

1

Semiconductor fundamentals

In this chapter, we review the basic fundamentals of semiconductors that will be used throughout the text. Only the fundamental issues that we will need to begin our study of semiconductor devices utilized in computing and telecommunications systems are discussed.

Before we begin our study it is useful to point out how semiconductor devices are instrumental in many applications. In this book we will mainly examine the application of semiconductor devices to computing and telecommunications systems. Specifically, we will examine the primary device used in integrated circuits for digital systems, the metal oxide semiconductor field effect transistor, MOSFET. The discussion will focus on state-of-the-art MOSFET devices and future approaches that extend conventional MOSFETs and revolutionary approaches that go well beyond MOSFETs. It is expected that computing hardware will continue to improve, providing faster and more powerful computers in the future using either some or all of the techniques discussed here or perhaps using completely new technologies. In any event, there is almost certainly going to be a large growth in computing hardware in order to maintain the pace of computer development and this book will help introduce the student to emerging technologies that may play a role in future computing platforms.

The second major topic of this book involves discussion of semiconductor devices for telecommunications applications. We will examine devices of use in lightwave communications as well as wireless communications networks. Among these devices are emitters, detectors, amplifiers, and repeaters.

Some mention should be made of the various commercial products that are and will be greatly impacted by semiconductor devices. The development of blue and blue-green light emitting diodes (LEDs) and lasers foment the evolution of new, highly efficient, rugged, ultra-long-life illumination elements. White light emitters using LEDs are now becoming commercially available. These emitters are far more efficient than incandescent bulbs, cost about the same or less, have lifetimes measured in years rather than months, are rugged and durable. It is expected that replacing incandescent lighting by LEDs worldwide can result in a substantial energy savings and potentially reduce consumption of fossil fuels. Perhaps this will lead to a reduction in greenhouse gas emission and help combat global warming and environmental decay in general. Blue lasers enable the development of very small compact discs for data storage, video and audio systems thus greatly expanding the storage capacity of CDs.

New semiconductor materials, such as gallium nitride (GaN) and silicon carbide (SiC), are emerging that are far more tolerant of high temperatures, and operate at significantly higher current densities and frequencies than existing devices. Devices

2 Semiconductor fundamentals

made from these materials are highly attractive for high power, high frequency, and high temperature operation. Specific applications are as power amplifiers for base stations in wireless telecommunications systems, hybrid electric vehicles, switching elements for electric power grids, and high power amplification for radar and satellite communications. Thus GaN and SiC may emerge as important semiconductor materials for many important applications.

1.1 Definition of a semiconductor

The first question one might raise is why are semiconductor materials important in electrical engineering? To answer this question let us first consider a useful characterization scheme for solids based on their electrical properties, specifically their electrical conductivity. Generally, all crystalline solids can be classified into one of four categories. These categories, arranged from highest electrical conductivity to lowest, are metals, semimetals, semiconductors, and insulators. The distinction among these four categories is of course, somewhat vague. For instance, some materials can be either metallic or semimetallic depending upon the form into which they crystallize. Additionally, the distinction between semiconductors and insulators can often become blurred for the wide band gap materials. Nevertheless, we will find it convenient to classify solids into one of these four categories.

Of the four classes of materials, semiconductors are arguably the most important in electrical engineering. The principal reason underlying the importance of semiconductors is the fact that their electrical properties can be readily engineered. Semiconductors are unique in that their conductivity can be significantly altered in several different ways. For the other three types of solids, metals, semimetals, and insulators, their conductivity cannot be readily and significantly altered making them far less attractive for electrical engineering.

There are numerous ways in which the conductivity of a semiconductor can be altered. In this book, we will address most of these approaches and how they can be utilized to make useful semiconductor devices. Before we outline the approaches to manipulating the electrical conductivity of a semiconductor, we should first review what a semiconductor is.

The most commonly used semiconductors are the elemental semiconductors silicon and germanium, and the compound semiconductors, consisting of compound materials. There are numerous compound semiconductors but they are generally formed from two, three, or four different elements and are referred to as binary, ternary, and quaternary compounds respectively. The most important compound semiconductors are based on Column IIIA and Column VA elements in the Periodic Table. For this reason, these compounds are called the III–V compound semiconductors or III–Vs. Examples of the III–V compounds are gallium arsenide (GaAs), indium phosphide (InP), aluminum arsenide (AlAs), indium arsenide (InAs), etc. Notice that in each case the cation is a Column III element while the anion is a Column V element. Ternary compounds can be formed using three elements such as $\text{Al}_x\text{Ga}_{1-x}\text{As}$, where the subscript x represents the mole fraction of aluminum present in the compound. Similarly,

quaternary compounds can be formed in which four elements are combined. An example of a quaternary compound semiconductor is $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$.

How though can we identify which materials are semiconductors? To answer this question we must first consider a fundamental result in the physics of solids. Every crystalline solid has translational symmetry. A system is said to have translational symmetry if it can be broken into a set of identical basic unit cells such that when the system is translated by a distance equal to the length of one unit cell it remains invariant. An obvious example is that of a uniform brick wall. If one translates a row of bricks by a length equal to that of a single brick, the wall looks precisely the same as before. The wall is said to be invariant under a linear translation. A similar situation holds for a crystalline solid. The arrangements of atoms forming a crystalline solid are like the bricks of a uniform wall. The atoms, much like the bricks, are arranged in periodic intervals. Therefore, when the system is translated by a distance equal to the separation between two adjacent atom centers, called the lattice constant, the system remains the same and is said to be invariant. Since the arrangement of the positions of the atoms in a crystalline solid is periodic, the electrostatic potential corresponding to the atoms is also periodic. The potential of the solid is thus also translationally symmetric. The fact that all crystalline solids have a periodic potential is extremely important. There is a fundamental result from quantum mechanics that applies to any system with a periodic potential. This result (Brennan, 1999, Chapter 8) states that for a system with a translationally symmetric potential, the electron energy levels are arranged in bands. These bands can either be conducting or forbidden. As the name implies a conduction band is one in which the electrons can propagate or conduct. Conversely, a forbidden band is one in which no conducting states exist. Electrons cannot be placed into a forbidden band.

In addition to the formation of energy bands, the presence of a periodic potential introduces energy gaps in the allowed energy spectrum. These gaps are called forbidden gaps. Forbidden gaps correspond to energy ranges wherein no allowed electronic states exist. A typical diagram showing a valence band, forbidden energy band and conduction band is shown in Fig. 1.1. As can be seen from the figure, allowed energy states exist only within the conduction and valence energy bands. As mentioned above, electrons within the conduction band can propagate through the crystal and thus carry a current. Electrons cannot be located within the forbidden band. In the valence band, electronic states exist but these states are not free. In other words, electrons within the valence band are localized into bound states that are formed by the molecular bonds between the constituent host atoms of the crystal.

A completely empty band cannot conduct a current. This should be obvious since an empty band has no carriers within it and thus there is nothing to carry the current. A less obvious fact is that a completely filled energy band also cannot conduct a current. This follows from the fact that no two electrons can simultaneously occupy the same quantum state. The general formal statement of this is the Pauli Principle, which applies to the class of particles called fermions, and includes electrons, protons, and neutrons. The Pauli Principle plays a strong role in the formation of atoms. As the reader is aware from fundamental chemistry, each atom in the Periodic Table is formed by

Cambridge University Press

978-0-521-15361-4 - Introduction to Semiconductor Devices: For Computing and Telecommunications Applications

Kevin F. Brennan

Excerpt

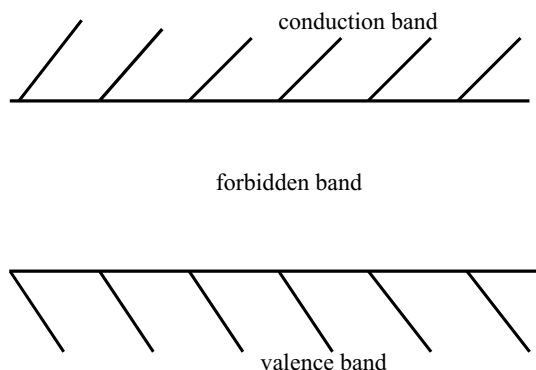
[More information](#)**4** Semiconductor fundamentals

Figure 1.1 Sketch of the conduction, forbidden, and valence bands within a semiconductor. Electrons within the conduction band can freely propagate through the crystal and thus can carry a current. Electrons within the valence band are localized into bound electronic states formed by the molecular bonding of the constituent atoms of the crystal. In the forbidden band, no electronic states exist and thus electrons cannot exist within the forbidden band. The forbidden band is also called the energy gap.

progressively adding an electron and proton (and possibly neutrons) to each previous atom starting with hydrogen. In the case of hydrogen the only charged particles present are one electron and one proton. The electron is placed into the lowest lying energy state of the atom. The next element is helium which comprises two electrons and two protons as well as two neutrons. The additional electron cannot be added to the same quantum state as the first electron and is placed into the first level, 1s, but with a different spin state. The 1s level is completely filled by two electrons. Thus for the next element, lithium with three electrons and three protons plus neutrons, the third electron in lithium must go into a higher energy state than that of the first two electrons, the 2s orbital. Thus ever larger atoms containing more electrons and protons are configured such that the additional electrons enter higher energy states. If electrons did not obey the Pauli Principle, then all of the electrons in an atom, no matter how many electrons are present, would be put into the lowest energy, 1s state. As a result, chemistry would be very different from what is observed.

According to the Pauli Principle, an electron cannot move into an already occupied state. This situation is similar to that of parking automobiles in a parking lot. No two cars can be put into the same parking spot simultaneously. Obviously, a parking spot must initially be unoccupied in order to place a car into it. Electrons behave in much the same way. In the case of electrons, quantum states assume the same role as parking spaces do for cars. It is important to further recognize that a filled parking lot cannot accept any more cars without removing one and similarly a filled energy band cannot accept any more electrons without removing one. Now we can understand why a filled energy band does not conduct a current. For a current to flow, electrons must move from one state to another. In a filled band there are no vacancies into which the electrons can move since all possible states are filled. Hence, no current can flow.

The distinction among each of the four categories of solids can now be made based on the energy bands in the material. An insulator is a material in which the highest occupied band, called the valence band, is completely filled and the next available energy band, called the conduction band, is completely empty. The energy separation between the filled and empty bands is called the energy gap. In order for a material to be insulating, it is also necessary that the energy gap be very high such that carriers cannot be readily promoted from the valence band into the conduction band. Therefore in an insulator, the valence band is completely filled and the conduction band is completely empty and no current can flow in the material. Conversely, a metal is a highly conductive material. Metals are solids in which the conduction band is only partially filled. The conduction band consists then of many electrons and many empty states. A large current can be supported within a metal since most of the electrons within the conduction band can contribute to the current conduction since there exist many vacancies into which the electrons can move under the action of a driving field. Consequently, metals have a very high electrical conductivity. The other two categories of materials, semimetals and semiconductors, are somewhat intermediate between metals and insulators. Semimetals are materials like insulators in that the conduction band is unoccupied and the valence band is fully occupied at zero temperature. However, in semimetals the energy gap vanishes in part such that the conduction and valence bands intersect. Electrons from the valence band can be readily accelerated into the conduction band at the point or points of intersection of the two bands and the material can thus support a current. In this way, semimetals exhibit a relatively high conductivity but not as high as that of a metal. Finally, a semiconductor is something like an insulator but with a relatively small energy gap separating the conduction and valence bands. At absolute zero temperature within a semiconductor the conduction band is completely empty and the valence band is completely filled. However, as the temperature is raised to room temperature, the energy gap is sufficiently small that some measurable population of the conduction band occurs. Therefore, a semiconductor will conduct a current at room temperature but with a much higher resistance than that of a metal.

The electrical resistance of a crystal is a function of the electron concentration in the conduction band. In a metal, the electron concentration within the conduction band is extremely high, on the order of $\sim 10^{23} \text{ cm}^{-3}$. In a semiconductor the electron concentration within the conduction band is many orders of magnitude lower. Therefore, the conductivity of a semiconductor is much less than that of a metal. To quantify the conductivity it is essential to determine the electron concentration. In the next section the technique used to determine the electron concentration within a semiconductor will be discussed.

Before we end this section, it is useful to discuss the shape of the energy bands in a crystal. One of the basic concepts of quantum mechanics is that fundamental particles have a wave-particle duality. This implies that a fundamental particle like an electron for example sometimes manifests itself as a wave and sometimes as a particle, but never simultaneously. Therefore, an electron has a wavelength associated with it, called the de Broglie wavelength, that accounts for its wavelike behavior. The momentum of an electron can be described using its wavelike behavior as

6 Semiconductor fundamentals

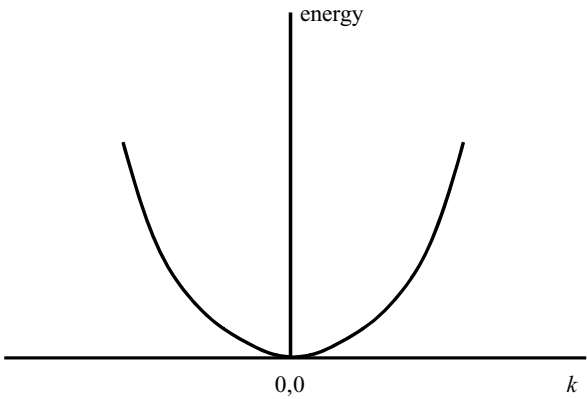


Figure 1.2 Sketch of the energy vs. k relationship for free electrons. Energy bands that obey this relationship are called parabolic energy bands. To a good approximation the energy bands within a semiconductor, at least near the band edge (bottom of the conduction band and top of the valence band), are parabolic.

$$p = \hbar k \tag{1.1}$$

where $\hbar$ is Planck's constant divided by 2π and k is defined as

$$k = \frac{2\pi}{\lambda} \tag{1.2}$$

λ is the electron wavelength and k is called the electron wavevector. A free electron has only kinetic energy given by

$$E = \frac{p^2}{2m} \tag{1.3}$$

Substituting into (1.3) for p the expression given by (1.1) obtains

$$E = \frac{\hbar^2 k^2}{2m} \tag{1.4}$$

The energy of the electron varies quadratically with the wavevector, k . The relationship between E and k given by (1.4) is called a parabolic energy vs. k relationship and is sketched in Fig. 1.2. Notice that the energy vs. k diagram shown in Fig. 1.2 is a parabola with vertex at $E = 0, k = 0$. Since the electron energy varies with respect to the electron wavevector, the $E(k)$ relationship is very important in semiconductors. The behavior of the electron as a function of k is referred to as the electron motion in k -space. In general, the wavevector $\vec{k}$ for an electron in a crystal is a three-dimensional vector. In free space, we can replace the vector $\vec{k}$ by its one-dimensional scalar magnitude, k . We can also often use this scalar one-dimensional model to gain insight into the behaviour of an actual semiconductor.

Typically, the mass that appears in the denominator of (1.4) is quite different from the free space mass and is referred to as the effective mass, usually written as m^* . The effective mass is usually less than the free space mass and takes into account

the motion of the electron within the crystalline lattice. The electron effective mass is defined as

$$\frac{1}{m^*} = \frac{1}{\hbar^2} \frac{d^2 E}{dk^2} \quad (1.5)$$

Notice that (1.5) implies that the curvature of the $E(k)$ relationship determines the effective mass of the electron. If the curvature is high, meaning that E changes greatly with a small change in k , then the effective mass of the electron is small. Conversely, if the curvature is low, implying that the energy E changes slowly with change in k , then the effective mass of the electron is large. In the limit of a horizontal line in the $E(k)$ relationship, the effective mass is infinite; the energy never changes for any change in momentum or k .

The energy bands within most semiconductors deviate from the simple parabolic energy relationship given by (1.4) at high energy, defined as several kT above the conduction band minimum or edge or several kT below the valence band edge.[†] The valence band edge is the point of minimum hole energy within the valence band and typically lies at $k = 0$ in k -space. The energy band structure in general is very complicated in most semiconductors, yet can have a profound effect on device operation, as will be seen in later chapters.

1.2 Equilibrium carrier concentrations and intrinsic material

It is important first to understand the concept of equilibrium. A full discussion of equilibrium can be found in the book by Brennan (1999). The most exacting definition of a system in equilibrium is that a closed system, isolated from the external environment, if left to itself over time will evolve towards equilibrium. Under equilibrium conditions there are no external agents, i.e., external voltages, external fields, radiative excitations, or any other external perturbation acting on the system. The system is completely isolated from the external world and as such is unperturbed. There is an important difference between equilibrium and steady-state. In steady-state the system does not change with time, but it is not isolated from the external world. In equilibrium the system is completely isolated from the external world and thus does not change with time but also has no net current flow. A system in steady-state though it does not change with time still has a net current flow. One simple way to view the difference between equilibrium and steady-state is to imagine a partially filled sink. In equilibrium the water level does not change and remains constant. Additionally, there is no net current flow. There is no input or output of water from the sink, the faucet is off and the drain is closed. For a sink in steady-state the water level also doesn't change. However, there is a net current flow. The faucet is on and the drain is open such that the input matches the output and thus the water level does not change. However, the

[†] Here, $k = k_B$, Boltzmann's constant. It is usually multiplied by T , the absolute temperature. The factor kT appears in the Fermi–Dirac distribution function, discussed in Section 1.2.

8 Semiconductor fundamentals

system interacts with its external environment and thus is not in equilibrium but in steady-state instead.

In order to calculate the electron concentration within the conduction band of a semiconductor in equilibrium it is useful to again draw an analogy to parking spaces and cars. In order to park one's car two conditions must be met. First, there must be a parking space. One cannot park one's car, at least legally and safely, in the middle of the road. There must be a parking space. However, the mere presence of a parking space does not ensure that one can park one's car. The second condition is that the space must be unoccupied. The obvious statement that one must have a vacant parking space available to park one's car has an analogy for electrons. In order to put an electron into an energy state, a similar set of two conditions must exist. These are that there must exist a state matching the energy of the electron into which it can be put and this state must be unoccupied. The total number of electrons in the conduction band depends upon the number of available states at a given energy multiplied by the probability that each state is occupied. Summing this product over all possible energies will give the total number of electrons within the conduction band. Mathematically, we can determine the electron concentration in the conduction band by integrating the product of the function that describes the number of available states at a given energy, called the density of states, $D(E)$, and the function that gives the probability that a state at that energy will be occupied, called the distribution function, $f(E)$. The electron concentration, n , is given then as

$$n = \int D(E)f(E)dE \tag{1.6}$$

where the integration is taken over the full range of energy values. In order to evaluate this expression it is necessary to determine both $D(E)$ and $f(E)$. The density of states function $D(E)$ for a three-dimensional system is given as (Brennan, 1999, Section 5.1),

$$D(E) = \frac{1}{2\pi^2} \left(\frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \sqrt{E} \tag{1.7}$$

where $\hbar$ is the reduced Planck constant, $\hbar/2\pi$.

The probability distribution function, $f(E)$, depends upon whether the system is in equilibrium or not. What then is the form of the equilibrium probability distribution function for electrons? To answer this question let us consider Fig. 1.3. Figure 1.3 shows a collection of bins, arranged in ascending energy into which one can place an electron. Let each bin represent an allowed energy state. It is important to recall that no two electrons can occupy the same quantum state simultaneously in accordance with the Pauli Principle. Therefore, once an electron has been placed into a bin, no additional electrons can be put into that bin. To attain the minimum energy configuration of the system the first electron must be put into the first bin. The next electron must then be placed into the second bin, the third electron into the third bin and so forth. This process continues until all of the electrons are placed into a bin. For example, in Fig. 1.3, if only six electrons are present they are placed into the first six bins as shown in the diagram. The normalized probability of each of the first six

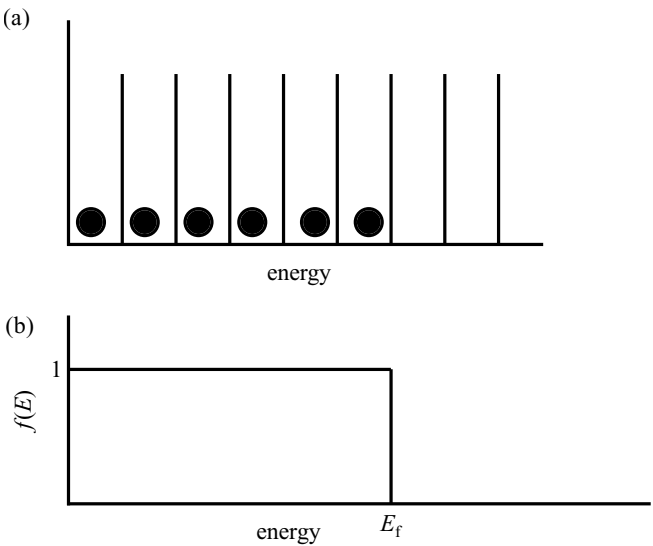


Figure 1.3 (a) Collection of energy bins representing energy states arranged in ascending energy. Into each bin only one electron can be placed in accordance with the Pauli Principle. Each circle represents an electron. The figure shows the minimum energy configuration of an arrangement of six electrons. (b) Corresponding probability distribution function, $f(E)$.

bins being occupied is thus 1. All bins above the sixth bin are empty in the example since no additional electrons are present. Hence the normalized probability of the bins higher than six being occupied is zero. The resulting probability distribution function is shown in Fig. 1.3(b). Note that the probability distribution shown in Fig. 1.3(b) holds for $T = 0$ K. Clearly, the probability distribution function reflects the physical situation, each of the first six states or bins is occupied, while those above six are empty. Inspection of Fig. 1.3 shows that the distribution has the value of 1 until an energy, E_f , is reached. This energy is called the Fermi level and is related to the number of electrons present in the system. For the present example, the energy corresponding to the Fermi level lies at the energy corresponding to the sixth bin.

What happens though at temperatures greater than absolute zero? Temperature is a measure of the internal energy of the system. At temperatures greater than zero, the total energy of the system must be greater than that corresponding to $T = 0$ K. Let us again consider a system with only six electrons. For simplicity let us set the energy of each bin to be an integer multiple of E . Thus for the system shown in Fig. 1.3(a), the total energy is given as the sum of the occupied bins as $E + 2E + 3E + 4E + 5E + 6E = 21E$. The next highest energy configuration, or higher temperature of the system is obtained by promoting the sixth electron into the seventh bin. The corresponding energy of the resulting configuration is then equal to $E + 2E + 3E + 4E + 5E + 7E = 22E$, which is obviously higher than that of the $T = 0$ K configuration. Higher temperature configurations are similarly achieved, i.e. by promoting electrons from the lower energy states into higher energy states. An example system is shown

10 Semiconductor fundamentals

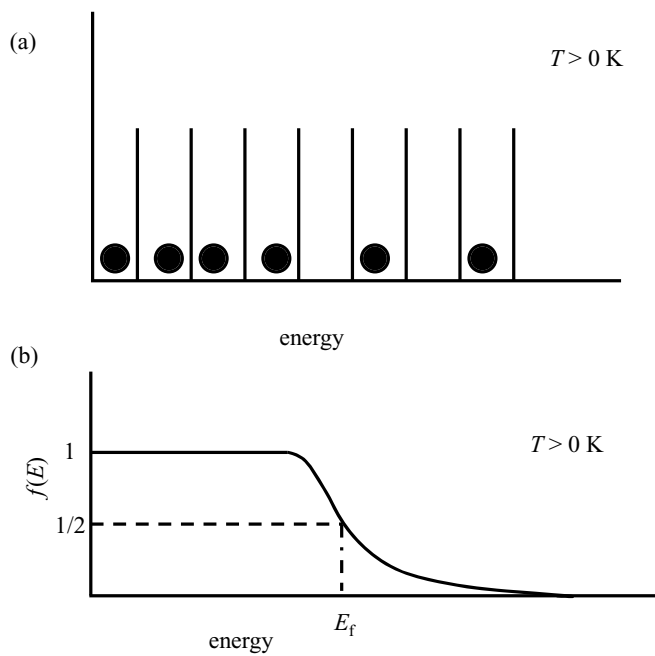


Figure 1.4 (a) Distribution of a collection of six electrons in energy bins corresponding to a collective energy or temperature greater than 0 K. (b) Corresponding probability distribution function for a $T > 0\text{ K}$ distribution.

in Fig. 1.4 along with the corresponding probability distribution function. As can be seen from the figure, at higher energies there exists a tail in the distribution function. This implies that there is a nonzero probability of a state with energy greater than E_f being occupied and correspondingly a nonzero probability that states below E_f are unoccupied.

The equilibrium probability distribution function, $f(E)$, called the Fermi–Dirac distribution can be expressed mathematically. Its derivation is rather complicated and will not be repeated here. The interested reader is referred to the books by Brennan (1999, Chapter 5) or Kittel and Kroemer (1980). The Fermi–Dirac distribution is given as

$$f(E) = \frac{1}{1 + e^{\frac{(E-E_f)}{kT}}} \tag{1.8}$$

where k is Boltzmann’s constant.

It is instructive to examine how $f(E)$ behaves and to show it replicates the distributions shown in Figs. 1.3 and 1.4. Consider first its behavior at $T = 0\text{ K}$. There are two conditions, $E < E_f$ and $E > E_f$. For $E < E_f$, the exponent in (1.8) is negative infinity (due to the division by zero), and exp of negative infinity is zero. Thus $f(E)$ for $E < E_f$, is $1/(1 + 0)$ or simply 1. This is of course exactly what is expected; for energies less than the Fermi level, at $T = 0\text{ K}$, $f(E) = 1$. The second case, $E > E_f$ at $T = 0\text{ K}$ leads to the following. Notice that in this case, the exponent is now positive infinity,