

Lecture 17

Introduction to Semiconductors

From this lecture on, we will be studying various phenomena involving solid-state materials and their interfaces. Most importantly, we will be discussing about the topic of semiconductors. In this lecture, the basic concept of solid-state and semiconductor physics will be introduced. For more detailed discussions about the physics of semiconductors, we recommend the readers to more specialized textbooks such as Ref. [1] and [2].

17.1 From molecule orbitals to solid state bands

The key physical properties of semiconductors are highly related with their electronic structures. Unlike single molecules, the behaviors (distribution, energy, momentum, etc.) of electrons in a semiconductor solid are periodic. In solid state physics, such periodic electronic structure is usually represented by the band theory, as a consequence of quantum mechanism. The rigorous derivation of the band theory is beyond the scope of this course, however the essence of the band theory can be well understood starting from our knowledge of molecule orbitals.

Let's start from the most common semiconductor, Si. The electronic structure of Si atom is $[\text{Ne}]3s^23p^2$. When the two Si atoms form a diatomic Si_2 molecule, the atomic orbitals (AO) split and form molecule orbitals (MO), as seen in Figure 17.1a.

Among all the Mos, the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) contribute most to the physical and chemical properties of a molecule. In the case of Si_2 , the HOMO is a π -type orbital and the LUMO is a σ -type orbital. When the number of atoms increases, more energy levels are created as a result of atomic orbital splitting, while the spacing between the individual energy levels are small enough, such that they can be regarded as a continuous distributions, which are named as "bands" (Figure 17.1b). Similar to the concept of HOMO and LUMO in individual molecules, the highest occupied and the lowest unoccupied bands are named as the conduction band (CB) and valence band (VB), respectively. For a semiconductor or insulator, the VB is fully occupied (Figure 17.1c). At 0 K, the energy of an electron cannot stay between the CB and VB. The energy difference between CB and VB is called the bandgap (usually noted as E_g). We will study in the next lectures, that the bandgap is the key quan-

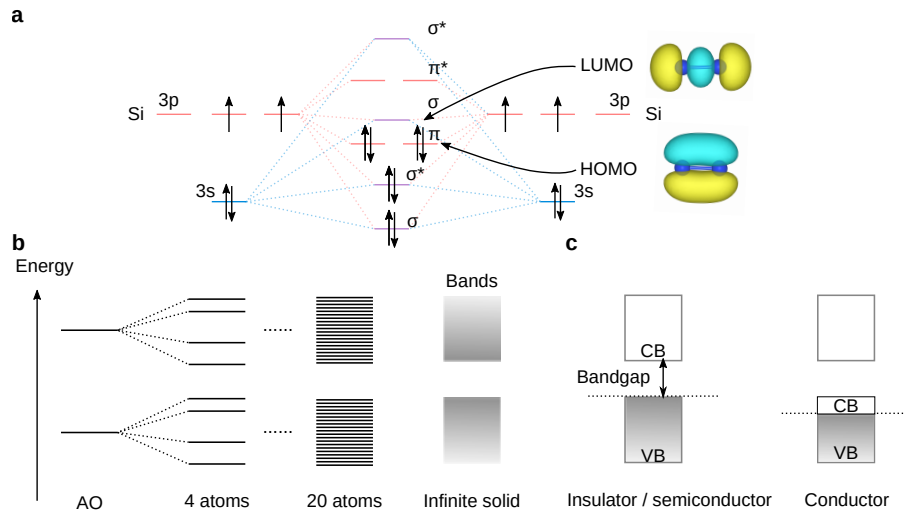


Figure 17.1: From molecule orbitals to bands. (a) MO diagram of Si_2 dimer, with σ and π orbital types shown. The shape of HOMO and LUMO are calculated using first principles simulations. (b) Evolution of electronic bands by increasing number of atoms. (c) The conduction band (CB) and valence band (VB) of semiconductor and conductors. A bandgap exists in semiconductor while CB and VB overlaps in a conductor.

tity that determines the property of a semiconductor. On the other hand, there is no gap between the CB and VB of a conductor, allowing conduction to occur.

The electrons in a solid material exhibit wave-like behavior, with its energy E varying as a function of the momentum $p = \hbar k$, where \hbar is the reduced Planck constant and k is the wave vector. In solid state physics, people usually plot the relation of E vs k (or p) for a typical material, and such plots are known as the “**band structure**” of a semiconductor. A typical band structure of the CB and VB of a semiconductor can be seen in Figure 17.2a, such that the energy E and momentum p has a parabolic relation.

This is similar to the case of free carriers: $E = E_0 + p^2/2m^*$, where E_0 is the potential energy and m^* is the “effective” mass of the carrier (will be discussed in following sections). The band structures of CB and VB are generally different: VB is “flatter” compared with CB and may also have splitting due to the effect known as spin-orbit-coupling (SOC). When we study the motion of carriers in a semiconductor, a simpler representation is often adapted. Consider a chain of atoms in a semiconductor without external electric field (Figure 17.2b). The energy of a carrier at a certain position, is described by the “local band structure” of individual atoms. At a larger length scale, the details of the band structure can be ignored, and the energy of the carriers have a continuous distribution, leading to a representation called “band diagram”. The bottom of CB in a band diagram is a line connecting all the minima of the CBs in the local band structure, and the height of CB represents the kinetic energy of electron. The situation is similar for VB.

Using the band diagram, we can easily determine the motion of carriers. The slope of the edge of CB and VB in a band diagram, equals the opposite of external electric field, $-\mathcal{E}$. As a result, an electron move downwards the CB in a band diagram, while a hole moves

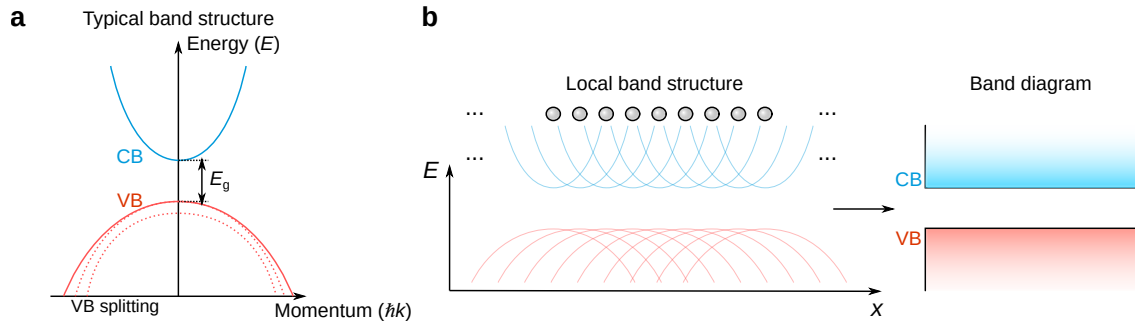


Figure 17.2: Band structure and band diagram of semiconductors. (a) Typical energy-momentum relation of a semiconductor, showing the CB and VB as parabolic bands. (b) Band diagram representation of semiconductors. The extrema of CB and VB in local band structures of individual atoms are linked to form continuous bands.

upwards the VB (Figure 17.3a). The equation of motion of carriers in the semiconductor, is:

$$\begin{aligned}
 m_n^* \frac{dv}{dt} &= -e\mathcal{E}, & \text{Electron} \\
 m_p^* \frac{dv}{dt} &= e\mathcal{E}, & \text{Hole}
 \end{aligned}
 \tag{17.1}$$

where m_n^* and m_p^* are the effective masses of electron and hole, respectively. Near the extrema of CB, the energy-momentum relation in a band structure (Figure 17.3b) gives $E = E_c + \frac{p^2}{2m^*}$, where E_c the energy of the CB edge, we can calculate m_n^* . Similarly, we can get m_p^* . For silicon, $m_n^* = 0.26 m_0$ and $m_p^* = 0.39 m_0$, where m_0 is the mass of electron in vacuum. In other words, in silicon, the carriers travel faster than in vacuum and electrons travel faster than holes.

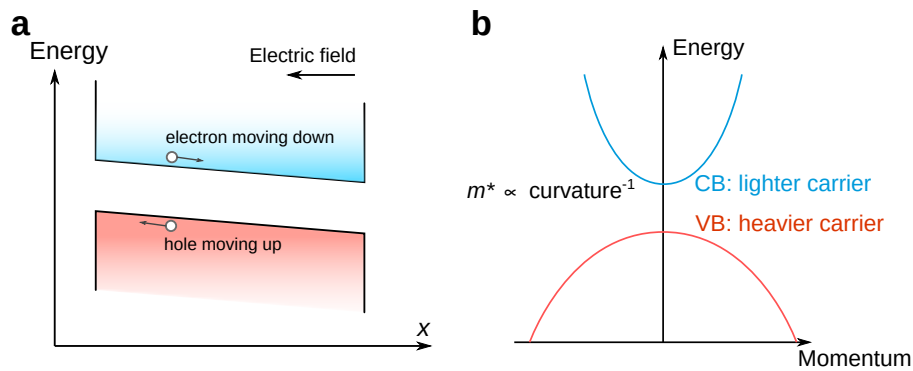


Figure 17.3: Motion of carriers in a band diagram. (a) Direction of carrier movement in a band diagram, with electron moving upwards the CB and holes moving downwards the VB. (b) Explanation of different carrier effective mass using band structure.

17.2 Carrier distribution in semiconductors

Due to the existence of the bandgap, a semiconductor cannot conductor current at 0 K. However when $T > 0$ K, a small portion of valence electrons can be thermally excited to the CB, leading to finite conductivity. As a result, positively-charged “holes” (usually represented by h^+) are left in the VB. Note that holes are not real particles, but rather movable vacancies of electron. In an **intrinsic** semiconductor (i.e. “pure” semiconductor without any impurity), the concentration of ionized electrons, n , always equals to the concentration of ionized holes, p , and is a function of temperature T . When we introduce other atoms (called dopants) in a semiconductor, the concentrations of electron and hole can be different. For instance, replacing a Si atom with P ($[\text{Ne}]3s^23p^3$) introduces an additional electron that can be easily ionized, and a positively charged, immobile ion P^+ . In this case, the semiconductor is **n-doped**, and P is called the **donor**. As a result, at equilibrium, we have more electrons than holes, such that $n > p$. On the other hand, replacing Si with B ($[\text{Ne}]2s^22p\{1\}$) creates additional holes and $n < p$. We call such semiconductor **p-doped** and B is called the **acceptor**. The additional carriers in doped semiconductors, compared with the intrinsic semiconductor, provides higher conductivity. Furthermore, the degree of conductivity, can be controlled by the concentration of donor (N_d) or acceptor (N_a). The effect of doping on the carriers of a semiconductor, can be seen in Figure 17.4.

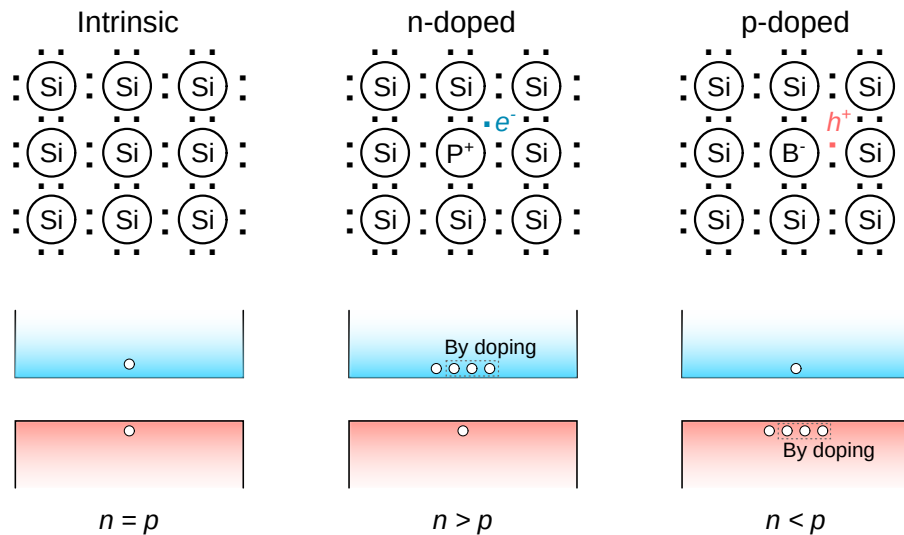


Figure 17.4: Effect of doping in a semiconductor. The black dots represents the electrons that form covalent bonds and the excess carriers (electron and hole) are shown in color.

To calculate the concentration of electrons and holes, we need to introduce the concept of **density of states** (DOS). The DOS at energy E , $D(E)$, is defined as the number of states divided by the energy spacing ΔE per unit volume Ω :

$$D(E) = \lim_{\Delta E \rightarrow 0} \frac{\text{Number of states}}{\Delta E \Omega} \quad (17.2)$$

and usually expressed in $\text{eV}^{-1}\text{cm}^{-3}$. An intuitive way to understand the origin of DOS, is to start from the single molecule picture again. For a single molecule, the energy levels are discrete. The number of states can be viewed as how many orbitals have the same energy level. When the number of atoms approaches infinity, the energy levels in a semiconductor becomes continuous, and the number of states at each energy level, is described by the DOS (Figure 17.5).

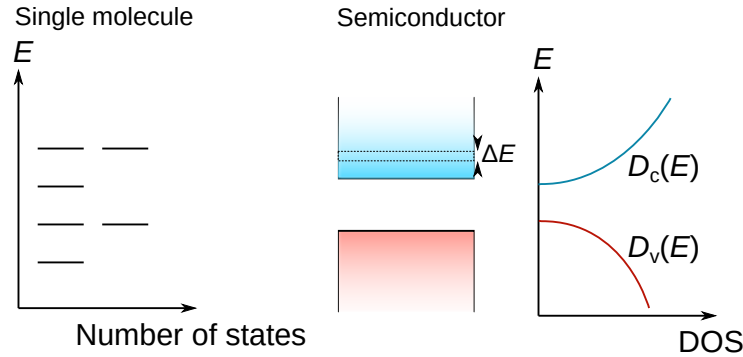


Figure 17.5: Evolution of discrete energy levels in single molecule to continuous density of states in semiconductor solid.

For a bulk semiconductor, the DOSs of electrons ($D_c(E)$) and holes ($D_v(E)$) have parabolic distribution, and are calculated by:

$$D_c(E) = \frac{8\pi m_n^* \sqrt{2m_n^*(E - E_c)}}{h^3} \quad (17.3)$$

$$D_v(E) = \frac{8\pi m_p^* \sqrt{2m_p^*(E_v - E)}}{h^3} \quad (17.4)$$

where E_c and E_v are the energies of the CB and VB edges, respectively, and h is the Planck constant. Note that such parabolic relation is only valid for bulk (3D) semiconductors, while the DOS of lower dimension semiconductors takes completely different forms (Figure 17.6).

Now let's calculate the density of carriers in a semiconductor. We define the occupation number $f(E)$ as the probability that states at energy E is occupied. $f(E) = 0$ means that the states are fully unoccupied while $f(E) = 1$ means the states are fully occupied. The total number of carriers in a semiconductor can be calculated using the DOS and occupation number:

$$n = \int_{E_c}^{\infty} D_c(E) f(E) dE \quad (17.5)$$

$$p = \int_{-\infty}^{E_v} D_v(E) [1 - f(E)] dE \quad (17.6)$$

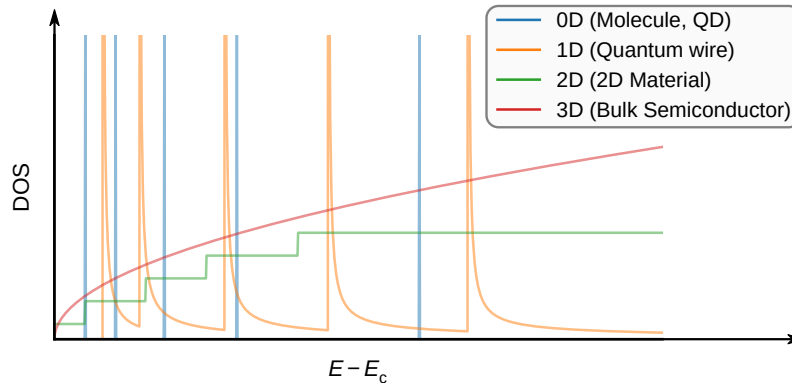


Figure 17.6: DOS of semiconductors with different dimensions.

The quantity $f(E)$ is described by the Fermi-Dirac distribution as:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)} \tag{17.7}$$

where E_F is the chemical potential (or known as the Fermi level) of the semiconductor, and k_B is the Boltzmann constant. The shape of Fermi-Dirac distribution is shown in Figure 17.7a.

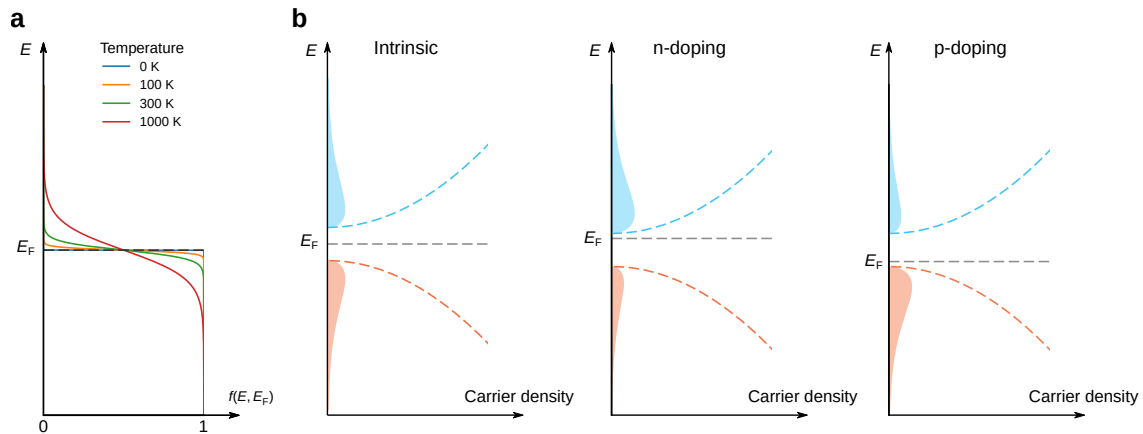


Figure 17.7: Carrier distribution in semiconductors. (a) The Fermi-Dirac distribution functions at different temperatures. (b) Energy-dependent distribution of electron (blue shaded region) and hole (red shaded region) for intrinsic and doped semiconductors. The dash curves correspond to the DOS.

At 0 K, $f(E)$ is a step function, with all states below E_F occupied and all states above E_F unoccupied. With the temperature increasing, we see that the shape of $f(E)$ is gradually smeared and the probability to get a occupied state above E_F increases, which corresponds to the thermal activation we discussed before. There are several concepts you need to remember about the Fermi-Dirac distribution:

1. The location of E_F does not necessary mean there is a state. For instance in intrinsic semiconductors, E_F is in the middle of the bandgap, where $D(E) = 0$.
2. Although the distribution function changes with temperature, the position of E_F is independent of the temperature.
3. The states that $E > E_F$ can be occupied as a result of thermal activation, described by the Fermi-Dirac distribution.

In comparison with semiconductors, the CB and VB of a conductor overlap. Therefore E_F coincide with the top of the VB. In practice, people usually use the quantity “work function” ϕ to define E_F . The definition of work function is the energy difference between E_F and the vacuum energy E_{vac} , such that:

$$e\phi = E_{\text{vac}} - E_F \quad (17.8)$$

and have unit of V. The values of work functions in several common metals are listed in Table 17.1.

Table 17.1: Work function $e\phi$ for several metals.

	$e\phi$ in eV
Al	~4.0
Ag	~4.5
Au	~5.0

Finally we can derive the expressions for n and p in a semiconductor. Using Equation 17.5 and the Fermi-Dirac distribution, we get:

$$\begin{aligned}
 n &= \int_{E_c}^{\infty} \frac{8\sqrt{2\pi}m_n^{*3/2}}{h^3} \sqrt{E - E_c} \frac{1}{1 + \exp\left(\frac{E - E_F}{k_B T}\right)} dE \\
 &\approx \int_{E_c}^{\infty} \frac{8\sqrt{2\pi}m_n^{*3/2}}{h^3} \sqrt{E - E_c} \exp\left(-\frac{E - E_F}{k_B T}\right) dE \\
 &= N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right)
 \end{aligned} \quad (17.9)$$

where the approximation is valid when $E_c - E_F \gg k_B T$. The value N_c is called the effective density of states in the CB, and has the form:

$$N_c = 2 \left[\frac{2\pi m_n^* k_B T}{h^2} \right]^{3/2} \quad (17.10)$$

Similarly, we can derive the concentration of holes:

$$p = N_v \exp\left(-\frac{E_F - E_v}{k_B T}\right) \quad (17.11)$$

with N_p expressed as:

$$N_v = 2 \left[\frac{2\pi m_p^* k_B T}{h^2} \right]^{3/2} \quad (17.12)$$

An important feature is that the product of n and p is constant for a semiconductor, such that:

$$n \times p = N_c N_v \exp\left(-\frac{E_g}{k_B T}\right) = n_i^2 \quad (17.13)$$

For intrinsic semiconductors, we know that $n = p = n_i = \sqrt{N_c N_v} \exp\left(-\frac{E_g}{2k_B T}\right)$. This is about 10^{10} cm^{-3} for silicon ($E_g = 1.12 \text{ eV}$) at 300 K.

With the help of Equations 17.9 and 17.11, we are able to calculate the carrier density using the value of E_F . In an intrinsic semiconductor, we have $E_F = (E_c + E_v)/2$. What about doped semiconductors that we do not know the position of E_F ? Qualitatively speaking, in an n-doped semiconductor, we have $n > p$, meaning the E_F shifts towards the edge of CB, E_c , such that $(E_c - E_F) < (E_F - E_v)$. Similarly, in p-doped semiconductors, E_F shifts towards E_v , such that $(E_c - E_F) > (E_F - E_v)$. The charge distribution described by the Fermi-Dirac model in intrinsic and doped semiconductors, can be seen in Figure 17.7b. We can describe the carrier density in a doped semiconductor in a generalized picture: the overall charge is contributed by the mobile carriers (n and p), as well as immobile ionic dopants (N_d and N_a). They are governed by the following equations:

$$\begin{aligned} n + N_a &= p + N_d \\ n \times p &= n_i^2 \end{aligned} \quad (17.14)$$

giving the solutions:

$$n = \frac{N_d - N_a}{2} + \left[\left(\frac{N_d - N_a}{2} \right)^2 + n_i^2 \right]^{1/2} \quad (17.15)$$

$$p = \frac{N_a - N_d}{2} + \left[\left(\frac{N_a - N_d}{2} \right)^2 + n_i^2 \right]^{1/2} \quad (17.16)$$

When the majority carrier is dominated by doping, the carrier concentrations can be simplified:

1. n-doped, $N_d - N_a \gg n_i$

We have:

$$\begin{aligned} n &\approx N_d - N_a = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) \\ p &\approx \frac{n_i^2}{n} \end{aligned} \quad (17.17)$$

2. p-doped, $N_a - N_d \gg n_i$

We have:

$$\begin{aligned}n &\approx \frac{n_i^2}{p} \\p &\approx N_a - N_d = N_v \exp\left(-\frac{E_F - E_v}{k_B T}\right)\end{aligned}\tag{17.18}$$

Using Equations 17.17 and 17.18, we can estimate the position of E_F if we know the dopant concentration, or calculate the carrier concentration when we know the position of E_F .

References

- (1) Sze, S.; Ng, K. K. In *Physics of Semiconductor Devices*; John Wiley & Sons, Inc.: 2006, pp 5–75.
- (2) Yu, P. Y.; Cardona, M., *Fundamentals of semiconductors: physics and materials properties*; Springer: 2010.

