

FoP3B Part II Lecture 1: Band gaps in Semiconductors

Without semiconductor materials modern day electronics would be impossible. A key physical property of a semiconductor is its **band gap**. The band gap is often used to determine if a given semiconductor is suitable for device applications, such as, for example, **solar cells** or **light emitting diodes (LED)**¹. Here we will explore the origin of band gaps in semiconductors, how it affects device applications and introduce the concept of **effective mass** of charge carriers (e.g. electrons).

Band gap and its origin

Consider a **free electron material** where an electron experiences a **uniform potential** (we are essentially ignoring the periodic potential of the atomic nuclei). The electron energy varies as k^2 , where k is the magnitude of the wavevector. When the periodic potential due to the atomic nuclei is introduced energy gaps appear at the **Brillouin zone boundaries** (Figure 1a). This is variously described as being due to Bragg reflection and interference of the $\pm\mathbf{G}/2$ Fourier components or lifting of the degeneracy of electronic states (FoP3B Part 1).

Now due to **Bloch's theorem** (FoP3B Part 1) the electronic structure can be fully described using only wavevectors within the first Brillouin zone. This gives rise to the **Reduced Zone scheme** and **electronic bands** for a given wavevector (Figure 1b).

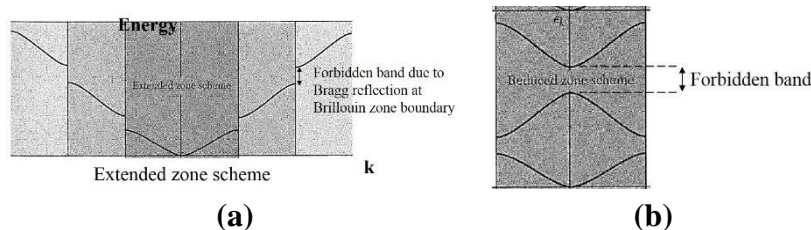


Figure 1: (a) Energy vs. wavenumber k for a periodic crystal and (b) the same diagram in the Reduced Zone scheme. The forbidden band is the semiconductor band gap.

The electronic bands are filled such that the energy is minimised, i.e. from the bottom up. Importantly it can be shown that **each primitive cell can only have two of its electrons in a given energy band**. Consider a typical semiconductor such as silicon. Silicon has an atomic number of 14 and therefore 14 electrons. With 2 atoms in a primitive cell this means that the first 14 electronic bands are completely full. The next available band is empty and separated by an energy gap. We call the highest *occupied* band the **valence band** and the lowest *unoccupied* band the **conduction band**. The **band gap** is the *minimum* energy separation between the two bands, and corresponds to the energy gap between the *valence band maximum* and *conduction band minimum*. **There are no electronic states within the band gap. Typical values for semiconductor band gaps are in the range 1-4 eV.**

Interaction of light with a semiconductor

The presence of a band gap means that a semiconductor interacts with light in a unique way. The key point to bear in mind is that **a photon has energy but negligible momentum**. When light is *absorbed* an electron in the solid is promoted to a higher energy level that must also be

¹ A solar cell converts sunlight into electricity while an LED does the opposite.

unoccupied. This means that **the minimum photon energy for light absorption is equal to the band gap**. If photons have lower energy then the light can only be *transmitted* through the solid or *reflected*. In a **direct band gap** semiconductor the valence band maximum has the same wavevector as the conduction band minimum (Figure 2a). For an electronic transition to take place in a direct band gap semiconductor only energy needs to be supplied, since the electron momentum is unchanged. These conditions are readily satisfied by photons of the appropriate energy, and consequently light absorption (as well as the reverse process of light emission) is strong in direct band gap semiconductors. This makes them useful for optoelectronic applications, such as solar cells and LEDs.

Not all semiconductors have a direct band gap. In some cases the wavevector for the conduction band minimum does not overlap with the valence band maximum (Figure 2b). These are known as **indirect band gap** semiconductors². An electronic transition here involves a change in electron momentum. To conserve momentum during light absorption **phonons** from the crystal must therefore be either created or destroyed. **Phonons have momentum but negligible energy (meV)**, which is why the phonon interaction is represented by a horizontal transition in Figure 2b (conversely the light interaction is a vertical transition). Since photons *and* phonons are involved light absorption and light emission are relatively weak in indirect bandgap semiconductors.

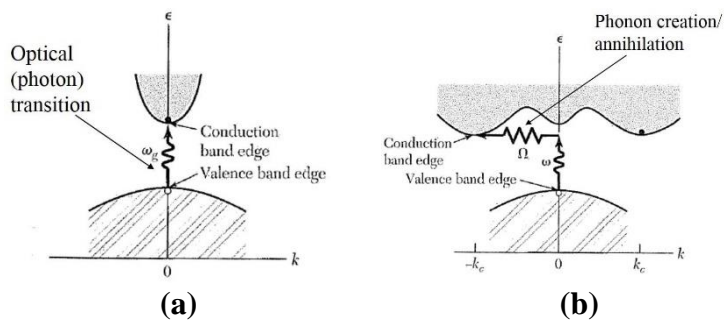


Figure 2: Light absorption in (a) direct and (b) indirect band gap semiconductor.

The band gap of semiconductor materials can usually be tuned to the desired energy through **alloying**, which is the process of mixing two different components to form a **solid solution**. For example, the alloy $\text{Al}_x\text{Ga}_{1-x}\text{As}$ can be formed by combining x fraction of AlAs with $(1-x)$ fraction of GaAs. The band gap (E_g) for an alloy of two components ‘A’ and ‘B’ usually has the following empirical relationship:

$$E_{g,AB} = xE_{g,A} + (1-x)E_{g,B} - bx(1-x)$$

where $E_{g,A}$ and $E_{g,B}$ are the band gaps for the pure components ‘A’ and ‘B’ and b is a constant known as the **bowing parameter**.

A further consideration in device applications is the **lattice parameter** of the semiconductor material. Frequently the active semiconductor material must be deposited as a nanometre or micrometre thick film on a support substrate (this is often due to the financial cost of producing such devices, but not always). If there is a large difference in lattice parameter between

² You might be wondering why the band structure in Figure 2b looks different to Figure 1b. The answer is that the latter assumes a weak periodic potential and is therefore typically accurate at only small values of k . The band structures of real materials are more complex!

substrate and film then significant strain can build up in the latter, leading to poorer quality devices. The lattice parameter (a) for an AB-alloy frequently obeys **Vegard's law**:

$$a_{AB} = xa_A + (1 - x)a_B$$

For an ideal optoelectronic device the semiconductor must have a direct band gap with energy that matches the intended application. A suitable substrate must also be found that is lattice matched to the semiconductor.

Effective mass of electrons

Charge transport is another key property governing device applications. An important concept is the **effective mass** of the charge carrying electrons. A simple way to think about effective mass is as follows. Consider first an electron in free space. We apply a force F and measure the acceleration a . The mass can then be determined from $F = ma$. Now consider an electron in a real crystal subject to the same external force. As the electron moves through the crystal it will undergo Coulomb scattering with other electrons as well as atomic nuclei. The presence of these *internal* forces means that $F \neq ma$ (recall that F is the *external* applied force). To satisfy Newton's second law we can therefore write $F = m^*a$, where m^* is the *effective* mass. It can be shown that for an isotropic crystal (see DUO supplementary material):

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

where d^2E/dk^2 is the curvature of electronic band (note that for anisotropic crystals we have to define an effective mass *tensor*).