storage Time (t_s), and Fall Time (t_f). Other types of devices list characteristics important to operation of that specific device.

When selecting equivalent parts for replacement of specified devices, the data sheet provides the necessary information to tell if a given part will perform the functions of another. Lists of *cross-references* and *substitution guides* generally only specify devices that have nearly identical parameters. There are usually a large number of additional devices that can be chosen as replacements. Knowledge of the circuit requirements adds even more to the list of possible replacements. The device parameters should be compared individually to make sure that the replacement part meets or exceeds the parameter values of the original part required by the circuit. Be aware that in some applications a far superior part may fail as a replacement, however. A transistor with too much gain could easily oscillate if there were insufficient negative feedback to ensure stability.

4.6.2 Diodes

Although many types of semiconductor diodes are available, they share many common characteristics. The different types of diodes have been developed to optimize particular characteristics for one type of application. You will find many examples of diode applications throughout this book.

The diode symbol is shown in **Figure 4.34**. Forward current flows in the direction from anode to cathode, in the direction of the arrow. Reverse current flows from cathode to anode. (Current is considered to be conventional current as described in the **Electrical Fundamentals** chapter.) The anode of a semiconductor junction diode is made of P-type material and the cathode is made of N-type material, as indicated in Figure 4.34.

Most diodes are marked with a part number and some means of identifying the anode or cathode. A thick band or stripe is commonly used to identify the cathode lead or terminal. Stud-mount diodes are usually labeled with a small diode symbol to indicate anode and cathode. Diodes in axial lead packages are sometimes identified with a color scheme as shown in **Figure 4.35**. (Diode packaging dimensions and marking are given in the **Construction Techniques** chapter's section on PCB layout.) Many surface mount diodes are packaged in the same SMT packages as resistors.

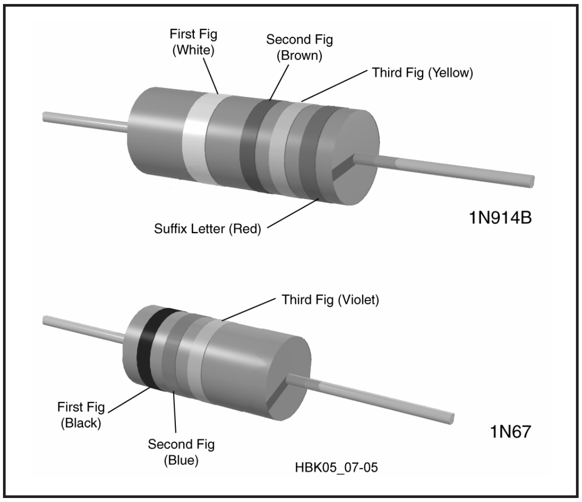


Figure 4.35 — Color-coding for semiconductor diodes. At A, the cathode is identified by the double-width first band. At B, the bands are grouped toward the cathode. Two-Figure designations are signified by a black first band. The color code is given in Table 4.2. The suffix-letter code is A-Brown, B-red, C-orange, D-yellow, E-green, F-blue. The 1N prefix is assumed.

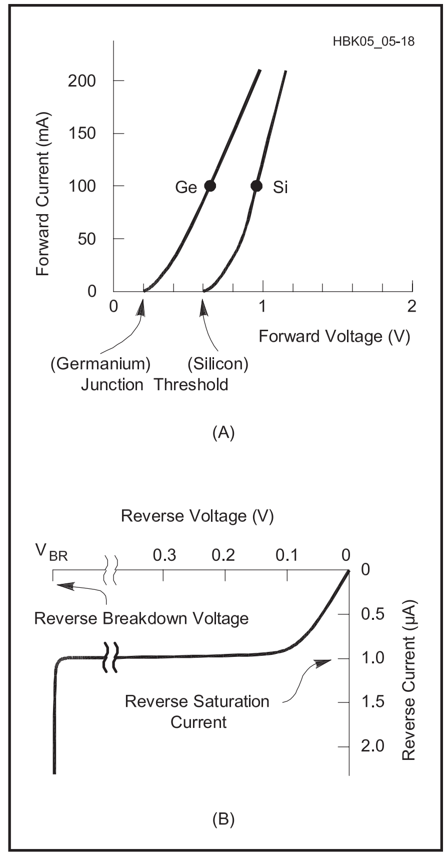


Figure 4.36 — Semiconductor diode (PN junction) characteristic curve. (A) Forward-biased (anode voltage higher than cathode) response for Germanium (Ge) and Silicon (Si) devices. Each curve breaks away from the X-axis at its junction threshold voltage. The slope of each curve is its forward resistance. (B) 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}.

Packages containing multiple diodes and rectifier bridge configurations are also commonly available. Full-wave bridge packages are labeled with tildes (~) for the ac inputs and + and – symbols for the rectifier outputs. High-power diodes are often packaged in TO-220 packages with two leads. The package may be labeled with a diode symbol but if not, you will have to obtain the manufacturer's data sheet to identify the anode and cathode leads.

DIODE RATINGS

Five major characteristics distinguish standard junction diodes from one another: current handling capacity, maximum voltage rating, response speed, reverse leakage current and junction forward voltage. Each of these characteristics can be manipulated during manufacture to produce special purpose diodes.

Current Capacity

The ideal diode would have zero resistance in the forward direction and infinite resistance in the reverse direction. This is not the case for actual devices, which behave as shown in the plot of a diode response in **Figure 4.36A**. Note that the scales of the two graphs are drastically different. The inverse of the slope of the line (the change in voltage between two points on a straight portion of the line divided by the corresponding change in current) on the upper right is the resistance of the diode in the forward direction, R_F.

The range of voltages is small and the range of currents is large since the forward resistance is very small (in this example, about 2 Ω). Nevertheless, this resistance causes heat dissipation according to $P = I_F^2 \times R_F$.

In addition, there is a forward voltage, V_F, whenever the forward current is flowing. This also results in heat dissipation as $P = I \times V_F$. In power applications where the average forward current is high, heating from forward resistance and the forward voltage drop can be significant. Since forward current determines the amount of heat dissipation, the diode's power rating is stated as a *maximum average current*. Exceeding the current rating in a diode will cause excessive heating that leads to PN junction failure as described earlier.

Although fixed voltages are often used for diodes in small-signal applications (0.6 V for silicon PN-junction diodes and 0.3 V for germanium, for example), the actual forward voltage at higher currents can be significantly higher and must be taken into account in high-current applications, such as power supplies.

Peak Inverse Voltage (PIV)

In Figure 4.36B, the lower left portion of the curve illustrates a much higher resistance that increases from tens of kilohms to thousands of megohms as the reverse voltage gets larger, and then decreases to near zero (a nearly vertical line) very suddenly. This sudden change occurs because the diode enters *reverse breakdown* or when the reverse voltage becomes high enough to push current across the junction. The voltage at which this occurs is the *reverse breakdown voltage*. Unless the current is so large that the diode fails from overheating, breakdown is not destructive and the diode will again behave normally when the bias is removed. The maximum reverse voltage that the diode can withstand under normal use is the *peak inverse voltage* (*PIV*) rating. A related effect is *avalanche breakdown* in which the voltage across a device is greater than its ability to control or block current flow.

Response Speed

The speed of a diode's response to a change in voltage polarity limits the frequency of ac current that the diode can rectify. The diode response in Figure 4.36 shows how that diode will act at dc. As the frequency increases, the diode may not be able to turn current on and off as fast as the changing polarity of the signal.

Diode response speed mainly depends on *charge storage* in the depletion region. When forward current is flowing, electrons and holes fill the region near the junction to recombine. When the applied voltage reverses, these excess charges move away from the junction so that no recombination can take place. As reverse bias empties the depletion region of excess charge, it begins to act like a small capacitor formed by the regions containing majority carriers on either side of the junction and the depletion region acting as the dielectric. This *junction capacitance* is inversely proportional to the width of the depletion region and directly proportional to the cross-sectional surface area of the junction.

The effect of junction capacitance is to allow current to flow for a short period after the applied voltage changes from positive to negative. To halt current flow requires that the junction capacitance be charged. Charging this capacitance takes some time; a few µs for regular rectifier diodes and a few hundred nanoseconds for *fast-recovery* diodes. This is the diode's *charge-storage time*. The amount of time required for current flow to cease is the diode's *recovery time*.

Reverse Leakage Current

Because the depletion region is very thin, reverse bias causes a small amount of reverse leakage or reverse saturation current to flow from cathode to anode. This is typically 1 µA or less until reverse breakdown voltage is reached. Silicon diodes have lower reverse leakage currents than diodes made from other materials with higher carrier mobility, such as germanium.

The reverse saturation current I_s is not constant but is affected by temperature, with higher temperatures increasing the mobility of the majority carriers so that more of them cross the depletion region for a given amount of reverse bias. For silicon diodes (and transistors) near room temperature, I_s increases by a factor of 2 every 4.8 °C. This means that for every 4.8 °C rise in temperature, either the current doubles (if the voltage across it is constant), or if the current is held constant by other resistances in the circuit, the diode voltage will *decrease* by $V_T \times \ln 2 = 18$ mV. For germanium, the current doubles every 8 °C and for gallium-arsenide (GaAs), 3.7 °C. This dependence is highly reproducible and may actually be exploited to produce temperature-measuring circuits.

While the change resulting from a rise of several degrees may be tolerable in a circuit design, that from 20 or 30 degrees may not. Therefore, it's a good idea with diodes, just as with other components, to specify power ratings conservatively (2 to 4 times margin) to prevent self-heating.

While component derating does reduce self-heating effects, circuits must be designed for the expected operating environment. For example, mobile radios may face temperatures from –20 to +140 °F (–29 to 60 °C).

Forward Voltage

The amount of voltage required to cause majority carriers to enter the depletion region and recombine, creating full current flow, is called a diode's *forward voltage*, V_F. It depends on the type of material used to create the junction and the amount of current. For silicon junction diodes at normal currents, $V_F = 0.7$ V, and for germanium diodes, $V_F = 0.3$ V. As stated earlier, V_F also affects power dissipation in the diode. Schottky diodes are used when low forward voltages are required, particularly at high-currents, and exhibit forward voltages of 0.1 – 0.2 V at low currents.

POINT-CONTACT DIODES

One way to decrease charge storage time in the depletion region is to form a metal-semiconductor junction for which the depletion is very thin. This can be accomplished with a *point-contact diode*, where a thin piece of aluminum wire, often called a *whisker*, is placed in contact with one face of a piece of lightly doped N-type material. In fact, the original diodes used for detecting radio signals ("cat's whisker diodes") were made with a steel wire in contact with a crystal of impure lead (galena). Point-contact diodes have high response speed but poor PIV and current-handling ratings. The 1N34 germanium point-contact diode is the best-known example of a point-contact diode still in common use.

SCHOTTKY DIODES

An improvement to point-contact diodes, the *hot-carrier diode* is similar to a point-contact diode, but with more ideal characteristics attained by using more efficient metals, such as platinum and gold, that act to lower forward resistance and increase PIV. This type of contact is known as a *Schottky barrier*, and diodes made this way are called *Schottky diodes*. The junctions of Schottky diodes, being smaller, store less charge and as a result have shorter switching times and junction capacitances than standard PN-junction diodes. Their forward voltage is also lower, typically less than 0.4 V at low currents (see the device data sheet). In most other respects they behave similarly to PN diodes.

PIN DIODES

The PIN diode, shown in Figure 4.34C, is a *slow response* diode that is capable of passing RF and microwave signals when it is forward biased. This device is constructed with a layer of *intrinsic* (undoped) semiconductor placed between very highly doped P-type and N-type material (called P^+-type and N^+-type material to indicate the extra amount of doping), creating a *PIN junction*. These devices provide very effective switches for RF signals and are often used in transmit-receive switches in transceivers and amplifiers. The majority carriers in PIN diodes have longer than normal lifetimes before recombination, resulting in a slow switching process that causes them to act more like resistors than diodes at high radio frequencies. The amount of resistance can be controlled by the amount of forward bias applied to the PIN diode, which allows them to act as current-controlled attenuators. (For additional discussion of PIN diodes and projects in which they are used, see the chapters on **Transmitting**, **RF Power Amplifiers**, and **Test Equipment and Measurements**.)

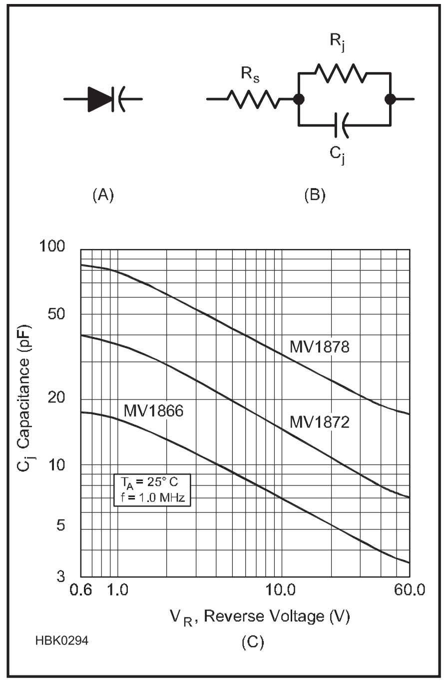


Figure 4.37 — Varactor diode. (A) Schematic symbol. (B) Equivalent circuit of the reverse biased varactor diode. R_S is the junction resistance, R_J is the leakage resistance and C_J is the junction capacitance, which is a function of the magnitude of the reverse bias voltage. (C) Plot of junction capacitance, C_J, as a function of reverse voltage, V_R, for three different varactor devices. Both axes are plotted on a logarithmic scale.

VARACTOR DIODES

Junction capacitance can be used as a circuit element by controlling the reverse bias voltage across the junction, creating a small variable capacitor. A *varactor* (also known by the trade name Varicap) is a diode with a junction specially formulated to have a relatively large range of capacitance values for a modest range of reverse bias voltages (**Figure 4.37**). They are used in oscillator and tuned circuits where a variable capacitor is needed. Operated with reverse bias, the depletion region forms a capacitor of variable width with a well-controlled voltage vs. capacitance function. Standard tuning diodes produce capacitances in the range of 5 to 40 pF. Hyper-abrupt tuning diodes produce variable capacitances to 100 pF or more for low frequency or wide-tuning-range applications.

The diode junction capacitance (C_j) under a reverse bias of V volts is given by

$$C_j = \frac{C_{j0}}{\sqrt{V_{on} - V}}$$

where C_{j0} = measured capacitance with zero applied voltage. As the reverse bias voltage on a diode increases, the width of the depletion region increases, decreasing its capacitance. Maximum capacitance occurs with minimum reverse voltage.

Note that the quantity under the radical is a large *positive* quantity for reverse bias. As seen from the equation, for large reverse biases C_j is inversely proportional to the square root of the voltage. Varactors are usually operated over a smaller range of bias so that the change in capacitance versus voltage is fairly linear.

Tuning diodes are specified by the capacitance produced at two reverse voltages, usually 2 and 30 V. This is called the *capacitance ratio* and is specified in units of pF per volt. Beyond this range, capacitance change with voltage can become non-linear and may cause signal distortion.

All diodes exhibit some capacitance when reverse biased. Amateurs have learned to use reverse biased Zener and rectifier diodes to form tuning diodes with 20 – 30 pF maximum capacitance. These "poor man's" tuning diodes are widely used in homebrew projects. However, because the capacitance ratio varies widely from one diode to the next, requiring experimentation to find a suitable diode, they are seldom used in published construction articles. When designing with varactor diodes, the reverse bias voltage must be absolutely free of noise since any variations in the bias voltage will cause changes in capacitance. For example, if the varactor is used to tune an oscillator, unwanted frequency shifts or instability will result if the reverse bias voltage is noisy. It is possible to frequency modulate a signal by adding the audio signal to the reverse bias on a varactor diode used in the carrier oscillator. (For examples of the use of varactors in oscillators and modulators, see the chapters on **Transmitting** and **Oscillators and Synthesizers**.)

ZENER DIODES

When the PIV of a reverse-biased diode is exceeded, the diode begins to conduct current as it does when it is forward biased. This current will not destroy the diode if it is limited to less than the device's maximum allowable value. By using heavy levels of doping during manufacture, a diode's PIV can be precisely controlled to be at a specific level, called the *Zener voltage*, creating a type of voltage reference. These diodes are called Zener diodes after their inventor, American physicist Clarence Zener.

When the Zener voltage is reached, the reverse voltage across the Zener diode remains constant even as the current through it changes. With an appropriate series current-limiting resistor, the Zener diode provides an accurate voltage reference (see **Figure 4.38**).

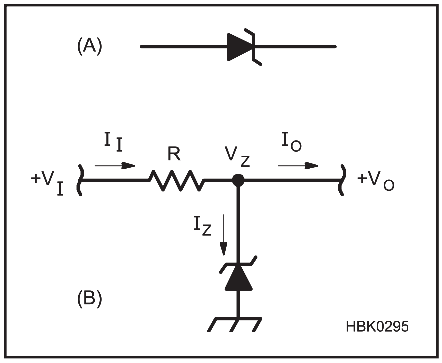


Figure 4.38 — Zener diode. (A) Schematic symbol. (B) Basic voltage regulating circuit. V_Z is the Zener reverse breakdown voltage. Above V_Z, the diode draws current until $V_I – I_IR = V_Z$. The circuit design should select R so that when the maximum current is drawn, $R < (V_I – V_Z) / I_O$. The diode should be capable of passing the same current when there is no output current drawn.

Zener diodes are rated by their reverse-breakdown voltage and their power-handling capacity, where $P = V_Z \times I_Z$. Since the same current must always pass though the resistor to drop the source voltage down to the reference voltage, with that current divided between the Zener diode and the load, this type of power source is very wasteful of current.

The Zener diode does make an excellent and efficient voltage reference in a larger voltage regulating circuit where the load current is provided from another device whose voltage is set by the reference. (See the **Power Sources** chapter for more information about using Zener diodes as voltage regulators.) When operating in the breakdown region, Zener diodes can be modeled as a simple voltage source.

Zener diodes are manufactured in a wide range of voltages and power handling capacities. Power dissipation in a Zener diode is equal to the Zener voltage multiplied by the average reverse current. The Zener voltage has a significant temperature coefficient and also varies with reverse current. To avoid excessive variations in Zener voltage, limit the diode's power dissipation to no more than 1/2 of the rated value; for precision uses, 1/5 to 1/10 of the rated power dissipation is recommended. Temperature-compensated Zener diodes are available with temperature coefficients as low as 5 parts per million per °C. If this is unacceptable, voltage reference integrated circuits based on Zener diodes have been developed that include additional circuitry to counteract temperature effects.

A variation of Zener diodes, *transient voltage suppressor* (*TVS*) diodes are designed to dissipate the energy in short-duration, high-voltage transients that would otherwise damage equipment or circuits. TVS diodes have large junction cross-sections so that they can handle large currents without damage. These diodes are also known by the trade name TransZorbs. Since the polarity of the transient can be positive, negative, or both, transient protection circuits can be designed with two devices connected with opposite polarities.

RECTIFIERS

The most common application of a diode is to perform rectification; that is, permitting current flow in only one direction. Power rectification converts ac current into pulsating dc current. There are three basic forms of power rectification using semiconductor diodes: half wave (1 diode), full-wave center-tapped (2 diodes) and full-wave bridge (4 diodes). These applications are shown in **Figure 4.39A**, **B** and **C** and are more fully described in the **Power Sources** chapter.

The most important diode parameters to consider for power rectification are the PIV and current ratings. The peak negative voltages that are blocked by the diode must be smaller in magnitude than the PIV and the peak current through the diode when it is forward biased must be less than the maximum average forward current.

Rectification is also used at much lower current levels in modulation and demodulation and other types of analog signal processing circuits. For these applications, the diode's response speed and junction forward voltage are the most important ratings.

4.6.3 Bipolar Junction Transistors (BJT)

The bipolar junction transistor is a *current-controlled device* with three basic terminals; *emitter*, *collector* and *base*. The current between the emitter and the collector is controlled by the current between the base and emitter. The convention when discussing transistor operation is that the three currents into the device are positive (I_c into the collector, I_b into the base and I_e into the emitter). Kirchhoff's Current Law applies to transistors just as it does to passive electrical networks: the total current entering the device must be zero. Thus, the relationship between the currents into a transistor can be generalized as

$$I_c + I_b + I_e = 0$$

which can be rearranged as necessary. For example, if we are interested in the emitter current,

$$I_e = –(I_c + I_b)$$

The back-to-back diode model shown in Figure 2.36 is appropriate for visualization of transistor construction. In actual transistors, however, the relative sizes of the collector, base and emitter regions differ. A common transistor configuration that spans a distance of 3 mm between the collector and emitter contacts typically has a base region that is only 25 μm across.

The operation of the bipolar transistor is described graphically by characteristic curves as shown in **Figure 4.40**. These are similar to the I-V characteristic curves for the two-terminal devices described in the preceding sections. The parameters shown by the curves depend on the type of circuit in which they are measured, such as common emitter or common collector. The output characteristic shows a set of curves for either collector or emitter current versus collector-emitter voltage at various values of input current (either base or emitter). The input characteristic shows the voltage between the input and common terminals (such as base-emitter) versus the input current for different values of output voltage.

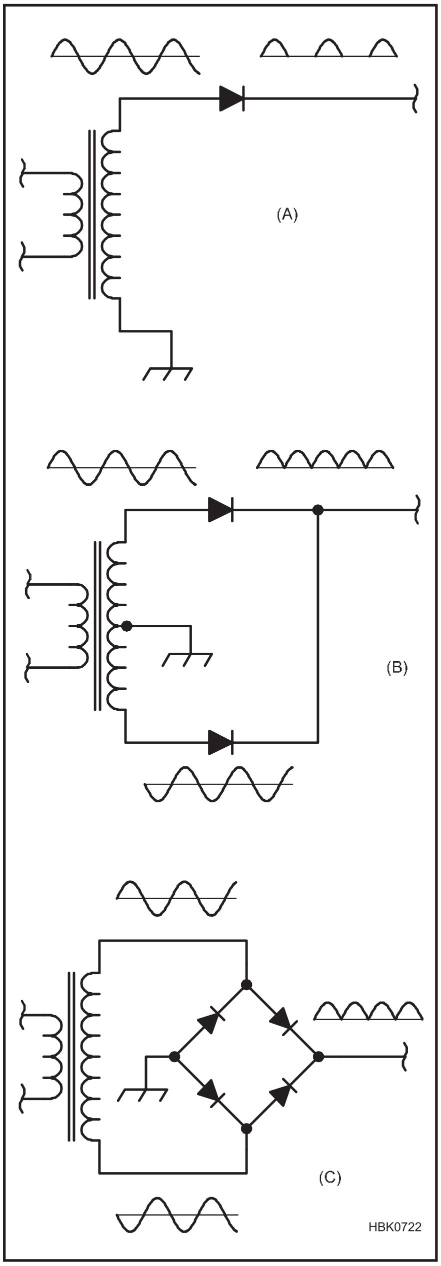


Figure 4.39 — Diode rectifier circuits. (A) Half wave rectifier circuit. Only when the ac voltage is positive does current pass through the diode. Current flows only during half of the cycle. (B) Full-wave center-tapped rectifier circuit. Center-tap on the transformer secondary is grounded and the two ends of the secondary are 180° out of phase. (C) Full-wave bridge rectifier circuit. In each half of the cycle two diodes conduct.

CURRENT GAIN

Two parameters describe the relationships between the three transistor currents at low frequencies:

$$\alpha = -\frac{\Delta I_C}{\Delta I_E} = 1$$

$$\beta = \frac{\Delta I_C}{\Delta I_B}$$

The relationship between α and β is defined as

$$\alpha = -\frac{\beta}{1+\beta}$$

Another designation for β is often used: h_{FE}, the *forward dc current gain*. (The "h" refers to "h parameters," a set of transfer parameters for describing a two-port network and described in more detail in the **RF Techniques** chapter.) The symbol h_{fe}, in which the subscript is in lower case, is used for the forward current gain of ac signals.

OPERATING REGIONS

Current conduction between collector and emitter is described by *regions* of the transistor's characteristic curves in Figure 4.40. (References such as *common-emitter* or *common-base* refer to the configuration of the circuit in which the parameter is measured.) The transistor is in its *active* or *linear region* when the base-collector junction is reverse biased and the base-emitter junction is forward biased. The slope of the output current, I_O, versus the output voltage, V_O, is virtually flat, indicating that the output current is nearly independent of the output voltage. In this region, the output circuit of the transistor can be modeled as a constant-current source controlled by the input current. The slight slope that does exist is due to base-width modulation (known as the *Early effect*).

When both the junctions in the transistor are forward biased, the transistor is said to be in its *saturation region*. In this region, V_O is nearly zero and large changes in I_O occur for very small changes in V_O. Both junctions in the transistor are reverse-biased in the *cutoff region*. Under this condition, there is very little current in the output, only the nanoamperes or microamperes that result from the very small leakage across the input-to-output junction. Finally, if V_O is increased to very high values, avalanche breakdown begins as in a PN-junction diode and output current increases rapidly. This is the *breakdown region*, not shown in Figure 4.40.

These descriptions of junction conditions are the basis for the use of transistors. Various configurations of the transistor in circuitry make use of the properties of the junctions to

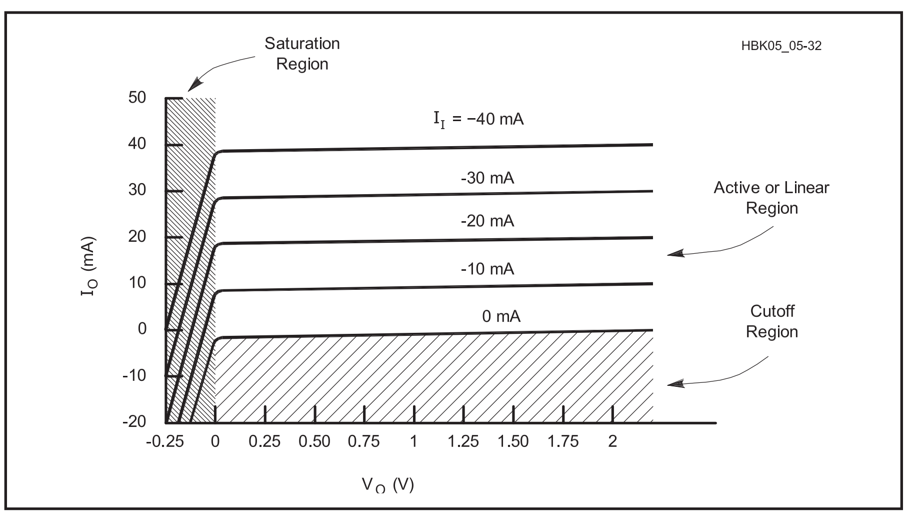


Figure 4.40 — Transistor response curve output characteristics. The X-axis is the output voltage, and the Y-axis is the output current. Different curves are plotted for various values of input current. The three regions of the transistor are its cutoff region, where no current flows in any terminal, its active region, where the output current is nearly independent of the output voltage and there is a linear relationship between the input current and the output current, and the saturation region, where the output current has large changes for small changes in output voltage.

serve different purposes in analog signal processing.

OPERATING PARAMETERS

A typical general-purpose bipolar-transistor data sheet lists important device specifications. Parameters listed in the ABSOLUTE MAXIMUM RATINGS section are the three junction voltages (V_{CEO}, V_{CBO} and V_{EBO}), the continuous collector current (I_C), the total device power dissipation (P_D) and the operating and storage temperature range. Exceeding any of these parameters is likely to cause the transistor to be destroyed. (The "O" in the suffixes of the junction voltages indicates that the remaining terminal is not connected, or open.)

In the OPERATING PARAMETERS section, three guaranteed minimum junction breakdown voltages are listed $V_{(BR)CEO}$, $V_{(BR)CBO}$ and $V_{(BR)EBO}$. Exceeding these voltages is likely to cause the transistor to enter avalanche breakdown, but if current is limited, permanent damage may not result.

Under ON CHARACTERISTICS are the guaranteed minimum dc current gain (β or h_{FE}), guaranteed maximum collector-emitter saturation voltage, $V_{CE(SAT)}$, and the guaranteed maximum base-emitter on voltage, $V_{BE(ON)}$. Two guaranteed maximum collector cutoff currents, I_{CEO} and I_{CBO}, are listed under OFF CHARACTERISTICS.

The next section is SMALL-SIGNAL CHARACTERISTICS, where the guaranteed minimum current gain-bandwidth product, BW or f_T, the guaranteed maximum output capacitance, C_{obo}, the guaranteed maximum input capacitance, C_{ibo}, the guaranteed range of input impedance, h_{ie}, the small-signal current gain, h_{fe}, the guaranteed maximum voltage feedback ratio, h_{re} and output admittance, h_{oe} are listed.

Finally, the SWITCHING CHARACTERISTICS section lists absolute maximum ratings for delay time, t_d; rise time, t_r; storage time, t_s; and fall time, t_f.

4.6.4 Field-Effect Transistors (FET)

FET devices are controlled by the voltage level of the input rather than the input current, as in the bipolar transistor. FETs have three basic terminals, the *gate*, the *source* and the *drain*. They are analogous to bipolar transistor terminals: the gate to the base, the source to the emitter, and the drain to the collector. Symbols for the various forms of FET devices are pictured in **Figure 4.41**.

The FET gate has a very high impedance, so the input can be modeled as an open circuit. The voltage between gate and source, V_{GS},

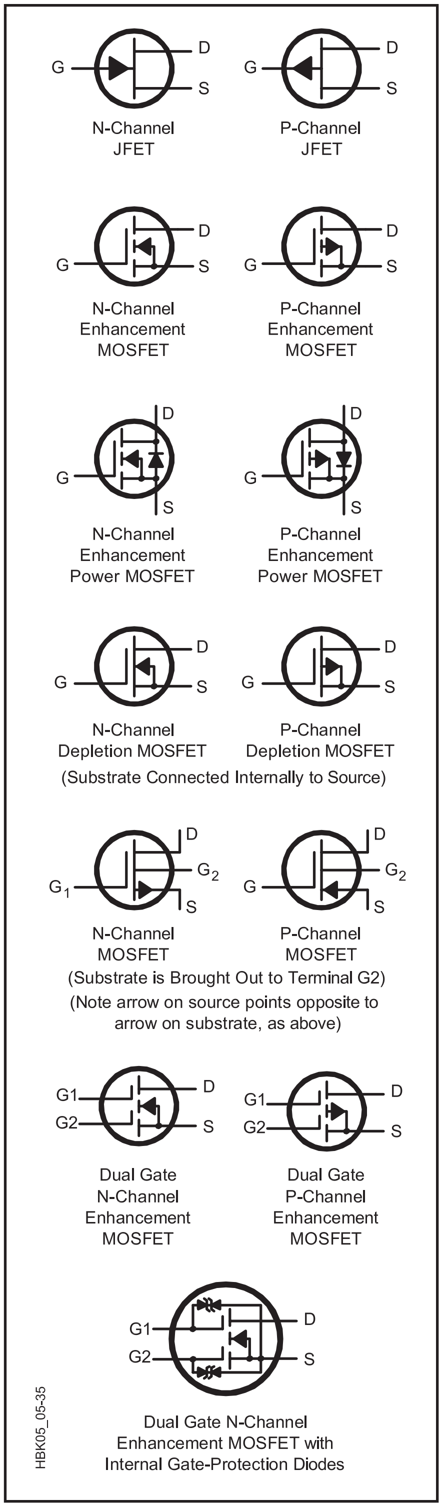


Figure 4.41 —FET schematic symbols.

controls the resistance of the drain-source channel, r_{DS}, and so the output of the FET is modeled as a current source, whose output current is controlled by the input voltage.

The action of the FET channel is so nearly

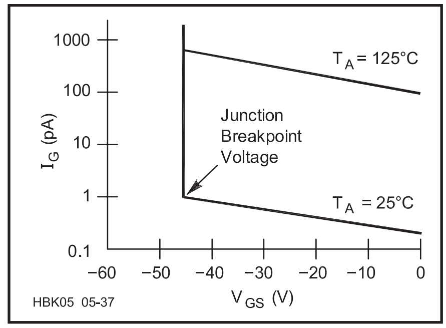


Figure 4.42 — JFET input leakage curves for common source amplifier configuration. Input voltage (V_{GS}) on the X-axis versus input current (I_G) on the Y-axis, with two curves plotted for different operating temperatures, 25 °C and 125 °C. Input current increases greatly when the gate voltage exceeds the junction breakpoint voltage.

ideal that, as long as the JFET gate does not become forward biased and inject current from the base into the channel, the drain and source currents are virtually identical. For JFETs the *gate leakage current,* I_G, is a function of V_{GS} and this is often expressed with an *input curve* (see **Figure 4.42**). The point at which there is a significant increase in I_G is called the *junction breakpoint voltage*. Because the gate of MOSFETs is insulated from the channel, gate leakage current is insignificant in these devices.

The dc channel resistance, r_{DS}, is specified in data sheets to be less than a maximum value when the device is biased on ($r_{DS(on)}$). When the gate voltage is maximum (V_{GS} = 0 for a JFET), $r_{DS(on)}$ is minimum. This describes the effectiveness of the device as an analog switch. Channel resistance is approximately the same for ac and dc signals until at high frequencies the capacitive reactances inherent in the FET structure become significant.

FETs also have strong similarities to vacuum tubes in that input voltage between the grid and cathode controls an output current between the plate and cathode. (See the chapter on **RF Power Amplifiers** for more information on vacuum tubes.)

Both JFET and MOSFET devices can be used in a wide variety of voltage-controlled circuits from variable attenuators to variable-gain amplifiers to servo circuits. A complete tutorial on these circuits is beyond the scope of this section but the subject is well-covered by the five-part series of articles by Ron Quan of Linear Integrated Circuits. It was published by *EDN* at **www.edn.com/a-guide-to-using-fets-for-voltage-controlled-circuits-part-1**. (Replace -1 with the part you wish to read.)

FORWARD TRANSCONDUCTANCE

The change in FET drain current caused by a change in gate-to-source voltage is called *forward transconductance, g_m*.

$$g_m = \frac{\Delta I_{DS}}{\Delta V_{GS}}$$

or

$$\Delta I_{DS} = g_m \, DV_{GS}$$

The input voltage, $V_{GS,}$ is measured between the FET gate and source and drain current, I_{DS}, flows from drain to source. Analogous to a bipolar transistor's current gain, the units of transconductance are siemens (S) because it is the ratio of current to voltage. (Both g_m and g_{fs} are used interchangeably to indicate transconductance. Some sources specify g_{fs} as the *common-source forward transconductance*. This chapter uses g_m, the most common convention in the reference literature.)

OPERATING REGIONS

The most useful relationships for FETs are the output and transconductance response characteristic curves in **Figure 4.43**. (References such as *common-source* or *common-gate* refer to the configuration of the circuit in which the parameter is measured.) Transconductance curves relate the drain current, I_D, to gate-to-source voltage, V_{GS}, at various drain-source voltages, V_{DS}. The FET's forward transconductance, g_m, is the slope of the lines in the forward transconductance curve. The same parameters are interrelated in a different way in the output characteristic, in which I_D is shown versus V_{DS} for different values of V_{GS}.

Like the bipolar transistor, FET operation can be characterized by regions. The *ohmic region* is shown at the left of the FET output characteristic curve in **Figure 4.44** where I_D is increasing nearly linearly with V_{DS} and the FET is acting like a resistance controlled by V_{GS}. As V_{DS} continues to increase, I_D saturates and becomes nearly constant. This is the FET's *saturation region* in which the channel of the FET can be modeled as a constant-current source. V_{DS} can become so large that V_{GS} no longer controls the conduction of the device and avalanche breakdown occurs as in bipolar transistors and PN-junction diodes. This is the *breakdown region*, shown in Figure 4.44 where the curves for I_D break sharply upward. If V_{GS} is less than V_P, so that transconductance is zero, the FET is in the *cutoff region*.

OPERATING PARAMETERS

A typical FET data sheet gives ABSOLUTE MAXIMUM RATINGS for V_{DS}, V_{DG}, V_{GS} and I_D, along with the usual device dissipation (P_D) and storage temperature range. Exceeding these limits usually results in destruction of the FET.

Under OPERATING PARAMTERS the OFF CHARACTERISTICS list the gate-source breakdown voltage, $V_{GS(BR)}$, the reverse gate current, I_{GSS} and the gate-source cutoff voltage, $V_{GS(OFF)}$. Exceeding $V_{GS(BR)}$ will not permanently damage the device if current is limited. The primary ON CHARACTERISTIC parameters are the channel resistance, r_{DS}, and the zero-gate-voltage drain current (I_{DSS}). An FET's dc channel resistance, r_{DS}, is specified in data sheets to be less than a maximum value when the device is biased on ($r_{DS(on)}$). For ac signals, $r_{ds(on)}$ is not necessarily the same as $r_{DS(on)}$, but it is not very different as long as the frequency is not so high that capacitive reactance in the FET becomes significant.

The SMALL SIGNAL CHARACTERISTICS include the forward transfer admittance, y_{fs}, the output admittance, y_{os}, the static drain-source on resistance, $r_{ds(on)}$ and various capacitances such as input capacitance, C_{iss}, reverse transfer capacitance, C_{rss}, and the drain-substrate capacitance, $C_{d(sub)}$. FUNCTIONAL CHARACTERISTICS include the noise figure, NF, and the common source power gain, G_{ps}.

MOSFETS

As described earlier, the MOSFET's gate is insulated from the channel by a thin layer of nonconductive oxide, doing away with any appreciable gate leakage current. Because of this isolation of the gate, MOSFETs do not need input and reverse transconductance curves. Their output curves (**Figure 4.45**) are similar to those of the JFET. The gate acts as a small capacitance between the gate and both the source and drain.

The output and transconductance curves in Figure 4.45A and 4.45B show that the depletion-mode N-channel MOSFET's transconductance is positive at $V_{GS} = 0$, like that of the N-channel JFET. Unlike the JFET, however, increasing V_{GS} does not forward-bias the gate-source junction and so the device can be operated with $V_{GS} > 0$.

In the enhancement-mode MOSFET, transconductance is zero at $V_{GS} = 0$. As V_{GS} is increased, the MOSFET enters the ohmic region. If V_{GS} increases further, the saturation region is reached and the MOSFET is said to be *fully-on*, with r_{DS} at its minimum value. The behavior of the enhancement-mode MOSFET is similar to that of the bipolar transistor in this regard.

The relatively flat regions in the MOSFET output curves are often used to provide a con-

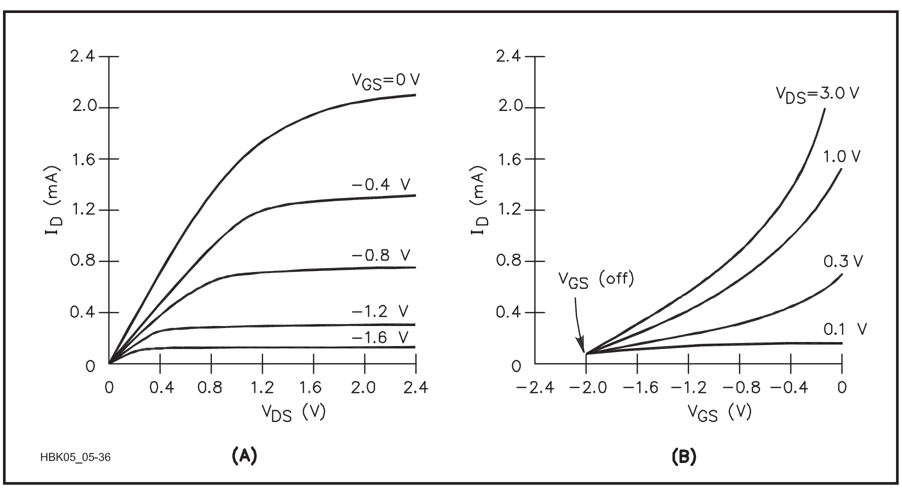


Figure 4.43 — JFET output and transconductance response curves for common source amplifier configuration. (A) Output voltage (V_{DS}) on the X-axis versus output current (I_D) on the Y-axis, with different curves plotted for various values of input voltage (V_{GS}). (B) Transconductance curve with the same three variables rearranged: V_{GS} on the X-axis, I_D on the Y-axis and curves plotted for different values of V_{DS}.

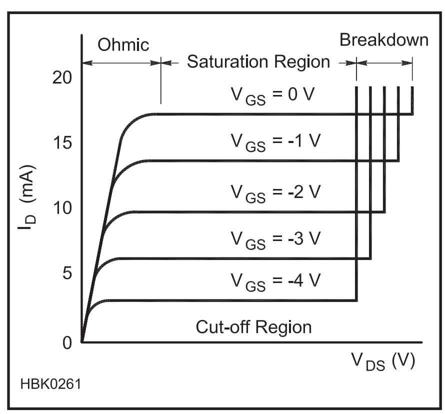


Figure 4.44 — JFET operating regions. At the left, I_D is increasing rapidly with V_{GS} and the JFET can be treated as resistance (R_{DS}) controlled by V_{GS}. In the saturation region, drain current, I_D, is relatively independent of V_{GS}. As V_{DS} increases further, avalanche breakdown begins and I_D increases rapidly.

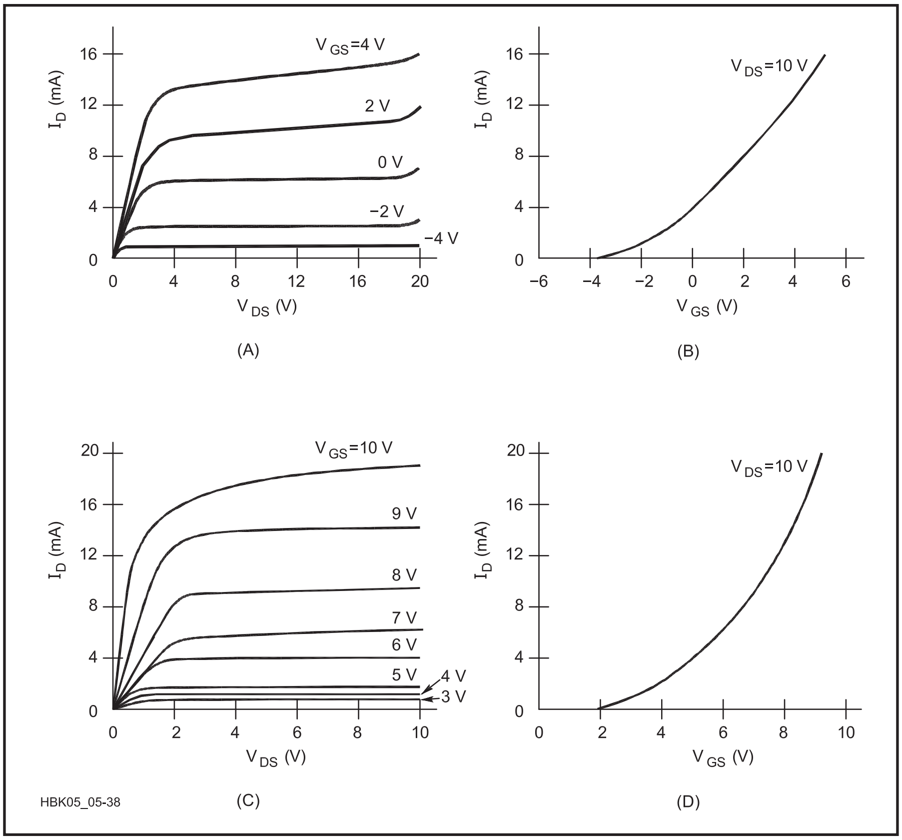


Figure 4.45 — MOSFET output [(A) and (C)] and transconductance [(B) and (D)] response curves. Plots (A) and (B) are for an N-channel depletion mode device. Note that V_{GS} varies from negative to positive values. Plots (C) and (D) are for an N-channel enhancement mode device. V_{GS} has only positive values.

stant current source. As is plotted in these curves, the drain current, I_D, changes very little as the drain-source voltage, V_{DS}, varies in this portion of the curve. Thus, for a fixed gate-source voltage, V_{GS}, the drain current can be considered to be constant over a wide range of drain-source voltages.

Multiple-gate MOSFETs are also available. Due to the insulating layer, the two gates are isolated from each other and allow two signals to control the channel simultaneously with virtually no loading of one signal by the other. A common application of this type of device is an automatic gain control (AGC) amplifier. The signal is applied to one gate and a rectified, low-pass filtered form of the output (the AGC voltage) is fed back to the other gate. Another common application is as a mixer in which the two input signals are applied to the pair of gates.

These two application notes are excellent sources of information for using these devices in radio applications. They can be found online by searching for the full name. *AN1020 — Active Mixer Design Using the NE25139 Dual Gate MESFET* from California Eastern Laboratories and *AN423 — Field Effect Transistor RF Amplifier Design Techniques* from Freescale Semiconductor.

MOSFET Gate Protection

The MOSFET is constructed with a very thin layer of SiO_2 for the gate insulator. This layer is extremely thin in order to improve the transconductance of the device, but this makes it susceptible to damage from high voltage levels, such as *electrostatic discharge* (ESD) from static electricity. If enough charge accumulates on the gate terminal, it can *punch through* the gate insulator and destroy it. The insulation of the gate terminal is so good that virtually none of this potential is eased by leakage of the charge into the device. While this condition makes for nearly ideal input impedance (approaching infinity), it puts the device at risk of destruction from even such seemingly innocuous electrical sources as static electrical discharges from handling.

Some MOSFET devices contain an internal Zener diode with its cathode connected to the gate and its anode to the substrate. If the voltage at the gate rises to a damaging level the Zener diode junction conducts, bleeding excess charges off to the substrate. When voltages are within normal operating limits the Zener has little effect on the signal at the gate, although it may decrease the input impedance of the MOSFET.

This solution will not work for all MOSFETs. The Zener diode must always be reverse biased to be effective. In the enhancement-mode MOSFET, $V_{GS} > 0$ for all valid uses of the part, keeping the Zener reverse biased. In depletion mode devices however, V_{GS} can be both positive and negative; when negative, a gate-protection Zener diode would be forward biased and the MOSFET gate would not be driven properly. In some depletion mode MOSFETs, back-to-back Zener diodes are used to protect the gate.

MOSFET devices are at greatest risk of damage from static electricity when they are out of circuit. Even though an electrostatic discharge is capable of delivering little energy, it can generate thousands of volts and high peak currents. When storing MOSFETs, the leads should be placed into conductive foam. When working with MOSFETs, it is a good idea to minimize static by wearing a grounded wrist strap and working on a grounded workbench or mat. A humidifier may help to decrease the static electricity in the air. Before inserting a MOSFET into a circuit board it helps to first touch the device leads with your hand and then touch the circuit board. This serves to equalize the excess charge so that little excess charge flows when the device is inserted into the circuit board.

Power MOSFETs

Power MOSFETs are designed for use as switches, with extremely low values of $r_{DS(on)}$; values of 50 milliohms (mΩ) are common. The largest devices of this type can switch tens of amps of current with V_{DS} voltage ratings of hundreds of volts. The schematic symbol for power MOSFETs (see Figure 4.30) includes a *body diode* that allows the FET to conduct in the reverse direction, regardless of V_{GS}. This is useful in many high-power switching applications. Power MOSFETs used for RF amplifiers are discussed in more detail in the **RF Power Amplifiers** chapter.

While the maximum ratings for current and voltage are high, the devices cannot withstand both high drain current and high drain-to-source voltage at the same time because of the power dissipated; $P = V_{DS} \times I_D$. It is important to drive the gate of a power MOSFET such that the device is fully on or fully off so that either V_{DS} or I_D is at or close to zero. When switching, the device should spend as little time as possible in the linear region where both current and voltage are nonzero because their product (P) can be substantial. This is not a big problem if switching only takes place occasionally, but if the switching is repetitive (such as in a switching power supply) care should be taken to drive the gate properly and remove excess heat from the device.

Because the gate of a power MOSFET is capacitive (up to several hundred pF for large devices), charging and discharging the gate quickly results in short current peaks of more than 100 mA. Whatever circuit is used to drive the gate of a power MOSFET must be able to handle that current level, such as an integrated circuit designed for driving the capacitive load an FET gate presents.

The gate of a power MOSFET should not be left open or connected to a high-impedance circuit. Use a pull-down or pull-up resistor connected between the gate and the appropriate power supply to ensure that the gate is placed at the right voltage when not being driven by the gate drive circuit.

GaAsFETs

FETs made from gallium-arsenide (GaAs) material are used at UHF and microwave frequencies because they have gain at these frequencies and add little noise to the signal. The reason GaAsFETs have gain at these frequencies is the high mobility of the electrons in GaAs material. Because the electrons are more mobile than in silicon, they respond to the gate-source input signal more quickly and strongly than silicon FETs, providing gain at higher frequencies (f_T is directly proportional to electron mobility). The higher electron mobility also reduces thermally generated noise in the FET, making the GaAsFET especially suitable for weak-signal preamps.

Because electron mobility is always higher than hole mobility, N-type material is used in GaAsFETs to maximize high-frequency gain. Since P-type material is not used to make a gate-channel junction, a metal Schottky junction is formed by depositing metal directly on the surface of the channel. This type of device is also called a *MESFET* (metal-semiconductor field-effect transistor).

4.6.5 Comparison of BJT and FET Devices

Analog signal processing deals with changing a signal to a desired form. The three primary types of devices — bipolar transistors, field-effect transistors and integrated circuits — perform similar functions, each with specific advantages and disadvantages. The vacuum tube, once the dominant signal processing component, is relegated to high-power amplifier and display applications and is found only in the **RF Power Amplifiers** chapter of this *Handbook*. *Cathode-ray tubes* (CRTs) are covered in a separate article in this book's online material.

Bipolar transistors, when treated properly, can have virtually unlimited life spans. They are relatively small and, if they do not handle high currents, do not generate much heat. They make excellent high-frequency amplifiers. Compared to MOSFET devices they are less susceptible to damage from electrostatic discharge. Bipolar transistors and ICs, like all semiconductors, are susceptible to damage from power and lightning transients.

There are many performance advantages to FET devices, particularly MOSFETs. The extremely low gate currents allow the design of analog stages with nearly infinite input resistance. Signal distortion due to loading is minimized in this way. FETs are less expensive to fabricate in ICs and so are gradually replacing bipolar transistors in many IC applications.

RF amplifiers are now designed almost exclusively using some variety of MOSFET in their final amplifiers. The transistors are often integrated into modules (a.k.a. "pallets") that include circuitry to protect the transistors from the high voltages generated by reflections under high SWR conditions. See the **RF Power Amplifier** chapter for more information on advances in RF amplifier technology.

An important consideration in the use of analog components is the future availability of parts. At an ever-increasing rate, as new components are developed to replace older technology, the older components are discontinued by the manufacturers and become unavailable for future use. ASIC and PGA technology, discussed along with integrated circuits, brings the power of custom electronics to the radio, but can make it nearly impossible to repair by replacing an IC, even if the problem is known. If field repair and service at the component level are to be performed, it is important to use standard ICs wherever possible. Even so, when demand for a particular component drops, a manufacturer will discontinue its production. This happens on an ever-decreasing timeline.

A further consideration is the trend toward digital signal processing and software-defined radio systems. (See the chapter on **DSP and SDR Fundamentals**.) More and more analog functions are being performed by microprocessors and the analog signals converted to digital at higher and higher frequencies. It is now common practice to digitize the incoming RF signal directly at the antenna system interface.

There will always be a need for analog circuits, but the balance point between analog and digital is accelerating towards the latter. In future years, radio and test equipment will consist of a powerful, general-purpose digital signal processor, surrounded by the necessary analog circuitry to convert the signals to digital form and supply the processor with power.

4.6.6 Optical Semiconductors

In addition to electrical energy and heat energy, light energy also affects the behavior of semiconductor materials. If a device is made to allow photons of light to strike the surface of the semiconductor material, the energy absorbed by electrons disrupts the bonds between atoms, creating free electrons and holes. This increases the conductivity of the material (*photoconductivity*). The photon can also transfer enough energy to an electron

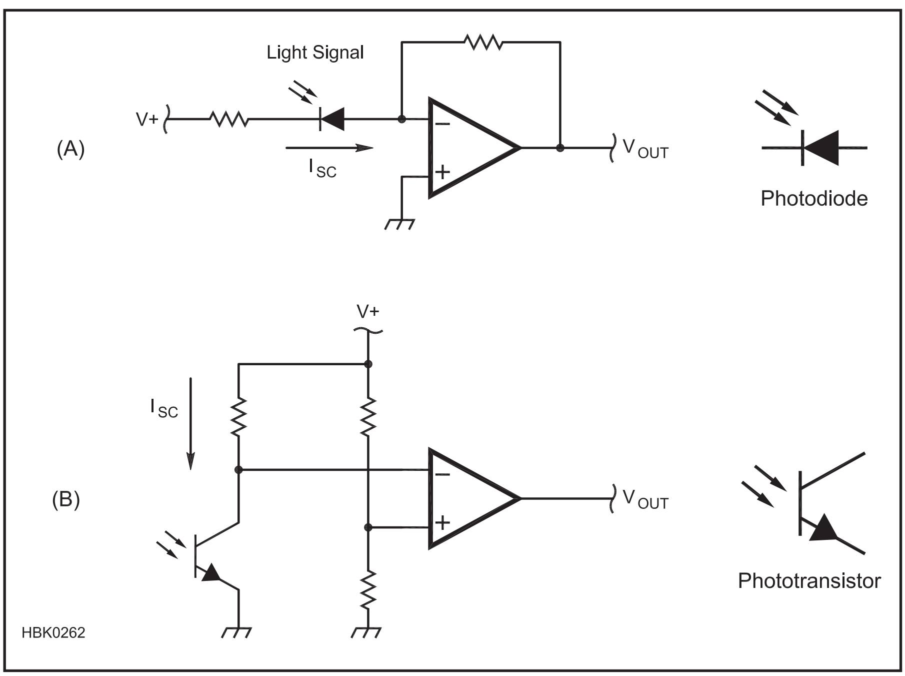


Figure 4.46 — The photodiode (A) is used to detect light. An amplifier circuit changes the variations in photodiode current to a change in output voltage. At (B), a photo-transistor conducts current when its base is illuminated. This causes the voltage at the collector to change causing the amplifier's output to switch between ON and OFF.

to allow it to cross a PN junction's depletion region as current flow through the semiconductor (*photoelectricity*).

PHOTOCONDUCTORS

In commercial *photoconductors* (also called *photoresistors*) the resistance can change by as much as several kilohms for a light intensity change of 100 ft-candles. The most common material used in photoconductors is cadmium sulfide (CdS), with a resistance range of more than 2 MΩ in total darkness to less than 10 Ω in bright light. Other materials used in photoconductors respond best at specific colors. Lead sulfide (PbS) is most sensitive to infrared light and selenium (Se) works best in the blue end of the visible spectrum.

PHOTODIODES

A similar effect is used in some diodes and transistors so that their operation can be controlled by light instead of electrical current biasing. These devices, shown in **Figure 4.46**, are called *photodiodes* and *phototransistors*. The flow of minority carriers across the reverse biased PN junction is increased by light falling on the doped semiconductor material. In the dark, the junction acts the same as any reverse biased PN junction, with a very low current, I_{SC}, (on the order of 10 µA) that is nearly independent of reverse voltage. The presence of light not only increases the current but also provides a resistance-like relationship (reverse current increases as reverse voltage increases). See **Figure 4.47** for the characteristic response of a photodiode. Even with no reverse voltage applied, the presence of light causes a small reverse current, as indicated by the points at which the lines in Figure 4.47 intersect the left side of the graph.

Photoconductors and photodiodes are generally used to produce light-related analog signals that require further processing. For example, a photodiode is used to detect infrared light signals from remote control devices as in Figure 4.46A. The light falling on the reverse-biased photodiode causes a change in I_{SC} that is detected as a change in output voltage.

Light falling on the phototransistor acts as base current to control a larger current between the collector and emitter. Thus, the phototransistor acts as an amplifier whose input signal is light and whose output is current. Phototransistors are more sensitive to light than the other devices. Phototransistors have lots of light-to-current gain, but photodiodes normally have less noise, so they make more sensitive detectors. The phototransistor in Figure 4.46B is being used as a detector. Light falling on the phototransistor causes collector current to flow, dropping the collector voltage below the voltage at the amplifier's + input and causing a change in V_{OUT}.

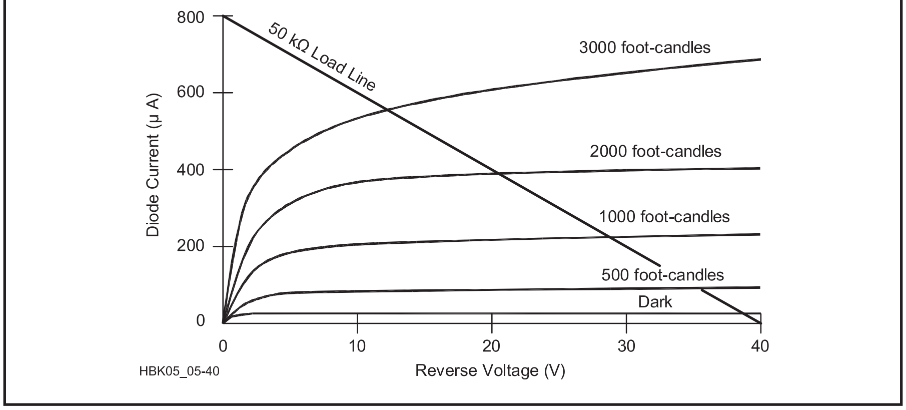


Figure 4.47 — Photodiode I-V curve. Reverse voltage is plotted on the X-axis and current through diode is plotted on the Y-axis. Various response lines are plotted for different illumination. Except for the zero illumination line, the response does not pass through the origin since there is current generated at the PN junction by the light energy. A load line is shown for a 50-kΩ resistor in series with the photodiode.

PHOTOVOLTAIC CELLS

When illuminated, the reverse-biased photodiode has a reverse current caused by excess minority carriers. As the reverse voltage is reduced, the potential barrier to the forward flow of majority carriers is also reduced. Since light energy leads to the generation of both majority and minority carriers, when the resistance to the flow of majority carriers is decreased these carriers form a forward current. The voltage at which the forward current equals the reverse current is called the *photovoltaic potential* of the junction. If the illuminated PN junction is not connected to a load, a voltage equal to the photovoltaic potential can be measured across it as the *terminal voltage,* V_T, or *open-circuit voltage,* V_{OC}.

Devices that use light from the sun to produce electricity in this way are called *photovoltaic (PV)* or *solar cells* or *solar batteries*. The symbol for a photovoltaic cell is shown in **Figure 4.48A**. The electrical equivalent circuit of the cell is shown in Figure 4.48B. The cell is basically a large, flat diode junction exposed to light. Metal electrodes on each side of the junction collect the current generated.

When illuminated, the cell acts like a current source, with some of the current flowing through a diode (made of the same material as the cell), a shunt resistance for leakage current and a series resistor that represents the resistance of the cell. Two quantities define the electrical characteristics of common silicon photovoltaic cells. These are an open-

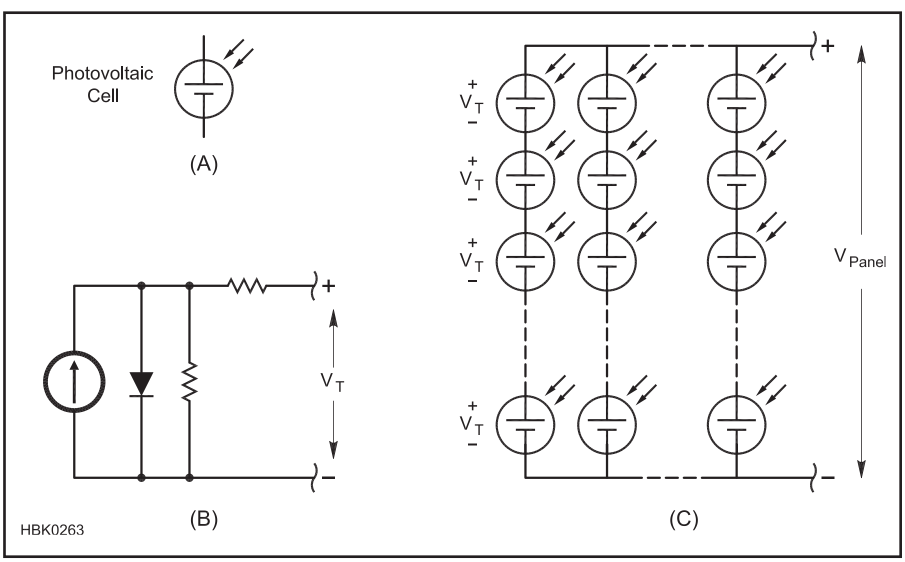


Figure 4.48 — A photovoltaic cell's symbol (A) is similar to a battery. Electrically, the cell can be modeled as the equivalent circuit at (B). Solar panels (C) consist of arrays of cells connected to supply power at a convenient voltage.

circuit voltage, V_{OC} of 0.5 to 0.6 V and the output *short-circuit current*, I_{SC} as above, that depends on the area of the cell exposed to light and the degree of illumination. A measure of the cell's effectiveness at converting light into current is the *conversion efficiency*. Typical silicon solar cells have a conversion efficiency of 10 to 15% although special cells with stacked junctions or using special light-absorbing materials have shown efficiencies as high as 40%.

Solar cells are primarily made from single-crystal slices of silicon, similar to diodes and transistors, but with a much greater area. *Polycrystalline silicon* and *thin-film* cells are less expensive but have lower conversion efficiency. Technology is advancing rapidly in the field of photovoltaic energy and there are a number of different types of materials and fabrication techniques that have promise in surpassing the effectiveness of the single-junction silicon cells.

Solar cells are assembled into arrays called *solar panels*, shown in Figure 4.48C. Cells are connected in series so that the combined output voltage is a more useful voltage, such as 12 V. Several strings of cells are then connected in parallel to increase the available output current. Solar panels are available with output powers from a few watts to hundreds of watts. Note that unlike batteries, strings of solar cells can be connected directly in parallel because they act as sources of constant current instead of voltage. (More information on the use of solar panels for powering radio equipment can be found in the chapter on **Power Sources**.)

LIGHT EMITTING DIODES (LEDs) AND LASER DIODES

In the photodiode, energy from light falling on the semiconductor material is absorbed to create additional electron-hole pairs. When the electrons and holes recombine, the same amount of energy is given off. In normal diodes, the energy from recombination of carriers is given off as heat. In certain forms of semiconductor material, the recombination energy is given off as light with a mechanism called *electroluminescence*. Devices made for this purpose are called *light emitting diodes* (*LEDs*).

The LED emits light when it is forward biased and excess carriers are present. As the carriers recombine, light is produced with a color that depends on the properties of the semiconductor material used. Gallium-arsenide (GaAs) generates light in the infrared region, gallium-phosphide (GaP) gives off red light when doped with oxygen or green light when doped with nitrogen. Orange light is attained with a mixture of GaAs and GaP (GaAsP). Silicon-carbide (SiC) creates a blue LED. White LEDs are made by coating the inside of the LED lens with a white-light emitting phosphor and illuminating the phosphor with light from a single-color LED.

LEDs have the advantages of low power requirements, fast switching times (on the order of 10 ns), and narrow spectra (relatively pure color). The primary application of low-power LEDs is as an illuminated visual indicator, replacing miniature incandescent bulbs for indicators and illumination. (High-power LEDs used for lighting and large displays are not discussed here.) LEDs are specified primarily by their color, size, shape, and output light intensity.

The "standard" size LED used in electronics is the T-13/4, which is 5 mm or 0.20 inches in diameter. The "miniature" size is the T-1, which is 3 mm or 0.125 inches in diameter. Today, the "standard" and "miniature" size is a bit of a misnomer due to the wide variety of LED sizes and shapes, including SMT varieties. However, the T-1 and T-13/4 remain the most common for homebrew projects due to their inexpensive availability and ease of mounting with a simple panel hole. Their long leads are ideal for prototyping. The LED, shown in **Figure 4.49**, is very simple to use. It is connected to a voltage source (V) with a series resistor (R) that limits the current to the desired level (I_F) for the amount of light to be generated.

$$R = \frac{V - V_F}{I_F}$$

where V_F is the forward voltage of the LED.

The cathode lead is connected to the lower potential and is specially marked as shown in the manufacturer's data sheet. LEDs may be connected in series for additional light, with the same current flowing in all of the diodes. Diodes connected in parallel without current-limiting resistors for each diode are likely to share the current unequally, thus the series connection is preferred.

The laser diode operates by similar principles to the LED except that all of the light produced is *monochromatic* (of the same color and wavelength) and it is *coherent*, meaning that all of the light waves emitted by the device are in phase. Laser diodes generally require higher current than an LED and will not emit light until the *lasing current* level is reached. Because the light is monochromatic and coherent, laser diodes can be used for applications requiring precise illumination and modulation, such as high-speed data links, and in data storage media such as CD-ROM and DVD. LEDs are not used for high-speed or high-frequency analog modulation because of recovery time limitations, just as in regular rectifiers.

OPTOISOLATORS

An interesting combination of optoelectronic components proves very useful in many analog signal processing applications. An *optoisolator* consists of an LED optically coupled to a phototransistor, usually in an enclosed package (see **Figure 4.50**). The optoisolator, as its name suggests, isolates different circuits from each other. Typically, isolation resistance is on the order of 10^{11} Ω and isolation capacitance is less than 1 pF. Maximum voltage isolation varies from 1,000 to 10,000 V ac. The most common optoisolators are available in six-pin DIP packages.

Optoisolators are primarily used for voltage level shifting and signal isolation. Voltage level shifting allows signals (usually digital signals) to pass between circuits operating at greatly different voltages. The isolation has two purposes: to protect circuitry (and operators) from excessive voltages and to isolate noisy circuitry from noise-sensitive circuitry.

Optoisolators also cannot transfer signals with high power levels. The power rating of the LED in a 4N25 device is 120 mW. Optoisolators have a limited frequency response due to the high capacitance of the LED. A typical bandwidth for the 4N25 series is 300 kHz. Optoisolators with bandwidths of

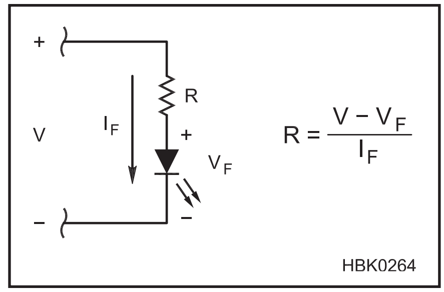


Figure 4.49 — A light-emitting diode (LED) emits light when conducting forward current. A series current-limiting resistor is used to set the current through the LED according to the equation.

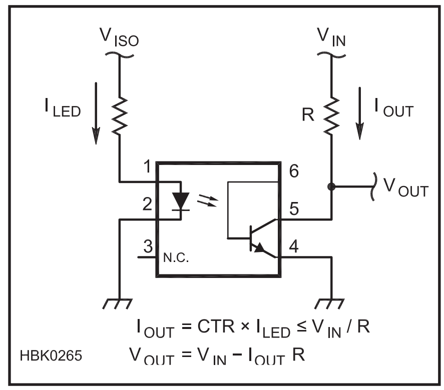


Figure 4.50 — The optoisolator consists of an LED (input) that illuminates the base of a phototransistor (output). The phototransistor then conducts current in the output circuit. CTR is the optoisolator's current transfer ratio.

several MHz are available, but are somewhat expensive.

As an example of voltage level shifting, an optoisolator can be used to allow a low-voltage, solid-state electronic Morse code keyer to activate a vacuum-tube grid-block keying circuit that operates at a high negative voltage (typically about –100 V) but low current. No common ground is required between the two pieces of equipment.

Optoisolators can act as input protection for circuits that are exposed to high voltages or transients. For example, a short 1,000-V transient that can destroy a semiconductor circuit will only saturate the LED in the optoisolator, preventing damage to the circuit. The worst that will happen is the LED in the optoisolator will be destroyed, but that is usually quite a bit less expensive than the circuit it is protecting.

Optoisolators are also useful for isolating different ground systems. The input and output signals are totally isolated from each other, even with respect to the references for each signal. A common application for optoisolators is when a computer is used to control radio equipment. The computer signal, and even its ground reference, typically contains considerable wide-band noise caused by the digital circuitry. The best way to keep this noise out of the radio is to isolate both the signal and its reference; this is easily done with an optoisolator.

The design of circuits with optoisolators is not greatly different from the design of circuits with LEDs and with transistors. On the input side, the LED is forward-biased and driven with a series current-limiting resistor whose value limits current to less than the maximum value for the device (for example, 60 mA is the maximum LED current for a 4N25). This is identical to designing with standalone LEDs.

On the output side, instead of current gain for a transistor, the optoisolator's *current transfer ratio* (*CTR*) is used. CTR is a ratio given in percent between the amount of current through the LED to the output transistor's maximum available collector current. For example, if an optoisolator's CTR = 25%, then an LED current of 20 mA results in the output transistor being able to conduct up to $20 \times 0.25 = 5$ mA of current in its collector circuit.

If the optoisolator is to be used for an analog signal, the input signal must be appropriately dc shifted so that the LED is always forward biased. A phototransistor with all three leads available for connection (as in Figure 4.50) is required. The base lead is used for biasing, allowing the optical signal to create variations above and below the transistor's operating point. The collector and emitter leads are used as they would be in any transistor amplifier circuit. (There are also linear optoisolators that include built-in linearizing circuitry.) The use of linear optoisolators is not common.

FIBER OPTICS

An interesting variation on the optoisolator is the *fiber-optic* connection. Like the optoisolator, the input signal is used to drive an LED or laser diode that produces modulated light (usually light pulses). The light is transmitted in a fiber optic cable, an extruded glass fiber that efficiently carries light over long distances and around fairly sharp bends. The signal is recovered by a photo detector (photoresistor, photodiode or phototransistor). Because the fiber optic cable is nonconductive, the transmitting and receiving systems are electrically isolated.

Fiber optic cables generally have far less loss than coaxial cable transmission lines. They do not leak RF energy, nor do they pick up electrical noise. Fiber optic cables are virtually immune to electromagnetic interference! Special forms of LEDs and phototransistors are available with the appropriate optical couplers for connecting to fiber optic cables. These devices are typically designed for higher frequency operation with gigahertz bandwidth.

4.6.7 Integrated Circuits (ICs)

If you look inside a transistor, the actual size of the semiconductor is quite small compared to the size of the packaging. For most semiconductors, the packaging takes considerably more space than the actual semiconductor device. Thus, an obvious way to reduce the physical size of circuitry is to combine more of the circuit inside a single package.

HYBRID INTEGRATED CIRCUITS

It is easy to imagine placing several small semiconductor chips in the same package. This is known as *hybrid circuitry*, a technology in which several semiconductor chips are placed in the same package and miniature wires are connected between them to make complete circuits.

Hybrid circuits miniaturize analog or analog/digital electronic circuits by eliminating much of the packaging and interconnections inherent in discrete electronics. The term *discrete* refers to the use of individual components to make a circuit, each in its own package. The individual components are attached together on a small circuit board or with bonding wires. Manufacturers often use hybrids when small size and specialized techniques are needed, but there is insufficient volume to justify the expense of a custom IC.

A current application for hybrid circuitry is UHF and microwave amplifiers — they are in wide use by the mobile phone industry. For example, the NXP MW7IC930N Wideband Integrated Power Amplifier is a complete transmitting module used in cellular base stations. Its TO-272 package in **Figure 4.51** is only about 1 inch long by 3/8-inch wide. This particular device is designed for use between 728 and 960 MHz and can be adapted for use on the amateur 902 MHz band. Other devices available as hybrid circuits include oscillators, signal processors, preamplifiers, and so forth. Surplus hybrids can be hard to adapt to amateur use unless they are clearly identified with manufacturing marks that will help in obtaining a data sheet.

MONOLITHIC INTEGRATED CIRCUITS

In order to build entire circuits on a single piece of semiconductor, it must be possible to fabricate resistors and capacitors, as well as transistors and diodes. Only then can the entire circuit be created on one piece of silicon called a *monolithic integrated circuit*. The following description is a greatly simplified illustration of monolithic integrated circuit structure and fabrication techniques.

An integrated circuit (IC) or "chip" is fabricated in layers. An example of a semiconductor circuit schematic and its implementation in an IC is pictured in **Figure 4.52**. The base layer of the circuit, the *substrate*, is made of P-type semiconductor material. Although less common, the polarity of the substrate can also be N-type material. Since the mobility of electrons is about three times higher than that of holes, bipolar transistors made with N-type collectors and FETs made with N-type channels are capable of higher speeds and power handling. Thus,

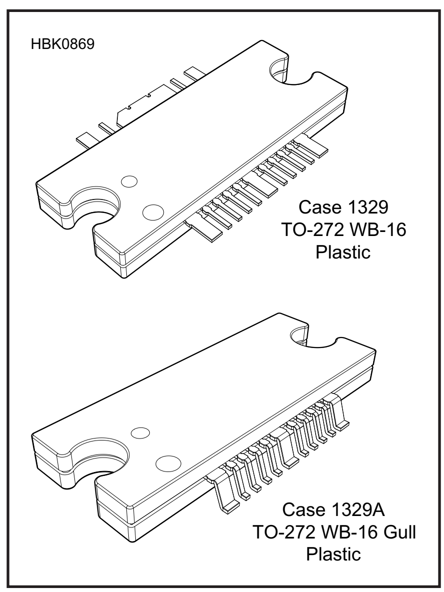


Figure 4.51 — The TO-272 is typical of packages used for hybrid IC RF amplifier modules at UHF and microwave frequencies.

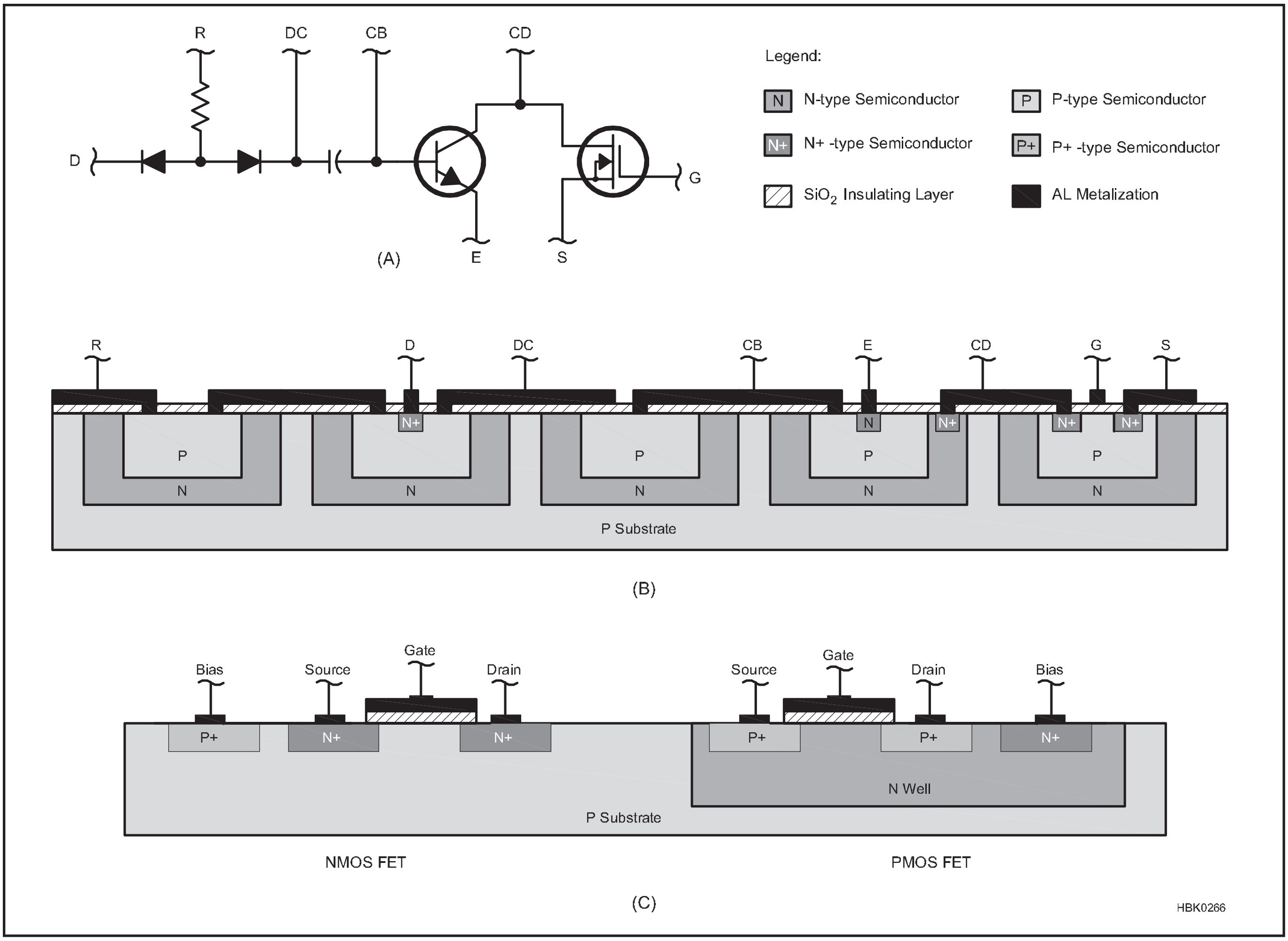


Figure 4.52— Integrated circuit construction. (A) Circuit containing two diodes, a resistor, a capacitor, an NPN transistor and an N-channel MOSFET. Labeled leads are D for diode, R for resistor, DC for diode-capacitor, E for emitter, S for source, CD for collector-drain and G for gate. (B) Integrated circuit that is identical to circuit in (A). Same leads are labeled for comparison. Circuit is built on a P-type semiconductor substrate with N-type wells diffused into it. An insulating layer of SiO_2 is above the semiconductor and is etched away where aluminum metal contacts are made with the semiconductor. Most metal-to-semiconductor contacts are made with heavily doped N-type material (N^+-type semiconductor).

P-type substrates are far more common. For devices with N-type substrates, all polarities in the ensuing discussion would be reversed. Many other types of substrates are employed in various special applications.

On top of the P-type substrate is a thin layer of N-type material in which the active and passive components are built. Impurities are diffused into this layer to form the appropriate component at each location. To prevent random diffusion of impurities into the N-layer, its upper surface must be protected. This is done by covering the N-layer with a layer of silicon dioxide (SiO_2). Wherever diffusion of impurities is desired, the SiO_2 is etched away. The precision of placing the components on the semiconductor material depends mainly on the fineness of the etching. The fourth layer of an IC is made of aluminum or copper and is used to make the interconnections between the components.

Different components are made in a single piece of semiconductor material by first diffusing a high concentration of acceptor impurities into the layer of N-type material. This process creates a P-type semiconductor — often referred to as a P^+-type semiconductor because of its high concentration of acceptor atoms — that isolates regions of N-type material. Each of these regions is then further processed to form single components.

A component is produced by the diffusion of a lesser concentration of acceptor atoms into the middle of each isolation region. This results in an N-type *isolation well* that contains P-type material, is surrounded on its sides by P^+-type material and has P-type material (substrate) below it. The cross-sectional view in Figure 4.52B illustrates the various layers. Connections to the metal layer are often made by diffusing high concentrations of donor atoms into small regions of the N-type well and the P-type material in the well. The material in these small regions is N^+-type and facilitates electron flow between the metal contact and the semiconductor. In some configurations, it is necessary to connect the metal directly to the P-type material in the well.

Fabricating Resistors and Capacitors

An isolation well can be made into a resistor by making two contacts into the P-type semiconductor in the well. Resistance is inversely proportional to the cross-sectional area of the well. An alternate type of resistor that can be integrated in a semiconductor circuit is a *thin-film resistor*, where a metallic film is deposited on the SiO_2 layer, masked on its upper surface by more SiO_2 and then etched to make the desired geometry, thus adjusting the resistance.

There are two ways to form capacitors in a semiconductor. One is to make use of the PN junction between the N-type well and the

P-type material that fills it. Much like a varactor diode, when this junction is reverse biased a capacitance is created. Since a bias voltage is required, this type of capacitor is polarized, like an electrolytic capacitor. Nonpolarized capacitors can also be formed in an integrated circuit by using thin film technology. In this case, a very high concentration of donor ions is diffused into the well, creating an N^+-type region. A thin metallic film is deposited over the SiO_2 layer covering the well and the capacitance is created between the metallic film and the well. The value of the capacitance is adjusted by varying the thickness of the SiO_2 layer and the cross-sectional size of the well. This type of thin film capacitor is also known as a metal oxide semiconductor (MOS) capacitor.

Unlike resistors and capacitors, it is very difficult to create inductors in integrated circuits. Generally, RF circuits that need inductance require external inductors to be connected to the IC. In some cases, particularly at lower frequencies, the behavior of an inductor can be mimicked by an amplifier circuit. In many cases the appropriate design of IC amplifiers can reduce or eliminate the need for external inductors. As the frequency of operation increases, however, the amount of inductance needed falls for an equivalent amount of reactance. This had led to several innovative techniques for creating small inductors using the metallization layers of the IC.

Fabricating Diodes and Transistors

The simplest form of diode is generated by connecting to an N^+-type connection point in the well for the cathode and to the P-type well material for the anode. Diodes are often converted from NPN transistor configurations. Integrated circuit diodes made this way can either short the collector to the base or leave the collector unconnected. The base contact is the anode and the emitter contact is the cathode.

Transistors are created in integrated circuitry in much the same way that they are fabricated in their discrete forms. The NPN transistor is the easiest to make since the wall of the well, made of N-type semiconductor, forms the collector, the P-type material in the well forms the base and a small region of N^+-type material formed in the center of the well becomes the emitter. A PNP transistor is made by diffusing donor ions into the P-type semiconductor in the well to make a pattern with P-type material in the center (emitter) surrounded by a ring of N-type material that connects all the way down to the well material (base), which is surrounded by another ring of P-type material (collector). This configuration results in a large base width separating the emitter and collector, causing these devices to have much lower current gain than the NPN form. This is one reason why integrated circuitry is designed to use many more NPN transistors than PNP transistors.

FETs can also be fabricated in IC form as shown in Figure 4.52C. Due to its many functional advantages, the MOSFET is the most common form used for digital ICs. MOSFETs are made in a semiconductor chip much the same way as MOS capacitors, described earlier. In addition to the signal processing advantages offered by MOSFETs over other transistors, the MOSFET device can be fabricated in 5% of the physical space required for bipolar transistors. CMOS ICs can contain 20 times more circuitry than bipolar ICs with the same chip size, making the devices more powerful and less expensive than those based on bipolar technology. CMOS is the most popular form of integrated circuit.

The final configuration of the switching circuit is CMOS as described in a previous section of this chapter. CMOS gates require two FETs, one of each form (NMOS and PMOS as shown in the figure). NMOS requires fewer processing steps, and the individual FETs have lower on-resistance than PMOS. The fabrication of NMOS FETs is the same as for individual semiconductors; P+ wells form the source and drain in a P-type substrate. A metal gate electrode is formed on top of an insulating SiO_2 layer so that the channel forms in the P-type substrate between the source and drain. For the PMOS FET, the process is similar but begins with an N-type well in the P-type substrate.

MOSFETs fabricated in this manner also have bias (B) terminals connected to the positive power supply to prevent destructive *latch-up*. This can occur in CMOS gates because the two MOSFETs form a *parasitic SCR*. If the SCR mode is triggered and both transistors conduct at the same time, large currents can flow through the FET and destroy the IC unless power is removed. Just as discrete MOSFETs are at risk of gate destruction, IC chips made with MOSFET devices have a similar risk. They should be treated with the same care to protect them from static electricity as discrete MOSFETs.

While CMOS is the most widely used technology, integrated circuits need not be made exclusively with MOSFETs or bipolar transistors. It is common to find IC chips designed with both technologies, taking advantage of the strengths of each.

4.6.8 Integrated Circuit Packaging

ICs come in a variety of packages, primarily dual and single in-line packages (DIPs and SIPs) and surface-mount packages. Most are marked with a part number and a four-digit manufacturer's date code indicating the year (first two digits) and week (last two digits) that the component was made. As mentioned in the introduction to this chapter, ICs are frequently house-marked and cross-reference directories can be helpful in identification and replacement.

IC part numbers provide a complete description of the device's function and ratings. For example, a 4066 IC contains four independent CMOS SPST switches. The 4066 is a CMOS device available from a number of different manufacturers in different package styles and ratings. The two- or three-letter prefix of the part number is generally associated with the part manufacturer. Next, the part type (4066 in this case) shows the function and pin assignments or "pin outs." Following the part type is an alphabetic suffix that describes the version of the part, package code, temperature range, reliability rating and possibly other information. For complete information on the part — any or all of which may be significant to circuit function — use the websites of the various manufacturers or do an online search for "data sheet" and the part number.

When choosing ICs that are not exact replacements, be wary of substituting "similar" devices, particularly in demanding applications, such as high-speed logic, sensitive receivers, precision instrumentation, and similar uses. In particular, substitution of one type of logic family for another — even if the device functions and *pin outs* (connections between the internal circuitry and external pins) are the same — can cause a circuit to not function or function erratically, particularly at temperature extremes. For example, substituting LS TTL devices for HCMOS devices will result in mismatches between logic level thresholds. Substituting a lower-power IC may result in problems supplying enough output current. Even using a faster or higher clock-speed part can cause problems if signals change faster or propagate more quickly than the circuit was designed for. Problems of this sort can be extremely difficult to troubleshoot unless you are skilled in circuit design. When necessary, you can add interface circuits or buffer amplifiers that improve the input and output capabilities of replacement ICs, but auxiliary circuits cannot improve basic device ratings, such as speed or bandwidth. Whenever possible, substitute ICs that are guaranteed or "direct" replacements and that are listed as such by the manufacturer.

4.6.9 Integrated Circuit Temperature Ranges

ICs are available in different operating temperature ranges. Four standard ranges are common:

- Commercial: 0 °C to 70 °C
- Industrial: –25 °C to 85 °C
- Automotive: –40 °C to 85 °C
- Military: –55 °C to 125 °C

In some cases, part numbers reflect the

temperature ratings. For example, an LM301A op amp is rated for the commercial temperature range, an LM201A op amp for the industrial range and an LM101A for the military range. It is usually acceptable, all other things being equal, to substitute ICs rated for a wider temperature range, but there are often other performance differences associated with the devices meeting wider temperature specifications that should be evaluated before making the substitution.

4.6.10 MMIC Amplifiers

Monolithic microwave integrated circuit (MMIC) amplifiers are single-supply 50-Ω wideband gain blocks offering high dynamic range for output powers to about +15 dBm. MMIC amplifiers are becoming increasingly popular in homebrew communications circuits. With bandwidths over 1 GHz, they are well suited for HF, VHF, UHF, and lower microwave frequencies.

MMIC amplifiers produce power gains from 10 dB to 30 dB. They also have a high third-order intercept point (IP3), usually in the +20 to +30 dBm range, easing the concerns about amplifier compression for most applications. They are used for RF and IF amplifiers, local oscillator amplifiers, transmitter drivers, and other medium power applications in 50-Ω systems. MMICs are especially well suited for driving 50-Ω double-balanced mixers (DBM). **Figure 4.53** shows the typical circuit arrangement for most MMIC amplifiers.

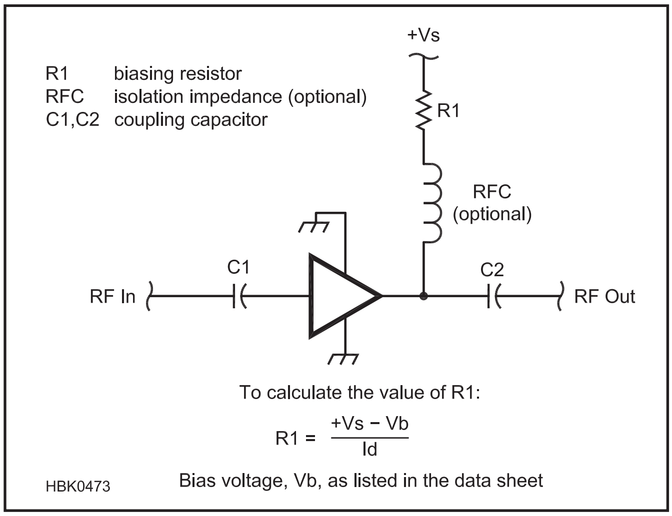


Figure 4.53 — MMIC application.

MMICs are available in a variety of surface mount packages requiring very few external components. (Package dimensions and outlines are provided in the **Construction Techniques** chapter on PCB layout.) Vendor data sheets and application notes, found on the manufacturer's websites, should be used for the proper selection of the biasing resistor, coupling capacitors, and other design criteria.

The main disadvantage of MMIC amplifiers is their relatively high current demands, usually in the 30 mA to 80 mA range per device, making them unsuitable for battery-powered portable equipment. On the other hand, the high current demand is what establishes their high gain and high IP3 characteristics with 50-Ω loads.

Their wide gain-bandwidth should be controlled by input and output tuned circuits or filters to reduce the gain outside the desired ranges. For example, for an HF amplifier, 30 MHz low-pass filters can be used to reduce the gain outside the HF spectrum, or a band-pass filter used for the frequency band of interest.

Selecting the proper MMIC amplifier is fairly straightforward. First, select a device for the desired frequency bandwidth, gain, and output power. Ensure device current is compatible with the design application. Calculate the value for the bias resistor (R1 in Figure 4.53) based on the device's biasing voltage (V_b) and whatever value of supply voltage (V_s) is available.

4.7 Amplifiers

By far, the most common type of analog circuit is the amplifier. The basic component of most electronics — the transistor — is an amplifier in which a small input signal controls a larger signal. Most of modern electronics, both analog and digital, are based on the transistor amplifier or switch, regardless of whether the input signal is amplified at the output.

4.7.1 Amplifier Configurations

Amplifier configurations are described by the *common* part of the device. The word "common" is used to describe the connection of a lead directly to a reference that is used by both the input and output ports of the circuit. The most common reference is ground, but positive and negative power sources are also valid references.

The type of circuit reference used depends on the type of device (transistor [NPN or PNP] or FET [P-channel or N-channel]), which lead is chosen as common, and the range of signal levels. Once a common lead is chosen, the other two leads are used for signal input and output. Based on the biasing conditions, there is only one way to select these leads. Thus, there are three possible amplifier configurations for each type of three-lead device. (Vacuum tube amplifiers are discussed in the chapter on **RF Power Amplifiers**.)

DC power sources are usually constructed so that ac signals at the output terminals are bypassed to ground through a very low impedance. This allows the power source to be treated as an *ac ground*, even though it may be supplying dc voltages to the circuit. When a circuit is being analyzed for its ac behavior, ac grounds are usually treated as ground, since dc bias is ignored in the ac analysis. Thus, a transistor's collector can be considered the "common" part of the circuit, even though in actual operation, a dc voltage is applied to it.

Figure 4.54 shows the three basic types of bipolar transistor amplifiers: the common-base, common-emitter, and common-collector. The common terminal is shown connected to ground, although as mentioned earlier, a dc bias voltage may be present. Each type of amplifier is described in the following sections. Following the description of the amplifier, additional discussion of biasing transistors and their operation at high frequencies and for large signals is presented.

4.7.2 Transistor Amplifiers

Creating a useful transistor amplifier depends on using an appropriate model for the transistor itself, choosing the right configuration of the amplifier, using the design equations for that configuration, and ensuring that the amplifier operates properly at different temperatures. This section follows that sequence, first introducing simple transistor

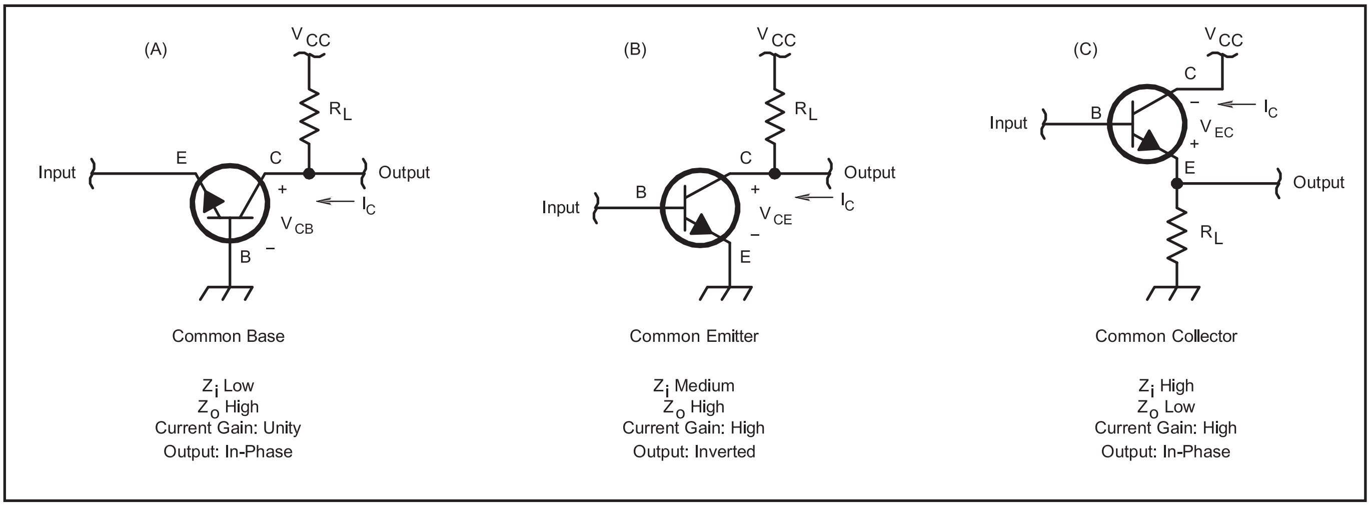


Figure 4.54 — The three configurations of bipolar transistor amplifiers with the basic characteristics of each circuit.

models and then extending that knowledge to the point of design guidelines for common circuits that use bipolar and FETs.

DEVICE MODELS AND CLASSES

Semiconductor circuit design is based on equivalent circuits that describe the physics of the devices. These circuits, made up of voltage and current sources and passive components such as resistors, capacitors and inductors, are called models. A complete model that describes a transistor exactly over a wide frequency range is a fairly complex circuit. As a result, simpler models are used in specific circumstances. For example, the *small-signal model* works well when the device is operated close to some nominal set of characteristics such that current and voltage interact fairly linearly. The *large-signal model* is used when the device is operated so that it enters its saturation or cut-off regions, for example.

Different frequency ranges also require different models. The *low-frequency models* used in this chapter can be used to develop circuits for dc, audio and very low RF applications. At higher frequencies, small capacitances and inductances that can be ignored at low frequencies begin to have significant effects on device behavior, such as gain or impedance. In addition, the physical structure of the device also becomes significant as gain begins to drop or phase shifts between input and output signals start to grow. In this region, *high-frequency models* are used.

Amplifiers are also grouped by their *operating class*, which describes the way in which the input signal is amplified. There are several classes of analog amplifiers; A, B, AB, AB1, AB2 and C.

The analog class designators specify over how much of the input cycle the active device is conducting current. A class-A amplifier's active device conducts current for 100 percent of the input signal cycle, such as shown in Figure 4.33. A class-B amplifier conducts during one-half of the input cycle, class-AB, AB1, and AB2 some fraction between 50 and 100 percent of the input cycle, and class-C for less than 50 percent of the input signal cycle.

Digital amplifiers, in which the active device is operated as a switch that is either fully-on or fully-off, similarly to switchmode power supplies, are also grouped by classes beginning with the letter D and beyond. Each different class uses a different method of converting the switch's output waveform to the desired RF waveform.

Amplifier classes, models, and their use at high-frequencies are discussed in more detail in the chapter on **RF Techniques**. In addition, the use of models for circuit simulation is discussed at length in the **Electronic Design Automation** chapter.

4.7.3 Bipolar Transistor Amplifiers

In this discussion, we will focus on simple models for bipolar transistors (BJTs). This discussion is centered on NPN BJTs but applies equally well to PNP BJTs if the bias voltage and current polarities are reversed. This section assumes the small-signal, low-frequency models for the transistors.

SMALL-SIGNAL BJT MODEL

The transistor is usually considered as a *current-controlled* device in which the base current controls the collector current:

$$I_c = \beta I_b$$

where

I_c = collector current,
I_b = base current, and
β = common-emitter current gain, beta.

(The term "common-emitter" refers to the type of transistor circuit described below, in which the transistor operates with base current as its input and collector circuit as its output.) Current is positive if it flows *into* a device terminal.

The transistor can also be treated as a voltage-controlled device in which the transistor's emitter current, I_e, is controlled by the base-emitter voltage, V_{be}:

$$I_c = I_{es}\left[e^{(qV_{be}/kT)} - 1\right] \approx I_{es}e^{(qV_{be}/kT)}$$

where

q = electronic charge,
k = Boltzmann's constant,
T = temperature in degrees Kelvin (K), and
I_{es} = emitter saturation current, typically 1×10^{-13} A.

The subscripts for voltages indicate the direction of positive voltage, so that V_{be} indicates positive is from the base to the emitter. It is simpler to design circuits using the current-controlled device, but accounting for the transistor's behavior with temperature requires an understanding of the voltage-controlled model.

Transistors are usually driven by both biasing and signal voltages. Both equations above for collector current apply to both transistor dc biasing and signal design. Both of these equations are approximations of the more complex behavior exhibited by actual transistors. The second equation applies to a simplification of the first *Ebers-Moll model* (see references). More sophisticated models for BJTs are described by Getreu (see references). Small-signal models treat only the signal components. We will consider bias later.

The next step is to use these basic equations to design circuits. We will begin with small-signal amplifier design and the limits of where the techniques can be applied. Later, we'll discuss large-signal amplifier design and the distortion that arises from operating the transistor in regions where the relationship between the input and output signals is nonlinear.

Common-Emitter Model

Figure 4.55 shows a BJT amplifier connected in the common-emitter configuration. (The emitter, shown connected to ground, is common to both the input circuit with the voltage source and the output circuit with the transistor's collector.) The performance of this circuit is adequately described by the simple equation $I_c = \beta I_b$. **Figure 4.56** shows the most common of all transistor small-signal models, a controlled current source with emitter resistance.

There are two variations of the model shown in the figure. Figure 4.56B shows the base as a direct connection to the junction of a current-controlled current source ($I_c = \beta I_b$) and a resistance, r_e, the *dynamic emitter resistance* representing the change in V_{be} with I_e. This resistance also changes with emitter current:

$$r_e = \frac{kT}{qI_c} \approx \frac{26}{I_e}$$

where I_e is the dc bias current in milliamperes.

The simplified approximation only applies at a typical ambient temperature of 300 K because r_e increases with temperature. In Figure 4.56A, the emitter resistance has been moved to the base connection, where it has the value $(\beta+1)r_e$. These models are electrically equivalent.

The transistor's output resistance (the Thevenin or Norton equivalent resistance between the collector and the grounded emitter) is infinite because of the current source. This is a good approximation for most silicon transistors at low frequencies (well below the transistor's gain-bandwidth product, F_T) and will be used for the design examples that follow.

As frequency increases, the capacitance inherent in BJT construction becomes significant and the *hybrid-pi model* shown in **Figure 4.57** is used, adding C_π in parallel with the input resistance. In this model the transfer parameter h_{ie} often represents the input impedance, shown here as a resistance at low frequencies.

THREE BASIC BJT AMPLIFIERS

Figure 4.58 shows a small-signal model applied to the three basic bipolar junction transistor (BJT) amplifier circuits: *common-emitter (*CE), *common-base* (CB) and *common-collector* (CC), more commonly known as the *emitter-follower* (EF). As defined earlier, the word "common" indicates that the referenced terminal is part of both the input and output circuits. The three transistor amplifier configurations are shown as simple circuits in Figure 4.44. The different transistor amplifier configurations have different gains, input and output impedances and phase relationships between the input and output signals.

In these simple models, transistors in both the CE and CB configurations have infinite output resistance because the collector current source is in series with the output current.

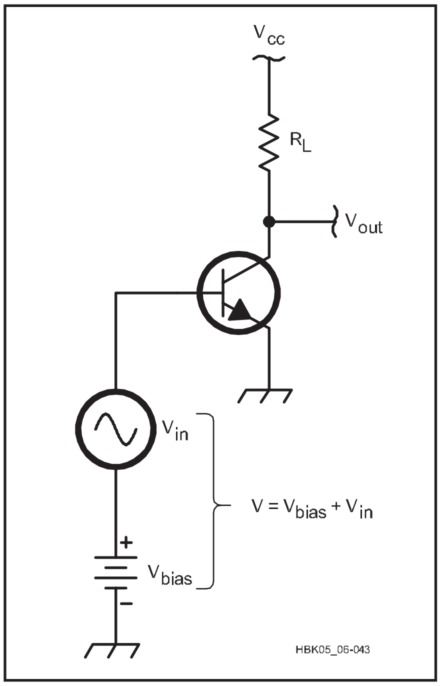


Figure 4.55 — Bipolar transistor with voltage bias and input signal.

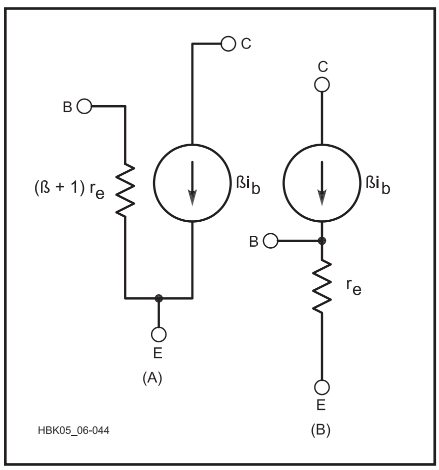


Figure 4.56 — Simplified low-frequency model for the bipolar transistor, a "beta generator with emitter resistance."
r_e = 26 / I_e (mA dc).

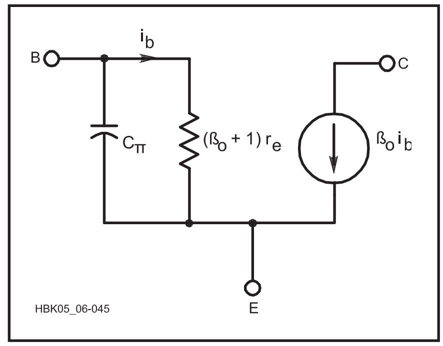


Figure 4.57 — The hybrid-pi model for the bipolar transistor.

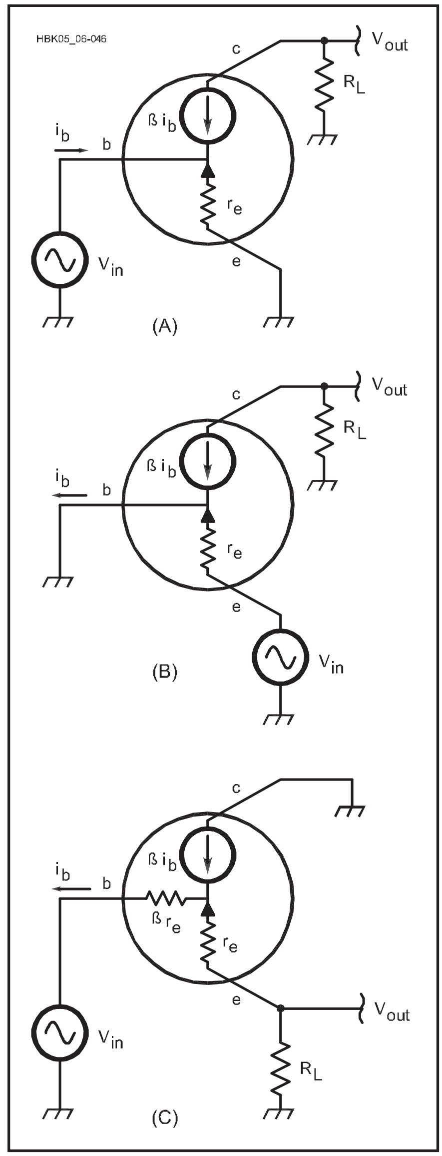


Figure 4.58 — Application of small-signal models for analysis of (A) the CE amplifier, (B) the CB and (C) the EF (CC) bipolar junction transistor amplifiers.

(The amplifier circuit's output impedance must include the effects of R_L.) The transistor connected in the EF configuration, on the other hand, has a finite output resistance because the current source is connected in parallel with the base circuit's equivalent resistance. Calculating the EF amplifier's output resistance requires including the input voltage source, V_s, and its impedance.

Examining the performance needs of the amplifier (engineers refer to these as the circuit's *performance requirements*) determines which of the three circuits is appropriate. Then, once the amplifier configuration is chosen, the equations that describe the circuit's behavior are used to turn the performance requirements into actual circuit component values.

This text presents design information for the CE amplifier in some detail, then summarizes designs for the CC and CB amplifiers. Detailed design analysis for all three amplifiers is described in the texts listed in the reference section for this chapter. All of the analysis in the following sections assume the small-signal, low-frequency model and ignore the effects of the coupling capacitors. High-frequency considerations are discussed in the **RF Techniques** chapter and some advanced discussion of biasing and large signal behavior of BJT amplifiers is available in the online content.

LOAD LINES AND Q-POINT

Transistors can operate with an infinite number of combinations of current (collector, emitter, and base) and voltage (collector-emitter, collector-base, or emitter-collector). Characteristic curves (see Figure 4.32 and the sections on Bipolar Junction and Field-Effect Transistors) show these combinations as a family of graphs in which one of the currents or voltages is varied in steps. A different set of curves can be drawn for each type of transistor and each type of amplifier circuit.

The particular combination at which the amplifier is operating is its *operating point*. The operating point is controlled by the selection of component values that make up the amplifier circuit so that it has the proper combination of gain, linearity and so forth. The result is that the operating point is restricted to a set of points that fall along a *load line*. The operating point with no input signal applied is the circuit's *quiescent point* or *Q-point*. As the input signal varies, the operating point moves along the load line, but returns to the Q-point when the input signal is removed.

More complete information on design using characteristic curves and load lines is available in texts: see the references for Banzhaf or Millman or the online example for BJT circuits at **www.zen22142.zen.co.uk/Design/bjtbias.htm**.

Figure 4.59 shows the load line and Q-point for an amplifier drawn on a transistor's set of characteristic curves for the CE amplifier circuit. The two end-points of the load line correspond to transistor saturation (I_{Csat} on the I_C current axis) and cutoff (V_{CC} on the V_{CE} voltage axis).

When a transistor is in saturation, further increases in base current do not cause a further increase in collector current. In the CE amplifier, this means that V_{CE} is very close to zero and I_C is at a maximum. In the circuit of Figure 4.44B, imagine a short circuit across the collector-to-emitter so that all of V_{CC} appears across R_L. Increasing base current will not result in any additional collector current. At cutoff, base current is so small that V_{CE} is at a maximum because no collector current is flowing and further reductions in base current cause no additional increase in V_{CE}.

In this simple circuit, $V_{CE} = V_{CC} - I_C R_L$ and the relationship between I_C and V_{CE} is a straight line between saturation and cutoff. This circuit's load line has a slope of $R_L = (V_{CC} - V_{CE}) / I_C$. No matter what value of base current is flowing in the transistor, the resulting combination of I_C and V_{CE} will be somewhere on the load line.

With no input signal to this simple circuit, the transistor is at cutoff where $I_C = 0$ and $V_{CE} = V_{CC}$. As the input signal increases so that base current gets larger, the operating point begins to move along the load line to the left, so that I_C increases and the voltage drop across the load, $I_C R_L$, increases, reducing V_{CE}. Eventually, the input signal will cause enough base current to flow that saturation is reached, where $V_{CE} \approx 0$ (typically 0.1 to 0.3 V for silicon transistors) and $I_C \approx V_{CC} / R_L$. If R_L is made smaller, the load line will become steeper and if R_L increases, the load line's slope is reduced.

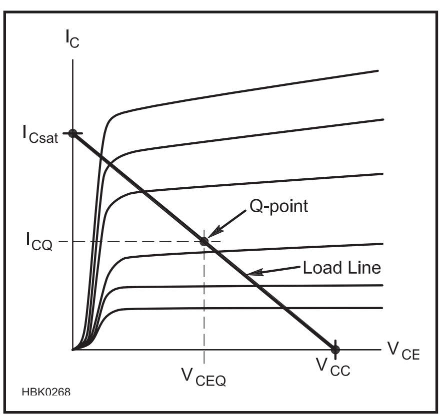


Figure 4.59 — A load line drawn on a transistor's characteristic curves. A circuit's load line shows all of the possible operating points with the specific component values chosen. If there is no input signal, the operating point is the quiescent or Q-point.

This simple circuit cannot reproduce negative input signals because the transistor is already in cutoff with no input signal. In addition, the shape and spacing of the characteristic curves show that the transistor responds nonlinearly when close to saturation and cutoff (the nonlinear regions) than it does in the middle of the curves (the linear or active region). Biasing is required so that the circuit does not operate in nonlinear regions, distorting the signal as shown in Figure 4.33.

If the circuit behaves differently for ac signals than for dc signals, a separate *ac load line* can be drawn as discussed below in the section "AC Performance" for the common-emitter amplifier. For example, in the preceding circuit, if R_L is replaced by a circuit that includes inductive or capacitive reactance, ac collector current will result in a different voltage drop across the circuit than will dc collector current. This causes the slope of the ac load line to be different than that of the dc load line.

The ac load line's slope will also vary with frequency, although it is generally treated as constant over the range of frequencies for which the circuit is designed to operate. The ac and dc load lines intersect at the circuit's Q-point because the circuit's ac and dc operation is the same if the ac input signal is zero.

COMMON-EMITTER AMPLIFIER

The *common-emitter amplifier* (*CE*) is the most common amplifier configuration of all — found in analog and digital circuits, from dc through microwaves, made of discrete components and fabricated in ICs. If you understand the CE amplifier, you've made a good start in electronics.

The CE amplifier is used when modest voltage gain is required along with an *input impedance* (the load presented to the circuit supplying the signal to be amplified) of a few hundred to a few kΩ. The current gain of the CE amplifier is the transistor's current gain, β.

The simplest practical CE amplifier circuit is shown in **Figure 4.60**. This circuit includes both coupling and biasing components. The

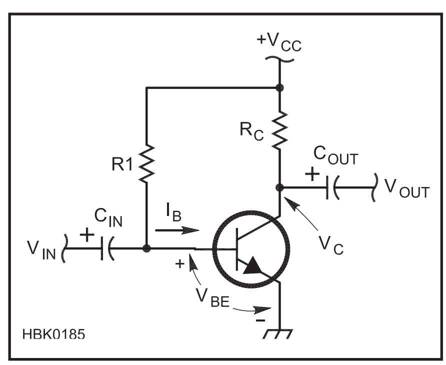


Figure 4.60 — Fixed-bias is the simplest common-emitter (CE) amplifier circuit.

capacitors at the input (C_{IN}) and output (C_{OUT}) block the flow of dc current to the load or to the circuit driving the amplifier. This is an ac-coupled design. These capacitors also cause the gain at very low frequencies to be reduced — gain at dc is zero, for example, because dc input current is blocked by C_{IN}. Resistor R_1 provides a path for bias current to flow into the base, offsetting the collector current from zero and establishing the Q-point for the circuit.

As the input signal swings positive, more current flows into the transistor's base through C_{IN}, causing more current to flow from the collector to emitter. This causes more voltage drop across R_L and so the voltage at the collector also drops. The reverse is true when the input signal swings negative. Thus, the output from the CE amplifier is inverted from its input.

Kirchoff's Voltage Law is used to analyze the circuit. We'll start with the collector circuit and treat the power supply as a voltage source.

$$V_{cc} = I_c R_c + V_{ce}$$

We can determine the circuit's voltage gain, A_V, from the variation in output voltage caused by variations in input voltage. The output voltage from the circuit at the transistor collector is

$$V_c = V_{CC} - I_c R_c = V_{CC} - \beta I_B R_C$$

It is also necessary to determine how base current varies with input voltage. Using the transistor's equivalent circuit of Figure 4.56A,

$$I_B = \frac{V_B}{(\beta + 1)\, r_e}$$

so that

$$V_c = V_{CC} - V_B \frac{\beta}{\beta + 1} \times \frac{R_C}{r_e}$$

We can now determine the circuit's *voltage gain*, the variation in output voltage, DV_C, due to variations in input voltage, DV_B. Since V_{CC} is constant and β is much greater than 1 in our model:

$$A_V \approx -\frac{R_C}{r_e}$$

Because r_e is quite small (typically a few ohms, see the equation for r_e in the section on the Common-Emitter Model), A_V for this circuit can be quite high.

The circuit load line's end-points are $V_{CE} = V_{CC}$ and $I_C = V_{CC} / R_C$. The circuit's Q-point is determined by the collector resistor, R_C, and resistor R_1 that causes bias current to flow into the base. To determine the Q-point, again use KVL starting at the power source and assuming that V_{BE} = 0.7 V for a silicon transistor's PN junction when forward-biased.

$$V_{CC} - I_B R_1 = V_B = V_{BE} = 0.7\text{ V}$$

so

$$I_B = \frac{V_{CC} - 0.7\text{V}}{R_1}$$

And the Q-point is therefore

$$V_{CEQ} = V_{CC} - \beta I_B R_C$$

and

$$I_{CQ} = \beta I_B$$

The actual V_{BE} of silicon transistors will vary from 0.6-0.75 V, depending on the level of base current, but 0.7 V is a good compromise value and widely used in small-signal, low-frequency design. Use 0.6 V for very low-power amplifiers and 0.75 V (or more) for high-current switch circuits.

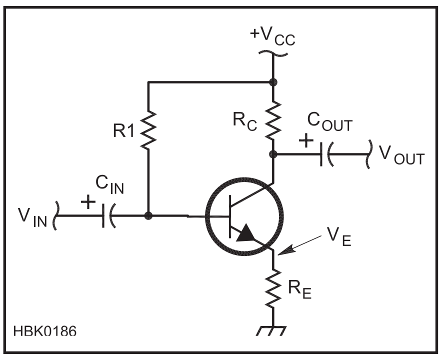


Figure 4.61 — Emitter degeneration. Adding R_E produces negative feedback to stabilize the bias point against changes due to temperature. As the bias current increases, the voltage drop across R_E also increases and causes a decrease in V_{BE}. This reduces bias current and stabilizes the operating point.

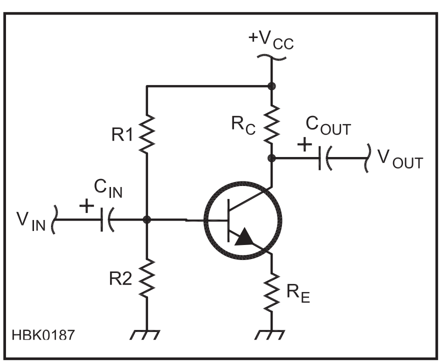


Figure 4.62 — Self-bias or self-emitter bias. R1 and R2 form a voltage divider to stabilize V_B and bias current. A good rule of thumb is for current flow through R1 and R2 to be 10 times the desired bias current. This stabilizes bias against changes in transistor parameters and component values.

This simple *fixed-bias* circuit is a good introduction to basic amplifiers, but is not entirely practical because the bias current will change due to the change of V_{BE} with temperature, leading to thermal instability. In addition, the high voltage gain can lead to instability due to positive feedback at high frequencies.

To stabilize the dc bias, **Figure 4.61** adds R_E, a technique called *emitter degeneration* because the extra emitter resistance creates negative feedback: as base current rises, so does V_E, the voltage drop across R_E. This reduces the base-emitter voltage and lowers base current. The benefit of emitter degeneration comes from stabilizing the circuit's dc behavior with temperature, but there is a reduction in gain because of the increased resistance in the emitter circuit. Ignoring the effect of R_L for the moment,

$$A_V \approx -\frac{R_C}{R_E}$$

In effect, the load resistor is now split between R_C and R_E, with part of the output voltage appearing across each because the changing current flows through both resistors. While somewhat lower than with the emitter connected directly to ground, voltage gain becomes easy to control because it is the ratio of two resistances.

Biasing the CE Amplifier

Figure 4.62 adds R_1 and R_2 from a voltage divider that controls bias current by fixing the base voltage at:

$$V_B = V_{CC} \frac{R_2}{R_1 + R_2}$$

Since

$$V_B = V_{BE} + (I_B + I_C)\, R_E = 0.7\text{ V} + (\beta + 1)\, I_B R_E$$

base current is

$$I_B = \frac{V_B - 0.7\text{ V}}{(\beta + 1)\, R_E}$$

and Q-point collector current becomes for high values of β

$$I_{CQ} = \beta I_B \approx \frac{V_{CC} \frac{R_2}{R_1 + R_2} - 0.7}{R_E}$$

This is referred to as *self-bias* or *self-emitter bias* in which the Q-point is much less sensitive to variations in temperature that affect β and V_{BE}.

A good rule-of-thumb for determining the sum of R_1 and R_2 is that the current flowing through the voltage divider, $V_{CC}/(R_1+R_2)$, should be at least 10 times the bias current, I_B. This keeps V_B relatively constant even with small changes in transistor parameters and temperature.

Q-point V_{CEQ} must now also account for the voltage drop across both R_C and R_E,

$$V_{CEQ} \approx V_{CC} - \beta I_B (R_C + R_E)$$

More sophisticated techniques for designing the bias networks of bipolar transistor circuits are described in reference texts listed at the end of this chapter.

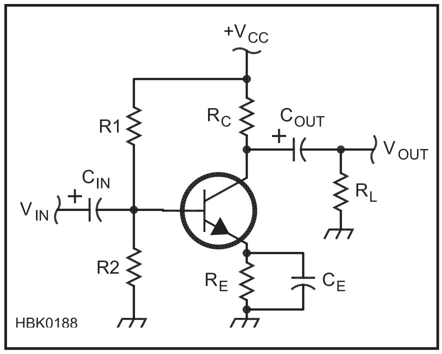


Figure 4.63 — Emitter bypass. Adding C_E allows ac currents to flow "around" R_E, returning ac gain to the value for the fixed-bias circuit while allowing R_E to stabilize the dc operating point.

Input and Output Impedance

With R_E in the circuit, the small changes in input current, I_B, when multiplied by the transistor's current gain, β, cause a large voltage change across R_E equal to $\beta I_B R_E$. This is the same voltage drop as if I_B were flowing through a resistance equal to bR_E. Thus, the effect of β on impedance at the base is to multiply the emitter resistance, R_E by β, as well. At the transistor's base,

$$Z_B \approx (\beta + 1) R_E$$

The input source doesn't just drive the base, of course, it also has to drive the combination of R1 and R2, the biasing resistors. From an ac point of view, both R1 and R2 can be considered as connected to ac ground, and they can be treated as if they were connected in parallel. When R1//R2 are considered along with the transistor base impedance, Z_B, the impedance presented to the input signal source is:

$$Z_{IN} = R1 // R2 // (\beta + 1) R_E$$

where // designates "in parallel with."

For both versions of the CE amplifier, the collector output impedance is high enough that

$$Z_{OUT} \approx R_C$$

CE Amplifier Design Example

The general process depends on the circuit's primary performance requirements, including voltage gain, impedances, power consumption, and so on. The most common situation in which a specific voltage gain is required and the circuit's Q-point has been selected based on the transistor to be used, and using the circuit of Figure 4.62, is as follows:

1) Start by determining the circuit's design constraints and assumptions: power supply V_{CC} = 12 V, transistor β = 150 and V_{BE} = 0.7 V. State the circuit's design requirements: $|A_V|$ = 5, Q-point of I_{CQ} = 4 mA and V_{CEQ} = 5 V. (A $V_{CEQ} \approx 1/2\ V_{CC}$ allows a wide swing in output voltage with the least distortion.)

2) Determine the values of R_C and R_E: $R_C + R_E = (V_{CC} - V_{CEQ})/I_{CQ}$ = 1.75 kΩ

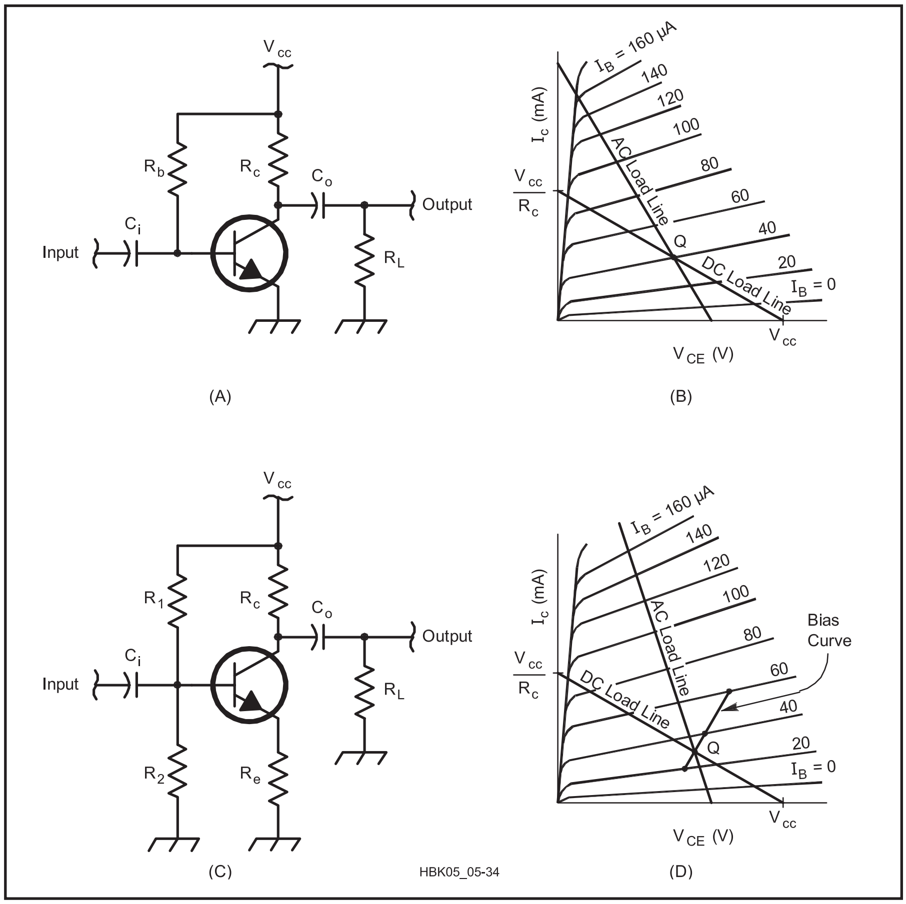


Figure 4.64 — Amplifier biasing and ac and dc load lines. (A) Fixed bias. Input signal is ac coupled through C_i. The output has a voltage that is equal to $V_{CC} - I_C \times R_C$. This signal is ac coupled to the load, R_L, through C_O. For dc signals, the entire output voltage is based on the value of R_C. For ac signals, the output voltage is based on the value of R_C in parallel with R_L. (B) Characteristic curve for the transistor amplifier pictured in (A). The slope of the dc load line is equal to $-1 / R_C$. For ac signals, the slope of the ac load line is equal to $-1 / (R_C // R_L)$. The quiescent-point, Q, is based on the base bias current with no input signal applied and the point where this characteristic line crosses the dc load line. The ac load line must also pass through point Q. (C) Self-bias. Similar to fixed bias circuit with the base bias resistor split into two: R1 connected to V_{CC} and R2 connected to ground. Also an emitter bias resistor, R_E, is included to compensate for changing device characteristics. (D) This is similar to the characteristic curve plotted in (B) but with an additional "bias curve" that shows how the base bias current varies as the device characteristics change with temperature. The operating point, Q, moves along this line and the load lines continue to intersect it as it changes. If C_E was added as in Figure 4.63, the slope of the ac load line would increase further.

3) $A_V = -5$, so $R_C = 5\ R_E$, thus $6R_E = 1.75$ kΩ and $R_E = 270\ \Omega$

4) Determine the base bias current, $I_B = I_{CQ}/\beta = 27\ \mu A$. By the rule of thumb, current through R_1 and $R_2 = 10\ I_B = 270\ \mu A$

5) Find the voltage across $R_2 = V_B = V_{BE} + I_C R_E = 0.7 + 4$ mA (0.27 kΩ) = 1.8 V. Thus, $R_2 = 1.8$ V / 270 μA = 6.7 kΩ

6) The voltage across $R_1 = V_{CC} - V_{R2} = 12 - 1.8 = 10.2$ V and $R_1 = 10.2$ V / 270 μA = 37.8 kΩ

Use the nearest standard values (R_E = 270 Ω, R_1 = 39 kΩ, R_2 = 6.8 kΩ) and circuit behavior will be close to predicted performance.

AC Performance

To achieve high gains for ac signals while maintaining dc bias stability, the *emitter-bypass capacitor*, C_E, is added in **Figure 4.63** to provide a low impedance path for ac signals around R_E. In addition, a more accurate formula for ac gain includes the effect of adding R_L through the dc blocking capacitor at the collector. In this circuit, the ac voltage gain is

$$A_V \approx -\frac{R_C\ //\ R_L}{r_e}$$

Because of the different signal paths for ac and dc signals, the ac performance of the circuit is different than its dc performance. This is illustrated in **Figure 4.64** by the intersecting load lines labeled "AC Load Line" and "DC Load Line." The load lines intersect at the Q-point because at that point dc performance is the same as ac performance if no ac signal is present.

The equation for ac voltage gain assumes that the reactances of C_{IN}, C_{OUT}, and C_E are small enough to be neglected (less than one-tenth that of the components to which they are connected at the frequency of interest). At low frequencies, where the capacitor reactances become increasingly large, voltage gain is reduced. Neglecting C_{IN} and C_{OUT}, the low-frequency 3 dB point of the amplifier, f_L, occurs where $X_{CE} = 0.414\ r_e$,

$$f_L = \frac{2.42}{2\pi r_e C_E}$$

This increases the emitter circuit impedance such that A_V is lowered to 0.707 of its mid-band value, lowering gain by 3 dB. (This ignores the effects of C_{IN} and C_{OUT}, which will also affect the low-frequency performance of the circuit.)

The ac input impedance of this version of the CE amplifier is lower because the effect of R_E on ac signals is removed by the bypass capacitor. This leaves only the internal emitter resistance, r_e, to be multiplied by the current gain,

$$Z_{IN} \approx R1\ //\ R2\ //\ \beta r_e$$

and

$$Z_{OUT} \approx R_C$$

again neglecting the reactance of the three capacitors.

The power gain, A_P, for the emitter-bypassed CE amplifier is the ratio of output power, V_O^2/Z_{OUT}, to input power, V_I^2/Z_{IN}. Since $V_O = V_I A_V$,

$$A_P = A_V^2\ \frac{R1\ //\ R2\ //\ \beta r_e}{R_C}$$

COMMON-COLLECTOR (EMITTER-FOLLOWER) AMPLIFIER

The common-collector (CC) amplifier in **Figure 4.65** is also known as the *emitter-follower* (EF) because the emitter voltage "follows" the input voltage. In fact, the amplifier has no voltage gain (voltage gain ≈ 1) but is used as a buffer amplifier to isolate sensitive circuits such as oscillators or to drive low-impedance loads, such as coaxial cables. As in the CE amplifier, the current gain of the emitter-follower is the transistor's current gain, β. It has relatively high input impedance with low output impedance and good power gain.

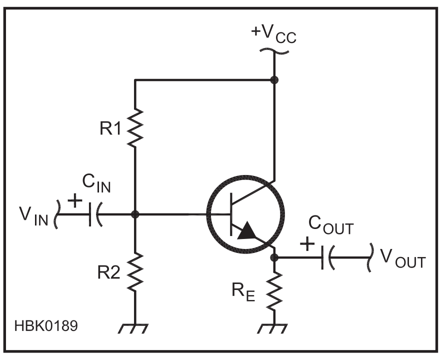


Figure 4.65 — Emitter follower (EF) amplifier. The voltage gain of the EF amplifier is unity. The amplifier has high input impedance and low output impedance, making it a good choice for use as a buffer amplifier.

The collector of the transistor is connected directly to the power supply without a resistor and the output signal is created by the voltage drop across the emitter resistor. There is no 180° phase shift as seen in the CE amplifier; the output voltage follows the input signal with 0° phase shift because increases in the input signal cause increases in emitter current and the voltage drop across the emitter resistor.

The EF amplifier has high input impedance: following the same reasoning as for the CE amplifier with an unbypassed emitter resistor,

$$Z_{IN} = R1\ //\ R2\ //\ (\beta + 1)\ R_E$$

The impedance at the EF amplifier's output consists of the emitter resistance, R_E, in parallel with the series combination of the internal emitter resistance, r_e, the parallel combination of biasing resistors R1 and R2, and the internal impedance of the source providing the input signal. In this case, current gain acts to *reduce* the effect of the input circuit's impedance on output impedance:

$$Z_{OUT} = \left[\frac{R_S\ //\ R1\ //\ R2}{(\beta + 1)}\right]//\ R_E$$

In practice, with transistor β of 100 or more, $Z_{OUT} \approx R_S/\beta$. However, if a very high impedance source is used, such as a crystal microphone element or photodetector, the effects of the biasing and emitter resistors must be considered.

Because the voltage gain of the EF amplifier is unity, the power gain is simply the ratio of input impedance to output impedance,

$$A_P \approx \frac{R1\ //\ R2\ //\ (\beta + 1)\ R_E}{R_E}$$

EF Amplifier Design Example

The following procedure is similar to the design procedure in the preceding section for the CE amplifier, except $A_V = 1$.

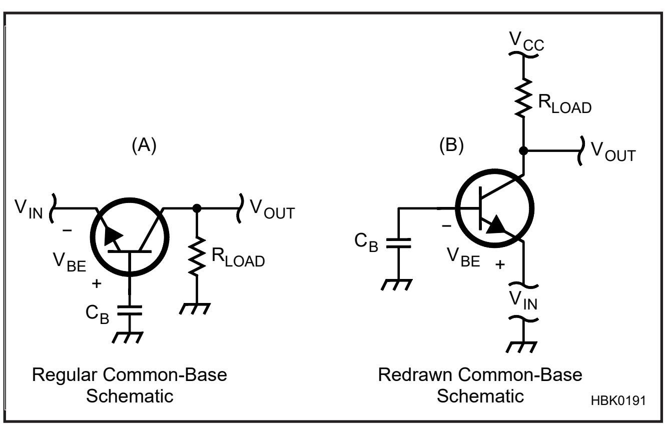


Figure 4.66 — The common-base (CB) amplifier is often drawn in an unfamiliar style (A), but is more easily understood when drawn similarly to the CE and EF amplifiers (B). The input signal to the CB amplifier is applied to the emitter instead of the base.

1) Start by determining the circuit's design constraints and assumptions: V_{cc} = 12 V (the power supply voltage), a transistor's β of 150 and V_{BE} = 0.7 V. State the circuit's design requirements: Q-point of I_{CQ} = 5 mA and V_{CEQ} = 6 V.

2) $R_E = (V_{CC} - V_{CEQ})/I_{CQ}$ = 1.2 kΩ

3) Base current, $I_B = I_{CQ}/\beta$ = 33 μA

4) Current through R1 and R2 = 10 I_B = 330 μA (10 I_B rule of thumb as with the CE amplifier)

5) Voltage across R2 = $V_{BE} + I_C R_E$ = 0.7 + 5 mA (1.2 kΩ) = 6.7 V and R2 = 6.7 V / 330 μA = 20.3 kΩ (use the standard value 22 kΩ)

6) Voltage across R1 = V_{CC} – 6.7 V = 5.3 V

7) R1 = 5.3 V / 330 μA = 16.1 kΩ (use 16 kΩ)

8) Z_{IN} = R1 // R2 // $R_E(\beta + 1)$ ≈ 8.5 kΩ

COMMON-BASE AMPLIFIER

The common-base (CB) amplifier of **Figure 4.66** is used where low input impedance is needed, such as for a receiver preamp with a coaxial feed line as the input signal source. Complementary to the EF amplifier, the CB amplifier has unity current gain and high output impedance.

Figure 4.66A shows the CB circuit as it is usually drawn, without the bias circuit resistors connected and with the transistor symbol turned on its side from the usual orientation so that the emitter faces the input. In order to better understand the amplifier's function, Figure 4.66B reorients the circuit in a more familiar style. We can now clearly see that the input has just moved from the base circuit to the emitter circuit.

Placing the input in the emitter circuit allows it to cause changes in the base-emitter current as for the CE and EF amplifiers, except that for the CB amplifier a positive change in input amplitude reduces base current by lowering V_{BE} and raising V_C. As a result, the CB amplifier is noninverting, just like the EF, with output and input signals in-phase.

A practical circuit for the CB amplifier is shown in **Figure 4.67**. From a dc point of view (replace the capacitors with open circuits), all of the same resistors are there as in the CE amplifier. The input capacitor, C_{IN}, allows the dc emitter current to bypass the ac input signal source and C_B places the base at ac ground while allowing a dc voltage for biasing. (All voltages and currents are labeled to aid in understanding the different orientation of the circuit.)

The CB amplifier's current gain,

$$A_I = \frac{i_C}{i_E} = \frac{\beta}{\beta + 1}$$

is relatively independent of input and output impedance, providing excellent isolation between the input and output circuits. Output impedance does not affect input impedance, allowing the CB amplifier to maintain stable input impedance, even with a changing load.

Following reasoning similar to that for the CE and EF amplifiers for the effect of current gain on R_E, we find that input impedance for the CB amp is

$$Z_{IN} = R_E \,//\, (\beta + 1)\, r_e$$

With high-gain transistors having a β > 100, for typical values of r_e (about 1 kΩ) the input impedance is approximately R_E. If R_E is chosen to be 50 Ω, the result will be a good input impedance match to 50 Ω feed lines and signal sources.

The output impedance for the CB amplifier is approximately

$$Z_{OUT} = R_C \,//\, \frac{1}{h_{oe}} \approx R_C$$

where h_{oe} is the transistor's collector output admittance. The reciprocal of h_{oe} is in the range of 100 kΩ at low frequencies.

Voltage gain for the CB amplifier is

$$A_V \approx \frac{R_C \,//\, R_L}{r_e}$$

As a result, the usual function of the CB amplifier is to convert input current from a low-impedance source into output voltage at a higher impedance.

Power gain for the CB amplifier is approximately the ratio of output to input impedance,

$$A_P \approx \frac{R_C}{R_E \,//\, (\beta + 1)\, r_e}$$

CB Amplifier Design Example

Because of its usual function as a current-to-voltage converter, the design process for the CB amplifier begins with selecting R_E and A_V, assuming that R_L is known.

1) Start by determining the circuit's design constraints and assumptions: V_{cc} = 12 V (the power supply voltage), a transistor's β of 150 and V_{BE} = 0.7 V. State the circuit's design requirements: R_E = 50 Ω, R_L = 1 kΩ, I_{CQ} = 5 mA, V_{CEQ} = 6 V.

2) Base current, $I_B = I_{CQ}/\beta$ = 33 μA

3) Current through R1 and R2 = 10 I_B = 330 μA (10 × I_B rule of thumb as with the CE amplifier)

4) Voltage across R2 = $V_{BE} + I_C R_E$ = 0.7 + 5 mA (50 Ω) = 0.95 V and R2 = 0.95 V / 330 μA = 2.87 kΩ (use the standard value 2.7 kΩ)

5) Voltage across R1 = V_{CC} – 0.95 V = 11.05 V

6) R1 = 11.05 V / 330 μA = 33.5 kΩ (use 33 kΩ)

7) $R_C = (V_{CC} - I_{CQ} R_E - V_{CEQ}) / I_{CQ}$ = (12 – 0.25 – 6) / 5 mA = 1.15 kΩ (use 1.2 kΩ)

8) A_V = (1.2 kΩ // 1 kΩ) / (26 mV / I_E) = 105

4.7.4 FET Amplifiers

The field-effect transistor (FET) is widely used in radio and RF applications. There are many types of FETs, with JFETs (junction FET) and MOSFETs (metal-oxide-semiconductor FET) being the most common types. In this section we will discuss JFETs, with the understanding that the use of MOSFETs is similar. (This discussion is based on N-channel JFETs, but the same discussion applies to P-channel devices if the bias voltages and currents are reversed.)

SMALL-SIGNAL FET MODEL

While bipolar transistors are most commonly viewed as current-controlled devices, the JFET, however, is purely a voltage-controlled device — at least at low frequencies. The input gate is treated as a reverse-biased diode junction with virtually no current flow. As with the bipolar transistor amplifier circuits, the circuits in this section are very basic and more thorough treatments of FET amplifier design can be found in the references at

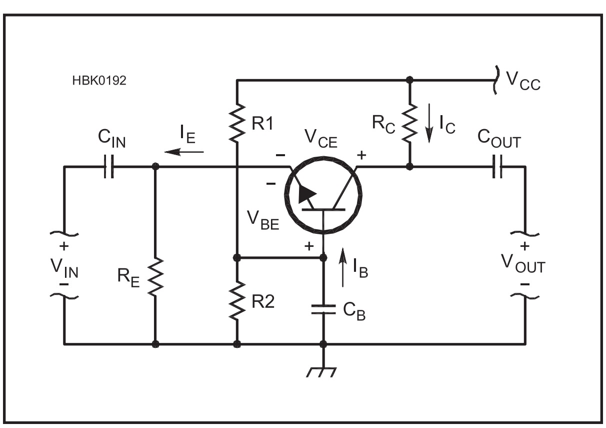


Figure 4.67 — A practical common-base (CB) amplifier. The current gain of the CB amplifier is unity. It has low input impedance and high output impedance, resulting in high voltage gain. The CB amplifier is used to amplify signals from low-impedance sources, such as coaxial cables.

the end of the chapter.

The operation of an N-channel JFET for both biasing and signal amplification can be characterized by the following equation:

$$I_D = I_{DSS}\left(\frac{1 - V_{SG}}{V_P}\right)^2$$

where

I_{DSS} = drain saturation current
V_{GS} = the gate-source voltage
V_P = the pinch-off voltage.

I_{DSS} is the maximum current that will flow between the drain and source for a given value of drain-to-source voltage, V_{DS}. Note that the FET is a *square-law* device in which output current is proportional to the square of an input voltage. (The bipolar transistor's output current is an exponential function of input current.)

Also note that V_{GS} in this equation has the opposite sense of the bipolar transistor's V_{BE}. For this device, as V_{GS} increases (making the source more positive than the gate), drain current decreases until at V_P the channel is completely "pinched-off" and no drain current flows at all. This equation applies only if V_{GS} is between 0 and V_P. JFETs are seldom used with the gate-to-channel diode forward-biased ($V_{GS} < 0$).

None of the terms in this equation depend explicitly on temperature. Thus, the FET is relatively free of the thermal instability exhibited by the bipolar transistor. As temperature increases, the overall effect on the JFET is to reduce drain current and to stabilize the operation.

The small-signal model used for the JFET is shown in **Figure 4.68**. The drain-source channel is treated as a current source whose output is controlled by the gate-to-source voltage so that $I_D = g_m V_{GS}$. The high input impedance allows the input to be modeled as an open circuit (at low frequencies). This simplifies circuit modeling considerably as biasing of the FET gate can be done by a simple voltage divider without having to consider the effects of bias current flowing in the JFET itself.

The FET has characteristic curves as shown in Figure 4.43 that are similar to those of a bipolar transistor. The output characteristic curves are similar to those of the bipolar transistor, with the horizontal axis showing V_{DS} instead of V_{CE} and the vertical axis showing I_D instead of I_C. Load lines, both ac and dc, can be developed and drawn on the output characteristic curves in the same way as for bipolar transistors.

The set of characteristic curves in Figure 4.43 are called *transconductance response curves* and they show the relationship between input voltage (V_{GS}), output current (I_D) and output voltage (V_{DS}). The output characteristic curves show I_D and V_{DS} for different values of V_{GS} and are similar to the BJT output characteristic curve. The input characteristic curves show I_D versus V_{GS} for different values of V_{DS}.

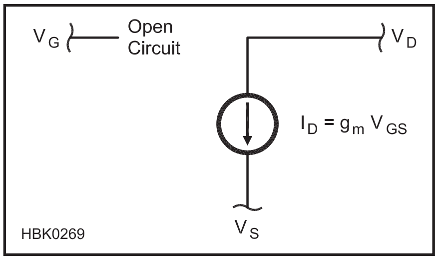


Figure 4.68 — Small-signal FET model. The FET can be modeled as a voltage-controlled current source in its saturation region. The gate is treated as an open-circuit due to the reverse-biased gate-channel junction.

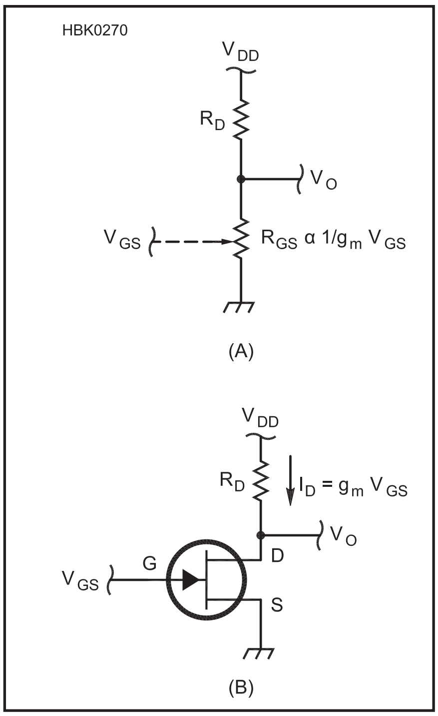


Figure 4.69 — In the ohmic region (A), the FET acts like a variable resistance, R_{DS}, with a value controlled by V_{GS}. The alpha symbol (α) means "is proportional to." In the saturation region (B), the drain-source channel of the FET can be treated like a current source with $I_D = g_m V_{GS}$.

MOSFETs act in much the same way as JFETs when used in an amplifier. They have a higher input impedance, due to the insulation between the gate and channel. The insulated gate also means that they can be operated with the polarity of V_{GS} such that a JFET's gate-channel junction would be forward biased, beyond V_P. Refer to the discussion of depletion- and enhancement-mode MOSFETs in the previous section on Practical Semiconductors.

THREE BASIC FET AMPLIFIERS

Just as for bipolar transistor amplifiers, there are three basic configurations of amplifiers using FETs: the *common-source* (CS) (corresponding to the common-emitter), *common-drain* (CD) or *source-follower* (corresponding to the emitter-follower) and the *common-gate* (CG) (corresponding to the common base). Simple circuits and design methods are presented here for each, assuming low-frequency operation and a simple, voltage-controlled current-source model for the FET. Discussion of the FET amplifier at high frequencies is available in the **RF Techniques** chapter and an advanced discussion of biasing FET amplifiers and their large-signal behavior is contained in this book's online information.

COMMON-SOURCE AMPLIFIER

The basic circuit for a common-source FET amplifier is shown in **Figure 4.69**. In the ohmic region (see the previous discussion on FET characteristics), the FET can be treated as a variable resistance as shown in Figure 4.69A where V_{GS} effectively varies the resistance between drain and source. However, most FET amplifiers are designed to operate in the saturation region and the model of Figure 4.68 is used in the circuit of Figure 4.69B in which,

$$I_D = g_m V_{GS}$$

where g_m is the FET's forward transconductance.

If V_O is measured at the drain terminal (just as the common-emitter output voltage is measured at the collector), then

$$\Delta V_O = -g_m \Delta V_{GS} R_D$$

The minus sign results from the output voltage decreasing as drain current and the voltage drop across R_D increases, just as in the CE amplifier. Like the CE amplifier, the input and output voltages are thus 180° out of phase. Voltage gain of the CS amplifier in terms of transconductance and the drain resistance is:

$$A_V = -g_m R_D$$

As long as $V_{GS} < 0$, this simple CS amplifier's input impedance at low frequencies is that of a reverse-biased diode — nearly infinite with a very small leakage current. Output impedance of the CS amplifier is approximately R_D because the FET drain-to-source channel acts like a current source with very high impedance.

$$Z_{IN} = \infty \text{ and } Z_{OUT} \approx R_D$$

As with the BJT, biasing is required to

create a Q-point for the amplifier that allows reproduction of ac signals. The practical circuit of Figure 4.69B is used to allow control of V_{GS} bias. A load line is drawn on the JFET output characteristic curves, just as for a bipolar transistor circuit. One end point of the load line is at $V_{DS} = V_{DD}$ and the other at $I_{DS} = V_{DD} / R_D$. The Q-point for the CS amplifier at I_{DQ} and V_{DSQ} is thus determined by the dc value of V_{GS}.

The practical JFET CS amplifier shown in **Figure 4.70** uses self-biasing in which the voltage developed across the source resistor, R_S, raises V_S above ground by $I_D R_S$ volts and $V_{GS} = -I_D R_S$ since there is no dc drop across R_G. This is also called *source degeneration*. The presence of R_S changes the equation of voltage gain to

$$A_V = -\frac{g_m R_D}{1 + g_m R_S} \approx -\frac{R_D}{R_S}$$

The value of R_S is obtained by substituting $V_{GS} = I_D R_S$ into the small-signal model's equation for I_D and solving for R_S as follows:

$$R_S = \frac{-V_P}{I_{DQ}} \left(1 - \sqrt{\frac{I_{DQ}}{I_{DSS}}}\right)$$

Once R_S is known, the equation for voltage gain can be used to find R_D.

The input impedance for the circuit of Figure 4.60 is essentially R_G. Since the gate of the JFET is often ac coupled to the input source through a dc blocking capacitor, C_{IN}, a value of 100 kΩ to 1 MΩ is often used for R_G to provide a path to ground for gate leakage current. If R_G is omitted in an ac-coupled JFET amplifier, a dc voltage can build up on the gate from leakage current or static electricity, affecting the channel conductivity.

$$Z_{IN} = R_G$$

Because of the high impedance of the drain-source channel in the saturation region, the output impedance of the circuit is:

$$Z_{OUT} \approx R_D$$

Designing the Common-Source Amplifier

The design of the CS amplifier begins with selection of a Q-point $I_{DQ} < I_{DSS}$. Because of variations in V_P and I_{DSS} from JFET to JFET, it may be necessary to select devices individually to obtain the desired performance.

1) Start by determining the circuit's design constraints and assumptions: V_{DD} = 12 V (the power supply voltage) and the JFET has an I_{DSS} of 35 mA and a V_P of –3.0V, typical of small-signal JFETs. State the circuit's design requirements: $|A_V|$ = 10 and I_{DQ} = 10 mA.

2) Use the equation above to find R_S = 139 Ω.

3) Since $|A_V|$ = 10, $R_D = 10\ R_S$ = 1390 Ω.

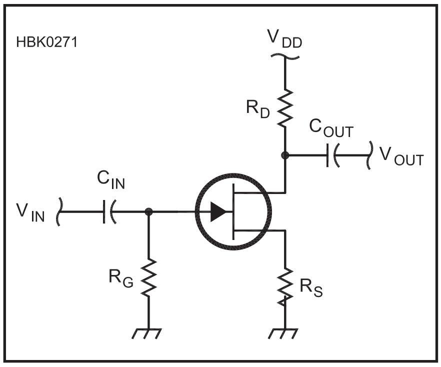


Figure 4.70 — Common-source (CS) amplifier with self-bias.

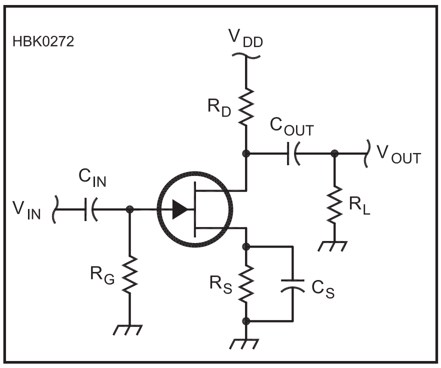


Figure 4.71 — Common-source amplifier with source bypass capacitor, C_S, to increase voltage gain without affecting the circuit's dc performance.

Use standard values for R_S = 150 Ω and R_D = 1.5 kΩ.

AC Performance

As with the CE bipolar transistor amplifier, a bypass capacitor can be used to increase ac gain while leaving dc bias conditions unchanged as shown in **Figure 4.71**. In the case of the CS amplifier, a *source bypass* capacitor is placed across R_S and the load, R_L, connected through a dc blocking capacitor. In this circuit voltage gain becomes:

$$A_V = -g_m\ (R_D\ //\ R_L)$$

Assuming C_{IN} and C_{OUT} are large enough to ignore their effects, the low-frequency cutoff frequency of the amplifier, f_L, is approximately where $X_{CS} = 0.707\ (R_D // R_L)$,

$$f_L = \frac{1.414}{2\pi\ (R_D\ //\ R_L)\ C_S}$$

as this reduces A_V to 0.707 of its mid-band value, resulting in a 3 dB drop in output amplitude.

The low-frequency ac input and output impedances of the CS amplifier remain

$$Z_{IN} = R_G \text{ and } Z_{OUT} \approx R_D$$

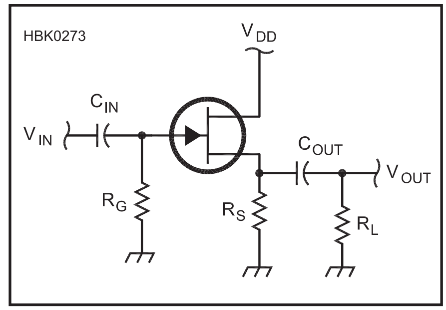


Figure 4.72 — Similar to the EF amplifier, the common-drain (CD) amplifier has a voltage gain of unity, but makes a good buffer with high input and low output impedances.

COMMON-DRAIN (SOURCE-FOLLOWER) AMPLIFIER

The *common-drain* amplifier in **Figure 4.72** is also known as a *source-follower* (SF) because the voltage gain of the amplifier is unity, similar to the emitter follower (EF) bipolar transistor amplifier. The SF amplifier is used primarily as a buffer stage and to drive low-impedance loads.

At low frequencies, the input impedance of the SF amplifier remains nearly infinite. The SF amplifier's output impedance is the source resistance, R_S, in parallel with the impedance of the controlled current source, $1/g_m$.

$$Z_{OUT} = R_S\ //\ \frac{1}{g_m}$$

$$= \frac{R_S}{g_m R_S + 1} \approx \frac{1}{g_m} \text{ for } g_m R_S \gg 1$$

Design of the SF amplifier follows essentially the same process as the CS amplifier, with $R_D = 0$.

THE COMMON-GATE AMPLIFIER

The *common-gate* amplifier in **Figure 4.73** has similar properties to the bipolar transistor common-base (CB) amplifier: unity current gain, high voltage gain, low input impedance and high output impedance. (Refer to the discussion of the CB amplifier regarding placement of the input and how the circuit schematic is drawn.) It is used as a voltage amplifier, particularly for low-impedance sources, such as coaxial cable inputs.

The CG amplifier's voltage gain is

$$A_V = g_m (R_D\ //\ R_L)$$

Since the output impedance of the CG amplifier is very high, we must take into account the output resistance of the controlled current source, r_o. This is analogous to the appearance of h_{oe} in the equation for output impedance of the bipolar transistor CG amplifier.

$$Z_O \approx r_o\ (g_m R_S + 1)\ //\ R_D$$

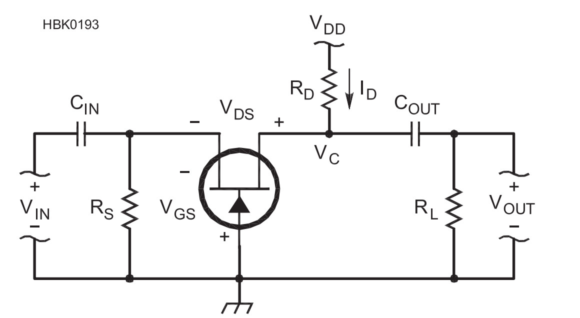


Figure 4.73 — FET common-gate (CG) amplifiers are often used as preamplifiers because of their high voltage gain and low input impedance. With the proper choice of transistor and quiescent-point current, the input impedance can match coaxial cable impedances directly.

The CG amplifier input impedance is approximately:

$$Z_I = R_S \,//\, \frac{1}{g_m}$$

Because the input impedance is quite low, the cascode circuit described later in the section on buffers is often used to present a higher-impedance input to the signal source.

Occasionally, the value of R_S must be fixed in order to provide a specific value of input impedance. Solving the small-signal FET model equation for I_{DQ} results in the following equation:

$$I_{DQ} = \frac{V_P}{2R_S^{\,2} I_{DSS}}\left(V_P + \sqrt{V_P^{\,2} - 4R_S I_{DSS} V_P}\right) - \frac{V_P}{R_S}$$

Designing the Common-Gate Amplifier

Follow the procedure for designing a CS amplifier, except determine the value of R_D as shown in the equation above for voltage gain, A_V.

1) Start by determining the circuit's design constraints and assumptions: V_{DD} = 12 V (the power supply voltage) and the JFET has a g_m of 15 mA/V, an I_{DSS} of 60 mA and V_P = –6 V. State the circuit's design requirements: A_V = 10, R_L = 1 kΩ and R_S = 50 Ω.

2) Solve the A_V equation for R_D: 10 = 0.015 × $R_D//R_L$, so $R_D//R_L$ = 667 Ω. R_D = 667 R_L / (R_L – 667) = 2 kΩ.

3) Use the equation above to find I_{DQ} = 10 mA. If I_{DQ} places the Q-point in the ohmic region, reduce A_V and repeat the calculations.

4.7.5 Buffer Amplifiers

Figure 4.74 shows common forms of buffers with low-impedance outputs: the *emitter follower* using a bipolar transistor, the *source follower* using a field-effect transistor and the *voltage follower* using an operational amplifier. (The operational amplifier is discussed later in this chapter.) These circuits are called "followers" because the output "follows" the input very closely with approximately the same voltage and little phase shift between the input and output signals.

4.7.6 Cascaded Buffers

THE DARLINGTON PAIR

Buffer stages that are made with single active devices can be more effective if cascaded. Two types of such buffers are in common use. The *Darlington pair* is a cascade of two transistors connected as emitter followers as shown in **Figure 4.75.** The current gain of the Darlington pair is the product of the current gains for the two transistors, $b_1 \times b_2$.

What makes the Darlington pair so useful is that its input impedance is equal to the load impedance times the current gain, effectively multiplying the load impedance:

$$Z_I = Z_{LOAD} \times \beta_1 \times \beta_2$$

For example, if a typical bipolar transistor has β = 100 and Z_{LOAD} = 15 kΩ, a pair of these transistors in the Darlington-pair configuration would have:

$$Z_I = 15\ \text{k}\Omega \times 100 \times 100 = 150\ \text{M}\Omega$$

This impedance places almost no load on the circuit connected to the Darlington pair's input. The shunt capacitance at the input of real transistors can lower the actual impedance as the frequency increases.

Drawbacks of the Darlington pair include lower bandwidth and switching speed. The extremely high dc gain makes biasing very sensitive to temperature and component tolerances. For these reasons, the circuit is usually used as a switch and not as a linear amplifier.

CASCODE AMPLIFIERS

A common-emitter amplifier followed by a common-base amplifier is called a *cascode buffer*, shown in its simplest form in

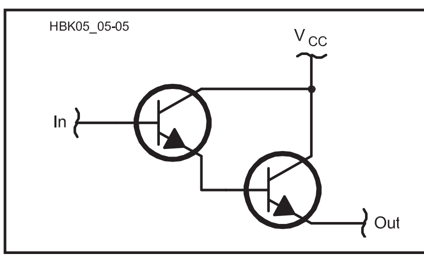


Figure 4.75 — Darlington pair made with two emitter followers. Input impedance, Z_I, is far higher than for a single transistor and output impedance, Z_O, is nearly the same as for a single transistor. DC biasing has been omitted for simplicity.

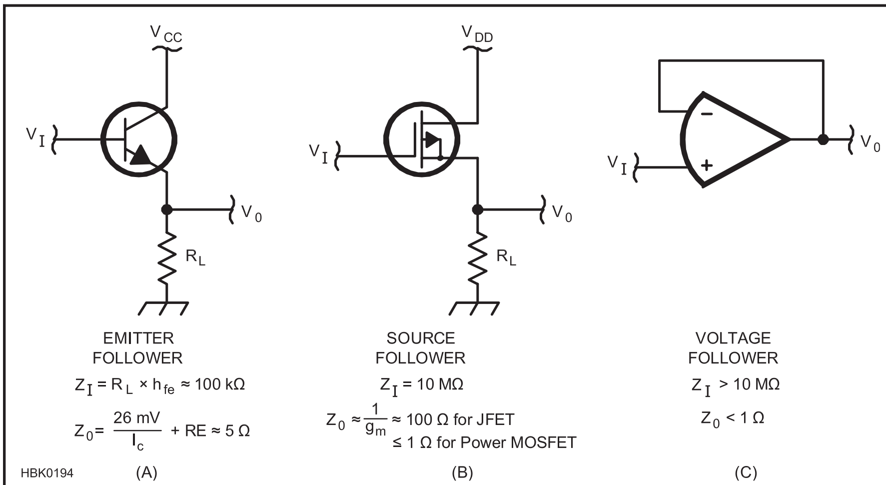


Figure 4.74 — Common buffer stages and some typical input (Z_I) and output (Z_O) impedances. (A) Emitter follower, made with an NPN bipolar transistor; (B) Source follower, made with an FET; and (C) Voltage follower, made with an operational amplifier. All of these buffers are terminated with a load resistance, R_L, and have an output voltage that is approximately equal to the input voltage (gain ≈ 1).

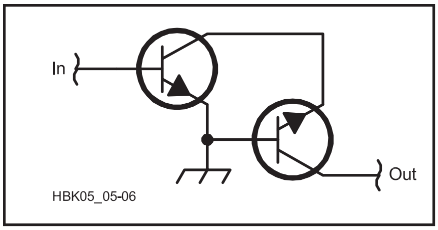


Figure 4.76 — Cascode buffer made with two NPN bipolar transistors has a medium input impedance and high output impedance. DC biasing has been omitted for simplicity.

Figure 4.76. (Biasing and dc blocking components are omitted for simplicity — replace the transistors with the practical circuits described earlier.) Cascode stages using FETs follow a common-source amplifier with a common-gate configuration. The input impedance and current gain of the cascode amplifier are approximately the same as those of the first stage. The output impedance of the common-base or -gate stage is much higher than that of the common-emitter or common-source amplifier. The power gain of the cascode amplifier is the product of the input stage current gain and the output stage voltage gain.

As an example, a typical cascode buffer made with BJTs has moderate input impedance (Z_{IN} = 1 kΩ), high current gain (h_{fe} = 50), and high output impedance (Z_{OUT} = 1 MΩ). Cascode amplifiers have excellent input/output isolation (very low unwanted internal feedback), resulting in high gain with good stability. Because of its excellent isolation, the cascode amplifier has little effect on external tuning components. Cascode circuits are often used in tuned amplifier designs for these reasons.

4.7.7 Using the Transistor as a Switch

When designing amplifiers, the goal was to make the transistor's output a replica of its input, requiring that the transistor stay within its linear region, conducting some current at all times. A switch circuit has completely different properties — its output current is either zero or some maximum value. **Figure 4.77** shows both a bipolar and metal-oxide semiconductor field-effect transistor (or MOSFET) switch circuit. Unlike the linear amplifier circuits, there are no bias resistors in either circuit. When using the bipolar transistor as a switch, it should operate in saturation or in cutoff. Similarly, an FET switch should be either fully-on or fully-off. The figure shows the waveforms associated with both types of switch circuits.

This discussion is written with power control in mind, such as to drive a relay or motor or lamp. The concepts, however, are equally applicable to the much lower-power circuits that control logic-level signals. The switch should behave just the same — switch between on and off quickly and completely — whether large or small.

More on Transistor Amplifiers at RF

See the **RF Techniques** chapter for more information on transistor amplifier design at RF, including a list of RF Amplifier Design resources in the References section.

DESIGNING SWITCHING CIRCUITS

First, select a transistor that can handle the load current and dissipate whatever power is dissipated as heat. Second, be sure that the input signal source can supply an adequate input signal to drive the transistor to the required states, both on and off. Both of these conditions must be met to insure reliable driver operation.

To choose the proper transistor, the load current and supply voltage must both be known. Supply voltage may be steady, but sometimes varies widely. For example, a car's 12 V power bus may vary from 9 to 18 V, depending on battery condition. The transistor must withstand the maximum supply voltage, V_{MAX}, when off. The load resistance, R_L, must also be known. The maximum steady-state current the switch must handle is:

$$I_{MAX} = \frac{V_{MAX}}{R_L}$$

If you are using a bipolar transistor, calculate how much base current is required to drive the transistor at this level of collector current. You'll need to inspect the transistor's data sheet because β decreases as collector current increases, so use a value for β specified at a collector current at or above I_{MAX}.

$$I_B = \frac{I_{MAX}}{\beta}$$

Now inspect the transistor's data sheet values for V_{CEsat} and make sure that this value of I_B is sufficient to drive the transistor fully into saturation at a collector current of I_{MAX}. Increase I_B if necessary — this is I_{Bsat}. The transistor must be fully saturated to minimize heating when conducting load current.

Using the *minimum* value for the input voltage, calculate the value of R_B:

$$R_B = \frac{V_{IN(min)} - V_{BE}}{I_{Bsat}}$$

The minimum value of input voltage must be used to accommodate the *worst-case* combination of circuit voltages and currents.

Designing with a MOSFET is a little easier because the manufacturer usually specifies the value V_{GS} must have for the transistor to be fully on, $V_{GS(on)}$. The MOSFET's gate, being insulated from the conducting channel, acts like a small capacitor of a few hundred pF and draws very little dc current. R_G in Figure 4.67 is required if the input voltage source does not actively drive its output to

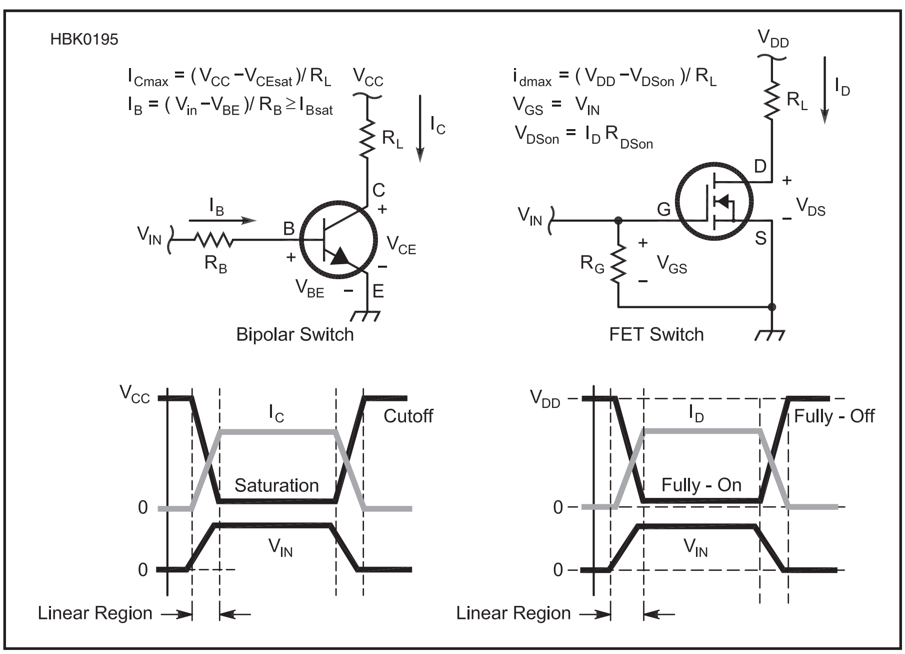


Figure 4.77 — A pair of transistor driver circuits using a bipolar transistor and a MOSFET. The input and output signals show the linear, cutoff and saturation regions.

zero volts when off, such as a switch connected to a positive voltage. The MOSFET won't turn off reliably if its gate is allowed to "float." R_G pulls the gate voltage to zero when the input is open-circuited.

Power dissipation is the next design hurdle. Even if the transistors are turned completely on, they will still dissipate some heat. Just as for a resistor, for a bipolar transistor switch the power dissipation is:

$$P_D = V_{CE} I_C = V_{CE(sat)}\ I_{MAX}$$

where $V_{CE(sat)}$ is the collector-to-emitter voltage when the transistor is saturated.

Power dissipation in a MOSFET switch is:

$$P_D = V_{DS} I_D = R_{DS(on)}\ I_{MAX}^2$$

$R_{DS(on)}$ is the resistance of the channel from drain to source when the MOSFET is on. MOSFETs are available with very low on-resistance, but still dissipate a fair amount of power when driving a heavy load. The transistor's data sheet will contain $R_{DS(on)}$ specification and the V_{GS} required for it to be reached.

Power dissipation is why a switching transistor needs to be kept out of its linear region. When turned completely off or on, either current through the transistor or voltage across it are low, also keeping the product of voltage and current (the power to be dissipated) low. As the waveform diagrams in Figure 4.67 show, while in the linear region, both voltage and current have significant values and so the transistor is generating heat when changing from off to on and vice versa. It's important to make the transition through the linear region quickly to keep the transistor cool.

The worst-case amount of power dissipated during each on-off transition is approximately

$$P_{transition} = \frac{1}{4} V_{MAX}\ I_{MAX}$$

assuming that the voltage and current increase and decrease linearly. If the circuit turns on and off at a rate of f, the total average power dissipation due to switching states is:

$$P_D = \frac{f}{2} V_{MAX}\ I_{MAX}$$

since there are two on-off transitions per switching cycle. This power must be added to the power dissipated when the switch is conducting current.

Once you have calculated the power the switch must dissipate, you must check to see whether the transistor can withstand it. The manufacturer of the transistor will specify a *free-air dissipation* that assumes no heat-sink and room temperature air circulating freely around the transistor. This rating should be at least 50% higher than your calculated power dissipation. If not, you must either use a larger transistor or provide some means of getting rid of the heat, such as heat sink. Methods of dissipating heat are discussed in this chapter's section on Heat Management.

INDUCTIVE AND CAPACITIVE LOADS

Voltage transients for inductive loads, such as solenoids or relays, can easily reach dozens of times the power supply voltage when load current is suddenly interrupted. To protect the transistor, the voltage transient must be clamped or its energy dissipated.

Where switching is frequent, such as for a variable frequency drive or switchmode power control, a series-RC *snubber* circuit (see **Figure 4.78**) is connected across the load. The RC circuit dissipates the transient's energy as heat. The following procedure will suffice for most amateur designs, and a more comprehensive analysis by Rudy Severns, N6LF, is listed in the References section.

1) Determine the transistor's output capacitance plus the device mounting capacitance. Multiply that value by 2 and select the nearest standard value for C_S. For example, if the transistor's output capacitance is 150 pF and the mounting capacitance is 70 pF, the total of 220 pF × 2 = 440 pF, and the nearest standard value is 470 pF. Use a low-loss film or ceramic component.

2) The required resistance for $R_S = V_{CC}$ / I_L where I_L is the load current. Power dissipation, $P_{DISS} = C_S V_{CC}^2 f_S$ where f_S is the switching frequency. Use a noninductive component such as carbon composition or metal oxide.

The most common method of dealing with the transient if switching is not frequent or continuous is to employ a "kickback" diode that is reverse-biased when the load is energized. (See Figure 4.68B.) When the load current is interrupted, the diode routes the transient's energy back to the power supply. See this chapter's section on Analog-Digital Interfacing for more information about kickback diode use. If the solenoid or relay is going to be used in an amateur station, add a small bypass capacitor (0.001 - 0.01 µF, value is not critical) across the diode to prevent it from generating harmonics or mixing products from strong RF.

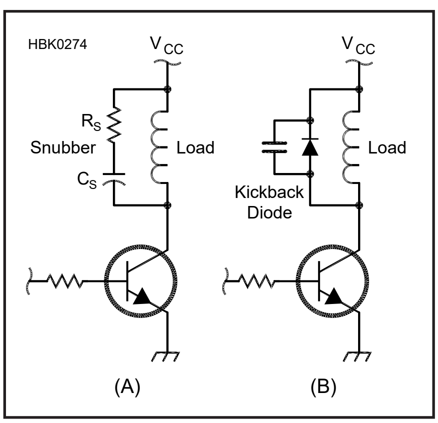


Figure 4.78 — The snubber RC circuit at (A) absorbs energy from transients with fast rise- and fall-times. At (B) a kickback diode protects the switching device when current is interrupted in the inductive load, causing a voltage transient, by conducting the energy back to the power source.

Capacitive loads such as heavily filtered power inputs may temporarily act like short circuits when the load is energized or de-energized. The surge current is only limited by the internal resistance of the load capacitance. The transistor will have to handle the temporary overloads without being damaged or overheating. The usual solution is to select a transistor with an I_{MAX} rating greater than the surge current. Sometimes a small cur-rent-limiting resistor can be placed in series with the load to reduce the peak surge current at the expense of dissipating power continuously when the load is drawing current.

HIGH-SIDE AND LOW-SIDE SWITCHING

The switching circuits shown in Figure 4.78 are *low-side switches*. This means the switch is connected between the load and ground. A *high-side switch* is connected between the power source and the load. The same concerns for power dissipation apply, but the methods of driving the switch change because the voltage of the emitter or source of the switching device will be at or near the power supply voltage when the switch is on.

To drive an NPN bipolar transistor or an N-channel MOSFET in a high-side circuit requires the switch input signal to be at least $V_{BE(sat)}$ or $V_{GS(on)}$ *above* the voltage supplied to the load. If the load expects to see the full power supply voltage, the switch input signal will have to be *greater* than the power supply voltage. A small step-up or boost dc-to-dc converter is often used to supply the extra voltage needed for the driver circuit.

One alternate method of high-side switching is to use a PNP bipolar transistor as the switching transistor. A small input transistor turns the main PNP transistor on by controlling the larger transistor's base current. Similarly, a P-channel MOSFET could also be employed with a bipolar transistor or FET acting as its driver. P-type material generally does not have the same high conductivity as N-type material, so these devices dissipate somewhat more power than N-type devices under the same load conditions.

4.7.8 Choosing a Transistor

With all the choices for transistors — websites and catalogs can list hundreds — selecting a suitable transistor can be intimidating. Start by determining the maximum voltage (V_{CEO} or $V_{DS(MAX)}$), current (I_{MAX}) and power dissipation ($P_{D(MAX)}$) the transistor must handle. Determine what dc current gain, β, or transconductance, g_m, is required. Then determine the highest frequency at which full gain is required and multiply it by either the voltage or current gain to obtain f_T or h_{fe}. This will reduce the number of choices dramatically.

4.8 Operational Amplifiers

An *operational amplifier*, or *op amp*, is one of the most useful linear devices. While it is possible to build an op amp with discrete components, and early versions were, the symmetry of the circuit demanded for high performance requires a close match of many components. It is more effective, and much easier, to implement as an integrated circuit. (The term "operational" comes from the op amp's origin in analog computers, where it was used to implement mathematical operations.)

The op amp's performance approaches that of an ideal analog circuit building block: an infinite input impedance (Z_i), a zero-output impedance (Z_o) and an open-loop voltage gain (A_v) of infinity. Obviously, practical op amps do not meet these specifications, but they do come closer than most other types of amplifiers. These attributes allow the circuit designer to implement many different functions with an op amp and only a few external components.

4.8.1 Characteristics of Practical Op-Amps

An op amp has three signal terminals (see **Figure 4.79**). There are two input terminals, the *noninverting input* marked with a + sign and the *inverting input* marked with a – sign. Voltages applied to the noninverting input cause the op amp output voltage to change with the same polarity.

The output of the amplifier is a single terminal with the output voltage referenced to the external circuit's reference voltage. Usually, that reference is ground, but the op amp's internal circuitry allows all voltages to *float*, that is, to be referenced to any arbitrary voltage between the op amp's power supply voltages. The reference can be negative, ground or positive. For example, an op amp powered from a single power supply voltage amplifies just as well if the circuit reference voltage is halfway between ground and the supply voltage.

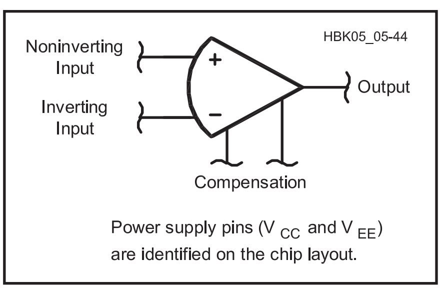


Figure 4.79 — Operational amplifier schematic symbol. The terminal marked with a + sign is the noninverting input. The terminal marked with a – sign is the inverting input. The output is to the right. On some op amps, external compensation is needed and leads are provided, pictured here below the device. Usually, the power supply leads are not shown on the op amp itself but are specified in the data sheet.

GAIN-BANDWIDTH PRODUCT AND COMPENSATION

An ideal op amp would have infinite frequency response, but just as transistors have an f_T that marks their upper frequency limit, the op amp has a *gain-bandwidth product* (GBW or BW). GBW represents the maximum product of gain and frequency available to any signal or circuit: voltage gain × frequency = GBW. If an op-amp with a GBW of 10 MHz is connected as a ×50 voltage amplifier, the maximum frequency at which that gain could be guaranteed is GBW / gain = 10 MHz / 50 = 200 kHz. GBW is an important consideration in high-performance filters and signal processing circuits whose design equations require high-gain at the frequencies over which they operate.

Older operational amplifiers, such as the LM301, have an additional two connections for *compensation*. To keep the amplifier from oscillating at very high gains it is often necessary to place a capacitor across the compensation terminals. This also decreases the frequency response of the op amp but increases its stability by making sure that the output signal cannot have the right phase to create positive feedback at its inputs. Most modern op amps are internally compensated and do not have separate pins to add compensation capacitance. Additional compensation can be created by connecting a capacitor between the op amp output and the inverting input.

CMRR AND PSRR

One of the major advantages of using an op amp is its very high *common mode rejection ratio* (CMRR). *Common mode* signals are those that appear equally at all terminals. For example, if both conductors of an audio cable pick up a few tenths of a volt of 60 Hz signal from a nearby power transformer, that 60 Hz signal is a common-mode signal to whatever device the cable is connected. Since the op amp only responds to *differences* in voltage at its inputs, it can ignore or reject common mode signals. CMRR is a measure of how well the op amp rejects the common mode signal. High CMRR results from the symmetry between the circuit halves. CMRR is important when designing circuits that process low-level signals, such as microphone audio or the mV-level dc signals from sensors or thermocouples.

The rejection of power-supply imbalance is also an important op amp parameter. Shifts in power supply voltage and noise or ripple on the power supply voltages are coupled directly to the op amp's internal circuitry. The op amp's ability to ignore those disturbances is expressed by the *power supply rejection ratio* (PSRR). A high PSRR means that the op amp circuit will continue to perform well even if the power supply is imbalanced or noisy.

INPUT AND OUTPUT VOLTAGE LIMITS

The op amp is capable of accepting and amplifying signals at levels limited by the power supply voltages, also called *rails*. The difference in voltages between the two rails limits the range of signal voltages that can be processed. The voltages can be symmetrical positive and negative voltages (±12 V), a positive voltage and ground, ground and a negative voltage or any two different, stable voltages.

In most op amps the signal levels that can be handled are one or two diode forward voltage drops (0.7 V to 1.4 V) away from each rail. Thus, if an op amp has 15 V connected as its upper rail (usually denoted V+) and ground connected as its lower rail (V–), input signals can be amplified to be as high as 13.6 V and as low as 1.4 V in most amplifiers. Any values that would be amplified beyond those limits are clamped (output voltages that should be 1.4 V or less appear as 1.4 V and those that should be 13.6 V or more appear as 13.6 V). This clamping action is illustrated in Figure 3.49 in the **Radio Fundamentals** chapter.

"Rail-to-rail" op amps have been developed to handle signal levels within a few tens of mV of rails (for example, the MAX406, from Maxim Integrated Products processes signals to within 10 mV of the power supply voltages). Rail-to-rail op-amps are often used in battery-powered products to allow the circuits to operate from low battery voltages for as long as possible.

INPUT BIAS AND OFFSET

The inputs of an op amp, while very high impedance, still allow some input current to flow. This is the *input bias current*, and it is in the range of nA in modern op amps. Slight asymmetries in the op amp's internal circuitry result in a slight offset in the op amp's out-put voltage, even with the input terminals shorted together. The amount of voltage difference between the op amp's inputs required to cause the output voltage to be exactly zero is the *input offset voltage*, generally a few mV or less. Some op amps, such as the LM741, have special terminals to which a potentiometer can be connected to *null* the offset by correcting the internal imbalance. Introduction of a small dc correction voltage to the noninverting terminal is sometimes used to apply an offset voltage that counteracts the internal mismatch and centers the signal in the rail-to-rail range.

DC offset is an important consideration in op amps for two reasons. Actual op amps have a slight mismatch between the inverting and noninverting terminals that can become a substantial dc offset in the output, depending on the amplifier gain. The op amp output voltage must not be too close to the clamping limits or distortion will occur.

A TYPICAL OP AMP

As an example of typical values for these parameters, one of today's garden-variety op amps, the TL084, which contains both JFET and bipolar transistors, has a guaranteed minimum CMRR of 80 dB, an input bias current guaranteed to be below 200 pA (1 pA = 1 millionth of a μA), and a gain-bandwidth product of 3 MHz. Its input offset voltage is 3 mV. CMRR and PSRR are 86 dB, meaning that an unwanted signal or power supply imbalance of 1 V will only result in a 2.5 nV change at the op amp's output! All this for less than 50 cents even purchased in single quantities, and there are four op-amps per package — that's a lot of performance.

4.8.2 Basic Op-Amp Circuits

If a signal is connected to the input terminals of an op amp without any other circuitry attached, it will be amplified at the device's *open-loop gain* (typically 200,000 for the TL084 at dc and low frequencies, or 106 dB). This will quickly saturate the output at the power supply rails. Such large gains are rarely used. In most applications, negative feedback is used to limit the circuit gain by providing a feedback path from the output terminal to the inverting input terminal. The resulting *closed-loop gain* of the circuit depends solely on the values of the passive components used to form the loop (usually resistors and, for frequency-selective circuits, capacitors). The higher the op-amp's open-loop gain, the closer the circuit's actual gain will approach that predicted from the component values. Note that the gain of the op amp itself has not changed — it is the configuration of the external components that determines the overall gain of the circuit. Some examples of different circuit configurations that manipulate the closed-loop gain follow.

INVERTING AND NONINVERTING AMPLIFIERS

The op amp is often used in either an *inverting* or a *noninverting* amplifier circuit as shown in **Figure 4.80**. (Inversion means that the output signal is inverted from the input signal about the circuit's voltage reference as described below.) The amount of amplification is determined by the two resistors: the feedback resistor, R_f, and the input resistor, R_i.

In the noninverting configuration shown in Figure 4.80A, the input signal is connected to the op-amp's noninverting input. The feedback resistor is connected between the output and the inverting input terminal. The inverting input terminal is connected to R_i, which is connected to ground (or the circuit reference voltage).

This circuit illustrates how op amp circuits use negative feedback, the high open-loop gain of the op amp itself, and the high input impedance of the op amp inputs to create a stable circuit with a fixed gain. The signal applied to the noninverting input causes the output voltage of the op-amp to change with the same polarity. That is, a positive input signal causes a positive change in the op amp's output voltage. This voltage causes current to flow in the voltage divider formed by R_f and R_i. Because the current into the inverting input is so low, the current through R_f is the same as R_i.

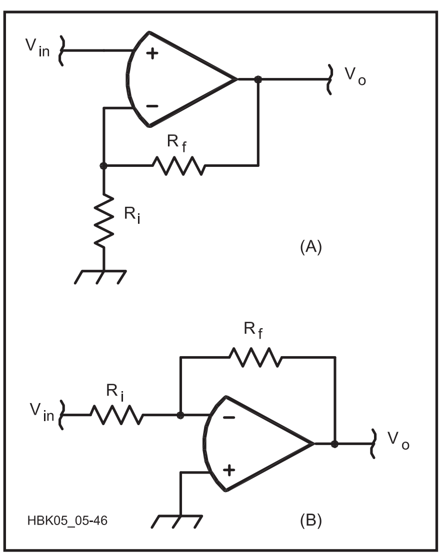


Figure 4.80 — Operational amplifier circuits. (A) Noninverting configuration. (B) Inverting configuration.

The voltage at the *summing junction*, the connection point for the two resistors and the inverting terminal, V_{INV}, is:

$$V_{INV} = V_O \left(\frac{R_i}{R_i + R_f} \right)$$

The op amp's output voltage will continue to rise until the *loop error signal*, the difference in voltage between the inverting and noninverting inputs, is close to zero. At this point, the voltage at the inverting terminal is approximately equal to the voltage at the noninverting terminal, V_{in}, so that $V_{INV} = V_{in}$. Substituting in the equation for V_{INV}, the gain of this circuit is:

$$\frac{V_O}{V_{in}} = \left(1 + \frac{R_f}{R_i} \right)$$

where
V_o = the output voltage, and
V_{in} = the input voltage.

The higher the op amp's open-loop gain, the closer will be the voltages at the inverting and noninverting terminals when the circuit is balanced and the more closely the circuit's closed-loop gain will equal that in the equation. So the negative feedback creates an electronic balancing act with the op amp increasing its output voltage so that the input error signal is as small as possible.

In the inverting configuration of Figure 4.80B, the input signal (V_{in}) is connected through R_i to the inverting terminal. The feedback resistor is again connected between the inverting terminal and the output. The noninverting terminal is connected to ground (or the circuit reference voltage). In this configuration the feedback action results in the output voltage changing to whatever value is needed such that the current through R_i is balanced by an equal and opposite current through R_f. The gain of this circuit is:

$$\frac{V_O}{V_{in}} = -\frac{R_f}{R_{in}}$$

where V_{in} represents the voltage input to R_{in}.

For the remainder of this section, "ground" or "zero voltage" should be understood to be the circuit reference voltage. That voltage may not be "earth ground potential." For example, if a single positive supply of 12 V is used, 6 V may be used as the circuit reference voltage. The circuit reference voltage is a fixed dc voltage that can be considered an ac ground because of the reference source's extremely low ac impedance.

The negative sign in the voltage gain equation indicates that the signal is inverted. For ac signals, inversion represents a 180° phase shift. The gain of the noninverting configuration can vary from a minimum of one to the maximum of which the op amp is capable, as indicated by A_v for dc signals, or the gain-bandwidth product for ac signals. The gain of the inverting configuration can vary from a minimum of zero (gains from zero to one attenuate the signal while gains of 1 and higher amplify the signal) to the maximum of which the device is capable.

The inverting amplifier configuration results in a special condition at the op amp's inverting input called *virtual ground*. Because the op amp's high open-loop gain drives the two inputs to be very close together, if the noninverting input is at ground potential, the inverting input will be very close to ground as well and the op amp's output will change with the input signal to maintain the inverting input at ground. Measured with a voltmeter, the input appears to be grounded, but it is merely maintained at ground potential by the action of the op amp and the feedback loop. This point in the circuit may not be connected to any other ground connection or circuit point because the resulting additional current flow will upset the balance of the circuit.

The *voltage follower* or *unity-gain buffer* circuit of **Figure 4.81** is commonly used as a buffer stage. The voltage follower has the input connected directly to the noninverting terminal and the output connected directly to the inverting terminal. This configuration has unity gain because the circuit is balanced when the output and input voltages are the same (error voltage equals zero). It also provides the maximum possible input impedance and the minimum possible output impedance of which the device is capable.

Differential and Difference Amplifier

A *differential amplifier* is a special application of an operational amplifier (see **Figure 4.82**). It amplifies the difference between two analog signals and is very useful to cancel noise under certain conditions. For instance, if an analog signal and a reference signal travel over the same cable they may pick up noise, and it is likely that both signals will have the same amount of noise. When the differential amplifier subtracts them, the signal will be unchanged but the noise will be completely removed, within the limits of the CMRR. The equation for differential amplifier operation is:

$$V_O = \frac{R_f}{R_i}\left[\frac{1}{\frac{R_n}{R_g}+1}\left(\frac{R_i}{R_f}+1\right) V_n - V_i\right]$$

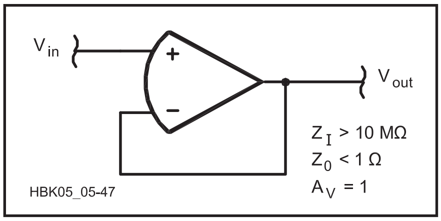


Figure 4.81 — Voltage follower. This operational amplifier circuit makes a nearly ideal buffer with a voltage gain of about one, and with extremely high input impedance and extremely low output impedance.

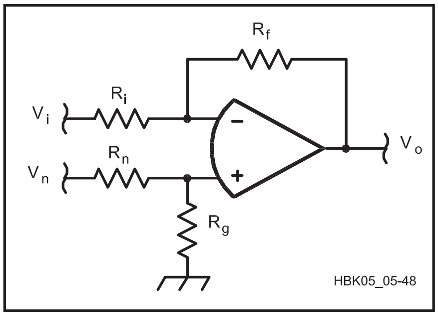


Figure 4.82 — Difference amplifier. This operational amplifier circuit amplifies the difference between the two input signals.

which, if the ratios R_i/R_f and R_n/R_g are equal, simplifies to:

$$V_O = \frac{R_f}{R_i}\left(V_n - V_1\right)$$

Note that the differential amplifier gain is identical to the inverting amplifier gain if the voltage applied to the noninverting terminal is equal to zero. If the voltage applied to the inverting terminal (V_i) is zero, the analysis is a little more complicated, but it is possible to derive the noninverting amplifier gain from the differential amplifier gain by taking into account the influence of R_n and R_g. If all four resistors have the *same* value the *difference amplifier* is created and V_O is just the difference of the two voltages:

$$V_O = V_n - V_1$$

Instrumentation Amplifier

Just as the symmetry of the transistors making up an op amp leads to a device with high values of Z_i, A_v and CMRR and a low value of Z_o, a symmetric combination of op amps is used to further improve these parameters. This circuit, shown in **Figure 4.83** is called an *instrumentation amplifier*. It has three parts; each of the two inputs is connected to a noninverting buffer amplifier with a gain of 1 + R2/R1. The outputs of these buffer amplifiers are then connected to a differential amplifier with a gain of R4/R3. V2 is the circuit's inverting input and V1 the noninverting input.

The three amplifier modules are usually all part of the same integrated circuit. This means that they have essentially the same temperature and the internal transistors and resistors are very well matched. This causes the subtle gain and tracking errors caused by temperature differences and mismatched components between individual op amps to be cancelled out or dramatically reduced. In addition, the external resistors using the same designators (R2, R3, R4) are carefully matched as well, sometimes being part of a single integrated *resistor pack*. The result is a circuit with better performance than any single-amplifier circuit over a wider temperature range.

Summing Amplifier

The high input impedance of an op amp makes it ideal for use as a *summing amplifier*. In either the inverting or noninverting configuration, the single input signal can be

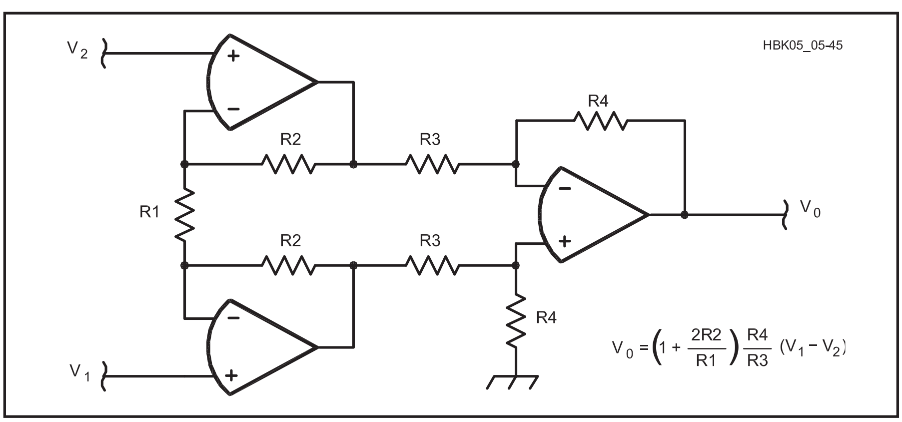


Figure 4.83 — Operational amplifiers arranged as an instrumentation amplifier. The balanced and cascaded series of op amps work together to perform differential amplification with good common-mode rejection and very high input impedance (no load resistor required) on both the inverting (V_1) and noninverting (V_2) inputs.

replaced by multiple input signals that are connected together through series resistors, as shown in **Figure 4.84**. For the inverting summing amplifier, the gain of each input signal can be calculated individually using the equation for inverting amplifier gain and, because of the superposition property of linear circuits, the output is the sum of each input signal multiplied by its gain. In the noninverting configuration, the output is the gain times the weighted sum of the m different input signals:

$$V_n = V_{n1}\frac{R_{p1}}{R_1 + R_{p1}} + V_{n2}\frac{R_{p2}}{R_2 + R_{p2}} + \ldots$$
$$+V_{nm}\frac{R_{pm}}{R_m + R_{pm}}$$

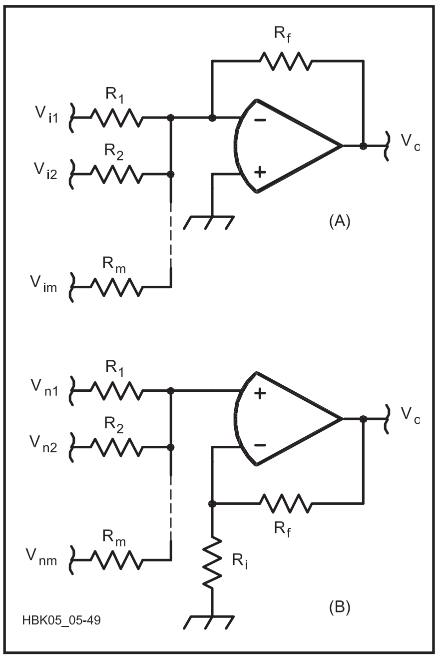


Figure 4.84 — Summing operational amplifier circuits. (A) Inverting configuration. (B) Noninverting configuration.

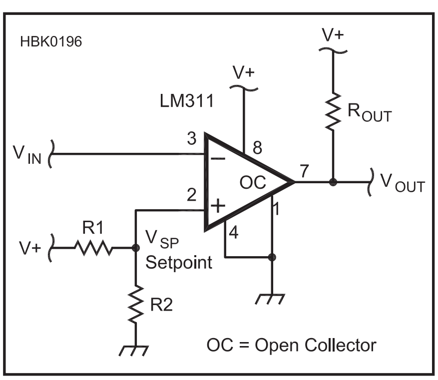


Figure 4.85 — A comparator circuit in which the output voltage is low when voltage at the inverting input is higher than the setpoint voltage at the noninverting input.

where R_{pm} is the parallel resistance of all m resistors excluding R_m. For example, with three signals being summed, R_{p1} is the parallel combination of R_2 and R_3.

Comparators

A *voltage comparator* is another special form of op amp circuit, shown in **Figure 4.85**. It has two analog signals as its inputs, and its output is either TRUE or FALSE depending on whether the noninverting or inverting signal voltage is higher, respectively. Thus, it "compares" the input voltages. TRUE generally corresponds to a positive output voltage and FALSE to a negative or zero voltage. The circuit in Figure 4.85 uses external resistors to generate a reference voltage, called the *setpoint*, to which the input signal is compared. A comparator can also compare two variable voltages.

A standard operational amplifier can be made to act as a comparator by connecting the two input voltages to the noninverting and inverting inputs with no input or feedback resistors. If the voltage of the noninverting input is higher than that of the inverting input, the output voltage will be driven to the positive clamping limit. If the inverting input is at a higher potential than the noninverting input, the output voltage will be driven to the negative clamping limit. If the comparator is comparing an unknown voltage to a known voltage, the known voltage is called the *setpoint* and the comparator output indicates whether the unknown voltage is above or below the setpoint.

An op amp that has been intended for use as a comparator, such as the LM311, is optimized to respond quickly to the input signals. In addition, comparators often have *open-collector outputs* that use an external *pull-up* resistor, R_{OUT}, connected to a positive power supply voltage. When the comparator output is TRUE, the output transistor is turned off and the pull-up resistor "pulls up" the output volt-

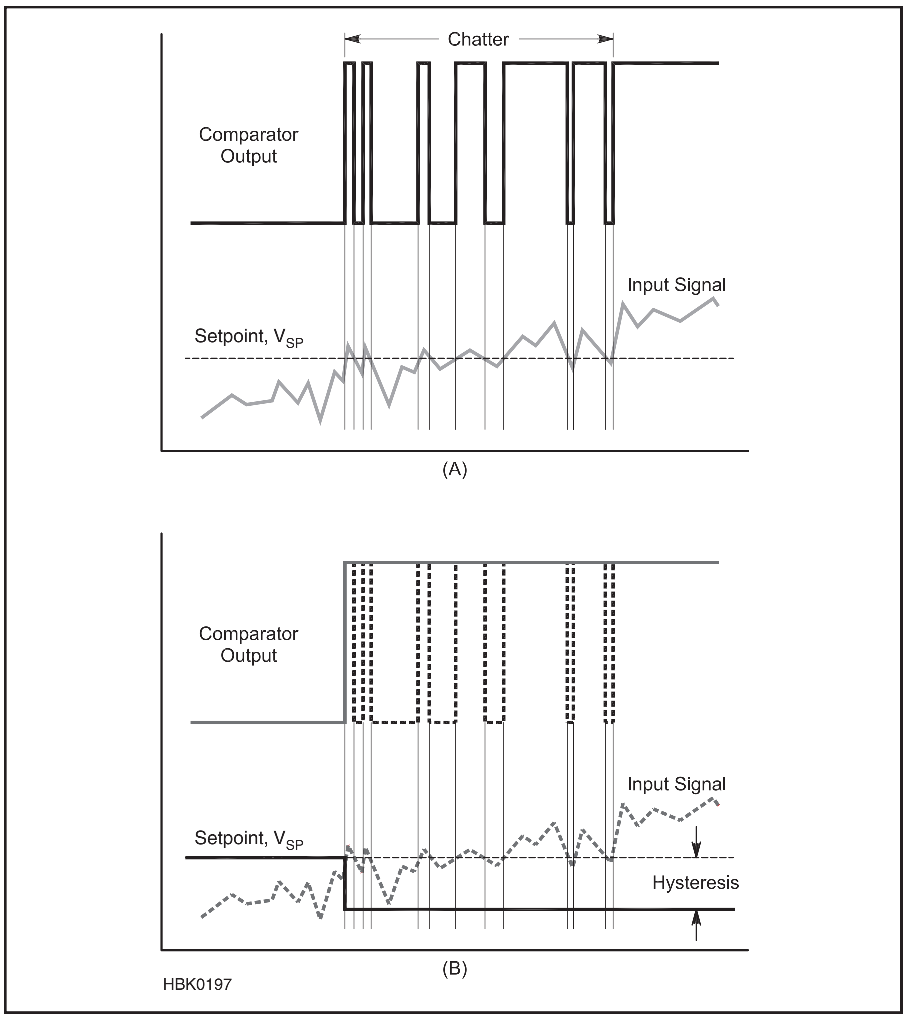


Figure 4.86 — Chatter (A) is caused by noise when the input signal is close to the setpoint. Chatter can also be caused by voltage shifts that occur when a heavy load is turned on and off. Hysteresis (B) shifts the setpoint a small amount by using positive feedback in which the output pulls the setpoint farther away from the input signal after switching.

age to the positive power supply voltage. When the comparator output is FALSE, the transistor is driven into the saturation and the output voltage is the transistor's $V_{CE(sat)}$.

Hysteresis

Comparator circuits also use *hysteresis* to prevent "chatter" — the output of the comparator switching rapidly back and forth when the input voltage is at or close to the setpoint voltage. There may be noise on the input signal, as shown in **Figure 4.86A**, that causes the input voltage to cross the setpoint threshold repeatedly. The rapid switching of the output can be confusing to the circuits monitoring the comparator output.

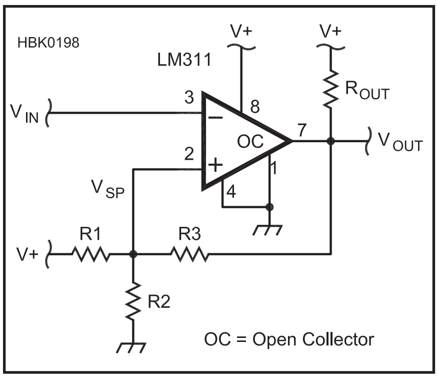


Figure 4.87 — Comparator circuit with hysteresis. R3 causes a shift in the comparator setpoint by allowing more current to flow through R1 when the comparator output is low.

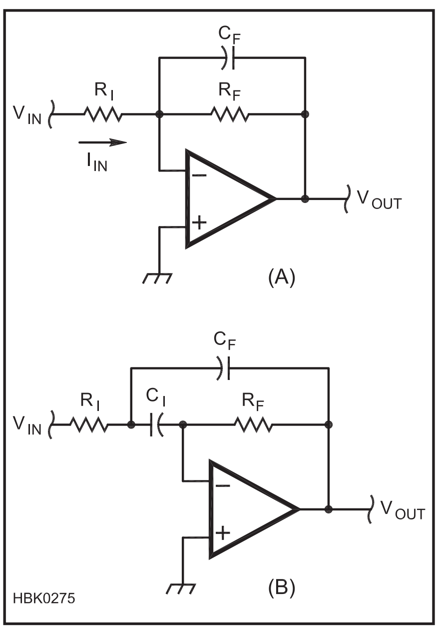


Figure 4.88 — Op amp active filters. The circuit at (A) has a low-pass response identical to an RC filter. The –3 dB frequency occurs when the reactance of C_F equals R_F. The band-pass filter at (B) is a multiple-feedback filter.

Hysteresis is a form of positive feedback that "moves" the setpoint by a few mV in the direction *opposite* to that in which the input signal crossed the setpoint threshold. As shown in Figure 4.86B, the slight shift in the setpoint tends to hold the comparator output in the new state and prevents switching back to the old state. **Figure 4.87** shows how the output of the comparator is fed back to the positive input through resistor R3, adding or subtracting a small amount of current from the divider and shifting the setpoint. If V_{HYS} is the amount of hysteresis desired (the shift in the setpoint voltage):

$$V_{HYS} \approx (V_{OH})\ (R1\ //\ R2)\ /\ [R3 + (R1\ //\ R2)]$$

where V_{OH} is the high-level output voltage with the comparator's output is off. Solving for R3 if the other values are known:

$$R3 \approx [(V_{OH})(R1\ //\ R2)\ /\ V_{HYS}] - (R1\ //\ R2)$$

Some applications of a voltage comparator are a zero crossing detector, a signal squarer (which turns other cyclical wave forms into square waves) and a peak detector. An amateur station application: Circuits that monitor the CI-V band data output voltage from ICOM HF radios use a series of comparators to sense the level of the voltage and indicate on which band the radio is operating.

FILTERS

One of the most important types of op amp circuits is the *active filter*. Two examples of op amp filter circuits are shown in **Figure 4.88.** The simple noninverting low-pass filter in Figure 4.88A has the same response as a passive single-pole RC low-pass filter, but unlike the passive filter, the op amp filter circuit has a very high input impedance and a very low output impedance, so that the filter's frequency and voltage response are relatively unaffected by the circuits connected to the filter input and output. This circuit is a low-pass filter because the reactance of the feedback capacitor decreases with frequency, requiring less output voltage to balance the voltages of the inverting and noninverting inputs.

The *multiple-feedback* circuit in Figure 4.88B results in a band-pass response while using only resistors and capacitors. This circuit is just one of many different types of active filters. Active filters are discussed in the **Analog and Digital Filtering** chapter.

RECTIFIERS AND PEAK DETECTORS

The high open-loop gain of the op amp can also be used to simulate the I-V characteristics of an ideal diode. A *precision rectifier* circuit is shown in **Figure 4.89** along with the I-V characteristics of a real (dashed lines) and ideal (solid line) diode. The high gain of the op amp compensates for the V_F forward voltage drop of the real diode in its feedback loop with an output voltage equal to the input voltage plus V_F. Remember that the op amp's output increases until its input voltages are balanced. When the input voltage is negative, which would reverse-bias the diode, the op amp's output can't balance the input because the diode blocks any current flow through the feedback loop. The resistor at the output holds the voltage at zero until the input voltage is positive once again. Precision half-wave and full-wave rectifier circuits are shown in **Figure 4.90**, and their operation is described in many reference texts.

One application of the precision rectifier circuit useful in radio is the *peak detector*, shown in **Figure 4.91**. A precision rectifier is used to charge the output capacitor which holds the peak voltage. The output resistor sets the time constant at which the capacitor discharges. The resistor can also be replaced by a transistor acting as a switch to reset the detector. This circuit is used in AGC loops,

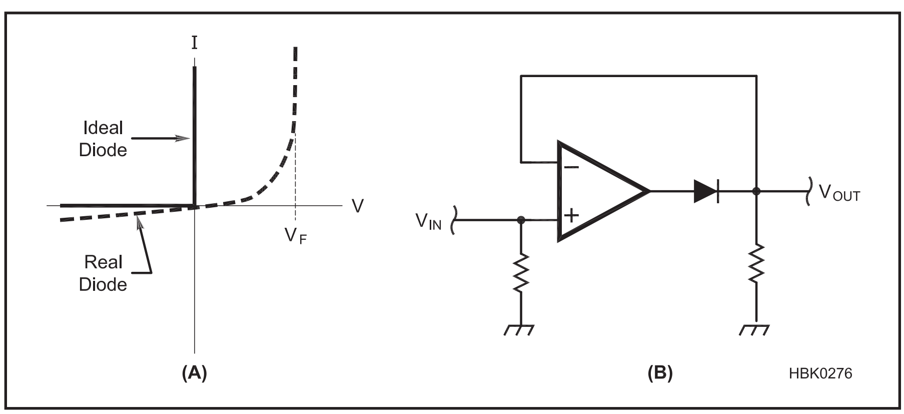


Figure 4.89 — Ideal and real diode I-V characteristics are shown at (A). The op amp precision rectifier circuit is shown at (B).

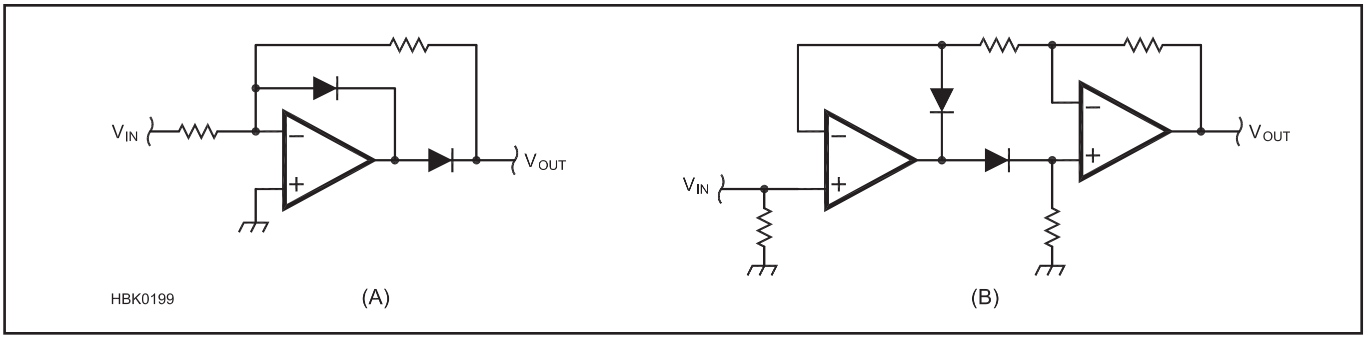


Figure 4.90 — Half-wave precision rectifier (A). The extra diode at the output of the op amp prevents the op amp from saturating on negative half-cycles and improves response time. The precision full-wave rectifier circuit at (B) reproduces both halves of the input waveform.

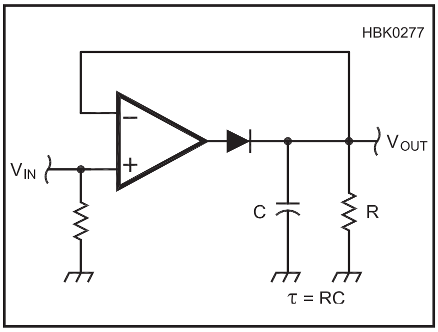


Figure 4.91 — Peak detector. Coupling a precision diode with a capacitor to store charge creates a peak detector. The capacitor will charge to the peak value of the input voltage. R discharges the capacitor with a time constant of τ = RC and can be omitted if it is desired for the output voltage to remain nearly constant.

spectrum analyzers, and other instruments that measure the peak value of ac waveforms.

LOG AMPLIFIER

There are a number of applications in radio in which it is useful for the gain of an amplifier to be higher for small input signals than for large input signals. For example, an audio compressor circuit is used to reduce the variations in a speech signal's amplitude so that the average power output of an AM or SSB transmitter is increased. A *log amplifier* circuit whose gain for large signals is proportional to the logarithm of the input signal's amplitude is shown in **Figure 4.92**. The log amp circuit is used in compressors and limiter circuits.

At signal levels that are too small to cause significant current flow through the diodes, the gain is set as in a regular inverting amplifier, $A_V = -R_f / R_i$. As the signal level increases, however, more current flows through the diode according to the Fundamental Diode Equation (see the **Electronic Fundamentals** chapter). That means the op amp output voltage has to increase less (lower gain) to cause enough current to flow through R_i such that the input voltages balance. The larger the input voltage, the more the diode conducts and the lower the gain of the circuit. Since the diode's current is exponential in response to voltage, the gain of the circuit for large input signals is logarithmic.

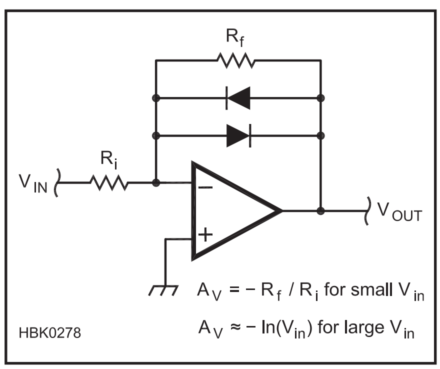


Figure 4.92 — Log amplifier. At low voltages, the gain of the circuit is $-R_f/R_i$, but as the diodes begin to conduct for higher-voltage signals, the gain changes to $-\ln(V_{in})$ in because of the diode's exponential current as described in the Fundamental Diode Equation.

Voltage-Current Converters

Another pair of useful op amp circuits convert voltage into current and current into voltage. These are frequently used to convert currents from sensors and detectors into voltages that are easier to measure. **Figure 4.93A** shows a voltage-to-current converter in which the output current is actually the current in the feedback loop. Because the op amp's high open-loop gain insures that its input voltages are equal, the current $I_{R1} = V_{IN} / R1$. Certainly, this could also be achieved with a resistor and Ohm's law, but the op amp circuit's high input impedance means there is little interaction between the input voltage source and the output current.

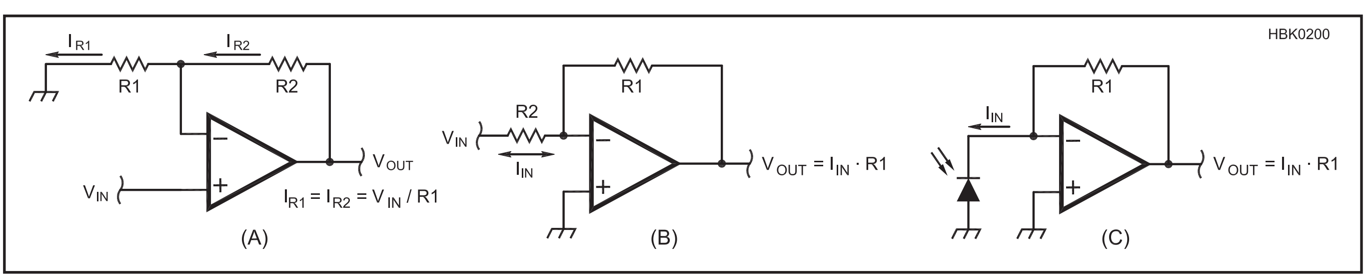


Figure 4.93 — Voltage-current converters. The current through R1 in (A) equals V_{in}/R1 because the op amp keeps both input terminals at approximately the same voltage. At (B), input current is balanced by the op amp, resulting in $V_{OUT} = I_{IN}R1$. Current through a photodiode (C) can be converted into a voltage in this way.

Going the other way, Figure 4.93B is a current-to-voltage converter. The op amp's output will change so that the current through the feedback resistor, R1, exactly equals the input current, keeping the inverting terminal at ground potential. The output voltage, $V_O = I_{IN}$ R1. Again, this could be done with just a resistor, but the op amp provides isolation between the source of input current and the output voltage. Figure 4.93C shows an application of a current-to-voltage converter in which the small currents from a photodiode are turned into voltage. This circuit can be used as a detector for amplitude modulated light pulses or waveforms.

4.9 Miscellaneous Analog ICs

The three main advantages of designing a circuit into an IC are to take advantage of the matched characteristics of its components, to make highly complex circuitry more economical, and to miniaturize the circuit and reduce power consumption. As circuits standardize and become widely used, they are often converted from discrete components to integrated circuits. Along with the op amp described earlier, there are many such classes of linear ICs.

4.9.1 Transistor and Driver Arrays

The most basic form of linear integrated circuit and one of the first to be implemented is the component array. The most common of these are the resistor, diode and transistor arrays. Though capacitor arrays are also possible, they are used less often. Component arrays usually provide space saving, but this is not the major advantage of these devices. They are the least densely packed of the integrated circuits because each device requires a separate off-chip connection. While it may be possible to place over a million transistors on a single semiconductor chip, individual access to these would require a total of three million pins, which is beyond the limits of practicability. More commonly, resistor and diode arrays contain from 5 to 16 individual devices and transistor arrays contain from three to six individual transistors. The advantage of these arrays is the very close matching of component values within the array. In a circuit that needs matched components, the component array is often a good method of obtaining this feature. The components within an array can be internally combined for special functions, such as termination resistors, diode bridges and Darlington pair transistors. A nearly infinite number of possibilities exists for these combinations of components, many of which are available in arrays.

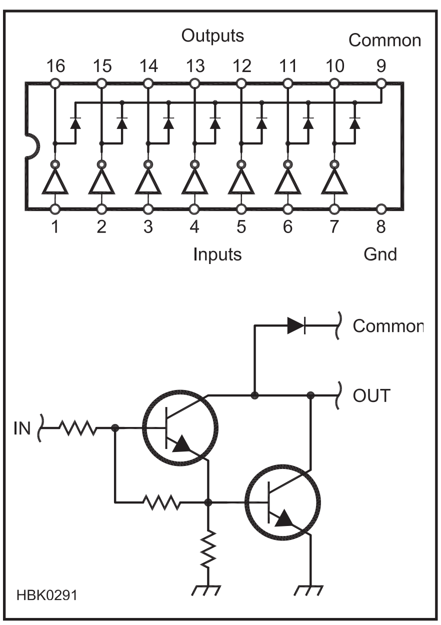


Figure 4.94 — Typical ULN2,000-series driver array configuration and internal circuit. The use of driver array ICs is very popular as an interface between microprocessor or other low-power digital circuits and loads such as relays, solenoids or lamps.

Driver arrays, such as the ULN2,000-series devices shown in **Figure 4.94** are very useful in creating an interface between low-power circuits such as microprocessors and higher-power loads and indicators. Each driver consists of a Darlington pair switching circuit as described earlier in this chapter. There are different versions with different types and arrangements of resistors and diodes.

Many manufacturers offer driver arrays. They are available with built-in kickback diodes to allow them to drive inductive loads, such as relays, and are heavy enough to source or sink current levels up to 1 A. (All of the drivers in the array cannot operate at full load at the same time, however. Read the data sheet carefully to determine what limitations on current and power dissipation may exist.)

4.9.2 Voltage Regulators and References

One of the most popular linear ICs is the voltage regulator. There are two basic types, the three-terminal regulator and the regulator controller. Examples of both are described in the **Power Sources** chapter.

The three-terminal regulator (input, ground, output) is a single package designed to perform all of the voltage regulation functions. The output voltage can be fixed, as in the 7800-series of regulators, or variable, as in the LM317 regulator. It contains a voltage reference, comparator circuits, current and temperature sensing protective circuits, and the main pass element. These ICs are usually contained in the same packages as power transistors and the same techniques of thermal management are used to remove excess heat.

Regulator controllers, such as the popular 723 device, contain all of the control and voltage reference circuitry but require external components for the main pass element, current sensing, and to configure some of their control functions.

Voltage references such as the Linear Technology LT1635 are special semiconductor diodes that have a precisely controlled I-V characteristic. A buffer amplifier isolates the sensitive diode and provides a low output impedance for the voltage signal. Voltage references are used as part of power regulators and by analog-digital converter circuits.

4.9.3 Timers (Multivibrators)

A *multivibrator* is a circuit that oscillates between two states, usually with a square wave or pulse train output. The frequency of oscillation is accurately controlled with the addition of appropriate values of external resistance and capacitance. The most common multivibrator in use today is the 555 timer IC (NE555 by Signetics [now Philips] or LM555 by National Semiconductor). This very simple eight-pin device has a frequency range from less than one hertz to several hundred kilohertz. Such a device can also be used in *monostable* operation, where an input pulse generates an output pulse of a different duration, or in *a stable* or *free-running* operation, where the device oscillates continuously. Other applications of a multivibrator include a frequency divider, a delay line, a pulse width modulator and a pulse position modulator. (These can be found in the IC's data sheet or in the reference listed at the end of this chapter.)

Figure 4.95 shows the basic components of a 555. Connected between power input (V_{cc}) and ground, the three resistors labeled "R" at the top left of the figure form a *voltage divider* that divides V_{CC} into two equal steps—one at 2/3 V_{CC} and one at 1/3 V_{CC}.

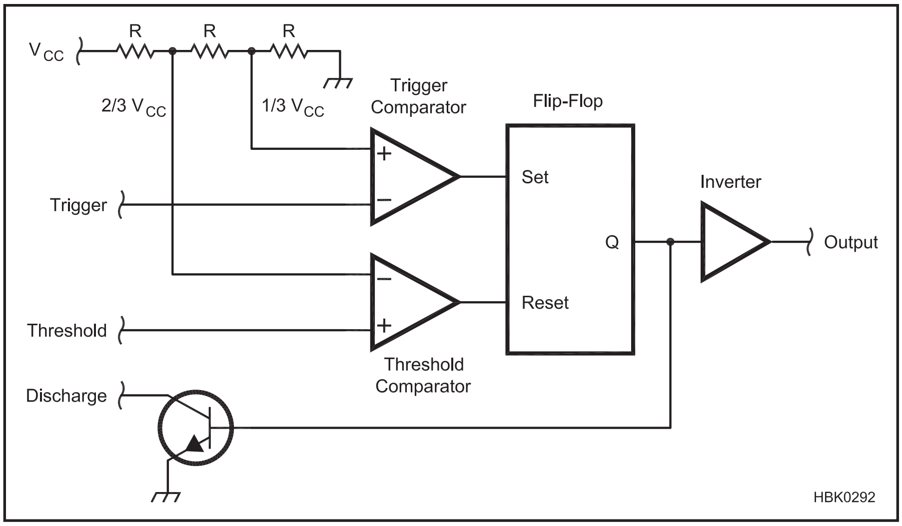


Figure 4.95 — Internal NE555 timer components. This simple array of components combines to make one of the most popular analog ICs. The 555 timer IC uses ratios of internal resistors to generate a precise voltage reference for generating time intervals based on charging and discharging a capacitor.

These serve as reference voltages for the rest of the circuit.

Connected to the reference voltages are blocks labeled *trigger comparator* and *threshold comparator*. (Comparators were discussed in a preceding section.) The trigger comparator in the 555 is wired so that its output is high whenever the trigger input is *less* than ⅓ V_{CC} and vice versa. Similarly, the threshold comparator output is high whenever the threshold input is *greater* than ⅔ V_{CC}.

These two outputs control a digital *flip-flop* circuit with an output, *Q,* that changes to high or low when the state of its *set* and *reset* input changes. The Q output stays high or low (it *latches* or *toggles*) until the opposite input changes. When the set input changes from low to high, Q goes low. When reset changes from low to high, Q goes high. The flip-flop ignores any other changes. An inverter makes the 555 output high when Q is low and vice versa — this makes the timer circuit easier to interface with external circuits.

The transistor connected to Q acts as a switch. When Q is high, the transistor is on and acts as a closed switch connected to ground. When Q is low, the transistor is off and the switch is open. These simple building blocks — voltage divider, comparator, flip-flop and switch — build a surprising number of useful circuits.

THE MONOSTABLE OR "ONE-SHOT" TIMER

The simplest 555 circuit is the monostable circuit. This configuration will output one fixed-length pulse when triggered by an input pulse. **Figure 4.96** shows the connections for this circuit.

Starting with capacitor C discharged, the flip-flop output, Q, is high, which keeps the discharge transistor turned on and the voltage on C below ⅔ V_{CC}. The circuit is in its stable state, waiting for a trigger pulse.

When the voltage at the trigger input drops below ⅓ V_{CC}, the trigger comparator output changes from low to high, which causes Q to toggle to the low state. This turns off the transistor (opens the switch) and allows C to begin charging toward V_{CC}.

When C reaches ⅔ V_{CC}, this causes the threshold comparator to switch its output from low to high, which resets the flip-flop. Q returns high, turning on the transistor and discharging C. The circuit has returned to its stable state. The output pulse length for the monostable configuration is:

$$T = 1.1R\,C_1$$

Notice that the timing is independent of the absolute value of V_{CC} — the output pulse width is the same with a 5 V supply as it is with a 15 V supply. This is because the 555 design is based on ratios and not absolute voltage levels.

THE ASTABLE MULTIVIBRATOR

The complement to the monostable circuit is the astable circuit in **Figure 4.97**. Pins 2, 6 and 7 are configured differently and the timing resistor is now split into two resistors, R1 and R2.

Start from the same state as the monostable circuit, with C completely discharged. The monostable circuit requires a trigger pulse to initiate the timing cycle. In the astable circuit, the trigger input is connected directly to the capacitor, so if the capacitor is discharged, then the trigger comparator output must be high. Q is low, turning off the discharge transistor, which allows C to immediately begin charging.

C charges toward V_{CC}, but now through the combination of R1 and R2. As the capacitor voltage passes ⅔ V_{CC}, the threshold comparator output changes from low to high, resetting Q to high. This turns on the discharge transistor and the capacitor starts to discharge through R2. When the capacitor is discharged below ⅓ V_{CC}, the trigger comparator changes from high to low and the cycle begins again, automatically. This happens over and over, causing a train of pulses at the output while C charges and discharges between ⅓ and ⅔ V_{CC} as seen in the figure.

The total time it takes for one complete cycle is the charge time, T_c, plus the discharge time, T_d:

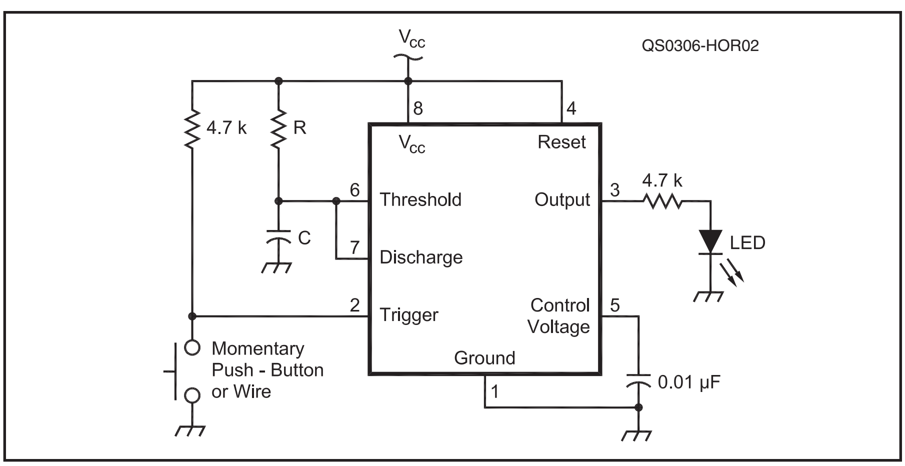


Figure 4.96 — Monostable timer. The timing capacitor is discharged until a trigger pulse initiates the charging process and turns the output on. When the capacitor has charged to ⅔ V_{CC}, the output is turned off, the capacitor is discharged and the timer awaits the next trigger pulse.

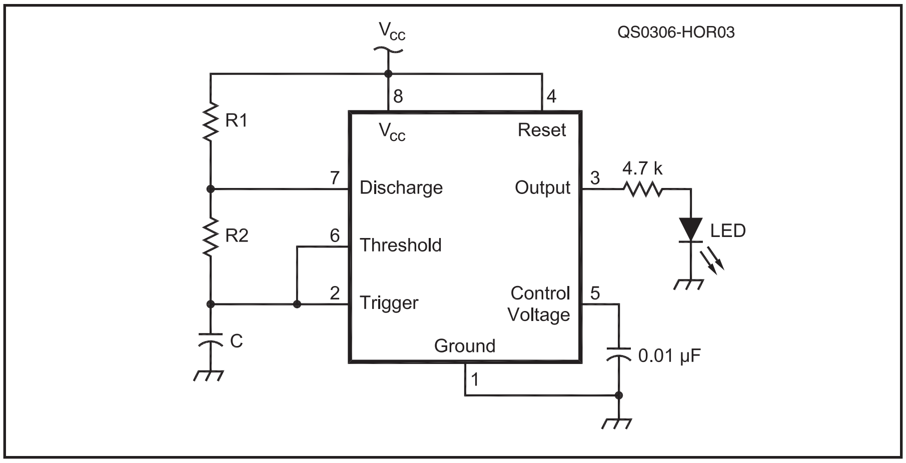


Figure 4.97 — Astable timer. If the capacitor discharge process initiates the next charge cycle, the timer will output a pulse train continuously.

$T = T_c + T_d = 0.693(R_1 + R_2)C + 0.693 R_2C = 0.693 (R_1 + 2R_2)C$

and the output frequency is:

$$f = \frac{1}{T} = \frac{1.443}{(R_1 + 2\,R_2)\,C}$$

When using the 555 in an application in or around radios, it is important to block any RF signals from the IC power supply or timing control inputs. Any unwanted signal present on these inputs, especially the Control Voltage input, will upset the timer's operation and cause it to operate improperly. The usual practice is to use a 0.01 µF bypass capacitor (shown on pin 5 in both Figure 4.96 and 4.97) to bypass ac signals such as noise or RF to ground. Abrupt changes in V_{CC} will also cause changes in timing, and these may be prevented by connecting filter capacitors at the V_{CC} input to ground.

4.9.4 Analog Switches and Multiplexers

Arrays of analog switches, such as the Maxim MAX312-series, allow routing of audio through lower frequency RF signals without mechanical switches. There are several types of switch arrays. Independent switches have isolated inputs and outputs and are turned on and off independently. Both SPST and SPDT configurations are available. Multiple switches can be wired with common control signals to implement multiple-pole configurations.

Use of analog switches at RF through microwave frequencies requires devices specifically designed for those frequencies. The Analog Devices ADG901 is a switch usable to 2.5 GHz. It absorbs the signal when off, acting as a terminating load. The ADG902 instead reflects the signal as an open circuit when off. Arrays of three switches called "T-switches" are used when very high isolation between the input and output is required.

Multiplexers or "muxes" are arrays of SPST switches configured to act as a multiposition switch that connects one of four to sixteen input signals to a single output. Demultiplexers ("demuxes") have a single input and multiple outputs. Multiplexer ICs are available as single N-to-1 switches (the MAX4617 is an 8-to-1 mux) or as groups of N-to-1 switches (the MAX4618 is a dual 4-to-1 mux).

Crosspoint switch arrays are arranged so that any of four to sixteen signal inputs can be connected to any of four to sixteen output signal lines. The Analog Devices AD8108 is an 8-by-8 crosspoint switch with eight inputs and eight outputs. These arrays are used when it is necessary to switch multiple signal sources among multiple signal receivers. They are most commonly used in telecommunications.

All analog switches use FET technology as the switching element. To switch ac signals, most analog switches require both positive and negative voltage power supplies. An alternative is to use a single power supply voltage and ground, but bias all inputs and output at one-half the power supply voltage. This requires dc blocking capacitors in all signal paths, both input and output, and loading resistors may be required at the device outputs. The blocking capacitors can also introduce low-frequency roll-off.

The impedance of the switching element varies from a few ohms to more than 100 ohms. Check the switch data sheet to determine the limits for how much power and current the switches can handle. Switch arrays, because of the physical size of the array, can have significant coupling or *crosstalk* between signal paths. Use caution when using analog switches for high-frequency signals as coupling generally increases with frequency and may compromise the isolation required for high selectivity in receivers and other RF signal processing equipment.

4.9.5 Audio Output Amplifiers

While it is possible to use op amps as low power audio output drivers for headphones, they generally have output impedances that are too high for most audio transducers such as speakers and headphones. The LM380 series of audio driver ICs have been used in radio circuits for many years, and a simple schematic for a speaker driver is shown in **Figure 4.98**.

The popularity of personal music players has resulted in the creation of many new and inexpensive audio driver ICs, such as the National Semiconductor LM4800- and LM4900-series. Drivers that operate from voltages as low as 1.5 V for battery-powered devices and up to 18 V for use in vehicles are now available.

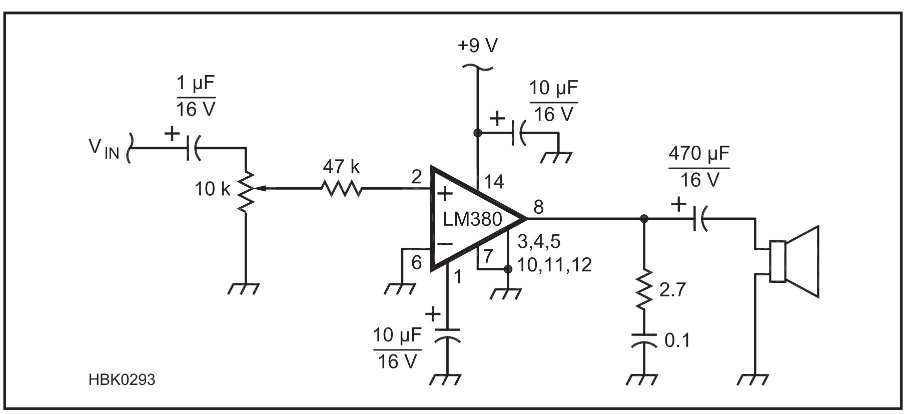


Figure 4.98 — Speaker driver. The LM380-series of audio output drivers are well-suited for low-power audio outputs, such as for headphones and small speakers. When using IC audio output drivers, be sure to refer to the manufacturer's data sheet for layout and power supply guidelines.

When choosing an audio driver IC for communications audio, the most important parameters to evaluate are its power requirements and power output capabilities. An overloaded or underpowered driver will result in distortion. Driver ICs intended for music players have frequency responses well in excess of the 3,000 Hz required for communications. This can lead to annoying and fatiguing hiss unless steps are taken to reduce the circuit's frequency response.

Audio power amplifiers should also be carefully decoupled from the power supply and the manufacturer may recommend specific circuit layouts to prevent oscillation or feedback. Check the device's data sheet for this information.

4.9.6 Temperature Sensors

Active temperature sensors use the temperature-dependent properties of semiconductor devices to create voltages that correspond to absolute temperature in degrees Fahrenheit (LM34) or degrees Celsius (LM35). These sensors (of which many more are available than the two examples given here) are available in small plastic packages, both leaded and surface-mount, that respond quickly to temperature changes. They are available with 1% and better accuracy, requiring only a source of voltage at very low current and ground. Complete application information is available in the manufacturer data sheets. Thermistors, a type of passive temperature sensor, are discussed in the Thermal Management section of this chapter. Temperature sensors are used in radio mostly in cooling and temperature control circuits.

4.10 Analog-Digital Interfacing

Quite often, logic circuits must either drive or be driven from non-logic sources. A very common requirement is sensing the presence or absence of a high (as compared to +5 V) voltage or perhaps turning on or off a 120 V ac device or moving the motor in an antenna rotator. A similar problem occurs when two different units in the shack must be interfaced because induced ac voltages or ground loops can cause problems with the desired signals.

A slow speed but safe way to interface such circuits is to use a relay. This provides absolute isolation between the logic circuits and the load. **Figure 4.99A** shows the correct way to provide this connection. The relay coil is selected to draw less than the available current from the driving logic circuit.

When current through the coil (or any inductive load) is suddenly interrupted by the switching transistor, a voltage transient called *back EMF* is generated at the switching device. The transient can easily reach dozens of times the power supply voltage, damaging or destroying a switching transistor. To clamp the voltage and return the stored energy to the power supply, a *kickback* or *flyback diode* is used. The diode limits the voltage at the switching transistor to the power supply voltage plus the diode's forward voltage drop.

Use a diode with a PIV rating of at least twice the operating voltage (5, 12, or 24 V in Figure 4.99). The diode's peak current rating should be several times the coil's steady-state current. In almost all solid-state circuits, an inexpensive 1N4001 (100 V, 1 A) diode will suffice.

The use of a kickback diode can increase the turn-off time of the relay. If the turn-off time is an important part of the design, a Zener diode in series with the kickback diode can be used. The Zener voltage should be approximately twice the supply voltage (a 1N5252 is a 24V Zener). This will still result in some turn-off delay. Experimentation may be necessary to find the best trade-off between Zener voltage and turn-off time for critical circuits.

Tranzorbs and MOVs can also be used.

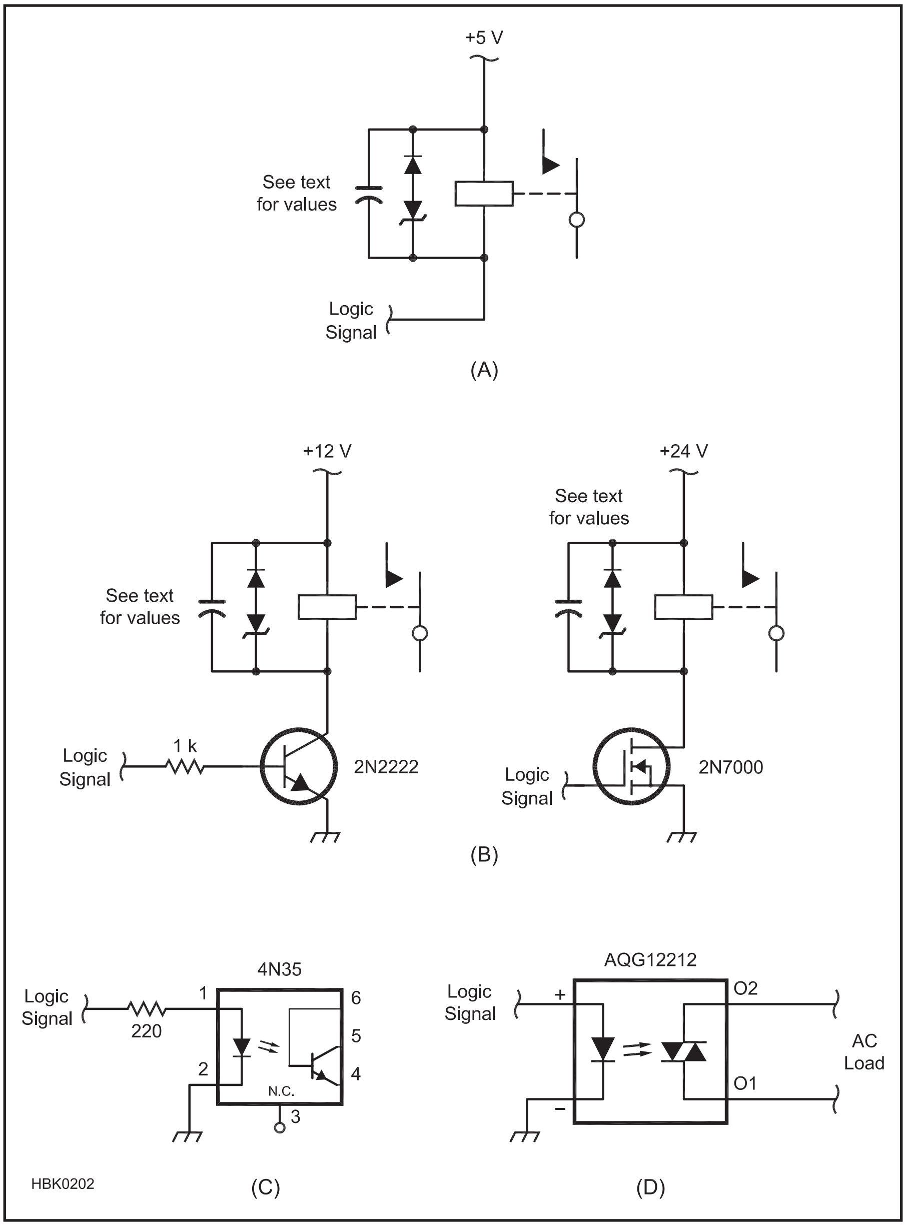


Figure 4.99 — Interface circuits for logic driving real-world loads. (A) driving a relay from a logic output; (B) using a bipolar transistor or MOSFET to boost current capacity; (C) using an optoisolator for electrical isolation; (D) using a solid-state relay for switching ac loads.

MOVs may change clamping voltages after extended use. A thorough discussion of relay drive protection can be found in STMicroelectronics Application Note AN-319, "Relay Drive Protection," which is available online from several sources.

For applications in which the relay is connected to circuits where strong RF might be present, such as antenna switches, rotator control cables, or tuning circuits, add a 0.01 to 0.001 µF ceramic capacitor (rated at 50 V or more) in parallel with the coil to prevent the diode(s) or clamping devices from generating mixing products or harmonics.

It is often not possible to find a relay that meets the load requirements *and* has a coil that can be driven directly from the logic output. Figure 4.99B shows two methods of using transistors to allow the use of higher power relays with logic gates.

Electro-optical couplers such as optoisolators and solid-state relays can also be used for this circuit interfacing. Figure 4.99C uses an optoisolator to interface two sets of logic circuits that must be kept electrically isolated, and Figure 4.99D uses a solid-state relay to control an ac line supply to a high current load. Note that this example uses a solid-state relay with internal current limiting on the input side; the LED input has an impedance of approximately 300 Ω. Some devices may need a series resistor to set the LED current; always consult the device data sheet to avoid exceeding device limits of the relay or the processor's I/O pin.

For safely using signals with voltages higher than logic levels as inputs, the same simple resistor and Zener diode circuit similar to that shown in Figure 4.38 can be used to clamp the input voltage to an acceptable level. Care must be used to choose a resistor value that will not load the input signal unacceptably.

4.11 Heat Management

While not strictly an electrical fundamental, managing the heat generated by electronic circuits is important in nearly all types of radio equipment. Thus, the topic is included in this chapter. Information on the devices and circuits discussed in this section may be found in other chapters.

Any actual energized circuit consumes electric power because any such circuit contains components that convert electricity into other forms of energy. This dissipated power appears in many forms. For example, a loudspeaker converts electrical energy into sound, the motion of air molecules. An antenna (or a light bulb) converts electricity into electromagnetic radiation. Charging a battery converts electrical energy into chemical energy (which is then converted back to electrical energy upon discharge). But the most common transformation by far is the conversion, through some form of resistance, of electricity into heat.

Sometimes the power lost to heat serves a useful purpose — toasters and hair dryers come to mind. But most of the time, this heat represents a power loss that is to be minimized wherever possible or at least taken into account. Since all real circuits contain resistance, even those circuits (such as a loudspeaker) whose primary purpose is to convert electricity to some *other* form of energy also convert some part of their input power to heat. Often, such losses are negligible, but sometimes they are not.

If unintended heat generation becomes significant, the involved components will get warm. Problems arise when the temperature increase affects circuit operation by either

- causing the component to fail, by explosion, melting, or other catastrophic event, or, more subtly,
- causing a slight change in the properties of the component, such as through a temperature coefficient (TC).

In the first case, we can design conservatively, ensuring that components are rated to safely handle two, three or more times the maximum power we expect them to dissipate. In the second case, we can specify components with low TCs, or we can design the circuit to minimize the effect of any one component. Occasionally we even exploit temperature effects (for example, using a resistor, capacitor or diode as a temperature sensor). Let's look more closely at the two main categories of thermal effects.

Not surprisingly, heat dissipation (more correctly, the efficient removal of generated heat) becomes important in medium- to high-power circuits: power supplies, transmitting circuits, and so on. While these are not the only examples where elevated temperatures and related failures are of concern, the techniques we will discuss here are applicable to all circuits.

4.11.1 Thermal Resistance

The transfer of heat energy, and thus the change in temperature, between two ends of a block of material is governed by the following heat flow equation and illustrated in **Figure 4.100**):

$$P = \frac{kA}{L}\Delta T = \frac{\Delta T}{\theta}$$

where

P = power (in the form of heat) conducted between the two points,

k = *thermal conductivity*, measured in W/(m °C), of the material between the two points, which may be steel, silicon, copper, PC board material, and so on,

L = length of the block,

A = area of the block, and

DT = *difference* in temperature between the two points.

Thermal conductivities of various common materials at room temperature are given in **Table 4.23**.

The heat flow equation has the same form as the variation of Ohm's law relating current flow to the ratio of a difference in potential to resistance; I = E/R. In this case, what's flowing is heat (P), the difference in potential

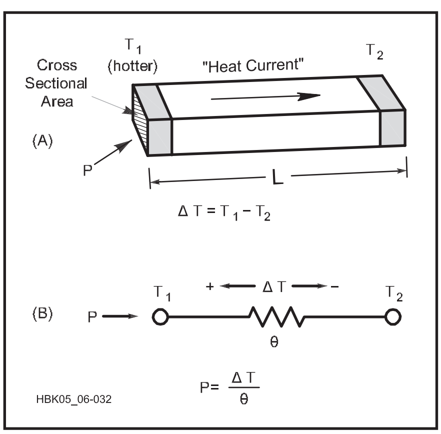


Figure 4.100 — Physical and "circuit" models for the heat-flow equation.

Table 4.23

Thermal Conductivities of Various Materials

Gases at 0 °C, Others at 25 °C; from *Physics,* by Halliday and Resnick, 3rd Ed.

Material	*k in units of W/m°C*
Aluminum	200
Brass	110
Copper	390
Lead	35
Silver	410
Steel	46
Silicon	150
Air	0.024
Glass	0.8

is a temperature difference (DT), and what's resisting the flow of heat is the *thermal resistance*:

$$\theta = \frac{L}{kA}$$

with units of °C/W. (The units of resistance are equivalent to V/A.) The analogy is so apt that the same principles and methods apply to heat flow problems as circuit problems. The following correspondences hold:

• Thermal conductivity W/(m °C) ↔ Electrical conductivity (S/m).

• Thermal resistance (°C/W) ↔ Electrical resistance (Ω).

• Thermal current (heat flow) (W) ↔ Electrical current (A).

• Thermal potential (T) ↔ Electrical potential (V).

• Heat source ↔ Power source.

For example, calculate the temperature of a 2-inch (0.05 m) long piece of #12 copper wire at the end that is being heated by a 25 W (input power) soldering iron, and whose other end is clamped to a large metal vise (assumed to be an infinite heat sink), if the ambient temperature is 25 °C (77 °F).

First, calculate the thermal resistance of the copper wire (diameter of #12 wire is 2.052 mm, cross-sectional area is 3.31×10^{-6} m^2)

$$\theta = \frac{L}{k\,A} = \frac{(0.05\text{ m})}{(390\text{ W}/(\text{m °C})) \times (3.31 \times 10^{-6}\text{ m}^2)}$$

= 38.7 ºC/W

Then, rearranging the heat flow equation above yields (after assuming the heat energy actually transferred to the wire is around 10 W)

DT = P θ = (10 W) × (38.7 °C/W) = 387 °C

So the wire temperature at the hot end is 25 C + DT = 412 °C (or 774 °F). If this sounds a little high, remember that this is for the steady state condition, where you've been holding the iron to the wire for a long time.

From this example, you can see that things can get very hot even with the application of moderate power levels. For this reason, circuits that generate sufficient heat to alter, not necessarily damage, the components must employ some method of cooling, either active or passive. Passive methods include heat sinks or careful component layout for good ventilation. Active methods include forced air (fans) or some sort of liquid cooling (in some high-power transmitters).

4.11.2 Heat Sink Selection and Use

The purpose of a heat sink is to provide a high-power component with a large surface area through which to dissipate heat. To use the models above, it provides a low thermal-resistance path to a cooler temperature, thus allowing the hot component to conduct a large "thermal current" away from itself.

Power supplies probably represent one of the most common high-power circuits amateurs are likely to encounter. Everyone has certainly noticed that power supplies get warm or even hot if not ventilated properly. Performing the thermal design for a properly cooled power supply is a very well-defined process and a good illustration of heat-flow concepts.

This material was originally prepared by ARRL Technical Advisor Dick Jansson, KD1K, during the design of a 28-V, 10-A power supply. (**Power Sources** chapter has more information on power supply design.) An outline of the design procedure shows the logic applied:

1. Determine the expected power dissipation (P_{in}).
2. Identify the requirements for the dissipating elements (maximum component temperature).
3. Estimate heat-sink requirements.
4. Rework the electronic device (if necessary) to meet the thermal requirements.
5. Select the heat exchanger (from heat sink data sheets).

The first step is to estimate the filtered, unregulated supply voltage under full load. Since the transformer secondary output is 32 V ac (RMS) and feeds a full-wave bridge rectifier, let's estimate 40 V as the filtered dc output at a 10-A load.

The next step is to determine the critical components and estimate their power dissipations. In a regulated power supply, the pass transistors are responsible for nearly all the power lost to heat. Under full load and allowing for some small voltage drops in the power-transistor emitter circuitry, the output of the series pass transistors is about 29 V for a delivered 28 V under a 10-A load. With an unregulated input voltage of 40 V, the total energy heat dissipated in the pass transistors is (40 V – 29 V) × 10 A = 110 W. The heat sink for this power supply must be able to handle that amount of dissipation and still keep the transistor junctions below the specified safe operating temperature limits. It is a good rule of thumb to select a transistor that has a maximum power dissipation of twice the desired output power.

Now, consider the ratings of the pass transistors to be used. This supply calls for 2N3055s as pass transistors. The data sheet shows that a 2N3055 is rated for 15-A service and 115-W dissipation. But the design uses *four* in parallel. Why? Here we must look past the big, bold type at the top of the data sheet to such subtle characteristics as the junction-to-case thermal resistance, q_{jc}, and the maximum allowable junction temperature, T_j.

The 2N3055 data sheet shows q_{jc} = 1.52 °C/W, and a maximum allowable case (and junction) temperature of 220 °C. While it seems that one 2N3055 could barely, on paper at least, handle the electrical requirements — at what temperature would it operate?

To answer that, we must model the entire "thermal circuit" of operation, starting with the transistor junction on one end and ending at some point with the ambient air. A reasonable model is shown in **Figure 4.101**. The ambient air is considered here as an infinite heat sink; that is, its temperature is assumed to be a constant 25 °C (77 °F). q_{jc} is the thermal resistance from the transistor junction to its case. q_{cs} is the resistance of the mounting interface between the transistor case and the heat sink. q_{sa} is the thermal resistance between the heat sink and the ambient air. In this "circuit," the generation of heat (the "thermal current source") occurs in the transistor at P_{in}.

Proper mounting of most TO-3 package power transistors such as the 2N3055 requires that they have an electrical insulator between the transistor case and the heat sink. However, this electrical insulator must at the same time

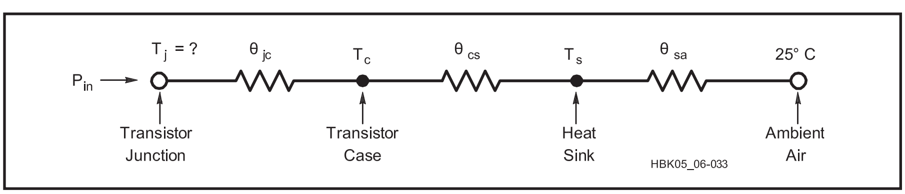


Figure 4.101 — Resistive model of thermal conduction in a power transistor and associated heat sink. See text for calculations.

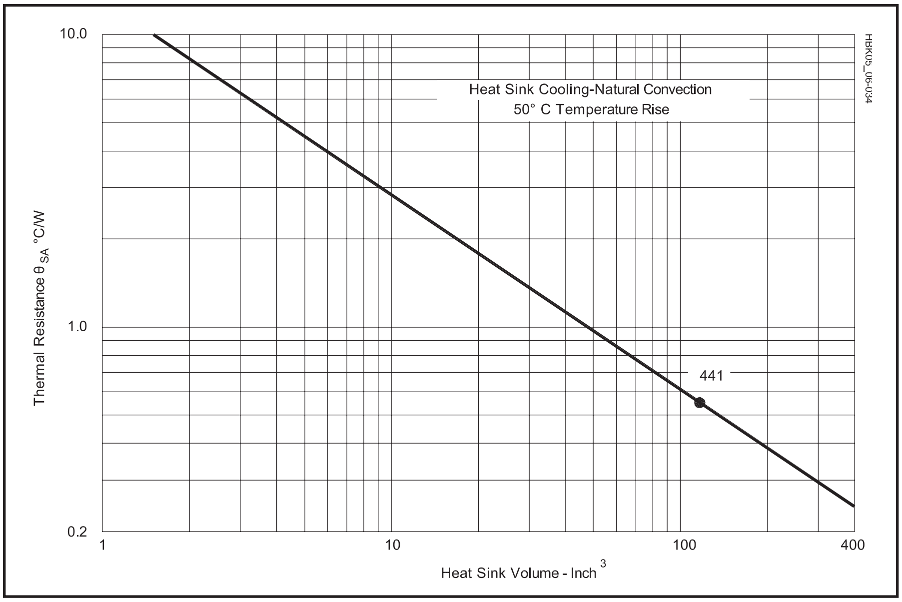


Figure 4.102 — Thermal resistance vs. heat-sink volume for natural convection cooling and 50 °C temperature rise. The graph is based on engineering data from Wakefield Thermal Solutions, Inc.

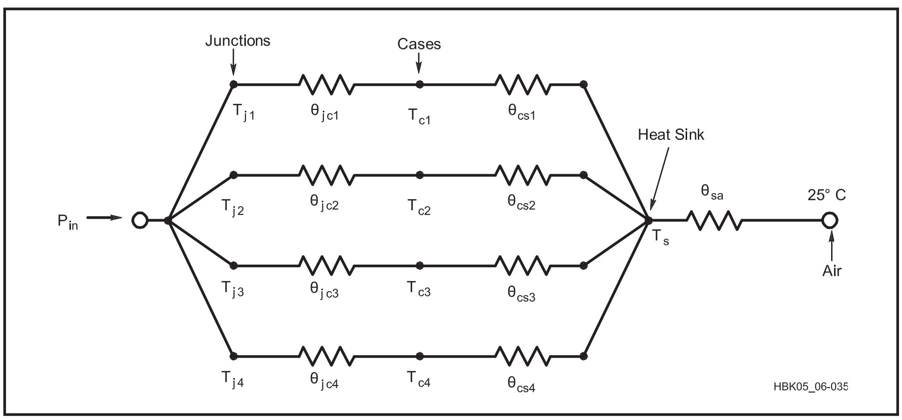


Figure 4.103 — Thermal model for multiple power transistors mounted on a common heat sink.

exhibit a low thermal resistance. To achieve a quality mounting, use thin polyimide or mica formed washers and a suitable thermal compound to exclude air from the interstitial space. "Thermal greases" are commonly available for this function. Any silicone grease may be used, but filled silicone oils made specifically for this purpose are better.

Using such techniques, a conservatively high value for q_{cs} is 0.50 °C/W. Lower values are possible, but the techniques needed to achieve them are expensive and not generally available to the average amateur. Furthermore, this value of q_{cs} is already much lower than q_{jc}, which cannot be lowered without going to a somewhat more exotic pass transistor.

Finally, we need an estimate of q_{sa}. **Figure 4.102** shows the relationship of heat-sink volume to thermal resistance for natural-convection cooling. This relationship presumes the use of suitably spaced fins (0.35 inch or greater) and provides a "rough order-of-magnitude" value for sizing a heat sink. For a first calculation, let's assume a heat sink of roughly 6 × 4 × 2 inch (48 cubic inches). From Figure 4.102, this yields a q_{sa} of about 1 °C/W.

Returning to Figure 4.101, we can now calculate the approximate temperature increase of a single 2N3055:

$dT = P\ q_{total}$

$= 110\ W \times (1.52\ °C/W + 0.5\ °C/W + 1.0\ °C/W)$

$= 332\ °C$

Given the ambient temperature of 25 °C, this puts the junction temperature T_j of the 2N3055 at 25 + 332 = 357 °C! This is clearly too high, so let's work backward from the air end and calculate just how many transistors we need to handle the heat.

First, putting more 2N3055s in parallel means that we will have the thermal model illustrated in **Figure 4.103**, with several identical q_{jc} and q_{cs} in parallel, all funneled through the same q_{as} (we have one heat sink).

Keeping in mind the physical size of the project, we could comfortably fit a heat sink of approximately 120 cubic inches (6 × 5 × 4 inches), well within the range of commercially available heat sinks. Furthermore, this application can use a heat sink where only "wire access" to the transistor connections is required. This allows the selection of a more efficient design. In contrast, RF designs require the transistor mounting surface to be completely exposed so that the PC board can be mounted close to the transistors to minimize parasitics. Looking at Figure 4.102, we see that a 120-cubic-inch heat sink yields a q_{sa} of 0.55 °C/W. This means that the temperature of the heat sink when dissipating 110 W will be 25 °C + (110 W × 0.55 °C/W) = 85.5 °C.

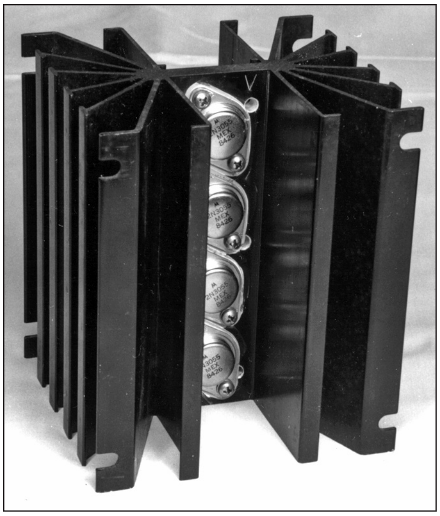
Industrial experience has shown that silicon transistors suffer substantial failure when junctions are operated at highly elevated temperatures. Most commercial and military specifications will usually not permit design junction temperatures to exceed 125 °C. To arrive at a safe figure for our maximum-allowed T_j, we must consider the intended use of the power supply. If we are using it in a 100% duty-cycle transmitting application such as RTTY or FM, the circuit will be dissipating 110 W continuously. For a lighter duty-cycle load such as CW or SSB, the "key-down" temperature can be slightly higher as long as the average is less than 125 °C. In this intermittent type of service, a good conservative figure to use is T_j = 150 °C.

Given this scenario, the temperature rise across each transistor can be 150 – 85.5 = 64.5 °C. Now, referencing Figure 2.4.93, remembering the total θ for each 2N3055 is 1.52 + 0.5 = 2.02 °C/W, we can calculate the maximum power each 2N3055 can safely dissipate:

$$P = \frac{\delta T}{\theta} = \frac{64.5\ °C}{2.02\ °C\ /\ W} = 31.9\ W$$

Thus, for 110 W full load, we need four 2N3055s to meet the thermal requirements of the design. Now comes the big question: What is the "right" heat sink to use? We have already established its requirements: it must be capable of dissipating 110 W and have a q_{sa} of 0.55 °C/W (see above).

A quick consultation with several manufacturer's catalogs reveals that Wakefield Thermal Solutions, Inc. model nos. 441 and



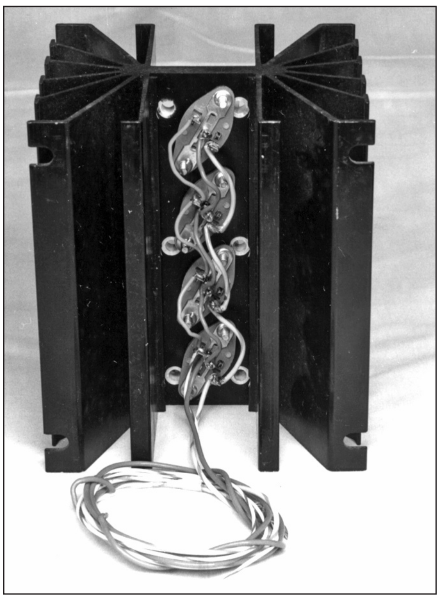


Figure 4.104 — A Wakefield 441 heat sink with four 2N3055 transistors mounted.

435 heat sinks meet the needs of this application. A Thermalloy model no. 6441 is suitable as well. Data published in the catalogs of these manufacturers show that in natural-convection service, the expected temperature rise for 100 W dissipation would be just under 60 °C, an almost perfect fit for this application. Moreover, the No. 441 heat sink can easily mount four TO-3-style 2N3055 transistors as shown in **Figure 4.104**. Remember: heat sinks should be mounted with the fins and transistor mounting area vertical to promote convection cooling.

The design procedure just described is applicable to any circuit where heat buildup is a potential problem. By using the thermal-resistance model, we can easily calculate whether or not an external means of cooling is necessary, and if so, how to choose it. Aside from heat sinks, forced air cooling (fans) is another common method. In commercial transceivers, heat sinks with forced-air cooling are common.

4.11.3 Semiconductor Temperature Effects

The number of excess holes and electrons in semiconductor material is increased as the temperature of a semiconductor increases. Since the conductivity of a semiconductor is related to the number of excess carriers, this also increases with temperature. With respect to resistance, semiconductors have a negative temperature coefficient. The resistance of silicon *decreases* by about 8% per °C and by about 6% per °C for germanium. Semiconductor temperature properties are the opposite of most metals, which *increase* their resistance by about 0.4% per °C. These opposing temperature characteristics permit the design of circuits with opposite temperature coefficients that cancel each other out, making a temperature insensitive circuit.

Semiconductor devices can experience an effect called *thermal runaway* as the current causes an increase in temperature. (This is primarily an issue with bipolar transistors.) The increased temperature decreases resistance and may lead to a further increase in current (depending on the circuit) that leads to an additional temperature increase. This sequence of events can continue until the semiconductor destroys itself, so circuit design must include measures that compensate for the effects of temperature.

Semiconductor Failure Caused by Heat

There are several common failure modes for semiconductors that are related to heat. The semiconductor material is connected to the outside world through metallic *bonding leads*. The point at which the lead and the semiconductor are connected is a common place for the semiconductor device to fail. As the device heats up and cools down, the materials expand and contract. The rate of expansion and contraction of semiconductor material is different from that of metal. Over many cycles of heating and cooling the bond between the semiconductor and the metal can break. Some experts have suggested that the lifetime of semiconductor equipment can be extended by leaving the devices powered on all the time, but this requires removal of the heat generated during normal operation.

A common failure mode of semiconductors is caused by the heat generated during semiconductor use. If the temperatures of the PN junctions remain at high enough levels for long enough periods of time, the impurities resume their diffusion across the PN junctions. When enough of the impurity atoms cross the depletion region, majority carrier recombination stops functioning properly and the semiconductor device fails permanently.

Excessive temperature can also cause failure anywhere in the semiconductor from heat generation within any current-carrying conductor, such as an FET channel or the bonding leads. Integrated circuits with more than one output may have power dissipation limits that depend on how many of the outputs are active at one time. The high temperature can cause localized melting or cracking of the semiconductor material, causing a permanent failure.

Another heat-driven failure mode, usually not fatal to the semiconductor, is excessive leakage current or a shift in operating point that causes the circuit to operate improperly. This is a particular problem in complex integrated circuits — analog and digital — dissipating significant amounts of heat under normal operating conditions. Computer microprocessors are a good example, often requiring their own cooling systems. Once the device cools, normal operation is usually restored.

To reduce the risk of thermal failures, the designer must comply with the limits stated in the manufacturer's data sheet, devising an adequate heat removal system. (Thermal issues are discussed in the **Electrical Fundamentals** chapter.)

4.11.4 Safe Operating Area (SOA)

Devices intended for use in circuits handling high currents or voltages are specified to have a *safe operating area* (*SOA*). This refers to the area drawn on the device's characteristic curve containing combinations of voltage and current that the device can be expected to control without damage under specific conditions. The SOA combines a number of limits — voltage, current, power, temperature and various breakdown mechanisms — in order to simplify the design of protective circuitry. The SOA is also specified to apply to specific durations of use — steady-state, long pulses, short pulses and so forth. The device may have separate SOAs for resistive and inductive loads.

You may also encounter two specialized types of SOA for turning the device on and off. *Reverse bias safe operating area* (*RBSOA*) applies when the device is turning off. *Forward bias safe operating* area (*FBSOA*) applies when turning the device on. These SOAs are used because the high rate-of-change of current and voltage places additional stresses on the semiconductor.

4.11.5 Semiconductor Derating

Maximum ratings for power transistors are usually based on a case temperature of 25 °C. These ratings will decrease with increasing operating temperature. Manufacturer's data sheets usually specify a *derating* figure or curve that indicates how the maximum ratings change per degree rise in temperature. If such information is not available (or even if it is!), it is a good rule of thumb to select a power transistor with a maximum power dissipation of at least twice the desired output power.

RECTIFIERS

Diodes are physically quite small, and they operate at high current densities. As a result their heat-handling capabilities are somewhat limited. Normally, this is not a problem in high-voltage, low-current supplies in which rectifiers in axial-lead DO-type packages are used. The use of high-current (2 A or greater) rectifiers at or near their maximum ratings, however, requires some form of heat sinking. The average power dissipated by a rectifier is

$$P = I_{AVG} \times V_F$$

where

I_{AVG} is the average current, and
V_F is the forward voltage drop.

Average current must account for the conduction duty cycle and the forward voltage drop must be determined at the average current level.

Rectifiers intended for such high-current applications are available in a variety of packages suitable for mounting to flat surfaces. Frequently, mounting the rectifier on the main chassis (directly, or with thin mica insulating washers) will suffice. If the diode is insulated from the chassis, thin layers of thermal compound or thermal insulating washers should be used to ensure good heat conduction. Large, high-current rectifiers often require special heat sinks to maintain a safe operating temperature. Forced-air cooling is sometimes used as a further aid.

4.11.6 RF Heating

RF current often causes component heating problems where the same level of dc current may not. An example is the tank circuit of an RF oscillator. If several small capacitors are connected in parallel to achieve a desired capacitance, skin effect will be reduced and the total surface area available for heat dissipation will be increased, thus significantly reducing the RF heating effects as compared to a single large capacitor. This technique can be applied to any similar situation; the general idea is to divide the heating among as many components as possible.

4.11.7 Forced-Air and Water Cooling

In amateur radio today, forced-air cooling is most commonly found in vacuum-tube circuits or in power supplies built in small enclosures, such as those in solid-state transceivers or computers. Fans or blowers are commonly specified in cubic feet per minute (CFM). While the nomenclature and specifications differ from those used for heat sinks, the idea remains the same: to offer a low thermal resistance between the inside of the enclosure and the (ambient) exterior.

For forced air cooling, we basically use the "one resistor" thermal model of Figure 4.100. The important quantity to be determined is heat generation, P_{in}. For a power supply, this can be easily estimated as the difference between the input power, measured at the transformer primary, and the output power at full load. For variable-voltage supplies, the worst-case output condition is minimum voltage with maximum current. A discussion of forced-air cooling for vacuum tube equipment appears in the **RF Power Amplifiers** chapter. For an in-depth discussion of fan cooling, see the References entry for Rabassa's February 2019 *QST* article.

Dust build-up is a common problem for forced-air cooling systems, even with powerful blowers and fans. If air intake grills and vents are not kept clean and free of lint and debris, air flow can be significantly reduced, leading to excessive equipment temperature and premature failure. Cleaning of air passageways should be included in regular equipment maintenance for good performance and maximum equipment life.

Water cooling systems are much less common in amateur equipment, used primarily for high duty cycle operating, such as RTTY, and at frequencies where the efficiency of the amplifier is relatively low, such as UHF and microwaves. In these situations, water cooling is used because water can absorb and transfer more than 3,000 times as much heat as the same volume of air!

The main disadvantage of water cooling is that it requires pumps, hoses, and reservoirs whereas a fan or blower is all that is required for forced-air cooling. For high-voltage circuits, using water cooling also requires special insulation techniques and materials to allow water to circulate in close contact with the heat source while remaining electrically isolated.

Nevertheless, the technique can be effective. The increased availability of inexpensive materials designed for home sprinkler and other low-pressure water distribution systems make water-cooling less difficult to implement. It is recommended that the interested reader review articles and projects in the amateur literature to observe the successful implementation of water cooling systems.

4.11.8 Heat Pipe Cooling

A heat pipe is a device containing a *working fluid* in a reservoir where heat is absorbed and a channel to a second reservoir where heat is dissipated. Heat pipes work by *evaporative cooling*. The heat-absorbing reservoir is placed in thermal contact with the heat source which transfers heat to the working fluid, usually a liquid substance with a boiling point just above room temperature. The working fluid vaporizes and the resulting vapor pressure pushes the hot vapor through the channel to the cooling reservoir.

In the cooling reservoir, the working fluid gives up its heat of vaporization, returning to the fluid state. The cooled fluid then flows back through the channel to the heat-absorbing reservoir where the process is repeated.

Heat pipes require no fans or pumps — movement of the working fluid is driven entirely from the temperature difference between the two reservoirs. The higher the temperature difference between the absorbing and dissipating reservoirs, the more effective the heat pump becomes, up to the limit of the dissipating reservoir to dissipate heat.

The primary application for heat pipes is for space operations. Heat pipe applications in terrestrial applications are limited due to the gravity gradient sensitivity of heat pipes. At present, computers and certain amplifier modules are the only amateur equipment making use of heat pipes. Nevertheless, as more general-purpose products become available, this technique will become more common.

4.11.9 Thermoelectric Cooling

Thermoelectric cooling makes use of the *Peltier effect* to create heat flow across the junction of two different types of materials. This process is related to the *thermoelectric effect* by which thermocouples generate voltages based on the temperature of a similar junction. A *thermoelectric cooler* or *TEC* (also known as a *Peltier cooler*) requires only a source of dc power to cause one side of the device to cool and the other side to warm. TECs are available with different sizes and power ratings for different applications.

TECs are not available with sufficient heat transfer capabilities that they can be used in high-power applications, such as RF amplifiers. However, they can be useful in lowering the temperature of sensitive receiver circuits, such as preamplifiers used at UHF and microwave frequencies, or imaging devices, such as charge-coupled devices (CCDs). Satellites use TECs as *radiative coolers* that dissipate heat directly as thermal or infrared radiation. TECs are also found in some computing equipment where they are used to remove heat from microprocessors and other large integrated circuits.

4.11.10 Temperature Compensation

Aside from catastrophic failure, temperature changes may also adversely affect circuits if the temperature coefficient (TC) of one or more components is too large. If the resultant change is not too critical, adequate temperature stability can often be achieved simply by using higher-precision components with low TCs (such as NP0/C0G capacitors or metal-film resistors). For applications where this is impractical or impossible (such as many solid-state circuits), we can minimize temperature sensitivity by *compensation* or *matching* — using temperature coefficients to our advantage.

Compensation is accomplished in one of two ways. If we wish to keep a certain circuit quantity constant, we can interconnect pairs of components that have equal but opposite TCs. For example, a resistor with a negative TC can be placed in series with a positive TC resistor to keep the total resistance constant. Conversely, if the important point is to keep the *difference* between two quantities constant, we can use components with the *same* TC so that the pair "tracks." That is, they both change by the same amount with temperature.

An example of this is a Zener reference circuit. Since a diode is strongly affected by operating temperature, circuits that use diodes or transistors to generate stable reference voltages must use some form of temperature compensation. Since, for a constant current, a reverse-biased PN junction has a negative voltage TC while a forward-biased junction has a positive voltage TC, a good way to temperature-compensate a Zener reference diode is to place one or more forward-biased diodes in series with it.

4.11.11 Thermistors

Thermistors can be used to control temperature and improve circuit behavior or protect against excessive temperatures, hot or cold. Circuit temperature variations can affect gain, distortion or control functions like receiver AGC or transmitter ALC. Thermistors can be used in circuits that compensate for temperature changes.

A *thermistor* is a small bit of intrinsic (no N or P doping) metal-oxide semiconductor compound material between two wire leads. As temperature increases, the number of liberated hole/electron pairs increases exponentially, causing the resistance to decrease exponentially. You can see this in the resistance equation:

$$R(T) = R(T0)e^{-\beta(1/T0-1/T)}$$

where T is some temperature in Kelvins and T0 is a reference temperature, usually 298 K (25°C), at which the manufacturer specifies R(T0).

The constant β is experimentally determined by measuring resistance at various temperatures and finding the value of β that best agrees with the measurements. A simple way to get an approximate value of β is to make two measurements, one at room temperature, say 25 °C (298 K) and one at 100 °C (373 K) in boiling water. Suppose the resistances are 10 kΩ and 938 Ω.

$$\beta = \frac{\ln\left(\frac{R(T)}{R(T0)}\right)}{\frac{1}{T} - \frac{1}{T0}} = \frac{\ln\left(\frac{938}{1000}\right)}{\frac{1}{373} - \frac{1}{298}} = 3507$$

With the behavior of the thermistor known — either by equation or calibration table — its change in resistance can be used to create an electronic circuit whose behavior is controlled by temperature in a known fashion. Such a circuit can be used for controlling temperature or detecting specific temperatures.

See "Termistors in Homebrew Projects" and "Thermistor Based Temperature Controller" both by Bill Sabin, WØIYH in this book's online information for practical projects using thermistors.

4.12 References and Bibliography

BOOKS AND PERIODICALS

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

ARRL Lab Staff, "Lab Notes — Capacitor Basics," *QST*, Jan. 1997, pp. 85 – 86.

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

Bergeron, B., NU1N, "Under the Hood II: Resistors," *QST,* Nov. 1993, pp. 41 – 44.

Bergeron, B., NU1N, "Under the Hood III: Capacitors," *QST,* Jan. 1994, pp. 45 – 48.

Bergeron, B., NU1N, "Under the Hood IV: Inductors," *QST,* Mar. 1994, pp 37 – 40.

Bergeron, B., NU1N, "Under the Hood: Lamps, Indicators, and Displays," *QST,* Sep. 1994, pp. 34 – 37.

Ebers, J., and Moll, J., "Large-Signal Behavior of Junction Transistors," *Proceedings of the IRE*, Vol. 42, No. 12, Dec. 1954, pp. 1761 – 1772.

Getreu, I., *Modeling the Bipolar Transistor* (Elsevier, 1979). Also available from Tektronix, Inc, Beaverton, Oregon, in paperback form. Must be ordered as Part Number 062-2841-00.

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

Grover, F., *Inductance Calculations*, Reprint (Dover Publications, 2009).

Hayward, W., W7ZOI, *Introduction to Radio Frequency Design* (ARRL, 2004).

Hayward, W., Campbell, R., and Larkin, B., *Experimental Methods in RF Design* (ARRL, 2009).

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

Jung, W., *IC Op Amp Cookbook* (Prentice Hall, 1986).

Kaiser, C., *The Capacitor Handbook*, 2nd Edition (CJ Publishing, 1995).

Kaiser, C., *The Diode Handbook* (CJ Publishing, 1999).

Kaiser, C., *The Inductor Handbook* (CJ Publishing, 1996).

Kaiser, C., *The Resistor Handbook*, 2nd Edition (CJ Publishing, 1998).

Kaiser, C., *The Transistor Handbook* (CJ Publishing, 1999).

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

McClanahan, J., W4JBM, "Understanding and Testing Capacitor ESR," *QST*, Sep. 2003, pp. 30 – 32.

Millman, J., and Grabel, A., *Microelectronics: Digital and Analog Circuits and Systems* (McGraw-Hill, 1988).

Mims, F., *Timer*, *Op Amp & Optoelectronic Circuits & Projects* (Master Publishing, 2004).

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

Rabassa, A., NW2M, "The Basics of Fan Cooling," *QST*, Feb. 2019, pp. 34 – 35.

Severns, R., N6LF, "Design of Snubbers for Power Circuits," Cornell-Dublier (CDE). **www.cde.com/resources/technical-papers/design.pdf**.

Silver, W., NØAX, "Experiment #24 — Heat Management," *QST*, Jan. 2005, pp. 64 –65.

Silver, W., NØAX, "Experiment #33 — The Transformer," *QST*, Oct. 2005, pp. 62 –63.

Silver, W., NØAX, "Experiment #62 — About Resistors," *QST*, Mar. 2008, pp. 66 –67.

Silver, W., NØAX, "Experiment #63 — About Capacitors," *QST*, Apr. 2008, pp. 70 – 71.

Silver, W., NØAX, "Experiment #132 — Resistor Networks," *QST*, Jan. 2014, pp. 63 – 64.

Smith, J., K8ZOA, "Carbon Composition, Carbon Film and Metal Oxide Film Resistors," *QEX*, Mar./Apr. 2008, pp. 46-57.

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

White, J., "Transistor Amplifier Design" *Microwaves & RF*, Parts 1 – 8, Jul. 2004 to Jan. 2005.

WEBSITES

"dB or not dB? Everything you ever wanted to know about decibels but were afraid to ask…," Rohde & Schwarz, Application Note 1MA98, **www.rohde-schwarz.us/en/applications/db-or-not-db-application-note_56280-15534.html**.

Hyperphysics Op-Amp Circuit Tutorials, **hyperphysics.phy-astr.gsu.edu/Hbase/Electronic/opampvar.html#c2**.

Munir, U., "Selecting the Right CMOS Analog Switch," Maxim Semiconductor, Application Note 5299, **www.maximintegrated.com/en/design/technical-documents/app-notes/5/5299.html**.

Schultz, W., "Power Transistor Safe Operating Area: Special Considerations for Switching Power Supplies," onsemi, AN-875-D, **www.onsemi.cn/pub/Collateral/AN875-D.PDF**.