

Chapter 2

Carrier Statistics in Equilibrium

Version 3.5 (July 31, 2012), ©Prentice Hall

Outline

- 2.1 Conduction and valence bands; bandgap; holes
- 2.2 Intrinsic semiconductor
- 2.3 Extrinsic semiconductor
 - 2.3.1 Donors and acceptors
 - 2.3.2 Charge neutrality
 - 2.3.3 Equilibrium carrier concentration in a doped semiconductor
- 2.4 Carrier statistics in equilibrium
 - 2.4.1 Conduction and valence band density of states
 - 2.4.2 Equilibrium electron concentration
 - 2.4.3 Equilibrium hole concentration
 - 2.4.4 np product in equilibrium
 - 2.4.5 Location of Fermi level
- 2.5 Summary
- 2.6 Further reading

Advanced Topics

- AT2.1 Temperature dependence of the bandgap
- AT2.2 Selected properties of the Fermi-Dirac integral
- AT2.3 Approximations for strongly degenerate semiconductor
- AT2.4 Statistics of donor and acceptor ionization
- AT2.5 Carrier freeze-out
- AT2.6 Heavy-doping effects
 - AT2.6.1 The Mott transition
 - AT2.6.2 Bandgap narrowing

Problems

This chapter discusses the statistics of electrons in a semiconductor in thermal equilibrium. Understanding the equilibrium situation is essential to appreciating the behavior of a semiconductor when perturbed. This is in fact a pervasive theme of this book. It is many times rather insightful to view semiconductor device operation as the process required to reestablish equilibrium following an external disturbance.

This chapter starts by discussing the distinct role that electrons in the conduction and valence bands have in semiconductors. The concept of "hole" is introduced. Basically, a hole is a missing electron in the valence band. Following this, "intrinsic" and "extrinsic" semiconductors are defined and the role of a special kind of foreign atoms, "dopants", play in controlling the equilibrium electron and hole concentrations in semiconductors is discussed. We continue with a rather general and fairly rigorous formulation to enable the computation of the equilibrium carrier concentrations in semiconductors in a wide range of situations. This formulation is based on an energy view of the situation and exploits the concept of Fermi level.

The results obtained in this chapter are very important to semiconductor device design and will be extensively used throughout this book.

2.1 Conduction and valence bands; bandgap; holes

We saw in Ch. 1 that a semiconductor is a crystalline solid with a band structure that at 0 K is characterized by several bands completely filled with electrons. The last full band is separated from the next empty band by a relatively small bandgap, the "fundamental" bandgap. The meaning of "relatively small" has to do with the actual temperature of operation of semiconductor devices. The proper scale of this statement is the thermal energy kT . For devices that operate near room temperature, the most widely used semiconductors have bandgaps of the order of 0.7 to 2 eV. Si for example has a bandgap at room temperature of 1.12 eV while GaAs has 1.42 eV. Wider bandgap semiconductors are currently under research for high-temperature applications, such as integrated circuits for jet engine control. A good example is SiC with a bandgap of 3 eV. Narrower bandgap semiconductors are also of interest for applications such as cooled infrared sensors. InSb with a bandgap of 0.17 eV is under investigation for this purpose.

Most of the behavior that we are likely to encounter in semiconductor devices involves an energy picture that is centered around the fundamental bandgap, or simply, "the bandgap". In semiconductor device operation we can safely ignore the complexities of the fully occupied deeper lying bands. In fact, the two key bands around the fundamental bandgap are given special names. The band immediately below is called the *valence band* while the band immediately above is called the *conduction band*. In an energy picture, E_v is used to denote the top edge of the valence band and E_c the location of the bottom edge of the conduction band. The symbol E_g is used to refer to the width of the energy gap that separates these two edges. The magnitude of the bandgap depends on temperature. Advanced Topic AT2.1 discusses this.

If we zoom into the energy band structure of a semiconductor around the fundamental bandgap, the picture at the left of Fig. 2.1 emerges. It shows the top of the valence band, the bandgap and the bottom of the conduction band. In the semiconductor literature it has

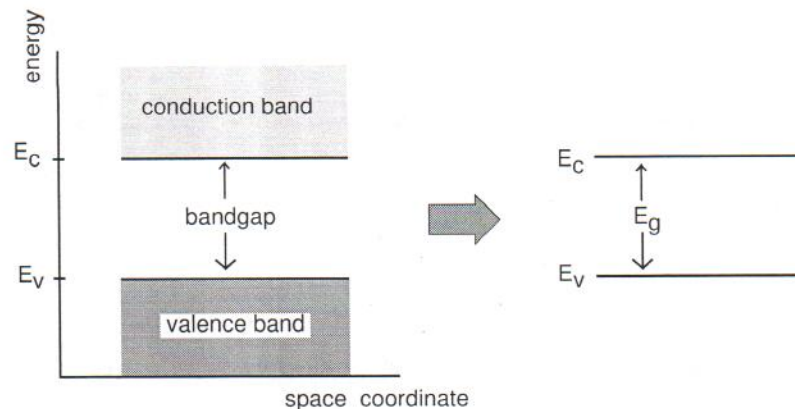


Figure 2.1: Left: band structure of a semiconductor around the fundamental energy gap. Dark shading represents a full band at 0 K. Light shading represents an empty band at 0 K. Right: simplified representation showing only band edges.

become common to extract the essence of this picture by just drawing the edges of the bands as represented on the right of Fig. 2.1. This is the first and simplest of many *energy band diagrams* that will be drawn in this book and that are widely used in the semiconductor device literature. The energy picture is a particularly intuitive and powerful way to represent many processes that take place in a semiconductor. It does not capture everything that is important about a semiconductor. For example, it does not contain enough information to understand light emission and absorption which requires an awareness of momentum. However, this simple energy picture is very useful in understanding transport in devices such as microelectronic transistors. For this reason, it is extensively used in microelectronics engineering.

It is important not to lose sight of the physical meaning of the conduction and valence bands. To refresh our memory we go back to the structural arrangement of atoms in a semiconductor. We discussed in the previous chapter that, in a semiconductor, atoms bond together mostly by sharing valence electrons. On average, an atom ends up with 8 valence electrons around it, a particularly low energy arrangement. This situation can be represented in a simple flat sketch as shown in Fig. 2.2a). In this figure, each box represents an atom and each stick represents a bonding valence electron. Each atom shares its four valence electrons with its four neighbors and therefore each pair of atoms shares two electrons between them. Although Fig. 2.2 shows a flattened array, it represents in reality a three-dimensional network of atoms. This simple flat picture does not change in any substantial way in the case of a compound semiconductor with two or more different kinds of atoms.

Fig. 2.2a) represents a 0 K situation, where all valence electrons are engaged in bonding between neighboring atoms. Fig. 2.2b) shows a picture at a finite temperature where one bonding electron has acquired enough energy from the finite thermal energy of the lattice to freely wander around the crystal. It becomes a “conduction” electron. The number of electrons that break off their ties to atoms is a function of temperature and other parameters. We will learn how to calculate this number later on in this chapter. Before that, it is important to note that in the place that was occupied by the electron that escaped, a *hole* is left. Since the electron is

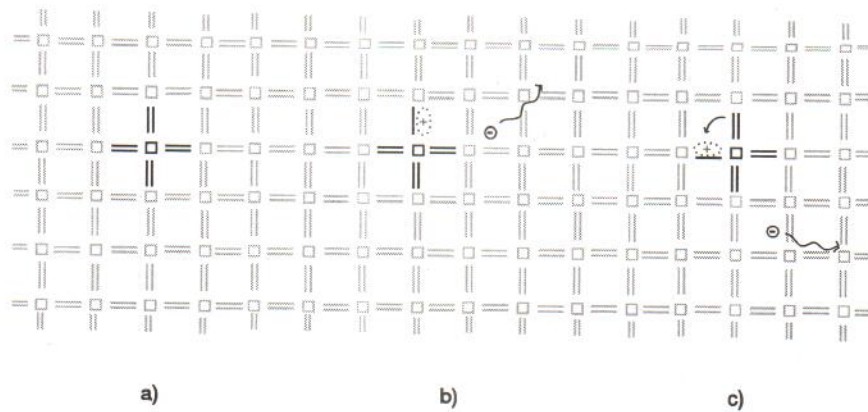


Figure 2.2: Flat model of the semiconductor lattice: a) at 0 K with all valence electrons participating in bonding, b) at finite temperature where one electron has become free leaving a hole behind, and c) after the hole moves to a neighboring location.

negatively charged with a charge of $-q$, this hole has a net positive charge of $+q$. The actual origin of this positive charge is the partially uncompensated charge associated with the protons in the two atoms involved. In spite of that, the hole can also move around. This happens when a neighboring bonding electron jumps in to satisfy the missing bond. This leaves two other atoms partially unbonded somewhere else. Effectively, this makes the hole move from one place to another (Fig. 2.2c). Both the conduction electron and the hole are called *carriers* because as they move, they "carry" their elemental electric charge along with them.

Energy band diagrams represent the situations depicted in Fig. 2.2 in a very physical way, as shown in Fig. 2.3. In order to understand how this is done, let us first think about the 0 K condition where no covalent bonds are broken. At 0 K the valence band is completely full of electrons while the conduction band is empty. This allows us to conclude that the bonding valence electrons are represented in the energy band diagram by levels or states inside the valence band. With a full valence band and a completely empty conduction band, no conduction is possible in an ideal semiconductor at 0 K.

At a finite temperature, the free electron depicted in Figs. 2.2b) and 2.2c) has acquired enough energy to be promoted from the valence band to the conduction band where the abundance of empty states allows it to move around easily. Free electrons therefore occupy states in the conduction band. The hole left behind, however, sits inside the valence band in the energy band picture. The hole is an empty state in the valence band. Whenever there is an empty state, conduction becomes possible. Since there are many electrons in the valence band (the bonding electrons), it will not be too difficult for one of these electrons to jump to the empty position where the hole resides thereby allowing the hole to move around. The hole is therefore also free. Similar to air bubbles in water where it is easier to focus on the simple movement of the bubble rather than on the complex flow of water around it, it is much more convenient to describe the complicated movement of bonding electrons in the valence band in terms of the hole dynamics. In energy band diagrams, only holes are typically represented in the valence band. Bonding electrons can be ignored in most circumstances.

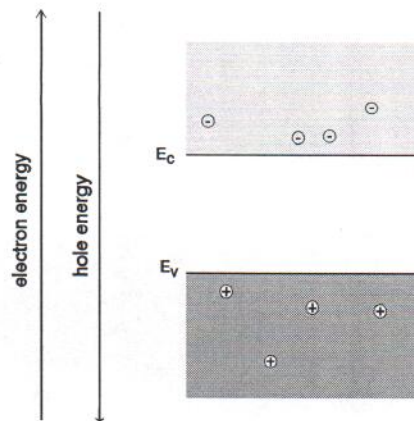


Figure 2.3: Sketch of free electrons and holes in energy band diagram. Note the different signs for electron and hole energy scales.

Before proceeding any further, there are two additional important observations that must be made about energy band diagrams. The first one is that at the edges of the bands, the kinetic energy of the carriers is zero. We can easily understand that with the help of Fig. 2.4a) which depicts a photon striking a semiconductor. If the energy of this particular photon is exactly the bandgap energy of the semiconductor, then the photogenerated electron and hole cannot have any extra kinetic energy. The entire photon energy is consumed to break the bond.¹ In the energy band picture, both carriers are represented at their respective band edges.

The second observation regarding energy band diagrams is that electron energies increase upwards while hole energies increase downwards, as Fig. 2.3 makes explicit. If a conduction electron has higher kinetic energy, it shows up at a high energy inside the conduction band. For holes, this picture is reversed - a high kinetic energy hole is represented deep down inside the valence band. This can be better understood with the help of Fig. 2.4b). Consider two photons of identical energy $h\nu > E_g$ impinging on different bonding electrons and producing electron-hole pairs. In one case, the electron that is promoted to the conduction band carries all the extra energy in excess of that required to create an electron-hole pair as kinetic energy. This electron is represented inside the conduction band at an energy $h\nu - E_g$ over the conduction band edge. Energy conservation leaves no extra kinetic energy for the hole which therefore sits at E_v . In a second case, the contrary situation occurs: the electron is ejected with no extra kinetic energy so it sits at E_c and the hole carries all the extra energy as kinetic energy and it is represented at an energy $h\nu - E_g$ below the band edge. Clearly hole energies increase downwards in the energy band diagram. A high kinetic energy fast-moving hole represents a situation in which bonding electrons are jumping from bonding site to bonding site very quickly.

The flat picture of a semiconductor shown in Fig. 2.2 allows us to visualize a conduction electron and a hole in a simple and intuitive way. It is important, however, to note two significant

¹We often refer to an electron-hole pair as resulting from the break up of a covalent bond. This term should not be interpreted to mean that the electron and hole coordinate their behavior in any way after they have been generated. Each carrier goes its own way.

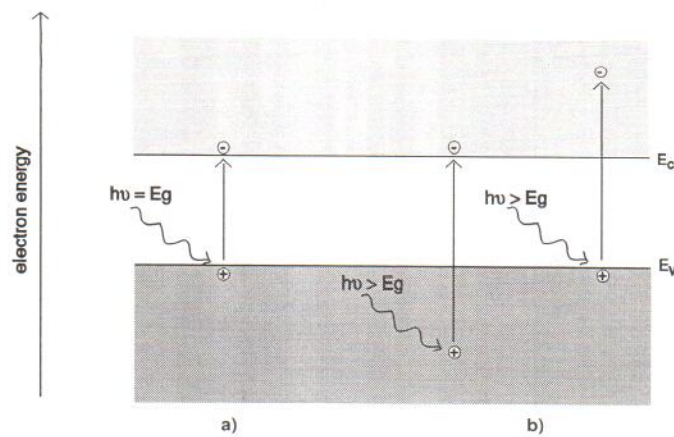


Figure 2.4: Energy band diagram depicting photons impinging on a semiconductor. In a) the photon has an energy identical to the bandgap energy. As a result, the photogenerated electron and hole sit at the respective band edges. In b), the two photons have more energy than the bandgap. In one case (right), the extra available energy beyond the bandgap is given as kinetic energy to the electron. In the other case (left) it is given to the hole.

inaccuracies that are implied in this figure. In Fig. 2.2 the electron and hole are sketched as having a “size” of the order of an interatomic distance in the semiconductor, about $0.2\text{--}0.3\text{ nm}$. This is wrong by at least an order of magnitude. In fact, in Ch. 1 we estimated the effective size of an electron at room temperature to be about 7.6 nm . Since a hole is not expected to be very different from an electron (we will get to better appreciate the differences between them as we advance in this book), this is also a good working number for the size of a hole. We also showed that a sphere of a diameter of 7.6 nm contains over 10,000 atoms. This illustrates how spread out electrons and holes really are in a semiconductor. A conduction electron is not the little lump represented in Fig. 2.2 but is spread out over thousands of atomic sites. Also, the notion that a hole jumps from a precise bonding location to a neighboring one is not very physical. A “fuzzier” probabilistic view is actually more appropriate.

The second inaccuracy implied in Fig. 2.2 comes from the view that the movement of a hole really represents the “retrograde” motion of a particular bonding electron. Not only does the large size of a hole make this view rather ambiguous, but it ignores the quantum mechanical nature of the bonding electrons in a semiconductor, which is very different from electrons in vacuum. The retrograde view of the motion of a hole can actually lead to rather unphysical conclusions. However, for the kinds of applications that we are interested in in this book, the collective dynamic behavior of the quantum-mechanical bonded electrons in a semiconductor in the presence of a broken bond is well represented by an ordinary classical particle of charge $+q$. The analogy of a bubble of air in a liquid representing the complex collective behavior of liquid molecules around it is really insightful in this regard.

As a final remark in this section, from here on in this text we will refer to conduction electrons as simply “electrons.” We will describe the behavior of bonding electrons in the valence band in terms of holes. We will totally ignore in the rest of our studies the core electrons that are tightly bound to each constituent atom of the crystal, as they play no role in semiconductor devices.

2.2 Intrinsic semiconductor

In a semiconductor in thermal equilibrium at a finite temperature, some of the bonds that tie the constituent atoms are broken. In consequence, there is a certain number of free electrons in the conduction band and a number of holes left in the valence band. It is inevitable to ask the question: in this situation, how many electrons and holes are there per unit volume? How do these concentrations depend on temperature? Rigorously answering these questions is the purpose of this chapter.

Let us first say a few words about what we mean by an "ideal" semiconductor. For the time being we will define this as a perfectly crystalline piece of a semiconductor that is 100% pure (*i.e.*, no foreign atoms) and that is unaffected by any surface effects. An ideal semiconductor defined this way is also called an *intrinsic semiconductor*. We will soon see that we can relax the second aspect of this stringent definition to just "sufficiently" pure, *i.e.*, some impurities might be present without changing the picture substantially, provided that their concentration is not too high. Also later on, we will understand the implications of the third restriction regarding surface effects. In subsequent Chapters we will learn how to deal with surfaces and to evaluate how far their influence extends into the body of a semiconductor.

So the question is again, in an intrinsic semiconductor in thermal equilibrium, how many electrons and holes are there? Answering this question requires a fairly elaborate model that is developed later on in this chapter. However, a number of important dependencies can be readily identified.

It is obvious that in an intrinsic semiconductor in thermal equilibrium, the number of electrons and holes has to be identical to each other at all temperatures. This is because every bond that gets broken produces precisely one electron and one hole. Also, every bond that is formed eliminates from circulation one electron and one hole. If we define n_o as the equilibrium electron concentration (number of electrons in the conduction band per unit volume) and p_o as the equilibrium hole concentration (number of holes in the valence band per unit volume), it is clear that n_o must equal p_o everywhere. This concentration is also called the *intrinsic carrier concentration*, n_i . Hence, in an intrinsic semiconductor in thermal equilibrium,

$$n_o = p_o = n_i \quad (2.1)$$

Typically, n_o , p_o , and n_i are given in units of cm^{-3} .

What are the key dependencies of n_i ? Intuitively we already know that n_i must have a direct dependence on temperature. The higher the temperature, the more vigorously the atoms vibrate in the semiconductor lattice and the easier it will be for a bond to break. We also expect n_i to exhibit some kind of inverse dependence on the bandgap of the semiconductor, E_g . This is because the higher the bandgap, the harder it is to break a bond and liberate an electron and a hole.

To be more specific about the dependence of n_i on T and E_g , we need a detailed model. This is presented below in Sect. 2.4.4. Yet, an analogy with chemical reactions, can reveal the leading

dependencies. Consider, for example, the well known reaction of decomposition of water into its two constituent ions:



When a water molecule decomposes, a hydrogen ion and a hydroxyl ion are produced. These two ions can combine again to form a water molecule. This chemical reaction is in some way similar to the break up of a crystalline bond in a semiconductor. When a crystalline bond breaks, an electron and a hole are produced. A bond is formed when an electron and a hole come together. Bond break up and formation in a crystalline semiconductor can be therefore thought of as the following chemical reaction:



This analogy is powerful because we can exploit what we have learned in elementary chemistry to understand a great deal about the statistics of electrons and holes in semiconductors. We know, for example, that in a chemical reaction the *law of mass action* establishes a relationship between the concentrations of reactants and reaction products, through a so-called "rate constant." We also know that the rate constant exhibits a peculiar dependence on temperature and on the energy required for the reaction to take place. For the water decomposition reaction, the law of mass action is written as:

$$K = \frac{[H^+][OH^-]}{[H_2O]} \sim \exp\left(-\frac{E}{kT}\right) \quad (2.4)$$

where E is the energy consumed or released in the water decomposition reaction.

Eq. 2.4 exhibits a peculiar dependence on T and E that is rather common in nature. Equations such as this apply to processes that are "thermally activated", that is, processes in which the rate of reaction is limited by the need to overcome a certain energy barrier. This threshold energy is called the "activation energy."

In analogy with Eq. 2.4, we could write a similar law of mass action for the crystalline bond dissociation reaction of 2.3, as follows:

$$K = \frac{n_o p_o}{[bonds]} \sim \exp\left(-\frac{E_g}{kT}\right) \quad (2.5)$$

where n_o and p_o have been defined above, and $[bonds]$ refers to the concentration of unbroken bonds. For the bond dissociation reaction, the activation energy is the bandgap energy since this is the energy that is required to break a bond.

Under typical conditions, the number of broken bonds in a semiconductor is a tiny fraction of the total number of bonds (this becomes clear when we put some numbers later on). Or, in other words, the concentration of bonds is never upset in a significant way (or the semiconductor

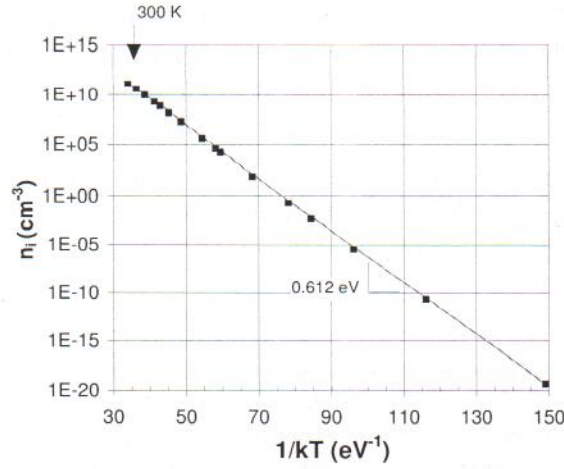


Figure 2.5: Arrhenius plot of n_i in Si. Measurements of Misiakos and Tsamakis [*Journal of Applied Physics* **74**, 3293 (1993)].

will melt). In consequence, the concentration of unbroken bonds ($[bonds]$ in Eq. 2.5) is to first order constant, and then we can conclude that:

$$n_o p_o \sim \exp\left(-\frac{E_g}{kT}\right) \quad (2.6)$$

The fact that the product of the equilibrium hole and electron concentrations has this peculiar dependence on E_g and T is a very significant result with important consequences. Although obtained here in a qualitative way, the rigorous treatment carried out later on in this chapter will lead to the same set of dependencies.

The result of Eq. 2.6, when combined with Eq. 2.1, allows us to obtain the key dependencies of n_i :

$$n_i \sim \exp\left(-\frac{E_g}{2kT}\right) \quad (2.7)$$

As expected, n_i increases with temperature and decreases with E_g . What is interesting is the specific negative exponential form of these dependencies. This is in fact readily observed in experiments. Fig. 2.5 shows experimental measurements of n_i for Si as a function of temperature. In this figure, the ordinate graphs n_i in a semilog scale, while the abscissa graphs $1/kT$ in a linear scale. T increases towards the left. When graphed this way, a straight line with a negative slope is obtained. This is consistent with Eq. 2.7 and is the characteristic signature of a thermally activated process. This kind of graph is called an *Arrhenius plot* in honor of Svante Arrhenius who showed around 1889 the pervasiveness of thermally activated processes in nature. From the slope of the straight line, the bandgap of Si can be obtained. For the data of Fig. 2.5, a bandgap energy of 1.224 eV is obtained. This is close but not exactly equal to the best known value of

1.124 eV for Si at room temperature. The reason for this small discrepancy is twofold. First, the missing prefactor in Eq. 2.7 is slightly temperature dependent. This prefactor will be derived later in Section 2.4. Additionally, E_g itself also depends on temperature, as discussed in Advanced Topic AT2.1. Issues around the temperature dependence of n_i are explored in detail in Problem 2.4.

The chemical reaction analogy does not give us a complete expression for n_i . The prefactor is missing and we therefore do not know how to compute its absolute value. To do this, we have to go through the detailed development described in Section 2.4. For reference, n_i for Si at room temperature is about 10^{10} cm^{-3} .

There is a second and hugely important result that emerges from Eq. 2.6. This equation states that the equilibrium np product in a semiconductor at a certain temperature is a constant that is specific to the semiconductor. Establishing a parallel again with the water dissociation reaction, this is equivalent to saying that the product of the concentration of H^+ and OH^- ions in pure water at a certain temperature is a constant. The implications of this are obvious. If we increase the concentration of H^+ in water (for example, by dissolving a small amount of acid), the concentration of OH^- will drop accordingly. This is because the equilibrium values of H^+ and OH^- are established from the balance of a dissociation reaction pointing to the right in 2.2 and a recombination reaction pointing to the left. Selectively adding H^+ to the bath enhances the rate of the left-pointing reaction bringing down the OH^- concentration.

The same arguments apply to electrons and holes in semiconductors. In the case of a perfectly pure semiconductor, the equilibrium electron and hole concentrations are identical and equal to n_i . This value arises from a balance between the rate of break up of bonds and the rate of formation of bonds. If we could find a way to selectively introduce, say, electrons so that the electron concentration increases above n_i , the rate of formation of bonds will increase and the hole concentration will drop below n_i . The product will remain a constant. Since for an intrinsic semiconductor $n_o = p_o = n_i$, in general, then,

$$n_o p_o = n_i^2 \quad (2.8)$$

This is a very important equation that will be used extensively in this and subsequent Chapters. When we do our more detail analysis in Section 2.4 below, we will find that this equation is of broad applicability but it is not completely general. In the language of chemical reactions again, it works in the "dilute" regime, that is, if the concentration of carriers that is added is not too high.

How carriers can be added in a selective way to a semiconductor is described in the next Section.

2.3 Extrinsic semiconductor

So far we have considered only the introduction of electrons and holes in a semiconductor via the spontaneous break up of atomic bonds. There is another way in which electrons and holes

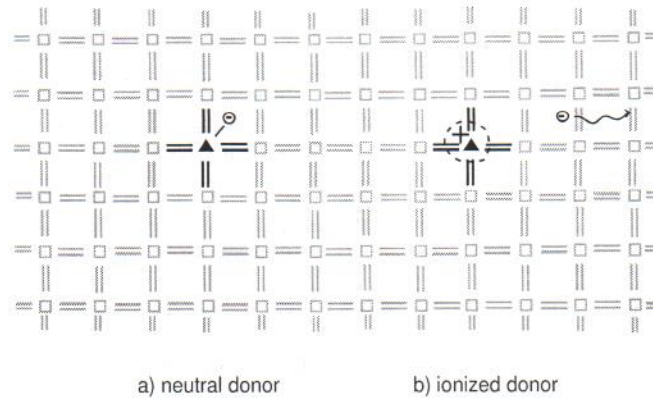


Figure 2.6: Sketch of a donor atom in a semiconductor lattice: a) in a neutral state, b) in an ionized state.

can be introduced and that is by "doping" it with small amounts of selected impurities. Key to the process of doping is the fact that it is possible to selectively introduce electrons into the conduction band without producing an equal amount of holes in the valence band and vice versa. If the resulting concentration of any one of the carriers overwhelms the thermal contribution that arises from the break up of covalent bonds, the semiconductor is said to be "extrinsic." When electrons are introduced in a preferential way, the semiconductor is said to be *n-type*. If holes are in the majority, the semiconductor is called *p-type*.

Doping is one of the pillars over which the microelectronics industry is built. Precise placement of dopant atoms in a semiconductor allows the creation of n-type and p-type regions with nearly arbitrary shape and doping distribution. It is largely through careful engineering of doping profiles that devices are designed to deliver the required specifications for different applications. Doping can be accomplished in a variety of ways, such as solid-state diffusion or ion implantation. In diffusion, dopant atoms diffuse from a dopant-rich source into the semiconductor. In ion implantation, ionized dopant species are accelerated to high energies and "slammed" against the semiconductor penetrating to a certain depth. A thermal activation step allows the incorporation of the dopants into substitutional locations in the lattice.

The following subsections describe how certain impurities are able to selectively introduce electrons or holes into a semiconductor. We will also learn to compute the equilibrium carrier concentrations in an extrinsic semiconductor.

2.3.1 Donors and acceptors

Consider what happens if, in a perfect Si lattice, a Si atom is substituted by a foreign atom, such as As or P, that comes from Column V of the periodic table (see Fig. 1.4). Since these atoms are located close to Si in the periodic table, they are not very different from Si. They can take the place of a Si atom without disrupting the lattice too much. The foreign atom is said in this case to be in a substitutional position. The uniqueness of column V atoms is that they have

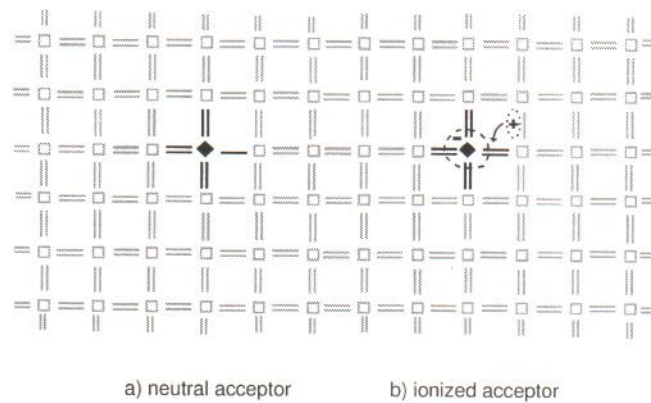


Figure 2.7: Sketch of an acceptor atom in a semiconductor lattice: a) in a neutral state, b) in an ionized state.

five electrons in the outer layer. This means that after consuming four electrons to bond to the four nearest neighbors, there is a fifth electron that remains loosely attached - its binding energy is typically in the $10 - 50 \text{ meV}$ range. At room temperature this fifth electron can easily pick up enough energy from the thermal energy of the lattice to escape the attraction of the donor atom and roam freely around the crystal. This electron is indistinguishable from a thermally created electron. The substitutional column V atom has thus "donated" one of its electrons to the conduction band. The "donor" is left positively charged since the number of protons in the nucleus is one unit higher than the number of electrons around it. This situation is depicted in Fig. 2.6.

A parallel situation takes place if Si is doped with column III elements such as B. These atoms have three electrons in the outer layer. A substitutional B atom bonds with three Si neighbors but is unable to bond with the fourth one. There is in effect a hole in the covalent bonding structure at the location of the impurity atom. This hole can easily migrate away from the impurity site if a nearby bonding electron jumps in and satisfies the missing covalent bond in a process that typically takes less than 0.1 eV . At this point the hole has moved to a neighboring location and is indistinguishable from a regular thermally created hole. In the process of releasing a hole to the valence band, the impurity has "accepted" a bonding electron and has become negatively charged. This is depicted in Fig. 2.7.

For GaAs and other III-V semiconductors the situation is slightly more complicated. Elements from Column VI, such as Se, are donors when placed substitutionally at an As site. Elements from Column II, such as Zn or Be, behave as acceptors at a Ga site. Elements from Column IV, such as Si or Ge, have an *amphoteric behavior*. This means that when placed at an As site, they behave as acceptors, but at a Ga site, they are donors. With amphoteric impurities, different doping conditions can result in either an n-type or a p-type semiconductor. In spite of this, Si is a widely used donor for III-V semiconductors. This is because it is not difficult to create conditions in which there is a high concentration of vacant Ga sites ("Ga vacancies") and a small concentration of vacant As sites ("As vacancies"). In this way, the Si atoms preferentially find themselves in substitutional positions at Ga sites where they behave as donors.

Electron states associated with donor and acceptor atoms can be represented in the energy band diagram as shown in Fig. 2.8. Donor states are graphed inside the forbidden gap slightly below the conduction band edge at E_D . The reason is that it only takes a small amount of energy to release the loosely bound electron to the conduction band. This energy is called the *donor ionization energy* and is represented with the symbol E_d . Similarly, acceptor states are represented slightly above the valence band, at E_A , as a little energy will result in the release of a hole to the valence band. This energy is called the *acceptor ionization energy*, E_a . Broken lines are used for the donor and acceptor states to indicate the fact that the impurity atoms are typically far apart from each other. Table 2.1 in Advanced Topic AT2.4 collects typical ionization energies for common Si and GaAs dopants. As is the case for Si atoms, the remaining bonding electrons of the impurity atoms are represented as full states inside the valence band.

2.3.2 Charge neutrality

A piece of a semiconductor in thermal equilibrium, as a whole, is charge neutral. This is regardless of whether it is elemental or compound, doped or undoped, at a finite temperature or at 0 K. Every atom of the semiconductor is charge neutral to begin with, that is, all atoms have matching numbers of electrons and protons. If no particles are allowed to escape from the semiconductor, overall charge neutrality must prevail.

Spatially, the picture can be more complex. First of all, the disruption to the perfect ordering of the crystal that is introduced by the surfaces can alter charge neutrality in their vicinity. This can happen even if the semiconductor is "pure", that is, without any dopant impurities. We will study the effect of surfaces later on in this book. Restricting ourselves to the bulk, far away from the surfaces, local charge neutrality might not exist either if the dopant distribution is not uniform in space. We will study this in greater detail in Ch. 4.

With these two caveats in mind, in the bulk of a uniformly doped semiconductor in thermal equilibrium, charge neutrality must prevail. Mathematically, for a general case in which the semiconductor is doped with donors and acceptors simultaneously, the balance of the negatively charged species and the positively charged species must be zero:

$$-n_o + p_o + N_D^+ - N_A^- = 0 \quad (2.9)$$

This is an important consideration when calculating equilibrium carrier concentrations in doped semiconductors, as done in the next subsection.

2.3.3 Equilibrium carrier concentration in a doped semiconductor

In order to compute the equilibrium carrier concentrations in the bulk of a uniformly doped semiconductor it is essential to first know the fraction of the dopant atoms that are ionized. Fortunately, the ionization energy of typical dopants is small enough that at room temperature a great majority of them are ionized. Denoting N_D as the donor concentration, N_D^+ as the

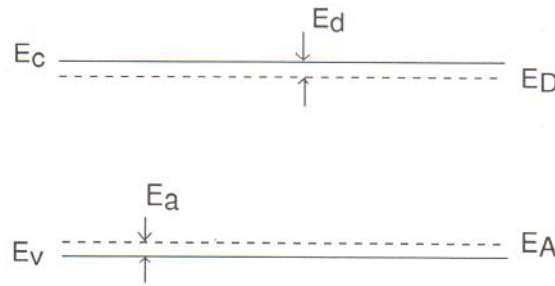


Figure 2.8: Representation of donor and acceptor states in the band diagram of a semiconductor.

ionized donor concentration, N_A as the acceptor concentration, and N_A^- as the ionized acceptor concentration, under many practical circumstances, we can write:

$$N_D^+ \simeq N_D \quad (2.10)$$

$$N_A^- \simeq N_A \quad (2.11)$$

At low temperatures, this cannot be the case since there is less thermal energy available and certain impurities will be in a neutral state. When designing devices to operate at low temperatures, it is crucial to correctly account for "carrier freeze-out," as this phenomenon is known. This is described in Advanced Topic AT2.5.

It might seem counter intuitive to state that dopant atoms with an ionization energy on the order of 40-60 meV can be all ionized at room temperature when the thermal energy is only 25 meV. This apparent contradiction is resolved by realizing that in most cases there are many more states close to the band edges ready to accommodate carriers than dopant states at a slightly lower energy. As a result, the slight energetic disadvantage of the band states is more than compensated by their large numbers. In other words, for typical doping concentrations, even though it is energetically favored, it is hard for electrons in the conduction band to find an ionized donor to fall into. The same applies to acceptors and holes in the valence band. One can view this as a golfer trying to get a golf ball in the hole. From potential energy arguments, it is certainly true that the hole is a more favorable position for the ball to be in. It does not follow, however, that regardless how the golfer hits the ball, it will neatly fall in the hole! For a typical doping level of 10^{17} cm^{-3} the ionization ratio of P in Si at room temperature is 97%, while this figure is 94% for B in Si. Advanced Topics AT2.4 and AT2.5 show how to perform these calculations.

Carrier production from dopant atoms proceeds in parallel with the thermal electron-hole pair formation from broken bonds. In a completely general case, both mechanisms must be accounted for in order to compute the equilibrium carrier concentrations. For most common semiconductors around room temperature, however, the doping concentrations overwhelm the intrinsic carrier concentration. In this instance, the equilibrium carrier concentrations are easy

to compute. For an n-type extrinsic semiconductor, the equilibrium electron concentration is determined by the donor concentration:

$$n_o \simeq N_D \quad (2.12)$$

and the hole concentration can be found from Eq. 2.8:

$$p_o = \frac{n_i^2}{n_o} \simeq \frac{n_i^2}{N_D} \quad (2.13)$$

Since n_o is much higher than n_i , it follows that p_o is much smaller than n_i . In an n-type semiconductor, then, $n_o \gg p_o$. In this case, electrons are said to be *majority carriers* while holes are called *minority carriers*. Eq. 2.12 could have also been obtained from Eq. 2.9 by setting $N_A = 0$ and $n_o \gg p_o$.

Exercise 2.1: Estimate the equilibrium electron and hole concentrations in Si uniformly doped with $N_D = 10^{16} \text{ cm}^{-3}$ at room temperature. State your assumptions.

Assuming that all donors are ionized, Eq. 2.12 gives:

$$n_o \simeq N_D = 10^{16} \text{ cm}^{-3}$$

Plugging in Eq. 2.13 we find:

$$p_o \simeq \frac{n_i^2}{N_D} \simeq \frac{(10^{10} \text{ cm}^{-3})^2}{10^{16} \text{ cm}^{-3}} = 10^4 \text{ cm}^{-3}$$

where we have used an approximate value for n_i (the error in the estimation of p_o is less than 10% as a result of this approximation).

This exercise shows just how different the equilibrium electron and hole concentrations are in a typical semiconductor under fairly standard conditions.

It is important to note here that while Eq. 2.12 assumes full dopant ionization, Eq. 2.13 is more restrictive since it additionally requires the validity of Eq. 2.8. As discussed below, there are some limitations to this. Nevertheless, for many situations, Eq. 2.13 is adequate and will be used extensively in this book.

In a parallel way, for a p-type semiconductor in equilibrium, we have:

$$p_o \simeq N_A \quad (2.14)$$

and

$$n_o \simeq \frac{n_i^2}{N_A} \quad (2.15)$$

In a p-type semiconductor, holes are the majority carriers and electrons are the minority carriers. Note again that Eq. 2.15 applies to the extent that Eq. 2.8 is valid.

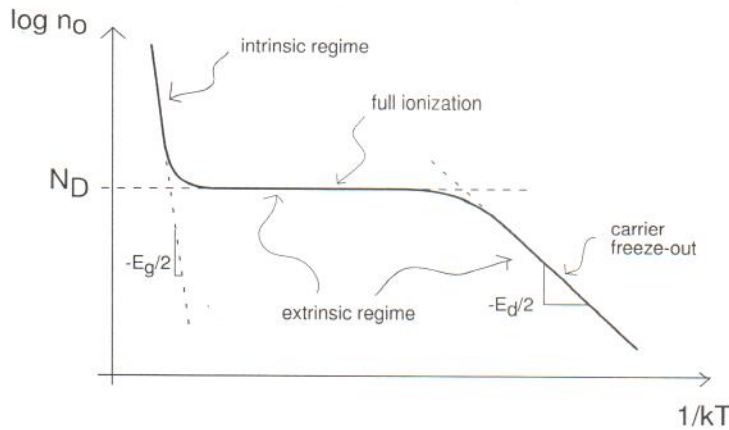


Figure 2.9: Arrhenius plot of n_o for an n-type semiconductor.

At high temperatures, an extrinsic semiconductor eventually becomes intrinsic as the thermal break up of covalent bonds overwhelms the doping concentration. In this instance, Eqs. 2.12-2.15 fail and Eq. 2.1 applies. In more general terms, if a semiconductor has a concentration of dopants much smaller than n_i the equilibrium electron and hole concentration is basically equal to n_i . We will still denote such a semiconductor as an *intrinsic semiconductor* even though it is not perfectly pure. However, it is important to recognize that the value of n_i for Si at room temperature is so low ($\sim 10^{10} \text{ cm}^{-3}$) that it is very hard to obtain a Si wafer that is intrinsic at room temperature. The residual dopant concentration that unavoidably exists almost always overwhelms n_i . This is also the case for most common semiconductors. This situation changes at higher temperatures since n_i increases rapidly with temperature. It is in fact not difficult to have intrinsic Si at say 600°C , as a numerical exercise in Advanced Topic AT2.1 shows.

The main results of this section are graphically summarized in the sketch of an Arrhenius plot of equilibrium electron concentration in an n-type semiconductor that is shown in Fig. 2.9. A similar picture applies for p_o in a p-type semiconductor. Fig. 2.9 shows that there is a broad intermediate regime in which $n_o \simeq N_D$. In this regime, the semiconductor is extrinsic and all donors are fully ionized. This is the most common regime of device operation. At low enough temperatures (towards the right in the diagram), the semiconductor is still extrinsic but not all dopants are ionized. In this freeze-out regime, the electron concentration drops very quickly with temperature (the activation energy of n_o is shown in Advanced Topic AT2.5 to be $E_d/2$). At high enough temperatures (towards the far left in the diagram), intrinsic carrier production overwhelms the doping level and the semiconductor becomes intrinsic.

Throughout this section we have considered semiconductors that are solely doped with donors or acceptors. It is very common in devices to have regions that contain both kinds of dopants. This happens, for example, in a PN junction fabricated by introducing n-type dopants into a p-type wafer. In this case, the n-type region is still n-type but it is partially compensated by the acceptor concentration. It is not difficult to derive general expressions for the equilibrium carrier concentrations for a *compensated semiconductor*, as this situation is known. If there is to be compensation, sound engineering practice, *i.e.* good process control, dictates that one type of

dopant should clearly overwhelm the other, that is either $N_D \gg N_A$ or $N_A \gg N_D$. It is easy to show that in this case, all equations derived above apply if we substitute $N_d = N_D - N_A$ for N_D in Eqs. 2.12 and 2.13, and $N_a = N_A - N_D$ for N_A in Eqs. 2.14 and 2.15 (see Problem 2.9). N_d and N_a are often called *net* donor and acceptor concentrations, respectively.

2.4 Carrier Statistics in Equilibrium

The results obtained in this chapter so far, though useful and relevant, are based on a qualitative analogy between electron-hole production in an ideal semiconductor and a chemical reaction. This view has left a few important voids, such as a complete expression for n_i . It is actually not very difficult to develop a rigorous set of models for the electron and hole concentrations in thermal equilibrium in intrinsic and doped semiconductors. The most physically meaningful way to do this is to adopt an energy view of semiconductors by exploiting energy band diagrams and using the concept of Fermi level introduced in the previous chapter.

In the following sections, we will develop a formulation for carrier statistics in semiconductors in thermal equilibrium in several steps. First, we will present a description of the density of states of the conduction and valence bands close to their respective band edges. Then, we will derive relationships between the equilibrium electron and hole concentrations and the relative location of the Fermi level within the band structure of the semiconductor. From this, we will be able to derive expressions for the $n_o p_o$ product and n_i . We will finish with a discussion about the location of the Fermi level in intrinsic and extrinsic semiconductors.

2.4.1 Conduction and valence band density of states

As introduced in the previous Chapter, the density of states (DOS) is a parameter that quantitatively describes the distribution of states available for electrons in a crystal. The DOS of the valence and conduction bands in semiconductors has in general a rather complicated shape. Fig. 2.10 shows, for example, calculations of the DOS for Si and GaAs as a function of energy.

As will become clear soon, a lot of the carrier action in a semiconductor takes place at energies close to the band edges. This implies that, in many device studies, there is no need for a detailed description of the entire conduction and valence bands. Accuracy at energies close to the edges of the bandgap will suffice.

A rather fundamental result of solid-state physics, and also an experimental observation, is that in the vicinity of the band edges, the density of states increases following a square-root dependence on energy. As a consequence, the conduction band density of states of a semiconductor, $g_c(E)$, and the valence band DOS, $g_v(E)$, close to the band edges can be expressed as:

$$g_c(E) = A_c \sqrt{E - E_c} \quad E \geq E_c \quad (2.16)$$

$$g_v(E) = A_v \sqrt{E_v - E} \quad E \leq E_v \quad (2.17)$$

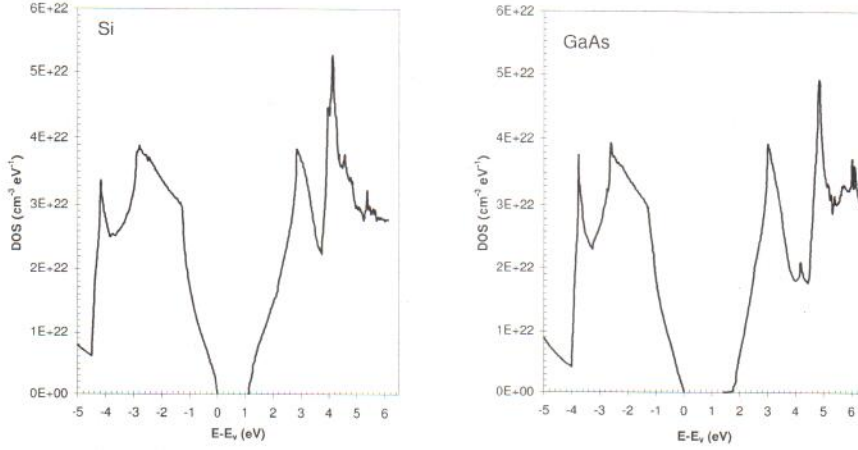


Figure 2.10: Density of states of the valence and conduction bands in Si (left) and GaAs (right) as a function of energy. Close to the band edges, the DOS have a square-root dependence with energy [data courtesy of M. Fischetti, IBM].

$g_c(E)$ and $g_v(E)$ have units of $\text{eV}^{-1} \cdot \text{cm}^{-3}$. They are sketched in Fig. 2.11.

A_c and A_v are two proportionality constants. For reasons that are best understood in a solid-state course, it is common practice to write these constants in the following way:

$$A_c = 4\pi \left(\frac{2m_{de}^*}{h^2} \right)^{3/2} \quad (2.18)$$

$$A_v = 4\pi \left(\frac{2m_{dh}^*}{h^2} \right)^{3/2} \quad (2.19)$$

m_{de}^* and m_{dh}^* defined this way have units of mass (in $\text{eV} \cdot \text{s}^2/\text{cm}^2$ in the "microelectronics" unit system). Because of this, they are called respectively, the *density of states electron effective mass* and the *density of states hole effective mass*. Three comments should be made about these two new entities.

- First, the values of m_{de}^* and m_{dh}^* , which are semiconductor dependent, are typically given in terms of the electron rest mass m_o . For example, $m_{de}^* = 1.09m_o$ and $m_{dh}^* = 1.15m_o$ for Si at room temperature (they depend weakly on temperature).
- Second, the term "effective mass" is used with widely different meanings in the semiconductor literature. For example, you are likely to come across a conductivity effective mass, longitudinal or transverse effective masses, light-hole or heavy-hole effective masses, and others. Although closely related, these are all different physical entities and it is important not to mix them. In order to compute the DOS of a semiconductor, the *DOS effective mass* must be used.

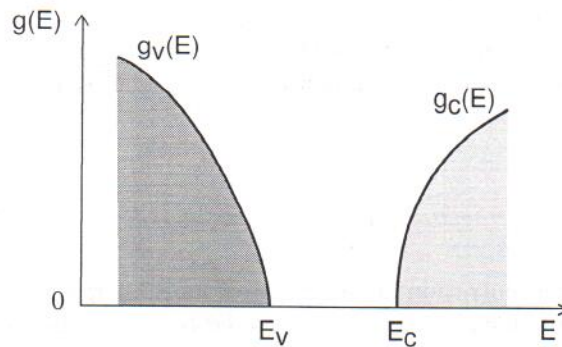


Figure 2.11: Sketch of conduction and valence band density of states close to the band edges.

- Finally, the DOS effective mass is definition sensitive. Some authors use slightly different expressions for Eqs. 2.18 and 2.19 (a difference of a factor of two is not uncommon). When using the DOS effective mass out of a particular source, it is important to check the expression that is being used by the author to compute the DOS and apply appropriate corrections if needed. Eqs. 2.18 and 2.19 are the most popular expressions.

Fig. 2.11 sketches the DOS for a semiconductor such as Si in which $m_{de}^* < m_{dh}^*$. In this case, at a given distance away from the band edge, the conduction band has fewer states than the valence band. That is the meaning of a "heavier" DOS effective mass.

2.4.2 Equilibrium electron concentration

We now turn our attention to the distribution of electrons in the conduction band in thermal equilibrium. In this section we not only seek to understand the shape of the electron distribution in the conduction band but also we aim at developing expressions that will give us the total electron concentration in the conduction band as a function of the relative position of the Fermi level with respect to the band edge.

Before we get launched in a detailed mathematical formulation of the problem, it is useful to develop some intuition as to the general dependencies that are to be expected. Assume for a moment that the Fermi level is below the conduction band edge. It is clear that the closer the Fermi level gets to the band edge, the higher the occupation probability of the conduction band states becomes. Therefore, the total electron concentration increases. Furthermore, if the Fermi level is not too close to the conduction band edge, this relationship has to be exponential since it is the exponential tail of the Fermi-Dirac distribution function that overlaps with the conduction band.

In a situation in which the Fermi level is inside the conduction band, the electron concentration also increases the more the Fermi level penetrates into the band. However, this relationship is much weaker than in the earlier case as a much bigger portion of the Fermi-Dirac distribution function overlaps with the conduction band and not just its exponential tail. The detailed model

that is developed in this section will show these dependencies.

The total electron concentration in the conduction band is computed by integrating the electron energy distribution across the entire conduction band. This is mathematically expressed as:

$$n_o = \int_{E_c}^{\infty} n_o(E) dE \quad (2.20)$$

where $n_o(E)$ represents the concentration of electrons per unit energy (in units of $cm^{-3} \cdot eV^{-1}$) in thermal equilibrium (the subindex "o" is frequently used to denote thermal equilibrium). Note that the integral extends from the bottom of the conduction band all the way up to infinity. This is of course strictly incorrect as the conduction band does not extend that far. The error for not being more precise with the upper integration limit is negligible since significant electron concentration only exists at the bottom of the band.

The calculation of $n_o(E)$ is simple. At a given energy, the electron concentration is obtained by multiplying the conduction band density of states at that energy times the probability that a state located there is occupied by an electron:

$$n_o(E) = g_c(E) f(E) \quad (2.21)$$

If we use in Eq. 2.21 the Fermi-Dirac distribution function $f(E)$ (Eq. 1.7), $g_c(E)$ from Eqs. 2.16 and 2.18, and we substitute everything into Eq. 2.20 we obtain:

$$n_o = 4\pi \left(\frac{2m_{de}^*}{h^2} \right)^{3/2} \int_{E_c}^{\infty} \frac{\sqrt{E - E_c}}{1 + \exp \frac{E - E_F}{kT}} dE \quad (2.22)$$

We can rewrite this integral in a more concise form by changing variables in the following way:

$$\eta = \frac{E - E_c}{kT} \quad (2.23)$$

and

$$\eta_c = \frac{E_F - E_c}{kT} \quad (2.24)$$

where η represents the normalized energy with respect to the conduction band edge in units of kT .

After this change of variables, Eq. 2.22 becomes:

$$n_o = 4\pi \left(\frac{2m_{de}^* kT}{h^2} \right)^{3/2} \int_0^{\infty} \frac{\sqrt{\eta}}{1 + e^{\eta - \eta_c}} d\eta \quad (2.25)$$

We can further define:

$$N_c = 2 \left(\frac{2\pi m_{de}^* kT}{h^2} \right)^{3/2} \quad (2.26)$$

N_c is called the **effective density of states of the conduction band**. Its units are cm^{-3} . The physical meaning of this parameter is explored in Problem 2.2.

Exercise 2.2: Compute N_c for Si at room temperature.

As in previous exercises, it is essential to be extremely careful with the units of the various fundamental constants. In computing N_c , the first step is to calculate the density of states effective mass for electrons m_{de}^* . From Appendix B at the end of the book, $m_{de}^*/m_o = 1.09$. Then $m_{de}^* = 1.09 \times 5.69 \times 10^{-16} \text{ eV} \cdot \text{s}^2/\text{cm}^2 = 6.20 \times 10^{-16} \text{ eV} \cdot \text{s}^2/\text{cm}^2$. We can now plug this in Eq. 2.26 to get:

$$N_c = 2 \left[\frac{2 \times 3.14 \times 6.20 \times 10^{-16} \text{ eV} \cdot \text{s}^2/\text{cm}^2 \times 25.9 \times 10^{-3} \text{ eV}}{(4.14 \times 10^{-15} \text{ eV} \cdot \text{s})^2} \right]^{3/2} = 2.85 \times 10^{19} \text{ cm}^{-3}$$

which is fairly close to the currently accepted value of $2.86 \times 10^{19} \text{ cm}^{-3}$ that is listed in Appendix B.

With the definition of N_c , Eq. 2.25 can be rewritten as:

$$n_o = N_c \mathcal{F}_{1/2}(\eta_c) \quad (2.27)$$

where $\mathcal{F}_{1/2}$ is called the *Fermi-Dirac integral of order 1/2*:

$$\mathcal{F}_{1/2}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{\eta}}{1 + e^{\eta-x}} d\eta \quad (2.28)$$

where x is an independent variable.

The function $\mathcal{F}_{1/2}$ has some interesting properties that are summarized in Advanced Topic AT2.2 at the end of this chapter. It is also tabulated in several books for different values of x . A plot of $\mathcal{F}_{1/2}$ is shown in Fig. 2.12.²

There are several interesting features to the result captured in Eq. 2.27. This equation only depends on η_c , the normalized Fermi energy with respect to the conduction band edge. Since $\mathcal{F}_{1/2}(x)$ is a monotonically increasing function of x (see Fig. 2.12), the higher the Fermi level is relative to the conduction band edge, the more electrons there are in the conduction band. This is to be expected since the occupation probability of any given state in the conduction band increases the higher the Fermi level lies with respect to it. The increase in n_o with increasing $E_F - E_c$ slows down, however, when the Fermi level penetrates into the conduction band (beyond

²In some textbooks, the Fermi-Dirac integral of order 1/2 is defined as $F_{1/2}(x) = \frac{\sqrt{\pi}}{2} \mathcal{F}_{1/2}(x)$. To make matters more confusing, the symbols are not used in a consistent way across the literature. It is important to always pay close attention to the definition of this integral.

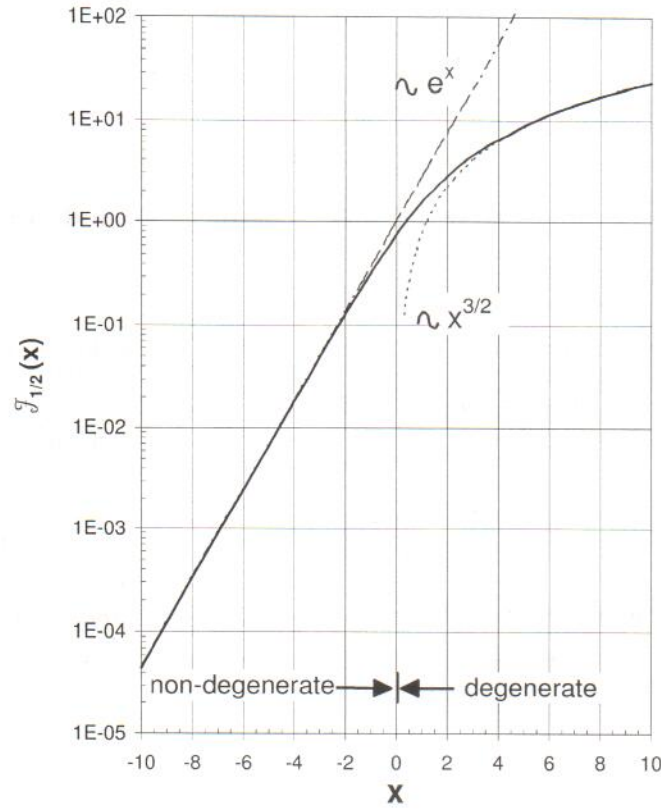


Figure 2.12: Fermi-Dirac integral of order 1/2. Also sketched are two simple analytical approximations.

$x = 0$ in Fig. 2.12). When this happens, the semiconductor is referred to as *degenerate*, as opposed to *non-degenerate* as it is called when E_F is below E_c .

With some restrictions in the value of x , it is possible to obtain accurate analytical approximations to $\mathcal{F}_{1/2}(x)$. Some of the most common ones are collected in Advanced Topic AT2.2 at the end of this chapter. The most useful of them is for $x \ll -1$ for which:

$$\mathcal{F}_{1/2}(x) \simeq e^x \quad \text{for } x \ll -1 \quad (2.29)$$

This approximation is graphed in Fig. 2.12.

Eq. 2.29 is useful because it permits us to write a simple analytical expression for n_o if the semiconductor is sufficiently non-degenerate. That is, if $\eta_c \ll -1$, n_o can be written as:

$$n_o \simeq N_c \exp \frac{E_F - E_c}{kT} \quad (2.30)$$

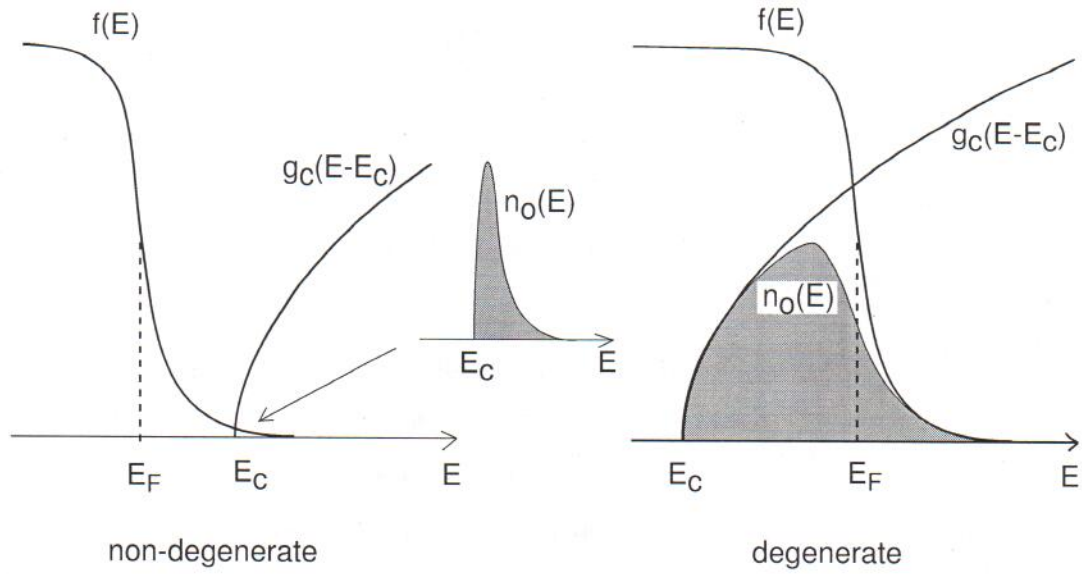


Figure 2.13: Sketch of electron distribution in conduction band in non-degenerate case (left) and degenerate case (right).

This equation is valid when $E_C - E_F \gg kT$, or in other words, when the Fermi level is far below the conduction band edge (about $3kT$ is usually enough). When this is the case, $n_0 \ll N_C$ which is a handy rule to remember.

An alternate analytical approximation can be obtained for a strongly degenerate semiconductor. This is also shown in Fig. 2.12. It is only valid if the Fermi level has penetrated substantially inside the conduction band. This approximation is discussed in more detail in Advanced Topic AT2.3.

Exercise 2.3: Estimate the location of the Fermi level with respect to the conduction band edge in a sample of Si at room temperature with: a) $n_o = 10^{17} \text{ cm}^{-3}$, b) $n_o = 10^{20} \text{ cm}^{-3}$.

For case a), $n_o = 10^{17} \text{ cm}^{-3} \ll N_c$. We can then use Maxwell-Boltzmann statistics. Solving for $E_F - E_c$ in Eq. 2.30, we get:

$$E_F - E_c = kT \ln \frac{n_o}{N_c} = 0.026 \ln \frac{10^{17}}{2.86 \times 10^{19}} = -0.15 \text{ eV}$$

Indeed E_F is below E_c by several kT s, as expected.

For case b), $n_o > N_c$ and the semiconductor is degenerate. We must use Fermi-Dirac statistics for electrons. Solving for $\mathcal{F}_{1/2}(\eta_c)$, we get:

$$\mathcal{F}_{1/2}(\eta_c) = \frac{n_o}{N_c} = \frac{10^{20}}{2.86 \times 10^{19}} = 3.5$$

To first order, we can use Fig. 2.12 to estimate $\eta_c \simeq 2.5$. Hence:

$$E_F - E_c = kT\eta_c = 0.026 \times 2.5 = 0.065 \text{ eV}$$

Clearly E_F has penetrated inside the conduction band, as expected in a degenerate n-type semiconductor.

We can more accurately estimate the location of E_F using Eqs. 2.56 and 2.57 in Advance Topic AT2.2. If we do this, we find $\eta_c = 2.44$, and $E_F - E_c = 63 \text{ meV}$.

Had we estimated the location of E_F using non-degenerate statistics, as in case a), we would have obtained $E_F - E_c = 33 \text{ meV}$, an error of nearly 50%.

Fig. 2.14 summarizes graphically the electron concentration in equilibrium in the conduction band as a function of the relative position of the Fermi level with respect to the band structure in a semilog scale. As the Fermi level moves up inside the bandgap, n_o increases exponentially. When E_F penetrates in the conduction band and the semiconductor becomes degenerate, the dependence flattens out.

Qualitatively, the key difference between the non-degenerate and degenerate regimes lies in the energy distribution of electrons inside the conduction band. This is sketched in Fig. 2.13 (see also Problem 2.11). If the Fermi level is sufficiently below the conduction band edge, only the exponentially decaying high-energy tail of the Fermi distribution penetrates into the conduction band. In this case few states are occupied and the energy distribution of conduction band electrons also decays exponentially with energy. In fact, one can directly obtain Eq. 2.30 by using Maxwell-Boltzmann statistics in Eq. 2.21.

In a strongly degenerate semiconductor, the Fermi level is deep inside the conduction band and many electrons occupy the bottom of the band. As Fig. 2.13 sketches, the energy distribution in this case is rather different from a non-degenerate situation. It has nearly a square-root shape at the bottom of the band with an exponentially decaying shape at higher energies.

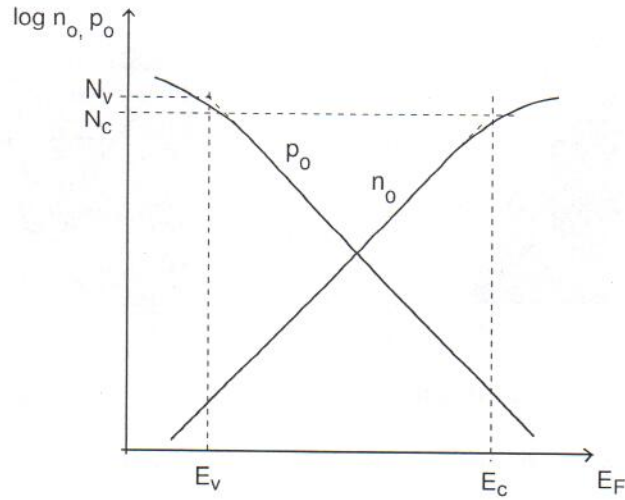


Figure 2.14: Sketch of electron and hole concentrations in equilibrium as a function of the relative position of the Fermi level with respect to the band structure.

2.4.3 Equilibrium hole concentration

Similar arguments as those presented in the previous section can be applied to compute the hole concentration in the valence band in equilibrium. This time we must integrate the hole distribution in energy across the entire valence band. For an equilibrium situation:

$$p_o = \int_{-\infty}^{E_v} p_o(E) dE \quad (2.31)$$

p_o is in this case given by the product of the density of states in the valence band times the probability that a given state is *empty*:

$$p_o(E) = g_v(E) [1 - f(E)] \quad (2.32)$$

If we proceed as above and define:

$$\eta_v = \frac{E_v - E_F}{kT} \quad (2.33)$$

and

$$N_v = 2 \left(\frac{2\pi m_{dh}^* kT}{h^2} \right)^{3/2} \quad (2.34)$$

we easily obtain

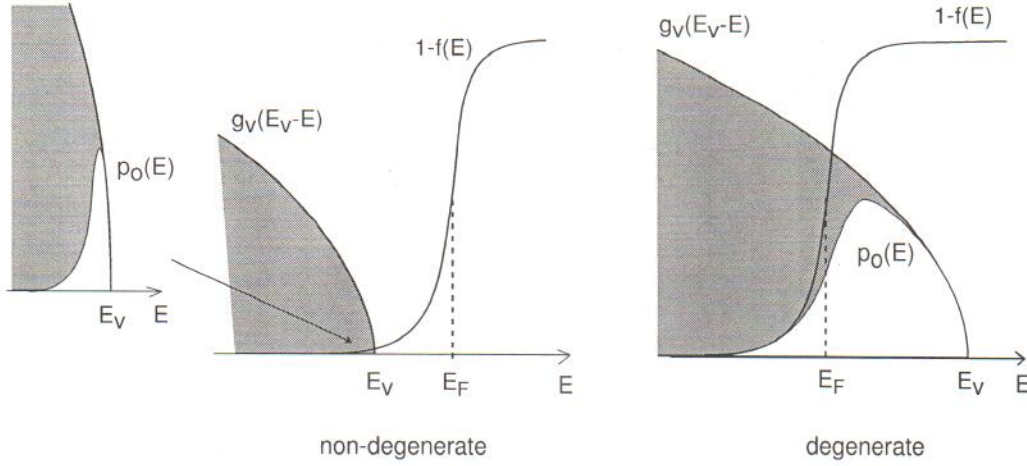


Figure 2.15: Sketch of hole distribution in valence band in non-degenerate (left) and degenerate case (right).

$$p_o = N_v \mathcal{F}_{1/2}(\eta_v) \quad (2.35)$$

N_v is called the **effective density of states of the valence band** and η_v gives the relative position of the Fermi level with respect to the valence band edge in units of kT . The hole concentration in equilibrium follows a similar relationship to what was obtained above for electrons. In this case, however, the lower the Fermi level with respect to the valence band edge, the higher the hole concentration. N_v in Si at room temperature is $3.10 \times 10^{19} \text{ cm}^{-3}$.

When the Fermi level is well above the valence band edge ($\eta_v \ll -1$, or $E_F - E_v \gg kT$, or $p_o \ll N_v$), the exponential approximation to the Fermi-Dirac integral applies. This allows us to write:

$$p_o \simeq N_v \exp \frac{E_v - E_F}{kT} \quad (2.36)$$

This is a case in which the hole distribution is *non-degenerate*. An analytical approximation can also be obtained for a strongly degenerate p-type semiconductor. This is shown in Advanced Topic AT2.3. The hole distribution in the valence band is sketched in Fig. 2.15 for these two cases.

Fig. 2.14 illustrates the dependence of p_o on the Fermi level position with respect to the band structure. As E_F moves up inside the bandgap, p_o drops exponentially. If the E_F is close or inside the valence band, the dependence of p_o on E_F is softer.

Fig. 2.16 offers a summary of the proper statistics that are to be used depending on the position of E_F with respect to the bands. If the Fermi level sits anywhere inside the forbidden gap, both electron and hole distributions are non-degenerate and Eqs. 2.30 and 2.36 are reasonably accurate. If the Fermi level penetrates inside anyone of the bands, then the semiconductor is degenerate. To be more specific, if the Fermi level is located inside the conduction band, then the electrons are degenerate but the holes are not. Fermi-Dirac statistics are necessary to describe the

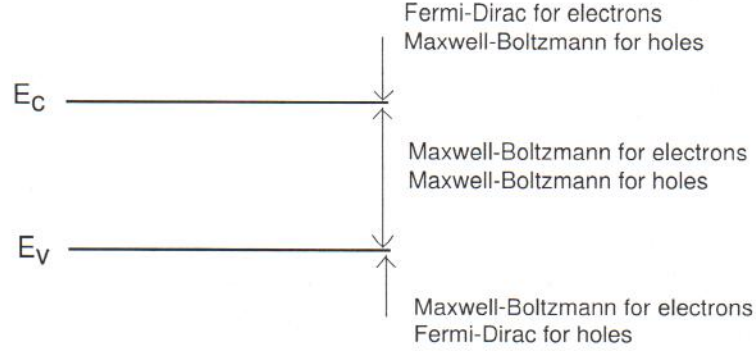


Figure 2.16: Summary of statistics for electrons and holes depending on the position of the Fermi level with respect to the band structure.

electrons but Maxwell-Boltzmann can be used for the holes. In this case, the exact expression 2.27 must be used. If the electrons are sufficiently degenerate, the strongly degenerate approximation 2.58 in Advanced Topic AT2.3 applies. If, conversely, the Fermi level has penetrated inside the valence band, the holes are degenerate but the electrons are not. A similar treatment applies when the Fermi level is in the vicinity or inside the valence band with the role of electrons and holes interchanged.

N_c and N_v represent practical benchmarks for the equilibrium carrier concentrations of electrons and holes in assessing whether the semiconductor is degenerate or not. Appendix B collects the values for Si and GaAs at room temperature.³ It is important to note that N_c and N_v increase with temperature in the form $\sim T^{3/2}$. The origin for this dependence is explored in Problem 2.2.

2.4.4 np product in equilibrium

Now that we have derived relationships between n_o and p_o and the relative location of the Fermi level with respect to the semiconductor band structure, we are in a position to derive an expression for the electron-hole product in equilibrium. Multiplying Eqs. 2.27 and 2.35, we get:

$$n_o p_o = N_c N_v \mathcal{F}_{1/2}(\eta_c) \mathcal{F}_{1/2}(\eta_v) \quad (2.37)$$

This equation is sketched in Fig. 2.17 as a function of the Fermi level position with respect to the band structure. Interestingly, for a broad range of Fermi level positions inside the bandgap, $n_o p_o$ is constant. This makes sense. As Fig. 2.14 illustrates, as E_F moves through the bandgap, n_o increases at the same rate at which p_o decreases. The product of these two parameters is then constant independent of the location of the Fermi level. It is easy to find an expression for this constant value by noting that in this intermediate region, both electrons and holes are

³In spite of their importance, there is still some degree of controversy about the numerical values of N_c and N_v in Si and GaAs. Different books quote different numbers. For these and other physical parameters, the values collected in this book are the most accurate in the judgment of the author.

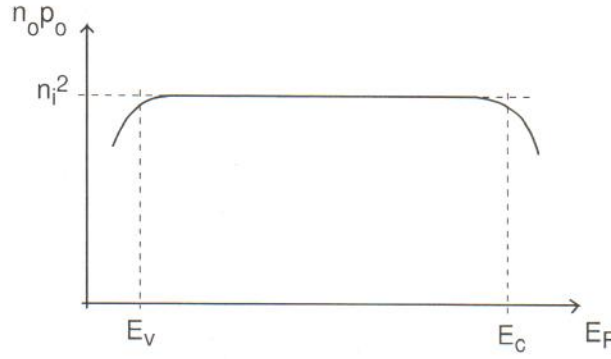


Figure 2.17: Sketch of $n_o p_o$ as a function of the position of the Fermi level with respect to the band structure.

non-degenerate and Maxwell-Boltzmann statistics apply for both carriers (see Fig. 2.16). Using 2.30 and 2.36 we get:

$$n_o p_o \simeq N_c N_v \exp\left(-\frac{E_g}{kT}\right) \quad (2.38)$$

This equation is of an identical form to that of Eq. 2.6 derived using the chemical reaction arguments.

E_g , N_c , and N_v are physical constants characteristic of each semiconductor that depend only on temperature. Eq. 2.38 then indeed confirms our intuitive understanding that for a given semiconductor at a given temperature, the equilibrium np product is a constant. Our rigorous treatment has shown, however, that this is only the case when neither the electron nor the hole distributions are degenerate. In other words, $n_o p_o$ is a constant in the "dilute" regime where the electron and hole concentrations are not too high. We will come back to this at the end of this section.

In an *intrinsic* semiconductor, Eq. 2.38 is very likely to apply under a broad temperature range since the carrier concentrations tend to be relatively low. Hence, the intrinsic carrier concentration is fairly accurately given by:

$$n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2kT}\right) \quad (2.39)$$

This equation is consistent with our intuitive understanding of the meaning of n_i , that is, the concentration of electrons and holes that a perfectly pure piece of semiconductor has in equilibrium at a certain temperature. In addition to the exponential dependence of n_i on E_g , already discussed in the previous section, we also see in Eq. 2.39 that n_i is proportional to N_c and N_v . This makes sense since the higher the effective density of states, the more states there are closer to the band edges and the higher the resulting carrier concentrations at a certain temperature.

Using Eq. 2.39 is only valid when both electrons and holes can be described by Maxwell-

Boltzmann statistics. This is a very good assumption for most intrinsic semiconductors. Only for semiconductors with very small bandgaps at relatively high temperatures might we need to be concerned with the accuracy of this expression. Problem 2.3 explores this issue.

Exercise 2.4: Compute n_i for Si at room temperature.

Appendix B lists all the needed parameters for Si at room temperature. Since E_g , k , and T all appear inside an exponential, it is important to have at least three significant digits in each of these parameters in order to have acceptable accuracy in the calculation of n_i . Using Eq. 2.39:

$$n_i = \sqrt{2.86 \times 10^{19} \text{ cm}^{-3} \times 3.10 \times 10^{19} \text{ cm}^{-3}} \exp\left(-\frac{1.124 \text{ eV}}{2 \times 25.9 \times 10^{-3} \text{ eV}}\right) = 1.12 \times 10^{10} \text{ cm}^{-3}$$

This value is fairly close to the conventionally accepted value of $1.07 \times 10^{10} \text{ cm}^{-3}$ listed in Appendix B. For "back-of-the-envelope" calculations, a value of 10^{10} cm^{-3} is adequate.

Coming back to Eq. 2.38, its validity does not extend to the degenerate regime (whether n- or p-type). It is reasonable that this be the case. As discussed above, when there are relatively few electrons and holes in a semiconductor, the energy required to form an extra electron-hole pair is E_g . If, say, the equilibrium electron concentration is so high that the Fermi level has penetrated inside the conduction band, the energy required to form an electron-hole pair is larger than E_g . This is because the bottom of the conduction band is full and empty states where electrons can be placed are only available at higher energies inside the band. As a result, the np product in equilibrium is reduced for high electron concentrations. The same happens for high hole concentrations.

An important consequence of this, is that Eqs. 2.13 and 2.15 do not apply in the degenerate regime as in this case $n_o p_o < n_i^2$. In this instance, the computation of the minority carrier concentrations are a bit more elaborate since the proper Fermi-Dirac statistics must be used for the majority carriers. Exercise 2.5 below illustrate how this can be done.

Exercise 2.5: Estimate the equilibrium hole concentration for the two cases of Exercise 2.3.

In case a), $n_o = 10^{17} \text{ cm}^{-3}$, which is a non-degenerate situation. Then:

$$p_o = \frac{n_i^2}{n_o} = \frac{1.1 \times 10^{20}}{10^{17}} = 1.1 \times 10^3 \text{ cm}^{-3}$$

In case b), $n_o = 10^{20} \text{ cm}^{-3}$. This is a degenerate situation and we can not use $n_o p_o = n_i^2$. In Exercise 2.3 we computed $E_F - E_c = 63 \text{ meV}$ for this case. This suggests that:

$$E_F - E_v = E_F - E_c + E_c - E_v = E_F - E_c + E_g = 0.063 + 1.124 = 1.187 \text{ eV}$$

Since the holes are non degenerate, we can compute p_o in the following way:

$$p_o = N_v \exp \frac{E_v - E_F}{kT} = 3.1 \times 10^{19} \exp \frac{-1.187}{0.0259} = 0.39 \text{ cm}^{-3}$$

Had we used simple non-degenerate statistics, as in case a), the result would have been 1.1 cm^{-3} nearly a factor of three off.

In this text, however, we do not place a lot of emphasis on the computation of minority carrier concentrations that accounts for the need for Fermi-Dirac statistics for the majority carriers

because, in reality, the situation for semiconductors with high carrier concentrations is all together more complex. In a heavily-doped semiconductor, a number of special "heavy-doping effects" actually occur. They make the computation of carrier concentrations even more complicated. This is discussed in detail in Advanced Topic AT2.6. Hence, in the main body of this book, Eqs. 2.13 and 2.15 will be used regardless of the doping level. Appropriate mention will be made in cases where the errors might be significant and the specialized treatment of Advanced Topic AT2.6 is required.

2.4.5 Location of Fermi level

In a semiconductor in thermal equilibrium, the location of the Fermi level with respect to the band edges completely characterizes the carrier concentrations. Hence, indicating the location of the Fermi level in an energy band diagram is a simple way to represent electron and hole concentrations. Let us distinguish between intrinsic and extrinsic situations.

Intrinsic semiconductor

It is easy to find the location of the Fermi energy in an intrinsic semiconductor. Assuming that Maxwell-Boltzmann statistics apply for both electrons and holes, n_o and p_o obey the simple relationship given by Eqs. 2.30 and 2.36. Equating them and solving for the Fermi level we find:

$$E_i = \frac{E_c + E_v}{2} + kT \ln \sqrt{\frac{N_v}{N_c}} \quad (2.40)$$

where the Fermi level in this intrinsic situation has been relabeled E_i , or *intrinsic Fermi level*.

The first term in Eq. 2.40 is precisely the middle of the bandgap. The second term adds a correction to it. The magnitude of the correction depends on the degree to which N_c and N_v differ from each other. For a semiconductor in which N_c was precisely identical to N_v , then E_i would sit exactly at the middle of the forbidden gap. This makes sense as this is the way to ensure that there are identical numbers of electrons and holes in equilibrium. If N_v is bigger than N_c , as is the case for Si, then equality of electron and hole concentrations forces E_i to be above the middle of the gap to compensate for the fact that there are more states at the top of the valence band than at the bottom of the conduction band (see exercise below). The logarithm in the second term of Eq. 2.40 makes this correction in general very small and graphically, E_i is nearly always close to the middle of the gap, as sketched in Fig. 2.18.

Exercise 2.6: Compute E_i in Si at 300 K.

Using Eq. 2.40:

$$kT \ln \sqrt{\frac{N_v}{N_c}} = 25.9 \text{ meV} \ln \sqrt{\frac{3.1 \times 10^{19} \text{ cm}^{-3}}{2.9 \times 10^{19} \text{ cm}^{-3}}} = 1 \text{ meV}$$

E_i is just 1 meV above the middle of the gap in Si at room temperature.

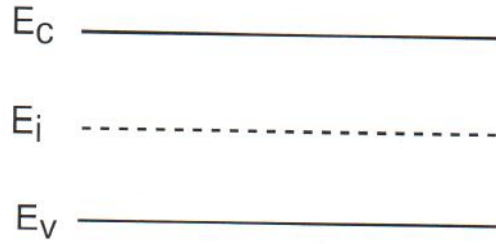


Figure 2.18: Sketch of the position of the Fermi level in an intrinsic semiconductor.

Extrinsic semiconductor

For an n-type non-degenerate semiconductor with all dopants ionized, Maxwell-Boltzmann statistics apply and we can either use Eqs. 2.12 and 2.30 or Eqs. 2.13, 2.36 and 2.39. In both cases we get:

$$E_F - E_c = kT \ln \frac{N_D}{N_c} \quad (2.41)$$

If needed, it is straightforward to refer E_F to the valence band edge. The logarithmic relationship between $E_F - E_c$ and N_D in Eq. 2.41 is a result of the fact that in a moderately doped n-type semiconductor, the electrons are non-degenerate and as a consequence they are well described by Maxwell-Boltzmann statistics.

Similarly for a non-degenerate p-type semiconductor:

$$E_F - E_v = kT \ln \frac{N_v}{N_A} \quad (2.42)$$

which can easily be referred to the edge of the conduction band if required. These equations can also be used for partially compensated cases in which one type of doping overwhelms the other by substituting N_d for N_D , and N_a for N_A (see Problem 2.9).

Fig. 2.19 sketches the position of the Fermi level as a function of N_D (plotted on a logarithmic horizontal scale). As N_D increases, E_F approaches the conduction band edge in a logarithmic manner following Eq. 2.41. For low doping levels, the Fermi level eventually converges to E_i as the semiconductor becomes intrinsic and Eq. 2.40 has to be used. For high doping levels, the semiconductor becomes degenerate and Eq. 2.41 also fails. To compute E_F in this case, one has to use the exact expression that accounts for Fermi-Dirac statistics, as was done in Exercise 2.3. In spite of these limitations, for n-type Si at around room temperature Eq. 2.41 (and also Eq. 2.42 for p-type Si) is fairly accurate over seven orders of magnitude of doping concentration (from $\sim 10^{11}$ to $\sim 10^{18} \text{ cm}^{-3}$).

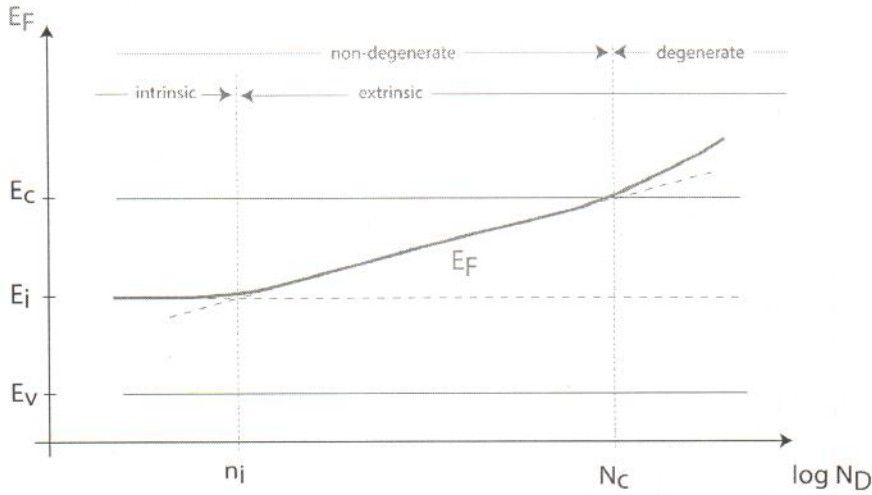


Figure 2.19: Sketch of the position of the Fermi level in an n-type semiconductor as a function of the doping level. Note abscissa is in a logarithmic scale.

Exercise 2.7: Compute the equilibrium carrier concentrations and the position of the Fermi level with respect to the conduction band edge for a uniformly doped Si region with $N_D = 1 \times 10^{16} \text{ cm}^{-3}$ and $N_A = 1 \times 10^{17} \text{ cm}^{-3}$ at room temperature.

This is a case of partial compensation. Since $N_A \gg N_D$ we can define a net acceptor concentration $N_a = N_A - N_D = 9 \times 10^{16} \text{ cm}^{-3}$. Using N_a instead of N_A in Eq. 2.14 we find that

$$p_o \simeq 9 \times 10^{16} \text{ cm}^{-3}$$

Since $p_o \ll N_v$, Eq. 2.15 can be used to compute the equilibrium electron concentration:

$$n_o \simeq \frac{(1.07 \times 10^{10} \text{ cm}^{-3})^2}{9 \times 10^{16} \text{ cm}^{-3}} = 1.3 \times 10^3 \text{ cm}^{-3}$$

The position of the Fermi level with respect to the top of the valence band is obtained from Eq. 2.42:

$$E_F - E_v = 0.0259 \text{ eV} \ln \frac{3.1 \times 10^{19} \text{ cm}^{-3}}{9 \times 10^{16} \text{ cm}^{-3}} = 0.15 \text{ eV}$$

The position of E_F with respect to E_c is simply obtained by subtracting E_g :

$$E_F - E_c = E_F - E_v - E_g = 0.151 - 1.124 = -0.97 \text{ eV}$$

Since the value of n_o has already been calculated, we could have also used Eq. 2.30 directly:

$$E_F - E_c = 0.0259 \text{ eV} \ln \frac{1.3 \times 10^3 \text{ cm}^{-3}}{2.9 \times 10^{19} \text{ cm}^{-3}} = -0.97 \text{ eV}$$

The high doping regime in which the semiconductor is degenerate or nearly degenerate is of very high practical importance. Degenerately doped n- and p-type regions are pervasive in

modern Si and GaAs microelectronic devices. It is not possible, for example, to correctly predict the current gain of a bipolar transistor without adequately dealing with heavy-doping effects. In spite of the need to use Fermi-Dirac statistics for the majority carriers, modeling heavily-doped regions is relatively simple at room temperature for many semiconductors. Advanced Topic AT2.6 discusses the idiosyncrasies of the high-doping regime and derives expressions for the equilibrium carrier concentrations and the position of the Fermi level in the band structure. As a simple rule, for Si at room temperature, if the carrier concentration exceeds about 10^{19} cm^{-3} , heavy doping effects are important.

2.5 Summary

- In a semiconductor, the equilibrium carrier concentrations are related to the location of the Fermi energy with respect to the band structure through:

$$n_o = N_c \mathcal{F}_{1/2}\left(\frac{E_F - E_c}{kT}\right)$$

$$p_o = N_v \mathcal{F}_{1/2}\left(\frac{E_v - E_F}{kT}\right)$$

- For the case in which the Fermi level does not get too close to either band edge (non-degenerate regime), the equilibrium carrier concentrations are related to the Fermi level location through simple exponential relations:

$$n_o \simeq N_c \exp \frac{E_F - E_c}{kT}$$

$$p_o \simeq N_v \exp \frac{E_v - E_F}{kT}$$

- For a given semiconductor, in the non-degenerate regime, the $n_o p_o$ product depends only on temperature:

$$n_o p_o \simeq N_c N_v \exp\left(-\frac{E_g}{kT}\right) = n_i^2$$

- Dopants preferentially introduce electrons or holes into a semiconductor.
- At around room temperature, under most practical circumstances, the majority carrier concentrations are set by the doping level and are independent of temperature. For an n-type semiconductor:

$$n_o \simeq N_D$$

For a p-type semiconductor:

$$p_o \simeq N_A$$

- At low temperatures, carrier freeze-out occurs. At high temperatures, intrinsic production of electron-hole pairs eventually overwhelms the doping level.

2.6 Further reading

There are several excellent books that discuss in considerable detail the statistics of electrons and holes in equilibrium.

Semiconductor Statistics by J. S. Blakemore, Dover, 1978. (ISBN 0-486-65362-5, QC611.B52). Ch. 2 and 3 of this book contain a rather comprehensive discussion of the statistics of electrons and holes in semiconductors in equilibrium in a variety of circumstances: intrinsic semiconductor, degenerate intrinsic semiconductors, as well as semiconductors with single and multiple impurity states. This text also has an appendix that describes several properties of the Fermi-Dirac integrals, presents analytical approximations, and includes tables of values for selected integrals.

Solid State and Semiconductor Physics by J. P. McKelvey, Krieger, 1986. (ISBN 0-89874-396-6, QC611.M495). Ch. 9 of this classic book contains a few sections with a treatment of the statistics of electrons and holes that is similar to the one presented here. This book is characterized by a good compromise between mathematical complexity and physical intuition.

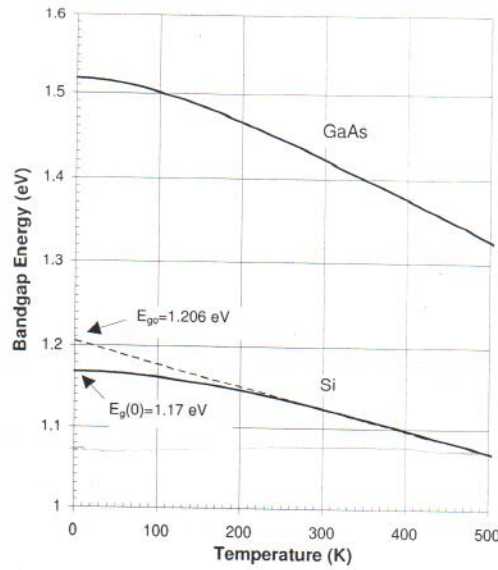


Figure 2.20: Bandgap energy vs. temperature for Si and GaAs.

AT2.1 Temperature dependence of the bandgap

The bandgap of a semiconductor decreases with temperature. This is a consequence of the expansion of the lattice constant as the temperature increases. This is shown for Si and GaAs in Fig. 2.20. The dependence is accurately captured by an equation of the form:

$$E_g(T) = E_g(0) - \frac{\alpha T^2}{T + \beta} \quad (2.43)$$

where $E_g(0)$ is the bandgap at 0 K and α and β are constants. The value of these constants for Si and GaAs is given in Appendix E.

The dependence of E_g with temperature is not very strong. It is important to consider it when computing expressions that contain the bandgap inside an exponential, as the next exercise shows.

Exercise 2.8: Compute n_i for Si at 600 °C.

The easiest way to compute n_i at any temperature is to use its value at 300 K since this is usually relatively well known. First, let's make explicit all temperature dependences of n_i . From Eq. 2.39, 2.26, and 2.34, we can rewrite n_i as:

$$n_i = CT^{3/2} \exp\left[-\frac{E_g(T)}{2kT}\right]$$

where C is a constant independent of temperature.

The ratio of n_i at two different temperatures is then:

$$\frac{n_i(T_1)}{n_i(T_2)} = \left(\frac{T_1}{T_2}\right)^{3/2} \exp\left\{-\frac{1}{2k}\left[\frac{E_g(T_1)}{T_1} - \frac{E_g(T_2)}{T_2}\right]\right\}$$

We can now plug in numbers for $T_1 = 873 \text{ K}$ (600 °C) and $T_2 = 300 \text{ K}$:

$$n_i(873 \text{ K}) = 1.07 \times 10^{10} \times \left(\frac{873}{300}\right)^{3/2} \exp\left[-\frac{1}{2 \times 8.62 \times 10^{-5}}\left(\frac{0.950}{873} - \frac{1.124}{300}\right)\right] = 2.6 \times 10^{17} \text{ cm}^{-3}$$

This is an interesting result. It shows that a Si region doped at the 10^{17} cm^{-3} level is becoming intrinsic at around 600 °C. It is easy to see that not accounting for the temperature dependence of E_g would have resulted in an error of about a factor of three.

Above 250 K, the bandgap of Si can be described to within experimental accuracy (about 1 meV) by the simpler expression:

$$E_g(T) = E_{go} - CT \quad (2.44)$$

where E_{go} is the *extrapolation* of E_g to 0 K and C is a constant (see Fig. 2.20). For Si, $E_{go} = 1.206 \text{ eV}$ and $C = 2.73 \times 10^{-4} \text{ eV} \cdot \text{K}^{-1}$. This is widely used around room temperature.

It is important not to confuse E_{go} with $E_g(0)$, the true bandgap at 0 K. Interestingly, E_{go} emerges frequently when carrying out experimental temperature-dependent studies of transport processes around room temperature. The reason for this is that the factor $\exp(-E_g/kT)$ that is pervasive in these types of processes has an activation energy around room temperature that is precisely E_{go} . This can be easily seen by using Eq. 2.44:

$$\exp\left(-\frac{E_g}{kT}\right) \simeq \exp\left(\frac{C}{k}\right) \cdot \exp\left(-\frac{E_{go}}{kT}\right) \quad (2.45)$$

A good example of this is the data for n_i in Si of Fig. 2.5 which, in fact, shows an activation energy very close to $E_{go}/2$. E_{go} frequently appears in the context of bandgap reference circuits.

AT2.2 Selected properties of the Fermi-Dirac integral

The Fermi-Dirac integral of order 1/2 belongs to a family of integrals called the *Fermi-Dirac integrals* that are defined as:

$$\mathcal{F}_j(x) = \frac{1}{\Gamma(j+1)} \int_0^\infty \frac{\eta^j}{1 + e^{\eta-x}} d\eta \quad (2.46)$$

The following relationship ties up all these integrals:

$$\frac{d}{dx} \mathcal{F}_j(x) = \mathcal{F}_{j-1}(x) \quad (2.47)$$

Other general properties of this class of integrals are described in several books (see, for example Blakemore's text cited above). Here we are almost exclusively interested in the Fermi-Dirac integral of order 1/2 for which:

$$\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2} \quad (2.48)$$

$\mathcal{F}_{1/2}(x)$ has been tabulated for $-4 \leq x \leq 10$ by Blakemore. Useful approximations are:

$$\mathcal{F}_{1/2}(x) \simeq e^x \quad \text{for } x \ll -1 \quad (2.49)$$

$$\mathcal{F}_{1/2}(x) \simeq \frac{4}{3\sqrt{\pi}} x^{3/2} \quad \text{for } x \gg 1 \quad (2.50)$$

These are particular expressions of the more general relationships:

$$\mathcal{F}_j(x) \simeq e^x \quad \text{for } x \ll -1 \quad (2.51)$$

$$\mathcal{F}_j(x) \simeq \frac{x^{j+1}}{(j+1)\Gamma(j+1)} \quad \text{for } x \gg 1 \quad (2.52)$$

Many numerical approximations have been developed for the Fermi-Dirac integrals. For the interested reader, a review of several of them was presented by J. S. Blakemore in *Solid-State Electronics* **25**, p. 1067 (1982). The following expression, for example, has an error smaller than 0.4 % over a range $-10 \leq x \leq +25$, which is sufficient for most applications:

$$\mathcal{F}_{1/2}(x) \simeq \frac{1}{e^{-x} + \xi(x)} \quad (2.53)$$

$$\xi(x) = \frac{3\sqrt{\pi}}{4\nu^{3/8}(x)} \quad (2.54)$$

$$\nu(x) = x^4 + 50 + 33.6x \{1 - 0.68 \exp[-0.17(x+1)^2]\} \quad (2.55)$$

When the carrier concentration is known and the position of the Fermi energy is to be calculated, the inverse function is needed. An approximation that has an error smaller than 1 % all the way up to $x = 20$ is:

$$u = \mathcal{F}_{1/2}(x) \quad (2.56)$$

$$x \simeq \ln u + u [64 + 0.05524 u(64 + \sqrt{u})]^{-1/4} \quad (2.57)$$

* AT2.3 Approximations for strongly degenerate semiconductor

In a degenerate semiconductor, the Fermi level is located inside one of the bands. In these circumstances, the Maxwell-Boltzmann approximation to the Fermi-Dirac integral leads to a very large error (see Fig. 2.12). A better analytical approximation is given in Advanced Topic AT2.2 in Eq. 2.50 that is valid for a sufficiently degenerate semiconductor. Using this approximation, for an n-type semiconductor we can write:

$$n_o \simeq \frac{4}{3\sqrt{\pi}} N_c \left(\frac{E_F - E_c}{kT} \right)^{3/2} \quad (2.58)$$

This equation is acceptable if $\eta_c \gg 1$, $E_F - E_c \gg kT$ or $n_o \gg N_c$.

Similarly, for a sufficiently degenerate p-type semiconductor, we can write:

$$p_o \simeq \frac{4}{3\sqrt{\pi}} N_v \left(\frac{E_v - E_F}{kT} \right)^{3/2} \quad (2.59)$$

valid if $\eta_v \gg 1$, $E_v - E_F \gg kT$, or $p_o \gg N_v$.

In many circumstances, the intermediate doping regime is of interest. For this, more accurate approximations are needed, such as those given in Advanced Topic AT2.2.

Notice that in Eqs. 2.58 and 2.59, the temperature dependence of N_c and N_v cancels out the $T^{-3/2}$ inside the brackets. The relationship between the majority carrier concentrations and the Fermi level is temperature independent. In fact, these equations can be easily obtained using the Fermi-Dirac distribution function in Eq. 2.21 at 0 K (see Problem 2.15 at end of Chapter).

AT2.4 Statistics of donor and acceptor ionization

A donor atom in a semiconductor can either be ionized or neutral. The donor is ionized when it has released its electron to the conduction band. It is neutral if it still holds on to it. Defining N_D as the total donor concentration, N_D^+ as the ionized donor concentration, and N_D^0 as the neutral donor concentration, mass conservation demands that:

semiconductor	dopant	E_d (meV)	β_d	E_a (meV)	β_a	N_{Mott} (cm^{-3})
Si	As	54	2	-	-	6.4×10^{18}
	P	45	2	-	-	3.5×10^{18}
	Sb	39	2	-	-	3.0×10^{18}
	B	-	-	45	4	$\sim 4 \times 10^{18}$
GaAs	C	5.9	2	26	4	$\sim 1 \times 10^{17}$
	Si	5.8	2	35	4	$\sim 2 \times 10^{17}$
	Be	-	-	28	4	$\sim 2 \times 10^{18}$
	Zn	-	-	31	4	$\sim 3 \times 10^{18}$

Table 2.1: Ionization energy, impurity degeneracy factor, and Mott transition of common dopants in Si and GaAs. The Mott concentration is not very well known in all cases.

$$N_D = N_D^+ + N_D^o \quad (2.60)$$

The relative number of neutral and ionized donors is in general a function of the donor ionization energy, the location of the Fermi level and temperature. If we define $f_d(E_D)$ as the probability that the donor energy level is occupied by an electron, the neutral donor concentration can be written as:

$$N_D^o = N_D f_d(E_D) \quad (2.61)$$

One might think that the occupation probability of the donor atom is equal to the Fermi-Dirac distribution function. However, such an approximation does not account for the fact that the fifth electron can become bound to the donor atom in more than one way. This is called *impurity level degeneracy*. When this happens, statistical mechanics tells us that the probability of occupation of the donor state becomes:

$$f_d(E_D) = \frac{1}{1 + \frac{1}{\beta_d} \exp \frac{E_D - E_F}{kT}} \quad (2.62)$$

where β_d is the *donor degeneracy factor* which is greater than or equal to unity.⁴ β_d depends on the host semiconductor and the type of dopant (p or n), but within a given type, it is independent of the specific dopant itself. Table 2.1 summarizes donor ionization energies for typical donors in Si and GaAs.

Using Eq. 2.62 in Eq. 2.61, the neutral donor concentration is:

$$N_D^o = \frac{N_D}{1 + \frac{1}{\beta_d} \exp \frac{E_D - E_F}{kT}} \quad (2.63)$$

⁴The technical literature is ambiguous in the definition of impurity level degeneracy. Sometimes, β_d is defined so as to be smaller than or equal to one. In this case, β_d appears in place of the $1/\beta_d$ factor in Eq. 2.62. This is a point where one must be careful.

The ionized donor concentration is, from 2.60:

$$N_D^+ = N_D - N_D^o = \frac{N_D}{1 + \beta_d \exp \frac{E_F - E_D}{kT}} \quad (2.64)$$

Similar arguments apply for acceptors. When an acceptor captures an electron, it becomes negatively charged. Defining the neutral acceptor concentration as N_A^o and the ionized acceptor concentration as N_A^- , mass conservation again implies:

$$N_A = N_A^- + N_A^o \quad (2.65)$$

In this case:

$$N_A^- = N_A f_a(E_A) \quad (2.66)$$

where $f_a(E_A)$ is the occupation probability of the acceptor state. Acceptor states are also typically degenerate. As a result, $f_a(E_A)$ is given by:

$$f_a(E_A) = \frac{1}{1 + \beta_a \exp \frac{E_A - E_F}{kT}} \quad (2.67)$$

where β_a is the *acceptor degeneracy factor* ($\beta_a \geq 1$). Values of β_a for several impurities in Si and GaAs are tabulated in Table 2.1.

In a similar way to the case of donors, using Eq. 2.67, the ionized and neutral acceptor concentrations are respectively given by:

$$N_A^- = \frac{N_A}{1 + \beta_a \exp \frac{E_A - E_F}{kT}} \quad (2.68)$$

$$N_A^o = \frac{N_A}{1 + \frac{1}{\beta_a} \exp \frac{E_F - E_A}{kT}} \quad (2.69)$$

In order to compute the actual ionization ratio of dopants, one must know the location of the Fermi energy. This has to be calculated in a self-consistent way. Advanced Topic AT2.5 shows how to do this in a specific case.

AT2.5 Carrier freeze-out

The energy required to ionize a dopant is small but finite. At low enough temperatures, the small thermal energy available precludes many dopants from becoming ionized. As a consequence, the carrier concentration drops with temperature. This is called carrier freeze-out.

The results obtained in Advanced Topic AT2.4 allow us to compute the carrier concentration as a function of temperature in the bulk of a semiconductor under charge neutrality conditions. Let us do it for an n-type semiconductor doped with a concentration N_D of donors (a similar approach can be used with a p-type semiconductor). The electron concentration in the bulk is equal to the ionized donor concentration:

$$n_o \simeq N_D^+ \quad (2.70)$$

In writing this equation, we have neglected the presence of a significant amount of electrons that arise from the natural break up of crystal bonds. This is because the energy required for this is E_g , many times larger than the typical donor ionization energy. As the temperature drops, bond break up freezes out much more quickly than donor ionization.

If we can assume Maxwell-Boltzmann statistics for the electrons in the conduction band, Eq. 2.64 can be rewritten as:

$$N_D^+ = \frac{N_D}{1 + \frac{\beta_d n_o}{N_c} \exp \frac{E_d}{kT}} \quad (2.71)$$

Plugging into Eq. 2.70, one obtains a quadratic equation in n_o that can easily be solved:

$$n_o = \frac{N_c}{2\beta_d} \exp\left(-\frac{E_d}{kT}\right) \left(\sqrt{1 + 4\beta_d \frac{N_D}{N_c} \exp \frac{E_d}{kT}} - 1 \right) \quad (2.72)$$

In the limit of high temperature, $kT \gg E_d$, the second term inside the square root is small and a simple Taylor series expansion leads to the result $n_o \simeq N_D$ already derived above.

For low enough temperatures, $kT \ll E_d$, and the second term inside the square root overwhelms the other two inside the brackets. We then obtain:

$$n_o \simeq \sqrt{\frac{N_D N_c}{\beta_d}} \exp\left(-\frac{E_d}{2kT}\right) \quad (2.73)$$

This result shows that the electron concentration at low temperatures is thermally activated with an activation energy of half of the donor ionization energy. This was plotted in Fig. 2.9 above.

In many situations, it is sufficient to compute the temperature at which freeze-out effects become significant. A good working criteria is when the two terms inside the square root in Eq. 2.72 are equal, that is:

$$T_{fo} \simeq \frac{E_d}{k \ln \frac{N_c}{4\beta_d N_D}} \quad (2.74)$$

Exercise 2.9: Estimate the temperature at which freeze-out effects become significant for Phosphorus-doped Si with $N_D = 5 \times 10^{17} \text{ cm}^{-3}$. Verify all your assumptions.

There is a small difficulty here because in order to use Eq. 2.74, we need to calculate N_c at T_{fo} , before T_{fo} is known. This situation can be resolved by writing a small computer program or using a calculator in an iterative way. In either case, N_c at T_{fo} is easily calculated starting from its room temperature value:

$$N_c(T_{fo}) = N_c(300 \text{ K}) \left(\frac{T_{fo}}{300} \right)^{3/2} = 2.86 \times 10^{19} \left(\frac{T_{fo}}{300} \right)^{3/2} \text{ cm}^{-3}$$

We now use Eq. 2.74:

$$T_{fo} \simeq \frac{0.045}{8.62 \times 10^{-5} \ln \frac{N_c(T_{fo})}{4 \times 2 \times 5 \times 10^{17}}} \text{ K}$$

Solving this system of two equations, we find $T_{fo} = 280 \text{ K}$, and $N_c(T_{fo}) = 2.6 \times 10^{19} \text{ cm}^{-3}$. Since at 280 K , $N_D \ll N_c$ and $n_o < N_D$, the assumption of Maxwell-Boltzmann statistics that is built into Eq. 2.74 is valid in this example. In order to verify the negligible impact of the hole concentration in Eq. 2.70, we must calculate n_o and p_o and show that $p_o \ll n_o$. This is done in the next exercise.

Exercise 2.10: Calculate the equilibrium electron and hole concentrations at 77 K in a sample of P-doped Si with $N_D = 5 \times 10^{17} \text{ cm}^{-3}$. Verify all your assumptions.

In the previous exercise we estimated that freeze-out effects are significant for temperatures lower than about 280 K . We can then use Eq. 2.73 to compute n_o at 77 K . Before that, we must first compute $N_c(77 \text{ K})$ which is found to be $3.7 \times 10^{18} \text{ cm}^{-3}$, and $kT(77 \text{ K})$ which is 6.6 meV . Then,

$$n_o = \sqrt{\frac{5 \times 10^{17} \times 3.7 \times 10^{18}}{2}} \exp\left(-\frac{0.045}{2 \times 0.0066}\right) = 3.2 \times 10^{16} \text{ cm}^{-3}.$$

In order to compute p_o , we must first calculate n_i at 77 K . Using the approach described in Advanced Topic AT2.1, we easily find $n_i(77 \text{ K}) = 2.8 \times 10^{-20} \text{ cm}^{-3}$. We then get:

$$p_o = \frac{n_i^2}{n_o} = \frac{7.9 \times 10^{-40}}{3.2 \times 10^{16}} = 2.5 \times 10^{-56} \text{ cm}^{-3}$$

a truly negligible number.

The use of non-degenerate statistics for the electrons is validated since we find that $n_o \ll N_c$ at 77 K . Furthermore, not accounting for the electron contribution from the break up of bonds is validated since we found that $p_o \ll n_o$.

✦ AT2.6 Heavy-doping effects

Heavily-doped regions are pervasive in microelectronic devices. For example, the source and drain of a MOSFET have doping levels of the order of 10^{20} cm^{-3} . Analogously, the emitter of a bipolar transistor has impurity levels of similar magnitude. At high doping levels, defined here as 10^{18} cm^{-3} and above, several special effects occur that deserve particular attention. Some of these phenomena are of relevance for important device figures of merit. For example, if heavy-doping effects are not taken into account, a calculation of the current gain of a bipolar transistor

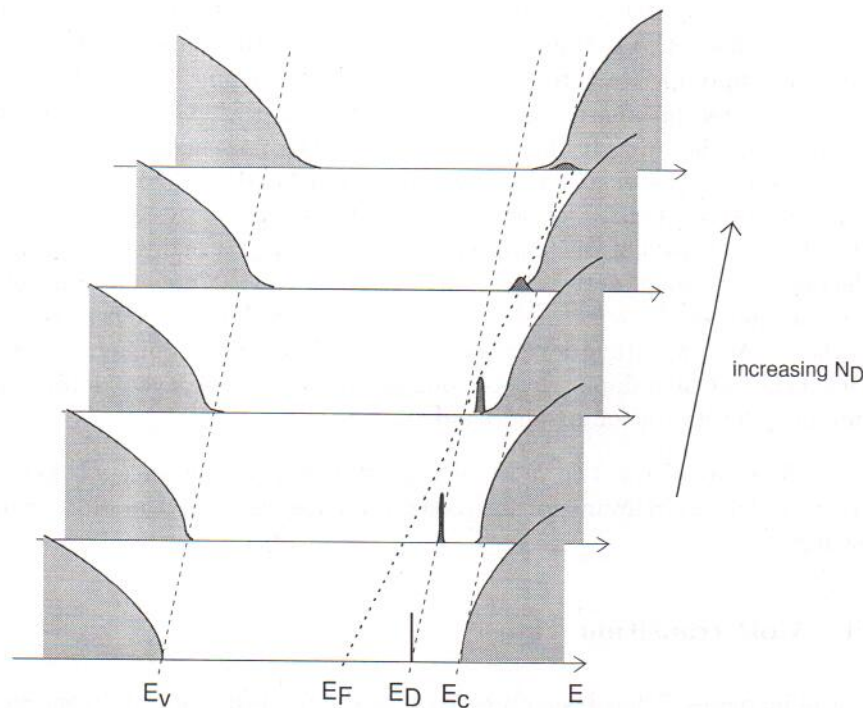


Figure 2.21: A sketch of heavy doping effects in an n-type semiconductor. As N_D increases, the impurity level broadens into an impurity band and merges into the conduction band, the conduction and valence bands develop band tails, and the bandgap narrows.

can easily yield an error exceeding a factor of 10.

Fig. 2.21 illustrates the special effects that occur at high doping levels in an n-type semiconductor:

1. The Fermi level penetrates into the majority-carrier band and the *majority carriers become degenerate*. This demands the use of Fermi-Dirac statistics for the majority carriers as discussed in the main body of this chapter and in Advanced Topic AT2.3.
2. *Bandgap narrowing* takes place, that is, the energy difference between the conduction and valence bands is reduced.
3. The impurity ionization energy drops as the doping level increases and eventually becomes zero. This is called the *Mott transition*.
4. The impurity level broadens into an *impurity band*. This is only relevant at low temperatures and will not be discussed here.
5. The density of states of the conduction and valence bands is deformed from its ideal square-root shape. *Band tailing* appears at the edges of the bands. The extent of these tails is only a few *meV* and as a result, this effect is not very significant for room temperature operation. We will not consider it any further here.

With the exception of majority carrier degeneracy, whose origin we already know, these effects arise from a variety of sources. A detailed discussion is beyond the scope of this book but a few words about them are appropriate here. First of all, at high doping levels, the dopant atoms can no longer be considered far apart from each other, in fact, their Coulombic potentials start to overlap. This reduces the impurity ionization energy. The presence of a high concentration of majority carriers shrinks the energy required to break a bond through *many-body effects*, a quantum-mechanical phenomenon. This results in bandgap narrowing. Many-body effects also contribute to a further reduction of the impurity ionization energy. Finally, at high doping levels, *doping level fluctuations* become very important. The fact is that dopants are not uniformly dispersed inside the bulk of the semiconductor. Some regions have a higher concentration of dopants than others. At low doping levels, these fluctuations in the doping concentration are not very significant, but at high doping levels, coupled with the other heavy doping effects, they contribute to impurity level broadening and band tailing.

We discuss in more detail the two most important heavy-doping effects, besides majority carrier degeneracy, which are relevant for microelectronic devices, *i.e.* the Mott transition and bandgap narrowing.

AT2.6.1 The Mott transition

As the doping level increases, it has been observed that the impurity ionization energy decreases. A simple expression that captures this phenomenon is:

$$E_i = E_{io}[1 - (\frac{N_i}{N_{Mi}})^{1/3}] \quad \text{for } N_i \leq N_{Mi} \quad (2.75)$$

$$E_i = 0 \quad \text{for } N_i \geq N_{Mi} \quad (2.76)$$

where "i" stands for the donor or acceptor subindex, E_{io} is the low-doping value of the ionization energy given in Table 2.1, and N_{Mi} is the Mott concentration, also listed in Table 2.1 (the Mott concentration is not very well known in all cases).

Eq. 2.76 shows that at a critical concentration, called the Mott transition, the ionization energy goes to zero. The implications of this should be immediately apparent. For doping levels higher than the Mott transition, all dopants are completely ionized at all temperatures. Carrier freeze-out does not occur at low temperatures for semiconductors doped beyond the Mott transition. As a result, the semiconductor remains conductive down to 0 K, just like a metal. For this reason, this critical concentration also receives the name of *Metal-Insulator Transition*. This is illustrated in Fig. 2.22 which displays the ratio of the measured electron concentration at 4.2 K over 300 K in Phosphorus-doped Si for several doping levels. Samples with doping levels below about $3.5 \times 10^{18} \text{ cm}^{-3}$ freeze out at low temperatures and $n_o(4.2 \text{ K}) < n_o(300 \text{ K})$. On the other hand, in samples with a doping level above $3.5 \times 10^{18} \text{ cm}^{-3}$ the electron concentration is independent of temperature, that is, $n_o(4.2 \text{ K}) \simeq n_o(300 \text{ K})$ (the small systematic discrepancy between $n_o(4.2 \text{ K})$ and $n_o(300 \text{ K})$ for high doping levels is a feature introduced in the interpretation of the data).

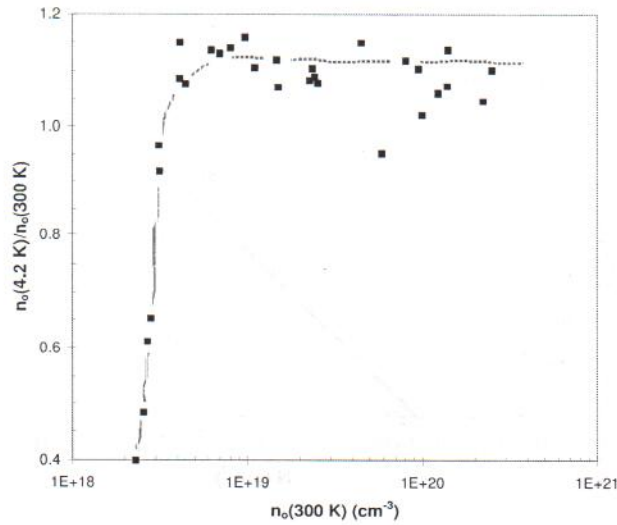


Figure 2.22: Ratio of the measured electron concentration at 4.2 K over 300 K in Phosphorus-doped Si for several doping levels. These data clearly reveal the Mott transition for Si:P at about $3.5 \times 10^{18} \text{ cm}^{-3}$ (the small systematic discrepancy between $n_o(4.2 \text{ K})$ and $n_o(300 \text{ K})$ for high doping levels is a feature introduced in the interpretation of the data) [data from C. Yamanouchi et al., *J. Phys. Soc. Jpn.* 22, 859 (1967)].

The implications of this phenomenon to microelectronic devices are profound. In devices designed to operate at low temperatures, regions doped below the Mott transition will significantly freeze-out (and their resistivity will substantially increase), while those regions that are doped above the Mott transition will not. The Mott transition has a noticeable influence even at room temperature. If it were not for the Mott transition, at high doping levels the Fermi level would cross the impurity level and significant dopant freeze-out would take place at room temperature. You can appreciate this in the next example.

Exercise 2.11: Calculate the equilibrium electron concentration at 300 K of a sample of P-doped Si with $N_D = 2 \times 10^{18} \text{ cm}^{-3}$. Do it with and without consideration of the Mott transition. Verify all your assumptions.

To be expeditious, we will use directly the exact expression 2.72. This equation is valid in both cases because the only assumption built into it is the Maxwell-Boltzmann approximation for the electron statistics. This is insured in this problem, since $N_D \ll N_c$.

Without consideration of the Mott transition, E_d for Phosphorus-doped Si is 0.045 eV. Plugging in Eq. 2.72, this yields:

$$n_o(E_{do} = 0.045 \text{ meV}) = 1.31 \times 10^{18} \text{ cm}^{-3}$$

The dopant ionization ratio is only 66%.

The Mott transition considerably reduces E_d . Using Eq. 2.75, we find that for $N_D = 2 \times 10^{18} \text{ cm}^{-3}$, $E_d = 0.005 \text{ eV}$. Plugging again into Eq. 2.72, we obtain:

$$n_o(E_d = 0.005 \text{ meV}) = 1.75 \times 10^{18} \text{ cm}^{-3}$$

The dopant ionization ratio is now 87%.

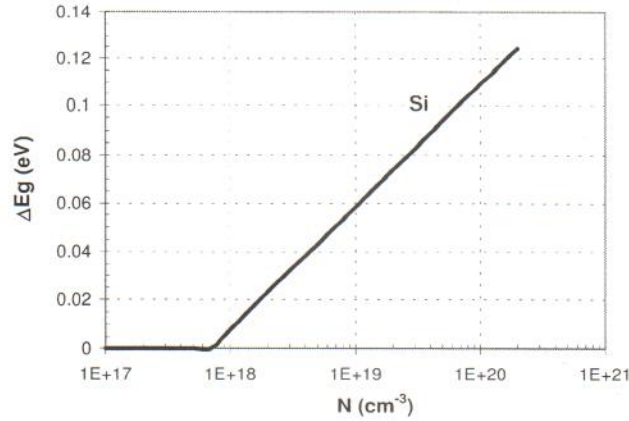


Figure 2.23: Bandgap narrowing vs. doping level in Si.

The dopant concentration at which the Mott transition takes place depends both on the host semiconductor and the dopant species. Values for the Mott transition for typical dopants in Si and GaAs are listed in Table 2.1. Some of these are not very accurately known.

AT2.6.2 Bandgap narrowing

At high doping levels, many body effects that arise from the high majority carrier concentration result in a rigid shrinkage of the bandgap. Experimentally it is found that bandgap narrowing increases with doping level but is otherwise a property of the host semiconductor, that is, it appears to be independent of the dopant species and temperature.

Bandgap narrowing, ΔE_g , is usually defined as a *positive* quantity, that is:

$$E_{g-HD} = E_{g-LD} - \Delta E_g \quad (2.77)$$

where E_{g-LD} is the low-doping value of the bandgap, and E_{g-HD} is the actual bandgap at a certain doping level. Fig. 2.23 shows bandgap narrowing as a function of doping level in Si. An analytical description of this function is given in Appendix E.

Bandgap narrowing does not affect the majority carrier concentration which is set by the doping level only. On the other hand, the minority carrier concentration is increased since the energy required to break a bond and produce an electron-hole pair is reduced. In the heavily-doped regime, this is usually accounted for by defining an *effective intrinsic carrier concentration*, n_{ie} , which in general is different from the low-doping value derived earlier in this chapter. Looking back at Eq. 2.39, a reduction of the bandgap results in an increase in n_i . The exponential dependence of n_i on E_g , makes small changes in the bandgap very significant.

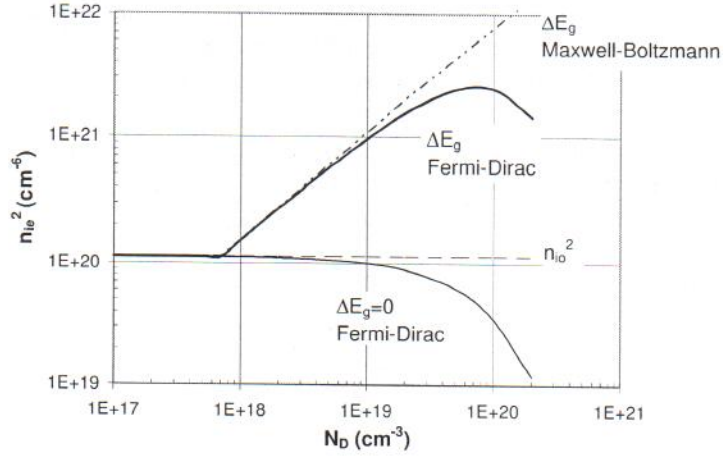


Figure 2.24: Effective intrinsic carrier concentration vs. doping level in the heavy doping regime for n-Si at 300 K. Also shown are the low-doped value n_{io}^2 , and n_{ie}^2 separately due to bandgap narrowing and electron degeneracy.

Unfortunately, we cannot just modify Eq. 2.39 since it was deduced under the assumption that Maxwell-Boltzmann statistics are valid for both types of carriers. In the heavy-doping regime, this is clearly not the case for the majority carriers. A new expression for the $n_o p_o$ product that is valid in the heavy-doping regime is required. The starting point is Eq. 2.37. Multiplying and dividing by $\exp \frac{E_g - LD}{kT}$, we get:

$$n_o p_o = n_{ie}^2 = n_{io}^2 \mathcal{F}_{1/2}(\eta_c) \mathcal{F}_{1/2}(\eta_v) \exp \frac{E_g - LD}{kT} \quad (2.78)$$

where, following conventional practice, we have relabeled the intrinsic carrier concentration in the low-doping regime as n_{io} and we have defined the $n_o p_o$ product in the highly-doped regime as n_{ie}^2 .

Eq. 2.78 can be further simplified if we focus on a specific doping polarity. For an n-type semiconductor, for example, the holes follow Maxwell-Boltzmann statistics. This allows us to use the exponential approximation to the Fermi-Dirac integral for the term in $\mathcal{F}_{1/2}(\eta_v)$. After some straightforward algebra, Eq. 2.78 can be rewritten as:

$$n_{ie}^2 = n_{io}^2 \frac{\mathcal{F}_{1/2}(\eta_c)}{e^{\eta_c}} \exp \frac{\Delta E_g}{kT} \quad (2.79)$$

where we have also used Eq. 2.77. A similar expression can be obtained for a p-type semiconductor.

Fig. 2.24 graphs Eq. 2.79 for n-Si at room temperature as a function of doping level. For low doping levels, n_{ie}^2 converges to n_{io}^2 . Beyond a doping level of around $N_D \simeq 10^{18} \text{ cm}^{-3}$, n_{ie}^2 starts to increase over the low-doped value. At about 10^{20} cm^{-3} , n_{ie}^2 peaks and starts decreasing.

This peculiar behavior of n_{ie}^2 can be understood by examining separately the impact of bandgap narrowing and electron degeneracy on n_{ie}^2 . These are also graphed separately in Fig. 2.24. The increase in bandgap narrowing with doping level makes n_{ie}^2 increase very quickly through the $\exp(\Delta E_g/kT)$ term in Eq. 2.79. But as the doping level increases, the Fermi level penetrates into the conduction band and the electrons become degenerate. This brings n_{ie}^2 down through the term in $\mathcal{F}_{1/2}(\eta_c)/e^{\eta_c}$ which is at most unity and decreases as η_c increases. A physical way to explain this is that as the doping level increases, the bottom of the conduction band fills up with electrons and the energy required to break a bond and produce an electron-hole pair increases. Everything else being equal, this reduces the $n_o p_o$ product.

Exercise 2.12: Calculate the equilibrium electron and hole concentrations at 300 K of a sample of P-doped Si with $N_D = 1 \times 10^{20} \text{ cm}^{-3}$. Do it with and without consideration of bandgap narrowing. Verify all your assumptions.

Since $N_D = 10^{20} \text{ cm}^{-3}$ is beyond the Mott transition for P-doped Si,

$$n_o = N_D = 10^{20} \text{ cm}^{-3}$$

In order to compute p_o , we must first obtain n_{ie}^2 . We could simply use Fig. 2.24. But the resolution of this graph is limited, and besides, it is of value to trace the steps of the computation of n_{ie}^2 .

First, we have to find the position of the Fermi level. This can be easily accomplished through the use of the approximation to the inverse Fermi-Dirac function given in Eqs. 2.56-2.57. This yields:

$$\eta_c = 2.43$$

This means that the Fermi level has penetrated over $2kT$'s into the conduction band.

Armed with η_c we can now compute the term $\mathcal{F}_{1/2}(\eta_c)/e^{\eta_c}$ in Eq. 2.79. Note that there is no need to carry out the Fermi-Dirac integral since $\mathcal{F}_{1/2}(\eta_c) = N_D/N_c = 3.50$. We then get:

$$\frac{\mathcal{F}_{1/2}(\eta_c)}{e^{\eta_c}} = 0.31$$

The degeneracy of the electron gas reduces $p_o n_o$ to 31% of its low-doped value.

To account for bandgap narrowing, we first, obtain the value of ΔE_g that corresponds to $N_D = 10^{20} \text{ cm}^{-3}$ in Fig. 2.23. We find $\Delta E_g = 109 \text{ meV}$. This allows us to compute the term that enters the equation of n_{ie}^2 :

$$\exp \frac{\Delta E_g}{kT} = 67.6$$

Bandgap narrowing alone increases the $p_o n_o$ product by over 67 times!

We now plug in all terms into Eq. 2.79, and get:

$$n_{ie}^2 = 2.4 \times 10^{21} \text{ cm}^{-3}$$

This allows us to obtain the equilibrium hole concentration as:

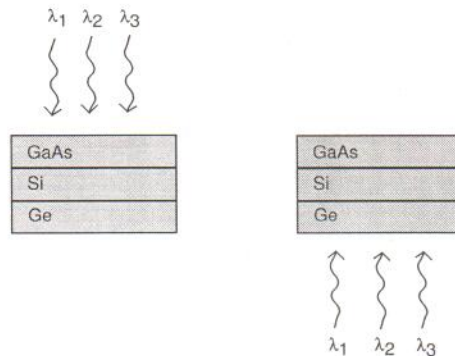
$$p_o = \frac{n_{ie}^2}{n_o} = 24 \text{ cm}^{-3}$$

Without any consideration of heavy doping effects, we would have obtained a value of p_o of 1.1, off by more than an order of magnitude.

Problems

[* denotes a problem that requires studying an Advanced Topic]

- 2.1 Consider a stack of three wafers at room temperature. The top one is GaAs ($E_{g1} = 1.42 \text{ eV}$), the middle one is Si ($E_{g2} = 1.12 \text{ eV}$) and the bottom one is Ge ($E_{g3} = 0.66 \text{ eV}$), as sketched below. To this stack we shine three lasers with wavelengths $\lambda_1 = 0.85 \mu\text{m}$, $\lambda_2 = 1.3 \mu\text{m}$, and $\lambda_3 = 1.55 \mu\text{m}$, once from above and a second time from below. For each illuminating condition, in which wafer is each laser beam absorbed? Explain.



- 2.2 This exercise explores the physical meaning of the effective density of states of a band.

- Assume a fictitious semiconductor with a conduction band density of states in which there are N_c states all located at E_c . Analytically, $g_c(E) = N_c \delta(E - E_c)$, where δ is the Dirac delta function. Calculate an expression for the electron concentration as a function of the relative position of E_F with respect to E_c . Simplify for a non-degenerate situation in which E_F is far below E_c .
 - Let us discuss now the origin of the temperature dependence $T^{3/2}$ deduced in class for N_c and N_v . Consider another fictitious semiconductor with a step-like conduction-band density of states $g_c(E) = A_c u(E - E_c)$, where A_c is a constant and u is the step function. Using Maxwell-Boltzmann statistics, calculate the electron concentration as a function of the position of E_F with respect to E_c . Why does the exponential prefactor (the effective density of states in this case) depend on T ?
 - Based on what you found above, what is the physical meaning of N_c ? Can you speculate on why it is called *effective* density of states? Is N_c the total integrated density of states of the conduction band? Why does it depend on T ? Does it make sense to talk about effective density of states in a degenerate case?
- 2.3 For high temperatures, Maxwell-Boltzmann statistics might not be suitable for an *intrinsic* semiconductor. How can this happen? In Si, which type of carrier, holes or electrons, suffers from this first as the temperature increases? *Estimate* the temperature dependence at which Maxwell-Boltzmann statistics will stop holding for intrinsic Si.
- 2.4 Misiakos and Tsamakis carried out measurements of the intrinsic carrier concentration in Si *vs.* temperature (see Fig. 2.5). The data they obtained is given on the table below:

T (K)	n_i (cm^{-3})
77.8	5.00×10^{-20}
100	2.00×10^{-11}
120.75	3.40×10^{-6}
137.5	4.60×10^{-3}
148.4	2.10×10^{-1}
169.7	8.40×10^1
195	1.78×10^4
199.5	4.30×10^4
213	4.16×10^5
239	1.92×10^7
256.5	1.45×10^8
270.6	6.70×10^8
281	1.79×10^9
300	9.70×10^9
319.5	4.51×10^{10}
340.5	1.89×10^{11}

Graph an Arrhenius plot and extract the activation energy.

Optional: do a least square fit of the data with the following equations:

a) $n_i(T) = K_1 \exp -\frac{K_2}{2kT}$

b) $n_i(T) = K_3 \left(\frac{T}{300}\right)^{3/2} \exp -\frac{K_4}{2kT}$

c) $n_i(T) = K_5 \left(\frac{T}{300}\right)^{K_6} \exp -\frac{K_7}{2kT}$

d) $n_i(T) = K_8 \left(\frac{T}{300}\right)^{K_9} \exp -\frac{E_g(T)}{2kT}$ with $E_g(T) = 1.17 - \frac{K_{10}T^2}{T+K_{11}}$

- 2.5 Consider a p-type Si wafer with an acceptor concentration of $N_A = 10^{17} cm^{-3}$. Compute at room temperature under equilibrium conditions:
- Hole concentration.
 - Electron concentration.
 - Position of the Fermi level with respect to the conduction band and the valence band edges.
 - Probability that a state at the bottom of the conduction band is occupied.
 - Probability that a state at the top of the valence band is empty.
- 2.6 At which temperature will a Si-sample with 10^{17} donors become an intrinsic semiconductor? Is the Maxwell-Boltzmann-approximation still valid at this temperature? If so, estimate at which temperature this approximation breaks down.
- 2.7 The *Fermi wavelength* is the de Broglie-wavelength of electrons located at the Fermi-energy. Estimate the Fermi wavelength of n-type Si with $3 \times 10^{20} cm^{-3}$ electrons. Assume the effective mass of an electron in Si to be the same as the density of states effective mass.
- 2.8 InP is a semiconductor with a bandgap at room temperature of 1.35 eV, an electron density of states effective mass of $0.077 m_o$, and a hole density of states effective mass of $0.64 m_o$. Calculate at room temperature:
- N_c , N_v , n_i and E_i . Verify all your assumptions.

- b) In a chunk of InP that sits in thermal equilibrium at 300 K, the Fermi level is 0.2 eV below the conduction band edge. Calculate n_o and p_o (without invoking $n_o p_o = n_i^2$). Calculate the np product and compare with n_i^2 calculated above. Comment.
- c) In a different chunk of InP also in thermal equilibrium at 300 K, the Fermi level is now 0.1 eV below the valence band edge. Calculate n_o and p_o (without invoking $n_o p_o = n_i^2$). Calculate the np product and compare with n_i^2 calculated above. Comment.

2.9 Consider a compensated semiconductor uniformly doped with a donor concentration N_D and an acceptor concentration N_A , with $N_D > N_A$. Assume that conditions are such that Maxwell-Boltzmann statistics and full impurity ionization apply.

- a) Derive a general expression for the electron and hole concentrations in equilibrium.
- b) Discuss the condition that needs to be satisfied for this semiconductor to be simply considered n-type with a net doping level $N_d \simeq N_D - N_A$.
- c) If $N_D = 10^{16} \text{ cm}^{-3}$ in Si at room temperature, what is the maximum compensation ratio N_A/N_D that allows the semiconductor to be simply considered n-type with $N_d = N_D - N_A$? Make explicit your error acceptance criteria.

2.10 It is possible to estimate the *donor ionization energy* in a semiconductor following a similar methodology to that followed in Problem 1.7. In very elemental terms, a donor can be considered as a Hydrogen atom that is immersed in a semiconductor. The fifth valence electron of the donor atom (the one that does not participate in bonding with neighboring Si atoms) behaves in a way as the electron of the Hydrogen atom. It effectively sees a positive charge of value $+q$ at the nucleus, since all the other protons are compensated by the rest of the electrons of the donor atom. The fundamental difference between a donor and a H atom is that the first is immersed in a solid. We can account for this by using the permittivity of the solid instead of that of vacuum and describe the electron by an effective mass that includes its quantum mechanical interactions with the lattice atoms.

Following the procedure of Problem 1.7, derive an expression for the radius and the binding energy of the fifth valence electron of a donor atom in Si. Use an effective mass for the electron equal to: $m_e^* = 0.28m_o$. Use also the permittivity of Si that is given in Appendix B. Give the result in terms of nm and eV . Comment. Calculate how many Si atoms are included in a sphere with a radius equal to the estimated radius of the fifth electron. Comment.

2.11 Compute and graph the *normalized* electron concentration per unit energy $n_o(E)/n_o$ in Si at room temperature for $E_F - E_c = -6kT, -4kT, -2kT, 0, 2kT, 4kT$, and $6kT$. Discuss the evolution of the electron distribution as E_F penetrates into the conduction band.

2.12 Specify the range of donor concentrations in which each of the three equations below are reasonably accurate for n-type GaAs at room temperature. Assume that all donors are fully ionized and that there are no other dopants. Explain (see suitable parameters for GaAs in Appendix B).

- a) $n_o p_o = n_i^2$
- b) $E_F - E_c = kT \ln \frac{N_D}{N_c}$
- c) $p_o = \frac{n_i^2}{N_D}$

2.13 Derive relationships for n_o and p_o in terms of n_i and the relative position of E_F and E_i under the constraint of Maxwell-Boltzmann statistics.

- 2.14 Following a procedure similar to that of Sec. 2.4.2, derive an expression for the average kinetic energy of all conduction band electrons of a non-degenerate semiconductor at room temperature. Comment on the result.
- 2.15* Derive Equation 2.58 directly from Eqs. 2.20 and 2.21 using the Fermi-Dirac distribution function at $0K$.
- 2.16 Consider an n-type Si sample with $N_D = 10^{16} \text{ cm}^{-3}$ at room temperature in thermal equilibrium. Calculate the position in energy of the peak of the electron concentration inside the conduction band.
- 2.17 Consider an n-type semiconductor with a concentration N_D of donors. Starting from Eqs. 2.8 and 2.9, derive general relationships for n_o and p_o in terms of N_D and n_i that are valid under Maxwell-Boltzmann statistics. Discuss the condition that needs to be satisfied for Eqs. 2.12 and 2.13 to be applicable.
- 2.18 Derive an expression for the transition temperature between the extrinsic and intrinsic regimes in a p-type semiconductor. Discuss the dependencies that are observed. Estimate the temperature at which a p-Si wafer doped with $N_A = 10^{17} \text{ cm}^{-3}$ becomes intrinsic.
- 2.19 In a certain Si sample at room temperature and in thermal equilibrium, the Fermi level is 0.1 eV below the valence band edge.
- Calculate the electron concentration, n_o .
 - Calculate the hole concentration, p_o .
- 2.20 Describe a semiconductor in which Maxwell-Boltzmann statistics cannot be used in thermal equilibrium for *neither electrons nor holes* at room temperature. Explain.
- 2.21 Other than through bandgap narrowing, does the threshold photon energy for carrier generation in a semiconductor depend on doping? Explain.
- 2.22 Fig. 2.9 sketches the electron concentration of a piece of non-degenerate n-type semiconductor in thermal equilibrium as a function of temperature in an Arrhenius plot. In the axes provided below, sketch the evolution of the location of the Fermi level with respect to the band edges as a function of temperature for the entire temperature range. For simplicity, assume that $N_C = N_V$, and both N_C and N_V as well as E_g are all temperature independent. Explain your drawing in each of the temperature regimes.

