

Notes for Lecture 8

Intrinsic and Extrinsic Semiconductors

In the previous lecture, we covered the rigorous concepts of a “hole” and an “electron.” These electrons and holes have light masses, $O(0.1)$ of the mass of the bare electron, and must be viewed as fundamentally different from the bare electron.

For a pure semiconductor crystal, the $T = 0$ state is a state in which every band is either completely filled (like the valence band) or completely empty (like the conduction band). This state can be thought of as a state with no electron and no hole. As T is raised, the system is excited to higher energy states. These higher energy states are characterized by some electrons in the conduction band and some holes in the valence band, where just how much electrons or holes are created is aptly summarized in the Fermi function. It is precisely in this sense that the electrons and the holes are fundamental particles of a solid: they are the fundamental ingredients for describing low¹ energy excitations of a solid. In contrast, a bare electron is not a fundamental particle in a solid.

In this lecture, we discuss Fermi energy and doping. When a pure semiconductor (intrinsic semiconductor) is doped, we call it an extrinsic semiconductor, which is far more important than an intrinsic semiconductor.

¹It may be that “higher” and “low” in this paragraph are confusing. The energy associated with kT is “higher” than the energy at $T = 0$. But, it is very “low” compared to the typical electronic energy scale (like band gap).

8.1 Fermi energy

The Fermi energy E_F has been defined already as a sharp boundary between occupied states and unoccupied states.

However, here is a more general definition of it².

The Fermi energy E_F is defined at $T = 0$, as the energy gain for adding an electron or the energy cost for removing an electron, where the initial state and the final state are in equilibrium.

Suppose that the initial number of electrons is N_e . Then, denoting the equilibrium energy as $E(N_e)$, the above definition corresponds to

$$E_F = E(N_e + 1) - E(N_e) \quad (8.1)$$

$$= E(N_e) - E(N_e - 1) \quad (8.2)$$

This definition implicitly assume that the two values – the electron removal energy ($E(N_e) - E(N_e - 1)$) and the electron addition energy ($E(N_e + 1) - E(N_e)$) – are equal to each other. This would be the case of a metal. In an intrinsic semiconductor, this is not true. Calling the valence band maximum (VBM) energy 0, and the conduction band minimum (CBM) energy E_G (energy gap), the electron removal energy is 0, and the electron addition energy is E_G . So, E_F is not well-defined for an intrinsic semiconductor. If one were to *insist* to have an E_F value even for an intrinsic semiconductor, then one would probably choose the average $E_G/2$ as E_F . We will see that this is a reasonable thing to do, later on. In practice, however, the Fermi level is “pinned” by dopants such as impurities, as a really pure semiconductor is impossible to obtain. So, now, let us consider the case of “doped” semiconductors, i.e. extrinsic semiconductors.

8.2 Doping and E_F Pinning

Doping of charge carriers occur when an atom of a difference valence replaces another atom.

²This definition is valid even if the band theory is not applicable. The band theory is however valid for semiconductors for the most part. For magnetic devices, however, the band theory is often not valid.

Let us look at the following sub-set of the periodic table.

row	III	IV	V
2	B	C	N
3	Al	Si	P
4	Ga	Ge	As
5	In	Sn	Sb

A Si crystal and a Ge crystal are well-known diamond structure semiconductors. Four valence electrons per atom. Each atom participates in four bondings, and so each bonding gets two electrons.

A GaAs crystal or a InP crystal are well-known so-called III-V semiconductors. The crystal structure is essentially the same as the diamond structure, except that the two fcc sub-lattices correspond to different atoms (zincblende structure). In these crystals each III element gives three valence electrons, while each V element gives five valence electrons. This means that again, there are exactly two electrons to participate in each bonding orbital.

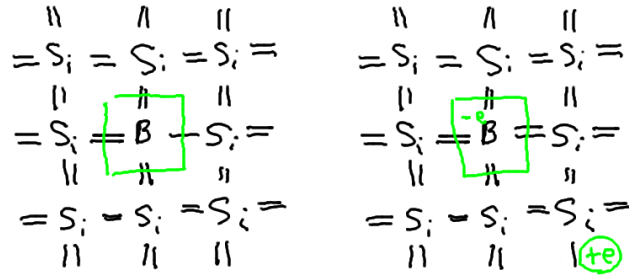
You can also see that a II-VI semiconductor must be possible too. Indeed. CdTe is an example.

In any case, when an atom in a semiconductor is replaced by another atom of different valency, then a doping can occur, and the replacing atom is called a **dopant** atom. For instance, if B is used as a dopant in a Si crystal, it will introduce a hole. In this case, we have a p -type semiconductor, and, B is called an **acceptor**. If, however, P is used as a dopant in a Si crystal, then we have a n -type semiconductor. In this case, P is called a **donor**.

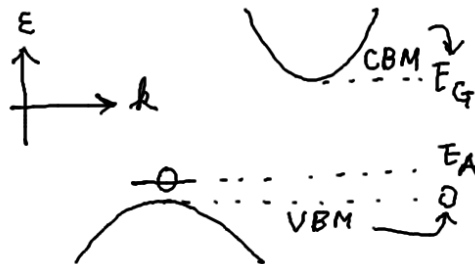
Consider the case of a B dopant, to be concrete. As shown below³, a B replacing a Si will resulting in a missing bond electron, i.e. a hole! But, of course, this hole will have a large mobility, due to the large hopping amplitude. So, after an instant ($\sim$ a femto second), the hole will already jumped couple of atoms, as shown on the right. What happened? As the result of this hole moving away from the B atom, the bonds at the B site have been “repaired” but a result of this is a negative charge, $-e$ for the B site (green square). So, what we have is then a heavy $-e$ charged B ion and a very light hole with charge $+e$. As we have seen in HW #1, problem 5, this leads to a Bohr-atom like binding, but with much less binding energy, on order of 10 meV,

³Note that this “bonding diagram” maps a three dimensional crystal structure to a convenient two dimensional structure. This is purely for investigating chemical bonds. The real structure is of course not a square two dimensional crystal like this.

instead of 13.6 eV (see Table 2.3 of the textbook), and with a correspondingly large orbit size.

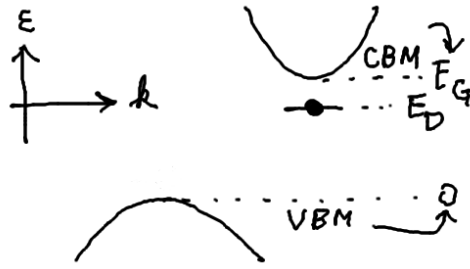


How would we mark this finding in an energy diagram? Note that the energy diagram almost always refers to the *electron* energy diagram, which means that the corresponding energy to the hole is reversed in sign. So, in the energy diagram, the binding of the hole to the positive B ion corresponds to energy increase! This is the so-called acceptor level, marked by E_A in the diagram below.



Here the empty circle at the acceptor level (E_A) means that a hole is there. This would be the ground state corresponding to $T = 0$, with the binding energy of the hole given by E_A . As the temperature rises, though, the hole can be excited to a higher energy level. Namely, in this *electron* energy diagram, the hole will jump *down* to create a free hole in the valence band.

A case of a donor impurity is equally easy. The donor level (E_B) forms just below the CBM (E_G) with the binding energy of the electron being $E_G - E_B$. At $T = 0$, the electron will be bound at the donor level, but it will become mobile if it gets excited to the conduction band at a finite temperature.



A good question is why these donor levels and acceptor levels are drawn flat? It is because they do not form a band if impurities are far from each other so that hopping amplitude for these orbitals is very small. These energy levels then remain essentially “atomic,” like core levels of atoms. You can think of this flat band as the limit of $t \rightarrow 0$ from the tight binding band theory.

Another good question may be “is this acceptor level or donor level something new?” – the answer is yes and no. The answer is yes, since they are new quantum states created mid-gap. The answer is no, in the sense that when a dopant level is created, it is not created out of the blue, but one state is “pulled out of” the conduction band (for the donor level) or the valence band (for the acceptor level)⁴. However, since the dopant density is small, the effect on the conduction band or the valence band is ignorable.

Common n type or p type semiconductors are not pure n type or pure p type. The reason is that there will be some impurities or some structural defects (vacancies, e.g.), which cause a small amount of doping even if there was no artificial doping intended. So, an n type semiconductor means that the n type dopant dominates any p type dopants. And a p type semiconductor means that the p type dopant dominates any n type dopants (intended or un-intended).

Now, let us consider the question (see the activity slides). At $T = 0$, would the conduction band ever have any electrons in a semiconductor? (Here, we focus on a neutral semiconductor, only! Also, we assume that impurity levels are non-dispersive, i.e. flat.) The answer would be no. The conduction band is always empty. Likewise, at $T = 0$, the valence band is always full.

A related question is where would the Fermi level be when one has a common n type semi-conductor, i.e. one with lots of donor impurities but with a small number of acceptor impurities. In this case, the electrons in the donor level will see that there are hole states available at the acceptor level. Since the acceptor level have much lower energy level than the donor level, some electrons will move there! Due to there

⁴This is because the number of possible wave functions that one can make is fixed for a given band and the number of atoms.

being much more donors than acceptors, the result is a partially filled donor level, and a completely filled acceptor level. So, the Fermi energy (E_F) lies exactly at E_D !

By the same token, for a common p type semi-conductor, E_F lies at E_A !

This is an important concept: “**Fermi level pinning**” or $T = 0$ chemical potential pinning!

8.3 Chemical Potential

The **chemical potential** (μ) is a concept very similar to the Fermi energy.

It is defined as the “energy” gain or loss when a particle is added or removed from the system, while the system is kept in equilibrium in the initial and final states.

However, the chemical potential applies to any finite temperature, and the “energy” above refers to the “free energy” ($E - TS$, where E is the energy, T the temperature, and S the entropy). At a finite temperature it is the minimization of the free energy, not the minimization of E alone, that determines the equilibrium state (competition of energy lowering and entropy maximization).

The importance of the chemical potential is the following. When two systems containing the same type of particles are in contact and in equilibrium with each other, then the chemical potential values of that particle (which can be Fermion or Boson) for the two systems must be equal.

To understand this is rather simple. Any two systems that are in contact are constantly exchanging particles, due to random events caused by thermal fluctuations. Say, you have two systems which contain the same kind of particles, but with two different chemical potential values, μ_1 and μ_2 . Let the corresponding number of particles be $N_1^{(0)}$ and $N_2^{(0)}$. When in contact, there will be a natural flow of particles from the high chemical potential to the low chemical potential side, since that will lower the total free energy. This flow can stop only if the chemical potential values for the two systems become equal. So, if we started with $\mu_1 > \mu_2$, then we will end up with the final chemical potential μ such that $\mu_1 > \mu > \mu_2$. In the final equilibrium state, the number of particles would be such that $N_1 < N_1^{(0)}$ and $N_2 > N_2^{(0)}$.

An analogy with a mechanical system can be made. If in a cylinder a freely moving membrane separates two gases with different pressures, then the high pressure gas

will increase its volume and the low pressure gas will decrease its volume, accordingly, until the pressure become equal so that there is no net force on the membrane. Here, the situation is very similar. When two systems are in contact with different chemical potential values, then the high μ system will decrease its number of particles, while the low μ system will, accordingly, increase its number of particles. Until, the two chemical potential values become equalized.

8.4 Fuzzy but Prevalent Convention

Up to now, I have been trying to be careful in distinguishing between the chemical potential (μ) and the Fermi energy/level (E_F).

μ applies to any kind of particle and at any temperature.

E_F applies to Fermions only and at $T = 0$.

Unfortunately, the semiconductor literature does not follow this convention that is widely accepted in the physics literature. Instead, in the semiconductor literature, “Fermi level (E_F)” means the chemical potential of the electron at any temperature!

We will follow this fuzzy convention, from now on! And only from now on! **So from now on, E_F is applicable at any temperature, not just at $T = 0$!** Just keep in mind you physics lovers, that you will benefit by remembering that the chemical potential is a much more general concept, and indeed, what semiconductor engineers call the “Fermi level” is in fact the “chemical potential” for the electron.

So, with this convention the following is a repeat of the last boxed paragraph (with a now narrowed meaning, focussed only on electron systems).

When two systems of electrons are in contact and in equilibrium, then the Fermi levels of the two systems must be equal.