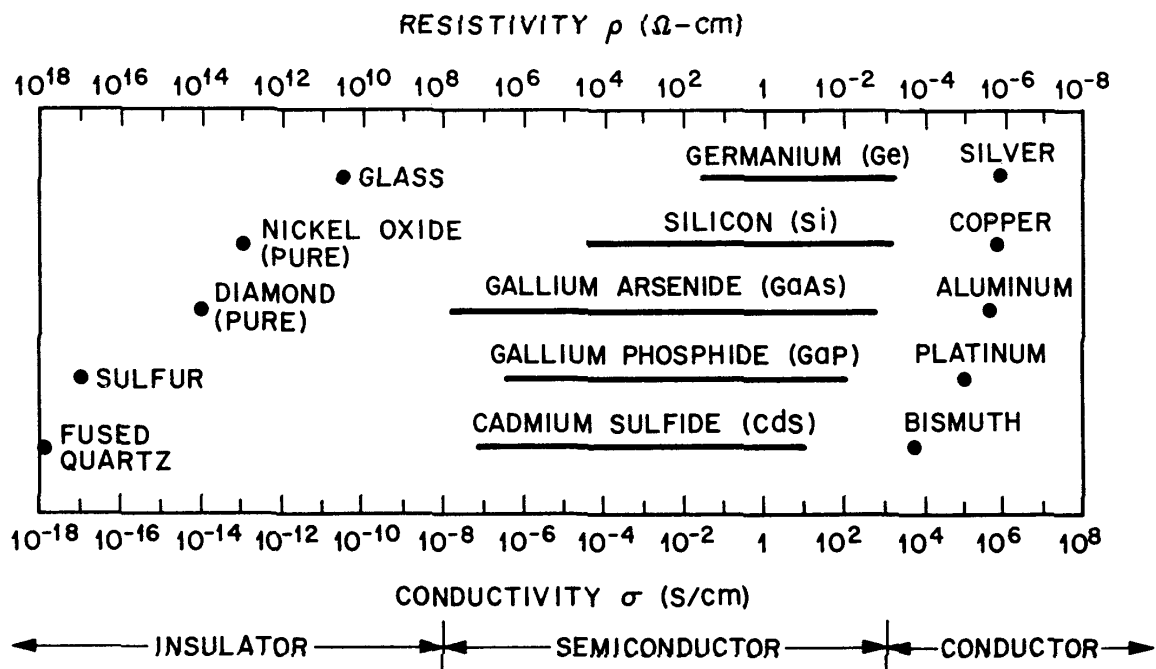


CHAPTER 1: Semiconductor Materials & Physics

In this chapter, the basic properties of semiconductors and microelectronic devices are discussed.

1.1 Semiconductor Materials

Solid-state materials can be categorized into three classes - insulators, semiconductors, and conductors. As shown in [Figure 1.1](#), the resistivity of semiconductors, ρ , is typically between 10^{-2} and $10^8 \Omega\text{-cm}$. The portion of the periodic table related to semiconductors is depicted in [Table 1.1](#).



[Figure 1.1](#): Typical range of conductivities for insulators, semiconductors, and conductors.

Semiconductors can be composed of a single element such as silicon and germanium or consist of two or more elements for compound semiconductors. A binary III-V semiconductor is one comprising one element from Column III (such as gallium) and another element from Column V (for instance, arsenic). The common element and compound semiconductors are displayed in [Table 1.2](#).

Table 1.1: Portion of the Periodic Table Related to Semiconductors.

Period	Column II	III	IV	V	VI
2		B Boron	C Carbon	N Nitrogen	
3	Mg Magnesium	Al Aluminum	Si Silicon	P Phosphorus	S Sulfur
4	Zn Zinc	Ga Gallium	Ge Germanium	As Arsenic	Se Selenium
5	Cd Cadmium	In Indium	Sn Tin	Sb Antimony	Te Tellurium
6	Hg Mercury		Pd Lead		

Table 1.2: Element and compound semiconductors.

Elements	IV-IV	III-V	II-VI	IV-VI
	Compounds	Compounds	Compounds	Compounds
Si	SiC	AlAs	CdS	PbS
Ge	SiGe	AlSb	CdSe	PbTe
	SiCGe	GaN	CdTe	
		GaAs	ZnS	
		GaP	ZnSe	
		GaSb	ZnTe	
		InAs	ZnO	
		InP		
		InSb		

1.2 Crystal Structure

Most semiconductor materials are single crystals. *Figure 1.2* exhibits three cubic-crystal unit cells - simple cubic, body-centered cubic, and face-centered cubic. The element semiconductors, silicon and germanium, have a diamond lattice structure as shown in *Figure 1.3*. This configuration belongs to the cubic-crystal family and can be envisaged as two interpenetrating fcc sublattices with one sublattice staggered from the other by one quarter of the distance along a diagonal of the cube. All atoms are identical in a diamond lattice, and each atom in the diamond lattice is surrounded by four equidistant nearest neighbors that lie at the corners of a tetrahedron. Most of the III-V semiconductors (e.g. GaAs) have a zincblende lattice (shown in *Figure 1.3b*) that is identical to a diamond lattice except that one fcc sublattice has Column III atoms (Ga) and the other has Column V atoms (As).

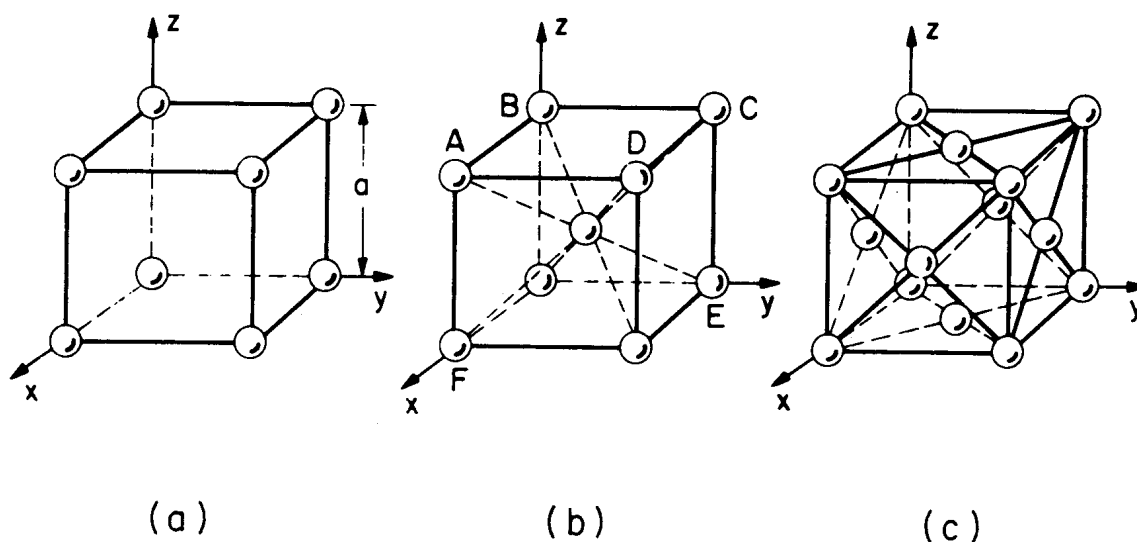


Figure 1.2: Three cubic-crystal unit cells – (a) Simple cubic (b) Body-centered cubic (c) Face-centered cubic.

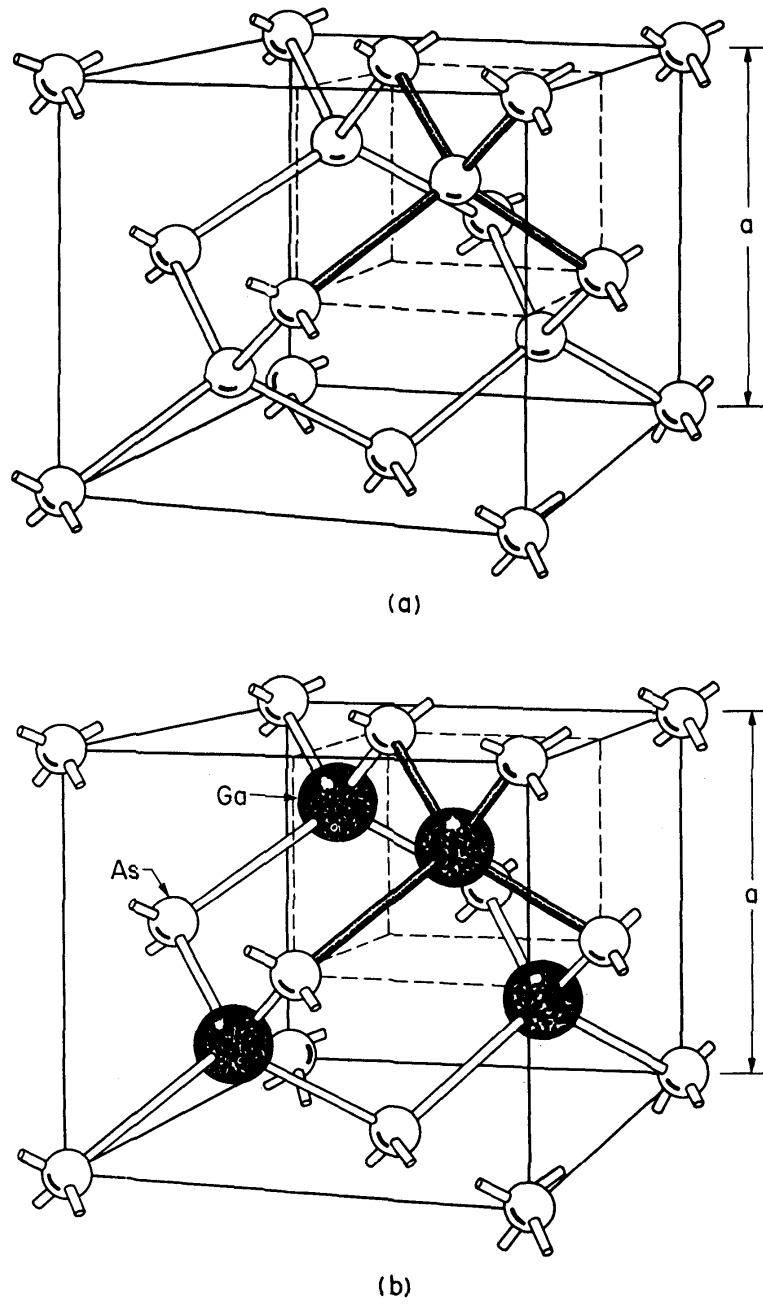


Figure 1.3: (a) Diamond lattice. (b) Zincblende lattice.

1.3 Energy Bands

According to the Bohr model, the energy levels of a hydrogen atom are given by *Equation 1.1*:

$$E_H = \frac{-m_o q^4}{8\epsilon_o^2 h^2 n^2} = \frac{-13.6}{n^2} \text{ eV} \quad (\text{Equation 1.1})$$

where m_o denotes the free electron mass
 q denotes the electronic charge
 ϵ_o denotes the free space permittivity
 h denotes the Plank constant
 n denotes the principal quantum number

Therefore, for $n = 1$, that is, ground state, $E_H = -13.6$ eV. For $n = 2$, the first excited state, $E_H = -3.4$ eV.

When two atoms approach one another, the energy level will split into two by the interaction between the atoms. When N atoms come together to form a crystal, the energy will be split into N separate but closely spaced levels, thereby resulting in an essentially continuous band of energy. The detailed energy band structures of crystalline solids can be calculated using quantum mechanics. *Figure 1.4* is a schematic diagram of the formation of a diamond lattice crystal from isolated silicon atoms. The energy band splits into two, the conduction band and the valence band, as the two atoms approach the equilibrium interatomic spacing. The region separating the conduction and valence bands is termed the forbidden gap or bandgap, E_g . *Figure 1.5* exhibits the energy band diagrams of three classes of solids: insulators, semiconductors, and conductors. In insulators, the bandgap is relatively large and thermal energy or an applied electric field cannot raise the uppermost electron in the valence band to the conduction band. In metals or conductors, the conduction band is either partially filled or overlaps the valence band such that there is no bandgap and current can readily flow in these materials. In semiconductors, the bandgap is smaller than that of insulators, and thermal energy can excite electrons to the conduction band. The bandgap of a semiconductor decreases with higher temperature. For instance, for silicon, E_g is 1.12 eV at room temperature and 1.17 eV at zero Kelvin.

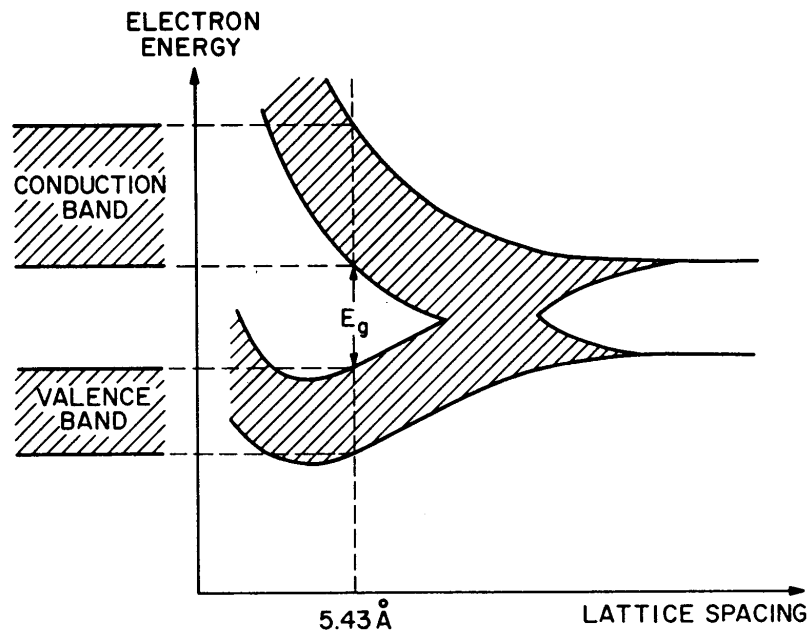


Figure 1.4: Formation of energy bands as a diamond lattice crystal by bringing together isolated silicon atoms.

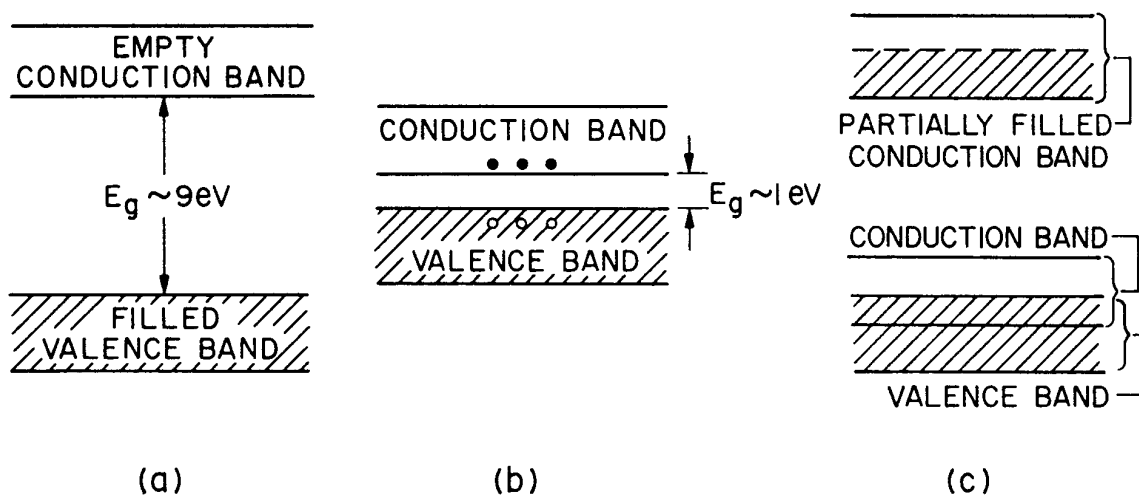


Figure 1.5: Schematic energy band representations of (a) an insulator, (b) a semiconductor, and (c) conductors.

Figure 1.6 shows a more detailed schematic of the energy band structures for silicon and gallium arsenide in which the energy is plotted against the crystal momentum for two crystal directions. For silicon, the minimum of the conduction band and the maximum of the valence band have different crystal momenta. Silicon is therefore an indirect bandgap semiconductor as a change in

crystal momentum is required for an electron transition between the valence and conduction bands. On the contrary, GaAs is a direct bandgap semiconductor and generation of photons is more efficient.

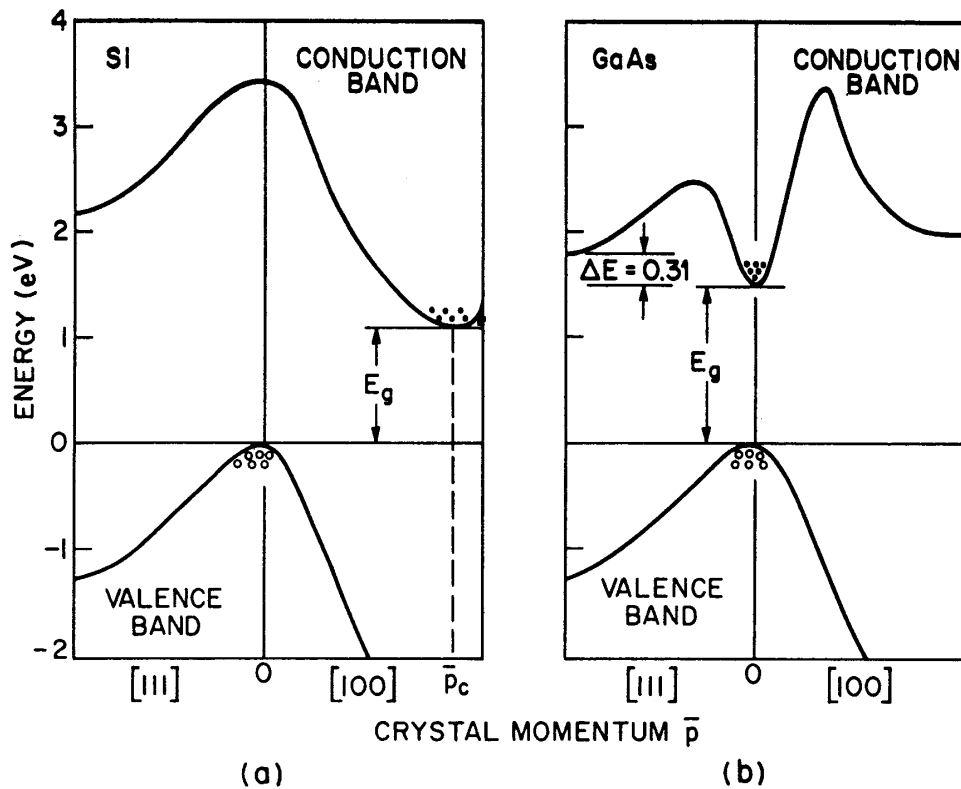


Figure 1.6: Energy band structures of Si and GaAs. Circles (o) indicate holes in the valence bands and dots (•) indicate electrons in the conductor bands.

1.4 Intrinsic Carrier Concentration

The probability that an electronic state with energy E is occupied by an electron is given by the Fermi-Dirac distribution function:

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \quad (\text{Equation 1.2})$$

where E_F is the Fermi level, the energy at which the probability of occupation by an electron is exactly one-half. At room temperature, the intrinsic Fermi level lies very close to the middle of the bandgap.

The effective density of states in the conduction band N_C is equal to $2[2\pi m_n kT/h^2]^{3/2}$. Similarly, the effective density of states in the valence band N_V is $2[2\pi m_p kT/h^2]^{3/2}$. At room temperature, N_C for silicon is 2.8×10^{19} atoms/cm³. For an intrinsic semiconductor, the number of electrons per unit volume in the conduction band is equal to the number of holes per unit volume in the valence band. That is, $n = p = n_i$ where n_i is the intrinsic carrier density. Since $n = N_C \exp\{-(E_C - E_F)/kT\}$ and $p = N_V \exp\{-(E_F - E_V)/kT\}$, where n is the electron density and p is the hole density, $np = n_i^2 = N_C N_V \exp\{(E_V - E_C)/kT\} = N_C N_V \exp\{-E_g/kT\}$. Therefore,

$$n_i = (N_C N_V)^{1/2} \exp\left\{\frac{-E_g}{2kT}\right\} \quad (\text{Equation 1.3})$$

For a doped, or extrinsic, semiconductor, the increase of one type of carriers reduces the number of the other type. Thus, the product of the two types of carriers remains constant at a given temperature. For Si, $n_i = 1.45 \times 10^{10}$ cm⁻³ and for GaAs, $n_i = 1.79 \times 10^6$ cm⁻³. GaAs has a lower intrinsic carrier density on account of its larger bandgap.

(For derivation of the equations described in this section, please peruse the recommended textbooks.)

1.5 Donors and Acceptors

Figure 1.7a shows schematically the doping of a silicon crystal with an arsenic atom. The arsenic atom forms covalent bonds with its four adjacent silicon atoms, and the fifth electron becomes a conduction electron, thereby giving rise to a positively charged arsenic atom. As a consequence, the silicon crystal becomes n-type and arsenic is called a donor. Boron, on the other hand, has only three outer shell electrons and is an acceptor in silicon. Impurities such as arsenic and boron have energy levels very close to the conduction band and valence band, respectively, as indicated in *Figure 1.8*. Shallow donors or acceptors such as these exist in ionized form at room temperature because thermal energy is sufficient to ionize them. This condition is called complete ionization, that is, $n = N_A$ or N_D . Since $n = N_C \exp\{-(E_C - E_F)/kT\}$, $N_D = N_C \exp\{-(E_C - E_F)/kT\}$ and

$$E_C - E_F = kT \ln[N_C/N_D] \quad (\text{Equation 1.4})$$

One of the implications of *Equation 1.4* is that $(E_C - E_F)$ becomes smaller with increasing N_D , or in other words, the Fermi level moves closer to the bottom of the conduction band. Similarly, for p-type semiconductors, the Fermi level moves towards the top of the valence band with increasing acceptor concentration.

When both donors and acceptors are present simultaneously, the impurity present at a higher concentration determines the type of conductivity in the semiconductor. The electron in an n-type semiconductor is called the majority carrier, whereas the hole in n-type semiconductor is termed the minority carrier. Conversely, in a p-type semiconductor, holes are majority carriers and electrons are minority carriers.

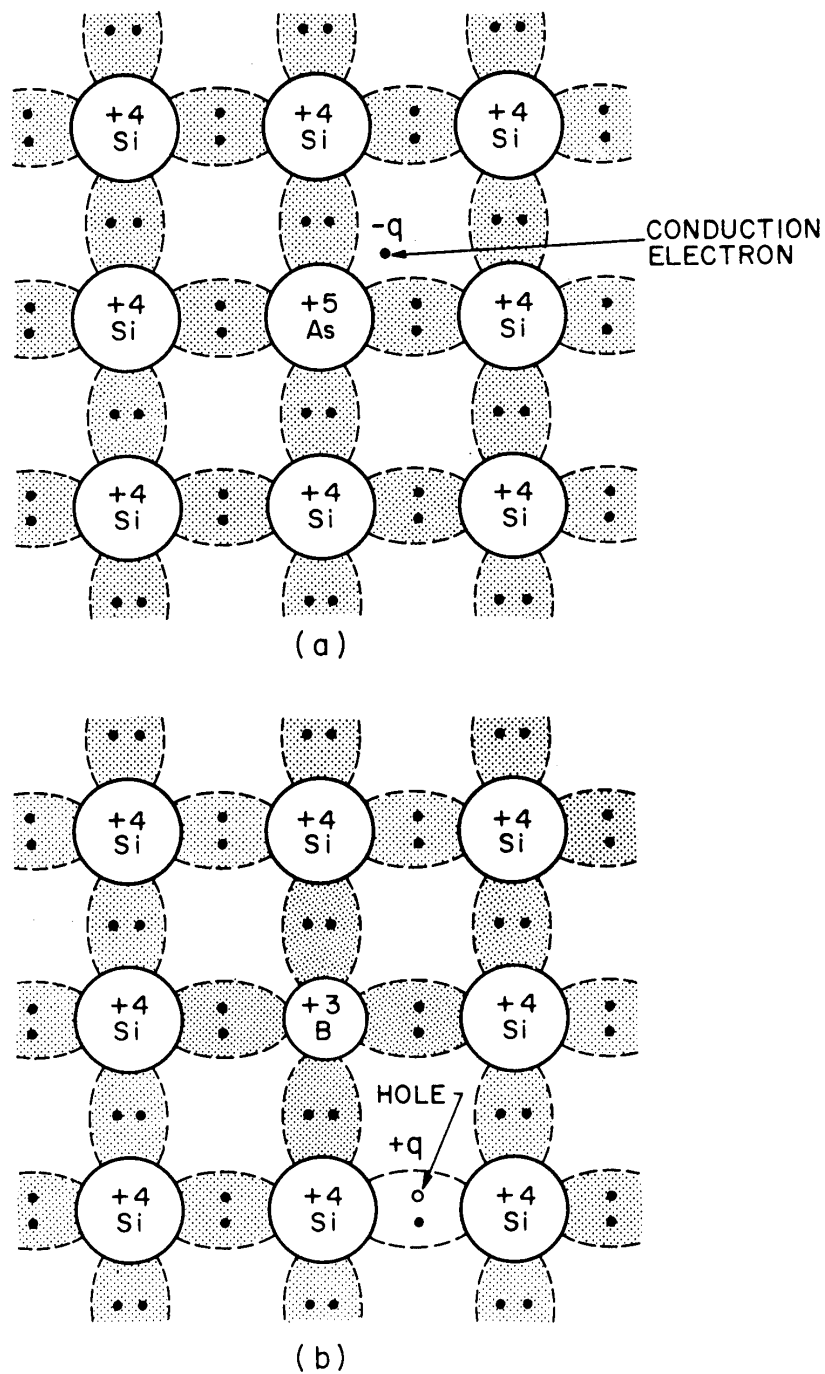


Figure 1.7: Schematic bond pictures of (a) n-type Si with donor (arsenic) and (b) p-type Si with acceptor (boron).

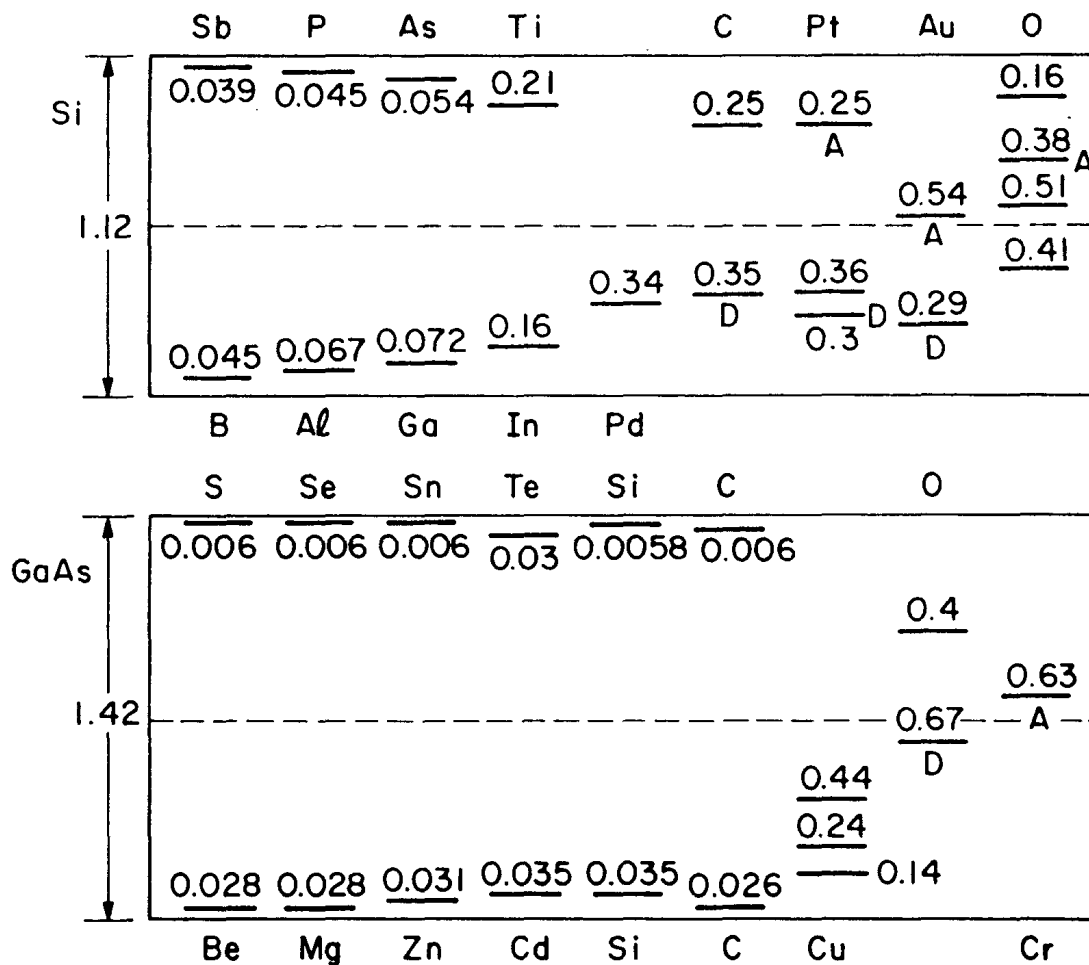


Figure 1.8: Measured ionization energies for various impurities in Si and GaAs. The levels below the gap center are measured from the top of the valence band and are acceptor levels unless indicated by D for donor level. The levels above the gap center are measured from the bottom of the conduction band and are donor levels unless indicated by A for acceptor level.

Example 1.1

A silicon wafer is doped with 10^{16} arsenic atoms/cm³. Find the carrier concentrations and the Fermi level at room temperature (300K).

Solution

At 300K, we can assume complete ionization of impurity atoms. We have:

$$n \approx N_D = 10^{16} \text{ cm}^{-3}$$

Thus,

$$p \approx \frac{n_i^2}{N_D} = \frac{(1.45 \times 10^{10})^2}{10^{16}} = 2.1 \times 10^4 \text{ cm}^{-3}$$

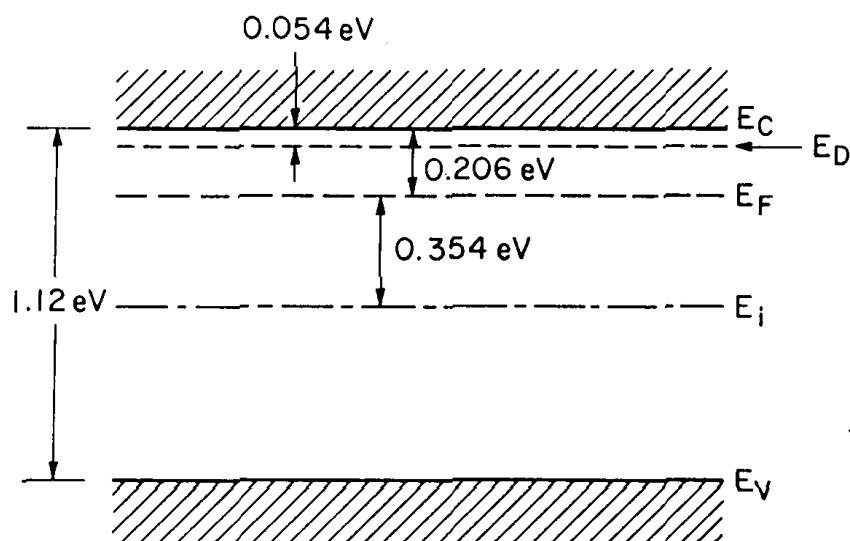
The Fermi level measured from the bottom of the conduction band is given by *Equation 1.8*:

$$E_C - E_F = kT \ln \left[\frac{N_C}{N_D} \right] = 0.0259 \ln \left[\frac{2.8 \times 10^{19}}{10^{16}} \right] = 0.206 \text{ eV}$$

The Fermi level measured from the intrinsic Fermi level is:

$$E_F - E_i = kT \ln \left[\frac{n}{n_i} \right] \approx kT \ln \left[\frac{N_D}{n_i} \right] = 0.0259 \ln \left[\frac{10^{16}}{1.45 \times 10^{10}} \right] = 0.354 \text{ eV}$$

Graphically, the band structure is:



1.6 Electron Mobility

Using the theorem of equipartition of energy, $m_n v_{th}^2/2 = 3kT/2$, where m_n is the electron effective mass and v_{th} is the average thermal velocity. Electrons in the semiconductor therefore move rapidly in all directions. The thermal motion of an individual electron can be visualized as a succession of random scattering from collisions with lattice atoms, impurity atoms, and other scattering centers. The average distance between collisions is called the mean free path, and the average time between collisions is termed the mean free time, τ_c . When a small electric field, ε , is applied to the semiconductor, each electron will experience a force equal to $-q\varepsilon$ and will be accelerated in opposite direction to the electric field with a drift velocity, v_n . By Newtonian physics, the momentum of the electron is force times time, that is, equal to $-q\varepsilon\tau_c$. Therefore, $m_n v_n = -q\varepsilon\tau_c$, or

$$v_n = -[q\tau_c/m_n]\varepsilon \quad (\text{Equation 1.5})$$

Equation 1.5 states that the drift velocity is proportional to the applied electric field. The proportionality factor is called the electron mobility, μ_n , in units of $\text{cm}^2/\text{V}\cdot\text{s}$. Hence,

$$v_n = -\mu_n\varepsilon \quad \text{where } \mu_n = q\tau_c/m_n \quad (\text{Equation 1.6})$$

A similar expression can be written for holes:

$$v_p = \mu_p\varepsilon \quad (\text{Equation 1.7})$$

The negative sign is removed from *Equation 1.7* because holes drift in the same direction as the electric field.

Carrier mobility depends on lattice scattering and impurity scattering. Lattice scattering results from thermal vibrations of the lattice atoms. As lattice vibration is more significant with increasing temperature, mobility decreases. In fact, at high temperature, lattice vibration dominates. Impurity scattering results when a charge carrier travels past an ionized donor or acceptor. The probability of impurity scattering depends on the total impurity concentration. Unlike lattice scattering, impurity scattering becomes less significant at high temperatures because the carriers move faster and are less effectively scattered. *Figure 1.9* illustrates these effects.

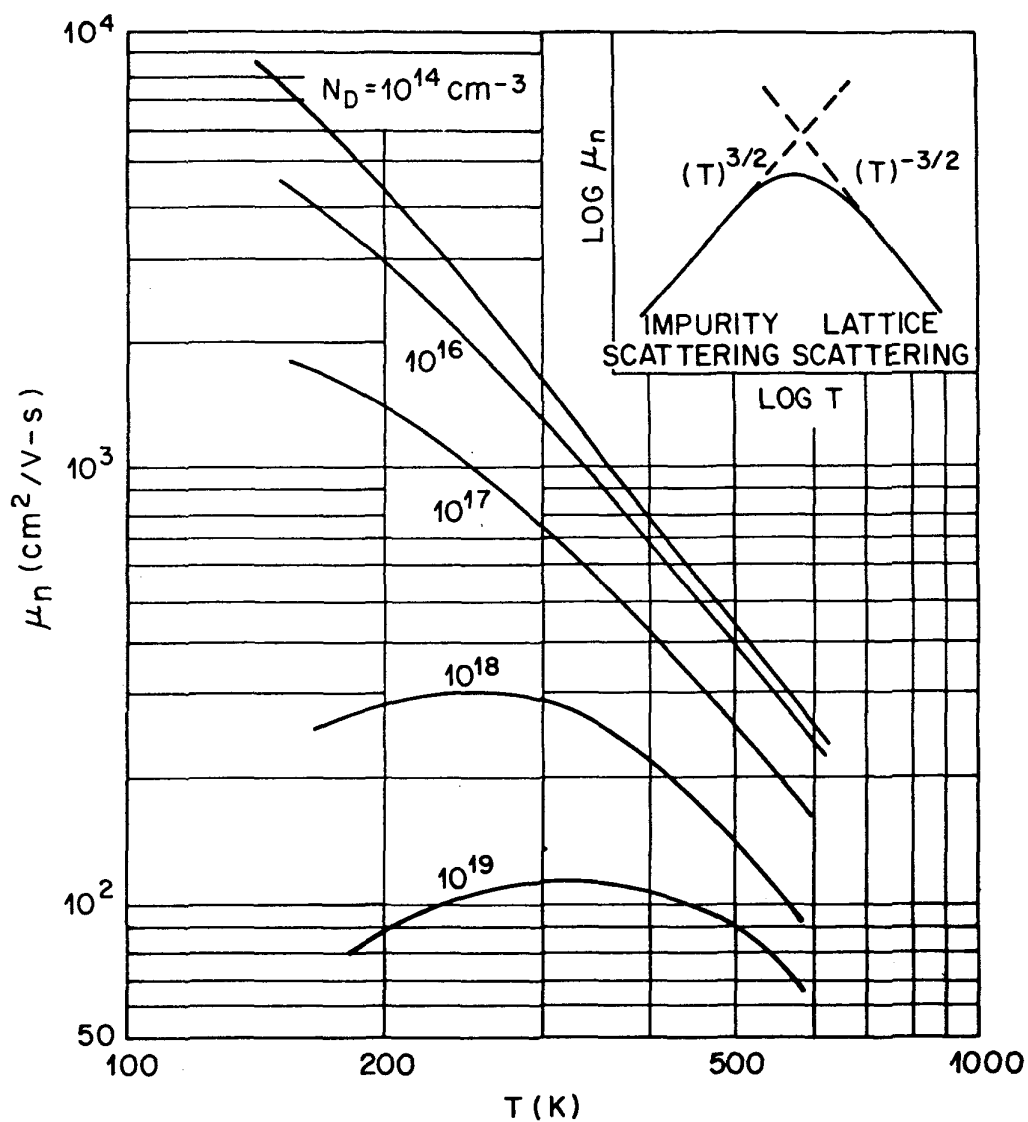


Figure 1.9: Electron mobility in silicon versus temperature for various donor concentrations. Insert shows the theoretical temperature dependence of electron mobility.

1.7 Resistivity and Conductivity

Consider a semiconductor sample shown in *Figure 1.10*, which has a cross-sectional area A , a length L , and a carrier concentration of n electrons/cm³. Suppose an electric field, ε , is applied to the sample. The electron current density, J_n , flowing in the sample can be calculated by summing the product of the charge on each electron times the electron's velocity over all electrons per unit volume n . That is,

$$J_n = I_n/A = \sum_{i=1}^n (qv_i) = -qnv_n = qn\mu_n\varepsilon \quad (\text{Equation 1.8})$$

The total current, J , is equal to J_n (electrons) + J_p (holes). Thus,

$$J = (qn\mu_n + qp\mu_p)\varepsilon \quad (\text{Equation 1.9})$$

This proportionality constant is known as the conductivity, σ . A more commonly used term is the resistivity, ρ , which is the reciprocal of the conductivity. Thus,

$$\rho = \sigma^{-1} = [q(n\mu_n + p\mu_p)]^{-1} \quad (\text{Equation 1.10})$$

Generally, in extrinsic semiconductors, one of the two components is dominant. Therefore, for n-type semiconductors,

$$\rho = (qn\mu_n)^{-1} \quad (\text{Equation 1.11})$$

For p-type semiconductors,

$$\rho = (qp\mu_p)^{-1} \quad (\text{Equation 1.12})$$

Figure 1.11 depicts the measured resistivity for silicon and gallium arsenide at 300K as a function of the impurity concentration.

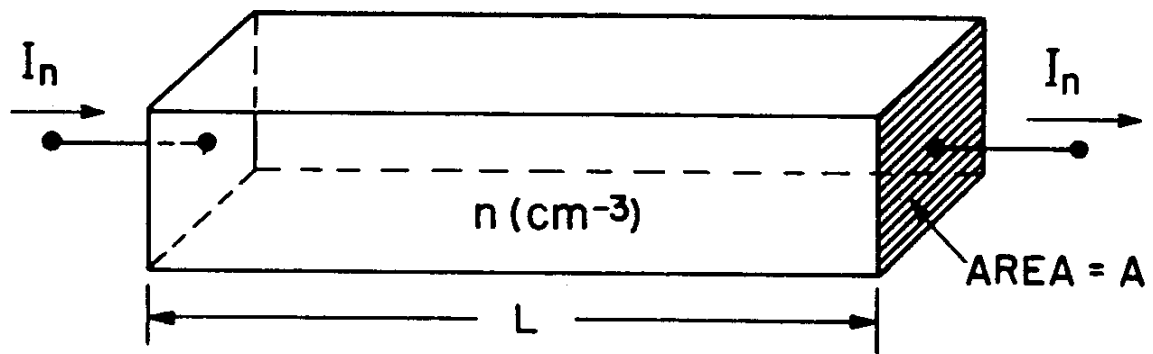


Figure 1.10: Current conduction in a uniformly doped semiconductor bar with length L and cross-sectional area A .

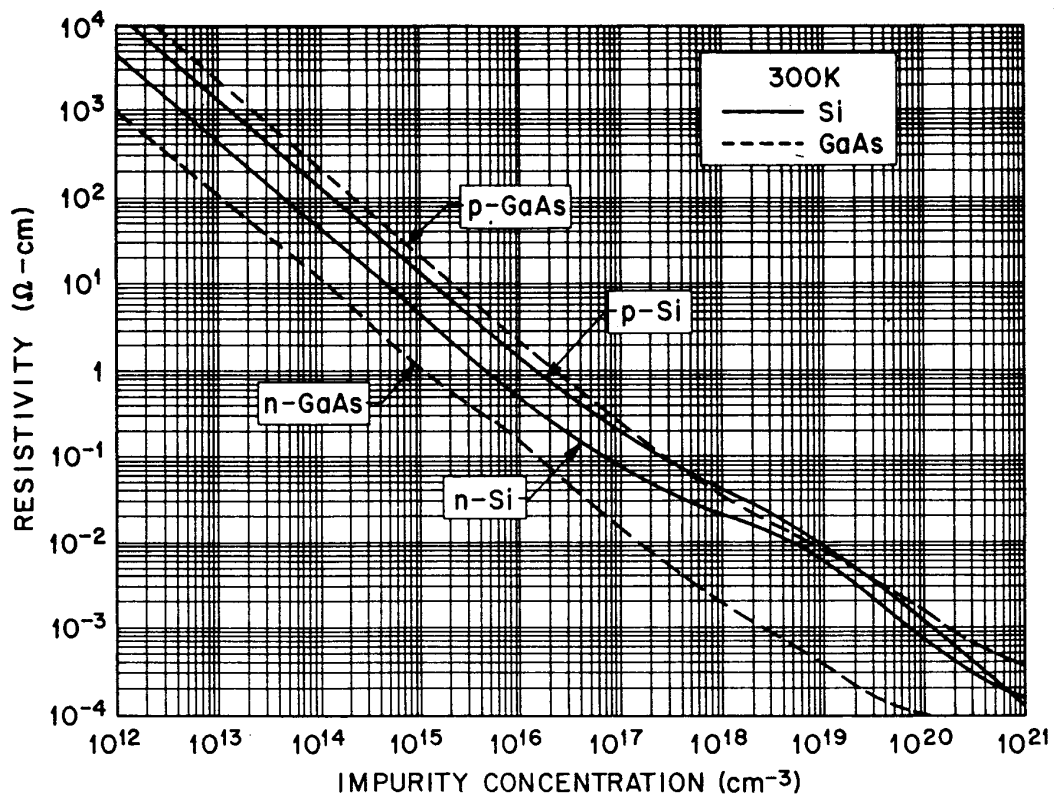
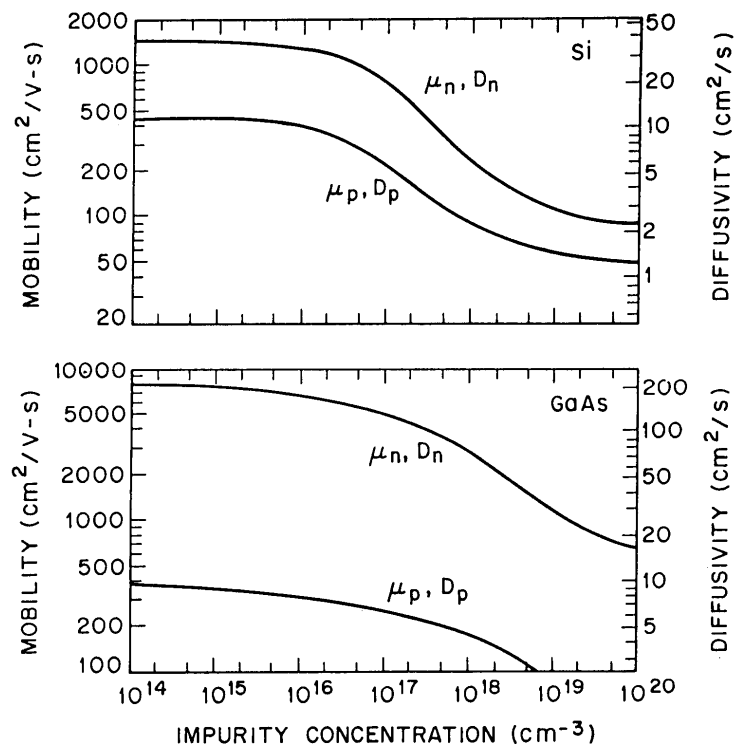


Figure 1.11: Resistivity versus impurity concentration for Si and GaAs.

Example 1.2

Using the mobility and diffusivity chart for Si at 300K as a function of impurity concentration (top graph), find the room temperature resistivity of an n-type silicon wafer doped with 10^{16} atoms/cm³ of phosphorus.

**Solution**

At room temperature, we assume that all donors are ionized. Thus,

$$n \approx N_D = 10^{16} \text{ cm}^{-3}$$

We can also calculate the resistivity using *Equation 1.11*:

$$\rho = \frac{1}{qn\mu_n} = \frac{1}{1.6 \times 10^{-19} \times 10^{16} \times 1300} = 0.48 \Omega\text{-cm}$$

The mobility μ_n is obtained from the mobility chart.

Alternatively, we can obtain the result graphically from *Figure 1.11* also giving $\rho = 0.48 \Omega\text{-cm}$.

1.8 P-N Junction

When a p-type semiconductor is joined together with an n-type semiconductor by processes such as epitaxy, ion implantation, or diffusion, a p-n junction is created. *Figure 1.12* illustrates the energy band structure for such a device. The large carrier concentration gradients at a p-n junction cause carrier diffusion. Holes from the p-side diffuse into the n-side, and electrons from the n-side diffuse into the p-side. This sets up an electric field, which in equilibrium, exactly counteracts these diffusion tendencies and thus permits no net transport of electrons or holes across the junction.

When a small positive voltage is applied to the p-side, there will be a net movement of holes flowing from the p-side to the n-side, thereby creating a forward bias situation. Conversely, if a negative voltage is applied to the p-side, i.e. reverse bias condition, the p-n junction becomes an open circuit. A p-n junction therefore acts as a diode.

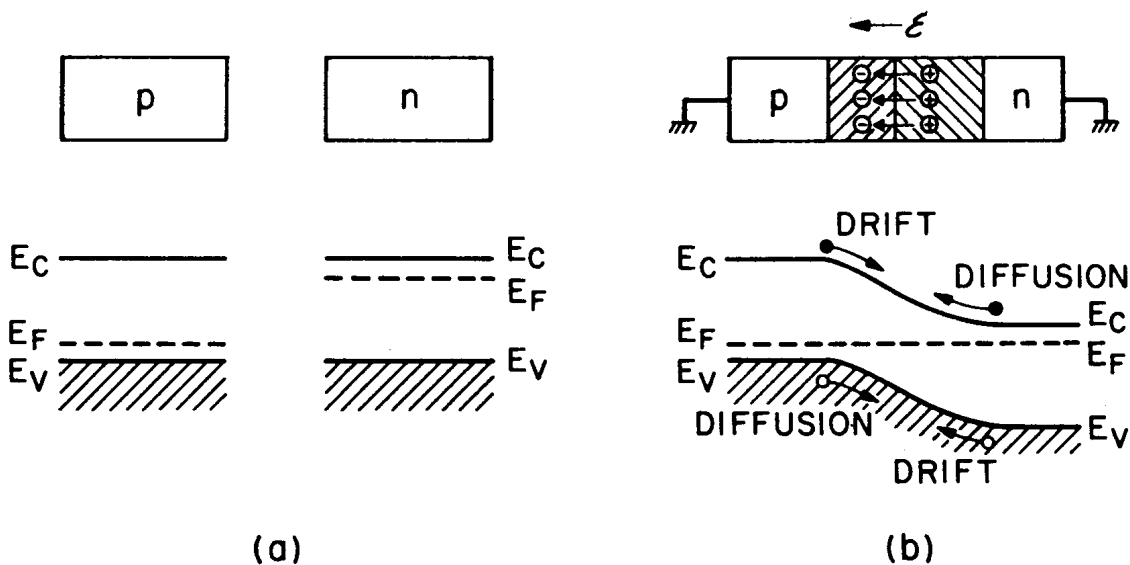


Figure 1.12: (a) Uniformly doped p-type and n-type semiconductors before junction formation. (b) Electric field in the depletion region and energy band diagram of a p-n junction in thermal equilibrium.

1.9 Generation and Recombination Processes

When the thermal equilibrium in a semiconductor is perturbed, that is, the product, np , is no longer equal to n_i^2 , by injection of excess carriers, processes exist to restore the system back to equilibrium ($np = n_i^2$). If the released energy results in the emission of a photon, the process is called radiative recombination (**Figure 1.13**), otherwise it is called nonradiative recombination. When excess carriers are introduced to a direct-bandgap semiconductor such as GaAs, the probability is high that electrons and holes will recombine directly and a photon is emitted.

Light emitting diodes (LEDs) are p-n junctions that emit spontaneous radiation in ultraviolet, visible, or infrared region under bias (**Figure 1.14**). The most common visible LEDs are made of alloys of $\text{GaAs}_{1-y}\text{P}_y$, where y is the atomic fraction. The bandgap of the alloy depends on its composition, i.e. y , thus implying that the emitted light frequency can be tailored by altering y . In the $\text{GaAs}_{1-y}\text{P}_y$ alloy, if y is greater than 0.45, the material becomes indirect-bandgap semiconductor and the probability for radiative transitions is quite small. However, special recombination centers can be incorporated to enhance the radiative probability, for instance, by doping the $\text{GaAs}_{1-y}\text{P}_y$ crystal with nitrogen creating an electron trap level close to the bottom of the conduction band. Nitrogen forms an isoelectronic center as the chemical properties of phosphorus and nitrogen are quite similar.

As shown in **Figure 1.13c**, when a photon of energy $h\nu_{12}$ impinges on an atom while it is in the excited state, the atom can be stimulated to make a transition to the ground state and gives off another photon of energy $h\nu_{12}$ that is in phase with the incident radiation. This process is called stimulated emission and forms the basis of solid-state lasers. The radiation thus produced is monochromatic and coherent.

For indirect-bandgap semiconductors like silicon, a direct recombination process is unlikely because electrons at the bottom of the conduction band have nonzero crystal momentum with respect to holes at the top of the valence band. A direct transition that conserves both energy and momentum is not possible without a simultaneous lattice interaction. Hence, the dominant recombination process is indirect transition, that is, via localized energy states in the forbidden bandgap. Impurities with energy levels in the bandgap can substantially enhance the recombination rate and reduce the lifetime of excess carriers.

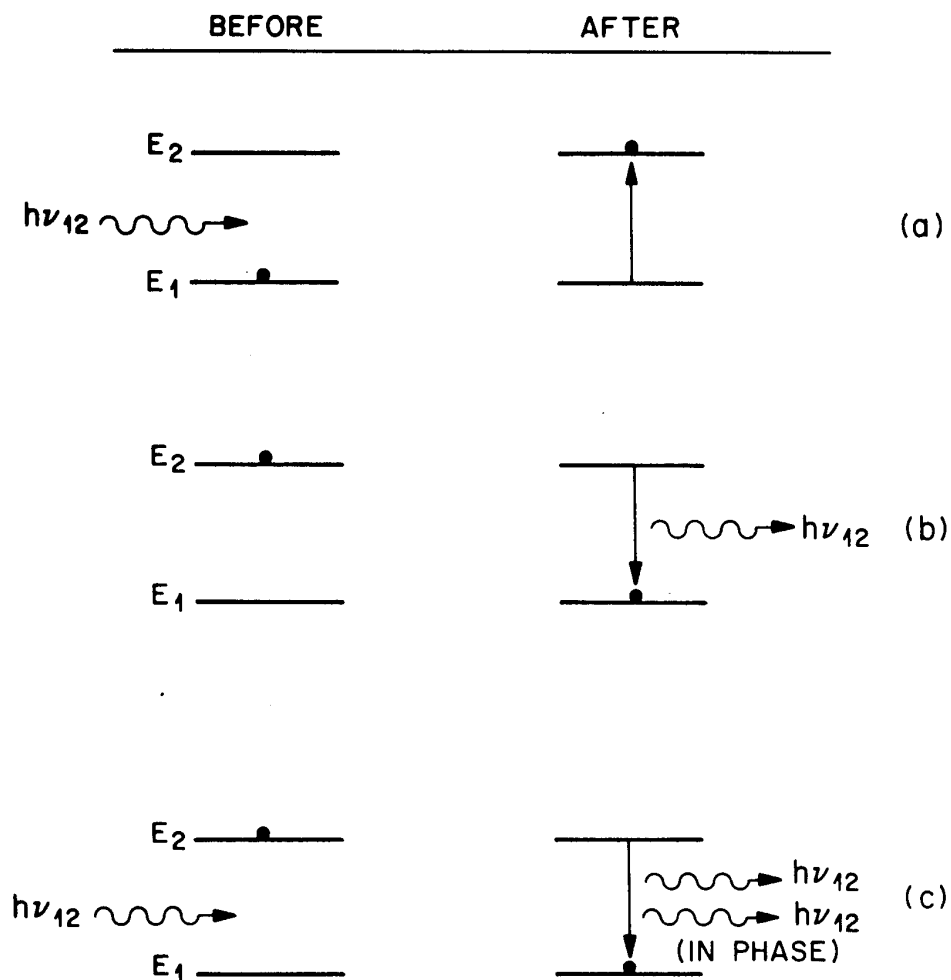


Figure 1.13: The three basic transition processes between two energy levels. Black dots indicate the state of the atom. The initial state is at the left; the final state, after the transition process, is at the right. (a) Absorption. (b) Spontaneous emission. (c) Stimulated emission.

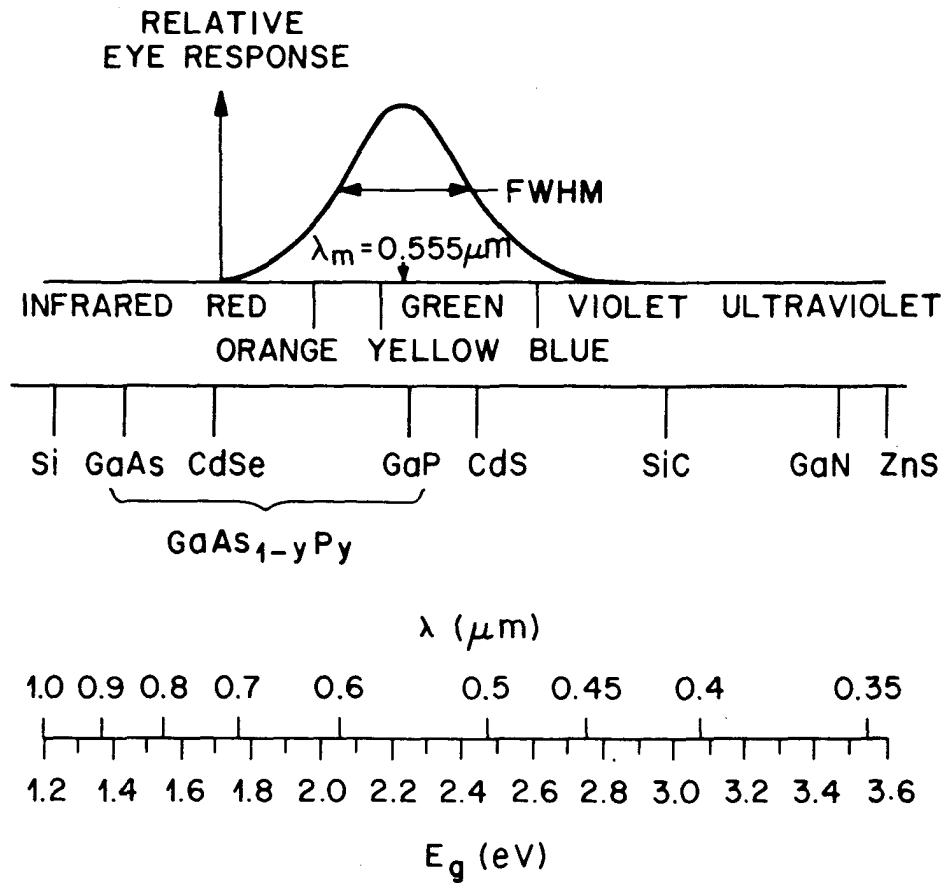


Figure 1.14: Semiconductors of interest as visible LEDs. Figure includes relative response of the human eye.

1.10 Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET)

The most common element in an integrated circuit is the metal-oxide-semiconductor field-effect transistor (MOSFET). *Fig. 1.15* shows the schematic cross section of two n-channel MOS transistors. In a MOSFET device, the channel current is controlled by a voltage applied to a gate that is separated from the channel by an insulator typically made of SiO_2 . It works as a switch in that when a positive voltage is applied to the gate, negative charges are attracted towards the gate insulator. If the voltage is large enough, enough negative charges accumulate underneath the gate dielectric to result in a conductive path between the source and drain. In the enhancement mode, (*Fig. 1.15a*), the transistor is normally off, and no current flows between the source and drain for a gate voltage (V_G) = 0 V. A conducting channel is then induced by applying a voltage of the appropriate polarity (positive for n-channel MOSFET or negative for p-channel MOSFET) to the gate. In the depletion-mode (*Fig. 1.15b*), a conducting channel already exists, and the device is on with no bias applied to the gate. The channel is depleted of mobile carriers by applying a gate voltage.

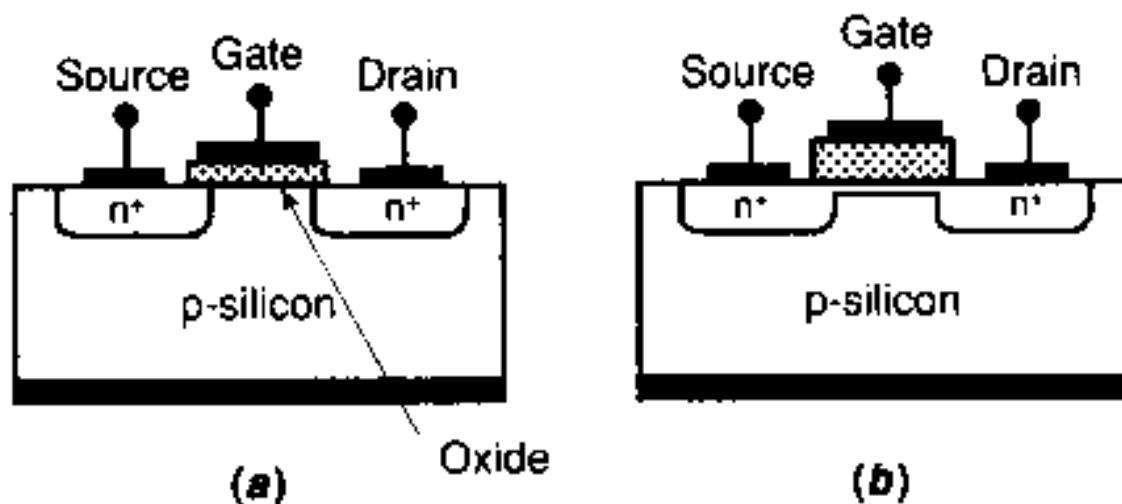


Figure 1.15: Schematic cross section of n-channel MOS transistors: (a) enhancement-type device and (b) depletion-type device.

The complementary MOS (CMOS) schematically shown in *Fig. 1.16* is of paramount importance in the IC industry. It consists of a combination of n-channel and p-channel MOS transistors in series. If both transistors are of the enhancement type, only one of the two transistors, that is, either the p-channel or n-channel device, is turned on at any one time by a common gate voltage. Since

the two transistors are in series and one of them is turned off, very little current is drawn in the steady state. This device is very useful in low-power microelectronics. A channel stop is shown in *Figure 1.16*. It prevents the formation of an inversion between the n- and p-channel devices under interconnecting metallization.

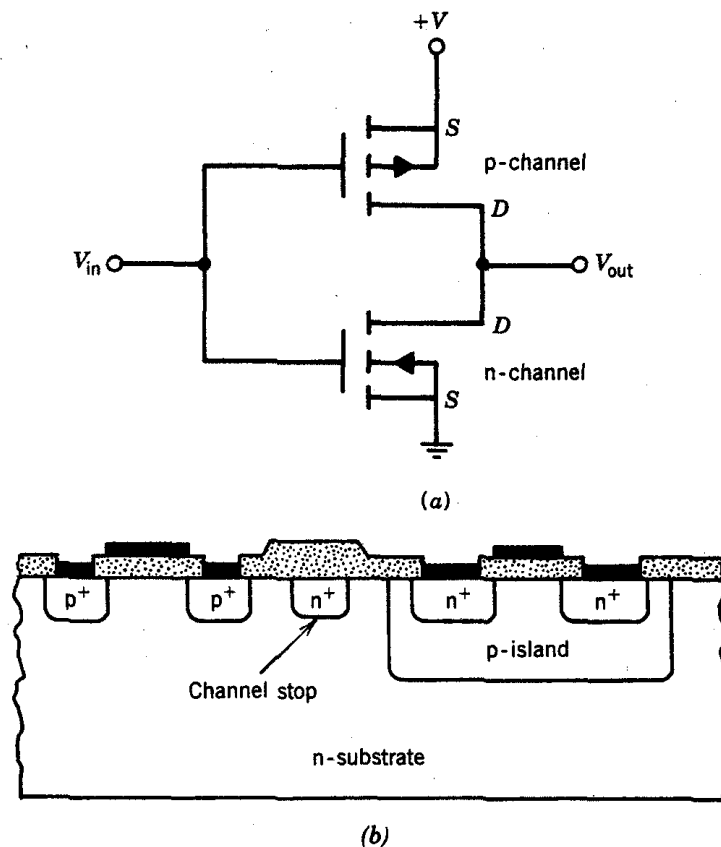


Figure 1.16: CMOS structure: (a) basic circuit and (b) cross-sections of p-channel and n-channel devices formed together.

1.11 Fabrication of Integrated Circuits

Figure 1.17 illustrates the sequence of forming a simple metal-oxide-semiconductor (MOS) capacitor. The fabrication process includes oxidation of silicon, patterning, etching, ion implantation, drive-in, and metallization, all of which will be reviewed in subsequent sections. Even for a simple MOS capacitor, the overall process can be quite onerous. To demonstrate the complexity of IC fabrication, **Figure 1.18** depicts a typical commercial process to produce a complementary MOS or CMOS logic gate device. In spite of its complexity over simple NMOS or PMOS, the CMOS structure reduces the overall power consumption and is used in many microprocessors and high-speed, low-voltage integrated circuits.

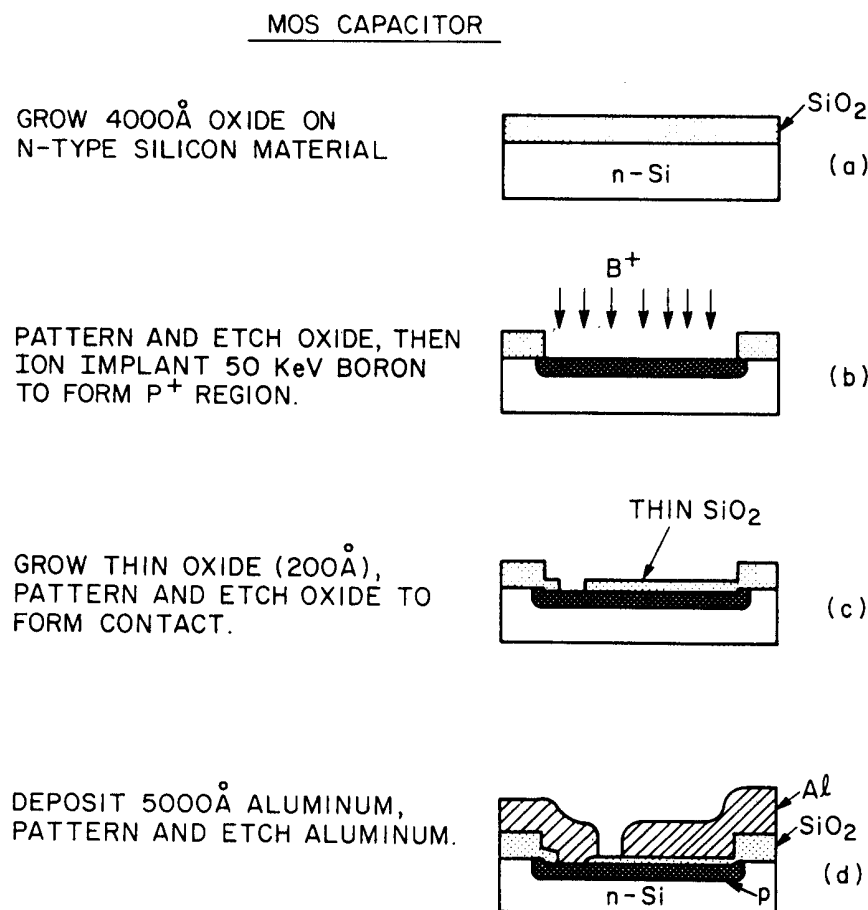


Figure 1.17: Steps to form an MOS capacitor.

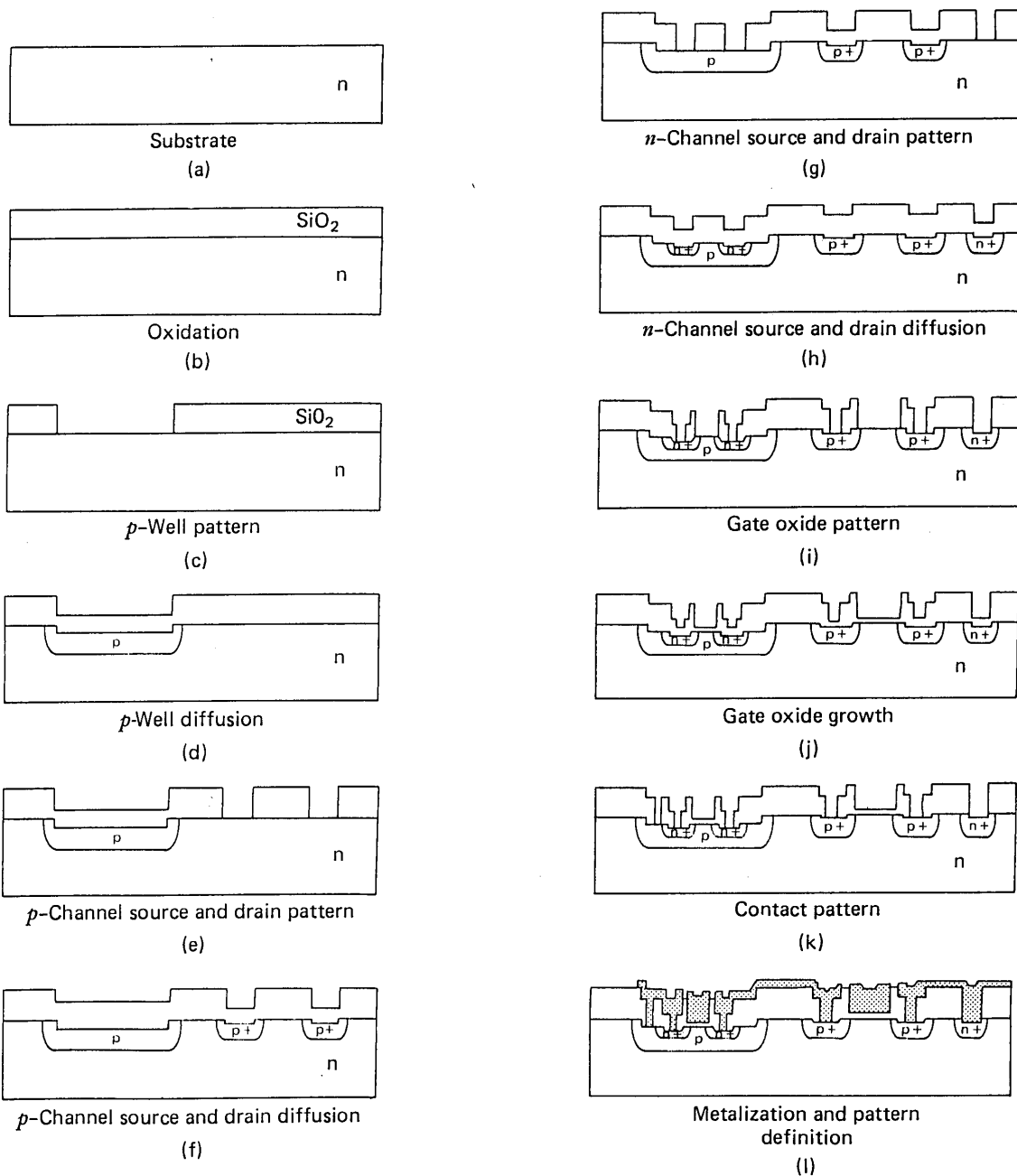


Figure 1.18: The CMOS process.