2 Carrier Modeling

Carriers are the entities that transport charge from place to place inside a material and hence give rise to electrical currents. In everyday life the most commonly encountered type of carrier is the electron, the subatomic particle responsible for charge transport in metallic wires. Within semiconductors one again encounters the familiar electron, but there is also a second equally important type of carrier—the hole. Electrons and holes are the focal point of this chapter, wherein we examine carrier related concepts, models, properties, and terminology.

Although reminders will be periodically interjected, it should be emphasized from the very start that the development throughout this chapter assumes *equilibrium* conditions exist within the semiconductor. "Equilibrium" is the term used to describe the unperturbed state of a system. Under equilibrium conditions there are no external voltages, magnetic fields, stresses, or other perturbing forces acting on the semiconductor. All observables are invariant with time. This "rest" condition, as we will see, provides an excellent *frame of reference.* Being able to characterize the semiconductor under equilibrium conditions permits one to extrapolate and ascertain the semiconductor's condition when a perturbation has been applied.

The reader should be forewarned that a number of formulas and facts will be presented without justification in this chapter. We would like to provide a complete discussion, properly developing every concept and deriving every formula. Unfortunately, this is not possible. It is our underlying philosophy, moreover, that being able to derive a result is secondary to knowing how to *interpret* and *use* a result. One can, of course, fill in any information gap through supplemental readings. For suggested references see the supplemental reading list at the end of Part I.

Lastly, please note that Appendix A provides a more in-depth introduction to the quantization concept and related topics considered in Section 2.1. Section 2.1 may be replaced by Appendix A with no loss in continuity.

2.1 THE QUANTIZATION CONCEPT

Instead of attempting to deal immediately with electrons in crystalline Si, where there are 14 electrons per atom and 5×10^{22} atoms/cm^3, we will take a more realistic approach and first establish certain ground rules by examining much simpler atomic systems. We begin

with the simplest of all atomic systems, the isolated hydrogen atom. This atom, as the reader may recall from a course in modern physics, was under intense scrutiny at the start of the twentieth century. Scientists of that time knew the hydrogen atom contained a negatively charged particle in orbit about a more massive positively charged nucleus. What they could not explain was the nature of the light emitted from the system when the hydrogen atom was heated to an elevated temperature. Specifically, the emitted light was observed at only certain discrete wavelengths; according to the prevailing theory of the time, scientists expected a continuum of wavelengths.

In 1913 Niels Bohr proposed a solution to the dilemma. Bohr hypothesized that the H-atom electron was restricted to certain well-defined orbits; or equivalently, Bohr assumed that the orbiting electron could take on only certain values of angular momentum. This "quantization" of the electron's angular momentum was, in turn, coupled directly to energy quantization. As can be readily established, if the electron's angular momentum is assumed to be $\mathbf{n}\hbar$, then

$$E_H = -\frac{m_0 q^4}{2(4\pi\varepsilon_0 \hbar \mathbf{n})^2} = -\frac{13.6}{\mathbf{n}^2}\ \text{eV}, \qquad \mathbf{n} = 1, 2, 3, \ldots \tag{2.1}$$

where (also see Fig. 2.1) E_H is the electron binding energy within the hydrogen atom, m_0 is the mass of a free electron, q is the magnitude of the electronic charge, ε_0 is the permittivity of free space, h is Planck's constant, $\hbar = h/2\pi$, and $\mathbf{n}$ is the energy quantum number or orbit identifier. The *electron volt* (eV) is a unit of energy equal to 1.6×10^{-19} joules. Now, with the electron limited to certain energies inside the hydrogen atom, it follows from the Bohr model that the transition from a higher $\mathbf{n}$ to a lower $\mathbf{n}$ orbit will release quantized energies of light; this explains the observation of emitted light at only certain discrete wavelengths.

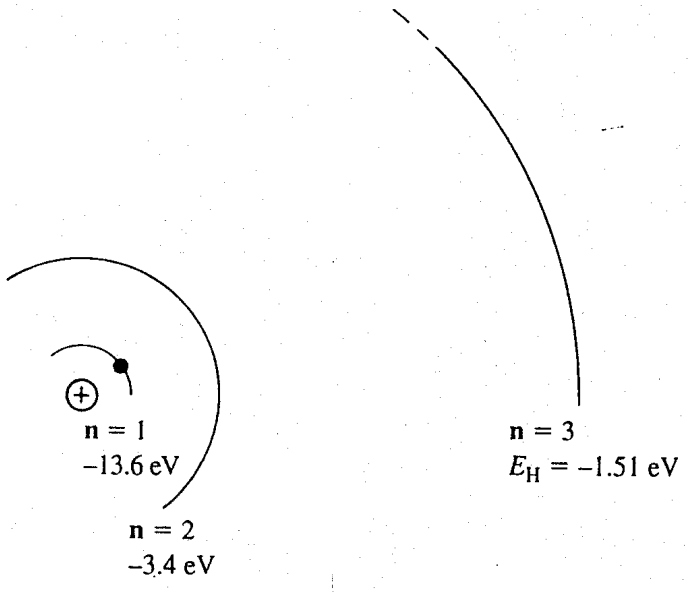


Figure 2.1 The hydrogen atom—idealized representation showing the first three allowed electron orbits and the associated energy quantization.

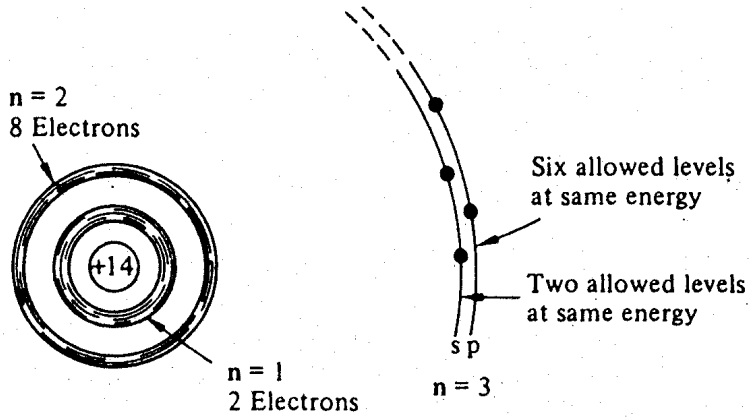


Figure 2.2 Schematic representation of an isolated Si atom.

For our purposes, the most important idea to be obtained from the Bohr model is that the energy of electrons in atomic systems is restricted to a limited set of values.† Relative to the hydrogen atom, the energy level scheme in a multi-electron atom like silicon is, as one might intuitively expect, decidedly more complex. It is still a relatively easy task, however, to describe the salient energy-related features of an isolated silicon atom. As pictured in Fig. 2.2, ten of the 14 Si-atom electrons occupy very deep-lying energy levels and are tightly bound to the nucleus of the atom. The binding is so strong, in fact, that these ten electrons remain essentially unperturbed during chemical reactions or normal atom–atom interactions, with the ten-electron-plus-nucleus combination often being referred to as the "core" of the atom. The remaining four Si-atom electrons, on the other hand, are rather weakly bound and are collectively called the *valence electrons* because of their strong participation in chemical reactions. As emphasized in Fig. 2.2, the four valence electrons, if unperturbed, occupy four of the eight allowed slots (or states) having the next highest energy above the deep-lying core levels. Finally, we should note, for completeness, that the electronic configuration in the 32-electron Ge-atom (germanium being the other elemental semiconductor) is essentially identical to the Si-atom configuration except that the Ge-core contains 28 electrons.

2.2 SEMICONDUCTOR MODELS

Building on the information presented in previous sections, we introduce and describe in this section two very important models or visualization aids that are used extensively in the analysis of semiconductor devices. The inclusion of semiconductor models in a chapter devoted to carrier modeling may appear odd at first glance but is actually quite appropriate. We are, in effect, modeling the carrier "container," the semiconductor crystal.

† Actually, not only energy but many other observables relating to atomic-sized particles are quantized. An entire field of study, Quantum Mechanics, has been developed to describe the properties and actions of atomic-sized particles and systems.

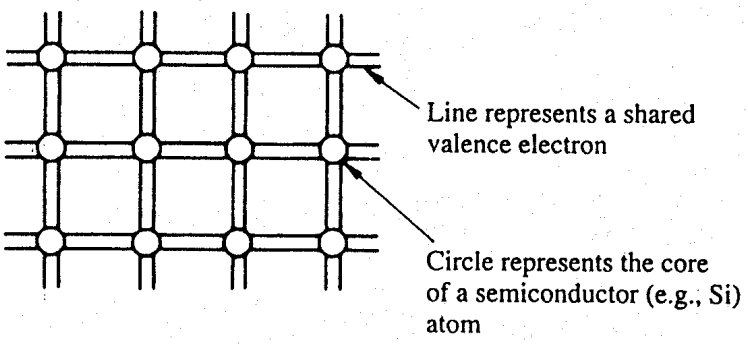


Figure 2.3 The bonding model.

2.2.1 Bonding Model

The isolated Si atom, or a Si atom not interacting with other atoms, was found to contain four valence electrons. Si atoms incorporated in the diamond lattice, on the other hand, exhibit a bonding that involves an attraction between each atom and its four nearest neighbors (refer to Fig. 1.4c). The implication here is that, in going from isolated atoms to the collective crystalline state, Si atoms come to share one of their valence electrons with each of the four nearest neighbors. This covalent bonding, or equal sharing of valence electrons with nearest neighbors, and the mere fact that atoms in the diamond lattice have four nearest neighbors, give rise to the idealized semiconductor representation, the bonding model, shown in Fig. 2.3. Each circle in the Fig. 2.3 bonding model represents the core of a semiconductor atom, while each line represents a shared valence electron. (There are eight lines connected to each atom because any given atom not only *contributes* four shared electrons but must also *accept* four shared electrons from adjacent atoms.) The two-dimensional nature of the model is, of course, an idealization that facilitates mental visualizations.

Although considerable use will be made of the bonding model in subsequent discussions, it is nevertheless worthwhile at this point to examine sample applications of the model to provide some indication of the model's utility. Two sample applications are presented in Fig. 2.4. In Fig. 2.4(a) we use the bonding model to picture a point defect, a missing atom, in the lattice structure. In Fig. 2.4(b) we visualize the breaking of an atom-to-atom bond and the associated release or freeing of an electron. [Bond breaking (at $T > 0$ K) and defects occur naturally in all semiconductors, and hence (if we may be somewhat overcritical) the basic model of Fig. 2.3 is strictly valid for an entire semiconductor only at $T \simeq 0$ K when the semiconductor in question contains no defects and no impurity atoms.]

2.2.2 Energy Band Model

If our interests were restricted to describing the spatial aspects of events taking place within a semiconductor, the bonding model alone would probably be adequate. Quite often, however, one is more interested in the energy-related aspects of an event. In such instances the bonding model, which says nothing about electron energies, is of little value and the energy band model becomes the primary visualization aid.

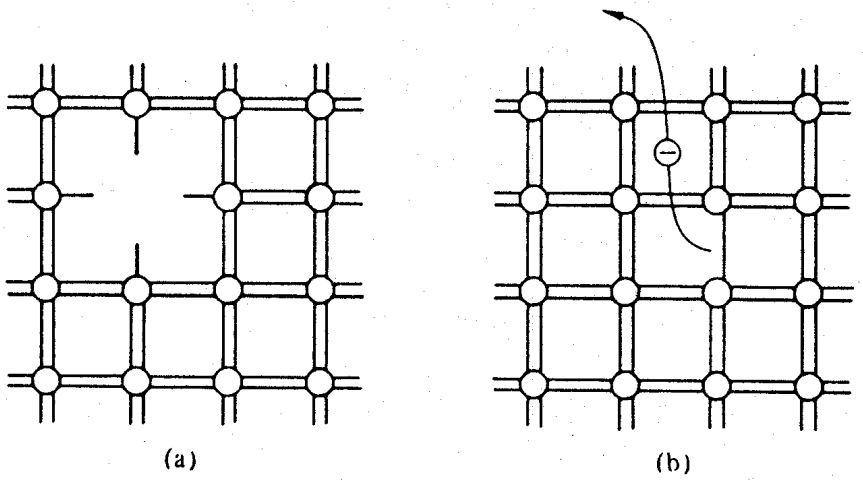


Figure 2.4 Sample utilization of the bonding model. (a) Visualization of a missing atom or point defect. (b) Breaking of an atom-to-atom bond and freeing of an electron.

Let us begin the conceptual path leading to the energy band model by recalling the situation inside an isolated Si atom. Reviewing the Section 2.1 discussion, ten of the 14 electrons inside an isolated Si atom are tightly bound to the nucleus and are unlikely to be significantly perturbed by normal atom–atom interactions. The remaining four electrons are rather weakly bound and, if unperturbed, occupy four of the eight allowed energy states immediately above the last core level. Moreover, it is implicitly understood that the electronic energy states within a group of Si atoms, say N Si atoms, are all identical—as long as the atoms are isolated, that is, far enough apart so that they are noninteracting.

Given the foregoing knowledge of the isolated atom situation, the question next arises as to whether we can use the knowledge to deduce information about the crystalline state. Assuredly we can omit any further mention of the core electrons because these electrons are not significantly perturbed by normal interatomic forces. The opposite, however, is true of the valence electrons. If N atoms are brought into close proximity (the case in crystalline Si), it is quite reasonable to expect a modification in the energy states of the valence electrons.

The modification in the valence-electron energy states actually known to take place is summarized in Fig. 2.5. Starting with N-isolated Si atoms, and conceptually bringing the atoms closer and closer together, one finds the interatomic forces lead to a progressive spread in the allowed energies. The spread in energies gives rise to closely spaced sets of allowed states known as *energy bands.* At the interatomic distance corresponding to the Si lattice spacing, the distribution of allowed states consists of two bands separated by an intervening energy gap. The upper band of allowed states is called the *conduction band;* the lower band of allowed states, the *valence band;* and the intervening energy gap, the *forbidden gap* or *band gap.* In filling the allowed energy band states, electrons tend to gravitate to the lowest possible energies. Noting that electrons are restricted to single occupancy in allowed states (the Pauli exclusion principle) and remembering that the $4N$ valence band states can just accommodate what were formerly $4N$ valence electrons, we typically find that the valence band is almost completely filled with electrons and the

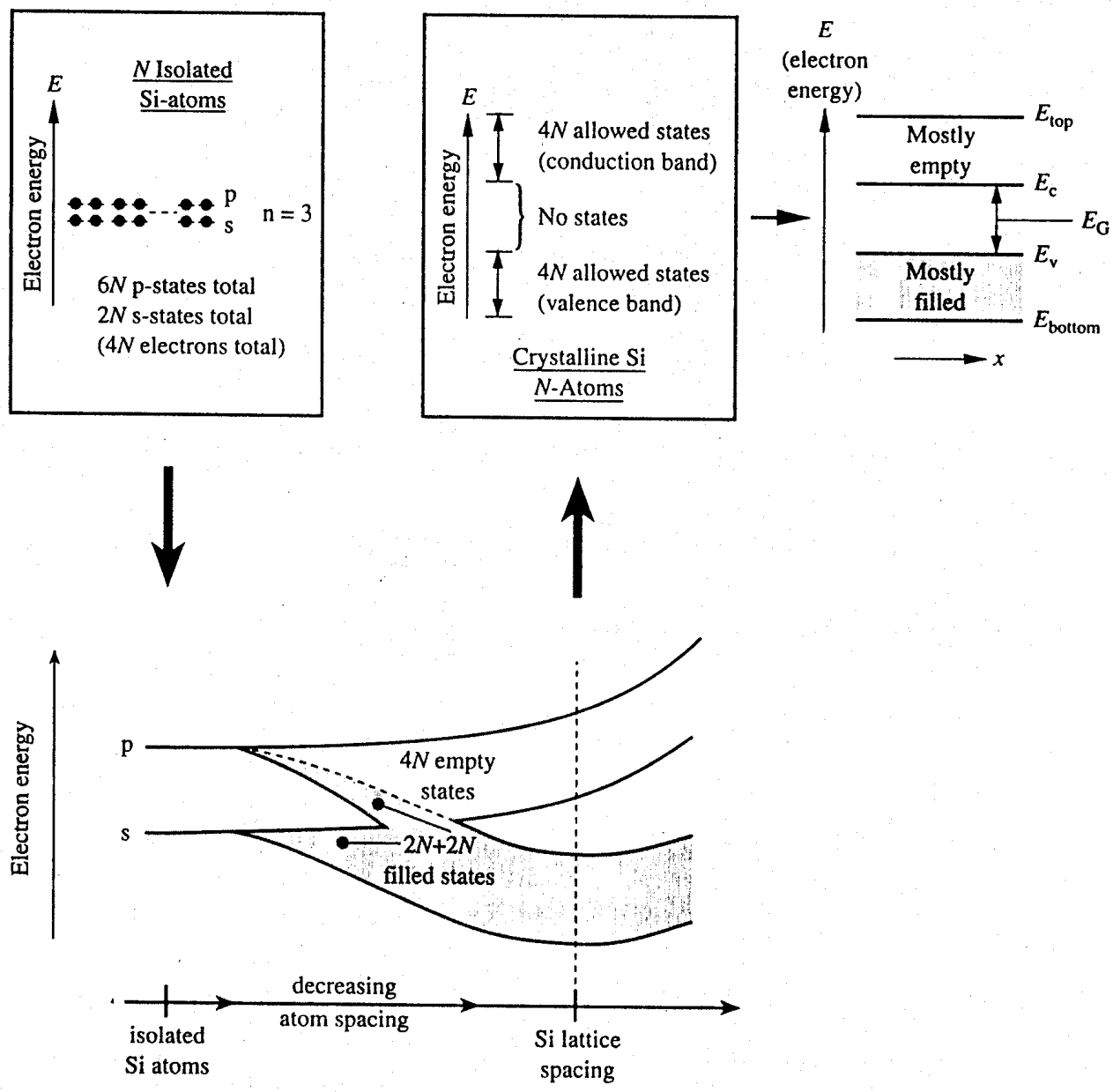


Figure 2.5 Conceptual development of the energy band model starting with N isolated Si atoms on the top left and concluding with a "dressed-up" version of the energy band model on the top right.

conduction band is all but devoid of electrons. Indeed, the valence band is completely filled and the conduction band completely empty at temperatures approaching $T = 0$ K.

To complete our plausibilization of the energy band model, we need to introduce and utilize one additional fact: Unlike the valence electrons in the isolated-atom case, the band electrons in crystalline silicon are not tied to or associated with any one particular atom. True, on the average, one will typically find four electrons being shared between any given Si atom and its four nearest neighbors (as in the bonding model). However, the identity of the shared electrons changes as a function of time, with the electrons moving around from point to point in the crystal. In other words, the allowed electronic states are no longer atomic states but are associated with the crystal as a whole; independent of the point ex-

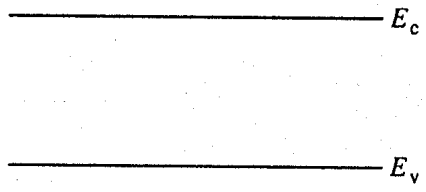


Figure 2.6 The energy band diagram—widely employed simplified version of the energy band model.

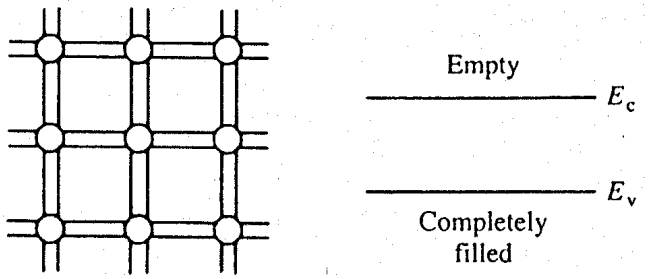
amined in a perfect crystal, one sees the same allowed-state configuration. We therefore conclude that for a perfect crystal under equilibrium conditions a plot of the allowed electron energies versus distance along any preselected crystalline direction (always called the x-direction) is as shown on the right-hand side of Fig. 2.5. The cited plot, a plot of allowed electron energy states as a function of position, is the basic energy band model. E_c introduced in the Fig. 2.5 plot is the lowest possible conduction band energy, E_v is the highest possible valence band energy, and $E_G = E_c - E_v$ is the band gap energy.

Finally, Fig. 2.6 displays the form of the basic energy band model encountered in practice. In this widely employed "shorthand" version, the line to indicate the top energy in the conduction band, the line to indicate the bottom energy in the valence band, the fill pattern drawing attention to mostly filled states in the valence band, the labeling of the y- or electron-energy axis, and the labeling of the x- or position axis are all understood to exist implicitly, but are not shown explicitly.

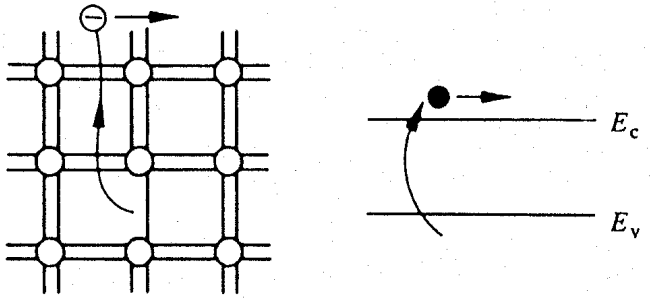
2.2.3 Carriers

With the semiconductor models having been firmly established, we are now in a position to introduce and to visualize the current-carrying entities within semiconductors. Referring to Fig. 2.7, we note first from part (a) that there are no carriers or possible current flow if the bonding model has no broken bonds. Equivalently, in the energy band model, if the valence band is completely filled with electrons and the conduction band is devoid of electrons, there are no carriers or possible current flow. This lack of carriers and associated current flow is easy to understand in terms of the bonding model where the shared electrons are viewed as being tied to the atomic cores. As accurately portrayed in the energy band model, however, the valence band electrons actually move about in the crystal. How is it then that no current can arise from this group of electrons? As it turns out, the momentum of the electrons is quantized in addition to their energy. Moreover, for each and every possible momentum state in a band, there is another state with an oppositely directed momentum of equal magnitude. Thus, if a band is completely filled with electrons, the *net* momentum of the electrons in the band is always identically zero. It follows that no current can arise from the electrons in a completely filled energy band.

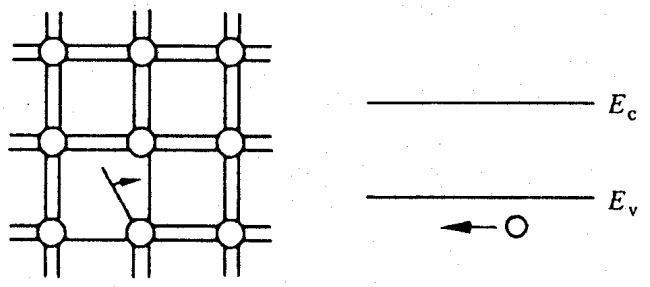
The electrons that do give rise to charge transport are visualized in Fig. 2.7(b). When a Si–Si bond is broken and the associated electron is free to wander about the lattice, the released electron is a carrier. Equivalently, in terms of the energy band model, excitation of valence band electrons into the conduction band creates carriers; that is, *electrons in the*



(a) No carriers



(b) The electron



(c) The hole

Figure 2.7 Visualization of carriers using the bonding model (left) and the energy band model (right). (a) No-carrier situation; (b) visualization of an electron; (c) visualization of a hole.

conduction band are carriers. Note that the energy required to break a bond in the bonding model and the band gap energy, E_G, are one and the same thing. Likewise, freed bonding-model electrons and conduction band electrons are just different names for the same electrons; in subsequent discussions the word "electrons," when used without a modifier, will be understood to refer to these conduction band electrons.

In addition to releasing an electron, the breaking of a Si–Si bond also creates a missing bond or void in the bonding structure. Thinking in terms of the bonding model, one can visualize the movement of this missing bond from place to place in the lattice as a result of nearby bound electrons jumping into the void (see Fig. 2.7c). Alternatively, one can think in terms of the energy band model where the removal of an electron from the valence band creates an empty state in an otherwise vast sea of filled states. The empty state, like a bubble in a liquid, moves about rather freely in the lattice because of the cooperative motion of the valence band electrons. What we have been describing, the missing bond in the bonding

scheme, the empty state in the valence band, is the second type of carrier found in semiconductors—the hole.

Although perhaps not obvious from the preceding introductory description, the valence band hole is an entity on an equal footing with the conduction band electron. Both electrons and holes participate in the operation of most semiconductor devices. Holes are even the primary carrier in some devices. As one's familiarity with carrier modeling grows, the comparable status of electrons and holes becomes more and more apparent. Ultimately, it is commonplace to conceive of the hole as a type of subatomic particle.

2.2.4 Band Gap and Material Classification

We conclude this section by citing a very important tie between the band gap of a material, the number of carriers available for transport in a material, and the overall nature of a material. As it turns out, although specifically established for semiconductors, the energy band model of Fig. 2.6 can be applied with only slight modification to all materials. The major difference between materials lies not in the nature of the energy bands, but rather in the magnitude of the energy gap between the bands.

Insulators, as illustrated in Fig. 2.8(a), are characterized by very wide band gaps, with E_G for the insulating materials diamond and SiO_2 being $\simeq$5 eV and $\simeq$8 eV, respectively. In these wide band gap materials the thermal energy available at room temperature excites very few electrons from the valence band into the conduction band; thus very few carriers exist inside the material and the material is therefore a poor conductor of current. The band

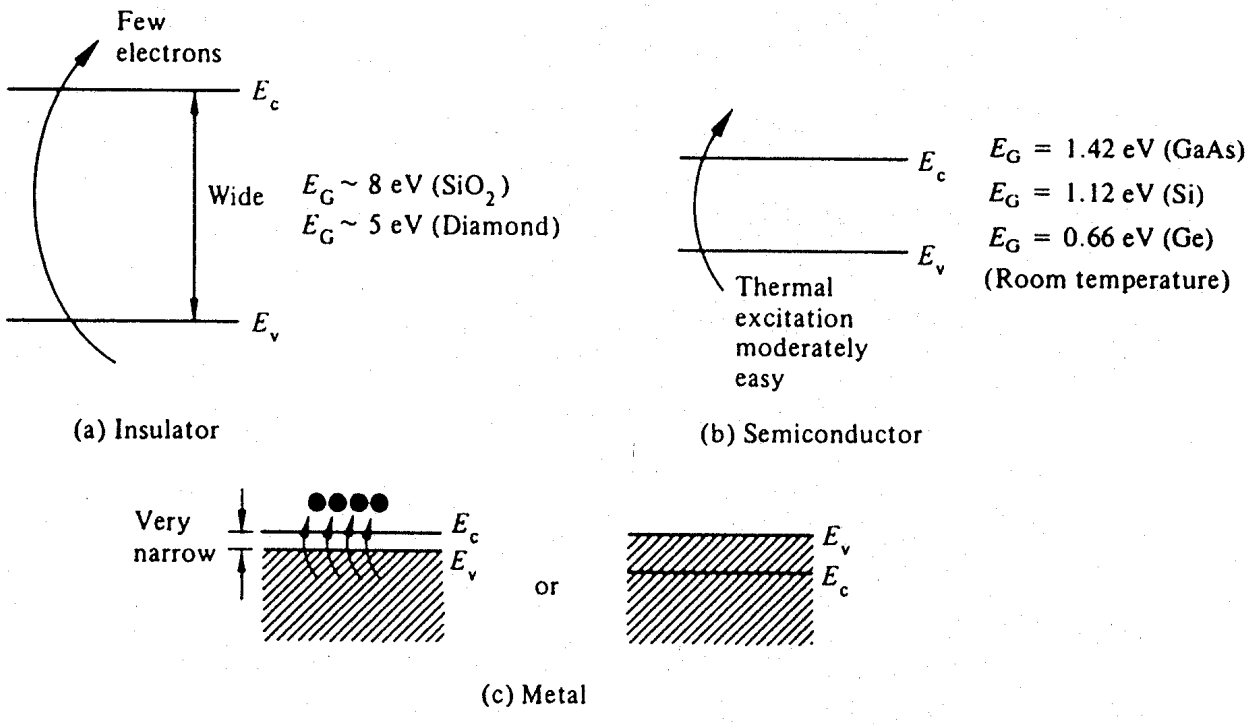


Figure 2.8 Explanation of the distinction between (a) insulators, (b) semiconductors, and (c) metals using the energy band model.

gap in metals, by way of comparison, is either very small, or no band gap exists at all due to an overlap of the valence and conduction bands (see Fig. 2.8c). There is always an abundance of carriers in metals, and hence metals are excellent conductors. Semiconductors present an intermediate case between insulators and metals. At room temperature ($T = 300$ K), $E_G = 1.42$ eV in GaAs, $E_G = 1.12$ eV in Si, and $E_G = 0.66$ eV in Ge. Thermal energy, by exciting electrons from the valence band into the conduction band, creates a moderate number of carriers in these materials, giving rise in turn to a current-carrying capability intermediate between poor and excellent.

2.3 CARRIER PROPERTIES

Having formally introduced the electron and the hole, we next seek to learn as much as possible about the nature of these carriers. In this particular section we examine a collage of carrier-related information, information of a general nature including general facts, properties, and terminology.

2.3.1 Charge

Both electrons and holes are charged entities. Electrons are negatively charged, holes are positively charged, and the *magnitude* of the carrier charge, q, is the same for the two types of carriers. To three-place accuracy in MKS units, $q = 1.60 \times 10^{-19}$ coul. Please note that, under the convention adopted herein, the electron and hole charges are $-q$ and $+q$, respectively; i.e., the sign of the charge is displayed explicitly.

2.3.2 Effective Mass

Mass, like charge, is another very basic property of electrons and holes. Unlike charge, however, the carrier mass is not a simple property and cannot be disposed of simply by quoting a number. Indeed, the apparent or "effective" mass of electrons within a crystal is a function of the semiconductor material (Si, Ge, etc.) and is different from the mass of electrons within a vacuum.

Seeking to obtain insight into the effective mass concept, let us first consider the motion of electrons in a vacuum. If, as illustrated in Fig. 2.9(a), an electron of rest mass m_0 is moving in a vacuum between two parallel plates under the influence of an electric field $\mathscr{E}$, then, according to Newton's second law, the force $\mathbf{F}$ on the electron will be

$$\mathbf{F} = -q\mathscr{E} = m_0 \frac{dv}{dt} \tag{2.2}$$

where v is the electron velocity and t is time. Next consider electrons (conduction band electrons) moving between the two parallel end faces of a semiconductor crystal under the influence of an applied electric field, as envisioned in Fig. 2.9(b). Does Eq. (2.2) also describe the overall motion of electrons within the semiconductor crystal? The answer is *no*.

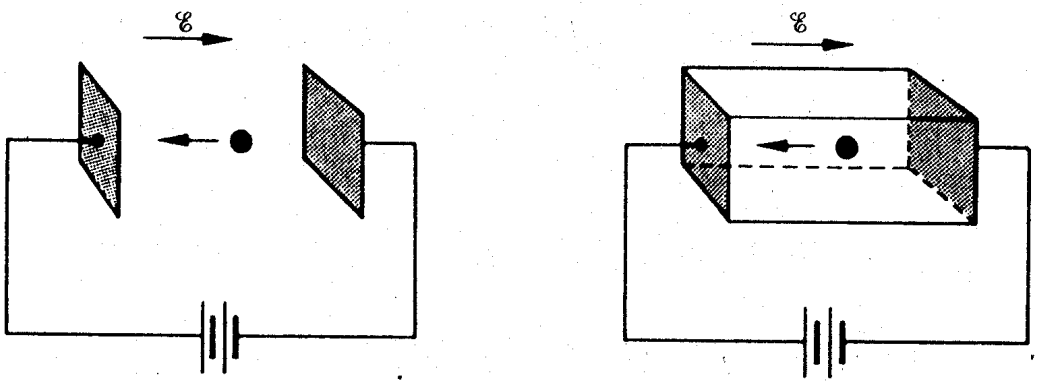


Figure 2.9 An electron moving in response to an applied electric field (a) within a vacuum, and (b) within a semiconductor crystal.

Electrons moving inside a semiconductor crystal will collide with semiconductor atoms, thereby causing a periodic deceleration of the carriers. However, should not Eq. (2.2) apply to the portion of the electronic motion occurring *between* collisions? The answer is again *no.* In addition to the applied electric field, electrons in a crystal are also subject to complex crystalline fields not specifically included in Eq. (2.2).

The foregoing discussion delineated certain important differences between electrons in a crystal and electrons in a vacuum, but it left unresolved the equally important question as to how one properly describes the motion of carriers in a crystal. Strictly speaking, the motion of carriers in a crystal can be described only by using Quantum Mechanics, the formalism appropriate for atomic-sized systems. Fortunately, if the dimensions of the crystal are large compared to atomic dimensions, the complex quantum mechanical formulation for the carrier motion between collisions simplifies to yield an equation of motion identical to Eq. (2.2), except that m_0 is replaced by an effective carrier mass. In other words, for the Fig. 2.9(b) electrons one can write

$$\mathbf{F} = -q\mathscr{E} = m_n^* \frac{dv}{dt} \tag{2.3}$$

where m_n^* is the electron effective mass. A similar equation can be written for holes with $-q \rightarrow q$ and $m_n^* \rightarrow m_p^*$. In each case the internal crystalline fields and quantum-mechanical effects are all suitably lumped into the effective mass factor multiplying dv/dt. This is a very significant result. *It allows us to conceive of electrons and holes as quasi-classical particles and to employ classical particle relationships in most device analyses.*

Although the effective mass formulation is a significant simplification, it should be mentioned that the carrier acceleration can vary with the direction of travel in a crystal; i.e., the effective masses can have multiple components. Moreover, depending on how a macroscopic observable is related to the carrier motion, a different grouping of mass components can lead to a different m^* being utilized in a particular relationship. There are, for example, cyclotron resonance effective masses, conductivity effective masses, density of

Table 2.1 Density of States Effective Masses at 300 K.

Material	m_n^*/m_0	m_p^*/m_0
Si	1.18	0.81
Ge	0.55	0.36
GaAs	0.066	0.52

states effective masses, among others. It is also probably not too surprising that the effective masses vary somewhat with temperature. Herein we make direct use of only the density of states effective masses (see Subsection 2.4.1). The density of states effective masses for electrons and holes in Si, Ge, and GaAs at 300 K are listed in Table 2.1.

2.3.3 Carrier Numbers in Intrinsic Material

The term *intrinsic semiconductor* in common usage refers to an extremely pure semiconductor sample containing an insignificant amount of impurity atoms. More precisely, an intrinsic semiconductor is a semiconductor whose properties are native to the material (that is, not caused by external additives). The number of carriers in an intrinsic semiconductor fits into the scheme of things in that it is an identifiable intrinsic property of the material.

Defining, quite generally,

$$n = \text{number of electrons/cm}^3$$

$$p = \text{number of holes/cm}^3$$

existing inside a semiconductor, then, given an intrinsic semiconductor under equilibrium conditions, one finds

$$n = p = n_i \tag{2.4}$$

and

$$\left.\begin{aligned} n_i &\simeq 2 \times 10^6/\text{cm}^3 \quad \text{in GaAs} \\ &\simeq 1 \times 10^{10}/\text{cm}^3 \quad \text{in Si} \\ &\simeq 2 \times 10^{13}/\text{cm}^3 \quad \text{in Ge} \end{aligned}\right\} \text{at room temperature}$$

The electron and hole concentrations in an intrinsic semiconductor are equal because carriers within a very pure material can be created only in pairs. Referring to Fig. 2.7, if a semiconductor bond is broken, a free electron and a missing bond or hole are created simultaneously. Likewise, the excitation of an electron from the valence band into the con-

duction band automatically creates a valence band hole along with the conduction band electron. Also note that the intrinsic carrier concentration, although large in an absolute sense, is relatively small compared with the number of bonds that could be broken. For example, in Si there are 5×10^{22} atoms/cm^3 and four bonds per atom, making a grand total of 2×10^{23} bonds or valence band electrons per cm^3. Since $n_i \simeq 10^{10}$/cm^3, one finds less than one bond in 10^{13} broken in Si at room temperature. To accurately represent the situation inside intrinsic Si at room temperature, we could cover all the university chalkboards in the world with the bonding model and possibly show only *one* broken bond.

2.3.4 Manipulation of Carrier Numbers—Doping

Doping, in semiconductor terminology, is the addition of controlled amounts of specific impurity atoms with the express purpose of increasing either the electron or the hole concentration. The addition of dopants in controlled amounts to semiconductor materials occurs routinely in the fabrication of almost all semiconductor devices. Common Si dopants are listed in Table 2.2. To increase the electron concentration, one can add either phosphorus, arsenic, or antimony atoms to the Si crystal, with phosphorus followed closely by arsenic being the most commonly employed donor (electron-increasing) dopants. To increase the hole concentration, one adds either boron, gallium, indium, or aluminum atoms to the Si crystal, with boron being the most commonly employed acceptor (hole-increasing) dopant.

To understand how the addition of impurity atoms can lead to a manipulation of carrier numbers, it is important to note that the Table 2.2 donors are all from Column V in the Periodic Table of Elements, while all of the cited acceptors are from Column III in the Periodic Table of Elements. As visualized in Fig. 2.10(a) using the bonding model, when a Column V element with five valence electrons is substituted for a Si atom in the semiconductor lattice, four of the five valence electrons fit snugly into the bonding structure. The fifth donor electron, however, does not fit into the bonding structure and is weakly bound to the donor site. At room temperature this electron is readily freed to wander about the lattice and hence becomes a carrier. Please note that this donation (hence the name "donor") of carrier electrons does not increase the hole concentration. The donor ion left behind when the fifth electron is released cannot move, and there are no broken atom–atom bonds associated with the release of the fifth electron.

Table 2.2 Common Silicon Dopants. Arrows indicate the most widely employed dopants.

Donors (Electron-increasing dopants)		*Acceptors (Hole-increasing dopants)*	
P ←	Column V elements	B ←	Column III elements
As ←		Ga	
Sb		In	
		Al	

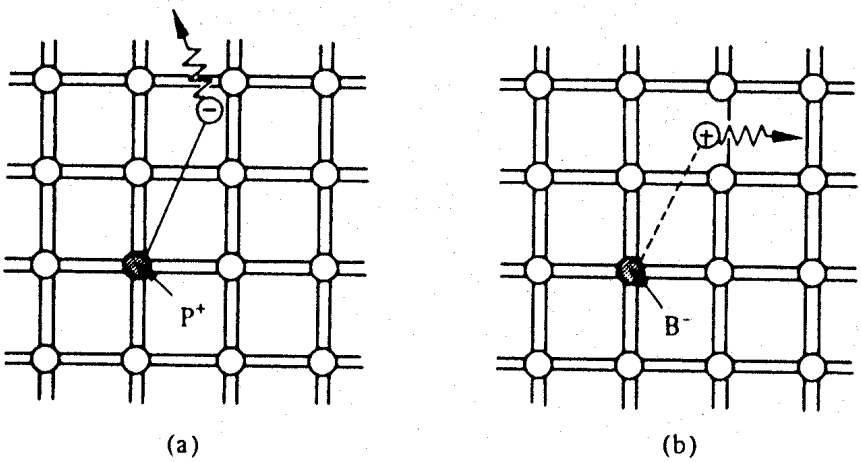


Figure 2.10 Visualization of (a) donor and (b) acceptor action using the bonding model. In (a) the Column V element P is substituted for a Si atom; in (b) the Column III element B is substituted for a Si atom.

The explanation of acceptor action follows a similar line of reasoning. The Column III acceptors have three valence electrons and cannot complete one of the semiconductor bonds when substituted for Si atoms in the semiconductor lattice (see Fig. 2.10b). The Column III atom, however, readily accepts (hence the name "acceptor") an electron from a nearby Si–Si bond, thereby completing its own bonding scheme and in the process creating a hole that can wander about the lattice. Here again there is an increase in only one type of carrier. The negatively charged acceptor ion (acceptor atom plus accepted electron) cannot move, and no electrons are released in the hole-creation process.

The foregoing bonding-model-based explanation of dopant action is reasonably understandable. There are, nevertheless, a few loose ends. For one, we noted that the fifth donor electron was rather weakly bound and readily freed at room temperature. How does one interpret the relative term "weakly bound"? It takes $\simeq$1 eV to break Si–Si bonds and very few of the Si–Si bonds are broken at room temperature. Perhaps "weakly bound" means a binding energy $\simeq$0.1 eV or less? The question also arises as to how one visualizes dopant action in terms of the energy band model. Both of the cited questions, as it turns out, involve energy considerations and are actually interrelated.

Let us concentrate first on the binding energy of the fifth donor electron. Crudely speaking, the positively charged donor-core-plus-fifth-electron may be likened to a hydrogen atom (see Fig. 2.11). Conceptually, the donor core replaces the hydrogen-atom nucleus and the fifth donor electron replaces the hydrogen-atom electron. In the real hydrogen atom the electron moves of course in a vacuum, can be characterized by the mass of a free electron, and, referring to Eq. (2.1), has a ground-state binding energy of -13.6 eV. In the pseudo-hydrogen atom, on the other hand, the orbiting electron moves through a sea of Si atoms and is characterized by an effective mass. Hence, in the donor or pseudo-atom case, the permittivity of free space must be replaced by the permittivity of Si and m_0 must be replaced by m_n^*. We therefore conclude that the binding energy (E_B) of the fifth donor electron is approximately

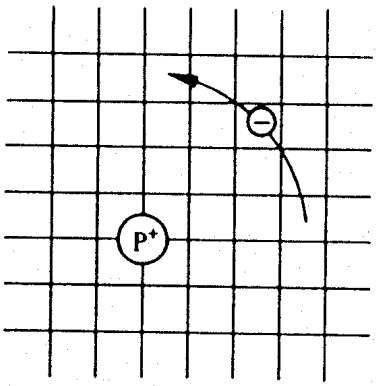


Figure 2.11 Pseudo-hydrogen atom model for the donor-site bond.

$$E_B \simeq -\frac{m_n^* q^4}{2(4\pi K_S \varepsilon_0 \hbar)^2} = \frac{m_n^*}{m_0}\frac{1}{K_S^2} E_{H|n=1} \simeq -0.1 \text{ eV} \tag{2.5}$$

where K_S is the Si dielectric constant ($K_S = 11.8$). Actual donor-site binding energies in Si are listed in Table 2.3. The observed binding energies are seen to be in good agreement with the Eq. (2.5) estimate and, confirming the earlier speculation, are roughly $\simeq 1/20$ the Si band gap energy.

Having established the strength of dopant-site bonds, we are now in a position to tackle the problem of how one visualizes dopant action using the energy band model. Working on the problem, we note first that when an electron is released from a donor it becomes a conduction band electron. If the energy absorbed at the donor is precisely equal to the electron binding energy, the released electron will moreover have the lowest possible energy in the conducting band—namely, E_c. Adding an energy $|E_B|$ to the bound electron, in other words, *raises* the electron's energy to E_c. Hence, we are led to conclude that the bound electron occupies an allowed electronic level an energy $|E_B|$ below the conduction band edge, or, as visualized in Fig. 2.12, donor sites can be incorporated into the energy band scheme by adding allowed electronic levels at an energy $E_D = E_c - |E_B|$. Note that the donor energy level is represented by a set of dashes, instead of a continuous line, because an electron bound to a donor site is localized in space; i.e., a bound electron does not leave the general Δx vicinity of the donor. The relative closeness of E_D to E_c of course reflects the fact that $E_c - E_D = |E_B| \simeq (1/20)E_G(\text{Si})$.

Table 2.3 Dopant-Site Binding Energies.

Donors	$\lvert E_B \rvert$	*Acceptors*	$\lvert E_B \rvert$
Sb	0.039 eV	B	0.045 eV
P	0.045 eV	Al	0.067 eV
As	0.054 eV	Ga	0.072 eV
		In	0.16 eV

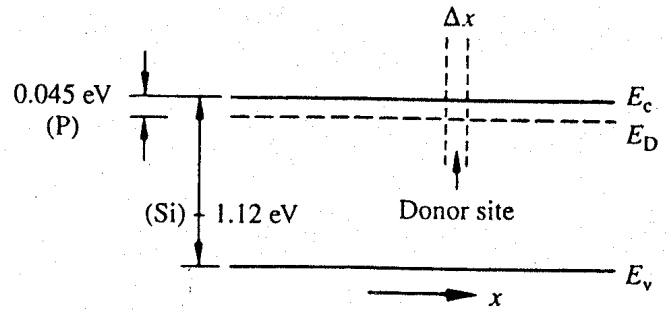


Figure 2.12 Addition of the $E = E_D$ donor levels to the energy band diagram. Dashes of width Δx emphasize the localized nature of the bound donor-site states.

The actual visualization of dopant action using the energy band model is pictured in Fig. 2.13. Examining the left-hand side of Fig. 2.13(a), one finds all the donor sites filled with bound electrons at temperatures $T \rightarrow 0$ K. This is true because very little thermal energy is available to excite electrons from the donor sites into the conduction band at these very low temperatures. The situation changes, of course, as the temperature is increased, with more and more of the weakly bound electrons being donated to the conduction band. At room temperature the ionization of the donors is all but total, giving rise to the situation pictured at the extreme right of Fig. 2.13(a). Although we have concentrated on donors, the situation for acceptors is completely analogous. As visualized in Fig. 2.13(b), acceptors introduce allowed electronic levels into the forbidden gap at an energy slightly above the valence band edge. At low temperatures, all of these sites will be empty—there is insuffi-

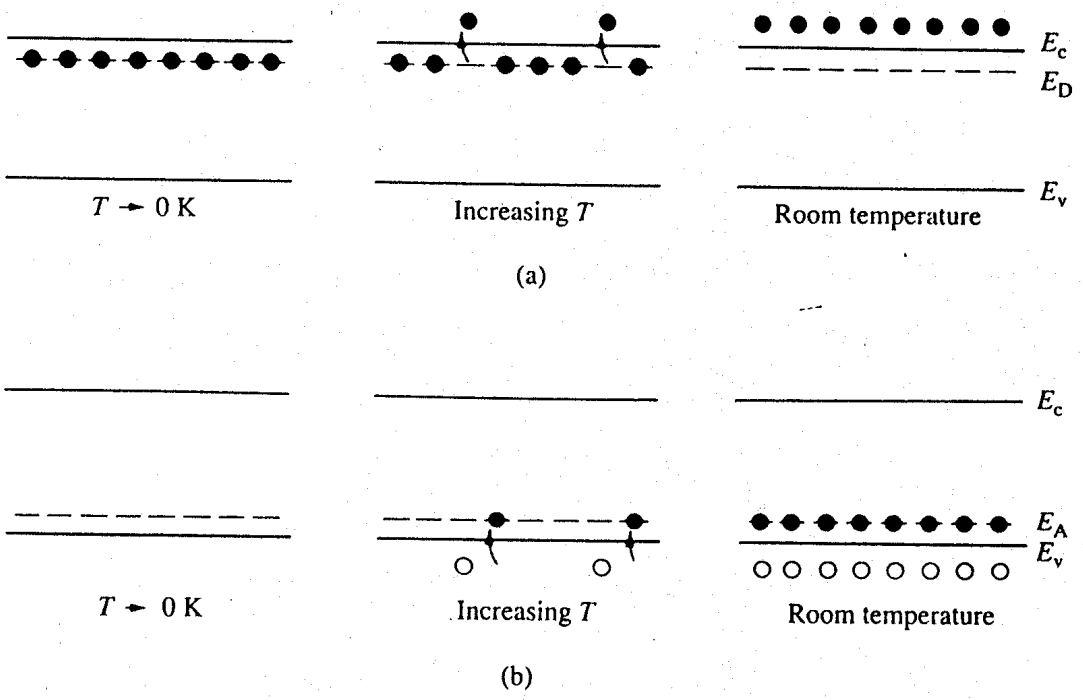


Figure 2.13 Visualization of (a) donor and (b) acceptor action using the energy band model.

cient energy at temperatures $T \rightarrow 0$ K for a valence band electron to make the transition to an acceptor site. Increasing temperature, implying an increased store of thermal energy, facilitates electrons jumping from the valence band onto the acceptor levels. The removal of electrons from the valence band of course creates holes. At room temperature, essentially all of the acceptor sites are filled with an excess electron and an increased hole concentration is effected in the material.

To complete this subsection, a few words are in order concerning the doping of non-elemental semiconductors such as GaAs. Dopant action in GaAs follows the same general principles but is slightly more involved because of the existence of two different lattice-site atoms. Completely analogous to the doping of a Si crystal, the Column VI elements S, Se, and Te act as donors when replacing the Column V element As in the GaAs lattice. Similarly, the Column II elements Be, Mg, and Zn act as acceptors when replacing the Column III element Ga. A new situation arises when Column IV elements such as Si and Ge are incorporated into GaAs. Si typically replaces Ga in the GaAs lattice and is a popular *n*-type dopant. However, under certain conditions Si can be made to replace As in the GaAs lattice, thereby functioning as an acceptor. In fact, GaAs *pn* junctions have been fabricated where Si was both the *p*-side and *n*-side dopant. An impurity that can act as either a donor or an acceptor is referred to as an *amphoteric* dopant.

Exercise 2.1

Energy Quiz

P: (a) 1 eV is equal to how many joules of energy?

(b) kT is equal to how many eV at 300 K?

(c) The ionization energy of acceptors and donors in Si is roughly equal to _____?

(d) $E_G(\text{Si}) =$ _____?

(e) $E_G(\text{SiO}_2) =$ _____?

(f) The energy required to ionize a hydrogen atom initially in the $\mathbf{n} = 1$ state is ___?

S: (a) 1 eV $= 1.60 \times 10^{-19}$ joules.

(b) $kT = (8.617 \times 10^{-5})(300) = 0.0259$ eV at 300 K.

(c) The ionization energy of dopants is equal to $|E_B|$ of the dopant sites. As discussed in Subsection 2.3.4, $|E_B| \cong 0.1$ eV for dopants in Si.

(d) $E_G(\text{Si}) = 1.12$ eV at 300 K (see Subsection 2.2.4).

(e) $E_G(\text{SiO}_2) \cong 8$ eV (see Subsection 2.2.4).

(f) Making use of Eq. (2.1) in Section 2.1, the hydrogen atom ionization energy is $|E_{H|n=1}| = 13.6$ eV.

2.3.5 Carrier-Related Terminology

Since terminology is often a stumbling block to understanding and since this particular section is replete with specialized terms, it seems appropriate to conclude the section with an overview of carrier-related terminology. Approximately one half of the carrier-related terms that follow were introduced and defined earlier in this section; the remainder of the terms are being listed here for the first time. All of the terms are widely employed and their definitions should be committed to memory.

Dopants—specific impurity atoms that are added to semiconductors in controlled amounts for the express purpose of increasing either the electron or the hole concentration.

Intrinsic semiconductor—undoped semiconductor; extremely pure semiconductor sample containing an insignificant amount of impurity atoms; a semiconductor whose properties are native to the material.

Extrinsic semiconductor—doped semiconductor; a semiconductor whose properties are controlled by added impurity atoms.

Donor—impurity atom that increases the electron concentration; *n*-type dopant.

Acceptor—impurity atom that increases the hole concentration; *p*-type dopant.

n-type material—a donor-doped material; a semiconductor containing more electrons than holes.

p-type material—an acceptor-doped material; a semiconductor containing more holes than electrons.

Majority carrier—the most abundant carrier in a given semiconductor sample; electrons in an *n*-type material, holes in a *p*-type material.

Minority carrier—the least abundant carrier in a given semiconductor sample; holes in an *n*-type material, electrons in a *p*-type material.

2.4 STATE AND CARRIER DISTRIBUTIONS

Up to this point in the modeling process we have concentrated on carrier properties and information of a conceptual, qualitative, or, at most, semiquantitative nature. Practically speaking, there is often a need for more detailed information. For example, most semiconductors are doped, and the precise numerical value of the carrier concentrations inside doped semiconductors is of routine interest. Another property of interest is the distribution of carriers as a function of energy in the respective energy bands. In this section we begin the process of developing a more detailed description of the carrier populations. The development will eventually lead to relationships for the carrier distributions and concentrations within semiconductors under equilibrium conditions.

2.4.1 Density of States

When the energy band model was first introduced in Section 2.2 we indicated that the total number of allowed states in each band was four times the number of atoms in the crystal. Not mentioned at the time was how the allowed states were distributed in energy; i.e., how many states were to be found at any given energy in the conduction and valence bands. We are now interested in this energy distribution of states, or *density of states,* as it is more commonly known, because the state distribution is an essential component in determining carrier distributions and concentrations.

To determine the desired density of states, it is necessary to perform an analysis based on quantum-mechanical considerations. Herein we will merely summarize the results of the analysis; namely, for energies not too far removed from the band edges, one finds

$$g_c(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}, \qquad E \geq E_c \tag{2.6a}$$

$$g_v(E) = \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}, \qquad E \leq E_v \tag{2.6b}$$

where $g_c(E)$ and $g_v(E)$ are the density of states at an energy E in the conduction and valence bands, respectively.

What exactly should be known and remembered about the cited density of states? For one, it is important to grasp the general density of states concept. The density of states can be likened to the description of the seating in a football stadium, with the number of seats in the stadium a given distance from the playing field corresponding to the number of states a specified energy interval from E_c or E_v. Second, the general form of the relationships should be noted. As illustrated in Fig. 2.14, $g_c(E)$ is zero at E_c and increases as the square root of energy when one proceeds upward into the conduction band. Similarly, $g_v(E)$ is precisely zero at E_v and increases with the square root of energy as one proceeds downward from E_v into the valence band. Also note that differences between $g_c(E)$ and $g_v(E)$ stem from differences in the carrier effective masses. If m_n^* were equal to m_p^*, the seating (states) on both sides of the football field (the band gap) would be mirror images of each other. Finally, considering closely spaced energies E and $E + dE$ in the respective bands, one can state

$g_c(E)dE$ represents the number of conduction band states/cm^3 lying in the energy range between E and $E + dE$ (if $E \geq E_c$),

$g_v(E)dE$ represents the number of valence band states/cm^3 lying in the energy range between E and $E + dE$ (if $E \leq E_v$).

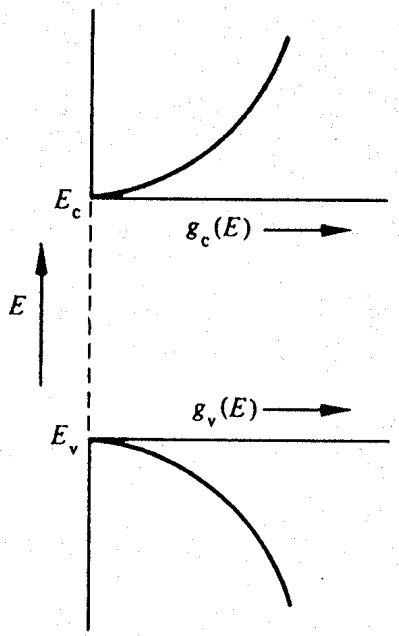


Figure 2.14 General energy dependence of $g_c(E)$ and $g_v(E)$ near the band edges. $g_c(E)$ and $g_v(E)$ are the density of states in the conduction and valence bands, respectively.

It therefore follows that $g_c(E)$ and $g_v(E)$ themselves are numbers/unit volume-unit energy, or typically, numbers/cm^3-eV.

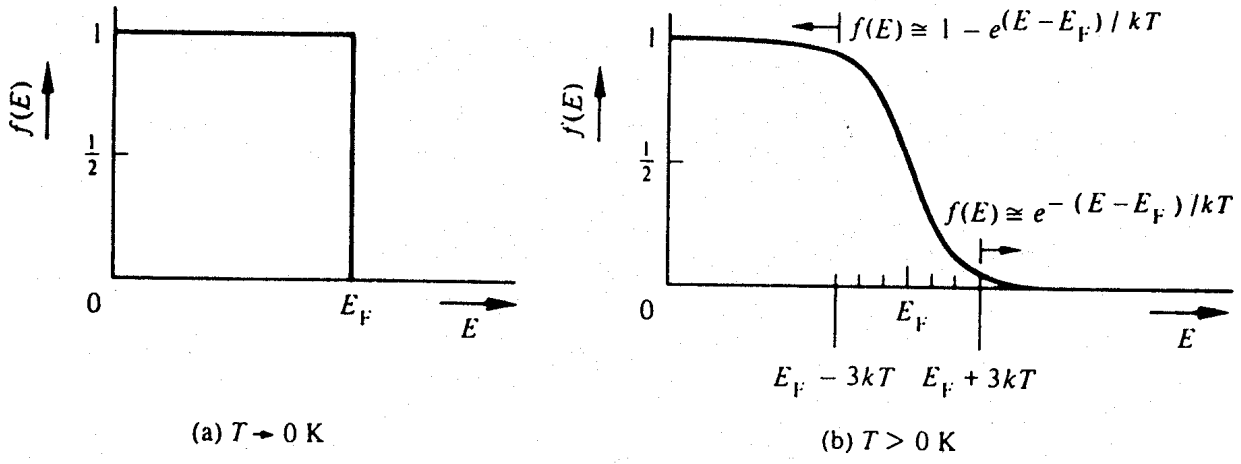
2.4.2 The Fermi Function

Whereas the density of states tells one how many states exist at a given energy E, the Fermi function $f(E)$ specifies how many of the existing states at the energy E will be filled with an electron, or equivalently,

$f(E)$ specifies, under equilibrium conditions, the probability that an available state at an energy E will be occupied by an electron.

Mathematically speaking, the Fermi function is simply a probability distribution function. In mathematical symbols,

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \tag{2.7}$$



(a) $T \to 0$ K (b) $T > 0$ K

Figure 2.15 Energy dependence of the Fermi function. (a) $T \to 0$ K; (b) generalized $T > 0$ K plot with the energy coordinate expressed in kT units.

where

E_F = Fermi energy or Fermi level

k = Boltzmann constant ($k = 8.617 \times 10^{-5}$ eV/K)

T = temperature in Kelvin (K)

Seeking insight into the nature of the newly introduced function, let us begin by investigating the Fermi function's energy dependence. Consider first temperatures where $T \to 0$ K. As $T \to 0$ K, $(E - E_F)/kT \to -\infty$ for all energies $E < E_F$ and $(E - E_F)/kT \to +\infty$ for all energies $E > E_F$. Hence $f(E < E_F) \to 1/[1 + \exp(-\infty)] = 1$ and $f(E > E_F) \to 1/[1 + \exp(+\infty)] = 0$. This result is plotted in Fig. 2.15(a) and is simply interpreted as meaning that all states at energies below E_F will be filled and all states at energies above E_F will be empty for temperatures $T \to 0$ K. In other words, there is a sharp cutoff in the filling of allowed energy states at the Fermi energy E_F when the system temperature approaches absolute zero.

Let us next consider temperatures $T > 0$ K. Examining the Fermi function, we make the following pertinent observations.

(i) If $E = E_F$, $f(E_F) = 1/2$.

(ii) If $E \geq E_F + 3kT$, $\exp[(E - E_F)/kT] \gg 1$ and $f(E) \simeq \exp[-(E - E_F)/kT]$. Consequently, above $E_F + 3kT$ the Fermi function or filled-state probability decays exponentially to zero with increasing energy. Moreover, most states at energies $3kT$ or more above E_F will be empty.

(iii) If $E \leq E_F - 3kT$, $\exp[(E - E_F)/kT] \ll 1$ and $f(E) \simeq 1 - \exp[(E - E_F)/kT]$. Below $E_F - 3kT$, therefore, $[1 - f(E)]$, the probability that a given state will be *empty*, decays exponentially to zero with decreasing energy. Most states at energies $3kT$ or more below E_F will be filled.

(iv) At room temperature ($T = 300$ K), $kT = 0.0259$ eV and $3kT = 0.0777$ eV $\ll E_G$(Si). Compared to the Si band gap, the $3kT$ energy interval that appears prominently in the $T > 0$ K formalism is typically quite small.

The properties just cited are reflected and summarized in the $T > 0$ K Fermi function plot displayed in Fig. 2.15(b).

Before concluding the discussion here, it is perhaps worthwhile to reemphasize that the Fermi function applies only under equilibrium conditions. The Fermi function, however, is universal in the sense that it applies with equal validity to all materials—insulators, semiconductors, and metals. Although introduced in relationship to semiconductors, the Fermi function is not dependent in any way on the special nature of semiconductors but is simply a statistical function associated with electrons in general. Finally, the relative positioning of the Fermi energy E_F compared to E_c (or E_v), an item of obvious concern, is treated in subsequent subsections.

Exercise 2.2

P: The probability that a state is filled at the conduction band edge (E_c) is precisely equal to the probability that a state is *empty* at the valence band edge (E_v). Where is the Fermi level located?

S: The Fermi function, $f(E)$, specifies the probability of electrons occupying states at a given energy E. The probability that a state is empty (not filled) at a given energy E is equal to $1 - f(E)$. Here we are told

$$f(E_c) = 1 - f(E_v)$$

Since

$$f(E_c) = \frac{1}{1 + e^{(E_c - E_F)/kT}}$$

and

$$1 - f(E_v) = 1 - \frac{1}{1 + e^{(E_v - E_F)/kT}} = \frac{1}{1 + e^{(E_F - E_v)/kT}}$$

we conclude

$$\frac{E_c - E_F}{kT} = \frac{E_F - E_v}{kT}$$

or

$$E_F = \frac{E_c + E_v}{2}$$

The Fermi level is positioned at midgap.

(C) Exercise 2.3

In this exercise we wish to examine in more detail how the Fermi function varies with temperature.

P: Successively setting $T = 100$, 200, 300, and then 400 K, compute and plot $f(E)$ versus $\Delta E = E - E_F$ for $-0.2\text{ eV} \leq \Delta E \leq 0.2\text{ eV}$. All $f(E)$ versus ΔE curves should be superimposed on a single set of coordinates.

S: MATLAB program script . . .

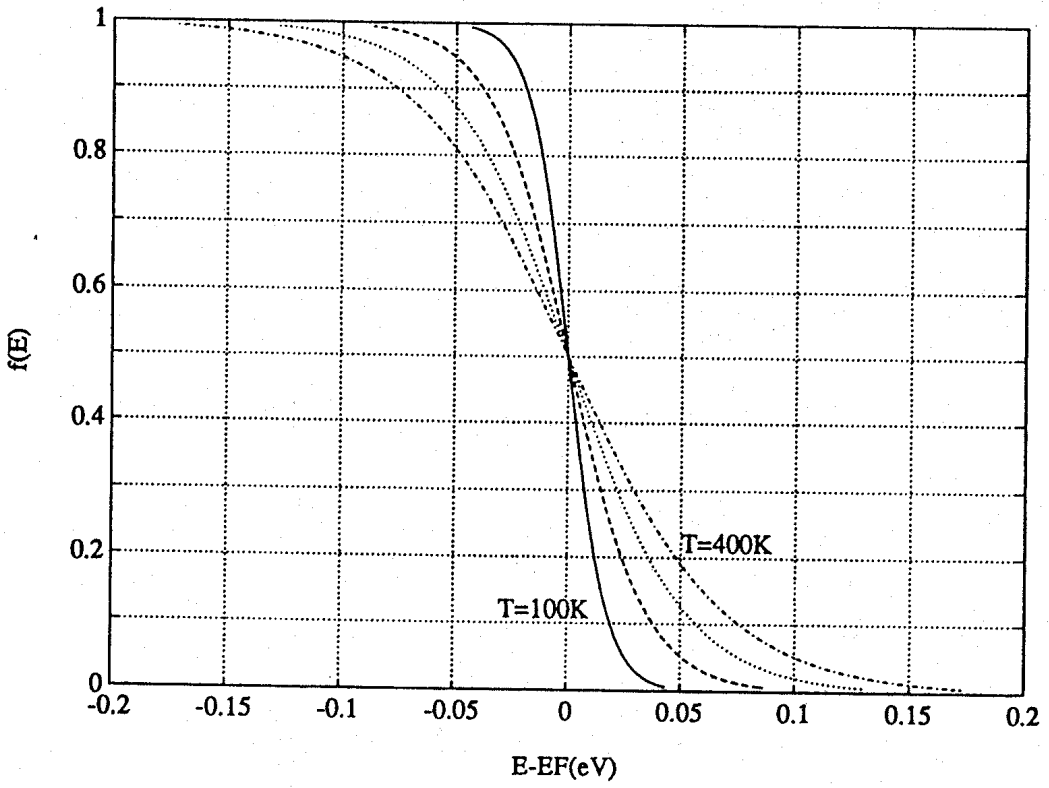
```
%Fermi Function Calculation, f(ΔE,T)

% Constant
k=8.617e-5;

%Computation proper
for ii=1:4;
   T=100*ii;
   kT=k*T;
   dE(ii,1)=-5*kT;
   for jj=1:101
        f(ii,jj)=1/(1+exp(dE(ii,jj)/kT));
        dE(ii,jj+1)=dE(ii,jj)+0.1*kT;
   end
end
dE=dE(:,1:jj);    %This step strips the extra dE value
```

```
%Plotting result
close
plot(dE',f'); grid;   %Note the transpose (') to form data columns
xlabel('E-EF(eV)');   ylabel('f(E)');
text(.05,.2,'T=400K');   text(-.03,.1,'T=100K');
```

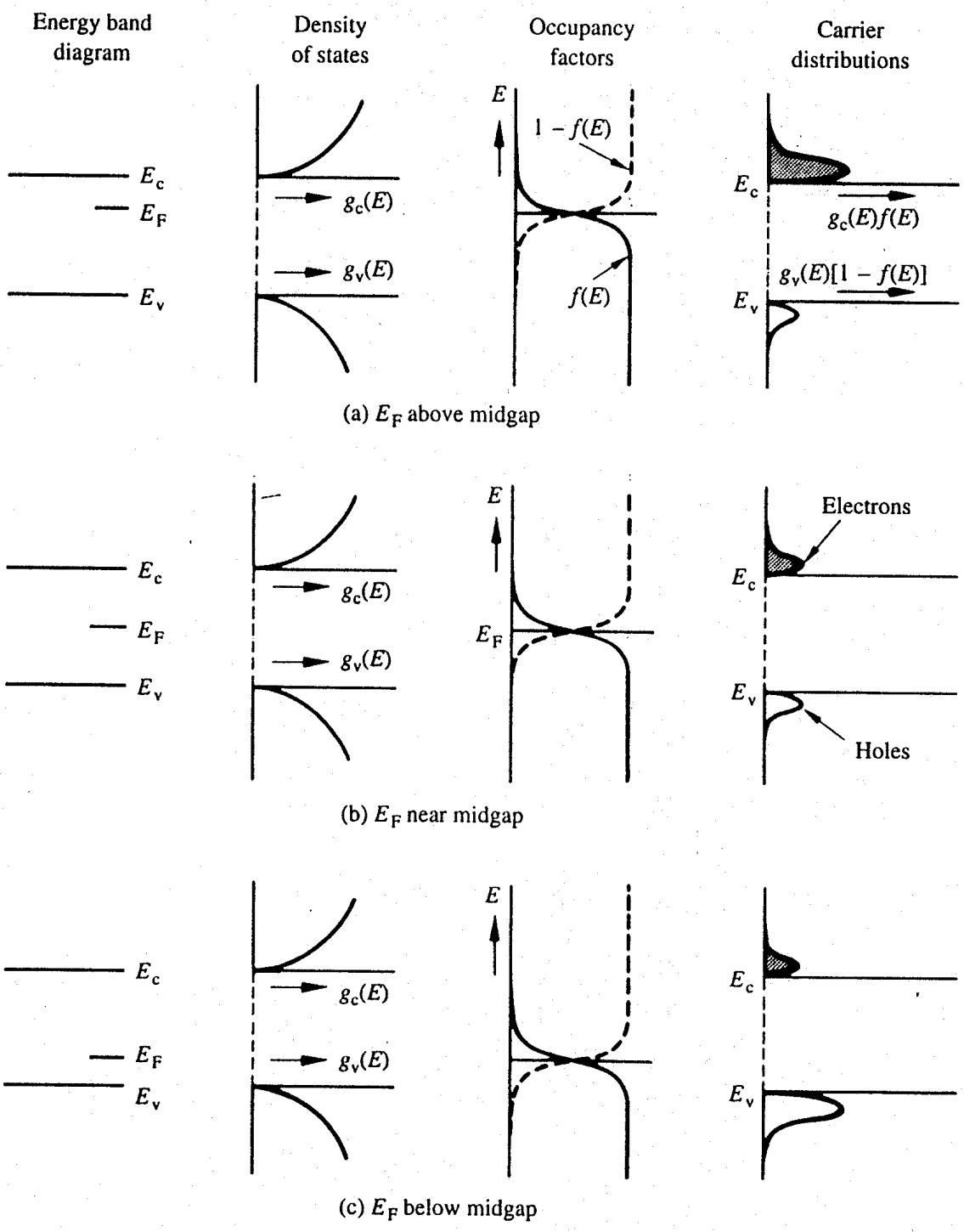
Program output . . .



2.4.3 Equilibrium Distribution of Carriers

Having established the distribution of available band states and the probability of filling those states under equilibrium conditions, we can now easily deduce the distribution of carriers in the respective energy bands. To be specific, the desired distribution is obtained by simply multiplying the appropriate density of states by the appropriate occupancy factor—$g_c(E)f(E)$ yields the distribution of electrons in the conduction band and $g_v(E)$ [1 − $f(E)$] yields the distribution of holes (unfilled states) in the valence band. Sample carrier distributions for three different assumed positions of the Fermi energy (along with associated energy band diagram, Fermi function, and density of states plots) are pictured in Fig. 2.16.

Examining Fig. 2.16, we note in general that all carrier distributions are zero at the



(a) E_F above midgap

(b) E_F near midgap

(c) E_F below midgap

Figure 2.16 Carrier distributions (not drawn to scale) in the respective bands when the Fermi level is positioned (a) above midgap, (b) near midgap, and (c) below midgap. Also shown in each case are coordinated sketches of the energy band diagram, density of states, and the occupancy factors (the Fermi function and one minus the Fermi function).

band edges, reach a peak value very close to E_c or E_v, and then decay very rapidly toward zero as one moves upward into the conduction band or downward into the valence band. In other words, most of the carriers are grouped energetically in the near vicinity of the band edges. Another general point to note is the effect of the Fermi level positioning on the relative magnitude of the carrier distributions. When E_F is positioned in the upper half of the band gap (or higher), the electron distribution greatly outweighs the hole distribution. Although both the filled-state occupancy factor $[f(E)]$ and the empty-state occupancy factor $[1 - f(E)]$ fall off exponentially as one proceeds from the band edges deeper into the conduction and valence bands, respectively, $[1 - f(E)]$ is clearly much smaller than $f(E)$ at corresponding energies if E_F lies in the upper half of the band gap. Lowering E_F effectively slides the occupancy plots downward, giving rise to a nearly equal number of carriers when E_F is at the middle of the gap. Likewise, a predominance of holes results when E_F lies below the middle of the gap. The argument here assumes, of course, that $g_c(E)$ and $g_v(E)$ are of the same order of magnitude at corresponding energies (as is the case for Si and Ge). Also, referring back to the previous subsection, the statements concerning the occupancy factors falling off exponentially in the respective bands are valid provided $E_c - 3kT \geq E_F \geq E_v + 3kT$.

The information just presented concerning the carrier distributions and the relative magnitudes of the carrier numbers finds widespread usage. The information, however, is often conveyed in an abbreviated or shorthand fashion. Figure 2.17, for example, shows a common way of representing the carrier energy distributions. The greatest number of circles or dots are drawn close to E_c and E_v, reflecting the peak in the carrier concentrations near the band edges. The smaller number of dots as one progresses upward into the conduction band crudely models the rapid falloff in the electron density with increasing energy. An extensively utilized means of conveying the relative magnitude of the carrier numbers is displayed in Fig. 2.18. To represent an intrinsic material, a dashed line is drawn in approximately the middle of the band gap and labeled E_i. The near-midgap positioning of E_i, the intrinsic Fermi level, is of course consistent with the previously cited fact that the electron and hole numbers are about equal when E_F is near the center of the band gap. Similarly, a solid line labeled E_F appearing above midgap tells one at a glance that the semiconductor in question is n-type; a solid line labeled E_F appearing below midgap signifies that the semiconductor is p-type. Note finally that the dashed E_i line also typically appears on the

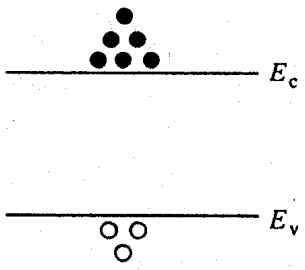


Figure 2.17 Schematic representation of carrier energy distributions.

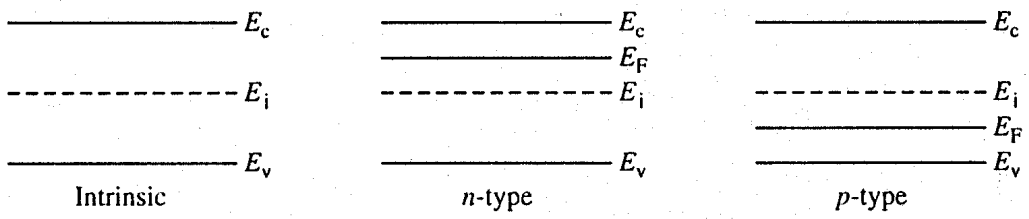


Figure 2.18 "At a glance" representation of intrinsic (left), *n*-type (middle), and *p*-type (right) semiconductor materials using the energy band diagram.

energy band diagrams characterizing extrinsic semiconductors. The E_i line in such cases represents the expected positioning of the Fermi level *if* the material were intrinsic, and it serves as a reference energy level dividing the upper and lower halves of the band gap.

2.5 EQUILIBRIUM CARRIER CONCENTRATIONS

We have arrived at an important point in the carrier modeling process. For the most part, this section simply embodies the culmination of our modeling efforts, with working relationships for the equilibrium carrier concentrations being established to complement the qualitative carrier information presented in previous sections. Unfortunately, the emphasis on the development of mathematical relationships makes the final assault on the carrier modeling summit a little more arduous (and perhaps just a bit boring). Hopefully, the reader can stay focused. A comment is also in order concerning the presentation herein of alternative forms for the carrier relationships. The alternative forms can be likened to the different kinds of wrenches used, for example, in home and automobile repairs. The open-end wrench, the box wrench, and the ratchet wrench all serve the same general purpose. In some applications one can use any of the wrenches. In other applications, however, a special situation restricts the type of wrench employed or favors the use of one wrench over another. The same is true of the alternative carrier relationships. Finally, boxes are drawn around expressions that find widespread usage. A single-walled box signifies a moderately important result; a double-walled box, a very important result.

2.5.1 Formulas for *n* and *p*

Since $g_c(E)\,dE$ represents the number of conduction band states/cm^3 lying in the E to $E + dE$ energy range, and $f(E)$ specifies the probability that an available state at an energy E will be occupied by an electron, it then follows that $g_c(E)f(E)dE$ gives the number of conduction band electrons/cm^3 lying in the E to $E + dE$ energy range, and $g_c(E)f(E)dE$ integrated over all conduction band energies must yield the total number of electrons in the conduction band. In other words, integration over the equilibrium distribution of electrons in the conduction band yields the equilibrium electron concentration. A similar statement can be made relative to the hole concentration. We therefore conclude

$$n = \int_{E_c}^{E_{top}} g_c(E) f(E) dE \tag{2.8a}$$

$$p = \int_{E_{bottom}}^{E_v} g_v(E)[1 - f(E)] dE \tag{2.8b}$$

Seeking explicit expressions for the carrier concentrations, let us focus our efforts on the n-integral. (The analogous p-integral manipulations are left to the reader as an exercise.) Substituting the Eq. (2.6a) expression for $g_c(E)$ and the Eq. (2.7) expression for $f(E)$ into Eq. (2.8a), one obtains

$$n = \frac{m_n^* \sqrt{2m_n^*}}{\pi^2 \hbar^3} \int_{E_c}^{E_{top}} \frac{\sqrt{E - E_c}\, dE}{1 + e^{(E - E_F)/kT}} \tag{2.9}$$

Now letting

$$\eta = \frac{(E - E_c)}{kT} \tag{2.10a}$$

$$\eta_c = \frac{(E_F - E_c)}{kT} \tag{2.10b}$$

$$E_{top} \to \infty \tag{2.10c}$$

yields

$$n = \frac{m_n^* \sqrt{2m_n^*}\, (kT)^{3/2}}{\pi^2 \hbar^3} \int_0^{\infty} \frac{\eta^{1/2} d\eta}{1 + e^{\eta - \eta_c}} \tag{2.11}$$

The (2.10c) simplification on the upper integration limit makes use of the fact that the integrand in question falls off rapidly with increasing energy and is essentially zero for energies only a few kT above E_c. Hence, extending the upper limit to ∞ has a totally negligible effect on the value of the integral.

Even with the cited simplification, the Eq. (2.11) integral cannot be expressed in a closed form containing simple functions. The integral itself is in fact a tabulated function that can be found in a variety of mathematical references. Identifying

$$F_{1/2}(\eta_c) \equiv \int_0^{\infty} \frac{\eta^{1/2} d\eta}{1 + e^{\eta - \eta_c}}, \quad \text{the Fermi-Dirac integral of order 1/2} \tag{2.12}$$

and also defining

$$N_C = 2\left[\frac{m_n^* kT}{2\pi\hbar^2}\right]^{3/2}, \quad \text{the ''effective'' density of conduction band states} \tag{2.13a}$$

$$N_V = 2\left[\frac{m_p^* kT}{2\pi\hbar^2}\right]^{3/2}, \quad \text{the ''effective'' density of valence band states} \tag{2.13b}$$

one obtains

$$\boxed{n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c)} \tag{2.14a}$$

and, by analogy,

$$\boxed{p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v)} \tag{2.14b}$$

where $\eta_v \equiv (E_v - E_F)/kT$.

The Eq. (2.14) relationships are a very general result, valid for any conceivable positioning of the Fermi level. The constants N_C and N_V are readily calculated; at 300 K, $N_{C,V} = (2.510 \times 10^{19}/\text{cm}^3)(m_{n,p}^*/m_0^*)^{3/2}$. The value of the Fermi integral can be obtained from available tables, from plots, or by direct computation. The general-form relationships, nonetheless, are admittedly cumbersome and inconvenient to use in routine analyses. Fortunately, simplified closed-form expressions do exist that can be employed in the vast majority of practical problems. To be specific, if E_F is restricted to values $E_F \leq E_c - 3kT$, then $1/[1 + \exp(\eta - \eta_c)] \simeq \exp[-(\eta - \eta_c)]$ for all $E \geq E_c (\eta \geq 0)$, and

$$F_{1/2}(\eta_c) = \frac{\sqrt{\pi}}{2} e^{(E_F - E_c)/kT} \tag{2.15a}$$

Likewise, if $E_F \geq E_v + 3kT$, then

$$F_{1/2}(\eta_v) = \frac{\sqrt{\pi}}{2} e^{(E_v - E_F)/kT} \tag{2.15b}$$

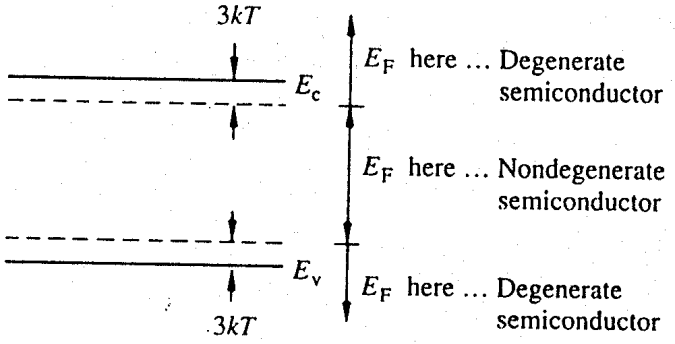


Figure 2.19 Definition of degenerate/nondegenerate semiconductors.

It therefore follows that, if $E_v + 3kT \leq E_F \leq E_c - 3kT$,

$$n = N_C e^{(E_F - E_c)/kT} \tag{2.16a}$$

$$p = N_V e^{(E_v - E_F)/kT} \tag{2.16b}$$

The mathematical simplification leading to Eqs. (2.16) is equivalent to approximating the occupancy factors, $f(E)$ and $1 - f(E)$, by simple exponential functions—an approximation earlier shown to be valid provided E_F was somewhere in the band gap no closer than $3kT$ to either band edge. Whenever E_F is confined, as noted, to $E_v + 3kT \leq E_F \leq E_c - 3kT$, instead of continually repeating the E_F restriction, the semiconductor is simply said to be *nondegenerate*. Whenever E_F lies in the band gap closer than $3kT$ to either band edge or actually penetrates one of the bands, the semiconductor is said to be *degenerate*. These very important terms are also defined pictorially in Fig. 2.19.

2.5.2 Alternative Expressions for *n* and *p*

Although in closed form, the Eq. (2.16) relationships are not in the simplest form possible, and, more often than not, it is the simpler alternative form of these relationships that one encounters in device analyses. The alternative-form relationships can be obtained by recalling that E_i, the Fermi level for an intrinsic semiconductor, lies close to midgap, and hence Eqs. (2.16) most assuredly apply to an intrinsic semiconductor. If this be the case, then specializing Eqs. (2.16) to an intrinsic semiconductor, i.e., setting $n = p = n_i$ and $E_i = E_F$, one obtains

$$n_i = N_C e^{(E_i - E_c)/kT} \tag{2.17a}$$

and

$$n_i = N_V e^{(E_v - E_i)/kT} \tag{2.17b}$$

Solving Eqs. (2.17) for N_C and N_V yields

$$N_C = n_i e^{(E_c - E_i)/kT} \tag{2.18a}$$

$$N_V = n_i e^{(E_i - E_v)/kT} \tag{2.18b}$$

Finally, eliminating N_C and N_V in the original Eq. (2.16) relationships using Eqs. (2.18) gives

$$\boxed{n = n_i e^{(E_F - E_i)/kT}} \tag{2.19a}$$

$$\boxed{p = n_i e^{(E_i - E_F)/kT}} \tag{2.19b}$$

Like Eqs. (2.16), the Eq. (2.19) expressions are valid for any semiconductor in equilibrium whose doping is such as to give rise to a nondegenerate positioning of the Fermi level. Whereas two constants and three energy levels appear in the original relationships, however, only one constant and two energy levels appear in the alternative relationships. Because of their symmetrical nature, the alternative expressions are also easier to remember, requiring merely an interchange of E_F and E_i in going from the n-formula to the p-formula.

2.5.3 n_i and the np Product

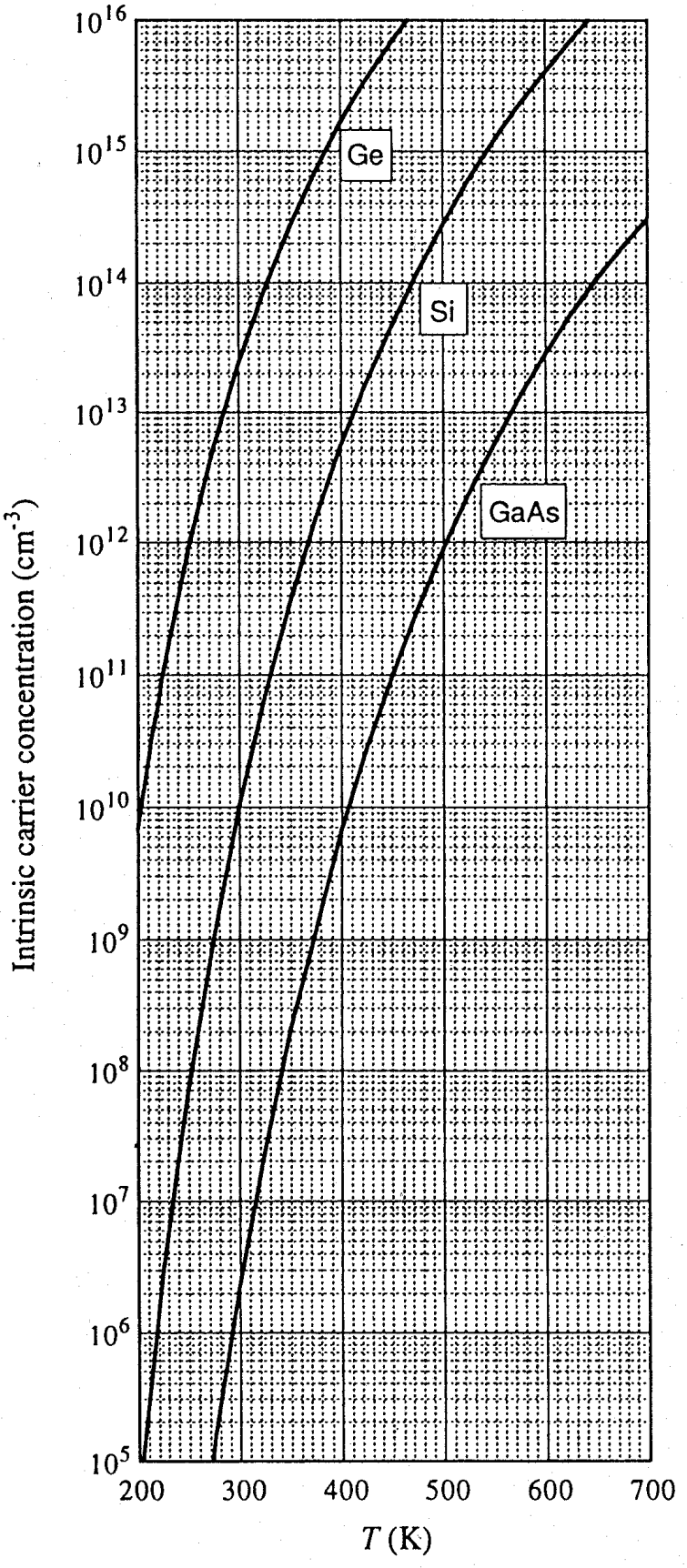
As can be inferred from its appearance in Eqs. (2.19), the intrinsic carrier concentration can figure prominently in the quantitative calculation of the carrier concentrations. Continuing to establish pertinent carrier concentration relationships, we next interject considerations specifically involving this important material parameter.

First, if the corresponding sides of Eqs. (2.17a) and (2.17b) are multiplied together, one obtains

$$n_i^2 = N_C N_V e^{-(E_c - E_v)/kT} = N_C N_V e^{-E_G/kT} \tag{2.20}$$

or

$$\boxed{n_i = \sqrt{N_C N_V}\, e^{-E_G/2kT}} \tag{2.21}$$



Si

T(°C)	n_i(cm^{-3})
0	8.86×10^8
5	1.44×10^9
10	2.30×10^9
15	3.62×10^9
20	5.62×10^9
25	8.60×10^9
30	1.30×10^{10}
35	1.93×10^{10}
40	2.85×10^{10}
45	4.15×10^{10}
50	5.97×10^{10}
300 K	1.00×10^{10}

GaAs

T(°C)	n_i(cm^{-3})
0	1.02×10^5
5	1.89×10^5
10	3.45×10^5
15	6.15×10^5
20	1.08×10^6
25	1.85×10^6
30	3.13×10^6
35	5.20×10^6
40	8.51×10^6
45	1.37×10^7
50	2.18×10^7
300 K	2.25×10^6

Figure 2.20 Intrinsic carrier concentrations in Ge, Si, and GaAs as a function of temperature.

Equation (2.21) expresses n_i as a function of known quantities and can be used to compute n_i at a specified temperature or as a function of temperature. Numerical values for the intrinsic carrier concentration in Si and Ge at room temperature were cited previously; the best available plots of n_i as functions of temperature in Si, Ge, and GaAs are displayed in Fig. 2.20.

A second very important n_i-based relationship follows directly from Eqs. (2.19). Multiplying the corresponding sides of Eqs. (2.19a) and (2.19b) together yields

$$\boxed{np = n_i^2} \tag{2.22}$$

Although appearing trivial, the *np* product relationship (Eq. 2.22) often proves to be extremely useful in practical computations. If one of the carrier concentrations is known, the remaining carrier concentration is readily determined using Eq. (2.22)—provided of course that the semiconductor is in equilibrium and nondegenerate.

(C) Exercise 2.4

Substituting the Eq. (2.13) definitions of N_C and N_V into Eq. (2.21), introducing normalization factors, and replacing known constants by their numerical values, one obtains

$$n_i = (2.510 \times 10^{19})\left(\frac{m_n^*}{m_0}\frac{m_p^*}{m_0}\right)^{3/4}\left(\frac{T}{300}\right)^{3/2} e^{-E_G/2kT}$$

E_G and the effective masses exhibit a weak but non-negligible temperature dependence. The E_G versus T variation can be modeled to four-place accuracy by the fit relationship noted in Problem 2.1(a). As deduced from the analysis by Barber [Solid-State Electronics, **10,** 1039 (1967)], the temperature dependence of the effective masses over the range 200 K $\leq T \leq$ 700 K can be approximated by

$$\frac{m_n^*}{m_0} = 1.028 + (6.11 \times 10^{-4})T - (3.09 \times 10^{-7})T^2$$

$$\frac{m_p^*}{m_0} = 0.610 + (7.83 \times 10^{-4})T - (4.46 \times 10^{-7})T^2$$

P: (a) Confirm that the n_i versus T curve for Si graphed in Fig. 2.20 is generated employing the relationships just cited, *provided* E_G in the n_i expression is replaced by $E_G - E_{ex}$, where $E_{ex} = 0.0074$ eV. (In the previously cited article, Barber suggested using an exciton correction factor of $E_{ex} = 0.007$ eV. The slightly larger value employed in the Fig. 2.20 computation was specifically chosen to give $n_i = 10^{10}$/cm^3 at 300 K.)

(b) The accepted value of n_i in Si at 300 K has been revised recently to be in agreement with the experimental n_i versus T data acquired by Sproul and Green [Journal of Applied Physics, **70,** 846 (July 1991)]. The authors concluded $n_i = 1.00 \pm 0.03 \times 10^{10}/\text{cm}^3$ in Si at 300 K. Over the probed temperature range of 275 K $\leq T \leq$ 375 K, their experimental data could be fitted by the relationship

$$n_i = (9.15 \times 10^{19})\left(\frac{T}{300}\right)^2 e^{-0.5928/kT}$$

Compare the n_i computed from this experimental-fit relationship to the part (a) output over the temperature range of mutual validity.

S: (a) MATLAB program script . . .

```
%ni vs. T calculation for Si (200K - 700K) used in Fig. 2.20

%Initialization
format short e

%Constants and T-range
k=8.617e-5;
A=2.510e19;
Eex=0.0074;   %Value was adjusted to match S&G ni(300K) value
T=200:25:700;

%Band Gap vs. T
EG0=1.17;
a=4.730e-4;
b=636;
EG=EG0-a.*(T.^2)./(T+b);

%Effective mass ratio (mnr=mn*/m0, mpr=mp*/m0)
mnr=1.028 + (6.11e-4).*T - (3.09e-7).*T.^2;
mpr=0.610 + (7.83e-4).*T - (4.46e-7).*T.^2;

%Computation of ni
ni=A.*((T./300).^(1.5)).*((mnr.*mpr).^(0.75)).*exp(-(EG-Eex)./(2 .*k.*T));

%Display output on screen
j=length(T);
fprintf('\n \n T      ni\n');   %There are ten spaces between T and ni.
for ii=1:j,
fprintf('%-10.f%-10.3e\n',T(ii),ni(ii));
end
```

(b) MATLAB program script . . .

```
%Experimental fit of Sproul-Green ni data (275K - 375K)

%ni calculation
T=275:25:375;
k=8.617e-5;
ni=(9.15e19).*(T./300).^2 .*exp(-0.5928./(k*T));

%Display result on screen
j=length(T);
fprintf('\n\n T       ni\n');   %There are ten spaces between T and ni.
for ii=1:j,
fprintf('%-10.f%-10.3e\n',T(ii),ni(ii));
end
```

Output from the two programs is reproduced below. Note the excellent agreement of the computational results (to within 2%) over the temperature range of mutual validity.

T (K)	(*a*) n_i (cm^{-3})	(*b*) n_i (cm^{-3})
200	5.246×10^{4}	—
275	1.059×10^{9}	1.051×10^{9}
300	1.000×10^{10}	1.006×10^{10}
325	6.798×10^{10}	6.887×10^{10}
350	3.565×10^{11}	3.623×10^{11}
375	1.518×10^{12}	1.542×10^{12}
400	5.449×10^{12}	—
500	2.716×10^{14}	—
600	3.988×10^{15}	—
700	2.865×10^{16}	—

2.5.4 Charge Neutrality Relationship

The relationships established to this point are devoid of an explicit dependence on the dopant concentrations introduced into a semiconductor. It is the charge neutrality relationship that provides the general tie between the carrier and dopant concentrations.

To establish the charge neutrality relationship, let us consider a *uniformly doped* semiconductor, a semiconductor where the number of dopant atoms/cm^3 is the same everywhere. Systematically examining little sections of the semiconductor far from any surfaces, and assuming equilibrium conditions prevail, one must invariably find that each and every section is charge-neutral, i.e., contains no net charge. If this were not the case, electric fields

would exist inside the semiconductor. The electric fields in turn would give rise to carrier motion and associated currents—a situation totally inconsistent with the assumed equilibrium conditions. There are, however, charged entities inside all semiconductors. Electrons, holes, ionized donors (donor atoms that have become positively charged by donating an electron to the conduction band) and negatively-charged ionized acceptors can all exist simultaneously inside any given semiconductor. For the uniformly doped material to be everywhere charge-neutral clearly requires

$$\frac{\text{charge}}{\text{cm}^3} = qp - qn + qN_D^+ - qN_A^- = 0 \tag{2.23}$$

or

$$\boxed{p - n + N_D^+ - N_A^- = 0} \tag{2.24}$$

where, by definition,

N_D^+ = number of ionized (positively charged) donors/cm^3,

N_A^- = number of ionized (negatively charged) acceptors/cm^3.

As previously discussed, there is sufficient thermal energy available in a semiconductor at room temperature to ionize almost all of the shallow-level donor and acceptor sites. Defining

N_D = total number of donors/cm^3,

N_A = total number of acceptors/cm^3,

and setting

$$N_D^+ = N_D$$

$$N_A^- = N_A$$

one then obtains

$$\boxed{p - n + N_D - N_A = 0} \quad \text{assumes total ionization of dopant atoms} \tag{2.25}$$

Equation (2.25) is the standard form of the charge neutrality relationship.

2.5.5 Carrier Concentration Calculations

We are finally in a position to calculate the carrier concentrations in a uniformly doped semiconductor under equilibrium conditions. In the computations to be presented we specifically make the assumptions of NONDEGENERACY (allowing us to use the *np* product relationship) and TOTAL IONIZATION of the dopant atoms. Note that n_i, which appears in the *np* product expression, has been calculated and plotted and must be considered a known quantity. Likewise, N_A and N_D, which appear in the charge neutrality relationship, are typically controlled and determined experimentally and should also be considered known quantities. The only other symbols used in the two equations are n and p. Thus, under the cited assumptions of nondegeneracy and total ionization of dopant atoms, we have two equations and two unknowns from which n and p can be deduced.

Starting with the *np* product expression, one can write

$$p = \frac{n_i^2}{n} \tag{2.26}$$

Eliminating p in Eq. (2.25) using Eq. (2.26) gives

$$\frac{n_i^2}{n} - n + N_D - N_A = 0 \tag{2.27}$$

or

$$n^2 - n(N_D - N_A) - n_i^2 = 0 \tag{2.28}$$

Solving the quadratic equation for n then yields

$$\boxed{n = \frac{N_D - N_A}{2} + \left[\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2\right]^{1/2}} \tag{2.29a}$$

and

$$\boxed{p = \frac{n_i^2}{n} = \frac{N_A - N_D}{2} + \left[\left(\frac{N_A - N_D}{2}\right)^2 + n_i^2\right]^{1/2}} \tag{2.29b}$$

Only the plus roots have been retained in Eqs. (2.29) because physically the carrier concentrations must be greater than or equal to zero.

Equations (2.29) are general-case solutions. In the vast majority of practical computations it is possible to simplify these equations prior to substituting in numerical values for N_D, N_A, and n_i. Special cases of specific interest are considered next.

(1) *Intrinsic Semiconductor* ($N_A = 0$, $N_D = 0$). With $N_A = 0$ and $N_D = 0$, Eqs. (2.29) simplify to $\boxed{n = n_i}$ and $\boxed{p = n_i}$. $n = p = n_i$ is of course the expected result for the equilibrium carrier concentrations in an intrinsic semiconductor.

(2) *Doped Semiconductor where either* $N_D - N_A \simeq N_D \gg n_i$ *or* $N_A - N_D \simeq N_A \gg n_i$. This is the special case of greatest practical interest. The unintentional doping levels in Si are such that the controlled addition of dopants routinely yields $N_D \gg N_A$ or $N_A \gg N_D$. Moreover, the intrinsic carrier concentration in Si at room temperature is about $10^{10}/\text{cm}^3$, while the dominant doping concentration (N_A or N_D) is seldom less than $10^{14}/\text{cm}^3$. Thus the special case considered here is the usual case encountered in practice. If $N_D - N_A \simeq N_D \gg n_i$, the square root in Eq. (2.29a) reduces to $N_D/2$ and

$$\boxed{\begin{aligned} n &\simeq N_D \\ p &\simeq n_i^2/N_D \end{aligned}} \qquad \begin{aligned} &N_D \gg N_A, \quad N_D \gg n_i \\ &\text{(nondegenerate, total ionization)} \end{aligned} \qquad \begin{aligned} &\text{(2.30a)} \\ &\text{(2.30b)} \end{aligned}$$

Similarly

$$\boxed{\begin{aligned} p &\simeq N_A \\ n &\simeq n_i^2/N_A \end{aligned}} \qquad \begin{aligned} &N_A \gg N_D, \quad N_A \gg n_i \\ &\text{(nondegenerate, total ionization)} \end{aligned} \qquad \begin{aligned} &\text{(2.31a)} \\ &\text{(2.31b)} \end{aligned}$$

As a numerical example, suppose a Si sample maintained at room temperature is uniformly doped with $N_D = 10^{15}/\text{cm}^3$ donors. Using Eqs. (2.30), one rapidly concludes $n \simeq 10^{15}/\text{cm}^3$ and $p \simeq 10^5/\text{cm}^3$.

(3) *Doped Semiconductor where* $n_i \gg |N_D - N_A|$. Systematically increasing the ambient temperature causes a monotonic rise in the intrinsic carrier concentration (see Fig. 2.20). At sufficiently elevated temperatures, n_i will eventually equal and then exceed the net doping concentration. If $n_i \gg |N_D - N_A|$, the square roots in Eqs. (2.29) reduce to n_i and $\boxed{n \simeq p \simeq n_i}$. In other words, *all semiconductors become intrinsic at sufficiently high temperatures where* $n_i \gg |N_D - N_A|$.

(4) *Compensated Semiconductor.* As is evident from Eqs. (2.29), donors and acceptors tend to negate each other. Indeed, it is possible to produce intrinsic-like material by making $N_D - N_A = 0$. In some materials, such as GaAs, N_A may be comparable to N_D in the as-grown crystal. When N_A and N_D are comparable and nonzero, the material is said to be *compensated.* If the semiconductor is compensated, both N_A and N_D must be retained in all carrier concentration expressions.

In summary, Eqs. (2.29) can always be used to compute the carrier concentrations if the semiconductor is nondegenerate and the dopant atoms are totally ionized. In the vast majority of practical situations, however, it is possible to simplify these equations prior to performing numerical computations. Equations (2.29) must be used to compute the carrier concentrations only in those rare instances when $|N_D - N_A| \sim n_i$. The simplified relationships of greatest practical utility are Eqs. (2.30) and (2.31).

Exercise 2.5

P: A Si sample is doped with 10^{14} boron atoms per cm^3.

(a) What are the carrier concentrations in the Si sample at 300 K?

(b) What are the carrier concentrations at 470 K?

S: (a) Boron in Si is an acceptor (see Table 2.2). Thus $N_A = 10^{14}/cm^3$. At 300 K, $n_i = 1.00 \times 10^{10}/cm^3$ and the given N_A is clearly much greater than n_i. Moreover, since the N_D doping was omitted from the problem statement, we infer $N_D \ll N_A$. With $N_A \gg n_i$ and $N_A \gg N_D$, Eqs. (2.31) may be used to calculate the carrier concentrations: $p = N_A = 10^{14}/cm^3$; $n = n_i^2/N_A = 10^6/cm^3$.

(b) As deduced from Fig. 2.20, $n_i \simeq 10^{14}/cm^3$ at 470 K. Because n_i is comparable to N_A, Eqs. (2.29) must be used to calculate at least one of the carrier concentrations. (Once one of the carrier concentrations is known, the second is more readily computed using the np product expression.) Performing the indicated calculations gives $p = N_A/2 + [(N_A/2)^2 + n_i^2]^{1/2} = 1.62 \times 10^{14}/cm^3$; $n = n_i^2/p = 6.18 \times 10^{13}/cm^3$.

2.5.6 Determination of E_F

Knowledge concerning the exact position of the Fermi level on the energy band diagram is often of interest. For example, when discussing the intrinsic Fermi level, we indicated that E_i was located somewhere near the middle of the band gap. It would be useful to know the *precise* positioning of E_i in the band gap. Moreover, we have developed computational formulas for n and p appropriate for nondegenerate semiconductors. Whether a doped semiconductor is nondegenerate or degenerate depends, of course, on the value or positioning of E_F.

Before running though the mechanics of finding the Fermi level in selected cases of interest, it is useful to make a general observation; namely, Eqs. (2.19) or (2.16) [or even more generally, Eqs. (2.14)] provide a one-to-one correspondence between the Fermi energy and the carrier concentrations. Thus, having computed any one of the three variables—n, p, or E_F—one can always determine the remaining two variables under equilibrium conditions.

(1) *Exact positioning of* E_i. In an intrinsic material

$$n = p \tag{2.32}$$

Substituting for n and p in Eq. (2.32) using Eqs. (2.16), and setting $E_F = E_i$, yields

$$N_C e^{(E_i - E_c)/kT} = N_V e^{(E_v - E_i)/kT} \tag{2.33}$$

Solving for E_i, one obtains

$$E_i = \frac{E_c + E_v}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) \tag{2.34}$$

But

$$\frac{N_V}{N_C} = \left(\frac{m_p^*}{m_n^*}\right)^{3/2} \tag{2.35}$$

Consequently,

$$\boxed{E_i = \frac{E_c + E_v}{2} + \frac{3}{4} kT \ln\left(\frac{m_p^*}{m_n^*}\right)} \tag{2.36}$$

According to Eq. (2.36), E_i lies precisely at midgap only if $m_p^* = m_n^*$ or if $T = 0$ K. For the more practical case of silicon at room temperature, Table 2.1 gives $m_p^*/m_n^* = 0.69$, $(3/4)kT \ln(m_p^*/m_n^*) = -0.0073$ eV, and E_i therefore lies 0.0073 eV below midgap. Although potentially significant in certain problems, this small deviation from midgap is typically neglected in drawing energy band diagrams, etc.

(2) *Doped semiconductors (nondegenerate, dopants totally ionized).* The general positioning of the Fermi level in donor- and acceptor-doped semiconductors assumed to be nondegenerate, in equilibrium, and maintained at temperatures where the dopants are fully ionized, is readily deduced from Eqs. (2.19). Specifically, solving Eqs. (2.19) for $E_F - E_i$, one obtains

$$E_F - E_i = kT \ln(n/n_i) = -kT \ln(p/n_i) \tag{2.37}$$

Depending on the simplifications inherent in a particular problem, the appropriate carrier concentration solution [Eqs. (2.29), (2.30), (2.31), etc.] is then substituted into Eq. (2.37) to determine the positioning of E_F. For example, per Eqs. (2.30a) and (2.31a), $n \simeq N_D$ in typical donor-doped semiconductors and $p \simeq N_A$ in typical accep-

tor-doped semiconductors maintained at or near room temperature. Substituting into Eq. (2.37), we therefore conclude

$$E_F - E_i = kT \ln(N_D/n_i) \quad \ldots N_D \gg N_A, \quad N_D \gg n_i \tag{2.38a}$$

$$E_i - E_F = kT \ln(N_A/n_i) \quad \ldots N_A \gg N_D, \quad N_A \gg n_i \tag{2.38b}$$

From Eqs. (2.38) it is obvious that the Fermi level moves systematically upward in energy from E_i with increasing donor doping and systematically downward in energy from E_i with increasing acceptor doping. The exact Fermi level positioning in Si at room temperature, nicely reinforcing the foregoing statement, is displayed in Fig. 2.21. Also note that for a given semiconductor material and ambient temperature, there exist maximum nondegenerate donor and acceptor concentrations, doping concentrations above which the material becomes degenerate. In Si at room temperature the maximum nondegenerate doping concentrations are $N_D \simeq 1.6 \times 10^{18}/\text{cm}^3$ and $N_A \simeq 9.1 \times 10^{17}/\text{cm}^3$. The large Si doping values required for degeneracy, we should interject, have led to the common usage of "highly doped" (or n^+-material/p^+-material) and "degenerate" as interchangeable terms.

Finally, the question may arise: What procedure should be followed in computing E_F when one is not sure whether a material is nondegenerate or degenerate? Unless a material is known to be degenerate, always assume nondegeneracy and compute E_F employing the appropriate nondegenerate relationship. If E_F derived from the nondegenerate formula lies in the degenerate zone, one must then, of course, recompute E_F using the more complex formalism valid for degenerate materials.

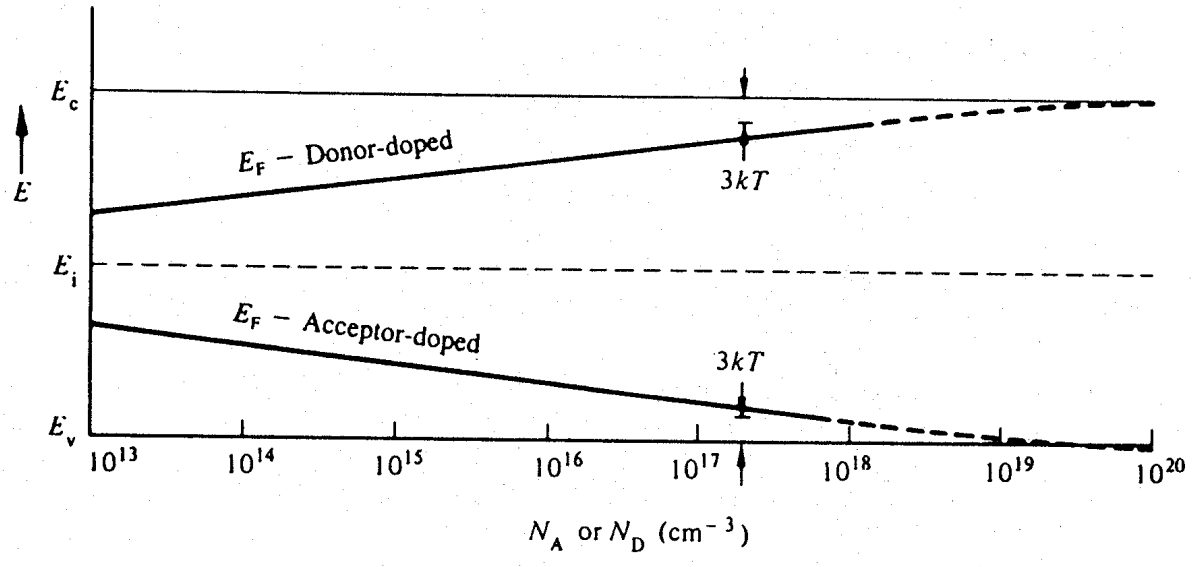


Figure 2.21 Fermi level positioning in Si at 300 K as a function of the doping concentration. The solid E_F lines were established using Eq. (2.38a) for donor-doped material and Eq. (2.38b) for acceptor-doped material ($kT = 0.0259$ eV, and $n_i = 10^{10}/\text{cm}^3$).

Exercise 2.6

P: For each of the conditions specified in Exercise 2.5, determine the position of E_i, compute $E_F - E_i$, and draw a carefully dimensioned energy band diagram for the Si sample. Note that $E_G(\text{Si}) = 1.08$ eV and $m_p^*/m_n^* = 0.71$ at 470 K.

S: (a) In part (a) of Exercise 2.5, the $N_A = 10^{14}/\text{cm}^3$ Si sample is mainted at 300 K. Using Eq. (2.36), we conclude E_i is located 0.0073 eV below midgap. (The positioning of E_i in Si at 300 K is also noted in the text following Eq. 2.36). Next applying Eq. (2.38b), we find

$$\begin{aligned} E_i - E_F &= kT \ln(N_A/n_i) \\ &= 0.0259 \ln(10^{14}/10^{10}) \\ &= 0.239 \text{ eV} \end{aligned}$$

The energy band diagram constructed from the deduced positioning of E_i and E_F is shown in Fig. E2.6(a).

(b) In part (b) of Exercise 2.5 the Si sample is heated to 470 K. With $m_p^*/m_n^* = 0.71$ and $kT = 0.0405$ eV at 470 K, $(3/4)kT \ln(m_p^*/m_n^*) = -0.0104$ eV and E_i is deduced to be located 0.0104 eV below midgap. Because N_A is comparable to n_i at 470 K, Eq. (2.37) must be used to compute the positioning of E_F. Specifically, with $n_i = 10^{14}/\text{cm}^3$ and $p = 1.62 \times 10^{14}/\text{cm}^3$,

$$\begin{aligned} E_i - E_F &= kT \ln(p/n_i) \\ &= 0.0405 \ln(1.62 \times 10^{14}/10^{14}) \\ &= 0.0195 \text{ eV} \end{aligned}$$

Here E_F is only slightly removed from E_i as pictured in Fig. E2.6(b).

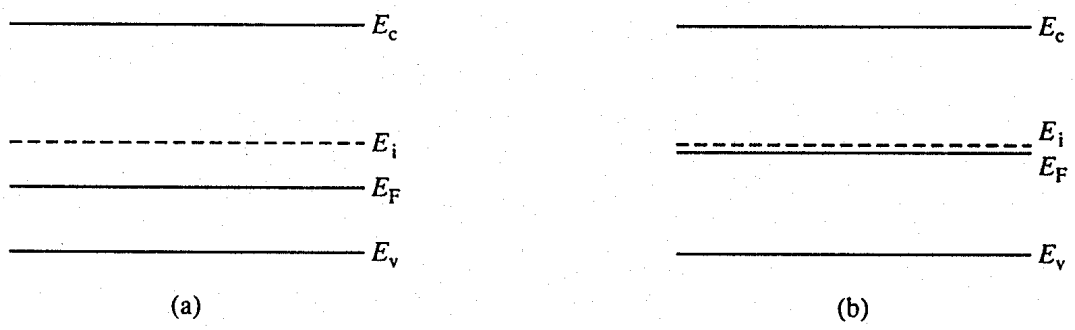


Figure E2.6

2.5.7 Carrier Concentration Temperature Dependence

A number of isolated facts about the carrier concentration temperature dependence have already been presented at various points in this chapter. The discussion in Section 2.3 concerned with dopant action, for example, described the increased ionization of dopant sites and the associated increase in the majority carrier concentration when the temperature of a semiconductor is raised from near $T = 0$ K toward room temperature. More recent subsections have included a plot of the intrinsic carrier concentration versus temperature (Fig. 2.20) and a calculation indicating that all semiconductors become intrinsic ($n \rightarrow n_i$, $p \rightarrow n_i$) at sufficiently high temperatures. In this subsection, which concludes the carrier concentration discussion, temperature-related facts are combined and embellished to provide a broader, more complete description of how the carrier concentrations vary with temperature.

Figure 2.22(a), a typical majority-carrier concentration-versus-temperature plot constructed assuming a phosphorus-doped $N_D = 10^{15}/cm^3$ Si sample, nicely illustrates the general features of the concentration-versus-temperature dependence. Examining Fig. 2.22(a), we find that n is fixed at approximately N_D over a broad temperature range extending roughly from 150 K to 450 K for the given Si sample. This $n \simeq N_D$ or "extrinsic temperature region" constitutes the normal operating range for most solid-state devices. Below 100 K or so, in the "freeze-out temperature region," n drops significantly below N_D and approaches zero as $T \rightarrow 0$ K. In the "intrinsic temperature region" at the opposite end of the temperature scale, n rises above N_D, asymptotically approaching n_i with increasing T.

To qualitatively explain the just-described concentration-versus-temperature dependence, it is important to recall that the equilibrium number of carriers within a material is affected by two separate mechanisms. Electrons donated to the conduction band from donor atoms and valence band electrons excited across the band gap into the conduction band (broken Si–Si bonds) both contribute to the majority-carrier electron concentration in a donor-doped material. At temperatures $T \rightarrow 0$ K the thermal energy available in the system is insufficient to release the weakly bound fifth electron on donor sites and totally insufficient to excite electrons across the band gap. Hence $n = 0$ at $T = 0$ K, as visualized on the left-hand side of Fig. 2.22(b). Slightly increasing the material temperature above $T = 0$ K "defrosts" or frees some of the electrons weakly bound to donor sites. Band-to-band excitation, however, remains extremely unlikely, and therefore the number of observed electrons in the freeze-out temperature region equals the number of ionized donors—$n = N_D^+$. Continuing to increase the system temperature eventually frees almost all of the weakly bound electrons on donor sites, n approaches N_D, and one enters the extrinsic temperature region. In progressing through the extrinsic temperature region, more and more electrons are excited across the band gap, but the number of electrons supplied in this fashion stays comfortably below N_D. Ultimately, of course, electrons excited across the band gap equal, then exceed, and, as pictured on the right-hand side of Fig. 2.22(b), finally swamp the fixed number of electrons derived from the donors.

As a practical note, it should be pointed out that the wider the band gap, the greater the energy required to excite electrons from the valence band into the conduction band, and

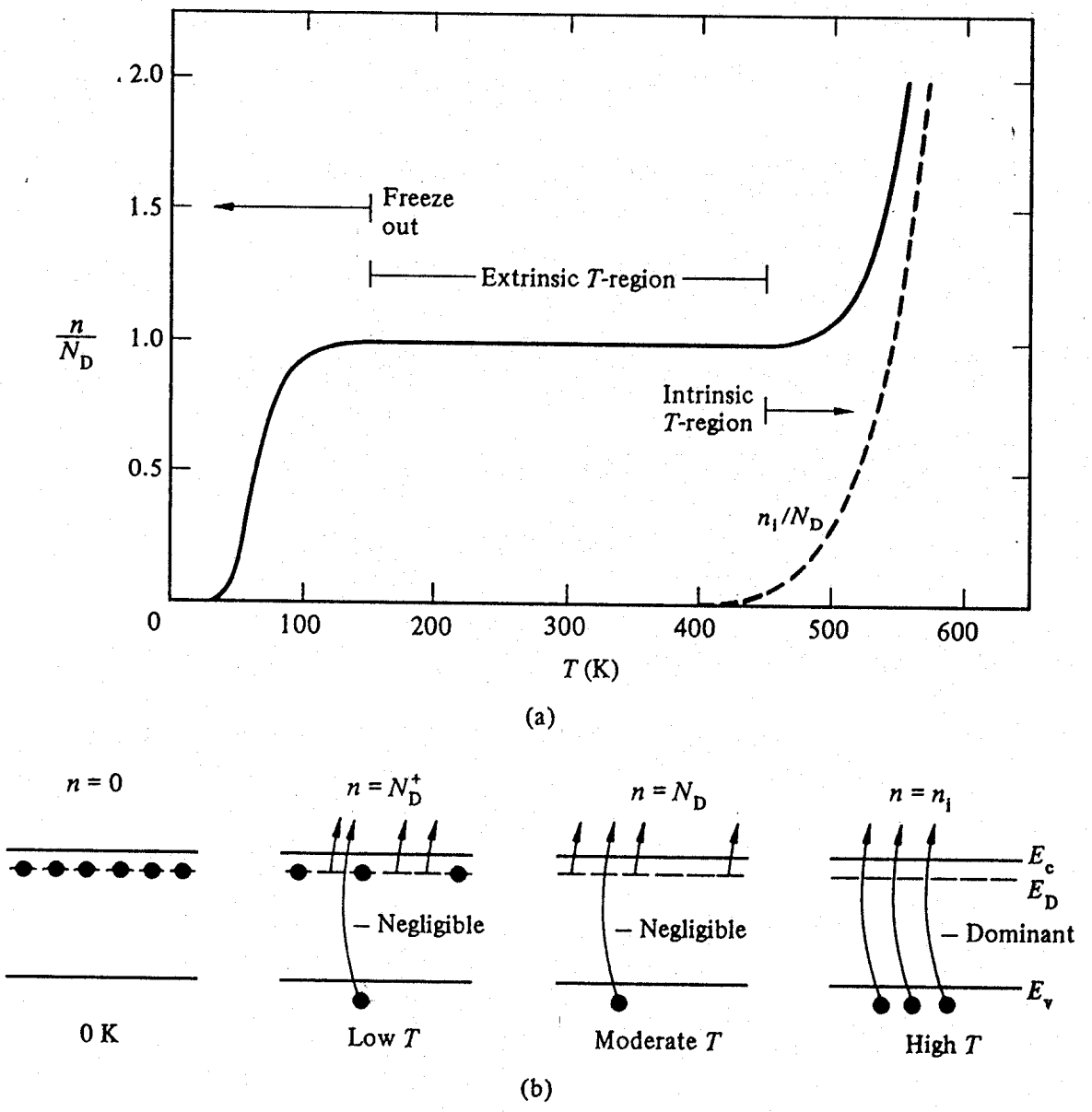


Figure 2.22 (a) Typical temperature dependence of the majority-carrier concentration in a doped semiconductor. The plot was constructed assuming a phosphorus-doped $N_D = 10^{15}/\text{cm}^3$ Si sample. n_i/N_D versus T (dashed line) has been included for comparison purposes. (b) Qualitative explanation of the concentration-versus-temperature dependence displayed in part (a).

the higher the temperature at the onset of the intrinsic temperature region. Since the temperature at the onset of the intrinsic temperature region corresponds to the upper end of the normal operating range for most solid-state devices, GaAs devices can inherently operate at a higher maximum temperature than similarly doped Si devices, which in turn can operate at a higher maximum temperature than similarly doped Ge devices. If we assume, for example, the critical doping concentration is $N_D = 10^{15}/\text{cm}^3$, and the onset of the intrinsic temperature region is approximated as the temperature where $n_i = N_D$, then from Fig. 2.20

the maximum operating temperatures are deduced to be 385 K, 540 K, and > 700 K for Ge, Si, and GaAs, respectively. Indeed, GaAs and SiC ($E_G > 2$ eV) devices continue to be under development for use in high temperature environments.

2.6 SUMMARY AND CONCLUDING COMMENTS

Under the general heading of carrier modeling we have described, examined, and characterized the carriers within a semiconductor under "rest" or equilibrium conditions. The many important topics addressed in this chapter included the introduction of two "visualization" models: the bonding model and the energy band model. The extremely useful energy band model is actually more than just a model—it is a sophisticated sign language providing a concise means of communicating on a nonverbal level. Relative to the carriers themselves, the reader by now has been successfully prodded into thinking of electrons and holes as classical ball-like "particles," where the charge on an electron is $-q$, the charge on a hole is $+q$, and the effective masses of the particles are m_n^* and m_p^*, respectively. The reader should also know that the carrier numbers in an intrinsic material are equal and relatively small; the carrier concentrations, however, can be selectively increased by adding special impurity atoms or dopants to the semiconductor.

In addressing the problem of determining the carrier concentrations in doped semiconductors, we developed or derived a number of useful mathematical relationships. The density of states functions (Eqs. 2.6), the Fermi function (Eq. 2.7), the symmetrical nondegenerate relationships for n and p (Eqs. 2.19), the np product (Eq. 2.22), the charge neutrality relationship (Eq. 2.25), and the simplified n and p expressions appropriate for a typical semiconductor maintained at room temperature (Eqs. 2.30 and 2.31) deserve special mention. The cited equations and a few others are collected in Table 2.4. With regard to the use of these relationships, the reader should be cautioned against "no-think plug and chug." Because semiconductor problems are replete with exceptions, special cases, and nonideal situations, it is imperative that the formula user be aware of derivational assumptions and the validity limits of any and all expressions used in an analysis or computation. In addition to the quantitative carrier relationships, the reader should also have a qualitative "feel" for the carrier distributions in the respective energy bands, the temperature dependence of the intrinsic concentration, and the typical temperature dependence of the majority carrier concentration in a doped semiconductor.

Finally, some attention should be given to the many technical terms and the key parametric values presented in this chapter. The terms extrinsic semiconductor, donor, acceptor, nondegenerate semiconductor, Fermi level, and so on, will be encountered again and again in the discussion of semiconductor devices. Likewise, a knowledge of typical numerical values for key parameters, such as $E_G = 1.12$ eV and $n_i = 10^{10}/\text{cm}^3$ in Si at 300 K, will be invaluable in subsequent work when performing both "back-of-the-envelope" and computer-assisted computations. Key parametric values also serve as yardsticks for gauging whether newly encountered quantities are relatively small or relatively large.

Table 2.4 Carrier Modeling Equation Summary.

Density of States and Fermi Function

$$g_c(E) = \frac{m_n^* \sqrt{2m_n^*(E - E_c)}}{\pi^2 \hbar^3}, \quad E \geq E_c$$

$$g_v(E) = \frac{m_p^* \sqrt{2m_p^*(E_v - E)}}{\pi^2 \hbar^3}, \quad E \leq E_v$$

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

Carrier Concentration Relationships

$$n = N_C \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_c) \qquad N_C = 2\left[\frac{m_n^* kT}{2\pi\hbar^2}\right]^{3/2}$$

$$p = N_V \frac{2}{\sqrt{\pi}} F_{1/2}(\eta_v) \qquad N_V = 2\left[\frac{m_p^* kT}{2\pi\hbar^2}\right]^{3/2}$$

$$n = N_C e^{(E_F - E_c)/kT}$$

$$p = N_V e^{(E_v - E_F)/kT}$$

$$n = n_i e^{(E_F - E_i)/kT}$$

$$p = n_i e^{(E_i - E_F)/kT}$$

n_i, np-Product, and Charge Neutrality

$$n_i = \sqrt{N_C N_V} e^{-E_G/2kT} \qquad np = n_i^2 \qquad p - n + N_D - N_A = 0$$

n, p, and Fermi Level Computational Relationships

$$n = \frac{N_D - N_A}{2} + \left[\left(\frac{N_D - N_A}{2}\right)^2 + n_i^2\right]^{1/2} \qquad E_i = \frac{E_c + E_v}{2} + \frac{3}{4} kT \ln\left(\frac{m_p^*}{m_n^*}\right)$$

$n \simeq N_D$ $\qquad N_D \gg N_A, N_D \gg n_i$ $\qquad E_F - E_i = kT\ln(n/n_i) = -kT\ln(p/n_i)$

$p \simeq n_i^2/N_D$

$p \simeq N_A$ $\qquad N_A \gg N_D, N_A \gg n_i$ $\qquad E_F - E_i = kT\ln(N_D/n_i) \quad N_D \gg N_A, N_D \gg n_i$

$n \simeq n_i^2/N_A$ $\qquad E_i - E_F = kT\ln(N_A/n_i) \quad N_A \gg N_D, N_A \gg n_i$

PROBLEMS

CHAPTER 2 PROBLEM INFORMATION TABLE				
Problem	*Complete After*	*Difficulty Level*	*Suggested Point Weighting*	*Short Description*
● 2.1	2.2.4	2	10 (5 each part)	E_G vs. T computation
2.2	2.3.4	1	10 (2 each part)	Bonding model applications
2.3	2.3.4 (a–d) 2.5.7 all	1	24 (2 each part)	E-band model applications
2.4	2.3.4	2	10 (2 each part)	Si-doped GaAs
2.5	2.4.1	2	5	States/cm^3 in ΔE
2.6	2.4.2	2	8 (a-2, b-3, c-3)	Fermi function questions
2.7	2.5.1	2–3	10	Find distribution peak
2.8	"	2	5	Population at $E_c+\Delta E$
● 2.9	"	2	15	Plot distributions in bands
● 2.10	"	3	20 (a-3, b-17..discuss-2)	T variation of distribution
2.11	"	2	10	Derive Eqs. (2.14b), (2.16b)
2.12	"	3	15 (a-3, b-12)	Hypothetical g_c = constant
2.13	"	2	10 (a-7, b-3)	Compute N_C, N_V
2.14	2.5.3	1	8 (4 each part)	n_i comparison
● 2.15	"	1	5	Plot n_i vs. T for Ge
2.16	2.5.5	2	12 (a::d-2, e-4)	Tricky conc. questions
2.17	"	1–2	10 (2 each part)	Compute n and p
2.18	2.5.6	2	15 (3 each part)	Determine E_i, $E_F - E_i$, etc.
● 2.19	"	2	15 (a-12, b-3)	Computer check P2.17/2.18
2.20	"	2	8	Verify nondegenerate limits
● 2.21	"	2	10	Plot $E_F - E_i$ vs. N_A, N_D
2.22	"	2–3	12 (a-3, b-3, c-2, d-4)	GaAs considerations
● 2.23	2.5.7	4	25 (a-8, b-10, c-2, d-3, e-2)	E_F variation with T

● **2.1** E_G versus T Computation

With increasing temperature an expansion of the crystal lattice usually leads to a weakening of the interatomic bonds and an associated decrease in the band gap energy. For many semiconductors the cited variation of the band gap energy with temperature can be modeled by the empirical relationship

$$E_G(T) = E_G(0) - \frac{\alpha T^2}{(T + \beta)}$$

where α and β are constants chosen to obtain the best fit to experimental data and $E_G(0)$ is the limiting value of the band gap at 0 K. As far as Si is concerned, a fit accurate to four places is obtained by employing

$$\begin{aligned} E_G(0) &= 1.170 \text{ eV} \\ \alpha &= 4.730 \times 10^{-4} \text{ eV/K} \\ \beta &= 636 \text{ K} \\ T &\text{ in Kelvin} \end{aligned}$$

(a) Make a plot of E_G versus T for Si spanning the temperature range from $T = 0$ K to $T = 600$ K. Specifically note the value of E_G at 300 K.

(b) For $T > 300$ K, the temperature variation is nearly linear. Noting this fact, some authors have employed

$$E_G(T) = 1.205 - 2.8 \times 10^{-4} T \qquad \ldots T > 300 \text{ K}$$

How does this simplified relationship compare with the more precise relationship over the temperature region of mutual validity?

2.2 Using the *bonding* model for a semiconductor, indicate how one visualizes (a) a missing atom, (b) an electron, (c) a hole, (d) a donor, (e) an acceptor.

2.3 Using the *energy band* model for a semiconductor, indicate how one visualizes (a) an electron, (b) a hole, (c) donor sites, (d) acceptor sites, (e) freeze-out of majority carrier electrons at donor sites as the temperature is lowered toward 0 K, (f) freeze-out of majority carrier holes at acceptor sites as the temperature is lowered toward 0 K, (g) the energy distribution of carriers in the respective bands, (h) an intrinsic semiconductor, (i) an n-type semiconductor, (j) a p-type semiconductor, (k) a nondegenerate semiconductor, (l) a degenerate semiconductor.

2.4 The bonding model for GaAs is pictured in Fig. P2.4.

(a) Draw the bonding model for GaAs depicting the removal of the shaded Ga and As atoms in Fig. P2.4. HINT: Ga and As take their bonding electrons with them when they are removed from the lattice. Also see Fig. 2.4(a) showing the results of removing an atom from the Si lattice.

(b) Redraw the bonding model for GaAs showing the insertion of Si atoms into the missing Ga and As atom sites.

(c) Is the GaAs doped p- or n-type when Si atoms replace Ga atoms? Explain.

(d) Is the GaAs doped p- or n-type when Si atoms replace As atoms? Explain.

(e) Draw the *energy band diagram* for GaAs when the GaAs is doped with Si on (i) Ga sites and (ii) on As sites.

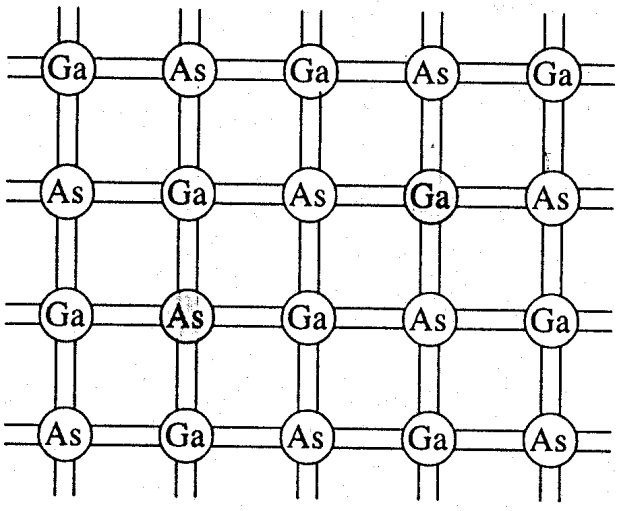


Figure P2.4

2.5 Develop an expression for the total number of available STATES/cm^3 in the conduction band between energies E_c and $E_c + \gamma kT$, where γ is an arbitrary constant.

2.6 (a) Under equilibrium conditions and $T > 0$ K, what is the probability of an electron state being occupied if it is located at the Fermi level?

(b) If E_F is positioned at E_c, determine (numerical answer required) the probability of finding electrons in states at $E_c + kT$.

(c) The probability a state is filled at $E_c + kT$ is equal to the probability a state is empty at $E_c + kT$. Where is the Fermi level located?

2.7 The carrier distributions or numbers of carriers as a function of energy in the conduction and valence bands were noted to peak at an energy very close to the band edges. (See the carrier distribution sketches in Fig. 2.16.) Taking the semiconductor to be nondegenerate, show that the energy at which the carrier distributions peak is $E_c + kT/2$ and $E_v - kT/2$ for the conduction and valence bands, respectively.

2.8 For a nondegenerate semiconductor, the peak in the electron distribution versus energy inside the conduction band noted in Fig. 2.16 occurs at $E_c + kT/2$. Expressed as a fraction of the electron population at the peak energy, what is the electron population in a nondegenerate semiconductor at $E = E_c + 5kT$?

● **2.9** The Fermi level in a Si sample maintained at $T = 300$ K is located at $E_c - E_G/4$. Compute and plot the electron and hole distributions (numbers/cm^3–eV) as a function of energy in the conduction and valence bands, respectively.

● **2.10** Let us investigate how the electron energy distribution in the conduction band varies as a function of temperature.

(a) Assuming the semiconductor to be nondegenerate and employing the Eq. (2.16a) expression for n, confirm that the electron distribution in the conduction band normalized to the total electron concentration is given by

$$\frac{g_c(E)f(E)}{n} = \frac{2\sqrt{E - E_c}}{\sqrt{\pi}\,(kT)^{3/2}}\, e^{-(E-E_c)/kT}$$

(b) Compute and plot the normalized electron distribution in the conduction band versus $E - E_c$ for temperatures $T = 300$ K, 600 K, and 1200 K. Plot the distribution values along the x-axis ($0 \leq g_c(E)f(E)/n \leq 20$ eV^{-1}) and $E - E_c$ ($0 \leq E - E_c \leq 0.4$ eV) along the y-axis on a single set of coordinates. Discuss your results.

2.11 Starting with Eq. (2.8b) and following a procedure analogous to that outlined in the text, present the intermediate steps and arguments leading to Eqs. (2.14b) and (2.16b).

2.12 The density of states in the conduction band of a hypothetical semiconductor is

$$g_c(E) = \text{constant} = N_C/kT \qquad \ldots E \geq E_c$$

(a) Assuming $E_F < E_c - 3kT$, sketch the electron distribution in the conduction band of the hypothetical semiconductor.

(b) Following the procedure outlined in the text, establish relationships for the electron concentration in the hypothetical semiconductor analogous to Eqs. (2.14a) and (2.16a).

2.13 (a) Verify the statement in Subsection 2.5.1 that, at 300 K,

$$N_{C,V} = (2.510 \times 10^{19}/\text{cm}^3)(m^*/m_0)^{3/2}$$

where one sets $m^* = m_n^*$ in computing N_C and $m^* = m_p^*$ in computing N_V. $m_0 = 9.109 \times 10^{-31}$ kg; $h = 6.625 \times 10^{-34}$ joule-sec; and $q = 1.602 \times 10^{-19}$ coul.

(b) Using the effective mass values recorded in Table 2.1, construct a table that lists the numerical values of N_C and N_V for Si, Ge, and GaAs at 300 K.

2.14 (a) Determine the temperature at which the intrinsic carrier concentration in (i) Si and (ii) GaAs are equal to the room temperature (300 K) intrinsic carrier concentration of Ge.

(b) Semiconductor A has a band gap of 1 eV, while semiconductor B has a band gap of 2 eV. What is the ratio of the intrinsic carrier concentrations in the two materials (n_{iA}/n_{iB}) at 300 K. Assume any differences in the carrier effective masses may be neglected.

• **2.15** Confirm that the n_i versus T curve for Ge graphed in Fig. 2.20 can be generated employing the empirical-fit relationship

$$n_i(\text{Ge}) = (1.76 \times 10^{16})T^{3/2}e^{-0.392/kT}$$

2.16 Concentration questions with a twist.

(a) A silicon wafer is uniformly doped p-type with $N_A = 10^{15}/\text{cm}^3$. At $T \simeq 0$ K, what are the equilibrium hole and electron concentrations?

(b) A semiconductor is doped with an impurity concentration N such that $N \gg n_i$ and all the impurities are ionized. Also, $n = N$ and $p = n_i^2/N$. Is the impurity a donor or an acceptor? Explain.

(c) The electron concentration in a piece of Si maintained at 300 K under equilibrium conditions is $10^5/\text{cm}^3$. What is the hole concentration?

(d) For a silicon sample maintained at $T = 300$ K, the Fermi level is located 0.259 eV above the intrinsic Fermi level. What are the hole and electron concentrations?

(e) In a nondegenerate germanium sample maintained under equilibrium conditions near room temperature, it is known that $n_i = 10^{13}/\text{cm}^3$, $n = 2p$, and $N_A = 0$. Determine n and N_D.

2.17 Determine the equilibrium electron and hole concentrations inside a uniformly doped sample of Si under the following conditions:

(a) $T = 300$ K, $N_A \ll N_D$, $N_D = 10^{15}/\text{cm}^3$.

(b) $T = 300$ K, $N_A = 10^{16}/\text{cm}^3$, $N_D \ll N_A$.

(c) $T = 300$ K, $N_A = 9 \times 10^{15}/\text{cm}^3$, $N_D = 10^{16}/\text{cm}^3$.

(d) $T = 450$ K, $N_A = 0$, $N_D = 10^{14}/\text{cm}^3$.

(e) $T = 650$ K, $N_A = 0$, $N_D = 10^{14}/\text{cm}^3$.

2.18 (a to e) For each of the conditions specified in Problem 2.17, determine the position of E_i, compute $E_F - E_i$, and draw a carefully dimensioned energy band diagram for the Si sample. NOTE: $E_G(\text{Si}) = 1.08$ eV at 450 K and 1.015 eV at 650 K.

• **2.19** (a) Assuming a nondegenerate Si sample and total impurity atom ionization, construct a MATLAB program that computes n, p, and $E_F - E_i$ given acceptable input values of T (temperature in Kelvin), N_D (cm^{-3}), and N_A (cm^{-3}). Incorporate the program presented in part (a) of Exercise 2.4 to compute n_i at a specified T. Make use of the MATLAB `input` function to enter the input variables from the command window.

(b) Use your program to check the relevant answers to Problems 2.17 and 2.18.

2.20 According to the text, the maximum nondegenerate donor and acceptor doping concentrations in Si at room temperature are $N_D \simeq 1.6 \times 10^{18}/\text{cm}^3$ and $N_A \simeq 9.1 \times 10^{17}/\text{cm}^3$, respectively. Verify the text statement.

• **2.21** Construct a computer program to produce a plot of $E_F - E_i$ versus N_A or N_D similar to Fig. 2.21. Use the program to verify the accuracy of the cited figure. (You may find it convenient to employ the MATLAB `logspace` function in constructing your program.)

2.22 GaAs considerations.

(a) Make a sketch similar to Fig. 2.14 that is specifically appropriate for GaAs. Be sure to take into account the fact that $m_n^* \ll m_p^*$ in GaAs.

(b) Based on your answer to part (a), would you expect E_i in GaAs to lie above or below midgap? Explain.

(c) Determine the precise position of the intrinsic Fermi level in GaAs at room temperature (300 K).

(d) Determine the maximum nondegenerate donor and acceptor doping concentrations in GaAs at room temperature.

2.23 Given an $N_A = 10^{14}/\text{cm}^3$ doped Si sample:

(a) Present a *qualitative* argument that leads to the approximate positioning of the Fermi level in the material as $T \to 0$ K.

• (b) Construct a MATLAB (computer) program to calculate and plot $E_F - E_i$ in the material as a function of temperature for $200 \text{ K} \leq T \leq 500$ K.

(c) What do you conclude relative to the general behavior of the Fermi level positioning as a function of temperature?

(d) Run your part (b) program to determine what happens if N_A is progressively increased in decade steps from $N_A = 10^{14}/\text{cm}^3$ to $N_A = 10^{18}/\text{cm}^3$. Summarize your observations.

(e) How would your foregoing answers be modified if the Si sample was doped with donors instead of acceptors?