Physics of Semiconductor Devices

Unit 2: Revision of Semiconductor Band Theory

Unit 2 Revision of Semiconductor Band Theory

Contents

Introduction

In the previous unit we discussed the concept of the band-structure of a crystalline material, using a diagram showing the relationship between an electron's energy E and its momentum $p = \hbar k$ (usually we drop the factor $\hbar$ and just talk of E-k diagrams). This enabled us to distinguish between metals, insulators, and semiconductors, depending on how many of the bands are occupied, and the fraction of the occupation.

In this unit we will now go on to examine the band-structure of real semiconductors, and discuss various experimental techniques that can be used to measure the semiconductor's band-gap.

Learning outcomes

After studying this unit you will be able to

- Explain the concept of the effective mass in semiconductors.
- Relate the effective mass to the semiconductor band-structure.
- Explain the concept of "holes" as positively charged particles in the valence band.
- Describe the difference between direct and indirect band-gap materials.
- Describe the key features of the band-structure of real semiconductors.
- Describe three different techniques for measuring the band-gap.

The Effective Mass

In Unit 1 we derived an ideal band diagram for a crystalline solid with a simple cubic lattice structure, using the nearly free electron model. Although we said that the electron interacts only weakly with the crystal lattice, now it is time to quantify how this interaction affects the electron. The detailed form of the $E(k)$ curve (which is different for different materials) determines the electron velocity and momentum – so it determines the electrical properties of the material.

Electrons in a vacuum

The energy of a free electron is related to the wavevector k by:

$$p = \hbar k \qquad\qquad\qquad \textbf{2-1.}$$

$$E = \frac{1}{2}m_e v^2 = \frac{p^2}{2m_e} = \frac{\hbar^2 k^2}{2m_e} \qquad\qquad\qquad \textbf{2-2.}$$
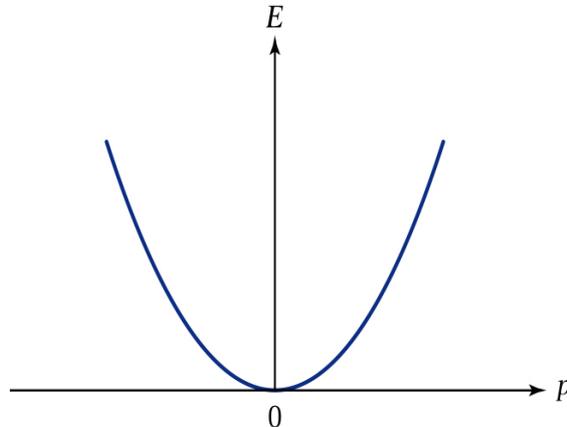

Figure 2.1: The parabolic energy (E) vs. momentum (p) curve for a free electron [reproduced from S.M. Sze, *Semiconductor Devices: Physics and Technology (2nd Edition),* (Figure 2.16), Wiley 2002].

Essentially what we are doing with these band diagrams is writing formulae for the electron velocity v and mass m in terms of the energy E and wavevector k :

For a free electron travelling as a wave, the group velocity of the electron wavepacket is given by

$$v = \frac{d\omega}{dk}.$$ **2-3.**

But $E = \hbar\omega$, so $v = \frac{1}{\hbar}\frac{\partial E}{\partial k}$ **2-4.**

In other words the electron velocity v depends on slope of $E(k)$.

We can also rewrite the electron mass as $m = \frac{\hbar^2}{d^2 E/dk^2}$. **2-5.**

Electrons in a crystal lattice

Now let us see how well the concepts of the electron mass and velocity work when the electron is travelling in a crystal. Firstly let's consider a more realistic band-structure, shown below. Note that the E(**k**) diagram is very similar to that for the free electron, except at the Brillouin Zone Boundaries, where it flattens off.
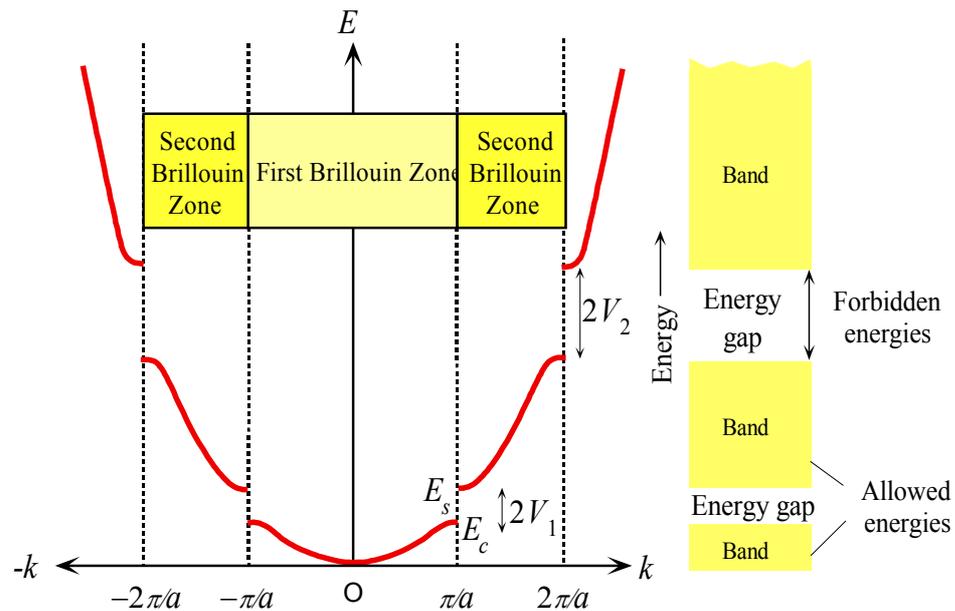

Figure 2.2: The energy of an electron as a function of its wavevector k inside a one-dimensional crystal. There are discontinuities in the energy at $k = \pm n\pi/a$ values where the waves suffer Bragg reflections in the crystal. There are energy gaps at these k values. Away from the critical k values, the $E-k$ behaviour is like that of a "free" electron, E increasing with k as $E = (\hbar k)^2/2m$.

Let's concentrate on the region near the centre of the zone, near k=0, and write an expression for *E(k)* by expanding about the minimum:

$$E = \frac{dE}{dk}\bigg|_0 k + \frac{1}{2}\frac{d^2E}{dk^2}\bigg|_0 k^2 + \frac{1}{6}\frac{d^3E}{dk^3}\bigg|_0 k^3 + \ldots \qquad \textbf{2-6.}$$

But we can drop all odd terms in k since the curve is symmetric. *i.e.*
$E(k) = E(-k)$

$$E \approx \frac{1}{2}\frac{d^2E}{dk^2}\bigg|_0 k^2 + O(k^4) \qquad \textbf{2-7.}$$

Therefore for small k, $E \approx \frac{1}{2}\frac{d^2E}{dk^2}\bigg|_0 k^2 \qquad \textbf{2-8.}$

In other words for small k the energy of electron in a semiconductor crystal has same quadratic form as free electron. So we write:

$$\frac{1}{2}\frac{d^2E}{dk^2}\bigg|_0 k^2 = \frac{\hbar^2 k^2}{2m^*} \qquad \textbf{2-9.}$$

where m^* **is the effective mass.** Note however that since the effective mass m^* depends on shape of *E-k* it will be different from the free electron m_e, and depends on how curved, or steep, the bands are around $k=0$, as shown in

Figure 2.3 below.

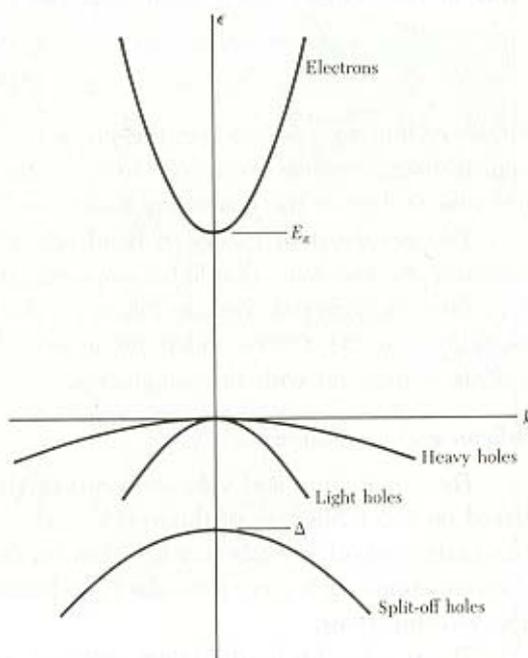$$\Rightarrow \qquad m^* = \frac{\hbar^2}{d^2E/dk^2\big|_0} \qquad \textbf{2-10.}$$

Figure 2.3: Schematic of the direct band-gap, conduction band, and valence bands. The curves in the valence band are called heavy hole, light hole, and split-off hole states, respectively. [reproduced from Kittel, *Introduction to Solid State Physics (7th Edition),* (Figure 8.13), Wiley 1995].

Electrons or Holes?

Before we move on to real band-structures, we will just consider the behaviour of electrons in the valence band. In a semiconductor almost all the allowed energy levels in the valence band are occupied by electrons, as shown in Figure 2.4 below
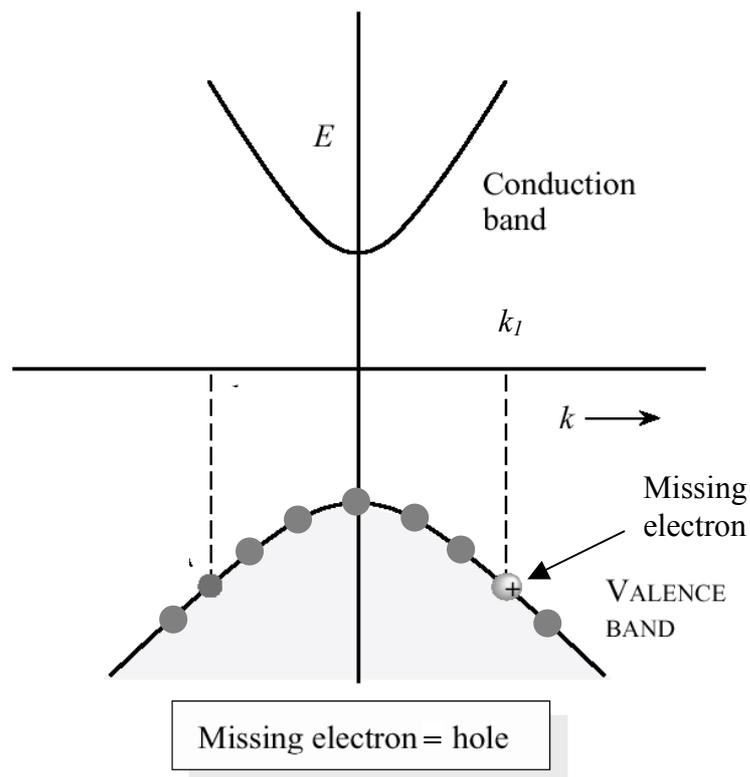

Figure 2.4: In the valance band, all states are filled except at the very top of the band, where electrical conduction can occur.

The usefulness of holes is that if we want to predict the optical or electrical properties, we don't have to consider the behaviour of all the electrons in the valence band – instead we only have to consider what happens to the comparatively few holes. This greatly simplifies the analysis of semiconductor devices. To quote from Sze:

> "The concept of a hole is analogous to that of a bubble in a liquid. Although it is actually the liquid that moves, it is much easier to talk about the motion of the bubble in the opposite direction."

In the valence band shown in Fig. 2.4 almost all the electrons moving in a given direction are "paired off" with electrons of equal and opposite k, and hence cannot contribute to any net current flow. The only exception is

the state at k_1. So of all the millions of electrons in the valence band, *only the electron/hole at k_1, and $-k_1$, contribute to the current I.*

$$I = 1(-|e|) \, \boldsymbol{v}(-k_1)$$
$$= +|e| \, \boldsymbol{v}(+k_1)$$

Remember that $v = \dfrac{1}{\hbar}\dfrac{dE}{dk}$ so $\boldsymbol{v}(-k_1) = -\boldsymbol{v}(+k_1)$

In other words the hole behaves as a positively charged particle with the velocity of the missing electron.

If we apply an electric field $E_x = +\beta\hat{x}$, directed to the right $\rightarrow$, (where $\hat{x}$ is a unit vector), how does a hole move?
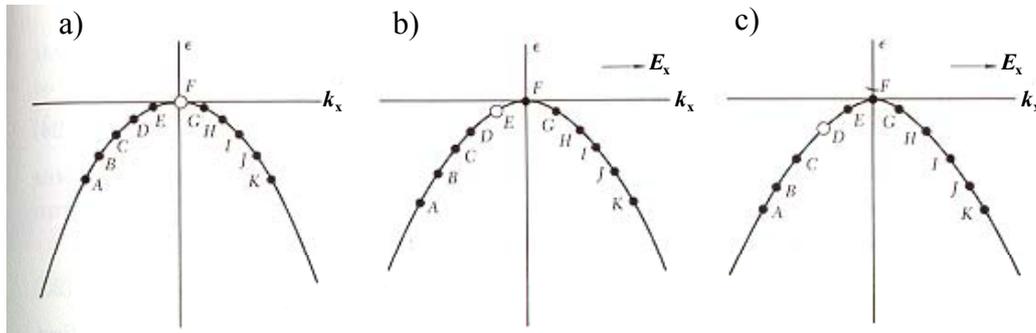

Figure 2.5: (a) At $t = 0$ all states are filled except F at the top of the band; the velocity v_x is zero at F because $dE_x/dk_x = 0$. (b) an electric field E_x is applied in the $\pm x$ direction. The force on the electrons is in the $-k_x$ direction and all electrons, make transitions together in the $-k_x$ direction, moving the hole into state E. (c) After a further interval the electrons move farther along in k space and the hole is now at D [reproduced from Kittel, *Introduction to Solid State Physics (7th Edition)*, Wiley 1995].

The force on the electrons is:

$$F = -|e|E_x = -|e|\beta, \quad \text{i.e. the force is to the left, } \leftarrow$$

So what happens to the electrons' momentum?

$$\text{Newton's 3}^{\text{rd}} \text{ law: } F = \frac{dp}{dt} = \hbar\frac{dk}{dt}$$

Force is negative $\Rightarrow$ electrons decrease k
                                  $\Rightarrow$ all electrons jump one state to the left
                                  $\Rightarrow$ so does the hole

In other words, the wavevector k of the hole decreases as well. What does this tell us about the velocity of the hole?

The hole wavepacket travels with a velocity $v = \dfrac{d\omega}{dk} = \dfrac{1}{\hbar}\dfrac{\partial E}{\partial k}$

For this hole in the valence band: $\quad \dfrac{\partial E}{\partial k} > 0$ (for negative k), and

$$\dfrac{\partial^2 E}{\partial k^2} < 0, \left(i.e. \dfrac{\partial v}{\partial k} < 0\right).$$

So in this band, as k becomes more negative, the velocity of hole *becomes more positive*, i.e. *v increases*.

$\Rightarrow$ The hole accelerates in the direction of the electric field.

$\therefore$ **The hole behaves as a *positively charged* particle with a *positive effective mass*:**

$$m^* = \dfrac{-\hbar^2}{d^2 E/dk^2\Big|_0} \qquad\qquad \textbf{2-12.}$$

Band-structure of Real Semiconductors

In the previous section we looked at an ideal band diagram for a crystalline solid with a simple cubic lattice structure. However things are more complex in real life, and every semiconductor has its own unique band-structure. In this course we are interested in the electronic and optical properties of the semiconductors, and so need only consider the structure of the conduction and valence bands.

Band-structure of Three Important Semiconductors

Figure 2.6 shows the band-structure of three important semiconductors, Ge, Si and GaAs, plotting the electron energy E against the wavevector k. Note that the two halves of each diagram do not match; the left shows wavevectors k (or electron momentum p) in the (111) crystal direction, while the right side refers to the (100) crystal direction (for a description of these different directions in the crystal lattice see §2.2.3 of Sze, *Crystal Planes and Miller Indices)*. There are several bands shown for each semiconductor, but the key features are all centred around the forbidden energy gap (or *band-gap*).
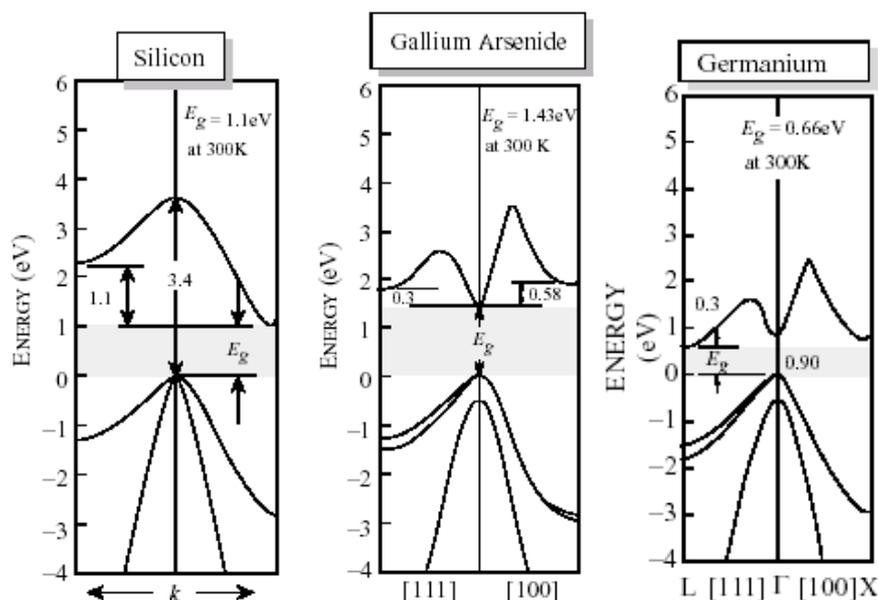

Figure 2.6: Realistic band-structures in semiconductors: Conduction and valence bands in Si, GaAs and Ge along [111] and [100]. [reproduced from Singh, *Semiconductor Devices: Basic Principles* (Figures 3.6-8), Wiley 2001]

Key features of the semiconductor band diagrams include:

- For the three materials shown above, there are two valence bands with different curvatures (i.e. different values of $\partial^2 E/\partial k^2$), and therefore different effective masses. The sharply curved band has a small effective mass, and holes in this band are called *light holes*. Conversely holes in the less curved band are referred to as *heavy holes*.

- Note that the curvature of the valence bands is different in the (111) and (100) directions – in other words the **hole effective mass varies with direction**. Although this difference is not normally important, it is exploited in devices such as p-type silicon strain gauges.

- The **band-gap** is defined as the energy difference between the lowest point of the conduction band and the highest point of the valence band, and is the single most important number for characterising a semiconductor. Large band-gaps are essential for making devices that operate at high temperatures, for the emission of visible light, and for very high resistance devices. Small band-gaps are useful for detecting low energy infra-red photons (thermal imaging), and for low voltage devices.

- **Direct and indirect band-gaps:** The relative position along the *k*-axis of the conduction band minimum and the valence band maximum determines whether a material has a direct or indirect band-gap. In materials such as GaAs the minimum in one band is directly above the maximum in the other (at *k*=0) – this is a direct gap material. In germanium and silicon the conduction band minimum is not directly above the valence band maximum. Direct band-gap materials are far more efficient at producing light than indirect gap materials, and are used in semiconductor light emitting diodes and solid state lasers.

Band-gaps and masses for different semiconductors

The effective masses of electrons and holes vary considerably between different semiconductors. For example, in silicon the electron effective mass is approximately $0.3m_e$ (where m_e is the free electron mass), whereas in GaAs it is $0.07\ m_e$. Since the electrons are lighter in GaAs, they can travel faster, and this makes GaAs an important material for high speed electronic devices (such as the microwave transistors in mobile phones and satellite communication systems).

In general holes have much larger effective masses than electrons. In GaAs the electron mass is $0.07\ m_e$, compared to masses of $m_{hh}^* \approx 0.40m_e$ and $m_{lh}^* \approx 0.09m_e$ for the heavy and light holes respectively. As we will see later, this is why electron based devices tend to be faster than hole based devices.

The table below summarises the band-gap properties of several different semiconductors. Note that as we move down the periodic table, the band-gap E_g decreases (Why might this be?).

Group	Material	E_g (eV)	Direct/Indirect
IV	C	5.47	I
	Si	1.14	I
	Ge	0.68	I
III-V	GaAs	1.42	D
	InSb	0.16	D
	GaP	2.24	I
II-VI	ZnS	3.60	D
	CdTe	1.6	D
	HgTe	0.0	D
	$Hg_xCd_{1-x}Te$	$0 - 1.6$	D

Measuring the Band-gap

We have seen that the band-gap is the most important physical constant when describing a semiconductor. In the following paragraphs we discuss different techniques for measuring the band-gap energy E_g.

1. Optical Absorption

This is a fairly simple technique. We just shine light on a thin sample and look at what comes out the other side.

Each incoming photon has energy $E = \hbar\omega$. Electrons in the conduction band can only absorb these photons if the photons can give them enough energy to excite them into the conduction band:
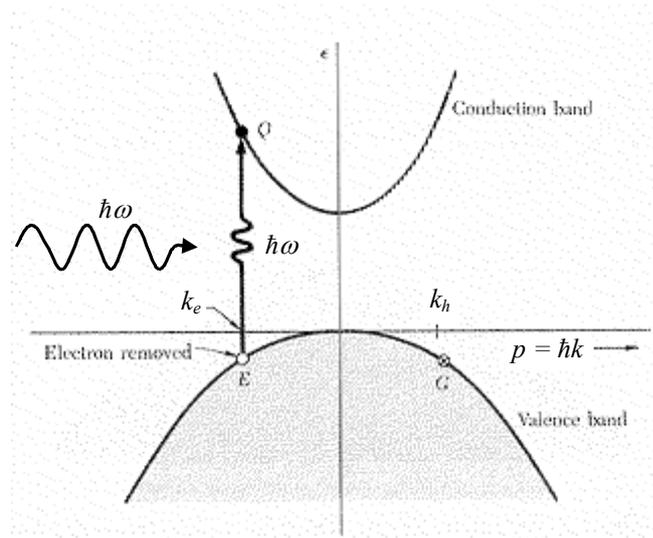

Figure 2.7: Absorption of a photon of energy $\hbar\omega$ and negligible wavevector takes an electron from state E in the filled valance band to state Q in the conduction band. If $\hbar\omega$ is less than the band-gap E_g no excitation can occur therefore absorption cannot take place. [reproduced from Kittel, *Introduction to Solid State Physics (7th Edition)*, (Figure 8.7), Wiley 1995]

So if the photon energy $\hbar\omega < E_g$ no excitation can occur $\Rightarrow$ no absorption (the photon passes straight through the semiconductor). Light can only be absorbed if $\hbar\omega \geq E_g$.

In the experiment, we shine light at intensity I_0 onto the sample, and measure the intensities of the transmitted (I_1) and reflected (I_2) light.
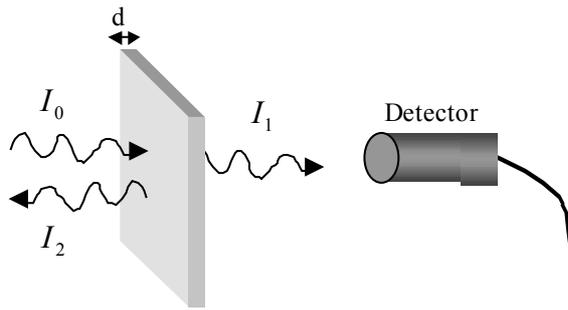
Then we normalise these to the intensity of the light sent in, and define the following two parameters:

$$R = \frac{I_1}{I_0},\qquad \text{Reflection Coefficient}\qquad\qquad \textbf{2-13.}$$

$$T = \frac{I_2}{I_0},\qquad \text{Transmission Coefficient}\qquad\qquad \textbf{2-14.}$$

As the light goes through the semiconductor, it will be absorbed, and the intensity will decay as $I = I_0\, e^{-\alpha x}$, where α **is called the absorption coefficient**.

For simplicity, we assume $\alpha d \gg 1$ (i.e. our sample is quite thick). We send in light at intensity I_0. Immediately some of it, RI_0, is reflected off the front face, so that only (1-$R)I_0$ is transmitted into the semiconductor. As this light travels through the semiconductor it is gradually absorbed, so that (1-$R)I_0\, e^{-\alpha d}$ reaches the back side of the semiconductor. But not all of this light escapes, due to reflection, so that the amount leaving is $(1-R)\times(1-R)I_0 e^{-\alpha d}$. This is sketched below:
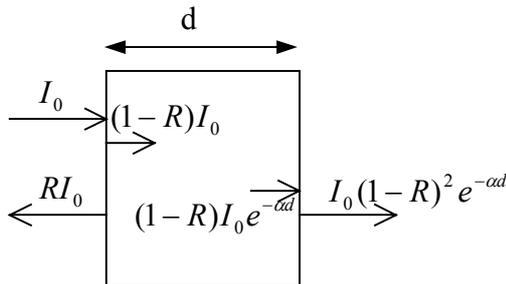
We can now use the reflection and transmission coefficients to extract the absorption coefficient α, which is what we are after:

$$T = \frac{I_2}{I_0} = (1-R)^2 e^{-\alpha d}\qquad\qquad\qquad \textbf{2-15.}$$

$$\Rightarrow \ln T = -\alpha d + \ln(1-R)^2\qquad\qquad\qquad \textbf{2-16.}$$

$$\Rightarrow \alpha d = \ln(1-R)^2 - \ln T\qquad\qquad\qquad \textbf{2-17.}$$

So by measuring *R* & *T,* and knowing the thickness of our sample, we obtain α:

$$\alpha = \frac{1}{d} \ln\left[\frac{(1-R)^2}{T}\right]$$

2-18.

Figure 2.8 below shows a schematic of how the absorption varies with the incident photon energy for a direct band-gap semiconductor.

- At very low energies ($\hbar\omega < E_g^{direct}$) no photons can be absorbed, so α=0.

- At $E_g^{direct} = \hbar\omega$ photons can only be absorbed at *k*=0. At higher energies, the absorption increases because there are more electrons available to be excited into the conduction band (in other words the *density of states* increases with increasing energy, as we will discuss in the next unit).

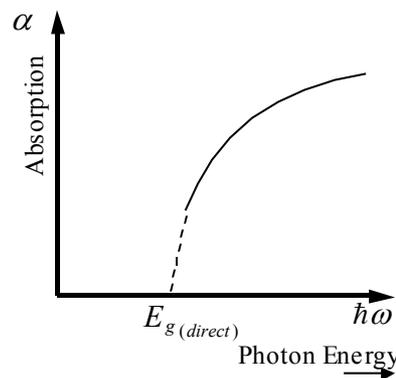The gap E_g is obtained by extrapolating the trace to α=0 (dotted line).

Figure 2.8: Optical absorption in direct band-gap materials. The absorption increases dramatically when the photon energy exceeds that required to create a free electron and a free hole.

Conservation of Crystal Momentum

We will now address the most important difference, in terms of optical properties, between direct and indirect band-gap materials.

Direct band-gap

When an electron in the valence band absorbs or emits a photon, the momentum of the electron-photon system must be conserved:

The momentum of an electron is $p = \hbar k$. So when the electron in an initial state $k_{initial}$ absorbs a photon with momentum $\hbar k_{photon}$, momentum conservation requires that:

Initial total momentum = final total momentum

$$\hbar k_{initial} + \hbar k_{photon} = \hbar k_{final} \qquad \textbf{2-19.}$$

But the photon carries very little momentum, much less than the momentum of the electron – i.e. $k_{photon} << k_{final}, k_{initial}$, so we can make the approximation that $k_{photon} \sim 0$

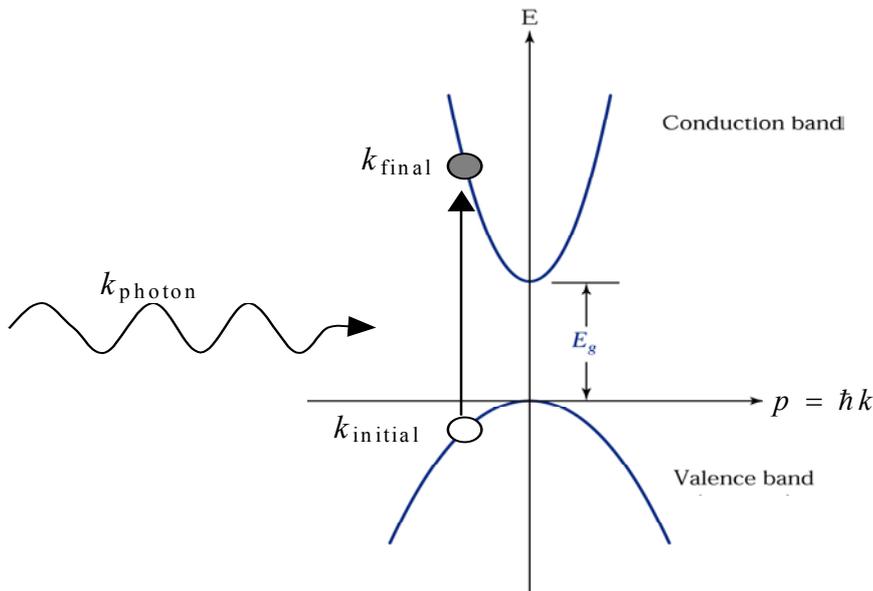$$\therefore \boxed{k_{initial} = k_{final}} \qquad \textbf{2-20.}$$

Figure 2.9: Photon absorption in a direct gap semiconductor. Momentum conservation requires that only vertical transitions are allowed.

Indirect Band-gap

For an indirect band-gap material, a photon with energy of $\hbar\omega = E_g$ cannot be absorbed, since that would promote a valence electron into the forbidden band. Because of momentum conservation, the only way for this photon to be absorbed is if the electron *simultaneously* absorbs a *photon* (which has much energy but little momentum) and a phonon (a vibration of the lattice, which has much momentum due to the large mass of the atoms in the lattice, but little energy). In this case:

$$\boxed{\hbar k_{initial} + \hbar k_{photon} = \hbar k_{initial}} \qquad (\text{since } k_{photon} \sim 0) \qquad \textbf{2-21.}$$

so $\boxed{k_{initial} + k_{phonon} = k_{final}}$ **2-22.**

(Note that an alternative way to get indirect semiconductors to emit and absorb light is to add impurities. Then photon absorption/emission occurs via the impurity level, which relaxes the momentum conservation rules.)
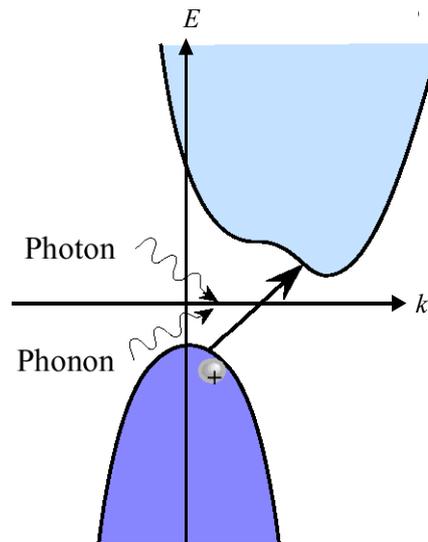

Figure 2.10: Indirect electron transitions in semiconductors [reproduced from Singh, *Semiconductor Devices: Basic Principles*, Wiley 2001].

Therefore the absorption for an indirect gap material has two components:

- At very low energies $(E_g^{indirect}, E_g^{direct} > \hbar\omega)$ no photons can be absorbed, so α=0.
- When the photon energy is larger than the indirect gap $(E_g^{direct} > \hbar\omega > E_g^{indirect})$ these is some absorption but it is weak, because if an electron to wants to absorb an incoming photon it also has to find a matching phonon to conserve momentum (or find an impurity atom).
- For very large photon energies, $(\hbar\omega > E_g^{direct})$, direct (vertical) transitions become possible, so the absorption increases steeply with energy as transitions can now occur without the need for phonons or impurities.
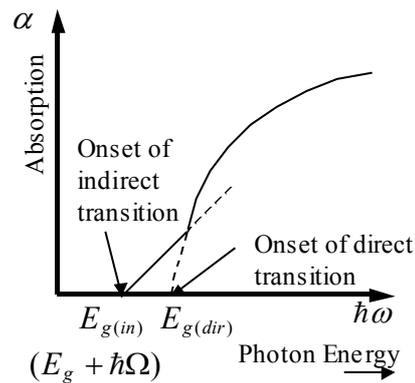
Figure 2.11: Dependence of optical absorption coefficient α on the energy of incident light, in an indirect band-gap semiconductor.

# 2. Electrical Conductivity vs. Temperature

In this technique we measure the resistance of the semiconductor as a function of temperature, and extract the band-gap E_g. In this section we will briefly discuss how this is done (although we will revisit many of the concepts such as probabilities, conductivities and mobilities in the next unit, so don't worry about the details too much at present).

For a pure, undoped (*intrinsic*) semiconductor, thermal energy can excite electrons from the completely full valence band into the conduction band. The probability of an electron being excited will depend on the band-gap E_g, and will increase with increasing temperature according to the equation: probability $\propto e^{-\Delta E/k_B T}$. Also, the number of thermally excited electrons n must be equal to the number of empty holes left in the valence band p:

$$n \approx p \approx A e^{-E_g/2k_B t} \quad \text{(the reason for the factor of 2 will be described in the next unit)}$$

If we apply a voltage V across the semiconductor, this creates an electric field $\vec{E}$. The current that flows as a result of this electric field is determined by the number of electrons and holes, and how fast they move:

$$\frac{I}{A} = j = nev_n + pev_p$$
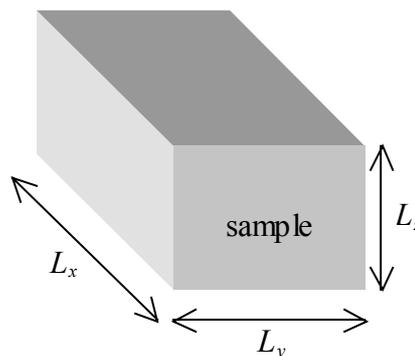
where $A = L_y \times L_z$ is the cross-sectional area of our piece of semiconductor, j is the current density, n and p are the electron and hole densities, and v_n and v_p are the electron and hole velocities. The velocity is related to the applied electric field $\mathbf{E}$ by a constant called the mobility μ:

$$\mathbf{v} = \mu\mathbf{E} \quad \text{(we will cover this in more detail later)},$$

So that the conductivity is:

$$\sigma = ne\mu_n + pe\mu_p \tag{2-23.}$$

In our intrinsic semiconductor $n = p$, so

$$\sigma = n(e\mu_e + e\mu_p) \tag{2-24.}$$

$$\sigma = Ae^{-E_g/2k_BT}(e\mu_e + e\mu_p) \tag{2-25.}$$

Now n and p depend exponentially on the temperature, whereas μ_n and μ_p have a much weaker dependence. So to a good approximation,

$$\ln\sigma \sim -\frac{E_g}{2k_BT} + \text{constant} \tag{2-26.}$$

In other words the conductivity mainly depends on the number of thermally excited carriers, which is exponentially sensitive to the temperature.
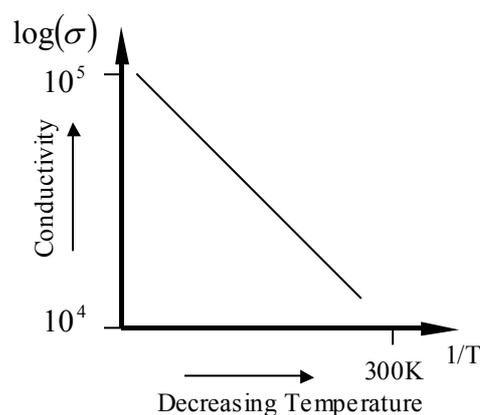

Figure 2.12: Temperature dependence of electrical conductivity in n-type germanium.

Therefore by measuring the conductivity as a function of temperature, we can obtain the band-gap E_g from the slope of a plot of $\log(\sigma)$ against $1/T$.

Exercise 2.1

Sketch (don't worry too much about the numerical values) $\frac{\partial E}{\partial k}$, $\frac{\partial^2 E}{\partial k^2}$, and m* for electrons in the conduction band of Si between L, Γ and X (use Figure 2.6 as a reference). Do the same for holes in the heavy hole valence band. Note that masses of hole and electrons are positive near $k=0$ in the conduction and valence bands respectively, this is the most important region of the band.

Exercise 2.2

Why are electronic transitions involving photons normally regarded as "vertical" while those involving phonons are normally taken to be "horizontal"?

_____

_____

_____

_____

_____

_____

_____

Exercise 2.3

Estimate the wavelengths of light at which Si, GaAs, and Ge become transparent. How big would the band-gap need to be in order that a semiconductor was transparent to the naked eye?

_____

_____

_____

_____

_____

_____

_____

Exercise 2.4

Photoluminescence is another technique by which the band-gap of a semiconductor may be measured. Light is shone onto the material ("photo"), exciting electrons from the valence band into the conduction band. The electrons relax to the lowest energy allowed in the conduction band by interacting with phonons. They then transit back into the valence band, emitting light as they do so ("-luminescence"), which can be analysed to find the transition energy. Interpret the photoluminescence spectrum below: what transitions do you think the two large peaks correspond to?
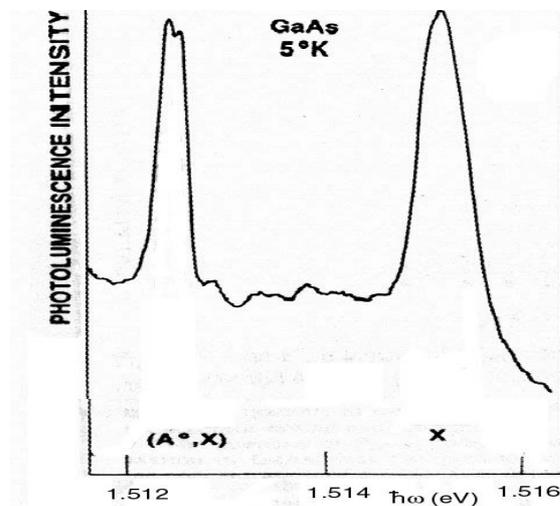
Readings

1. An unusual description of the concepts of band-structure, and some of the topics to be covered in the next unit, can be found at Britney's Guide to Semiconductor Physics:

    http://britneyspears.ac/lasers.htm

2. Further reading on the concepts covered in this unit can be found in:

    ➢ S.M. Sze, *Semiconductor Devices Physics and Technology,* Sections 2.1-2.5 (included in Unit 1)

    ➢ Kittel's *Introduction to Solid State Physics* (published by Wiley) also covers this material, at a slower rate.

Conclusion

In this unit we have revised several key concepts in semiconductor band theory:

- The concept of electrons and holes in semiconductors.
- The effective masses of electrons and holes in the conduction and valence bands.
- The differences between direct and indirect band-gap semiconductors, which has a fundamental impact on their use for optoelectronic applications.
- The band-structures of real semiconductors, and their anisotropy.
- Most significantly we covered different ways in which the band-gap of a semiconductor can be measured experimentally.

In Unit 3 we will go on to discuss how we can control the number of electrons and holes in a semiconductor (by doping), and derive the detailed equations that describe how the number of carriers depends on both the temperature and the doping level.