ELECTRONIC COMPONENTS

MATERIALS AND PASSIVE COMPONENTS

Conductive Material and Components

Conductance and Resistance. When a voltage is applied across a conductor, a current will flow through the conductor. For a given voltage v (volts), the current i (amperes) will increase with the conductance G of the conductor. In the linear range of operation, this characteristic is expressed by the Ohm's law:

$$i = Gv$$

Resistance R (Ω) is the inverse of conductance:

$$R = 1/G$$

Silver, copper, gold, and aluminum are good conductors of electricity.

Resistivity. For a conductor, resistance increases with the length (L) and decreases with the area of cross-section (A). The corresponding relationship is:

$$R = \rho L/A$$

The constant of proportionality ρ is the resistivity of the conducting material. Hence, resistivity may be defined as the resistance of a conductor of unity length and unity cross-sectional area. It may be expressed in the units $\Omega \cdot \mathrm{cm}^2/\mathrm{cm}$ or $\Omega \cdot \mathrm{cm}$. A larger unit would be $\Omega \cdot \mathrm{m}^2/\mathrm{m}$ or $\Omega \cdot \mathrm{m}$.

Alternatively, resistivity may be defined as the resistance of a conductor of unity length and unity diameter. According to this definition,

$$R = \rho L/d^2$$

where d represents the wire diameter. If the wire diameter is 1 mil (or 1/1000 in), the wire area would be 1 circular mil (or cmil). Furthermore, if the wire length is 1 foot, the units of ρ would be $\Omega \cdot \mathrm{cmil/ft}$. Resistivities of several common materials are given in Table 1.

Effect of Temperature. Electrical resistance of a material can change with many factors. For example, the resistance of a typical metal increases with temperature, and the resistance decreases with temperature for many nonmetals and semiconductors. Typically, temperature effects on hardware have to be minimized in precision equipment, and temperature compensation or calibration would be necessary. On the other hand, high temperature sensitivity of resistance in some materials is exploited in temperature sensors such as resistance temperature detectors (RTDs) and thermistors. The sensing element of an RTD is made of a metal such as nickel, copper, platinum, or silver. For not too large variations in temperature, the following linear relationship could be used:

$$R = R_0(1 + \alpha \Delta t)$$

where R is the final resistance, R_0 is the initial resistance, ΔT is the change in temperature, and α is the temperature coefficient of resistance.

Values of α for several common materials are given in Table 2. These values can be expressed in ppm/°C (parts per million per degree centigrade) by multiplying each value by 10^6. Note that graphite has a negative temperature coefficient, and nichrome has a very low temperature coefficient of resistance. A platinum RTD can operate accurately over a wide temperature range and possesses a high sensitivity (typically 0.4Ω/°C).

Table 1. Resistivities of Some Materials

Material	Resistivity ρ ($\Omega \cdot$ m) at 20 °C (68 °F)
Aluminum	2.8×10^{-8}
Copper	1.7×10^{-8}
Ferrite (manganese-zinc)	20.0
Gold	2.4×10^{-8}
Graphite carbon	775.0×10^{-8}
Lead	9.6×10^{-8}
Magnesium	45.8×10^{-8}
Mercury	20.4×10^{-8}
Nichrome	112.0×10^{-8}
Polyester	1×10^{10}
Polystyrene	1×10^{16}
Porcelain	1×10^{16}
Silver	1.6×10^{-8}
Steel	15.9×10^{-8}
Tin	11.5×10^{-8}
Tungsten	5.5×10^{-8}

Note: Multiply by 6.0×10^8 to obtain the resistivity in $\Omega \cdot$ cmil/ft.

J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*. Copyright © 1999 John Wiley & Sons, Inc.

Table 2. Temperature Coefficients of Resistance for Several Materials

Material	Temp. Coeff. Resistance α (per °C) at 20 °C (68 °F)
Aluminum	0.0040
Brass	0.0015
Copper	0.0039
Gold	0.0034
Graphite carbon	−0.0005
Iron	0.0055
Lead	0.0039
Nichrome	0.0002
Silver	0.0038
Steel	0.0016
Tin	0.0042
Tungsten	0.0050

Thermistors are made of semiconductor material such as oxides of cobalt, copper, manganese, and nickel. Their resistance decreases with temperature. The relationship is nonlinear and is given approximately by

$$R = R_0 e^{-\beta(1/T_0 - 1/T)}$$

where the temperatures T and T_0 are in absolute degrees (K or R), and R and R_0 are the corresponding resistances. The parameter β is a material constant.

Effect of Strain. The property of resistance change with strain in materials, or piezoresistivity, is used in strain gauges. The foil strain gauges use metallic foils (e.g., a copper-nickel alloy called constantan) as their sensing elements. The semiconductor strain gauges use semiconductor elements (e.g., silicon with the trace impurity boron) in place of metal foils. An approximate relationship for a strain gauge is

$$\Delta R/R = S_s \epsilon$$

where ΔR is the change in resistance due to strain ϵ, R is initial resistance, and S_s is the sensitivity (gauge factor) of the strain gauge.

The gauge factor is of the order of 4.0 for a metal-foil strain gauge and can range from 40.0 to 200.0 for a semiconductor strain gauge (1).

Temperature effects have to be compensated for in high-precision measurements of strains. Compensation circuitry may be employed for this purpose. In semiconductor strain gauges, self compensation for temperature effects can be achieved due to the fact that the temperature coefficient of resistance varies nonlinearly with the concentration of the dope material (1). The temperature coefficient curve of a p-type semiconductor strain gauge is shown in Fig. 1.

Superconductivity. The resistivity of some materials drops virtually to zero when the temperature is decreased close to absolute zero, provided that the magnetic field strength of the environment is less than some critical value. Such materials are called superconducting materials. The superconducting temperature T (absolute) and the corresponding critical magnetic field strength H are related through

$$H = H_0(1 - T/T_c)^2$$

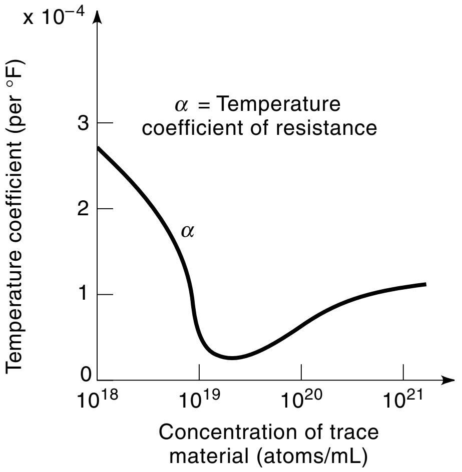


Figure 1. The temperature coefficient of resistance of a p-type semiconductor strain gauge.

where H_0 is the critical magnetic field strength for a superconducting temperature of absolute zero, and T_c is the superconducting temperature at zero magnetic field. The constants H_0 and T_c for several materials are listed in Table 3.

Superconducting elements can be used to produce high-frequency (e.g., 1×10^{11} Hz) switching elements (e.g., Josephson junctions) which can generate two stable states (e.g., zero voltage and a finite voltage, or zero magnetic field and a finite magnetic field). Hence, they are useful as computer memory elements. Other applications of superconductivity include powerful magnets with low dissipation (for medical imaging, magnetohydrodynamics, fusion reactors, particle accelerators, etc.), actuators (for motors, magnetically leviated vehicles, magnetic bearings, etc.), sensors, and in power systems.

Color Code for Fixed Resistors. Carbon, wound metallic wire, and conductive plastics are commonly used as commercial resistors. A wire-wound resistor element is usually encapsulated in a casing made of an insulating material such as porcelain or bakelite. Axial or radial leads are provided for external connection. The outer surface of a fixed resistor is color coded for the purpose of its specification. Four stripes are used for coding. The first stripe gives the first digit of a two digit number, and the second stripe gives the second digit. The third stripe specifies a multiplier which should be included with the two-digit number to give the resistance value in ohms. The fourth stripe gives the percentage tolerance of the resistance value. This color code is given in Table 4.

Table 3. Superconductivity Constants for Some Materials

Material	T_c (K)	H_0 (A/m)
Aluminum	1.2	0.8×10^4
Gallium	1.1	0.4×10^4
Indium	3.4	2.3×10^4
Lead	7.2	6.5×10^4
Mercury	4.0	3.0×10^4
Tin	3.7	2.5×10^4
Vanadium	5.3	10.5×10^4
Zinc	0.9	0.4×10^4

Table 4. Color Code for Fixed Resistors

	First Stripe	Second Stripe	Third Stripe	Fourth Stripe
Color	*First Digit*	*Second Digit*	*Multiplier*	*Tolerance* (%)
Silver	—	—	10^{-2}	±10
Gold	—	—	10^{-1}	±5
Black	0	0	1	—
Brown	1	1	10	±1
Red	2	2	10^{2}	±2
Orange	3	3	10^{3}	—
Yellow	4	4	10^{4}	—
Green	5	5	10^{5}	—
Blue	6	6	10^{6}	—
Violet	7	7	10^{7}	—
Gray	8	8	10^{8}	—
White	9	9	10^{9}	—

Dielectric Material and Components

Dielectrics and Capacitors. Dielectric materials are insulators, having resistivities larger than $1 \times 10^{12}\ \Omega \cdot \mathrm{m}$ and containing less than 1×10^{6} mobile electrons per m^3. When a voltage is applied across a medium of dielectric material sandwiched between two electrode plates, a charge polarization takes place at the two electrodes. The resulting charge depends on the capacitance of the capacitor formed in this manner. In the linear range, the following relationship holds:

$$q = Cv$$

where v is applied voltage (V), q is stored charge (C), and C is capacitance (F). Since current (i) is the rate of change of charge ($\mathrm{d}q/\mathrm{d}t$), we can write

$$i = C\,\mathrm{d}v/\mathrm{d}t$$

Hence, in the frequency domain (substitute $j\omega$ for the rate of change operator), we have

$$i = C\,j\omega\,v$$

and the electrical impedance (v/i in the frequency domain) of a capacitor is given by

$$1/(j\omega C)$$

where ω is the frequency variable, and $j = \sqrt{-1}$.

Permittivity. Consider a capacitor made of a dielectric plate of thickness d sandwiched between two conducting plates (electrodes) of common (facing) area A. Neglecting the fringe effect, its capacitance is given by

$$C = \epsilon A/d$$

where ϵ is the permittivity of the dielectric material. The relative permittivity (or dielectric constant) ϵ_r is defined as

$$\epsilon_r = \epsilon/\epsilon_0$$

where ϵ_0 = permittivity of vacuum (approx. 8.85×10^{-12} F/m). Relative permittivities of some materials are given in Table 5.

Table 5. Dielectric Constants of Some Materials

Material	Relative Permittivity ϵ_r
Air	1.0006
Carbon dioxide gas	1.001
Ceramic (high permittivity)	8000.0
Cloth	5.0
Common salt	5.9
Diamond	5.7
Glass	6.0
Hydrogen (liquid)	1.2
Mica	6.0
Oil (mineral)	3.0
Paper (dry)	3.0
Paraffin wax	2.2
Polythene	2.3
PVC	6.0
Porcelain	6.0
Quartz (SiO_2)	4.0
Vacuum	1.0
Water	80.0
Wood	4.0

Capacitor Types. The capacitance of a capacitor is increased by increasing the common surface area of the electrode plates. This increase can be achieved, without excessively increasing the size of the capacitor, by employing a rolled-tube construction. Here, a dielectric sheet (e.g., paper or a polyester film) is placed between two metal foils, and the composite is rolled into a tube. Axial or radial leads are provided for external connection. If the dielectric material is not flexible (e.g., mica), a stacked-plate construction may be employed in place of the rolled construction to obtain compact capacitors having high capacitance. High permittivity ceramic disks are used as the dielectric plates in miniature, single-plate, high-capacitance capacitors. Electrolytic capacitors can be constructed using the rolled-tube method, using a paper soaked in an electrolyte in place of the dielectric sheet. When a voltage is applied across the capacitor, the paper becomes coated with a deposit of dielectric oxide which is formed through electrolysis. This becomes the dielectric medium of the capacitor. Capacitors having low capacitances of the order of 1×10^{-12} F (1 pF), and high capacitances of the order of 4×10^{-3} F are commercially available.

An important specification for a capacitor is the breakdown voltage which is the voltage at which discharge will occur through the dielectric medium (i.e., the dielectric medium ceases to function as an insulator). This is measured in terms of the dielectric strength, which is defined as the breakdown voltage for a dielectric element of thickness 1 mil (1×10^{-3} in). Approximate dielectric strengths of several useful materials are given in Table 6.

Table 6. Approximate Dielectric Strengths of Several Materials

Material	Dielectric Strength (V/mil)
Air	25
Ceramics	1000
Glass	2000
Mica	3000
Oil	400
Paper	1500

Table 7. Color Code for Ceramic and Paper Capacitors

	End Color	First Dot	Second Dot	Third Dot	Fourth Dot Tolerance	
Color	*Temp. Coeff. ppm/°C*	*First Digit*	*Second Digit*	*Multiplier*	*for* ≤10 pF	*for* >10 pF
Black	0	0	0	1	±2 pF	±20%
Brown	−30	1	1	10	±0.1 pF	±1%
Red	−80	2	2	1×10^2	—	±2%
Orange	−150	3	3	1×10^3	—	±2.5%
Yellow	−220	4	4	1×10^4	—	—
Green	−330	5	5	—	±0.5 pF	±5%
Blue	−470	6	6	—	—	—
Violet	−750	7	7	—	—	—
Gray	30	8	8	0.01	±0.25 pF	—
White	100	9	9	0.1	±1 pF	±10%

Color Code for Fixed Capacitors. Color codes are used to indicate the specifications of a paper or ceramic capacitor. The code consists of a colored end followed by a series of four dots printed on the outer surface of the capacitor. The end color gives the temperature coefficient of the capacitance in parts per million per degree centigrade (ppm/°C). The first two dots specify a two-digit number. The third dot specifies a multiplier which, together with the two-digit number, gives the capacitance value of the capacitor in pF. The fourth dot gives the tolerance of the capacitance. This code is shown in Table 7.

Piezoelectricity. Some materials, when subjected to a stress (strain), produce an electric charge. These are termed piezoelectric materials, and the effect is called piezoelectricity. Most materials that possess a nonsymmetric crystal structure are known to exhibit the piezoelectric effect. Examples are barium titanate, cadmium sulphide, lead zirconate titanate, quartz, and rochelle salt. The reverse piezoelectric effect (the material deforms in an electric field) is also useful in practice.

The piezoelectric characteristic of a material may be represented by its piezoelectric coefficient, k_p, which is defined as

$$k_p = \frac{\text{Change in strain (m/m)}}{\text{Change in electric field strength (V/m)}}$$

with no applied stress. Piezoelectric coefficients of some common materials are given in Table 8.

Applications of piezoelectric materials include actuators for ink-jet printers, miniature step motors, force sensors, precision shakers, high-frequency oscillators, and acoustic amplifiers. Note that large k_p values are desirable in piezoelectric actuators. For instance, PZT (lead zirconate titanate) is used in microminiature step motors (1). On the other hand, small k_p values are desirable in piezoelectric sensors (e.g., quartz accelerometers).

Table 8. Piezoelectric Coefficients of Some Materials

Material	Piezoelectric Coefficient k_p (m/V)
Barium titanate	2.5×10^{-10}
Lead zirconate titanate (PZT)	6.0×10^{-10}
Quartz	0.02×10^{-10}
Rochelle salt	3.5×10^{-10}

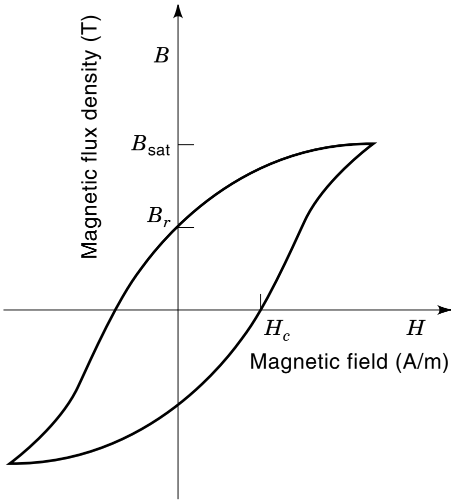


Figure 2. Hysteresis curve (magnetization curve) of a magnetic material.

Magnetic Material and Components

Magnetism and Permeability. When electrons move (or spin), a magnetic field is generated. The combined effect of such electron movements is the cause of magnetic properties of material.

In the linear range of operation of a magnetic element, we can write

$$B = \mu H$$

where B is the magnetic flux density (Wb/m^2 or T), H is magnetic field strength (A/m), and μ is the permeability of the magnetic material. The relative permeability μ_r of a magnetic material is defined as

$$\mu = \mu/\mu_0$$

where μ_0 is the permeability of a vacuum (approx. $4\pi \times 10^{-7}$ H/m). (Note: 1 T = 1 Wb/m^2; 1 H = 1 Wb/A.)

Hysteresis Loop. The B versus H curve of a magnetic material is not linear and exhibits a hysteresis loop as shown in Fig. 2. It follows that μ is not a constant. Initial values (when magnetization is started at the demagnetized state of $H = 0$ and $B = 0$) are usually specified. Some representative values are given in Table 9.

Table 9. Initial Relative Permeability (Approximate) of Some Materials

Material	Relative Permeability μ_r
Alnico (Fe$_2$ Ni Al)	6.5
Carbon steel	20
Cobalt steel (35% Co)	12
Ferrite (manganese-zinc)	800–10,000
Iron	200
Permalloy (78% Ni, 22% Fe)	3000
Silicon iron (grain oriented)	500–1,500

Table 10. Parameters of Some Magnetic Materials

Material	H_c (A/m)	B_r (Wb/m^2)
Alnico	4.6×10^4	1.25
Ferrites	14.0×10^4	0.65
Steel (carbon)	0.4×10^4	0.9
Steel (35% Co)	2.0×10^4	1.4

Properties of magnetic materials can be specified in terms of parameters of the hysteresis curve. Some important parameters are shown in Fig. 2:

$$H_c = \text{coercive field or coercive force (A/m)}$$

$$B_r = \text{remnant flux density (Wb/m}^2\text{ or T)}$$

$$B_{\text{sat}} = \text{saturation flux density (T)}$$

Magnetic parameters of a few permanent-magnetic materials are given in Table 10. Note that high values of H_c and B_r are desirable for high-strength permanent magnets. Furthermore, high values of μ are desirable for core materials that are used to concentrate magnetic flux.

Magnetic Materials. Magnetic characteristics of a material can be imagined as if contributed by a matrix of microminiature magnetic dipoles. Paramagnetic materials (e.g., platinum and tungsten) have their magnetic dipoles arranged in a somewhat random manner. These materials have a μ_r value approximately equal to 1, (i.e., no magnetization). Ferromagnetic materials (e.g., iron, cobalt, nickel, and some manganese alloys) have their magnetic dipoles aligned in one direction (parallel) with virtually no cancellation of polarity. These materials have a high μ_r (of the order of 1000) in general. At low H values, μ_r will be correspondingly low. Antiferromagnetic materials (e.g., chromium and manganese) have their magnetic dipoles arranged in parallel, but in an alternately opposing manner thereby virtually canceling the magnetization ($\mu_r = 1$). Ferrites have parallel magnetic dipoles arranged alternately opposing, as in antiferromagnetic materials, but the adjacent dipoles have unequal strengths. Hence, there is a resultant magnetization (μ_r is of the order of 1000).

Applications of magnets and magnetic materials include actuators (e.g., motors, magnetically leviated vehicles, tools, magnetic bearings), sensors and transducers, relays, resonators, and cores of inductors and transformers. Also, see the applications of superconductivity.

Piezomagnetism. When a stress (strain) is applied to a piezomagnetic material, the degree of magnetization of the material changes. Conversely, a piezomagnetic material undergoes deformation when the magnetic field in which the material is situated is changed.

Hall Effect Sensors. Suppose that a dc voltage v_{ref} is applied to a semiconductor element that is placed in a magnetic field in an orthogonal direction, as shown in Fig. 3. A voltage v_0 is generated in the third orthogonal direction, as indicated in the figure (1). This is known as the Hall effect. Hall effect sensors use this phenomenon. For example, the motion of a ferromagnetic element can be detected in this manner since the magnetic field in which the sensor is mounted would vary as a result of the motion of the ferromagnetic element. Hall effect sensors are useful as position sensors, speed sensors, commutation devices for motors, and instrument transformers for power transmission systems.

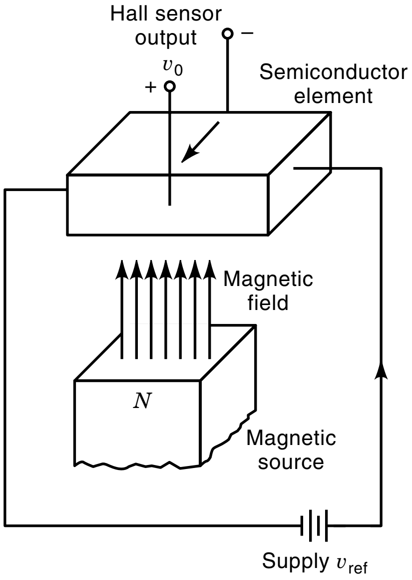


Figure 3. A Hall effect sensor.

Magnetic Bubble Memories. Consider a film of magnetic material such as gadolinium gallium oxide ($Gd_3Ga_5O_{12}$) deposited on a nonmagnetic garnet layer (substrate). The direction of magnetization will be perpendicular to the surface of the film. Initially, some regions of the film will be N poles, and the remaining regions will be S poles. An external magnetic field can shrink either the N regions or the S regions, depending on the direction of the field. The size of the individual magnetic regions can be reduced to the order of 1 μm in this manner. These tiny magnetic bubbles are the means with which information is stored in a magnetic bubble memory.

Inductance. Suppose that a conducting coil having n turns is placed in a magnetic field of flux ϕ (Wb). The resulting flux linkage is $n\phi$. If the flux linkage is changed, a voltage is induced in the coil. This induced voltage (v) is given by

$$v = \mathrm{d}(n\phi)/\mathrm{d}t = n\,\mathrm{d}\phi/\mathrm{d}t$$

If the change in magnetic flux is brought about by a change in current (i), we can write

$$v = L\,\mathrm{d}i/\mathrm{d}t$$

where L is the inductance of the coil (H).

In the frequency domain, we have

$$v = L\,j\omega\,i$$

where ω = frequency and $j = \sqrt{-1}$.

It follows that the electrical impedance of an inductor is given by $j\omega\,L$.

ACTIVE COMPONENTS

Active components made of semiconductor junctions, and field effect components are considered in this section. Junction diodes, bipolar junction transistors, and field-effect transistors are of particular interest here. Active components are widely used in the monolithic (integrated-circuit) form as well as in the form of discrete elements.

pn Junctions

A pure semiconductor can be doped to form either a *p*-type semiconductor or an *n*-type semiconductor. A *pn* junction is formed by joining a *p*-type semiconductor element and an *n*-type semiconductor element.

Semiconductors. Semiconductor materials have resistivities that are several million times larger than those of conductors and several billion times smaller than those of insulators. Crystalline materials such as silicon and germanium are semiconductors. For example, the resistivity of pure silicon is about 5×10^{10} times that of silver. Similarly, the resistivity of pure germanium is about 5×10^{7} times that of silver. Typically, semiconductors have resistivities ranging from 10^{-4} to $10^{7}\ \Omega \cdot \mathrm{m}$. Other examples of semiconductor materials are gallium arsenide, cadmium sulfide, and selenium.

A pure (intrinsic) semiconductor material has some free electrons (negative charge carriers) and holes (positive charge carriers). Note that a hole is formed in an atom when an electron is removed. Strictly, the holes cannot move. But suppose that an electron shared by two atoms (a covalent electron) enters an existing hole in an atom, leaving behind a hole at the point of origin. The resulting movement of the electron is interpreted as a movement of a hole in the direction opposite to the actual movement of the covalent electron.

The number of free electrons in a pure semiconductor is roughly equal to the number of holes. The number of free electrons or holes in a pure semiconductor can be drastically increased by adding traces of impurities in a controlled manner (doping) into the semiconductor during crystal growth (e.g., by alloying in a molten form, and by solid or gaseous diffusion of the trace). An atom of a pure semiconductor that has four electrons in its outer shell will need four more atoms to share in order to form a stable covalent bond. These covalent bonds are necessary to form a crystalline lattice structure of atoms which is typical of semiconductor materials. If the trace impurity is a material such as arsenic, phosphorus, or antimony whose atoms have five electrons in the outer shell (a donor impurity), a free electron will be left over after the formation of a bond with an impurity atom. The result will be an *n*-type semiconductor having a very large number of free electrons. If, on the other hand, the trace impurity is a material such as boron, gallium, aluminum or indium whose atoms have only three electrons in the outer shell (an acceptor impurity), a hole will result on formation of a bond. In this case, a *p*-type semiconductor, consisting of a very large number of holes, will result. Doped semiconductors are termed extrinsic.

Depletion Region. When a *p*-type semiconductor is joined with an *n*-type semiconductor, a *pn* junction is formed. A *pn* junction exhibits the diode effect, much larger resistance to current flow in one direction than in the opposite direction across the junction. As a *pn* junction is formed, electrons in the *n*-type material in the neighborhood of the common layer will diffuse across into the *p*-type material. Similarly, the holes in the *p*-type material near the junction will diffuse into the opposite side (strictly, the covalent electrons will diffuse in the opposite direction). The diffusion will proceed until an equilibrium state is reached. But, as a result of the loss of electrons and the gain of holes on the *n* side and the opposite process on the *p* side, a potential difference is generated across the *pn* junction, with a negative potential on the *p* side and a positive potential on the *n* side. Due to the diffusion of carriers across the junction, the small region surrounding the common area will be virtually free of carriers (free electrons and holes). Hence, this region is called the depletion region. The potential difference that exists in the depletion region is mainly responsible for the diode effect of a *pn* junction.

Biasing. The forward biasing and the reverse biasing of a *pn* junction are shown in Fig. 4. In the case of forward biasing, a positive potential is connected to the *p* side of the junction, and a negative potential is connected to the *n* side. The polarities are reversed for reverse biasing. Note that in forward biasing, the external voltage (bias voltage v) complements the potential difference of the depletion region [Figure 4(a)]. The free electrons that crossed over to the *p* side from the *n* side will continue to flow toward the positive terminal of the external supply, thereby generating a current (junction current i). The junction current increases with the bias voltage, as shown in Fig. 4(c).

In reverse biasing, the potential in the depletion region is opposed by the bias voltage [Fig. 4(b)]. Hence, the diffusion of free electrons from the *n* side into the *p* side is resisted. Since there are some (very few) free electrons in the *p* side and some holes in the *n* side, the reverse bias will reinforce the

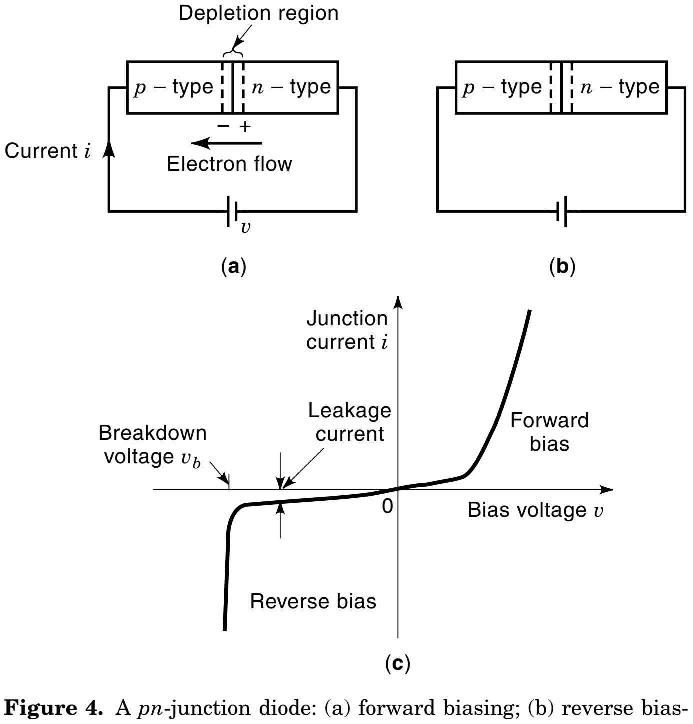


Figure 4. A *pn*-junction diode: (a) forward biasing; (b) reverse biasing; (c) characteristic curve.

Table 11. Typical Breakdown Voltage of *pn* Junction at Room Temperature

Semiconductor	Breakdown Voltage (V) *Dope Concentration* = 10^{15} atoms/cm^3	 *Dope Concentration* = 10^{17} atoms/cm^3
Germanium	400	5.0
Silicon	300	11.0
Gallium arsenide	150	16.0

flow of these minority electrons and holes. This will create a very small current (about 10^{-9} A for silicon and 10^{-6} A for germanium at room temperature), known as the *leakage current,* in the opposite direction to the forward-bias current. If the reverse bias is increased, at some voltage [breakdown voltage v_b in Fig. 4(c)] the junction will break down, generating a sudden increase in the reverse current. There are two main causes of this breakdown. First, the intense electric field of the external voltage can cause electrons to break away from neutral atoms in large numbers. This is known as zener breakdown. Second, the external voltage will accelerate the minority free electrons on the *p* side (and minority holes on the *n* side), creating collisions that will cause electrons on the outer shells of neutral atoms to break away in large numbers. This is known as the avalanche breakdown. In some applications (e.g., rectifier circuits), junction breakdown is detrimental. In some other types of applications (e.g., as constant voltage sources and in some digital circuits), the breakdown state of specially designed diodes is practically utilized. Typical breakdown voltages of *pn* junctions made of three common semiconductor materials are given in Table 11. Note that the breakdown voltage decreases with the concentration of the trace material.

The current through a reverse-biased *pn* junction will increase exponentially with temperature. For a forward-biased *pn* junction, current will increase with temperature at low to moderate voltages and will decrease with temperature at high levels of voltage.

Diodes

A semiconductor diode is formed by joining a *p*-type semiconductor with an *n*-type semiconductor. A diode offers much less resistance to current flow in one direction (forward) than in the opposite direction (reverse). There are many varieties of diodes. Zener diodes, VVC diodes, tunnel diodes, microwave power diodes, pin diodes, photodiodes, and light-emitting diodes (LED) are examples. The last two varieties will be discussed in separate sections.

Zener Diodes. Zener diodes are a particular type of diodes that are designed to operate in the neighborhood of the reverse breakdown (both zener and avalanche breakdowns). In this manner, a somewhat constant voltage output (the breakdown voltage) can be generated. This voltage depends on the concentration of the trace impurity. By varying the impurity concentration, output voltages in the range of 2 to 200 V may be realized from a zener diode. Special circuits would be needed to divert large currents that are generated at the breakdown point of the diode. The rated power dissipation of a zener diode should take into consideration the current levels that are possible in the breakdown region. Applications of zener diodes include constant voltage sources, voltage clipper circuits, filter circuits for voltage transients, digital circuits, and two-state devices.

VVC Diodes. Voltage variable capacitor (VVC) diodes use the property of a diode that, in reverse bias, the capacitance decreases (nonlinearly) with the bias voltage. The depletion region of a *pn* junction is practically free of carriers (free electrons and holes) and, hence, behaves like the dielectric medium of a capacitor. The adjoining *p*-region and *n*-region serve as the two plates of the capacitor. The width of the depletion region increases with the bias voltage. Consequently, the capacitance of a reverse biased *pn* junction decreases as the bias voltage is increased. The obtainable range of capacitance can be varied by changing the dope concentration and also by distributing the dope concentration nonuniformly along the diode. For example, a capacitance variation of 5 to 500 pF may be obtained in this manner (note: 1 pF = 1×10^{-12} F). VVC diodes are also known as varactor diodes and varicaps and are useful in voltage-controlled tuners and oscillators.

Tunnel Diodes. The depletion of a *pn* junction can be made very thin by using very high dope concentrations (in both the *p* and *n* sides). The result is a tunnel diode. Since the depletion region is very narrow, charge carriers (free electrons and holes) in the *n* and *p* sides of the diode can tunnel through the region into the opposite side on application of a relatively small voltage. The voltage-current characteristic of a tunnel diode is quite linear at low (forward and reverse) voltages. When the forward bias is further increased, however, the behavior will become very nonlinear; the junction current will peak, then drop (a negative conductance) to a minimum (valley), and finally rise again, as the voltage is increased. Due to the linear behavior of the tunnel diode at low voltages, almost instantaneous current reversal (i.e., very low reverse recovery time) can be achieved by switching the bias voltage. Tunnel diodes are useful in high-frequency switching devices, sensors, and signal conditioning circuits.

***pin* Diodes.** The width of the depletion region of a conventional *pn* junction varies with many factors, primarily the applied (bias) voltage. The capacitance of a junction depends on this width and will vary due to such factors. A diode with practically a constant capacitance is obtained by adding a layer of silicon in between the *p* and *n* elements. The sandwiched silicon layer is called the intrinsic layer, and the diode is called a *pin* diode. The resistance of a *pin* diode varies inversely with junction current. *Pin* diodes are useful as current-controlled resistors at constant capacitance.

Schottky Barrier Diodes. Most diodes consist of semiconductor-semiconductor junctions. An exception is a Schottky barrier diode which consists of a metal-semiconductor (*n*-type) junction. A metal such as gold, silver, platinum, or palladium and a semiconductor such as silicon or gallium arsenide may be used in the construction. Since no holes exist in the metal, a depletion region cannot be formed at the metal-semiconductor junction. Instead, an electron barrier is formed by the free electrons from the *n*-type semiconductor. Consequently, the junction capacitance will be negligible, and the reverse recovery time will be very small. For this reason, Schottky diodes

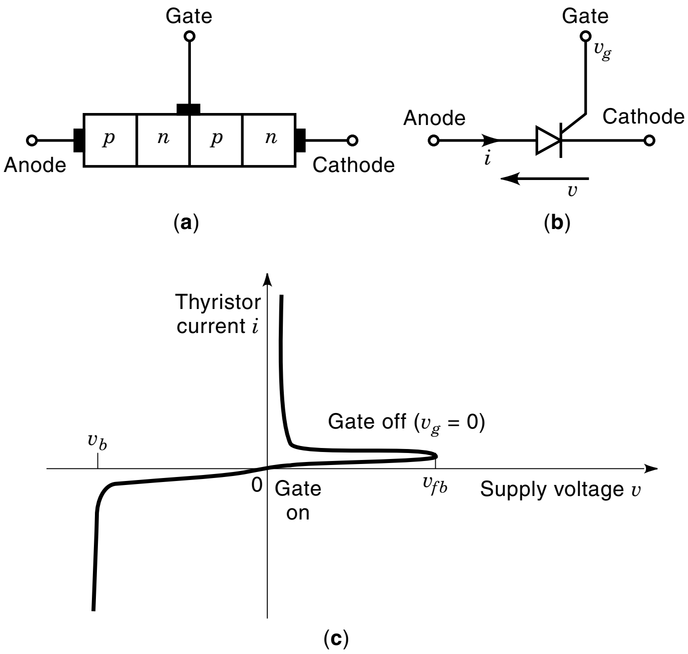


Figure 5. The thyristor: (a) schematic representation; (b) circuit symbol; (c) characteristic curve.

can handle very high switching frequencies (10^9 Hz range). Since the electron barrier is easier to penetrate than a depletion region, by using a reverse bias, Schotky diodes exhibit much lower breakdown voltages. Operating noise is also lower than for semiconductor-semiconductor diodes.

Thyristors. A thyristor, also known as a silicon-controlled rectifier, a solid-state controlled rectifier, a semiconductor-controlled rectifier, or simply an SCR, possesses some of the characteristics of a semiconductor diode. It consists of four layers (*pnpn*) of semiconductor and has three terminals—the anode, the cathode, and the gate—as shown in Fig. 5(a). The circuit symbol for a thyristor is shown in Fig. 5(b). The thyristor current is denoted by i, the external voltage is v, and the gate potential is v_g. The characteristic curve of a thyristor is shown in Fig. 5(c). Note that a thyristor cannot conduct in either direction (i almost zero) until either the reverse voltage reaches the reverse breakdown voltage (v_b), or the forward voltage reaches the forward breakover voltage (v_{fb}). The forward breakover is a bistable state, and once this voltage is reached, the voltage drops significantly, and the thyristor begins to conduct like a forward-biased diode. When v_g is less than or equal to zero with respect to the cathode, v_{fb} becomes quite high. When v_g is made positive, v_{fb} becomes small, and v_{fb} will decrease as the gate current (i_g) is increased. A small positive v_g can make v_{fb} very small, and then the thyristor will conduct from anode to cathode but not in the opposite direction (i.e., it behaves like a diode). It follows that a thyristor behaves like a voltage-triggered switch; a positive firing signal (a positive v_g) will close the switch. The switch will be opened when both i and v_g are made zero. When the supply voltage v is dc and nonzero, the thyristor will not be able to turn itself off. In this case a commutating circuit that can make the trigger voltage v_g slightly negative has to be employed. Thyristors are commonly used in control circuits for dc and ac motors.

Parameter values for diodes are given in data sheets provided by the manufacturer. Commonly used variables and characteristic parameters in association with diodes are described in Table 12. For thyristors, as mentioned before, several other quantities such as v_{fb}, v_g, and i_g should be included. The time required for a thyristor to be turned on by the trigger signal (turn-on time) and the time for it to be turned off through commutation (turn-off time) determine the maximum switching frequency (bandwidth) for a thyristor. Another variable that is important is the holding current or latching current, which denotes the small forward current that exists at the breakover voltage.

Bipolar Junction Transistors

A bipolar junction transistor (BJT) has two junctions which are formed by joining p regions and n regions. Two types of transistors, *npn* and *pnp*, are possible with this structure. A BJT has three terminals, as indicated in Fig. 6(a). The middle (sandwiched) region of a BJT is thinner than the end regions, and this region is known as the base. The end regions are termed the emitter and the collector. Under normal conditions, the emitter-base junction is forward biased, and the collector-base junction is reverse biased, as shown in Fig. 6(b).

Table 12. Characteristic Variables and Parameters for Diodes

Diode Variable/Parameter		Description
Forward bias	(v_f)	A positive external voltage at p with respect to n
Reverse bias	(v_r)	A positive external voltage at n with respect to p
Breakdown voltage	(v_b)	The minimum reverse bias that will break down the junction resistance
Junction current	(i_f)	Forward current through a forward-biased diode
Leakage current	(i_r)	Reverse current through a reverse-biased diode
Transition capacitance	(C_t)	Capacitance (in the depletion region) of a reverse-biased diode
Diffusion capacitance	(C_d)	Capacitance exhibited while a forward biased diode is switched off
Forward resistance	(R_f)	Resistance of a forward-biased diode
Reverse recovery time	(t_{rr})	Time needed for the reverse current to reach a specified level when the diode is switched from forward to reverse
Operating temperature range	(T_A)	Allowable temperature range for a diode during operation
Storage temperature range	(T_{srg})	Temperature that should be maintained during storage of a diode
Power dissipation	(P)	The maximum power dissipation allowed for a diode at a specified temperature

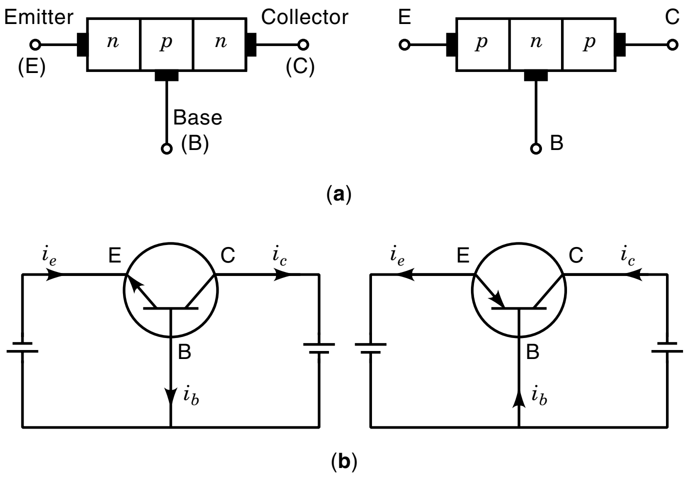


Figure 6. Bipolar junction transistors: (a) *npn* and *pnp* transistors; (b) circuit symbols and biasing.

To explain the behavior of a BJT, consider an *npn* transistor under normal biasing. The forward bias at the emitter-base junction will cause free electrons in the emitter to flow into the base region, thereby creating the emitter current (i_e). The reverse bias at the collector-base junction will increase the depletion region there. The associated potential difference at the collector-base junction will accelerate the free electrons in the base into the collector and will form the collector current (i_c). Holes that are created in the base, for recombination with some free electrons that entered the base, will form the base current (i_b). Usually, i_c is slightly smaller than i_e. Furthermore, i_b is much smaller than i_c.

Transistor Characteristics. The common-emitter connection is widely used for transistors in amplifier applications. In this configuration, the emitter terminal will be common to the input side and the output side of the circuit. Transistor characteristics are usually specified for this configuration. Figure 7 shows typical characteristic curves for a junction transistor in the common-emitter connection. In this configuration, both voltage gain (output voltage/input voltage) and current gain (collector current/base current) will be greater than unity, thereby providing a voltage amplification as well as a current amplification. Note from Fig. 7 that the control signal is the base current (i_b), and the characteristic of the transistor depends on i_b. This is generally true for any bipolar junction transistor; a BJT is a current-controlled transistor. In the common-base configuration, the base terminal is common to both input and output.

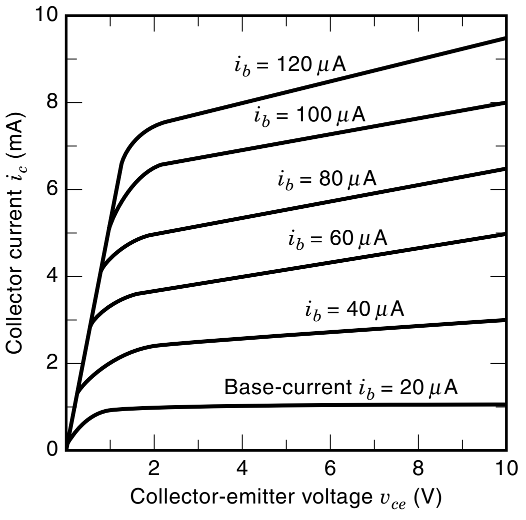


Figure 7. Characteristic curves of a common emitter BJT.

Table 13. Rating Parameters for Transistors

Transistor Parameter		Description
Collector to base voltage	(v_{cb})	Voltage limit across collector and base with emitter open
Collector to emitter voltage	(v_{ce})	Voltage limit across collector and emitter with base connected to emitter
Emitter to base voltage	(v_{eb})	Voltage limit across emitter and base with collector open
Collector cut off current	(i_{co})	Reverse saturation current at collector with either emitter open (i_{cbo}) or base open (i_{ceo})
Transistor dissipation	(P_T)	Power dissipated by the transistor at rated conditions
Input impedance	(h_i)	Input voltage/input current with output voltage = 0 (Defined for both common emitter and common base configurations, h_{ie}, h_{ib})
Output admittance	(h_o)	Output current/output voltage with input current = 0 (h_{oe}, h_{ob} are defined)
Forward current transfer ratio	(h_f)	Output current/input current with output voltage = 0 ($h_{fe,}$ h_{fb} are defined)
Reverse voltage transfer ratio	(h_r)	Input voltage/output voltage with input current = 0 ($h_{re,}$ h_{rb} are defined)
Rise time	(t_r)	Time taken to reach the full current level for the first time when turned on
Storage time	(t_s)	Time taken to reach the steady current level when turned on
Fall time	(t_f)	Time taken for the current to reach zero when turned off

Maximum frequency of operation and allowable switching rate for a transistor are determined by parameters such as rise time, storage time, and fall time. These and some other useful ratings and characteristic parameters for bipolar junction transistors are defined in Table 13. Values for these parameters are normally given in the manufacturer's data sheet for a particular transistor.

Fabrication Process. The actual manufacturing process for a transistor is complex and delicate. For example, an *npn* transistor can be fabricated by starting with a crystal of *n*-type silicon. This starting element is called the wafer or substrate. The *npn* transistor is formed, by using the planar diffusion method, in the top half of the substrate as follows: The substrate is heated to about 1000°C. A gas stream containing a donor-type impurity (which forms *n*-type regions) is im-

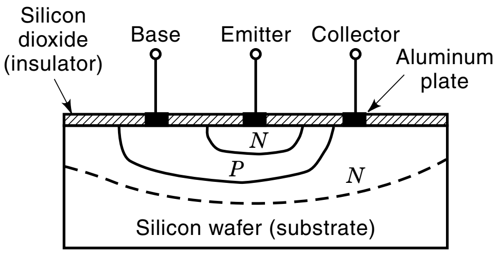


Figure 8. An *npn* transistor manufactured by the planar diffusion method.

pinged on the crystal surface. This produces an *n*-type layer on the crystal. Next the crystal is oxidized by heating to a high temperature. The resulting layer of silicon dioxide acts as an insulating surface. A small area of this layer is then dissolved off using hydrofluoric acid. The crystal is again heated to 1000°C, and a gas stream containing acceptor-type impurity (which forms *p*-type regions) is impinged on the window thus formed. This produces a *p* region under the window on top of the *n* region which was formed earlier.

Oxidation is repeated to cover the newly formed *p* region. Using hydrofluoric acid, a smaller window is cut on the latest silicon dioxide layer, and a new *n* region is formed, as before, on top of the *p* region. The entire manufacturing process has to be properly controlled so as to control the properties of the resulting transistor. Aluminum contacts have to be deposited on the uppermost *n* region, the second *p* region (in a suitable annular window cut on the silicon dioxide layer), and on the *n* region below it or on the crystal substrate. A pictorial representation of an *npn* transistor fabricated in this manner is shown in Fig. 8.

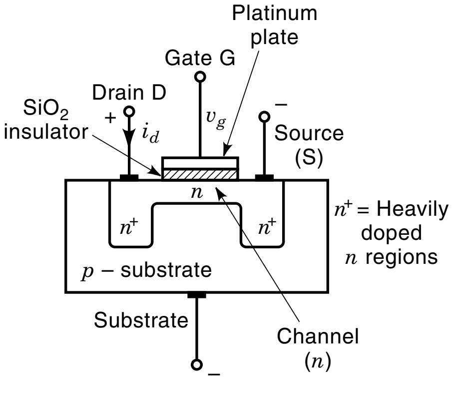
Field Effect Transistors

A field effect transistor (FET), unlike a bipolar junction transistor (BJT), is a voltage-controlled transistor. The electrostatic field generated by a voltage applied to the gate terminal of a FET controls the behavior of the FET. Since the device is voltage controlled at very low input current levels, the input impedance is very high, and the input power is very low. Other advantages of an FET over a BJT are that the former is cheaper and requires significantly less space on a chip in the monolithic form. FETs are somewhat slower (in terms of switching rates) and more nonlinear than BJTs, however.

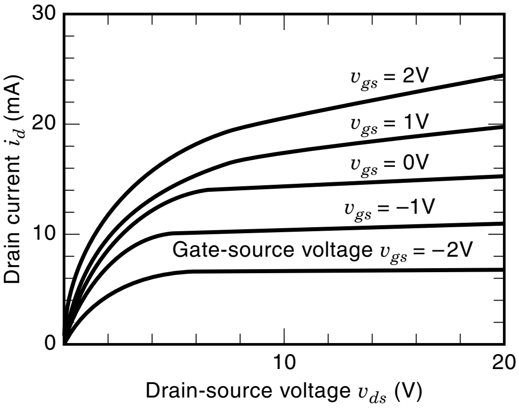
There are two primary types of FETs: metal oxide semiconductor field effect transistor (MOSFET) and junction field effect transistor (JFET). Even though the physical structure of the two types is somewhat different, their characteristics are quite similar. Insulated gate FET (or IGFET) is a general name given to MOSFETs.

The MOSFET. An *n*-channel MOSFET is produced using a *p*-type silicon substrate, and a *p*-channel MOSFET by an *n*-type substrate. An *n*-channel MOSFET is shown in Fig. 9(a). During manufacture, two heavily doped *n*-type regions are formed on the substrate. One region is termed source (S) and the other region drain (D). The two regions are connected by a moderately doped and narrow *n* region called a channel. A metal coating deposited over an insulating layer of silicon dioxide which is formed on the channel is the gate (G). The source lead is usually joined with the substrate lead. This is a depletion-type MOSFET (or D-MOSFET). Another type is the enhancement-type MOSFET (or E-MOSFET). In this type, a channel linking the drain and the source is not physically present in the substrate but is induced during operation of the transistor.

Consider the operation of the *n*-channel D-MOSFET shown in Fig. 9(a). Under normal operation, the drain is positively biased with respect to the source. Drain current i_d is considered the output of a MOSFET (analogous to the collector current of a BJT). The control signal of a MOSFET is the gate voltage v_{gs} with respect to the source (analogous to the base current of a BJT). It follows that a MOSFET is a voltage-controlled device. Since the source terminal is used as the reference for both input (gate voltage) and output (drain), this connection is called the common-source configuration. Suppose that the gate voltage is negative with respect to the source. This will induce holes in the channel, thereby decreasing the free electrons there through recombination. This, in turn, will reduce the concentration of free electrons in the drain region and, hence, will reduce the drain current i_d. Clearly, if the magnitude of the negative voltage at the gate is decreased, the drain current will increase, as indicated by the characteristic curves in Fig. 9(b). A positive bias at the gate will further increase the drain current of an *n*-channel MOSFET as shown. The opposite will be true for a *p*-channel MOSFET.



(a)



(b)

Figure 9. A metal oxide semiconductor FET: (a) an *n*-channel depletion-type MOSFET; (b) D-MOSFET characteristics.

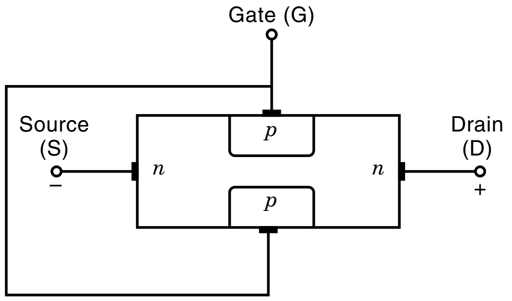


Figure 10. An *n*-channel JFET.

The JFET. A junction field effect transistor (JFET) is different in physical structure to a MOSFET but similar in characteristics. The structure of an *n*-channel JFET is shown in Fig. 10. It consists of two *p*-type regions formed inside an *n*-type region. The two *p* regions are separated by a narrow *n* region called a channel. The channel links two *n*-type regions called source (S) and drain (D). The two *p* regions are linked by a common terminal and form the gate (G). As for a MOSFET, drain current i_d is considered the output of the JFET, and gate voltage v_{gs}, with respect to the source, is considered the control signal. For normal operation, the drain is positively biased with respect to the source, as for an *n*-channel MOSFET, and the common-source configuration is used.

To explain the operation of a JFET, consider the *n*-channel JFET shown in Fig. 10. Depletion regions are present at the two *pn* junctions of the JFET (as for a semiconductor diode). If the gate voltage is made negative, the resulting field will weaken the *p* regions. As a result, the depletion regions will shrink. Some of the free electrons from the drain will diffuse toward the channel to occupy the growing *n* regions due to the shrinking depletion regions. This will reduce the drain current. It follows that drain current decreases as the magnitude of the negative voltage at the gate is increased. This behavior is similar to that of a MOSFET. A *p*-channel JFET has two *n* regions representing the gate and two *p* regions forming the source and the drain which are linked by a *p*-channel. Its characteristic is the reverse of an *n*-channel JFET.

Common types of transistor are summarized in Table 14. Semiconductor devices have numerous uses. A common use is as switching devices or as two-state elements. Typical two-state elements are schematically illustrated in Fig. 11.

LIGHT EMITTERS AND DISPLAYS

Visible light is part of the electromagnetic spectrum. Electromagnetic waves in the wave length range of 390 to 770 nm (Note: 1 nm = 1×10^{-9} m) form the visible light. Ultraviolet rays and X-rays are also electromagnetic waves, but have lower wave lengths (higher frequencies). Infrared rays, microwaves, and radio waves are electromagnetic waves having higher wave lengths. Table 15 lists wave lengths of several types of electromagnetic waves. Visible light occupies a broad range of wave lengths. For example, in optical coupling applications, the narrower the wave spectrum, the clearer (noise free) the coupling process. Consequently, it is advantageous to use special light sources in applications of that type. Furthermore, since visible light can be contaminated by environmental light, thereby introducing an error signal into the system, it is also useful to consider electromagnetic waves that are different from what is commonly present in operating environments in applications such as sensing, optical coupling, and processing.

Incandescent Lamps

Tungsten-filament incandescent lamps that are commonly used in household illumination emit visible light in a broad

Table 14. Common Transistor Types

Transistor Type		
Abbreviation	*Name*	Description
BJT	Bipolar Junction Transistor	A three-layer device (*npn* or *pnp*) Current controlled Control = base current Output = collector current
FET	Field Effect Transistor	A physical or induced channel (*n*-channel or *p*-channel) voltage controlled Control = gate voltage Output = drain current
MOSFET	Metal Oxide Semiconductor FET	*n*-channel or *p*-channel
D-MOSFET	Depletion-type MOSFET	A channel is physically present
E-MOSFET	Enhancement-type MOSFET	A channel is induced
VMOS	V-shaped Gate MOSFET or VFET	An E-MOSFET with increased power handling capacity
DG-MOS	Dual-gate MOSFET	A secondary gate is present between main gate and drain (lower capacitance)
D-MOS	Double-diffused MOSFET	A channel layer is formed on a high-resistivity substrate and then source and drain are formed (by diffusion). High breakdown voltage
CMOS	Complementary Symmetry MOSFET	Uses two E-MOSFETs (*n* channel and *p* channel). Symmetry is used to save space on chip. Cheaper and lower power consumption.
GaAs	Gallium Arsenide MOSFET	Uses gallium arsenide, aluminum gallium arsenide, (AlGaAs), indium gallium arsenide phosphide (InGaAsP), etc. in place of silicon substrate. Faster operation
JFET	Junction FET	*p*-channel or *n*-channel. Has two (*n* or *p*) regions in a (*p* or *n*) region linked by a channel (*p* or *n*) Control = gate voltage Output = drain current

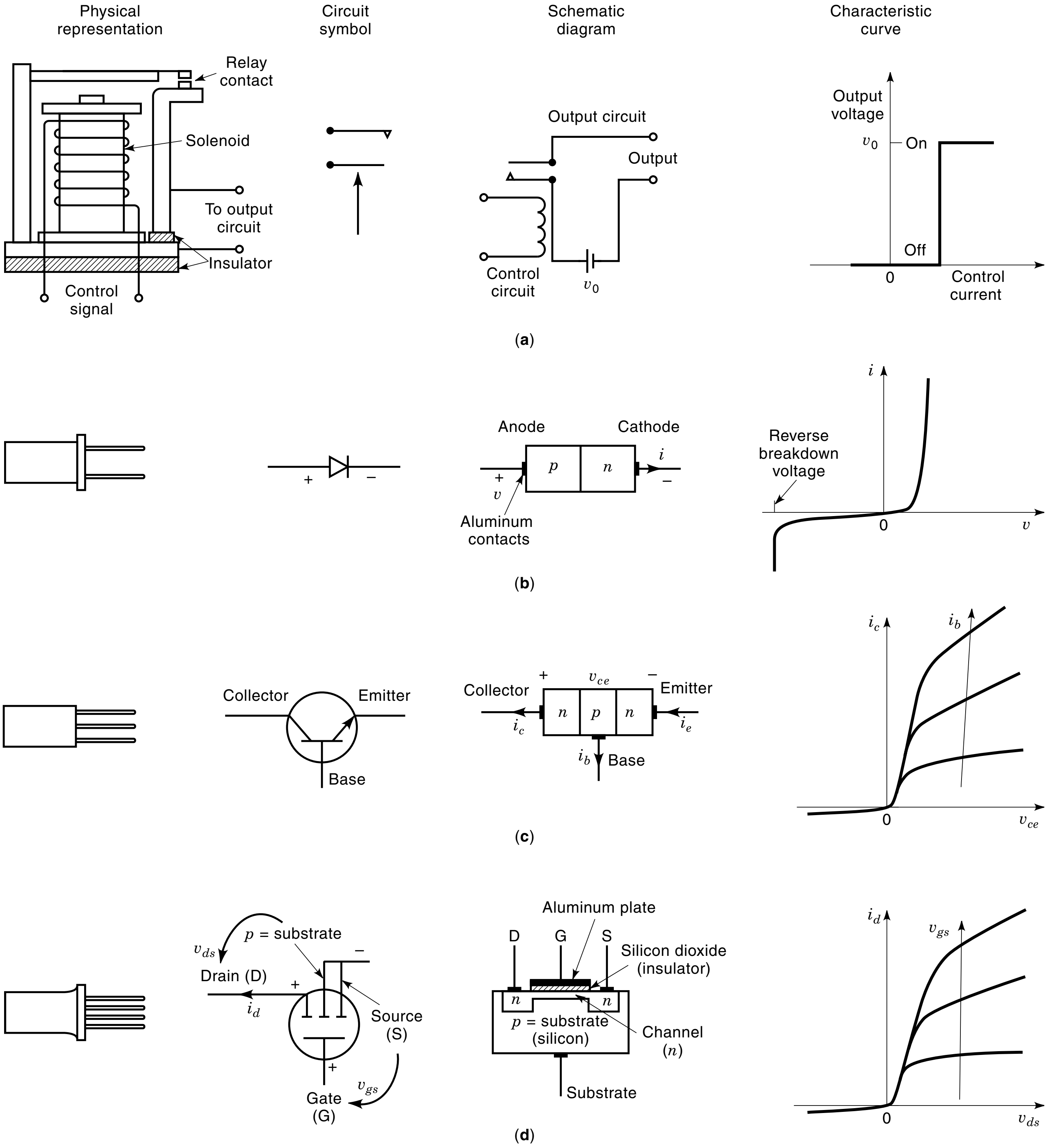


Figure 11. Discrete switching (two-state) elements: (a) electromagnetic relay; (b) zener diode; (c) bipolar junction transistor (*npn*); (d) *n*-channel MOSFET.

Table 15. Wave Lengths of Several Selected Components of the Electromagnetic Spectrum

Wave Type	Approximate Wave Length Range (μm)
Radio waves	$1 \times 10^6 - 5 \times 10^6$
Microwaves	$1 \times 10^3 - 1 \times 10^6$
Infrared rays	$0.8 - 1 \times 10^3$
Visible light	$0.4 - 0.8$
Ultraviolet rays	$1 \times 10^{-2} - 0.4$
X rays	$1 \times 10^{-6} - 5 \times 10^{-2}$

Table 16. Wave Length Characteristics of Common LEDs (1 Å = 1 × 10^{-10} m)

LED Type	Wave Length at Peak Intensity (Å)	Color
Gallium arsenide	5500	Green
	9300	Infrared
Gallium arsenide phosphide	5500	Green
	7000	Red
Gallium phosphide	5500	Green
Gallium aluminum arsenide	8000	Red
	8500	Infrared
Indium gallium arsenide phosphide	13000	Infrared

spectrum. Furthermore, they are not efficient because they emit more infrared radiation than useful visible light. Ionizing lamps filled with gases such as halogens, sodium-vapor, neon, or mercury vapor have much narrower spectra, and they emit very pure visible light (with negligible infrared radiation). Hence, these types of incandescent lamps are more efficient for illumination purposes. Regular fluorescent lamps are known to create a line-frequency (60 Hz or 50 Hz) flicker but are quite efficient and durable. All these types of light sources are usually not suitable in many applications primarily because of the following disadvantages:

1. They are bulky.
2. They cannot be operated at high switching rates (from both time constant and component life points of view).
3. Their spectral bandwidth can be very wide.

Note that a finite time is needed for an incandescent lamp to emit light once it is energized. That is, it has a large time constant. This limits the switching speed to less than 100 Hz. Furthermore, lamp life will decrease rapidly with increasing switching frequency.

Light-Emitting Diodes

The basic components of a light-emitting diode (LED) are shown in Fig. 12(a). The element symbol that is commonly used in electrical circuits is shown in Fig. 12(b). The main component of an LED is a semiconductor diode element, typically made of gallium compounds (e.g., gallium arsenide or GaAs and gallium arsenide phosphide or GaAsP). When a voltage is applied in the forward-bias direction to this semiconductor element, it emits visible light (and also other electromagnetic wave components, primarily infrared). In the forward-bias configuration, electrons are injected into the *p* region of the diode and recombined with holes. Radiation energy (including visible light) is released spontaneously in this process. This is the principle of operation of an LED. Suitable doping with trace elements such as nitrogen will produce the desired effect. The radiation energy generated at the junction of a diode has to be directly transmitted to a window of the diode in order to reduce absorption losses. Two types of construction are commonly used; edge emitters emit radiation along the edges of the *pn* junction, and surface emitters emit radiation normal to the junction surface.

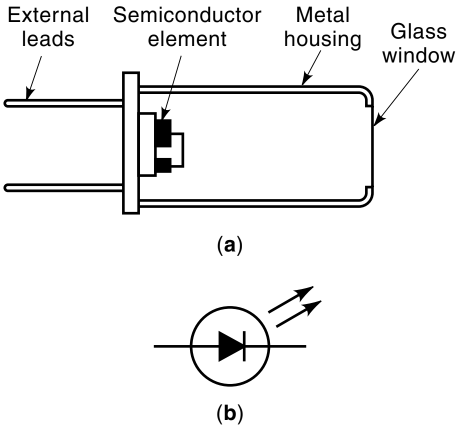


Figure 12. A light-emitting diode (LED): (a) physical construction; (b) circuit symbol.

Infrared light-emitting diodes (IRED) are LEDs that emit infrared radiation at a reasonable level of power. Gallium arsenide (GaAs), gallium aluminum arsenide (GaAlAs), and indium gallium arsenide phosphide (InGaAsP) are the commonly used IRED material. Gallium compounds and not silicon or germanium are used in LEDs for reasons of efficiency and intensity characteristics. (Gallium compounds exhibit sharp peaks of spectral output in the desired frequency bands.) Table 16 gives wave length characteristics of common LED and IRED types (1 Å = 1 × 10^{-10} m = 0.1 nm). Note that Å denotes the unit angstrom.

Light-emitting diodes are widely used in optical electronics because they can be constructed in miniature sizes, they have small time constants and low impedances, they can provide high switching rates (typically over 1000 Hz), and they have much longer component life than incandescent lamps. They are useful as both light sources and displays.

Lasers

Laser (light amplification by stimulated emission of radiation) is a light source that emits a concentrated beam of light which will propagate typically at one or two frequencies (wave lengths) and in phase. Usually, the frequency band is extremely narrow (i.e., monochromatic), and the waves in each frequency are in phase (i.e., coherent). Furthermore, the energy of a laser is highly concentrated (power densities of the order of one billion watts/cm^2). Consequently, a laser beam can travel in a straight line over a long distance with very little dispersion. Hence, it is useful in gauging and aligning applications. Lasers can be used in a wide variety of sensors (e.g., motion sensors, tactile sensors, laser-doppler velocity sensors) that employ photosensing and fiber optics. Also, lasers are used in medical applications, microsurgery in particular. Lasers have been used in manufacturing and material removal applications such as precision welding, cutting, and drilling of different types of materials, including metals, glass, plastics, ceramics, leather, and cloth. Lasers are used in inspection (detection of faults and irregularities) and gauging (measurement of dimensions) of parts. Other applications of lasers include heat treatment of alloys, holographic methods of nondestructive testing, communication, information processing, and high-quality printing.

Lasers may be classified as solid, liquid, gas, and semiconductor. In a solid laser (e.g., ruby laser, glass laser), a solid rod with reflecting ends is used as the laser medium. The laser medium of a liquid laser (e.g., dye laser, salt-solution laser) is a liquid such as an organic solvent with a dye or an inorganic solvent with dissolved salt compound. Very high peak power levels are possible with liquid lasers. Gas lasers (e.g., helium–neon or He–Ne laser, helium–cadmium or He–Cd laser, carbon dioxide or CO_2 laser) use a gas as the laser medium. Semiconductor lasers (e.g., gallium arsenide laser) use a semiconductor diode similar to an edge-emitting LED. Some lasers have their main radiation components outside the visible spectrum of light. For example, a CO_2 laser (wavelength of about 110,000 Å) primarily emits infrared radiation.

In a conventional laser unit, the laser beam is generated by first originating an excitation to create a light flash. This will initiate a process of emitting photons from molecules within the laser medium. This light is then reflected back and forth between two reflecting surfaces before the light beam is finally emitted as a laser. These waves will be limited to a very narrow frequency band (monochromatic) and will be in phase (coherent). For example, consider the He–Ne laser unit schematically shown in Fig. 13. The helium and neon gas mixture in the cavity resonator is heated by a filament lamp and ionized using a high dc voltage (2000 V). Electrons released in the process will be accelerated by the high voltage and will collide with the atoms, thereby releasing photons (light). These photons will collide with other molecules, releasing more photons. This process is known as lasing. The light generated in this manner is reflected back and forth by the silvered surface and the partially-reflective lens (beam splitter) in the cavity resonator, thereby stimulating it. This is somewhat similar to a resonant action. The stimulated light is concentrated into a narrow beam by a glass tube and emitted as a laser beam through the partially silvered lens.

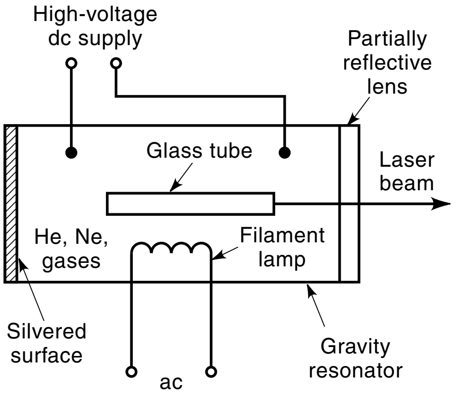


Figure 13. Helium–neon (He–Ne) laser.

Table 17. Properties of Several Types of Lasers (1 Å = 1 × 10^{-10} m)

Laser Type	Wave Length (Å)	Output Power (W/cm^2)
Solid:		
Ruby	7000	0.1 to 100
Glass	1000	0.1 to 500
Liquid:		
Dye	4000 to 10,000	0.001 to 1
Gas:		
Helium–neon	6330	0.001 to 2
Helium–cadmium	4000	0.001 to 1
Carbondioxide	110,000	1 to 1 × 10^4
Semiconductor:		
GaAs	9000	0.002 to 0.01
InGaAsP	13,000	0.001 to 0.005

A semiconductor laser is somewhat similar to an LED. The laser element is typically made of a *pn* junction (diode) of semiconductor material such as gallium arsenide (GaAs) or indium gallium arsenide phosphide (InGaAsP). The edges of the junction are reflective (naturally or by depositing a film of silver). As a voltage is applied to the semiconductor laser, the ionic injection and spontaneous recombination that take place near the *pn* junction will emit light as in an LED. This light will be reflected back and forth between the reflective surfaces, passing along the depletion region many times and creating more photons. The stimulated light (laser) beam is emitted through an edge of the *pn* junction. Semiconductor lasers are often maintained at very low temperatures in order to obtain a reasonable component life. Semiconductor lasers can be manufactured in very small sizes. They are lower in cost and require less power in comparison to the conventional lasers. Wave length and power output characteristics of several types of lasers are given in Table 17.

Liquid Crystal Displays (LCDs)

A liquid crystal display (LCD) consists of a medium of liquid crystal material (e.g., organic compounds such as cholesteryl nonanote and p-azoxyanisole) trapped between a glass sheet and a mirrored surface, as shown in Fig. 14. Pairs of transparent electrodes (e.g., indium tin oxide), arranged in a planar matrix, are deposited on the inner surfaces of the sandwiching plates. In the absence of an electric field across an

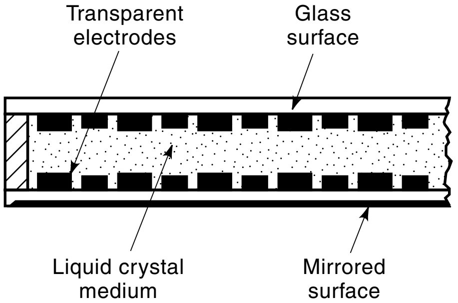


Figure 14. A liquid crystal display (LCD) element.

electrode pair, the atoms of liquid crystal medium in that region will have a parallel orientation. As a result, any light that falls on the glass sheet will first travel through the liquid crystal, then will be reflected back by the mirrored surface, and finally will return unscattered. Once an electrode pair is energized, the molecular alignment of the entrapped medium will change, causing some scattering. As a result, a dark region in the shape of the electrode will be visible. Alphanumeric characters and other graphic images can be displayed in this manner by energizing a particular pattern of electrodes.

Other types of LCD construction are available. In one type, polarized glass sheets are used to entrap the liquid crystal. In addition, a special coating is applied on the inner surfaces of the two sheets that will polarize the liquid crystal medium in different directions. This polarization structure is altered by an electric field (supplied by an electrode pair), thereby displaying an image element. LCDs require external light to function. But they need significantly low currents and power levels to operate. For example, an LED display might need a watt of power, whereas a comparable LCD might require just a small fraction of a milliwatt. Similarly, the current requirement for an LCD will be in the microampere range. LCDs usually need an ac biasing, however. An image resolution on the order of 5 lines/mm is possible with an LCD.

Plasma Displays

A plasma display is somewhat similar to an LCD in construction. The medium used in a plasma display is an ionizing gas (e.g., neon with traces of argon or xenon). A planar matrix of electrode pairs is used on the inner surfaces of entrapping glass. When a voltage above the ionizing voltage of the medium is applied to the electrode pair, the gas will break down, and a discharge will result. The electron impacts that are generated at the cathode as a result will cause further release of electrons to sustain the discharge. A characteristic orange glow will result. The pattern of energized electrodes will determine the graphic image.

The electrodes could be either dc coupled or ac coupled. In the case of the latter, the electrodes are coated with a layer of dielectric material to introduce a capacitor at the gas interface. The power efficiency of a plasma display is higher than that of an LED display. A typical image resolution of 2 lines/mm is obtainable.

Cathode Ray Tubes (CRT)

A schematic representation of a cathode ray tube (CRT) is given in Fig. 15. In a CRT, an electron beam is used to trace lines, characters, and other graphic images on the CRT screen. The electron beam is generated by an electron gun. A cathode made of a metal such as nickel coated with an oxide such as barium strontium calcium oxide forms the electron gun and is heated (say, using a tungsten coil heater) to generate electrons. Electrons are accelerated toward the inner surface of the CRT screen using a series of anodes, biased in increasing steps. The CRT screen is made of glass. Its inner surface is coated with a crystalline phosphor material. The electrons that impinge on the screen will excite the phosphor layer which will result in the release of additional electrons and radiation. As a result, the point of impingement will be illuminated. The electron beam is focused using either electrostatic (a pair of electrode plates) or magnetic (a coil) means. The position of the luminous spot on the screen is controlled using a similar method. Two pairs of electrodes (or two coils) will be needed to deflect the electron to an arbitrary position on the screen.

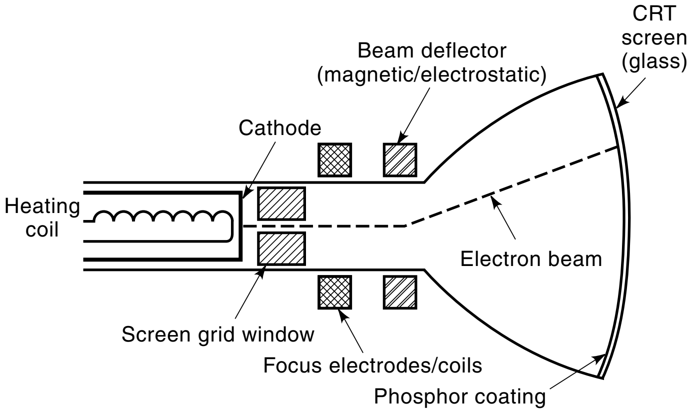


Figure 15. Schematic representation of a cathode ray tube (CRT).

Different types of phosphor material will provide different colors (red, green, blue, white, etc.). The color of a monochrome display is determined by this. Color displays employ one of two common techniques. In one method (masking), three guns are used for the three basic colors (red, green, and blue). The three beams pass through a small masking window and fall on the faceplate. The faceplate has a matrix of miniature phosphor spots (e.g., at 0.1 mm spacing). The matrix consists of a regular pattern of R–G–B phosphor elements. The three electron beams fall on three adjacent spots of R–G–B phosphor. A particular color is obtained as a mixture of the three basic colors by properly adjusting the intensity of the three beams. In the second method (penetration), the faceplate has several layers of phosphor. The color emitted will depend on the depth of penetration of the electron beam into the phosphor.

Flicker in a CRT display, at low frequencies, will strain the eye and also can deteriorate dynamic images. Usually a minimum flicker frequency of 40 Hz will be satisfactory, and even higher frequencies can be achieved with most types of phosphor coatings. Flicker effect worsens with the brightness of an image. The efficiency of a phosphor screen is determined by the light flux density per unit power input (measured in lumens/watt). A typical value is 40 lm/W. Time constant determines the time of decay of an image when power is turned off. Common types of phosphor and their time constants are given in Table 18.

Table 18. Time Constants of CRT Phosphor

Phosphor	Color	Time Constant (ms)
P1	Green	30.0
P4	White	0.1
P22	Red	2.0
	Green	8.0
	Blue	6.0
RP20	Yellow–green	5.0

CRTs have numerous uses. Computer display screens, television picture tubes, radar displays, and oscilloscope tubes are common applications. The raster-scan method is a common way of generating an image on a computer or television screen. In this method, the electron beam continuously sweeps the screen (say, starting from the top left corner of the screen and tracing horizontal lines up to the bottom right corner, continuously repeating the process). The spot is turned on or off using a controller according to some logic which will determine the image that is generated on the screen. In another method used in computer screens, the beam is directly moved to trace the curves that form the image. In oscilloscopes, the horizontal deflection of the beam can be time sequenced and cycled in order to enable the display of time signals.

LIGHT SENSORS

A light sensor (also known as a photodetector or photosensor) is a device that is sensitive to light. Usually, it is a part of an electrical circuit with associated signal conditioning (amplification, filtering, etc.) so that an electrical signal representative of the intensity of light falling on the photosensor is obtained. Some photosensors can serve as energy sources (cells) as well. A photosensor may be an integral component of an optoisolator or other optically-coupled system. In particular, a commercial optical coupler typically has an LED light source and a photosensor in the same package, with a pair of leads for connecting it to other circuits, and perhaps power leads.

By definition, the purpose of a photodetector or photosensor is to sense visible light. But there are many applications where sensing of adjoining bands of the electromagnetic spectrum, namely infrared radiation and ultraviolet radiation, would be useful. For instance, since objects emit reasonable levels of infrared radiation even at low temperatures, infrared sensing can be used in applications where imaging of an object in the dark is needed. Applications include infrared photography, security systems, and missile guidance. Also, since infrared radiation is essentially thermal energy, infrared sensing can be effectively used in thermal control systems. Ultraviolet sensing is not as widely applied as infrared sensing.

Typically, a photosensor is a resistor, diode, or transistor element that brings about a change (e.g., generation of a potential or a change in resistance) into an electrical circuit in response to light that is falling on the sensor element. The power of the output signal may be derived primarily from the power source that energizes the electrical circuit. Alternatively, a photocell can be used as a photosensor. In this latter case, the energy of the light falling on the cell is converted into electrical energy of the output signal. Typically, a photosensor is available as a tiny cylindrical element with a sensor head consisting of a circular window (lens). Several types of photosensors are described below.

Photoresistors

A photoresistor (or photoconductor) has the property of decreasing resistance (increasing conductivity) as the intensity of light falling on it increases. Typically, the resistance of a photoresistor could change from very high values (megohms) in the dark to reasonably low values (less than 100 Ω) in bright light. As a result, very high sensitivity to light is possible. Some photocells can function as photoresistors because their impedance decreases (output increases) as the light intensity increases. Photocells used in this manner are termed photoconductive cells. The circuit symbol of a photoresistor is given in Fig. 16(a).

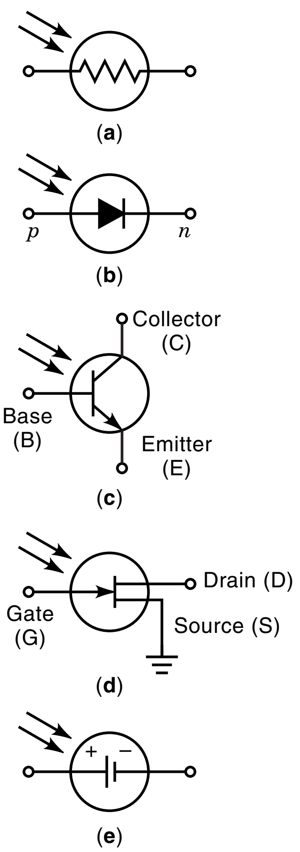


Figure 16. Circuit symbols of some photosensors: (a) photoresistor; (b) photodiode; (c) phototransistor (*npn*); (d) photo-FET (*n*-channel); (e) photocell.

A photoresistor may be formed by sandwiching a photoconductive crystalline material such as cadmium sulfide (CdS) or cadmium selenide (CdSe) between two electrodes. Lead sulfide (PbS) or lead selenide (PbSe) may be used in infrared photoresistors.

Photodiodes

A photodiode is a *pn* junction of semiconductor material that produces electron-hole pairs in response to light. The symbol for a photodiode is shown in Fig. 16(b). Two types of photodiodes are available. A photovoltaic diode generates a sufficient potential at its junction in response to light (photons) falling on it. Hence, an external bias source is not necessary for a photovoltaic diode. A photoconductive diode undergoes a resistance change at its junction in response to photons. This type of photodiode is usually operated in reverse-biased form; the *p*-lead of the diode is connected to the negative lead of the circuit, and *n*-lead is connected to the positive lead of the circuit. The breakdown condition may occur at about 10 V, and the corresponding current will be nearly proportional to the intensity of light falling on the photodiode. Hence, this current can be used as a measure of the light intensity. Since the current level is usually low (a fraction of a milliampere), amplification might be necessary before using it in the subsequent application (e.g., actuation, control, display). Semiconductor materials such as silicon, germanium, cadmium sulfide, and cadmium selenide are commonly used in photodiodes. A diode with an intrinsic layer (a pin diode) can provide faster response than with a regular *pn* diode.

Phototransistor

Any semiconductor photosensor with amplification circuitry built into the same package (chip) is popularly called a phototransistor. Hence, a photodiode with an amplifier circuit in a single unit might be called a phototransistor. Strictly, a phototransistor is manufactured in the form of a conventional bipolar junction transistor with base (B), collector (C) and emitter (E) leads.

Symbolic representation of a phototransistor is shown in Fig. 16(c). This is an *npn* transistor. The base is the central (*p*) region of the transistor element. The collector and the emitter are the two end regions (*n*) of the element. Under operating conditions of the phototransistor, the collector–base junction is reverse biased (i.e., a positive lead of the circuit is connected to the collector, and a negative lead of the circuit is connected to the base of an *npn* transistor). Alternatively, a phototransistor may be connected as a two-terminal device with its base terminal floated and the collector terminal properly biased (positive for an *npn* transistor). For a given level of source voltage (usually applied between the emitter lead of the transistor and load, the negative potential being at the emitter load), the collector current (current through the collector lead) i_c is nearly proportional to the intensity of the light falling on the collector–base junction of the transistor. Hence, i_c can be used as a measure of the light intensity. Germanium or silicon is the semiconductor material that is commonly used in phototransistors.

Photo-FET

A photo-field effect transistor is similar to a conventional FET. The symbol shown in Fig. 16(d) is for an *n*-channel photo-FET. This consists of an *n*-type semiconductor element (e.g., silicon doped with boron), called channel. A much smaller element of *p*-type material is attached to the *n*-type element. The lead on the *p*-type element forms the gate (G). The drain (D) and the source (S) are the two leads on the channel. The operation of an FET depends on the electrostatic fields created by the potentials applied to the leads of the FET.

Under operating conditions of a photo-FET, the gate is reverse-biased (i.e., a negative potential is applied to the gate of an *n*-channel photo-FET). When light is projected at the gate, the drain current i_d will increase. Hence, drain current (current at the D lead) can be used as a measure of light intensity.

Photocells

Photocells are similar to photosensors except that a photocell is used as an electricity source rather than a sensor of radiation. Solar cells, which are more effective in sunlight, are commonly available. A typical photocell is a semiconductor junction element made of a material such as single-crystal silicon, polycrystalline silicon, and cadmium sulfide. Cell arrays are used in moderate-power applications. Typical power output is 10 mW per square centimeter of surface area, with a potential of about 1.0 V. The circuit symbol of a photocell is given in Fig. 16(e).

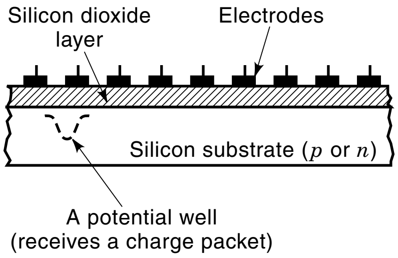


Figure 17. A charge-coupled device (CCD).

Charge-Coupled Device

A Charge-coupled device (CCD) is an integrated circuit (a monolith device) element of semiconductor material. A CCD made from silicon is schematically represented in Fig. 17. A silicon wafer (*p* type or *n* type) is oxidized to generate a layer of SiO_2 on its surface. A matrix of metal electrodes is deposited on the oxide layer and is linked to the CCD output leads. When light falls onto the CCD element, charge packets are generated within the substrate silicon wafer. Now if an external potential is applied to a particular electrode of the CCD, a potential well is formed under the electrode, and a charge packet is deposited here. This charge packet can be moved across the CCD to an output circuit by sequentially energizing the electrodes using pulses of external voltage. Such a charge packet corresponds to a pixel (a picture element). The circuit output is the video signal. The pulsing rate could be higher than 10 MHz. CCDs are commonly used in imaging application, particularly in video cameras. A typical CCD element with a facial area of a few square centimeters may detect 576 × 485 pixels, but larger elements (e.g., 4096 × 4096 pixels) are available for specialized applications. A charge injection device (CID) is similar to a CCD. In a CID, however, there is a matrix of semiconductor capacitor pairs. Each capacitor pair can be directly addressed through voltage pulses. When a particular element is addressed, the potential well there will shrink, thereby injecting minority carriers into the substrate. The corresponding signal, tapped from the substrate, forms the video signal. The signal level of a CID is substantially smaller than that of a CCD, as a result of higher capacitance.

Applications of Optically Coupled Devices

One direct application is in the isolation of electric circuitry. When two circuits are directly connected through electrical connections (cables, wires, etc.), a two-way path is created at the interface for the electrical signals. In other words, signals in circuit A will affect circuit B and signals in circuit B, will affect circuit A. This interaction means that noise in one circuit will directly affect the other. Furthermore, there will be loading problems; the source will be affected by the load. Both

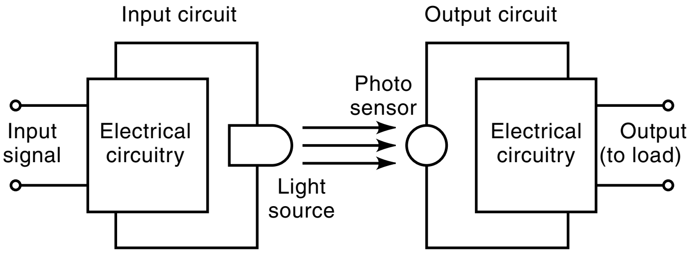


Figure 18. An optically coupled device.

these situations are undesirable. If the two circuits are optically coupled, however, there is only a one-way interaction between the two circuits (see Fig. 18). Variations in the output circuit (load circuit) will not affect the input circuit. Hence, the input circuit is isolated from the output circuit. The connecting cables in an electrical circuit can introduce noise components such as electromagnetic interference, line noise, and ground-loop noise. The likelihood of these noise components affecting the overall system is also reduced by using optical coupling. In summary, isolation between two circuits and isolation of a circuit from noise can be achieved by optical coupling. Optical coupling is widely used in communication networks (telephones, computers, etc.) and in circuitry for high-precision signal conditioning (e.g., for sophisticated sensors and control systems) for these reasons.

The medium through which light passes from the light source to the photosensor can create noise problems, however. If the medium is open (see Fig. 18), then ambient lighting conditions will affect the output circuit, resulting in an error. Also, environmental impurities (dust, smoke, moisture, etc.) will affect the light received by the photosensor. Hence, a more controlled medium of transmission would be desirable. Linking the light source and the photosensor using optical fibers is a good way to reduce problems due to ambient conditions in optically coupled systems.

Optical coupling may be used in relay circuits where a low-power circuit is used to operate a high-power circuit. If the relay that operates the high-power circuit is activated using an optical coupler, reaction effects (noise and loading) on the low-power circuit can be eliminated. Optical coupling is used in power electronics and control systems in this manner.

Many types of sensors and transducers that are based on optical methods do, indeed, employ optical coupling. (e.g., optical encoders, fiberoptic tactile sensors). Optical sensors are widely used in industry for parts counting, parts detection, and level detection. In these sensors, a light beam is projected from a source to a photodetector, both units being stationary. An interruption of the beam through the passage of a part will generate a pulse at the detector, and this pulse is read by a counter or a parts detector. Furthermore, if the light beam is located horizontally at a required height, its interruption when the material filled into a container reaches that level could be used for filling control in the packaging industry. Note that the light source and the sensor could be located within a single package if a mirror is used to reflect light from the source back onto the detector. Further applications are within computer disk drive systems, for example, to detect the write protect notch as well as the position of the recording head.

BIBLIOGRAPHY

1. C. W. de Silva, *Control Sensors and Actuators,* Englewood Cliffs, NJ: Prentice-Hall, 1989.

Reading List

M. G. Abele, *Structures of Permanent Magnets,* New York: Wiley, 1993.

C. Belove (ed.), *Handbook of Modern Electronics and Electrical Engineering,* New York: Wiley, 1986.

P. F. Dahl, *Superconductivity,* New York: American Institute of Physics, 1992.

D. G. Fink and D. Christiansen (eds.), *Electronics Engineers' Handbook,* New York: McGraw-Hill, 1982.

P. M. Fishbane, S. Gasiorowicz, and S. T. Thornton, *Physics for Scientists and Engineers,* Upper Saddle River, NJ: Prentice-Hall, 1996.

D. C. Giancoli, *Physics Principles and Applications,* Englewood Cliffs, NJ: Prentice-Hall, 1991.

A. Holmes-Siedle and L. Adams, *Handbook of Radiation Effects,* New York: Oxford University Press, 1993.

L. V. Interrante, L. A. Casper, and A. B. Ellis (eds.), *Materials Chemistry,* Washington, DC: American Chemical Society, 1995.

M. Kaufaman and A. H. Seidman (eds.), *Handbook for Electronic Engineering Technicians,* New York: McGraw-Hill, 1984.

D. Metzger, *Electronic Components, Instruments, and Troubleshooting,* Englewood Cliffs, NJ: Prentice-Hall, 1981.

M. Rosenstein and P. Morris, *Modern Electronic Devices: Circuit Design and Application,* Reston, VA: Reston Publishing Company, 1985.

M. N. O. Sadiku, *Elements of Electromagnetics,* Orlando, FL: Saunders College Publishing, 1994.

J. Seymour, *Electronic Devices and Components,* London: Pitman Publishing, 1981.

L. Solymar and D. Walsh, *Lectures on the Electrical Properties of Materials,* 3rd ed., Oxford: Oxford University Press, 1984.

C. M. Thomson, *Fundamentals of Electronics,* Englewood Cliffs, NJ: Prentice-Hall, 1979.

Clarence W. de Silva
University of British Columbia