Notes for Lecture 5

Valence Band, Conduction Band

5.1 Energy Band – H crystal in 1D, cont.

Let us summarize what we figured out in the last lecture.

We learned that for the $1s$ band of a 1D H crystal, we can take neighboring "$1s$" orbitals orthogonal to each other (due to their being Wannier orbitals), and that the following two orbitals represent the states at the "energy band edges."

$$\psi_L = \frac{1}{\sqrt{N}} \left(\psi_1 + \psi_2 + \cdots + \psi_N\right)$$

$$\psi_H = \frac{1}{\sqrt{N}} \left(\psi_1 - \psi_2 + \psi_3 - \psi_4 + \cdots + (-1)^{N-1}\psi_N\right)$$

Here, we generalized the result from the last lecture so that N means any positive integer, not just a power of 2. This is correct according to the fully quantum mechanical result.

The energy separation between ψ_H and ψ_L, i.e. the **band width**, is given by $4t$.

Clearly the high energy state, ψ_H, has the periodicity $2a$, while the low energy state, ψ_L, has the periodicity a.

Here is a paradoxical statement.

ψ_L has the periodicity ∞ as well as a.

How is this possible? Well, here is how. Consider

$$\psi_k = \frac{1}{\sqrt{N}} \sum_{s=0}^{N-1} \exp(iksa)\psi_i \tag{5.1}$$

Here, s is the "site" index, $s = 0, 1, 2, \cdots$, each corresponding to exactly one H.

You will see that when we plug in $k = \pi/a$ (thus $ksa = s\pi$), we get ψ_H, as we expected. When we plug in $k = 2\pi/a$ (thus $ksa = 2\pi s$), we also get ψ_L, also as we expected. Well, we also get ψ_L, obviously when we plug in $k = 0$ (i.e. $\lambda = \infty$)!

This may seem like "just math" – let us think physically about this. Given the state ψ_H, it is rather easy to consider shortening the wavelength to approach ψ_L. The opposite, however, is equally easy. Namely, as we increase the wave length steadily, we see that the state will approach ψ_L also! In terms of the sign of the coefficients, $\psi_H \sim + - + - + - \cdots$. As we increase the wave length, we see that this pattern will evolve, e.g., to $+ + - - + + - - + + - - \cdots$ and $+ + + - - - + + + - - - \cdots$. As the wave length increases, the region for the first "positive sign block" will increase in size indefinitely. Furthermore, near $s = 0$, all ψ_s's will have a near-unit coefficient. This near $s = 0$ region will increase in size. I.e., in the infinite wave length limit, we get ψ_L!

This is not a coincidence, but the reflection of a deeper truth.

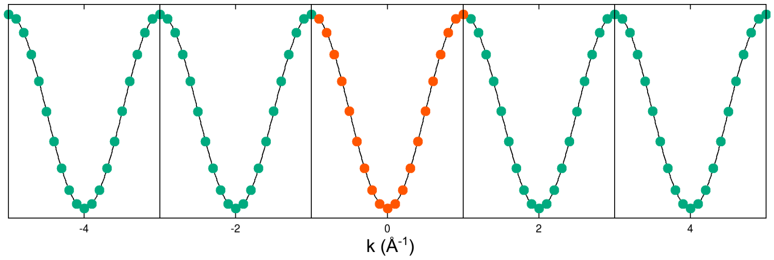
ψ_k and $\psi_{k+\frac{2\pi n}{a}}$ are physically indistinguishable for any wave vector k and any integer n, in a one dimensional crystal. This means that $\hbar k$ is ambiguous modulo $\hbar\ 2\pi/a$. This is due to the nature of a crystal environment. We call $\hbar k$ the **crystal momentum**.

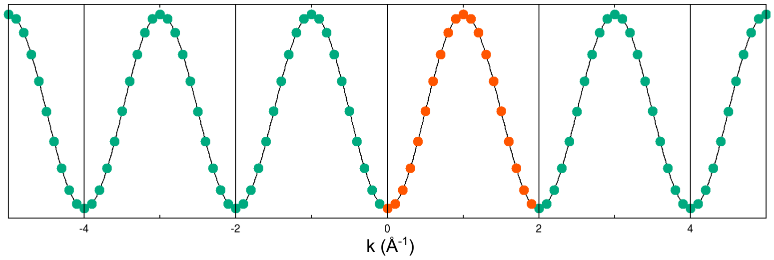
This is a very important fact[1]. At an intuitive level, it can be understood as a "k-space description" of a crystal. We defined a crystal as basis + lattice, i.e. something that repeats in real space. This means that a crystal is something that repeats in k

---

[1] The generalization of this to three dimensions is straightforward in the case of cubic crystal: we can simply apply the above to the three different components (k_x, k_y and k_z). For a more general case, refer to Kittel or any solid state physics book.

space also! The k space periodicity is given by $2\pi/a$ if a is the real space periodicity. This "periodicity in k space" is a fundamental property that must be satisfied by any physical wave function in a crystal[2]. Now, Eq. 5.1 is not just any function. It is indeed the exact tight binding wave function for the current simple 1s band problem in 1D H. It is left as an exercise to show explicitly that Eq. 5.1 does satisfy $\psi_k = \psi_{k+\frac{2\pi n}{a}}$.





The above figures must be studied very carefully. They illustrate the 1s tight binding band energy eigenvalues, corresponding to the eigenfunction Eq. 5.1, with $N = \infty$ (line) and $N = 20$ (points). You will find it very beneficial to *always* think N to be finite, however large it is, as it really is finite. Especially, in these days, when nano-crystals and nano-structures can be made routinely.

The above figure illustrates two ways to identify those unique k values. Because k and $k + 2\pi/a$ are equivalent (and so $k + 2\pi/a$ is not unique any more once k is chosen), the set of unique k values (red points) span the length $2\pi/a$, which is 2 Å^{-1} in this example. This span is the size of the unit cell in k space.

The above figure simply plots the eigenvalues corresponding to ψ_k of Eq. 5.1.

$$\varepsilon_k = -2t\cos(ka) \tag{5.2}$$

with a taken to be, just for convenience, π Å. In relation to what we did in the last lecture (section 4.2), we ignore the α term, which gives a uniform energy shift only. Note the following property, which can be deduced from what we did in the previous lecture (section 4.2).

---

[2] This is the famous "**Bloch's theorem**," a major principle of solid state physics. Its physical origin is Bragg diffraction.

For a H crystal consisting of N atoms, there are exactly N unique equally spaced k values per band. The 2nd column of the table in page 11 of LN 4 corresponds precisely to the number of unique k values. Each unique k point corresponds to a unique eigenfunction ψ_k that represent a unique eigenstate.

It follows that:

For a H crystal, the $1s$ band is exactly half-filled at $T = 0$.

This is because there is one electron per H atom and there is a spin degeneracy for each k. Here, T is the temperature. We specify $T = 0$, to avoid complication in the argument due to the thermal excitation of electrons[3].

5.2 Why Energy Band?

Why is the dispersion relationship like Eq. 5.2 called an "energy band"[4]? It can be understood in two ways.

First, notice that atomic orbitals form bands when the atomic orbitals in neighboring atoms have a significant overlap so that t – the hopping amplitude – is finite. In the limit where $t \to 0$ (e.g. imagine a very stretched out crystal), then, it can be seen that the energy levels will remain the same as the atomic orbital energy values. As the hopping amplitude becomes finite, atomic energy levels become "energy bands."

Second, it turns out a complimentary view is also possible. In this complementary view, one first turns off all the atomic potential. I.e., consider a free electron state. In this free electron state, all energy values are available – an infinite continuum of energy values starting from zero. However, as the atomic potentials of the crystal are "turned on," one can show that forbidden energy regions occur, and the free electron dispersion gets "chopped off" into pieces of dispersion relations, each of which is now periodic in $\vec{k}$ space and resembles the shape of the free electron dispersion only partially. This is another meaning of "energy bands."

---

[3] We will be talking much about the finite T effect in subsequent lectures

[4] Of course, one should not think that a real band is always as simple as Eq. 5.2. A real band is generally a rather complicated function of k.

These two views, starting from opposite limits, are in fact equivalent to each other in describing the energy band. I.e., when done correctly, they both give rise to the same correct energy bands. If you like to find out more, please read standard solid state physics textbook, such as Kittel.

5.3 Metal, Semi-Metal, Semi-Conductor, Insulator

The above theory gives us an excellent chance to have a precise definition for metals, insulators, and such. Notice that the distinction between metals and insulators is a triumphant achievement of quantum mechanics. In classical mechanics, any material with free charged particles, like valence electrons, must be metals.

A partially-filled band, like half-filled band here, at $T = 0$ means that the material is a **metal**. Cu, Al, etc. are examples. A partially filled band necessarily has a sharp boundary between the occupied states and the unoccupied states – the energy at which this happens is called the **Fermi energy**.

If all bands are completely filled with electrons at $T = 0$, and the next band is a finite energy ("energy gap") away from the top of the filled band, then the material is defined as an **insulator**. NaCl, LiF, etc. are examples. An important insulator in the semiconductor industry is SiO_2 – band gap is about 8 eV.

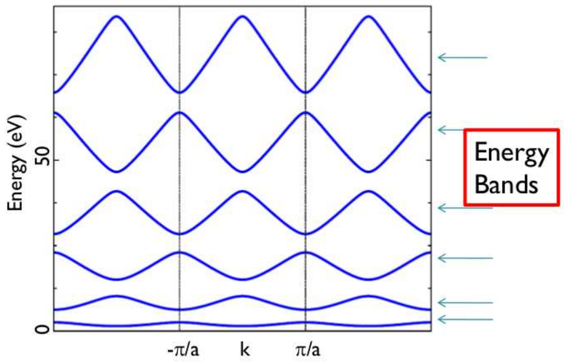
A **semiconductor** is an insulator with a modest gap ($\sim$ 1 eV or less). Si, Ge, GaAs, HgTe etc. are examples. C (diamond) can be viewed as belonging here, although it can be also called an insulator (band gap $\approx$ 5 eV).

Where is the Fermi energy for a semiconductor (or an insulator)? As we shall see, this is a *very* important question in terms of how semiconductor devices work. At this point, we are considering a pure crystal without any impurities. This is what is defined as an **intrinsic semiconductor**. Here, it suffices to note that Fermi energy is not well-defined, because there is no sharp boundary of energy below which all states are occupied and above which all states are unoccupied. Instead, any energy value that falls within the energy gap will do fine!

Finally, if all bands are completely filled with electrons at $T = 0$, but the next band continues on with zero energy gap, then the material is defined as a **semi-metal**. C (Graphene) is such an example.

The following figure shows a series of bands in a 1D crystal. As an exercise, mark the position of the Fermi energy for H, He, or Li. Keep in mind a possibility that the

Fermi energy is not defined uniquely.



From the above, we conclude that a H crystal in 1D would be a metal. Similarly, a He crystal in 1D would be an insulator, assuming the next band ($2s$ band or, better, $2sp$ bands) is at a finite energy gap away. And a Li crystal is a metal. Note that in reality these 1D crystals do not exist in Nature. However, the above theory in terms of predicting which material would be a metal and which an insulator turns out to be correct also for a 3D crystal of H, He, Li, if those crystals exist. 3D crystals of H and Li exist. Li is a metal but H is not. Crystals of He can exist only under pressure. They are indeed insulators. So, the simple theory got 2 out of 3! Why did the band theory go wrong for H? That is a long fascinating story, but is completely out of scope here. Here, it suffices to say that nobody said that the band theory is perfect. However, for all semiconductors that we will discuss, the band theory turns out to work very well!

5.4 Valence Band, Conduction Band

A **valence band** in a semiconductor means a band that originates from the atomic valence shell and that is completely filled at zero temperature.

A **conduction band** in a semiconductor means a band that originates from the atomic valence shell and that is is completely empty at zero temperature.

Note that these definitions are precise, because we are considering only an intrinsic semiconductor at this point.

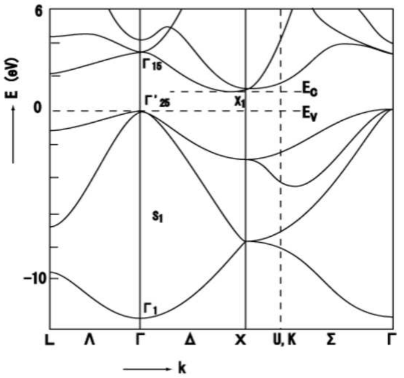
For instance, in Si, the valence shell consists of $3s$ and $3p$ orbitals. What happens is that these orbital mix to form sp^3 atomic orbitals, four "legs" of a tetrahedron formed around each Si. Then, for each Si-Si bonding, each leg is hybridized with another leg. This gives a strong σ bonding molecular orbital (MO). And a σ^* anti-bonding MO. Two electrons go into the σ MO while the σ^* MO is empty.

Now, it is worth pointing out a slight difference between the Si case and the 1D H case. In the latter case, all H atoms were equivalent. What this meant is that we could think of an electron in the *atomic* $1s$ orbital to hop between lattice points to make a band. In the Si case, however, remember how we we talked of a Si crystal as an fcc with two Si atoms (Lecture 1)? Well, what that means is that a Si crystal = two atom basis (Si_2) + lattice. In this case, what we must do is to think of the orbitals for Si_2 and make those orbitals to hybridize with those of the neighboring Si_2. In other words, the band that we get in a Si crystal better be called σ band and σ^* band. Both are sp^3 bands, but their molecular character is completely different. Since each Si_2 has 4 σ orbitals, and since Si_2 is the basis, there are a total of 4 σ bands. All of them are completely filled at $T = 0$. Likewise, there are 4 σ^* bands. They are completely empty at $T = 0$. This is why Si is a semiconductor. Also, why C (diamond), Ge, GaAs, InP, etc. are semiconductors.

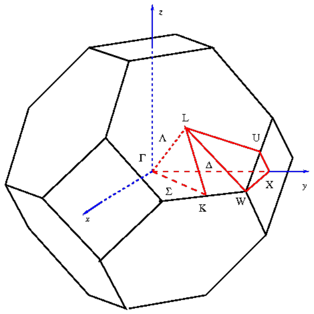
In this example of Si and related semiconductors, the σ bands are valence bands. The σ^* bands are conduction bands.

Here is the band structure for Si. It's got an "indirect gap" of magnitude 1.12 eV.

Band Structure of Si

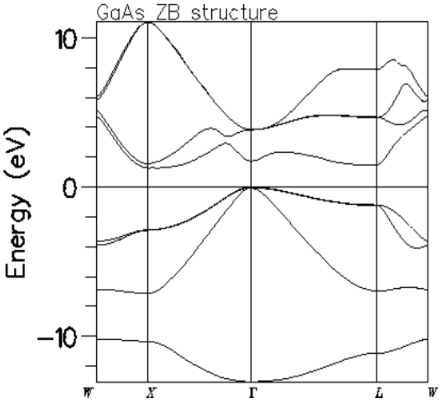


http://www2.warwick.ac.uk/

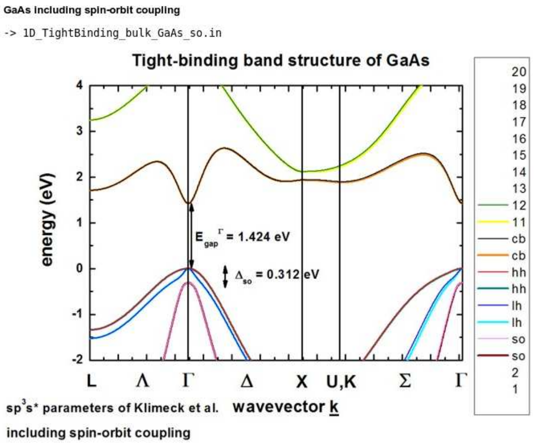


k space unit cell for fcc Brillouin zone

Here is the band structure for GaAs. It's got a "direct gap" of 1.42 eV. The one with the "spin-orbit splitting" is a realistic band structure.

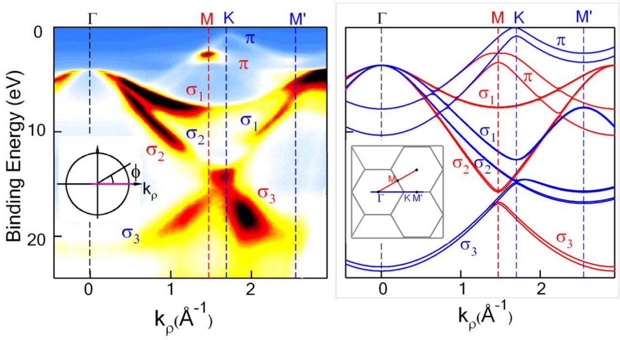


http://cmt.dur.ac.uk/sjc/thesis/thesis/node16.html

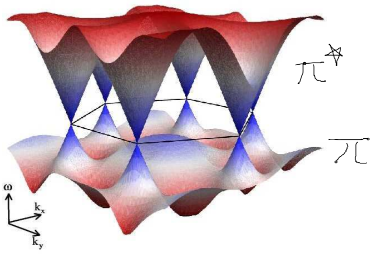


http://www.nextnano.de/nextnano3/tutorial/1Dtutorial_TightBinding_bulk_GaAs_GaP.htm

Below is the band structure of graphene – the top one is the real one that is measured by myself – measuring bands is part of my work. Graphene crystal = 2 atom basis + lattice. The bonding pattern is different between two C's: strong σ bonding between sp^2 orbitals, and weak π bonding between p_z orbitals. The π and π^* bands meet at the Fermi energy, while π bands are completely filled and π^* bands are completely empty at $T = 0$ – this peculiar electronic structure, which makes this material a semi-metal, is the origin of many interesting properties of graphene.



Zhou, GHG *et al*, Phys. Rev. B, 2005



π and π^* meet
at six points.
π: occupied
π^*: un-occupied

My technique (ARPES)
meaures occupied
states.

http://www.physics.leidenuniv.nl/institute/