Notes for Lecture 7

Holes, Electrons

In the previous lecture, we learned how electrons move in response to an electric field to generate current. In this lecture, we will see why the “hole” is a natural concept.

7.1 Effective Mass

Before we go to the topic of “hole,” we discuss this important topic. The **effective mass** m^* is defined as

$$m^* = \left(\frac{1}{\hbar^2} \frac{d^2\varepsilon(k)}{dk^2} \right)^{-1} \tag{7.1}$$

Here, we use a simple notation (k instead of $\vec{k}$) suitable for 1D, or for an isotropic energy dispersion in 3D. The “isotropic energy dispersion” means that ε is dependent only on $k = |\vec{k}|$. In the most general case, the effective mass is defined as a “tensor” but we won’t go there[1].

Note also that we are concerned with a non-relativistic regime only.

It is helpful to check that the above definition is reasonable for a free electron. In this case, $\varepsilon(k) = \hbar^2 k^2/(2m_e)$, where for clarity[2] we use m_e (same as m_0 of the

---

[1] We do not go there, because it is not important for learning important concepts. You may need to deal with the tensor aspect of the effective mass if you are doing research, though.

[2] In previous lectures, we often used m to mean m_e, the electron mass. We won’t do it any more, since from now on we have to deal with different masses.

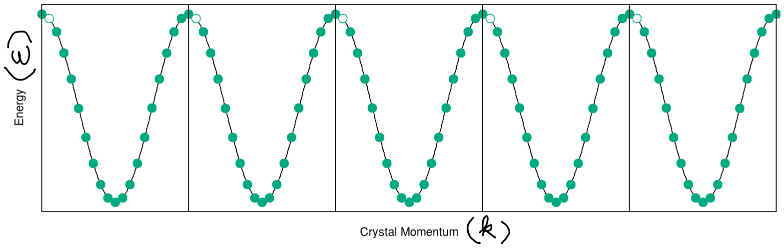
book). Here, we are taking only the kinetic energy part, since the rest energy part, $m_e c^2$ is just a constant. Given this free electron dispersion, it is easy to see that $d^2\varepsilon(k)/dk^2 = \hbar^2/m_e$. So, the effective mass defined as Eq. 7.1 is, in this case, exactly m_e.

The energy dispersion $\varepsilon(k)$ of the electron in a crystal contains information about the quantum mechanical interaction between the "bare" electron with the crystal lattice. Due to this interaction, the electron might find itself a bit more difficult to move around, or a bit easier to move around. As we shall see later on, this concept of "how easily the electron moves around" is the concept of the *mobility*, which is an important concept and is related partly to the effective mass.

7.2 Hole

Recall the following diagram from the previous lecture.

$t = \frac{\gamma}{2}$



Note that in this diagram the important states are near the top of the dispersion curve.

Let us ask the question, what is the effective mass of these states near the top of the $\varepsilon(k)$ curve? The answer is somewhat disturbing. Since, the dispersion curve is concave downwards, the effective mass is negative! Another way of seeing this is that the slope decreases near the top of the dispersion curve, and so $d^2\varepsilon/dk^2$ is negative, giving a negative effective mass, according to Eq. 7.1!

While a negative mass may not be the end of the world, it definitely feels strange.

Another related thing is that any physical quantity for a nearly filled band such as the above involves a sum over all k values except one, $\sum_{k'}$ in the notation of the last class. Such a sum seems unnecessary.

The solution to these two "problems" is rather simple. **We reference the configuration above relative to the filled band. Such a state of absence of an**

electron is defined as a hole. The quantum numbers of a hole is then obtained by this referencing rule.

For instance, the total crystal momentum of the above state is given by $\hbar \sum_{k'} k$. When we reference this to the total crystal momentum of the filled band, i.e. when we subtract the total crystal momentum of the filled band from the above, then we have $\hbar \left(\sum_{k'} k - \sum_{k} k\right)$. This is the negative of the crystal momentum of the empty state! How so? It is because the only difference between the two sums is the fact the empty state is included in the second sum, but not in the first sum.

By the same token, **any quantum number**, the total energy or the total spin angular momentum or the electrical charge etc., **is the negative of that for the empty state!**

We summarize this as follows

$$\text{charge} : q_h = +e = -(\text{charge of electron}) \tag{7.2}$$
$$\text{energy} : \varepsilon_h = -\varepsilon_e \tag{7.3}$$
$$\text{crystal momentum} : k_h = -k_e \tag{7.4}$$

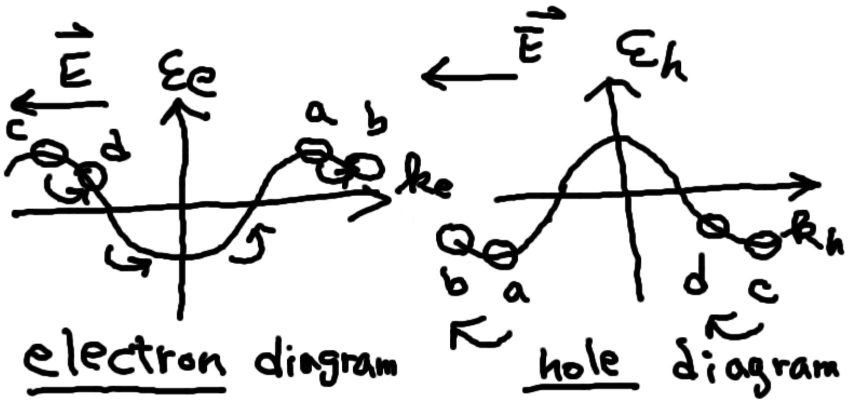
However, not all physical quantities are negated as we go from the electron picture to the hole picture. An example is the group velocity: $v_g = \frac{1}{\hbar}\frac{d\varepsilon}{dk}$ remains the same for the hole, since both ε and k change sign. A quantity that *does* change the sign is the effective mass, as m^* has the same sign as $d^2\varepsilon/dk^2$. On change of sign for k and ε, m^* changes the sign, due to the second derivative being equal to $\frac{d}{dk}\frac{d}{dk}\varepsilon$, where the sign change of k has no effect, while the sign change of ε changes the overall sign.

So, we have nicely dealt with the "dilemma" that the mass is negative[3], by defining a new particle, **hole**. It is easy to see that when two or more electrons are missing, then the state can be thought of as two holes or more holes, respectively.

Here are important diagrams for the one electron missing in a band. On the left is an "electron energy diagram." On the right is a "hole energy diagram." Keep in mind that most energy diagrams that you encounter will be the electron kind. When you go from the electron diagram to the hole diagram, you need to flip both k and ε. When this is done, you can see that the hole move in the opposite way to the electron!

---

[3] Another way to avoid the negative mass is to consider the "hole" as a time-reversed state of the electron, a la Feynman.



A hole is a real physical entity, not just some mathematical convenience. What do I mean by this? Suppose you are given two materials. One has holes in them, while the other has electrons in them. What I mean is that there are clearcut differences in certain measurements. Thermo-power ("hot probe" measurement), or Hall effect measurements are such examples. In the first case, the sign of "thermopower"[4] gives the sign of the charge carrier. In the second case, the sign of "Hall coefficient" (R_H)[5] determines the sign of the charge carrier. Please refer to the notes that you have taken during the lecture on these points. (However, change $\Delta T \rightarrow \nabla T$ for the thermopower definition.)

In fact, a hole is as real as a positron is real! One may also say that a hole in a semiconductor (in solid state physics) is concrete conceptual underpinning for a position in vacuum (in particle physics).

7.3 Electron

So, we have a new particle – a "hole" particle. Also, you must keep in mind that the "electron" that we speak of is actually unlike the "bare" electron that is found in vacuum. All this is because a crystalline environment is effectively modifying the nature of the electron.

The "electron" that we speak of is quite different from the electron in vacuum, which we call a "bare electron." This is because the wave function in a crystal is a combination of all Bragg-diffracted waves. In some crude analogy, consider yourself in

---

[4] The thermopower is the generated voltage divided by the thermal gradient applied.

[5] This is the transverse voltage development over a thin "Hall bar" of sample, divided by the product of the magnetic field and the longitudinal current applied.

a room filled with mirror. Your motion corresponds to the motion of the bare electron. However, when you move all your images move as well, and the total movement can be quite different from your movement alone!

It turns out that for semiconductors, the electron becomes "lighter" by a typical factor of 1/10. What I mean here is that the effective mass is typically 1/10 times the bare electron mass. More so in GaAs or InAs than in Si. So, the electron becomes more *mobile*

7.4 Effective Mass, Revisited

All in all, what we have learned so far is that **electrons and holes are distinct physical entities with their own effective masses**. In contrast, a bare electron is *not* a very useful concept! One can go so far as to say that a bare electron is not a physically meaningful object for a crystal. This is because low energy excited states must be thought of as being populated with electrons and holes, not bare electrons.

In terms of the band dispersion $\varepsilon(k)$, here is precisely how we define the electron effective mass m_n^* and the hole effective mass m_p^*. It is equivalent to the definition of Eq. 7.1, but here we just have more information about where the conduction band minimum is (at $k = k_b$) and where the valence band maximum is (at $k = k_t$). Also, in order for the definition of Eq. 7.1 to work well, we require that the dispersion to be (approximately) parabolic, which will be the case near k_b or k_t.

$$\varepsilon(k) = \varepsilon(k_b) + \frac{\hbar^2 (k - k_b)^2}{2m_n^*} \tag{7.5}$$

$$\varepsilon(k) = \varepsilon(k_t) - \frac{\hbar^2 (k - k_t)^2}{2m_p^*} \tag{7.6}$$

Keep in mind that all ε values here refer to as the electron energy. To get to the hole energy, we need to negate both ε and k values. Because the energy and the momentum are usually given from the electron view, what one has to do is then to mentally go to the hole view by doing these negations in one's mind.

Effective mass of Si (fraction of m_e): 0.16 (light hole), 0.52 (heavy hole), 0.19 (electron along $\langle 100 \rangle$), 0.93 (electron, when moving perpendicular to $\langle 100 \rangle$). The anisotropy in the electron case is an example of the "tensor" character of the effective mass being important. For GaAs: 0.082 (light hole), 0.5 (heavy hole), 0.066 (electron).

Note that, as mentioned in page 5 of LN 06, the crystal momentum $\hbar k$ is not necessarily equal to mass times velocity, in general. However, now that we have defined the effective mass for the bottom of the conduction band and the top of the valence band, we can see that for *these* regions the crystal momentum is equal to mass times group velocity in the following sense:

$$m_n^* v_{g,n} = \hbar(k - k_b) \qquad \text{Crystal momentum for electron} \qquad (7.7)$$

$$m_p^* v_{g,p} = -\hbar(k - k_t) \qquad \text{Crystal momentum for hole} \qquad (7.8)$$

These are derived from, Eqs. 7.5 and 7.6, and the definition of the group velocity $\hbar v_g = d\varepsilon/dk$.

7.5 Fermi Function

The "quasi-particles" such as electrons and holes in a semiconductor are important since they play fundamental roles in describing low energy excited states relevant for thermal and transport properties.

The Fermi function

$$f(E) = \frac{1}{\exp\left[(E-\mu)/(k_B T)\right] + 1} \qquad (7.9)$$

is an important quantity for considering the thermally excited states.

The physical meaning of this is **the average occupancy of a unique quantum state for a Fermion characterized by energy E.** "Unique" in the sense that if there are more than one states with the same energy (e.g. with different spins), then $f(E)$ applies to each of them. "Fermion" means a particle like an electron or a hole. "Occupancy" is literally that – how many fermions occupy the state. Note that mathematically $0 < f(E) < 1$ at any finite temperature. The fact that $f(E)$ cannot be greater than 1 is due to the Pauli exclusion principle.

T is the temperature. k_B is the Boltzmann constant. It is 26 meV / 300 K $\approx$ 8.6E-5 eV/K. I like you to use the former way of remembering (RT corresponds to $\approx$ 26 meV in energy).

Here, μ is the so-called the **chemical potential**. It is defined as the energy[6] gain

---

[6] Strictly speaking this is the Gibbs free energy not just the energy.

or loss when a particle is added or removed from the system. We will say a few more words about this important quantity later.

In order to examine this function, it is useful to consider the following function

$$h(x) = \frac{1}{\exp(x) + 1} \tag{7.10}$$

The first thing to note is that $h(x) \to 0$ as $x \to \infty$ and $h(x) \to 1$ as $x \to -\infty$. Also, $h(x = 0) = 1/2$. So, it is a function that starts at 1, and decreases to 0 as x increases, passing through 1/2 at $x = 0$.

It is helpful to consider the negative derivative of this function, $-h'(x)$. This function shows a peak centered at $x = 0$. The width of the peak is about 4 (FWHM: full width at half maximum). What this means is that $h(x)$ is changing rapidly for $|x| < 2$, but it is changing rather slowly for $|x| > 2$. Because the left (right) limiting value of $h(x)$ is 1 (0), this means that $h(x)$ is almost 1 (0) if $x \lesssim -2$ ($x \gtrsim 2$)). To put it more quantitatively, $h(x = -3) = 0.95$ and $h(x = 3) = 0.05$. **Correction rel. to lecture:** it is not 0.01, but 0.05! So, one may say that, approximately, if $|x| > 3$, then $h(x) \approx 1$ (negative x) or 0 (positive x), while, if $|x| < 3$, then $h(x)$ is quite different from 1 or 0.

Coming back to $f(E)$, this means that when $|E - \mu| \lesssim 3k_BT$, there is a significant departure of $f(E)$ from 0 or 1, while if $|E - \mu| \gtrsim 3k_BT$, then the function basically remains to be 0 or 1.

Why is this consideration important? Firstly, at $T = 0$, it is easy to see that $f(E) = 0$ if $E > \mu$ and $f(E) = 1$ if $E < \mu$. (At $E = \mu$, we still have $f(E = \mu) = 1/2$.) What does this mean? It means at $T = 0$, a state is either definitely filled with an electron $f(E) = 1$ or definitely empty ($f(E) = 0$), unless the energy happens to be at the boundary (μ). As we shall see, in an (intrinsic) semiconductor, the chemical potential lies between the valence band (VB) and the conduction band (CB), and thus at $T = 0$, the valence band is completely filled, and the conduction band is completely empty. In other words, at $T = 0$, there is no electron or hole! There cannot be any action at $T = 0$!

Now, turn on the temperature, then $f(E)$ will depart from 1 ($E < \mu$) and 0 ($E > \mu$). This means some holes created for $E < \mu$, and some electrons created for $E > \mu$! However, the numbers of such particles will be very small if $|E - \mu| \lesssim 3k_BT$. This is the case of an "intrinsic" semiconductor, a pure crystal without any impurities. For instance in the Si crystal, electrons and holes are generated due to finite temperature. At 300 K, the electron density (which is equal to the hole density for an intrinsic semiconductor, due to charge neutrality) is given by 1×10^{10} per cm^3.

Recognizing that the volume per Si is 20 $\text{Å}^3 = 2\times10^{-23}$ cm^3, this carrier density corresponds to 2 $\times10^{-13}$ per Si, or one in 5×10^{12}, or 5 trillion! That is far too low a carrier density to do anything useful!