Contents

Handout 1

The Drude and Sommerfeld models of metals

This series of lecture notes covers all the material that I will present via transparencies during this part of the C3 major option. However, thoughout the course **the material that I write on the blackboard will not necessarily be covered in this transcript**. Therefore I strongly advise you to take good notes on the material that I present on the board. In some cases a different approach to a topic may be presented in these notes – this is designed to help you understand a topic more fully.

I recommend the book *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) as a primary textbook for this part of the course. Dr Singleton lectured this course for a number of years and the book is very good. You should also read other accounts in particular those in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976)and *Fundamentals of semiconductors*, by P. Yu and M. Cardona (Springer, Berlin, 1996). You may also find *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) and *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) useful. The lecture notes and both problem sets will be added to the website http://www-thz.physics.ox.ac.uk/cmpoption.html as the lectures are given. You will find that this first lecture is mainly revision, but it will get you ready from the new material in the following lectures.

We start the course by examining metals, the class of solids in which the presence of charge-carrying electrons is most obvious. By discovering why and how metals exhibit high electrical conductivity, we can, with luck, start to understand why other materials (glass, diamond, wood) do not.

1.1 What do we know about metals?

To understand metals, it is useful to list some of their properties, and contrast them with other classes of solids. I hope that you enjoy the course!

1. The metallic state is favoured by elements; $> \frac{2}{3}$ are metals.

2. Metals tend to come from the left-hand side of the periodic table; therefore, a metal atom will consist of a rather tightly bound "noble-gas-like" ionic core surrounded by a small number of more loosely-bound valence electrons.

3. Metals form in crystal structures which have relatively large numbers $n_{\rm nn}$ of nearest neighbours, *e.g.*

   hexagonal close-packed $n_{\rm nn} = 12$;

   face-centred cubic $n_{\rm nn} = 12$

   body-centred cubic $n_{\rm nn} = 8$ at distance d (the nearest-neighbour distance) with another 6 at $1.15d$.

   These figures may be compared with typical ionic and covalent systems in which $n_{\rm nn} \sim 4 - 6$).

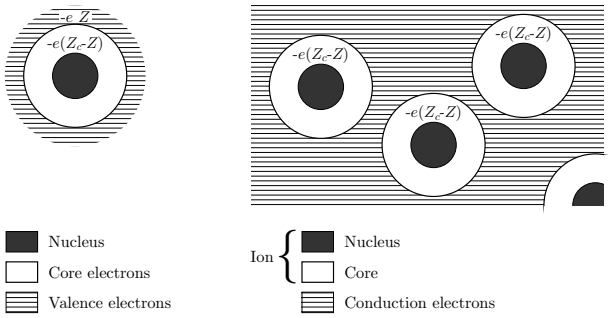


Figure 1.1: Schematic representations of a single, isolated metal atom and a solid metal. The atom (left) consists of a nucleus (size greatly exaggerated!) of charge $+eZ_c$ surrounded by the core electrons, which provide charge $-e(Z_c - Z)$, and Z valence electrons, of charge $-eZ$. In the solid metal (right), the core electrons remain bound to the nucleus, but the valence electrons move throughout the solid, becoming conduction electrons.

The large coordination numbers and the small numbers of valence electrons (~ 1) in metals imply that the outer electrons of the metal atoms occupy space between the ionic cores rather uniformly. This suggests the following things.

4. The bonds that bind the metal together are rather undirectional; this is supported by the malleability of metals. By contrast, ionic and covalent solids, with their more directional bonds (deduced from the smaller number of nearest neighbours mentioned above) are brittle.

5. There is a lot of "empty space" in metals; *e.g.* Li has an interatomic distance of 3 Å, but the ionic radius is only ~ 0.5 Å. This implies that there is a great deal of volume available for the valence (conduction) electrons to move around in.[1]

We therefore picture the metal as an array of widely spaced, small ionic cores, with the mobile valence electrons spread through the volume between (see Figure 1.1). The positive charges of the ionic cores provide charge neutrality for the valence electrons, confining the electrons within the solid (see Figure 1.2).

1.2 The Drude model

1.2.1 Assumptions

The Drude model was the first attempt to use the idea of a "gas" of electrons, free to move between positively charged ionic cores; the assumptions were

- a *collision* indicates the scattering of an electron by (and only by) an ionic core; *i.e.* the electrons do not "collide" with anything else;
- between collisions, electrons do not interact with each other *(the independent electron approximation)* or with ions *(the free electron approximation)*;
- collisions are instantaneous and result in a change in electron velocity;
- an electron suffers a collision with probability per unit time τ^{-1} (*the relaxation-time approximation)*; *i.e.* τ^{-1} is the scattering rate;
- electrons achieve thermal equilibrium with their surroundings only through collisions.

[1] This is actually one of the reasons why solid metals are stable; the valence electrons have a much lower zero-point energy when they can spread themselves out through these large volumes than when they are confined to one atom.

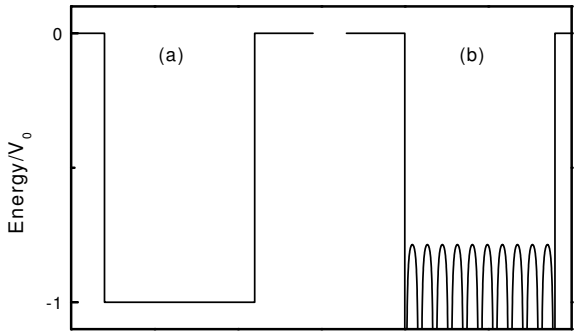


Figure 1.2: Schematic of the electronic potential assumed in the Drude and Sommerfeld models of metals (a). The small-scale details of the potential due to the ionic cores (shown schematically in (b)) are replaced by an average potential $-V_0$. In such a picture, the ionic cores merely maintain charge neutrality, hence keeping the electrons within the metal; the metal sample acts as a "box" containing electrons which are free to move within it.

These assumptions should be contrasted with those of the kinetic theory of conventional gases.[2] For example, in conventional kinetic theory, equilibrium is achieved by collisions *between* gas molecules; in the Drude model, the "molecules" (electrons) do not interact with each other at all!

Before discussing the results of the Drude model, we examine how the scattering rate can be included in the equations of motion of the electrons. This approach will also be useful in more sophisticated treatments.

1.2.2 The relaxation-time approximation

The current density $\mathbf{J}$ due to electrons is

$$\mathbf{J} = -ne\mathbf{v} = -\frac{ne}{m_e}\mathbf{p}, \tag{1.1}$$

where m_e is the electron mass, n is the electron density and $\mathbf{p}$ is the average electron momentum. Consider the evolution of $\mathbf{p}$ in time δt under the action of an external force $\mathbf{f}(t)$. The probability of a collision during δt is $\frac{\delta t}{\tau}$, *i.e.* the probability of surviving without colliding is $(1-\frac{\delta t}{\tau})$. For electrons that don't collide, the increase in momentum $\delta\mathbf{p}$ is given by[3]

$$\delta\mathbf{p} = \mathbf{f}(t)\delta t + O(\delta t)^2. \tag{1.2}$$

Therefore, the contribution to the average electron momentum from the electrons that do not collide is

$$\mathbf{p}(t+\delta t) = (1-\frac{\delta t}{\tau})(\mathbf{p}(t) + \mathbf{f}(t)\delta t + O(\delta t)^2). \tag{1.3}$$

Note that the contribution to the average momentum from electrons which *have* collided will be of order $(\delta t)^2$ and therefore negligible. This is because they constitute a fraction $\sim \delta t/\tau$ of the electrons and because the momentum that they will have acquired since colliding (each collision effectively randomises their momentum) will be $\sim \mathbf{f}(t)\delta t$.

Equation 1.3 can then be rearranged to give, in the limit $\delta t \to 0$

$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = -\frac{\mathbf{p}(t)}{\tau} + \mathbf{f}(t). \tag{1.4}$$

[2] See *e.g. Three phases of matter*, by Alan J. Walton, second edition (Clarendon Press, Oxford 1983) Chapters 5-7.

[3] The $O(\delta t)^2$ comes from the fact that I have allowed the force to vary with time. In other words, the force at time $t+\delta t$ will be slightly different than that at time t; *e.g.* if δt is very small, then $\mathbf{f}(t+\delta t) \approx \mathbf{f}(t) + (\mathrm{d}\mathbf{f}/\mathrm{d}t)\delta t$.

In other words, the collisions produce a frictional damping term. This idea will have many applications, even beyond the Drude model; we now use it to derive the electrical conductivity.

The electrical conductivity σ is defined by

$$\mathbf{J} = \sigma \mathbf{E}, \tag{1.5}$$

where $\mathbf{E}$ is the electric field. To find σ, we substitute $\mathbf{f} = -e\mathbf{E}$ into Equation 1.4; we also set $\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = 0$, as we are looking for a steady-state solution. The momentum can then be substituted into Equation 1.1, to give

$$\sigma = ne^2\tau/m_\mathrm{e}. \tag{1.6}$$

If we substitute the room-temperature value of σ for a typical metal[4] along with a typical $n \sim 10^{22} - 10^{23}\mathrm{cm}^{-3}$ into this equation, a value of $\tau \sim 1 - 10$ fs emerges. In Drude's picture, the electrons are the particles of a classical gas, so that they will possess a mean kinetic energy

$$\frac{1}{2}m_\mathrm{e}\langle v^2\rangle = \frac{3}{2}k_\mathrm{B}T \tag{1.7}$$

(the brackets $\langle\ \rangle$ denote the mean value of a quantity). Using this expression to derive a typical classical room temperature electron speed, we arrive at a mean free path $v\tau \sim 0.1 - 1$ nm. This is roughly the same as the interatomic distances in metals, a result consistent with the Drude picture of electrons colliding with the ionic cores. However, we shall see later that the Drude model very seriously underestimates typical electronic velocities.

1.3 The failure of the Drude model

1.3.1 Electronic heat capacity

The Drude model predicts the electronic heat capacity to be the classical "equipartition of energy" result[5] *i.e.*

$$C_\mathrm{el} = \frac{3}{2}nk_\mathrm{B}. \tag{1.8}$$

This is independent of temperature. Experimentally, the low-temperature heat capacity of metals follows the relationship (see Figures 1.3 and 1.4)

$$C_\mathrm{V} = \gamma T + AT^3. \tag{1.9}$$

The second term is obviously the phonon (Debye) component, leading us to suspect that $C_\mathrm{el} = \gamma T$. Indeed, even at room temperature, the electronic component of the heat capacity of metals is much smaller than the Drude prediction. This is obviously a severe failing of the model.

1.3.2 Thermal conductivity and the Wiedemann-Franz ratio

The thermal conductivity κ is defined by the equation

$$\mathbf{J}_q = -\kappa\nabla T, \tag{1.10}$$

where $\mathbf{J}_q$ is the flux of heat (*i.e.* energy per second per unit area). The Drude model assumes that the conduction of heat in metals is almost entirely due to electrons, and uses the kinetic theory expression[6] for κ, *i.e.*

$$\kappa = \frac{1}{3}\langle v^2\rangle\tau C_\mathrm{el}. \tag{1.11}$$

[4] Most metals have resistivities $(1/\sigma)$ in the range $\sim 1 - 20$ $\mu\Omega$cm at room temperature. Some typical values are tabulated on page 8 of *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976).

[5] See any statistical mechanics book (*e.g. Three phases of matter*, by Alan J. Walton, second edition (Clarendon Press, Oxford 1983) page 126; *Statistical Physics*, by Tony Guenault (Routledge, London 1988) Section 3.2.2).

[6] See any kinetic theory book, *e.g. Three phases of matter*, by Alan J. Walton, second edition (Clarendon Press, Oxford 1983) Section 7.3.

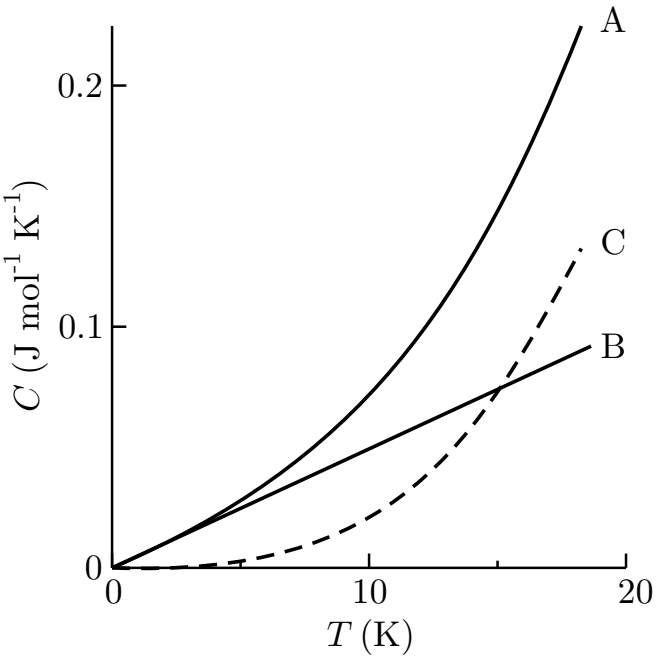


Figure 1.3: Heat capacity of Co at low temperatures. A is the experimental curve, B is the electronic contribution γT and C is the Debye component AT^3. (Data from G. Duyckaerts, *Physica* **6**, 817 (1939).)

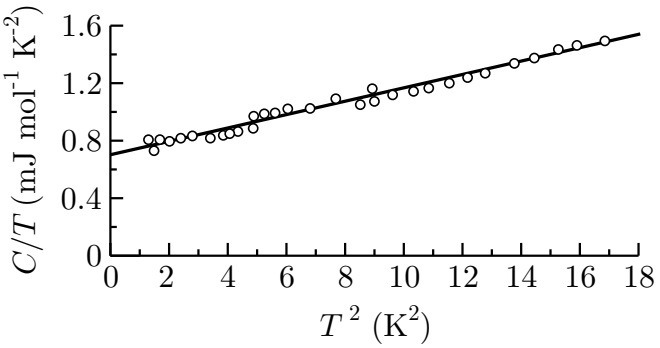


Figure 1.4: Plot of $C_{\rm V}/T$ versus T^2 for Cu; the intercept gives γ. (Data from W.S. Corak *et al.*, *Phys. Rev.* **98**, 1699 (1955).)

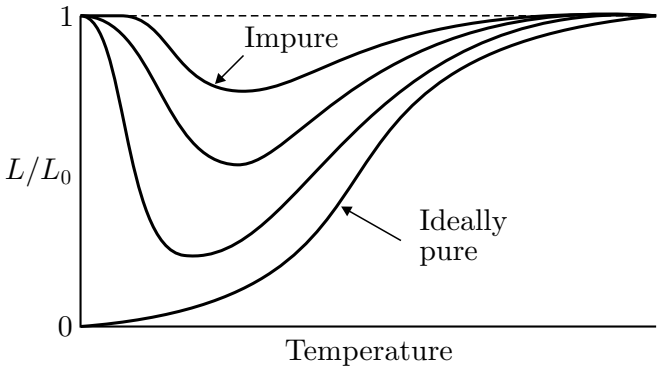


Figure 1.5: Schematic of experimental variation of $L = \kappa/(\sigma T)$ with temperature; L_0 is the theoretical value of L derived in later lectures. The high temperature limit of the figure represents room temperature.

$C_{\rm el}$ is taken from Equation 1.8 and the speed takes the classical thermal value, *i.e.* $\frac{1}{2}m_{\rm e}\langle v^2\rangle = \frac{3}{2}k_{\rm B}T$.

The fortuitous success of this approach came with the prediction of the Wiedemann-Franz ratio, κ/σ. It had been found that the Wiedemann-Franz ratio divided by the temperature, $L = \kappa/(\sigma T)$,[7] was close to the constant value $\sim 2.5 \times 10^{-8}\mathrm{W\Omega K^{-2}}$ for many metals at room temperature.[8] Substituting Equations 1.6 and 1.11 into this ratio leads to

$$\frac{\kappa}{\sigma T} = \frac{3}{2}\frac{k_{\rm B}^2}{e^2} \approx 1.1 \times 10^{-8}\mathrm{W\Omega K^{-2}}. \tag{1.12}$$

However, in spite of this apparent success, the individual components of the model are very wrong; *e.g.* $C_{\rm el}$ in the Drude model is at least two orders of magnitude bigger than the experimental values at room temperature! Furthermore, experimentally $\kappa/(\sigma T)$ drops away from its constant value at temperatures below room temperature (see Figure 1.5); the Drude model cannot explain this behaviour.

1.3.3 Hall effect

The Hall effect occurs when a magnetic field $\mathbf{B}$ is applied perpendicular to a current density $\mathbf{J}$ flowing in a sample. A *Hall voltage* is developed in the direction perpendicular to both $\mathbf{J}$ and $\mathbf{B}$. We assume that the current flows parallel to the x direction and that $\mathbf{B}$ is parallel to to z direction; the Hall voltage will then be developed in the y direction.

The presence of both electric and magnetic fields means that the force on an electron is now $\mathbf{f} = -e\mathbf{E} - e\mathbf{v}\times\mathbf{B}$. Equation 1.4 becomes

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\frac{e}{m_{\rm e}}\mathbf{E} - \frac{e}{m_{\rm e}}\mathbf{v}\times\mathbf{B} - \frac{\mathbf{v}}{\tau}. \tag{1.13}$$

In steady state, the left-hand side vanishes, and the two components of the equation read

$$v_x = -\frac{e\tau}{m_{\rm e}}E_x - \omega_{\rm c}\tau v_y \tag{1.14}$$

and

$$v_y = -\frac{e\tau}{m_{\rm e}}E_y + \omega_{\rm c}\tau v_x, \tag{1.15}$$

where we have written $\omega_{\rm c} = eB/m_{\rm e}$ ($\omega_{\rm c}$ is of course the classical cyclotron angular frequency). If we impose the condition $v_y = 0$ (no current in the y direction) we have

$$\frac{E_y}{E_x} = -\omega_{\rm c}\tau. \tag{1.16}$$

[7] $L = \kappa/(\sigma T)$ is often referred to as the *Lorenz number*.

[8] Typical values of κ and σ for metals are available in *Tables of Physical and Chemical Constants*, by G.W.C. Kaye and T.H. Laby (Longmans, London, 1966); find the book and check out the Lorenz number for yourself!

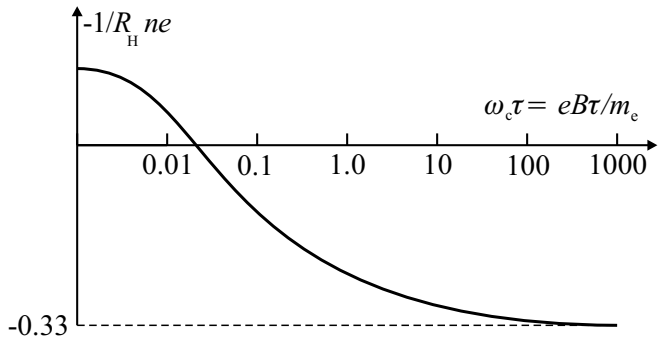


Figure 1.6: $-1/(R_{\rm H}ne)$ for Aluminium as a function of magnetic field.

Metal	Observed $R_{\rm H}$	Calculated $R_{\rm H}$	j
Li	-17.0	-13.1	1
Na	-25.0	-24.4	1
Cu	-5.5	-7.4	1
Ag	-8.4	-10.4	1
Zn	+4.1	-4.6	2
Cd	+6.0	-6.5	2

Table 1.1: Observed and calculated values of the Hall coefficient $R_{\rm H}$ in units of $10^{-11}{\rm m}^3{\rm C}^{-1}$ for several metals. The calculated values assume j free electrons per atom.

Writing $J_x = -envx$, Equations 1.14 and 1.16 can be combined to give

$$R_{\rm H} \equiv \frac{E_y}{J_x B} = -\frac{1}{ne}. \tag{1.17}$$

$R_{\rm H}$ is known as the *Hall coefficient*.

The Drude model therefore predicts the surprising fact that $R_{\rm H}$ is independent of both the magnitude of B and the scattering time τ. In reality, however, $R_{\rm H}$ is found to vary with magnetic field (see Figure 1.6) In many cases (see Table 1.1) the Hall coefficient is even *positive*!

There will an explanation of these effects in later lectures, where the Hall effect will be treated much more rigorously.

1.3.4 Summary

Although the Drude model was a brave attempt to apply classical Kinetic Theory to the electrons in a metal, it is plainly flawed in may respects. In order to start to remedy the most severe failings (*e.g.* the electronic heat capacity, the thermal conductivity) we must treat the electrons as quantum-mechanical particles (fermions), rather than the molecules of a classical gas.

1.4 The Sommerfeld model

1.4.1 The introduction of quantum mechanics

Quantum mechanics tells us that we can only have a certain number of states N per unit volume of phase space, where

$$N = (\frac{1}{2\pi})^j V_{kj} V_{rj}, \tag{1.18}$$

where j is the number of dimensions, V_{kj} is the k-space volume and V_{rj} is the r-space (*i.e.* real) volume. For example, in three dimensions

$$N = (\frac{1}{2\pi})^3 V_{k3} V_{r3}. \tag{1.19}$$

Now we assume electrons free to move within the solid; once again we ignore the details of the periodic potential due to the ionic cores (see Figure 1.2(b)),[9] replacing it with a mean potential $-V_0$ (see Figure 1.2(a)). The energy of the electrons is then

$$E(\mathbf{k}) = -V_0 + \frac{p^2}{2m_e} = -V_0 + \frac{\hbar^2 k^2}{2m_e}. \tag{1.20}$$

However, we are at liberty to set the origin of energy; for convenience we choose $E = 0$ to correspond to the average potential within the metal (Figure 1.2(a)), so that

$$E(\mathbf{k}) = \frac{\hbar^2 k^2}{2m_e}. \tag{1.21}$$

Let us now imagine gradually putting free electrons into a metal. At $T = 0$, the first electrons will occupy the lowest energy (*i.e.* lowest $|\mathbf{k}|$) states; subsequent electrons will be forced to occupy higher and higher energy states (because of Pauli's Exclusion Principle). Eventually, when all N electrons are accommodated, we will have filled a sphere of k-space, with

$$N = 2(\frac{1}{2\pi})^3 \frac{4}{3}\pi k_F^3 V_{r3}, \tag{1.22}$$

where we have included a factor 2 to cope with the electrons' spin degeneracy and where k_F is the k-space radius (the *Fermi wavevector*) of the sphere of filled states. Rearranging, remembering that $n = N/V_{r3}$, gives

$$k_F = (3\pi^2 n)^{\frac{1}{3}}. \tag{1.23}$$

The corresponding electron energy is

$$E_F = \frac{\hbar^2 k_F^2}{2m_e} = \frac{\hbar^2}{2m_e}(3\pi^2 n)^{\frac{2}{3}}. \tag{1.24}$$

This energy is called the *Fermi energy.*

The substitution of typical metallic carrier densities into Equation 1.24,[10] produces values of E_F in the range $\sim 1.5 - 15$ eV (*i.e.* $\sim$ atomic energies) or $E_F/k_B \sim 20 - 100 \times 10^3$ K (*i.e.* $\gg$ room temperature). Typical Fermi wavevectors are $k_F \sim 1/$(the atomic spacing) $\sim$ the Brillouin zone size $\sim$ typical X-ray k-vectors. The velocities of electrons at the Fermi surface are $v_F = \hbar k_F/m_e \sim 0.01c$. Although the Fermi surface is the $T = 0$ groundstate of the electron system, the electrons present are enormously energetic!

The presence of the Fermi surface has many consequences for the properties of metals; *e.g.* the main contribution to the bulk modulus of metals comes from the Fermi "gas" of electrons.[11]

1.4.2 The Fermi-Dirac distribution function.

In order to treat the thermal properties of electrons at higher temperatures, we need an appropriate distribution function for electrons from Statistical Mechanics. The Fermi-Dirac distribution function is

$$f_D(E,T) = \frac{1}{e^{(E-\mu)/k_B T} + 1}, \tag{1.25}$$

[9] Many condensed matter physics texts spend time on hand-waving arguments which supposedly justify the fact that electrons do not interact much with the ionic cores. Justifications used include "the ionic cores are very small"; "the electron–ion interactions are strongest at small separations but the Pauli exclusion principle prevents electrons from entering this region"; "screening of ionic charge by the mobile valence electrons"; "electrons have higher average kinetic energies whilst traversing the deep ionic potential wells and therefore spend less time there". We shall see in later lectures that it is the *translational symmetry* of the potential due to the ions which allows the electrons to travel for long distances without scattering from anything.

[10] Metallic carrier densities are usually in the range $\sim 10^{22} - 10^{23}\text{cm}^{-3}$. Values for several metals are tabulated in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976), page 5. See also Problem 1.

[11] See Problem 2; selected bulk moduli for metals have been tabulated on page 39 of *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976).

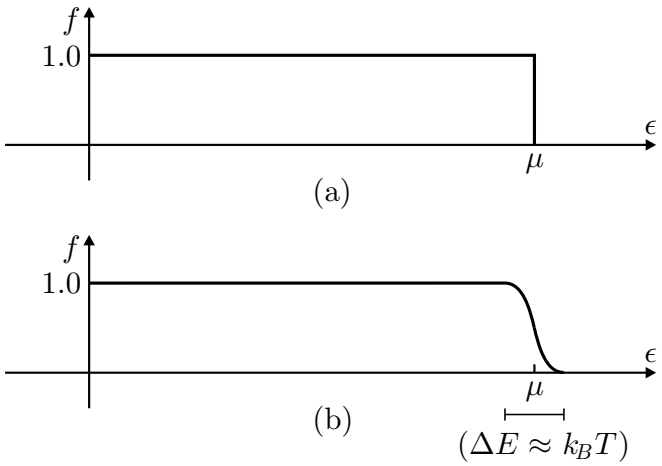


Figure 1.7: The Fermi-Dirac distribution function at $T = 0$ and at a finite temperature $T \ll E_{\rm F}/k_{\rm B}$.

where μ is the *chemical potential*; $f_{\rm D}$ gives the probability of occupation of a state of energy E. One of the definitions of the chemical potential[12] is that it is the energy at which the probability of occupation is $\frac{1}{2}$.

Figure 1.7 shows $f_{\rm D}$ at $T = 0$ and at a finite temperature $T \ll E_{\rm F}/k_{\rm B}$ (*e.g.* room temperature). Considering first the $T = 0$ figure, we see that

$$E_{\rm F} \equiv \mu(T = 0). \tag{1.26}$$

This defines $E_{\rm F}$.

Turning to the finite T figure, $f_{\rm D}$ only varies significantly within $k_{\rm B}T$ of μ. This has two implications.

1. As $k_{\rm B}T \ll E_{\rm F}$, this implies that $\mu \approx E_{\rm F}$.

   In much of what follows, we shall make use of Fermi-Dirac statistics. Strictly the chemical potential μ is the fundamental energy used in the Fermi–Dirac distribution function $f_{\rm D}$. However, for a typical metal at all accessible temperatures (*i.e.* until it melts) $\mu \approx E_{\rm F} \equiv \mu(T = 0)$. The substitution $\mu \to E_{\rm F}$ made in the thermodynamic analysis below (for convenience) results in minimal errors.

2. Only electrons with energies within $k_{\rm B}T$ of μ, *i.e.* $E_{\rm F}$, will be able to contribute to thermal processes, transport *etc.*. Electrons further below μ will be unable to acquire sufficient thermal energy to be excited into empty states; states more than $\sim k_{\rm B}T$ above μ will be empty.

1.4.3 The electronic density of states

It will be useful to have a function that describes the number of electron states in a particular energy range. We start by substituting a general value of k in Equation 1.22 instead of $k_{\rm F}$. This gives $n(\mathbf{k})$, the number of states per unit volume of r-space with wavevectors less than $|\mathbf{k}|$. For free electrons, with energy $E = \hbar^2\mathbf{k}^2/2m_{\rm e}$, one can then define a density of states, $g(E)$, where $g(E)\mathrm{d}E$ is the number of electrons per unit volume of r-space with energies between E and $E + \mathrm{d}E$:

$$g(E) \equiv \frac{\mathrm{d}n}{\mathrm{d}E} = \frac{\mathrm{d}n}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}} E^{\frac{1}{2}}. \tag{1.27}$$

[12] The chemical potential is often (confusingly) referred to as *the Fermi level*. Although you will see this usage in other books, I have tried to stick to *chemical potential*, to avoid confusion with *the Fermi energy*. Furthermore, I find the term "Fermi level" defective, in that it implies that there is a quantum-mechanical state (level) present at that energy. As we shall see when we come to treat semiconductors, there need be no such state at that energy.

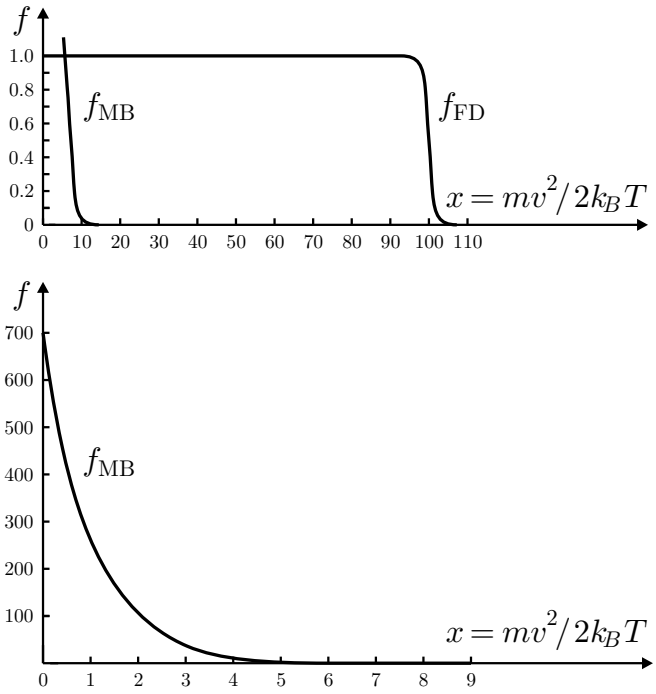


Figure 1.8: Comparison of the Maxwell–Boltzmann and Fermi–Dirac distribution functions using parameters relevant for a typical metal at room temperature. The concentration of filled states at low energies in the Maxwell–Boltzmann distribution is the reason why the Drude model fails to describe the thermal properties of metals.

1.4.4 The electronic density of states at $E \approx E_{\mathrm{F}}$

As only the electrons within $\sim k_{\mathrm{B}}T$ of E_{F} are able to take part in thermal processes, only the density of electron states at the Fermi energy, $g(E_{\mathrm{F}})$, will be particularly important in many calculations. In the free-electron approximation, the Fermi energy is given by $E_{\mathrm{F}} = \frac{\hbar^2}{2m_{\mathrm{e}}}(3\pi^2 n)^{\frac{2}{3}}$ (see Equation 1.24). Taking natural logarithms gives

$$\ln(E_{\mathrm{F}}) = \frac{2}{3}\ln(n) + \text{const.} \tag{1.28}$$

Differentiating (*i.e.* making infinitessimal changes in E_{F} and n) gives

$$\frac{\mathrm{d}E_{\mathrm{F}}}{E_{\mathrm{F}}} = \frac{2}{3}\frac{\mathrm{d}n}{n}. \tag{1.29}$$

We rearrange to obtain[13]

$$\frac{\mathrm{d}n}{\mathrm{d}E_{\mathrm{F}}} \equiv g(E_{\mathrm{F}}) = \frac{3}{2}\frac{n}{E_{\mathrm{F}}}. \tag{1.30}$$

1.4.5 The electronic heat capacity

Let $U(T)$ be the energy of the electron system at temperature T. At absolute zero, the electronic states are filled up to $E = E_{\mathrm{F}}$ and empty for higher energies. $U(0)$ can therefore be evaluated rather easily:

$$U(0) = \int_0^{E_{\mathrm{F}}} Eg(E)\mathrm{d}E = \frac{E_{\mathrm{F}}^{\frac{5}{2}}}{5\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}, \tag{1.31}$$

where $g(E)$ has been taken from Equation 1.27.

[13]Note that this expression contains only n and E_{F}; the mass of the electron and the spin degeneracy (factor 2) which feature in the prefactor in Equation 1.24 do not show up. This means that Equation 1.30 has a rather wider applicability than its simple derivation might suggest; remember this when you come to do some of the questions!

At finite temperature, electrons are excited into higher levels, and the expression for the energy of the electron system becomes slightly more complicated

$$U(T) = \int_0^\infty E\ g(E)\ f_{\mathrm{D}}(E,T)\mathrm{d}E$$

$$= \frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}\int_0^\infty \frac{E^{\frac{3}{2}}}{(\mathrm{e}^{(E-\mu)/k_{\mathrm{B}}T}+1)}\mathrm{d}E. \tag{1.32}$$

The integral in Equation 1.32 is a member of the family of so-called *Fermi–Dirac integrals*

$$F_j(y_0) = \int_0^\infty \frac{y^j}{\mathrm{e}^{(y-y_0)}+1}\mathrm{d}y. \tag{1.33}$$

There are no general analytical expressions for such integrals, but certain asymptotic forms exist. Comparison with Equation 1.32 shows that $y_0 \equiv \mu/k_{\mathrm{B}}T$; we have seen that for all practical temperatures, $\mu \approx E_{\mathrm{F}} \gg k_{\mathrm{B}}T$. Hence, the asymptotic form that is of interest for us is the one for y_0 very large and positive:

$$F_j(y_0) \approx \frac{y_0^{j+1}}{j+1}\left(1+\frac{\pi^2 j(j+1)}{6y_0^2}+O(y_0^{-4})+...\right) \tag{1.34}$$

Using Equation 1.34 to evaluate Equation 1.32 in the limit $\mu \gg k_{\mathrm{B}}T$ yields

$$U(T) = \frac{2}{5}\mu^{\frac{5}{2}}\{1+\frac{5}{8}\left(\frac{\pi k_{\mathrm{B}}T}{\mu}\right)^2\}\frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}. \tag{1.35}$$

We are left with the problem that μ is temperature dependent; μ is determined by the constraint that the total number of electrons must remain constant:

$$n = \int_0^\infty g(E)\ f_{\mathrm{D}}(E,T)\mathrm{d}E$$

$$= \frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}\int_0^\infty \frac{E^{\frac{1}{2}}}{(\mathrm{e}^{(E-\mu)/k_{\mathrm{B}}T}+1)}\mathrm{d}E. \tag{1.36}$$

This contains the $j = \frac{1}{2}$ Fermi–Dirac integral, which may also be evaluated using Equation 1.34 (as $\mu \gg k_{\mathrm{B}}T$) to yield

$$\mu \approx E_{\mathrm{F}}\{1-\frac{\pi^2}{12}\left(\frac{k_{\mathrm{B}}T}{\mu}\right)^2\}. \tag{1.37}$$

Therefore, using a polynomial expansion

$$\mu^{\frac{5}{2}} \approx E_{\mathrm{F}}^{\frac{5}{2}}\{1-\frac{5\pi^2}{24}\left(\frac{k_{\mathrm{B}}T}{\mu}\right)^2 .....\}, \tag{1.38}$$

Combining Equations 1.35 and 1.38 and neglecting terms of order $(k_{\mathrm{B}}T/E_{\mathrm{F}})^4$ gives

$$U(T) \approx \frac{2}{5}E_{\mathrm{F}}^{\frac{5}{2}}\{1-\frac{5\pi^2}{24}\left(\frac{k_{\mathrm{B}}T}{\mu}\right)^2\}\{1+\frac{5}{8}\left(\frac{\pi k_{\mathrm{B}}T}{\mu}\right)^2\}\frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}.$$

$$\approx \frac{2}{5}E_{\mathrm{F}}^{\frac{5}{2}}\{1+\frac{5}{12}\left(\frac{\pi k_{\mathrm{B}}T}{\mu}\right)^2\}\frac{1}{2\pi^2}\left(\frac{2m_{\mathrm{e}}}{\hbar^2}\right)^{\frac{3}{2}}. \tag{1.39}$$

Thus far, it was essential to keep the temperature dependence of μ in the algebra in order to get the prefactor of the temperature-dependent term in Equation 1.39 correct. However, Equation 1.37 shows that at temperatures $k_{\mathrm{B}}T \ll \mu$ (*e.g.* at room temperature in typical metal), $\mu \approx E_{\mathrm{F}}$ to a high degree of accuracy. In order to get a reasonably accurate estimate for $U(T)$ we can therefore make the substitution $\mu \to E_{\mathrm{F}}$ in denominator of the second term of the bracket of Equation 1.39 to give

$$U(T) = U(0) + \frac{n\pi^2 k_{\mathrm{B}}^2 T^2}{4E_{\mathrm{F}}}, \tag{1.40}$$

where we have substituted in the value for $U(0)$ from Equation 1.31. Differentiating, we obtain

$$C_{\rm el} \equiv \frac{\partial U}{\partial T} = \frac{1}{2}\pi^2 n \frac{k_{\rm B}^2 T}{E_{\rm F}}. \tag{1.41}$$

Equation 1.41 looks very good as it

1. is proportional to T, as are experimental data (see Figures 1.3 and 1.4);
2. is a factor $\sim \frac{k_{\rm B}T}{E_{\rm F}}$ smaller than the classical (Drude) value, as are experimental data.

Just in case the algebra of this section has seemed like hard work,[14] note that a reasonably accurate estimate of $C_{\rm el}$ can be obtained using the following reasoning. At all practical temperatures, $k_{\rm B}T \ll E_{\rm F}$, so that we can say that only the electrons in an energy range $\sim k_{\rm B}T$ on either side of $\mu \approx E_{\rm F}$ will be involved in thermal processes. The number density of these electrons will be $\sim k_{\rm B}T g(E_{\rm F})$. Each electron will be excited to a state $\sim k_{\rm B}T$ above its groundstate ($T = 0$) energy. A reasonable estimate of the thermal energy of the system will therefore be

$$U(T) - U(0) \sim (k_{\rm B}T)^2 g(E_{\rm F}) = \frac{3}{2} n k_{\rm B} T \left(\frac{k_{\rm B}T}{E_{\rm F}}\right),$$

where I have substituted the value of $g(E_{\rm F})$ from Equation 1.30. Differentiating with respect to T, we obtain

$$C_{\rm el} = 3 n k_{\rm B} \left(\frac{k_{\rm B}T}{E_{\rm F}}\right),$$

which is within a factor 2 of the more accurate method.

1.5 Successes and failures of the Sommerfeld model

The Sommerfeld model is a great improvement on the Drude model; it can successfully explain

- the temperature dependence and magnitude of $C_{\rm el}$;
- the approximate temperature dependence and magnitudes of the thermal and electrical conductivities of metals, and the Wiedemann–Franz ratio (see later lectures);
- the fact that the electronic magnetic susceptibility is temperature independent.[15]

It cannot explain

- the Hall coefficients of many metals (see Figure 1.6 and Table 1.1; Sommerfeld predicts $R_{\rm H} = -1/ne$);
- the magnetoresistance exhibited by metals (see subsequent lectures);
- other parameters such as the thermopower;
- the shapes of the Fermi surfaces in many real metals;
- the fact that some materials are insulators and semiconductors (*i.e.* not metals).

Furthermore, it seems very intellectually unsatisfying to completely disregard the interactions between the electrons and the ionic cores, except as a source of instantaneous "collisions". This is actually the underlying source of the difficulty; to remedy the failures of the Sommerfeld model, we must reintroduce the interactions between the ionic cores and the electrons. In other words, we must introduce the periodic potential of the lattice.

[14] The method for deriving the electronic heat capacity used in books such as *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) looks at first sight rather simpler than the one that I have followed. However, Kittel's method contains a hidden trap which is not mentioned; it ignores $({\rm d}\mu/{\rm d}T)$, which is said to be negligible. This term is actually of comparable size to all of the others in Kittel's integral (see *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapter 2), but happens to disappear because of a cunning change of variables. By the time all of this has explained in depth, Kittel's method looks rather more tedious and less clear, and I do not blame him for wishing to gloss over the point!

[15] The susceptibility of metals in the Sommerfeld model is derived in Chapter 7 of *Magnetism in Condensed Matter*, by S.J. Blundell (OUP 2000).

1.6 Reading.

A more detailed treatment of the topics in this Chapter is given in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapters 1-3 and *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) Chapter 1. Other useful information can be found in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Chapter 11, *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Sections 9.1-9.14, *Electrons in Metals and Semiconductors*, by R.G. Chambers (Chapman and Hall, London 1990) Chapters 1 and 2, and *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapters 6 and 7.

Handout 2

The quantum mechanics of particles in a periodic potential: Bloch's theorem

2.1 Introduction and health warning

We are going to set up the formalism for dealing with a periodic potential; this is known as *Bloch's theorem*. The next two-three lectures are going to appear to be *hard work* from a conceptual point of view. However, although the algebra *looks* complicated, the underlying ideas are really quite simple; you should be able to reproduce the various derivations yourself (make good notes!).

I am going to justify the Bloch theorem fairly rigorously. This formalism will then be used to treat two opposite limits, a very weak periodic potential and a potential which is so strong that the electrons can hardly move. You will see that both limits give qualitatively similar answers, *i.e.* reality, which lies somewhere in between, must also be like this!

For this part of the course these notes provide a slightly different (Fourier) approach to the results that I will derive in the lectures. I recommend that you work through and understand both methods, as the Bloch theorem forms the foundation on which the rest of the course is based.

2.2 Introducing the periodic potential

We have been treating the electrons as totally free. We now introduce a periodic potential $V(\mathbf{r})$. The underlying translational periodicity of the lattice is defined by the *primitive lattice translation vectors*

$$\mathbf{T} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{2.1}$$

where n_1, n_2 and n_3 are integers and $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$ are three noncoplanar vectors. [1] Now $V(\mathbf{r})$ must be periodic, *i.e.*

$$V(\mathbf{r}+\mathbf{T}) = V(\mathbf{r}). \tag{2.2}$$

The periodic nature of $V(\mathbf{r})$ also implies that the potential may be expressed as a Fourier series

$$V(\mathbf{r}) = \sum_{\mathbf{G}} V_{\mathbf{G}} \mathrm{e}^{\mathrm{i}\mathbf{G}.\mathbf{r}}, \tag{2.3}$$

where the $\mathbf{G}$ are a set of vectors and the $V_{\mathbf{G}}$ are Fourier coefficients.[2]

Equations 2.2 and 2.3 imply that

$$\mathrm{e}^{\mathrm{i}\mathbf{G}.\mathbf{T}} = 1, \ \ i.e. \ \mathbf{G}.\mathbf{T} = 2p\pi, \tag{2.4}$$

[1] See *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) pages 4-7.

[2] Notice that the units of $\mathbf{G}$ are the same as those of the wavevector $\mathbf{k}$ of a particle; $|\mathbf{k}| = 2\pi/\lambda$, where λ is the de Broglie wavelength. The $\mathbf{G}$ are vectors in k-space.

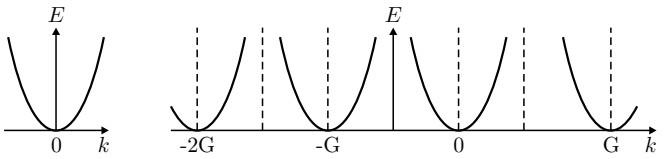


Figure 2.1: The effect of introducing the reciprocal lattice; instead of dealing with just one electron dispersion relationship (left) we have an infinite number of copies (right).

where p is an integer. As $\mathbf{T} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, this implies that

$$\mathbf{G} = m_1\mathbf{A}_1 + m_2\mathbf{A}_2 + m_3\mathbf{A}_3, \tag{2.5}$$

where the m_j are integers, and the $\mathbf{A}_j$ are three noncoplanar vectors defined by

$$\mathbf{a}_j.\mathbf{A}_l = 2\pi\delta_{jl}. \tag{2.6}$$

Take a few moments to convince yourself that this is the only way of defining $\mathbf{G}$ which can satisfy Equation 2.4 for all possible $\mathbf{T}$.

Using very simple reasoning, we have shown that the existence of a lattice in r-space automatically implies the existence of a lattice in k-space. The vectors $\mathbf{G}$ define the *reciprocal lattice*; the $\mathbf{A}_j$ are its primitive translation vectors.

The reciprocal lattice has extraordinary consequences for the electronic motion, even before we "switch on" the lattice potential. Instead of dealing with just one electron dispersion relationship $E(\mathbf{k})$ there must be an infinite number of equivalent dispersion relationships (see Figure 2.1) such that $E(\mathbf{k}) = E(\mathbf{k} + \mathbf{G})$ for all $\mathbf{G}$ (*c.f.* Equation 2.2). However, the k-space periodicity also implies that all information will be contained in the primitive unit cell of the reciprocal lattice, known as *the first Brillouin zone*.[3] The first Brillouin zone has a k-space volume

$$V_{k3} = \mathbf{A}_1.\mathbf{A}_2 \times \mathbf{A}_3. \tag{2.7}$$

2.3 Born–von Karman boundary conditions

We need to derive a suitable set of functions with which we can describe the motion of the electrons through the periodic potential; "motion" implies that we do not want standing waves. The functions should reflect the translational symmetry properties of the lattice; to do this we use *Born–von Karman periodic boundary conditions*.

We choose a plane wave

$$\phi(\mathbf{r}) = \mathrm{e}^{\mathrm{i}(\mathbf{k}.\mathbf{r}-\omega t)} \tag{2.8}$$

subject to boundary conditions which include the symmetry of the crystal

$$\phi(\mathbf{r} + N_j\mathbf{a}_j) = \phi(\mathbf{r}), \tag{2.9}$$

where $j = 1,\ 2,\ 3$ and $N = N_1N_2N_3$ is the number of primitive unit cells in the crystal; N_j is the number of unit cells in the jth direction.

The boundary condition (Equation 2.9) implies that

$$\mathrm{e}^{\mathrm{i}N_j\mathbf{k}.\mathbf{a}_j} = 1 \tag{2.10}$$

for $j = 1,\ 2,\ 3$. Comparing this with Equation 2.4 (and the discussion that follows it) suggests that the allowed wavevectors are

$$\mathbf{k} = \sum_{j=1}^{3} \frac{m_j}{N_j}\mathbf{A}_j. \tag{2.11}$$

[3]The first Brillouin zone is the *Wigner-Seitz primitive cell* of the reciprocal lattice.See *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Section 10.6 or *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapter 2.

Each time that all of the m_j change by one we generate a new state; therefore the volume of k-space occupied by one state is

$$\frac{\mathbf{A}_1}{N_1}.\frac{\mathbf{A}_2}{N_2}\times\frac{\mathbf{A}_3}{N_3}=\frac{1}{N}\mathbf{A}_1.\mathbf{A}_2\times\mathbf{A}_3. \tag{2.12}$$

Comparing this with Equation 2.7 shows that *the Brillouin zone always contains the same number of k-states as the number of primitive unit cells in the crystal.* This fact will be of immense importance later on (remember it!); it will be a key factor in determining whether a material is an insulator, semiconductor or metal.

2.4 The Schrödinger equation in a periodic potential.

The Schrödinger equation for a particle[4] of mass m in the periodic potential $V(\mathbf{r})$ may be written

$$H\psi=\{-\frac{\hbar^2\nabla^2}{2m}+V(\mathbf{r})\}\psi=E\psi. \tag{2.13}$$

As before (see Equation 2.3), we write the potential as a Fourier series

$$V(\mathbf{r})=\sum_{\mathbf{G}}V_{\mathbf{G}}\mathrm{e}^{\mathrm{i}\mathbf{G}.\mathbf{r}}, \tag{2.14}$$

where the $\mathbf{G}$ are the reciprocal lattice vectors. We are at liberty to set the origin of potential energy wherever we like; as a convenience for later derivations we set the uniform background potential to be zero, *i.e.*

$$V_0\equiv 0. \tag{2.15}$$

We can write the wavefunction ψ as a sum of plane waves obeying the Born–von Karman boundary conditions,[5]

$$\psi(\mathbf{r})=\sum_{\mathbf{k}}C_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}. \tag{2.16}$$

This ensures that ψ also obeys the Born-von Karman boundary conditions.

We now substitute the wavefunction (Equation 2.16) and the potential (Equation 2.14) into the Schrödinger equation (Equation 2.13) to give

$$\sum_{\mathbf{k}}\frac{\hbar^2k^2}{2m}C_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}+\{\sum_{\mathbf{G}}V_{\mathbf{G}}\mathrm{e}^{\mathrm{i}\mathbf{G}.\mathbf{r}}\}\{\sum_{\mathbf{k}}C_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}\}=E\sum_{\mathbf{k}}C_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}. \tag{2.17}$$

The potential energy term can be rewritten

$$V(\mathbf{r})\psi=\sum_{\mathbf{G},\mathbf{k}}V_{\mathbf{G}}C_{\mathbf{k}}\mathrm{e}^{\mathrm{i}(\mathbf{G}+\mathbf{k}).\mathbf{r}}, \tag{2.18}$$

where the sum on the right-hand side is over all $\mathbf{G}$ and $\mathbf{k}$. As the sum is over all possible values of $\mathbf{G}$ and $\mathbf{k}$, it can be rewritten as[6]

$$V(\mathbf{r})\psi=\sum_{\mathbf{G},\mathbf{k}}V_{\mathbf{G}}C_{\mathbf{k}-\mathbf{G}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}. \tag{2.19}$$

Therefore the Schrödinger equation (Equation 2.17) becomes

$$\sum_{\mathbf{k}}\mathrm{e}^{\mathrm{i}\mathbf{k}.\mathbf{r}}\{(\frac{\hbar^2k^2}{2m}-E)C_{\mathbf{k}}+\sum_{\mathbf{G}}V_{\mathbf{G}}C_{\mathbf{k}-\mathbf{G}}\}=0. \tag{2.20}$$

[4]Note that I have written "a particle"; in the proof of Bloch's theorem that follows, I do not assume any specific form for the energy of the particles in the potential. The derivation will be equally true for photons with $\omega=ck$, electrons with $E=\hbar^2k^2/2m_{\mathrm{e}}$ *etc., etc.*. Thus, the conclusions will be seen to be true for *any* particle in *any* periodic potential.

[5]That is, the $\mathbf{k}$ in Equation 2.16 are given by Equation 2.11.

[6]Notice that the $\mathbf{G}$ also obey the Born–von Karman boundary conditions; this may be easily seen if values of m_j that are integer multiples of N_j are substituted in Equation 2.11. As the sum in Equation 2.16 is over all $\mathbf{k}$ that obey the Born–von Karmann boundary conditions, it automatically encompasses all $\mathbf{k}-\mathbf{G}$.

As the Born-von Karman plane waves are an orthogonal set of functions, the coefficient of each term in the sum must vanish (one can prove this by multiplying by a plane wave and integrating), *i.e.*

$$\left(\frac{\hbar^2 k^2}{2m} - E\right) C_{\mathbf{k}} + \sum_{\mathbf{G}} V_{\mathbf{G}} C_{\mathbf{k}-\mathbf{G}} = 0. \tag{2.21}$$

(Note that we get the Sommerfeld result if we set $V_{\mathbf{G}} = 0$.)

It is going to be convenient to deal just with solutions in the first Brillouin zone (we have already seen that this contains all useful information about k-space). So, we write $\mathbf{k} = (\mathbf{q} - \mathbf{G}')$, where $\mathbf{q}$ lies in the first Brillouin zone and $\mathbf{G}'$ is a reciprocal lattice vector. Equation 2.21 can then be rewritten

$$\left(\frac{\hbar^2 (\mathbf{q} - \mathbf{G}')^2}{2m} - E\right) C_{\mathbf{q}-\mathbf{G}'} + \sum_{\mathbf{G}} V_{\mathbf{G}} C_{\mathbf{q}-\mathbf{G}'-\mathbf{G}} = 0. \tag{2.22}$$

Finally, we change variables so that $\mathbf{G}'' \to \mathbf{G} + \mathbf{G}'$, leaving the equation of coefficients in the form

$$\left(\frac{\hbar^2 (\mathbf{q} - \mathbf{G}')^2}{2m} - E\right) C_{\mathbf{q}-\mathbf{G}'} + \sum_{\mathbf{G}''} V_{\mathbf{G}''-\mathbf{G}'} C_{\mathbf{q}-\mathbf{G}''} = 0. \tag{2.23}$$

This equation of coefficients is very important, in that it specifies the $C_{\mathbf{k}}$ which are used to make up the wavefunction ψ in Equation 2.16.

2.5 Bloch's theorem

Equation 2.23 only involves coefficients $C_{\mathbf{k}}$ in which $\mathbf{k} = \mathbf{q} - \mathbf{G}$, with the $\mathbf{G}$ being general reciprocal lattice vectors. In other words, if we choose a particular value of $\mathbf{q}$, then the only $C_{\mathbf{k}}$ that feature in Equation 2.23 are of the form $C_{\mathbf{q}-\mathbf{G}}$; these coefficients specify the form that the the wavefunction ψ will take (see Equation 2.16).

Therefore, for each distinct value of $\mathbf{q}$, there is a wavefunction $\psi_{\mathbf{q}}(\mathbf{r})$ that takes the form

$$\psi_{\mathbf{q}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{\mathbf{q}-\mathbf{G}} \mathrm{e}^{\mathrm{i}(\mathbf{q}-\mathbf{G}).\mathbf{r}}, \tag{2.24}$$

where we have obtained the equation by substituting $\mathbf{k} = \mathbf{q} - \mathbf{G}$ into Equation 2.16. Equation 2.24 can be rewritten

$$\psi_{\mathrm{q}}(\mathbf{r}) = \mathrm{e}^{\mathrm{i}\mathbf{q}.\mathbf{r}} \sum_{\mathbf{G}} C_{\mathbf{q}-\mathbf{G}} \mathrm{e}^{-\mathrm{i}\mathbf{G}.\mathbf{r}} = \mathrm{e}^{\mathrm{i}\mathbf{q}.\mathbf{r}} u_{j,\mathbf{q}}, \tag{2.25}$$

i.e. (a plane wave with wavevector within the first Brillouin zone)×(a function $u_{j,\mathbf{q}}$ with the periodicity of the lattice).[7]

This leads us to *Bloch's theorem*. "The eigenstates ψ of a one-electron Hamiltonian $H = -\frac{\hbar^2 \nabla^2}{2m} + V(\mathbf{r})$, where $V(\mathbf{r} + \mathbf{T}) = V(\mathbf{r})$ for all Bravais lattice translation vectors $\mathbf{T}$ can be chosen to be a plane wave times a function with the periodicity of the Bravais lattice."

Note that Bloch's theorem

- is true for *any* particle propagating in a lattice (even though Bloch's theorem is traditionally stated in terms of electron states (as above), in the derivation we made no assumptions about what the particle *was*);
- makes no assumptions about the *strength* of the potential.

[7] You can check that $u_{j,\mathbf{q}}$ (*i.e.* the sum in Equation 2.25) has the periodicity of the lattice by making the substitution $\mathbf{r} \to \mathbf{r} + \mathbf{T}$ (see Equations 2.2, 2.3 and 2.4 if stuck).

2.6 Electronic bandstructure.

Equation 2.25 hints at the idea of *electronic bandstructure*. Each set of $u_{j,\mathbf{q}}$ will result in a set of electron states with a particular character (*e.g.* whose energies lie on a particular dispersion relationship[8]); this is the basis of our idea of an *electronic band*. The number of possible wavefunctions in this band is just going to be given by the number of distinct $\mathbf{q}$, *i.e.* the number of Born-von Karman wavevectors in the first Brillouin zone. Therefore the number of electron states in each band is just $2\times$(the number of primitive cells in the crystal), where the factor two has come from spin-degeneracy. This is going to be very important in our ideas about *band filling*, and the classification of materials into metals, semimetals, semiconductors and insulators.

We are now going to consider two tractable limits of Bloch's theorem, a very weak periodic potential and a very strong periodic potential (so strong that the electrons can hardly move from atom to atom). We shall see that both extreme limits give rise to *bands*, with *band gaps* between them. In both extreme cases, the bands are qualitatively very similar; *i.e.* real potentials, which must lie somewhere between the two extremes, must also give rise to qualitatively similar bands and band gaps.

2.7 Reading.

This topic is treated with some expansion in Chapter 2 of *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001). A simple justification of the Bloch theorem is given in *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) in the first few pages of Chapter 7 and in *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Section 10.4; an even more elementary one is found in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Section 12.3. More rigorous and general proofs are available in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) pages 133-140.

[8] A *dispersion relationship* is the function $E = E(\mathbf{q})$.

Handout 3

The nearly-free electron model

3.1 Introduction

Having derived Bloch's theorem we are now at a stage where we can start introducing the concept of bandstructure. When someone refers to the bandstructure of a crystal they are generally talking about its electronic *dispersion*, $E(\mathbf{k})$ (i.e. how the energy of an electron varies as a function of crystal wavevector). However, Bloch's theorem is very general and can be applied to any periodic interaction, not just to electrons in the periodic electric potential of ions. For example in recent years the power of band theory has been applied to photons in periodic dielectric media to study *photonic bandstructure* (i.e. dispersion relations for photons in a "photonic crystal").

In this lecture we will firstly take a look at dispersion for an electron in a periodic potential where the potential very weak (the *nearly free electron approximation*) and in the next lecture we will look at the case where the potential is very strong (*tight binding approximation*). Firstly let's take a closer look at dispersion.

3.2 Dispersion $E(\mathbf{k})$

You will recall from the Sommerfeld model that the dispersion of a free electron is $E(k) = \frac{\hbar^2 k^2}{2m}$. It is completely isotropic (hence the dispersion only depends on $k = |\mathbf{k}|$) and the Sommerfeld model produces exactly this bandstructure for every material – not very exciting! Now we want to understand how this parabolic relation changes when you consider the periodicity of the lattice.

Using Bloch's theorem you can show that translational symmetry in real space (characterised by the set translation vectors $\{\mathbf{T}\}$) leads to translational symmetry in k-space (characterised by the set of reciprocal lattice vectors $\{\mathbf{G}\}$). Knowing this we can take another look at Schrödinger's equation for a free electron in a periodic potential $V(\mathbf{r})$:

$$H\psi_{\nu\mathbf{k}}(\mathbf{r}) = \{-\frac{\hbar^2\nabla^2}{2m} + V(\mathbf{r})\}\psi_{\nu\mathbf{k}}(\mathbf{r}) = E_{\nu\mathbf{k}}\psi_{\nu\mathbf{k}}(\mathbf{r}). \tag{3.1}$$

and taking the limit $V(\mathbf{r}) \to 0$ we know that we have a plane wave solution. This implies that the Bloch function $u(\mathbf{r}) \to 1$. However considering the translational invariance in k-space the dispersion relation must satisfy:

$$E_{\nu\mathbf{k}} = \frac{\hbar^2|\mathbf{k}|^2}{2m} = \frac{\hbar^2|\mathbf{k}+\mathbf{G}|^2}{2m} \tag{3.2}$$

for the set of all reciprocal lattice vectors $\{\mathbf{G}\}$. This dispersion relation is show in Fig. 3.1

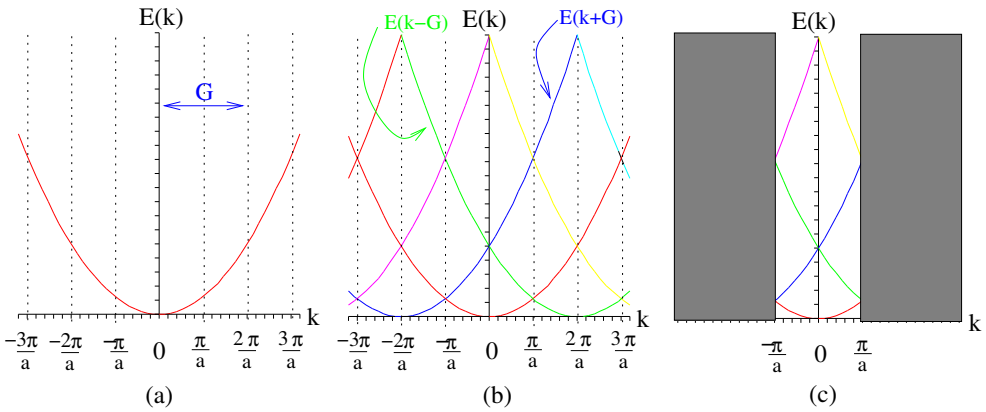


Figure 3.1: Simple bandstructure diagrams for a one dimensional periodic solid in the limit $V(\mathbf{r}) \to 0$ expressed in the extended zone (a), repeated zone (b), and reduced zone (c) schemes.

3.3 Nearly free electron model

Since we are in the weak potential limit we can treat the crystal potential as a weak perturbation added to the Hamiltonian of a free electron. Let's start with the Schrödinger equation for a free electron

$$\hat{\mathrm{H}}_{\mathrm{o}}\psi_{\nu\mathbf{k}}(\mathbf{r}) = E_{\nu\mathbf{k}}\psi_{\nu\mathbf{k}}(\mathbf{r}) \tag{3.3}$$

where

$$\hat{\mathrm{H}}_{\mathrm{o}} = \frac{\hat{\mathrm{p}}^2}{2m} \tag{3.4}$$

which has plane wave eigenstates

$$\psi_{\nu\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V_{r3}}}\exp(i\mathbf{k}\cdot\mathbf{r}) \tag{3.5}$$

We now introduce a small perturbation, $\hat{\mathrm{H}}'$ associated with the crystal potential

$$\hat{\mathrm{H}} = \hat{\mathrm{H}}_{\mathrm{o}} + \hat{\mathrm{H}}' \tag{3.6}$$

. Since the lattice is periodic we may expand the perturbation into a Fourier series where $\{\mathbf{G}\}$ are a set of vectors and $V_{\mathbf{G}}$ are Fourier coefficients[1]

$$\hat{\mathrm{H}}' = V(\mathbf{r}) = \sum_{\{\mathbf{G}\}} V_{\mathbf{G}}\exp(-i\mathbf{G}\cdot\mathbf{r}). \tag{3.7}$$

Since the lattice is periodic we may expand the perturbation into a Fourier series where $\mathbf{G}$ are a set of vectors and $V_{\mathbf{G}}$ are Fourier coefficients.

⋮

Treating the nearly free electron model using degenerate perturbation theory has been shown on the blackboard during lectures

3.4 Consequences of the nearly-free-electron model.

In the lectures we have derived two simple rules, which are

- away from Brillouin-zone boundaries the electronic bands (*i.e.* dispersion relationships) are very similar to those of a free electron;

[1] By considering $V(\mathbf{r}+\mathbf{T}_n) = V(\mathbf{r})$ you can show that $\mathbf{G}$ turns out to be the reciprocal lattice vector (see Section 2.2)

- bandgaps open up whenever $E(\mathbf{k})$ surfaces cross, which means in particular at the zone boundaries.

To see how these rules influence the properties of real metals, we must remember that each band in the Brillouin zone will contain $2N$ electron states (see Sections 2.3 and 2.6), where N is the number of primitive unit cells in the crystal. We now discuss a few specific cases.

3.4.1 The alkali metals

The alkali metals Na, K *et al.* are monovalent (*i.e.* have one electron per primitive cell). As a result, their Fermi surfaces, encompassing N states, have a volume which is half that of the first Brillouin zone. Let us examine the geometry of this situation a little more closely.

The alkali metals have a body-centred cubic lattice with a basis comprising a single atom. The conventional unit cell of the body-centred cubic lattice is a cube of side a containing two lattice points (and hence 2 alkali metal atoms). The electron density is therefore $n = 2/a^3$. Substituting this in the equation for free-electron Fermi wavevector (Equation 1.23) we find $k_\mathrm{F} = 1.24\pi/a$. The shortest distance to the Brillouin zone boundary is half the length of one of the $\mathbf{A}_j$ for the body-centred cubic lattice, which is

$$\frac{1}{2}\frac{2\pi}{a}(1^2+1^2+0^2)^{\frac{1}{2}} = 1.41\frac{\pi}{a}.$$

Hence the free-electron Fermi-surface reaches only $1.24/1.41 = 0.88$ of the way to the closest Brillouin-zone boundary. The populated electron states therefore have ks which lie well clear of any of the Brillouin-zone boundaries, thus avoiding the distortions of the band due to the bandgaps; hence, the alkali metals have properties which are quite close to the predictions of the Sommerfeld model (*e.g.* a Fermi surface which is spherical to one part in 10^3).

3.4.2 Elements with even numbers of valence electrons

These substances contain just the right number of electrons ($2Np$) to completely fill an integer number p of bands up to a band gap. The gap will energetically separate completely filled states from the next empty states; to drive a net current through such a system, one must be able to change the velocity of an electron, *i.e* move an electron into an unoccupied state of different velocity. However, there are no easily accessible empty states so that such substances should not conduct electricity at $T = 0$; at finite temperatures, electrons will be thermally excited across the gap, leaving filled and empty states in close energetic proximity both above and below the gap so that electrical conduction can occur. Diamond (an insulator), Ge and Si (semiconductors) are good examples.

However, the divalent metals Ca *et al.* plainly conduct electricity rather well. To see why this is the case, consider the Fermi surface of the two-dimensional divalent metal with a square lattice shown in Figure 3.2. Initially, the free-electron Fermi surface is a circle with an area equivalent to the first Brillouin zone (Figure 3.2(a)), which consequently straddles the Brillouin-zone boundary (Figures 3.2(b) and (c)). Figures 3.2 (d) and (e) show what happens when a weak periodic potential is "turned on" and bandgaps open up at the Brillouin-zone boundaries; the band gap raises the energy of the states close to the zone edge in Figure 3.2(c) and lowers those close to the zone edge in Figure 3.2(b) (see Figure **??**). For ease of reference, we shall call the former states "the upper band" and the latter states "the lower band". Hence some electrons will transfer back from the upper band (the states above the gap) to the lower band (the states below it), tending to distort the Fermi surface sections close to the Brillouin-zone boundaries.

In the situation shown in Figures 3.2(d) and (e), the material is obviously still an electrical conductor, as filled and empty states are adjacent in energy. Let us call the band gap at the centres of the Brillouin-zone edges $E_\mathrm{g}^\mathrm{cent}$ and that at the corners of the Brillouin zone $E_\mathrm{g}^\mathrm{corn}$.[2] The lowest energy states in the upper band will be at points $(\pm\frac{\pi}{a}, 0)$, $(0, \pm\frac{\pi}{a})$, where a is the lattice parameter of the square lattice, with energy

$$E_\mathrm{lowest}^\mathrm{u} = \frac{\hbar^2}{2m_\mathrm{e}}\frac{\pi^2}{a^2} + \frac{E_\mathrm{g}^\mathrm{cent}}{2}, \tag{3.8}$$

[2] $E_\mathrm{g}^\mathrm{cent}$ and $E_\mathrm{g}^\mathrm{corn}$ will in general not be the same; two plane waves contribute to the former and four to the latter. However, the gaps will be of similar magnitude.

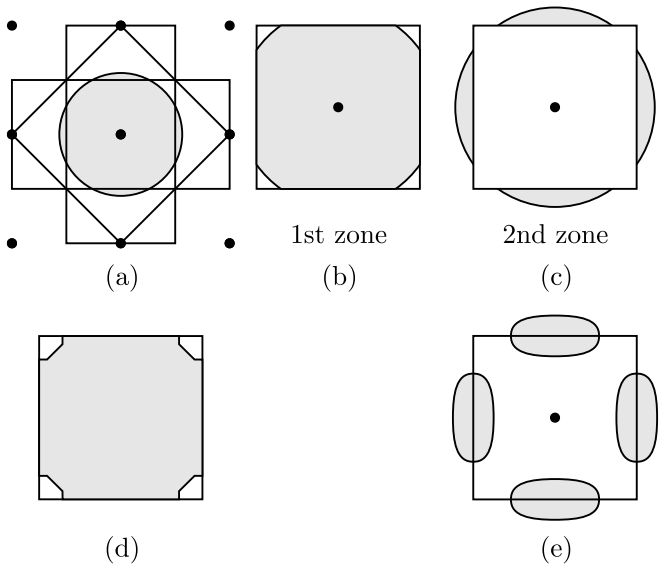


Figure 3.2: The evolution of the Fermi surface of a divalent two-dimensional metal with a square lattice as a band gap is opened at the Brillouin zone boundary: (a) free-electron Fermi surface (shaded circle), reciprocal lattice points (solid dots) and first (square) second (four isoceles triangles) and third (eight isoceles triangles) Brillouin zones; (b) the section of Fermi surface enclosed by the first Brillouin zone; (c) the sections of Fermi surface in the second Brillouin zone; (d) distortion of the Fermi-surface section shown in (b) due to formation of band gaps at the Brillouin-zone boundaries; (e) result of the distortion of the Fermi-surface section in (c) plus "folding back" of these sections due to the periodicity of k-space.

i.e. the free-electron energy plus half the energy gap. Similarly, the highest energy states in the lower band will be at the points $(\pm\frac{\pi}{a}, \pm\frac{\pi}{a})$, with energy

$$E^{\mathrm{l}}_{\mathrm{highest}} = \frac{\hbar^2}{2m_{\mathrm{e}}}\frac{2\pi^2}{a^2} - \frac{E^{\mathrm{corn}}_{\mathrm{g}}}{2}, \tag{3.9}$$

i.e. the free-electron energy minus half the energy gap. Therefore the material will be a conductor as long as $E^{\mathrm{l}}_{\mathrm{highest}} > E^{\mathrm{u}}_{\mathrm{lowest}}$. Only if the band gap is big enough for $E^{\mathrm{l}}_{\mathrm{highest}} < E^{\mathrm{u}}_{\mathrm{lowest}}$, will all of the electrons be in the lower band at $T = 0$, which will then be completely filled; filled and empty states will be separated in energy by a gap and the material will be an insulator at $T = 0$.

Thus, in general, in two and three dimensional divalent metals, the geometrical properties of the free-electron dispersion relationships allow the highest states of the lower band (at the corners of the first Brillouin zone boundary most distant from the zone centre) to be at a higher energy than the lowest states of the upper band (at the closest points on the zone boundary to the zone centre). Therefore both bands are partly filled, ensuring that such substances conduct electricity at $T = 0$; only one-dimensional divalent metals have no option but to be insulators.

We shall see later that the empty states at the top of the lowest band (*e.g.* the unshaded states in the corners in Figure 3.2(d)) act as *holes* behaving as though they have a *positive* charge. This is the reason for the positive Hall coefficients observed in many divalent metals (see Table 1.1).

3.4.3 More complex Fermi surface shapes

The Fermi surfaces of many simple di- and trivalent metals can be understood adequately by the following sequence of processes.

1. Construct a free-electron Fermi sphere corresponding to the number of valence electrons.

2. Construct a sufficient number of Brillouin zones to enclose the Fermi sphere.

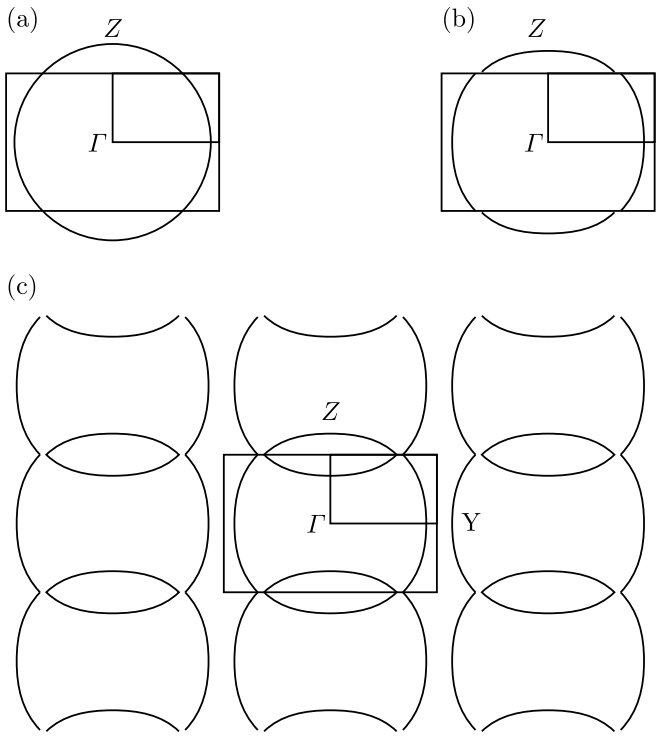


Figure 3.3: The evolution of the Fermi surface of a divalent two-dimensional metal as a band gap is opened at the Brillouin zone boundary. (a) Free-electron Fermi circle and rectangular Brillouin zone; (b) the effect of a band gap opening up at the Brillouin-zone boundary; (c) resulting Fermi surface sections in the extended-zone scheme.

3. Split and "round off" the edges of the Fermi surface wherever it cuts a Brillouin-zone boundary (*i.e.* points at which band-gaps opens up).

4. Apply the periodicity of k-space by replicating all of the Fermi-surface sections at equivalent points in the first Brillouin zone.

These steps are illustrated in great detail for a number of cases in *e.g. Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapter 9.

Note that the shape of the Brillouin zone has a profound effect on the Fermi surface sections generated, as shown by the Fermi surface of Figure 3.3. As in Figure 3.2, we have a divalent metal. However, in Figure 3.3, the Brillouin zone is *rectangular*, rather than square, so that the Fermi surface only cuts two of the Brillouin zone edges. Thus, after band gaps have opened up, the Fermi surface consists of a closed ellipse plus corrugated open lines, rather than the two closed sections of Figure 3.2.

3.5 Reading

An expansion of this material is given in *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001), Chapter 3. A more detailed treatement of traditional elemental metals is given in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapters 8, 9, 10 and 12 (even if you understand nothing of the discussion, the pictures are good). Simpler discussions are available in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Chapter 12, *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Sections 10.1-10.21, *Electrons in Metals and*

Semiconductors, by R.G. Chambers (Chapman and Hall, London 1990) Chapters 4-6, *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapters 8 and 9.

Handout 4

The tight-binding model

4.1 Introduction

In the tight-binding model we assume the opposite limit to that used for the nearly-free-electron approach, *i.e.* the potential is so large that the electrons spend most of their lives bound to ionic cores, only occasionally summoning the quantum-mechanical wherewithal to jump from atom to atom.

The atomic wavefunctions $\phi_j(\mathbf{r})$ are defined by

$$\mathcal{H}_{\rm at}\phi_j(\mathbf{r}) = E_j\phi_j(\mathbf{r}), \tag{4.1}$$

where $\mathcal{H}_{\rm at}$ is the Hamiltonian of a single atom.

The assumptions of the model are

1. close to each lattice point, the crystal Hamiltonian $\mathcal{H}$ can be approximated by $\mathcal{H}_{\rm at}$;
2. the bound levels of $\mathcal{H}_{\rm at}$ are well localised, *i.e.* the $\phi_j(\mathbf{r})$ are very small one lattice spacing away, implying that
3. $\phi_j(\mathbf{r})$ is quite a good approximation to a stationary state of the crystal, as will be $\phi_j(\mathbf{r}+\mathbf{T})$.

Reassured by the second and third assumptions, we make the Bloch functions $\psi_{j,\mathbf{k}}$ of the electrons in the crystal from linear combinations of the atomic wavefunctions.

4.2 Band arising from a single electronic level

A method of obtaining the dispersion relation for a single electronic level $E_\nu(\mathbf{k})$ in the tight binding approximation was given in lectures. Note that this topic is also covered in a number of textbooks. *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) considers the specific case of a crystal with mutually perpendicular (Cartesian) basis vectors, this may help you understand the more general result.

4.3 General points about the formation of tight-binding bands

The derivation given in the lecture illustrates several points about real bandstructure.

1. Figure 4.1 shows schematically the process involved in forming the tight-binding bands. N single atoms with p (doubly degenerate) atomic levels have become p $2N$-fold degenerate bands.
2. The transfer integrals give a direct measure of the width in energy of a band (the *bandwidth*); small transfer integrals will give a narrow bandwidth. In question 2 of problem set 1, we shall see that carriers close to the bottom of a tight-binding band have *effective masses* which are inversely proportional to the transfer integrals. The effective mass parameterises the ease with

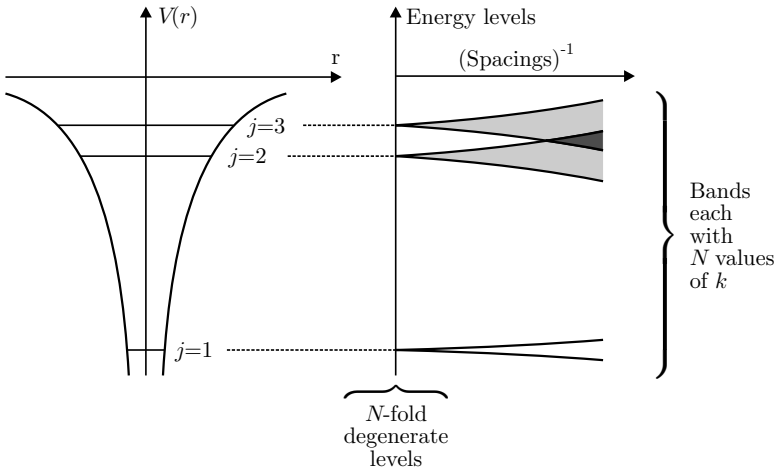


Figure 4.1: Schematic representation of the formation of tight-binding bands as the spacing between atoms is reduced.

which an electron can be accelerated. Thus, small transfer integrals lead to heavy effective masses (and narrow bandwidths), indicating that it is hard to move the electrons around; large transfer integrals lead to light effective masses (and large bandwidths), indicating that it is easy to move the electrons around. This ties in exactly with the real-space picture of the formation of tight-binding bands, in which the transfer integrals reflect the ease with which an electron can transfer from atom to atom.

3. The "shape" of the bands in k-space will be determined in part by the real-space crystal structure; if the atoms in a certain direction are far apart, then the bandwidth will be narrow for motion in that direction.

4. The bands will also reflect the character of the atomic levels which have gone to make them up.

The latter point will be illustrated in Section 4.3.1 below.

4.3.1 An example: the transition metals

The tight-binding model implies that bands will reflect the character of the atomic levels which have gone to make them up. This is illustrated by Figure 4.2, which shows the calculated tight-binding bands for Copper. Note that the character of the original atomic levels is reflected in the width and shape of the the bands; the more compact and anisotropic 3d orbitals give rise to five narrow bands of complex shape, whilst the single band derived from the larger, spherical 4s orbitals is wide and almost free-electron-like. The properties of a particular 3d metal are strongly dependent on the position of the Fermi energy amongst this mess.

Note that the colours of 3d metals such as Cu and Au result from the optical transitions which are possible between the occupied d-bands and the empty states at the top of the s-band.

4.4 Reading

Alternative treatments are given in *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapter 9 (simpler), *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Section 10.9 (see also Section 10.3) (simpler), *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) Chapter 4 (same level, more details) and *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) pages 176-190 (more general and therefore more complicated). The tight-binding model is related to the

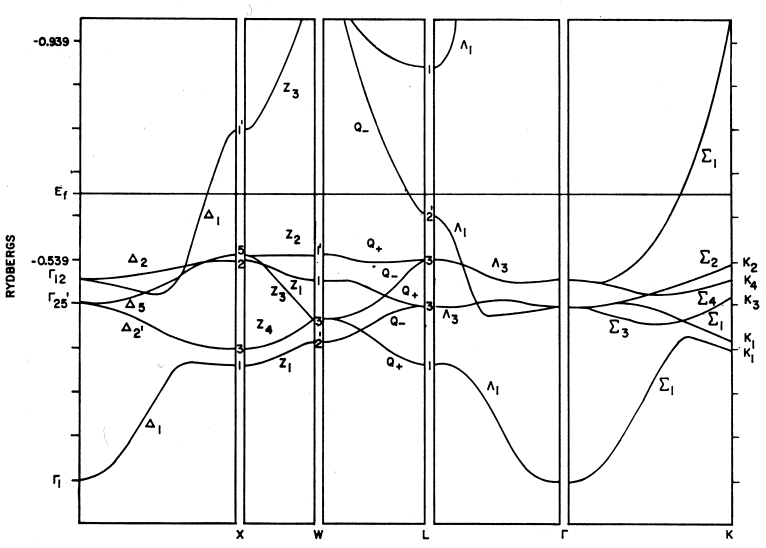


Figure 4.2: Theoretical calculation of the bandstructure of Copper. The electron energy (in units of Rydbergs, *i.e.* 13.6 eV) is plotted as a function of **k** for various directions in the Brillouin zone. Note that the band derived from the 4s orbitals (starting at Γ_1, it anticrosses with the bands derived from the 3d orbitals, before re-emerging as Δ_1) is very similar to the free-electron dispersion relationship. The Fermi energy is indicated for Cu; the d-bands are deeply buried, and Cu in many respects looks reasonably like a metal obeying the Sommerfeld model; apart from its "necks", the Fermi surface is quite spherical. On the other hand, in the case of a metal such as Co, the Fermi energy is in amongst the d-bands, resulting in a complex Fermi surface and very different properties. (Figure taken from G.A. Burdick, *Phys. Rev.* **129**, 138 (1963).)

Kronig-Penney Model, which is a very simple illustration of the formation of bands; see *e.g. Quantum Mechanics*, by Stephen Gasiorowicz (Wiley, New York 1974) page 98.

Handout 5

Some general points about bandstructure

5.1 Comparison of tight-binding and nearly-free-electron bandstructure

Let us compare a band of the nearly-free-electron model with a one-dimensional tight-binding band

$$E(k) = E_0 - 2t\cos(ka), \tag{5.1}$$

where E_0 is a constant. Note that

- both bands look qualitatively similar, *i.e.*
- both bands have minima and maxima (*i.e.* points where $\mathrm{d}E/\mathrm{d}k = 0$) at the Brillouin-zone centres and boundaries respectively;
- both bands have the same k-space periodicity;
- the tops and bottoms of both bands are approximately parabolic (*i.e.* E is proportional to $(\mathbf{k}-\mathbf{k}_0)^2$ plus a constant, where $\mathbf{k}_0$ is the point in k-space at which the band extremum occurs).

The last point is obvious for the bottom of the lowest nearly-free-electron band. To see that the same thing happens for the tight-binding approach, we use the approximation

$$\cos(ka) \approx 1 - \frac{(ka)^2}{2} \tag{5.2}$$

for small ka in Equation 5.1 to give

$$E(k) \approx E_0 - 2t + ta^2k^2. \tag{5.3}$$

The equivalent proof for the top of the band (close to π/a) is left to the reader.

The above discussion shows that both very weak and very strong periodic potentials give rise to qualitatively similar bands. We therefore expect that our ideas about bands derived from these simple models will apply to the real bands in solids, where the potential strengths will fall between the two extremes. The more complex arrangements of atoms and/or molecules often found in real solids will give rise to more complex bandshapes, but qualitatively the properties of the bands will be the same as those of our simple models.

5.2 The importance of k

5.2.1 $\hbar\mathbf{k}$ is *not* the momentum

Bloch's theorem has introduced wavevectors which we have labelled $\mathbf{k}$ and $\mathbf{q}$ (see *e.g.* Equation 2.25). Although at first sight $\mathbf{k}$ and $\mathbf{q}$ look similar to the straightforward electron momentum $\mathbf{p} = \hbar\mathbf{k}$ in the Sommerfeld model, it is easy to show that the Bloch functions are not eigenstates of the momentum operator;

$$\frac{\hbar}{i}\nabla\psi(\mathbf{r}) = \frac{\hbar}{i}\nabla e^{i\mathbf{q}.\mathbf{r}}u_{j,\mathbf{q}} = \hbar\mathbf{q}\psi(\mathbf{r}) + e^{i\mathbf{q}.\mathbf{r}}\frac{\hbar}{i}\nabla u_{j,\mathbf{q}} \neq \mathbf{p}\psi(\mathbf{r}) \tag{5.4}$$

(where the Bloch wavefunction used is defined in Equation 2.25).

The quantity $\hbar\mathbf{k}$ is instead a *crystal momentum*, the momentum of the system as a whole; it is better to think of $\mathbf{k}$ as a quantum number which describes a Bloch state.

5.2.2 Group velocity.

In the original formulation of the Bloch states (see Equation 2.16), the electron wavefunction was defined to be the superposition of a set of plane waves (a *wavepacket*). It is therefore possible to use the idea of a *group velocity* $\mathbf{v}$

$$\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}}E, \tag{5.5}$$

where $\nabla_{\mathbf{k}}$ is the gradient operator in k-space, to describe the real-space motion of the electron (*c.f.* $v = (\mathrm{d}\omega/\mathrm{d}k)$ for the wavepackets encountered in Optics). Bearing this in mind, we are now going to derive the *effective mass*, which is very useful in parameterising the dynamics of band electrons when they are subjected to external forces. The derivation also illustrates a very important point about $\mathbf{k}$. I shall follow the derivation through in one dimension; one of the problems involves doing the same thing in three dimensions.

5.2.3 The effective mass

Let an external force f be applied to a band electron. The force will do work

$$\delta E = fv\delta t \tag{5.6}$$

in time δt. In addition,

$$\delta E = \frac{\mathrm{d}E}{\mathrm{d}k}\delta k = \hbar v\delta k. \tag{5.7}$$

Equating Equations 5.6 and 5.7, dividing through by δt and considering the limit $\delta t \to 0$ gives

$$\hbar\frac{\mathrm{d}k}{\mathrm{d}t} = f. \tag{5.8}$$

The equivalent three dimensional formula in the derivation of one of the problems is

$$\hbar\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \mathbf{f}, \tag{5.9}$$

where $\mathbf{k}$ and $\mathbf{f}$ are now vectors. Equation 5.9 is enormously important; it shows that *in a crystal* $\hbar\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}$ *is equal to the external force on the electron.*

After this amazing fact, the rest of the effective mass derivation is almost an anticlimax. The rate of change of velocity with time is

$$\frac{\mathrm{d}v}{\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}^2E}{\mathrm{d}k\mathrm{d}t} = \frac{1}{\hbar}\frac{\mathrm{d}^2E}{\mathrm{d}k^2}\frac{\mathrm{d}k}{\mathrm{d}t}. \tag{5.10}$$

Substituting for $\frac{\mathrm{d}k}{\mathrm{d}t}$ from Equation 5.8 and rearranging gives

$$\frac{\hbar^2}{\frac{\mathrm{d}^2E}{\mathrm{d}k^2}}\frac{\mathrm{d}v}{\mathrm{d}t} = m^*\frac{\mathrm{d}v}{\mathrm{d}t} = f, \tag{5.11}$$

where the *effective mass* m^* is defined by

$$m^* = \frac{\hbar^2}{\frac{\mathrm{d}^2 E}{\mathrm{d}k^2}}. \tag{5.12}$$

(The equivalent tensor quantities for an anisotropic band are given in one of the problems.)

Equation 5.11 shows that the effective mass gives a convenient way of describing the motion of band electrons subjected to an external force; the identity "force=(mass×acceleration)" can be used with the effective mass substituted for the inertial mass.

In general, the effective mass will be energy-dependent. However, one often has to deal with almost empty or almost full bands, *i.e.* the states close to the minima and maxima of the $E_j(\mathbf{k})$ dispersion relationships. As we have already seen in Section 5.1, these regions are often approximately parabolic (this is actually a general property of any maximum or minimum– try expanding any function possessing such an extremum close to that point using a Taylor series). Close to these points, the electrons can therefore be treated as if they were free, but with an effective mass; *i.e.* their dispersion relationship is

$$E(\mathbf{k}) \approx E_0 + \frac{\hbar^2}{2m^*}(\mathbf{k} - \mathbf{k_0})^2, \tag{5.13}$$

where energy E_0 and wavevector $\mathbf{k_0}$ define the band extremum. Note that m^* can be either positive or negative. This approximation often simplifies calculations; m^* in effect contains all of the necessary information about the way in which the electron's motion is modified by the crystal potential through which it moves. The equivalent model for anisotropic band extrema is explored in the problem about the effective mass tensor.

5.2.4 The effective mass and the density of states

In the previous lectures, we have seen that it is most natural to count electron states by evaluating the volume of k-space occupied (*e.g.* the Brillouin-zone volume determines the number of electrons that a band can accommodate). The counting of states is easy, because the states are spread through phase-space uniformly. However, when one comes to consider the evaluation of quantities such as the electronic heat capacity, where thermal population of states is a significant factor, it is often convenient to work in terms of energy (see *e.g.* the derivation of the electronic heat capacity within the Sommerfeld model), requiring a knowledge of the number of states per unit energy range per unit volume, $g(E)$.

If one substitutes a general value of k in Equation 1.22 instead of k_F, one obtains $n(\mathbf{k})$, the number of states per unit volume of r-space with wavevectors less than $|\mathbf{k}|$. If the band in question is assumed to obey Equation 5.13 close to the region of interest (*i.e.* it is parabolic and isotropic), then

$$g(E) \equiv \frac{\mathrm{d}n}{\mathrm{d}E} = \frac{\mathrm{d}n}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{1}{2\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{\frac{3}{2}}(E - E_0)^{\frac{1}{2}}. \tag{5.14}$$

This shows that the effective mass is a very convenient way of parameterising the curvature of a band at a certain energy, and hence the density of states $g(E)$ at that energy. We shall use this technique in a number of subsequent derivations.

Finally, note that Equation 5.14 shows that a heavy effective mass results in a large density of states. The reason for this can be seen in Figure 5.1. which shows a parabolic band with a light effective mass (steep curve) compared to a band with a heavy effective mass (shallow curve). A fixed region of k-space will always contain a fixed number of states. In the case of the heavy-mass band, the k-space interval Δk corresponds to a small interval of energy δE_1; the same interval of k corresponds to a much larger energy interval δE_2 for the light-mass band. As the same number of states is accommodated in each energy region, the density of states $g(E)$ is much higher for the heavy-mass band.

5.2.5 Summary of the properties of k

The properties of the **k** used in the Bloch wavefunction (see Equations 2.16 and 2.25) may be summarised as follows.

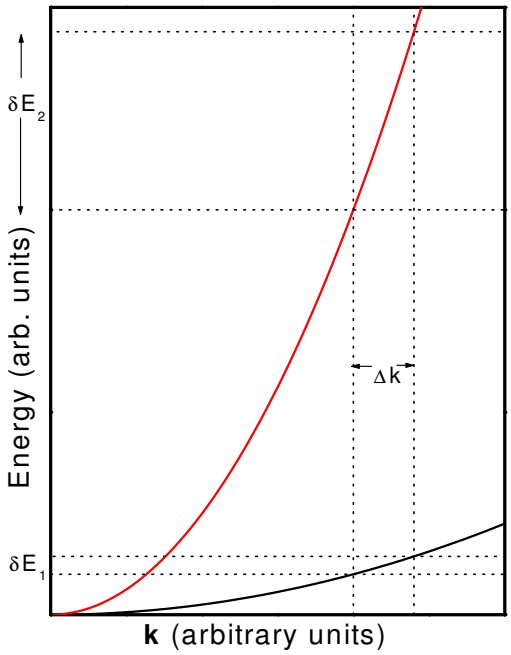


Figure 5.1: Comparison of a parabolic band with a light effective mass (steep curve) with a band with a heavy effective mass (shallow curve). In the case of the heavy-mass band, the k-space interval Δk corresponds to a small interval of energy δE_1; the same interval of k corresponds to a much larger energy interval δE_2 for the light-mass band.

1. $\mathbf{k}$ is not the (physical momentum/$\hbar$) of the electron; it is a quantum number describing the electron's state within a band. Each band is labelled using the index j.

2. For each j, $\mathbf{k}$ takes all values consistent with the Born–von Karman boundary conditions within the first Brillouin zone; in our notation $\mathbf{k}$ runs though all $\mathbf{q}$. j can run through an infinite number of discrete values.

3. For a given j, the electronic dispersion relationship $E_j(\mathbf{k})$ has no explicit form; the only constraint is that it must be periodic, *i.e.* $E_j(\mathbf{k}) = E_j(\mathbf{k} + \mathbf{G})$.

4. The velocity of an electron with energy $E_j(\mathbf{k})$ is given by $\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E_j(\mathbf{k})$.

5. The rate of change of $\mathbf{k}$ under the action of an external force $\mathbf{f}$ is given by $\hbar\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = \mathbf{f}$.

6. The electronic wavefunction is $\psi(\mathbf{r}) = e^{i\mathbf{k}.\mathbf{r}}u_{j,\mathbf{k}}$, where $u_{j,\mathbf{k}}(\mathbf{r}) = u_{j,\mathbf{k}}(\mathbf{r} + \mathbf{T})$. Here $u_{j,\mathbf{k}}(\mathbf{r})$ has no simple explicit form.

The above list tells us all we need to know about the dynamics of electrons in solids. Let us imagine for a line or two that we know nothing about bands; a calculation of the motion of an electron in a solid due to an external force seems intractable, as the electron is also subject to a complex array of forces from the ionic cores/molecules in the crystal. Band theory gets round this problem very effectively. The external force results in a time-dependent quantum number $\mathbf{k}(t)$. The resulting electronic velocity is obtained from the bandstructure using $\mathbf{v} = (1/\hbar)\nabla_{\mathbf{k}} E_j(\mathbf{k})$ and substituting $\mathbf{k} = \mathbf{k}(t)$. The bandstructure thus tells us all that we need to know about the interaction between the electrons and the relatively static components of the crystal (ions, molecules *etc.*).

5.2.6 Scattering in the Bloch approach

The Bloch wavefunctions describe states which are stationary solutions to the Schrödinger equation of a periodic potential. *Thus these states persist for ever and ever in a perfectly periodic infinite crystal*; there will be no scattering of electrons in the absence of disorder (which disturbs the perfect periodicity) or boundaries (which destroy the periodicity/infiniteness of the crystal). This is a manifestation of the wave-like nature of electrons; in a periodic array of scatterers a wave can propogate without attenuation because of the coherent constructive interference of the scattered waves. Scattering only takes place because of lack of periodicity in r-space.

This observation is in great contrast with the Drude model (where electrons were pictured crashing into virtually every ion) and the Sommerfeld approach (where scattering was acknowledged but its cause was a mystery). In addition

- it encompasses all possible causes of scattering (phonons, other electrons, impurities and boundaries, all of which disturb the local periodicity);

- it explains the enormous low-temperature scattering lengths observed in very pure single crystals of metals (often $\sim$ mm or even cm);

- it removes the need for the hand-waving explanations as to why the electrons in a metal "might not see the ionic cores very much" used to justify the Drude and Sommerfeld approaches.

5.3 General (non-isotopic) density of states

In Section 5.2.4 we considered the density of states $g(E)$ where $E(|\mathbf{k}|)$, i.e. in the special case of isotropic systems. However in non-isotropic materials we may still need to know the density of states at a particular point in k-space. In this section we will treat the more general case of non-isotropic materials, where the density of states at is expressed as $g(E(\mathbf{k}))$.

$$g(E(\mathbf{k})) = \sum_{\nu} g_\nu(E_\nu(\mathbf{k})) \tag{5.15}$$

where

$$g_\nu(E_\nu(\mathbf{k})) = \begin{cases} \frac{1}{\pi}\frac{1}{|\nabla_k E|} & \text{1D} \\ \frac{1}{2\pi^2}\int_l \frac{dl}{|\nabla_k E|} & \text{2D} \\ \frac{1}{4\pi^3}\int\int_S \frac{dS}{|\nabla_k E|} & \text{3D} \end{cases} \tag{5.16}$$

Discussion of and a method showing how to obtain these results will presented in the lecture

5.3.1 Van Hove singularities

This section will also be presented in lectures on the board

5.4 Holes

It is going to be useful to be able to describe a few empty states close to the top of an almost full band using the concept of *holes*. Consider a band, containing electrons with quantum numbers $\mathbf{k}_j$, velocities $\mathbf{v}_j$ and energies $E(\mathbf{k}_j)$, where $E = 0$ is at the top of the band. For a full band, the values of $\mathbf{k}$ should all sum to zero, *i.e.*

$$\sum_j \mathbf{k}_j = 0. \tag{5.17}$$

We consider removing one electron to create an excitation which we label *a hole*. Suppose that the lth electron is removed; the band then acquires a net $\mathbf{k}$ which we attribute to the presence of the hole. The hole has $\mathbf{k} = \mathbf{k}_\mathrm{h}$, with

$$\mathbf{k}_\mathrm{h} = \sum_{j\neq l} \mathbf{k}_j = -\mathbf{k}_l. \tag{5.18}$$

It is obvious that the lower down the band the empty state, the more excited the system. The hole's energy E_h must therefore take the form

$$E_\mathrm{h} = -E(\mathbf{k}_l). \tag{5.19}$$

The group velocity $\mathbf{v}_\mathrm{h}$ associated with the hole is

$$\mathbf{v}_\mathrm{h} = \frac{1}{\hbar}\nabla_{\mathbf{k}_\mathrm{h}} E_\mathrm{h} = \frac{1}{\hbar}\nabla_{-\mathbf{k}_l}(-E(\mathbf{k}_l)) = \mathbf{v}_l, \tag{5.20}$$

where the minus signs in Equations 5.18 and 5.19 have cancelled each other.

The full band will carry no current, *i.e.*

$$\sum_j (-e)\mathbf{v}_j = 0; \tag{5.21}$$

the removal of the pth electron produces a current

$$\sum_{j\neq l} (-e)\mathbf{v}_j = -(-e)\mathbf{v}_l = (+e)\mathbf{v}_\mathrm{h}, \tag{5.22}$$

i.e. the hole appears to have an associated *positive* charge. Finally, substitution of Equations 5.18 and 5.19 into Equation 5.12 shows that the effective mass m^*_h (or effective mass tensor components in a more complex band) associated with the hole is given by

$$m^*_\mathrm{h} = -m^*_l. \tag{5.23}$$

The importance of holes stems from the fact that bands are often well-characterised only close to the band extrema. Our knowledge of the dispersion relationships away from a particular extremum is often rather nebulous; such regions are not easily studied experimentally. It is therefore easier to deal with a small number of empty states close to the well-characterised maximum of an almost full band rather than attempt to treat the huge number of poorly-characterised states lower down in the band.

The fact that holes behave as though they have positive charge and positive effective mass explains why many di- and trivalent metals have positive Hall coefficients (see Section 1.3.3 and Table 1.1). A simple instance in which this could happen is shown in Figure 3.2. The section of Fermi surface in the corner of the Brillouin zone (Figure 3.2(d)) corresponds to a small number of empty states at the top of a band; the states are therefore hole-like. By contrast, the sections of Fermi surface straddling the zone boundaries (Figure 3.2(e)) represent a few filled states at the bottom of the upper band; these states are therefore electron-like. The observed Hall coefficient would depend on the relative contributions that the various Fermi-surface sections make to the electrical conductivity.

5.5 Postscript.

We have now derived all of the ideas needed for a reasonable understanding of the bandstructures of real solids. We shall take these ideas and use them to study the properties of semiconductors and insulators in the following lecture.

5.6 Reading.

A simple treatment is found in *e.g.* *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Sections 12.4 and 12.5, and a more comprehensive version is available in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapter 12.

A higher-level description of the general properties of several different types of bandstructure calculation is given in *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991) Chapter 2, *Electrons in metals*, by J.M. Ziman (Taylor and Francis) and (harder) *Fundamentals of semiconductors*, by P. Yu and M. Cardona (Springer, Berlin, 1996) Chapter 2.

Handout 6

Semiconductors and Insulators

The most important aspect of semiconductor[1] bandstructure may be summarised as follows; at absolute zero the highest completely filled band (*the valence band*) is separated from the lowest empty band (*the conduction band*) by an energy gap or band gap E_g of forbidden states. Therefore the material does not conduct electricity at $T = 0$. At finite temperatures a variety of processes enable electrons to be excited into the conduction band and empty states to occur in the valence band, thus allowing electrical conduction. However, we shall see that the presence of the energy gap still dominates the properties of the semiconductor.

Most of the technologically-important semiconductors, such as Ge, Si, GaAs and (Hg,Cd)Te, have a face-centred-cubic lattice with a two-atom basis; this is illustrated in Figure 6.1. In the case of Si and Ge, the atoms on the A and B sites in Figure 6.1 are identical (the so-called *diamond* structure). In the case of binary semiconductors such as (for example) GaAs the A sites will have Ga atoms and the B sites As atoms; the crystal is said to have the *zinc blende structure*. As stated above, in each case, the underlying lattice is face-centred cubic and there is a two atom basis. Thus, both the group IV elements and many of the binary semiconductors such as GaAs have the same shape of first Brillouin zone, which is shown in Figure 6.1 and which will be useful for understanding the band diagrams in the following discussion.

6.1 Bandstructure of Si and Ge

6.1.1 General points

Si and Ge were the first semiconductors to be exploited commercially, and Si continues to be very important as a technological material. [2] Figures 6.2 and 6.3 show calculations of the bandstructures of Si and Ge respectively. At first sight, the situation looks a complex, fearful mess. However, at the typical temperatures that will concern us ($0- \sim 300$ K), the only dramatic action (*i.e.* thermal excitation of electrons and holes, optical absorption edges) will occur close to the highest point in the valence band and the lowest points in the conduction band.

6.1.2 Heavy and light holes

In both Si and Ge, two bands converge at the valence band maximum in the Brillouin-zone centre (the Γ-point). These bands are known as *the heavy- and light-hole bands*; the flatter one, with its large value of $(\mathrm{d}^2E/\mathrm{d}k^2)^{-1}$, is the heavy hole band, and the steeper one the light hole band (for obvious reasons– see Figure 5.1).

The heavy holes tend to dominate the properties of the valence-band extremum; their heavier effective mass means that their density of states will be much larger than that of the light holes (see Section 5.2.4).

[1] For brevity I shall refer to such materials collectively as "semiconductors"; as we shall see, the distinction between a semiconductor and an insulator is fairly nebulous.

[2] See *e.g. Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991) Section 5.7 and references therein.

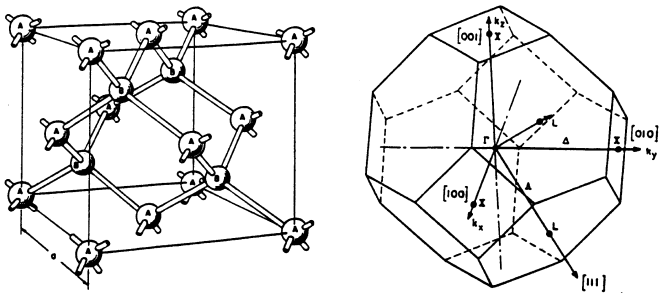


Figure 6.1: Left: an illustration of the crystal structures of the group IV elements and many of the binary semiconductors such as GaAs and CdTe. In the case of C (diamond), Si, Ge and Sn, the atoms on the A and B sites are identical (the so-called *diamond* structure). In the case of binary semiconductors such as for example GaAs, the A will be Ga and B As; the crystal is said to have the *zinc blende structure*. In each case, the underlying lattice is face-centred cubic and there is a two atom basis. Thus, both the group IV elements and the binary semiconductors have the same shape of first Brillouin zone, which is shown on the right.

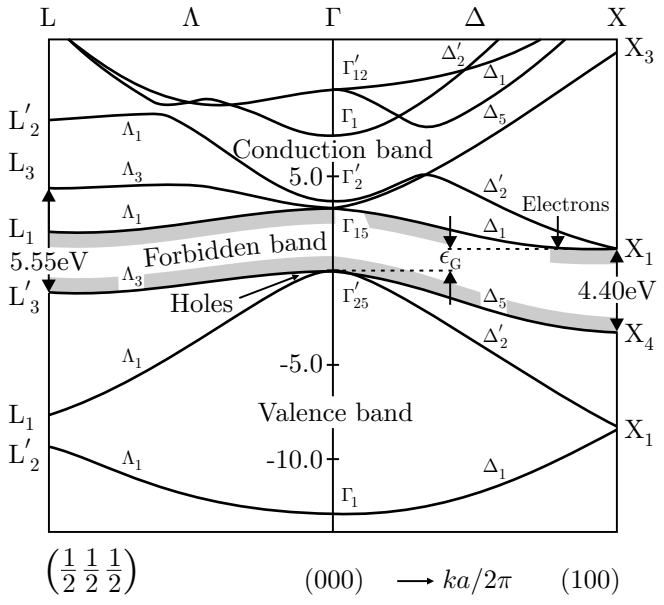


Figure 6.2: Calculated bandstructure of Si projected along the (100) and (111) directions. The shading indicates the lowest (unoccupied at $T = 0$) conduction band and the highest (occupied at $T = 0$) valence band, with the "forbidden band" or band gap in between. (Based on M. Cardona and F. Pollack, *Phys. Rev.* **142**, 530 (1966).)

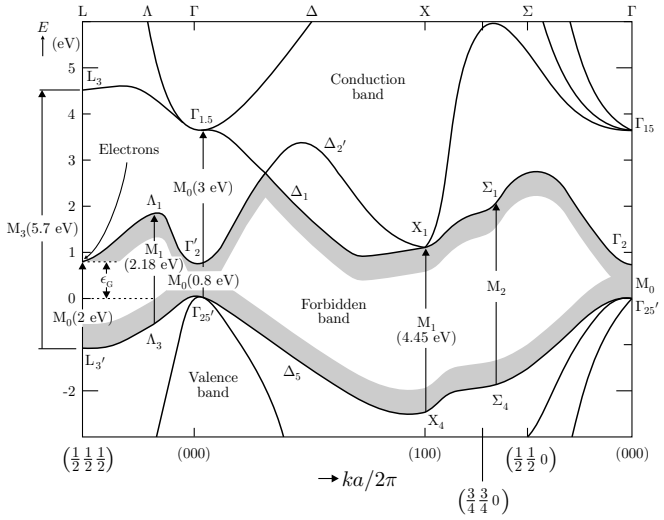


Figure 6.3: Calculation of the bandstructure of Ge along various projections. The shading indicates the lowest (unoccupied at $T = 0$) conduction band and the highest (occupied at $T = 0$) valence band. (Based on J.C. Philips *et al.*, *Proc. Int. Conf. Phys. Semiconductors* (Exeter, UK) (The Institute of Physics, London, 1962) 564.)

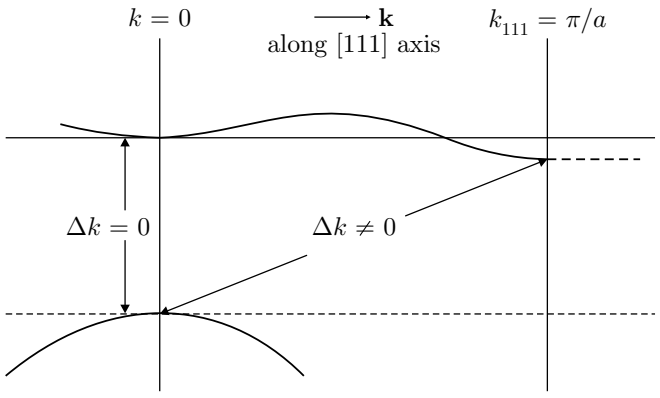


Figure 6.4: Schematic of direct and indirect optical transitions in Ge.

6.1.3 Optical absorption

Photons with energy $h\nu \sim E_g \sim 0-2$ eV will have wavevectors which are $\ll$ typical Brillouin-zone size (energetic X-rays ($h\nu \gg 1$ eV) have wavevectors $\sim$ Brillouin-zone size). This means that transitions involving the absorption of a photon to excite an electron from the valence band to the conduction band, leaving a hole behind, will be essentially vertical in k-space. The transition strengths will be greatest when the joint density of initial and final states is large, *i.e.* when conduction and valence bands are approximately parallel.

Note that Si and Ge are *indirect-gap semiconductors*; the smallest band separation (the thermodynamic band gap, which determines the thermal population of electron and hole states) is not vertical in k-space. Figure 6.4 shows a schematic representation of this situation in Ge; strong optical transitions will occur at the point labelled $\Delta k = 0$ on the diagram, at a higher energy than the thermodynamic band gap, which is labelled $\Delta k \neq 0$.

It turns out that optical transitions *can* occur at energies close to the thermodynamic band gap in Si and Ge if both a phonon and a photon are involved; emission or (at higher temperatures) absorption of a phonon with the appropriate wavevector allows momentum to be conserved so that the transitions labelled $\Delta k \neq 0$ in Figure 6.4 are possible.

Figure 6.5 shows the optical absorption coefficient of Ge at 300 K and 77 K. Both the indirect ($\Delta k \neq 0$) and direct ($\Delta k = 0$) transitions can be seen as absorption edges. The higher energy direct transition is some 100 times stronger than the indirect transition. In addition, the indirect transition is stronger at 300 K than at 77 K, as there is a substantial population of phonons of a suitable wavevector to take part in the transition at the higher temperature; optical transitions assisted by phonon absorption thus becomes possible in addition to those assisted by emission.

Finally, note that the energy gaps change with temperature. This is not unexpected in models of bandstructure; *e.g.* in the tight-binding model of bands, thermal expansion will influence the atomic separation and hence the transfer integrals.

6.1.4 Constant energy surfaces in the conduction bands of Si and Ge

The conduction band minima in Si and Ge (see Figure 6.2 and 6.3) are situated away from the Brillouin-zone centre (the Γ point) and are thus rather anisotropic. At energies close to (*i.e.* $<\sim 50-100$ meV from) the bottom of the conduction bands, effective mass tensors with constant coefficients (as discussed in the problems) can be used to describe the conduction-band minima in Si and Ge. By choosing suitable axes, the tensors can be made diagonal. The constant energy surfaces are approximate ellipsoids of rotation about one of these axes (see Figure 6.6).

6.2 Bandstructure of the direct-gap III-V and II-VI semiconductors

6.2.1 Introduction

Compound semiconductors such as GaAs and InSb are known as *III-V semiconductors* because they combine elements from groups III and V of the Periodic Table. In a similar way, semiconductors such as (Cd,Hg)Te are known as II-VI semiconductors. Many III-V and II-VI semiconductors have technological applications, often as emitters and detectors of electromagnetic radiation. It is important that we examine their bandstructure and consequent properties, in order to see why they are used for such applications. We shall see that one of the most important points is that the minimum energy band-gaps in these materials are direct, so that optical transitions with $\Delta k \approx 0$ can occur.

6.2.2 General points

Figure 6.7 shows the Brillouin zone and calculated bandstructure of GaAs, a typical direct-gap III-V semiconductor. Note that

- both heavy- and light-hole bands are present at the valence-band extremum, as before (*c.f.* Figure 6.3);

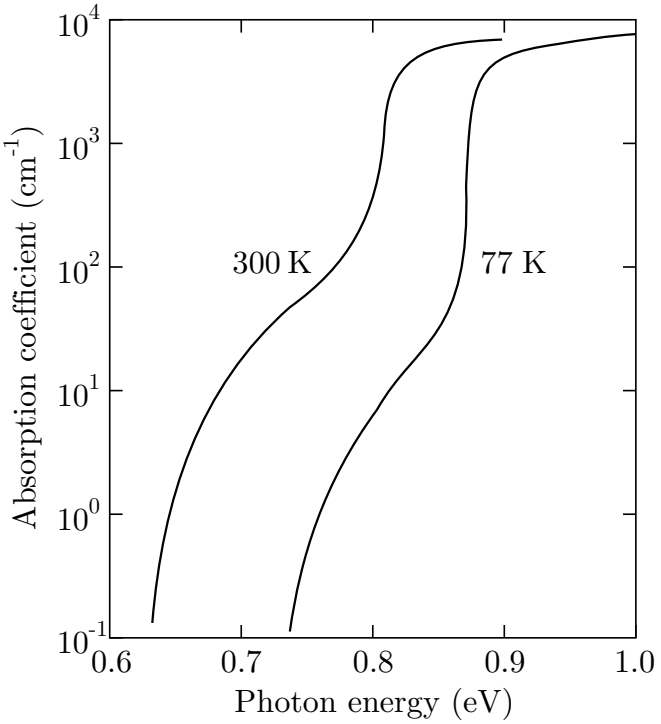


Figure 6.5: Optical absorption of Ge at 300 K and 77 K. (Data from R. Newman and W.W. Tyler, *Solid State Physics* **8**, 49 (1959).)

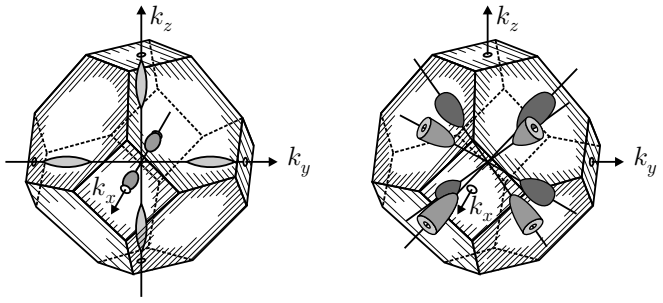


Figure 6.6: Constant energy surfaces of Si (left) and Ge (right).

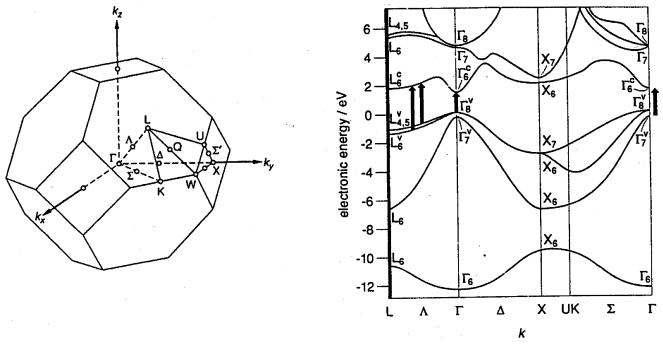


Figure 6.7: Left-hand side: Brillouin zone of GaAs. Right-hand side: bandstructure of GaAs. (After D. Long, *Energy bands in semiconductors* (Wiley, New York, 1968).)

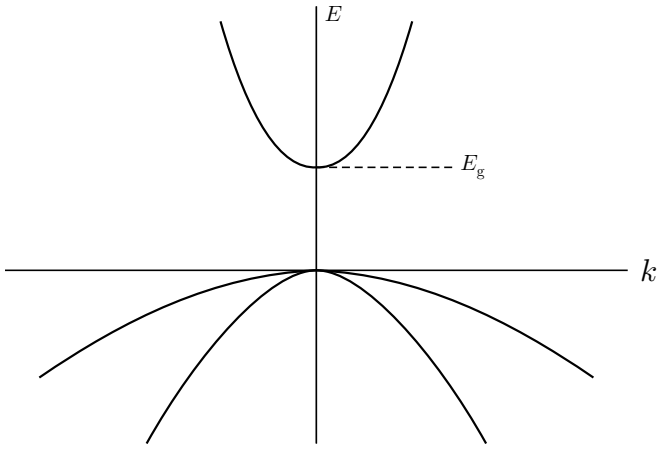


Figure 6.8: Schematic of bandstructure of GaAs close to the direct band gap. The origin of the energy scale is taken as the top of the valence band.

Material	E_g	m_c^*	m_{lh}^*	m_{hh}^*
GaAs	1.52	0.067	0.082	0.45
InAs	0.42	0.023	0.025	0.41
InSb	0.24	0.014	0.016	0.40

Table 6.1: Energy gaps (in eV), and effective masses (in units m_e) of typical III-V semiconductors.

- the thermodynamic band gap separating the highest point of the valence band from the lowest point of the conduction band is direct;
- Just below the valence-band maximum, there is another band, known as the *spin-orbit split-off band*; this is labelled Γ_7^V at the Brillouin-zone centre.

For most practical purposes, we need consider only the electronic dispersion relationships close to the valence band maximum and conduction band minimum; a schematic of this region of k-space is given in Figure 6.8. In fact Figure 6.8 is a good representation of this region for many direct-gap III-V (*e.g.* InSb, InAs) and II-VI (*e.g.* CdTe, $Cd_{0.2}Hg_{0.8}Te$) semiconductors. Table 6.1 shows band-edge effective masses for three examples. Note that

- the light-hole effective mass $m_{lh}^* \approx m_c^*$, the electron (conduction-band) effective mass;
- m_c^* and m_{lh}^* scale roughly with E_g;
- the heavy-hole effective mass m_{hh}^* is almost material-independent.

The fact that wider band gap semiconductors have larger conduction-band and light-hole effective masses is rather a general feature of semiconductors and insulators; it will be an important factor in the discussion below.

6.2.3 Optical absorption

Figure 6.9 shows the optical absorption coefficient of GaAs at several temperatures. Note that the lowest energy band-gap is direct ($\Delta k = 0$); in contrast to the case of Ge, there is no indirect band-gap absorption at energies below the direct gap. There is, nevertheless, one marked similarity to the case of Ge; the direct energy-gap rises as the temperature falls.

At lower temperatures, an extra peak emerges close to the onset of absorption. This is due to the formation of *an exciton*, a bound electron-heavy hole pair. [3] (As stated above, the heavy holes

[3]Excitons are also seen just below the direct band gaps of semiconductors such as Ge and Si at sufficiently low temperatures; see *e.g.* T.P. McLean, in Progress in Semiconductors, **5** (Wiley, 1960).

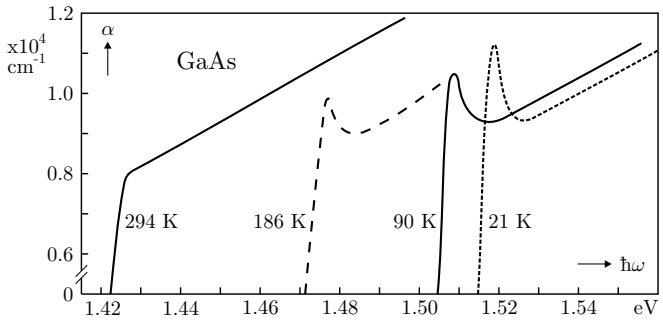


Figure 6.9: Optical absorption coefficient of GaAs at several different temperatures. Note that the joint density of states of the electrons and holes might be expected to give an absorption coefficient which is initially proportional to $(\hbar\omega - E_g)^{1/2}$ (see Section 5.2.4), as the conduction and valence bands are almost parabolic close to the band edges. However, this is masked by the contribution of the exciton to the joint density of states. (Data from M.D. Sturge, *Phys. Rev.* **127**, 768 (1962).)

dominate the spectra because of their large density of states.) Excitons and the optical properties of semiconductors will be covered in the *Optical Properties of Solids* part of the course.

6.2.4 Constant energy surfaces in direct-gap III-V semiconductors

Figure 6.7 shows that the conduction-band minimum and valence-band maximum of GaAs are at the Brillouin-zone centre (the Γ point), a point of very high symmetry. Close to these points, the bands are quite isotropic (especially the conduction band); for many purposes, each band can therefore be treated using a single, constant effective mass (see Table 6.1) to a reasonable degree of accuracy, as long as the carrier energies do not exceed a few tens of meV.

6.3 Thermal population of bands in semiconductors

6.3.1 The law of Mass-Action

The most notable thing about semiconductors (compared to metals) is the fact that the number of mobile charge carriers varies with temperature. We shall now derive the equation governing this behaviour. The number density of electrons $\mathrm{d}n$ with energy between E and $E + \mathrm{d}E$ will be given by

$$\mathrm{d}n = f_{\mathrm{D}}(E,T)g(E)\mathrm{d}E, \tag{6.1}$$

where f_{D} is the Fermi-Dirac distribution function

$$f_{\mathrm{D}} = \frac{1}{e^{\frac{(E-\mu)}{k_{\mathrm{B}}T}} + 1} \tag{6.2}$$

$g(E)$ is the density of states and μ is the chemical potential.[4]

Distribution functions for electrons and holes Now let us consider the case that μ is within the band gap, such that the chance of occupancy of a conduction-band state is $\ll 1$. This implies that $(E-\mu) \gg k_{\mathrm{B}}T$, so that the Fermi-Dirac distribution function f_{DC} for an electron in the conduction band can be approximated as

$$f_{\mathrm{DC}} \approx \frac{1}{e^{\frac{(E-\mu)}{k_{\mathrm{B}}T}}} = e^{-\frac{(E-\mu)}{k_{\mathrm{B}}T}}. \tag{6.3}$$

[4]In the next bit of the derivation we are going make a distinction between the Fermi-Dirac distribution function in the valence and conduction bands, as though it becomes two different functions. Of course it does not; however, differing approximations can be made for the two bands, so that some sort of short-hand labelling which distinguishes the two is desirable.

The chance of finding a hole in the valence band will be $1 - f_{\mathrm{DV}}$, where f_{DV} is the Fermi-Dirac distribution function within the valence band. Under similar conditions to those used in the derivation of Equation 6.3, $(E - \mu)$ will be negative and $|E - \mu| \gg k_{\mathrm{B}}T$. Therefore the approximation

$$1 - f_{\mathrm{DV}} = 1 - (1 + e^{\frac{(E-\mu)}{k_{\mathrm{B}}T}})^{-1} \approx 1 - (1 - e^{\frac{(E-\mu)}{k_{\mathrm{B}}T}}) = e^{-\frac{(\mu-E)}{k_{\mathrm{B}}T}} \tag{6.4}$$

can be made.

Density of states We assume that the conduction and valence bands near the band edges are parabolic, each characterised by a single effective mass. (Even though there are both heavy and light holes, this is not a bad approximation, as the heavy holes have a much bigger effective mass and hence density of states; the light holes can be neglected.) The densities of states for the conduction band (g_{c}) and valence band (g_{v}) are therefore

$$g_{\mathrm{c}} = C m_{\mathrm{c}}^{*\frac{3}{2}} (E - E_{\mathrm{c}})^{\frac{1}{2}} \tag{6.5}$$

and

$$g_{\mathrm{v}} = C m_{\mathrm{hh}}^{*\frac{3}{2}} (E_{\mathrm{v}} - E)^{\frac{1}{2}} \tag{6.6}$$

respectively, where m_{c}^* and m_{hh}^* are the electron and heavy hole effective masses respectively, E_{c} and E_{v} are the energies of the conduction and valence band edges respectively and $C = (1/2\pi^2)(2/\hbar^2)^{3/2}$ (see Section 5.2.4).

Combining Equations 6.1, 6.3 and 6.5, the number of electrons in the conduction band per unit volume is

$$n = \int_{E_{\mathrm{c}}}^{\infty} f_{\mathrm{DC}} g_{\mathrm{c}} \mathrm{d}E \approx C m_{\mathrm{c}}^{*\frac{3}{2}} \int_{E_{\mathrm{c}}}^{\infty} (E - E_{\mathrm{c}})^{\frac{1}{2}} e^{-\frac{(E-\mu)}{k_{\mathrm{B}}T}} \mathrm{d}E. \tag{6.7}$$

Making the substitution $y = (E - E_{\mathrm{c}})/(k_{\mathrm{B}}T)$ gives

$$n \approx C(m_{\mathrm{c}}^* k_{\mathrm{B}}T)^{\frac{3}{2}} e^{-\frac{(E_{\mathrm{c}}-\mu)}{k_{\mathrm{B}}T}} \int_0^{\infty} y^{\frac{1}{2}} e^{-y} \mathrm{d}y, \tag{6.8}$$

leading to

$$n = 2\frac{(2\pi m_{\mathrm{c}}^* k_{\mathrm{B}}T)^{\frac{3}{2}}}{h^3} e^{-\frac{(E_{\mathrm{c}}-\mu)}{k_{\mathrm{B}}T}} = N_{\mathrm{c}} e^{-\frac{(E_{\mathrm{c}}-\mu)}{k_{\mathrm{B}}T}}. \tag{6.9}$$

Here N_{c} is the effective number density of *accessible* states at the conduction band bottom.

Similarly, combining Equations 6.1, 6.4 and 6.6, p, the number of holes in the valence band per unit volume is

$$p = \int_{-\infty}^{E_{\mathrm{v}}} (1 - f_{\mathrm{DV}}) g_{\mathrm{v}} \mathrm{d}E \approx C m_{\mathrm{hh}}^{*\frac{3}{2}} \int_{-\infty}^{E_{\mathrm{v}}} (E_{\mathrm{v}} - E)^{\frac{1}{2}} e^{-\frac{(\mu-E)}{k_{\mathrm{B}}T}} \mathrm{d}E. \tag{6.10}$$

The same type of substitution as in the case of the electrons leads to

$$p \approx 2\frac{(2\pi m_{\mathrm{hh}}^* k_{\mathrm{B}}T)^{\frac{3}{2}}}{h^3} e^{-\frac{(\mu-E_{\mathrm{v}})}{k_{\mathrm{B}}T}} = N_{\mathrm{v}} e^{-\frac{(\mu-E_{\mathrm{v}})}{k_{\mathrm{B}}T}}, \tag{6.11}$$

where N_{v} is the effective number density of *accessible* states at the valence band top.

Combining Equations 6.9 and 6.11 gives

$$np \approx N_{\mathrm{c}} N_{\mathrm{v}} e^{-\frac{E_{\mathrm{g}}}{k_{\mathrm{B}}T}} = WT^3 e^{-\frac{E_{\mathrm{g}}}{k_{\mathrm{B}}T}}, \tag{6.12}$$

where the energy gap $E_{\mathrm{g}} = E_{\mathrm{c}} - E_{\mathrm{v}}$, and W is a constant depending on the details of the valence and conduction-band extrema. This is known as the *Law of Mass-Action*. The exponential dependence of np on the temperature is the chief mechanism determining the temperature dependence of the electrical conductivity of semiconductors.

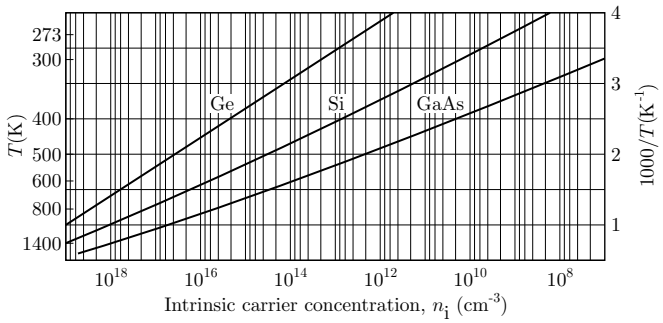


Figure 6.10: Calculated intrinsic carrier densities n_i versus temperature in Ge ($E_\mathrm{g} = 0.74$ eV), Si ($E_\mathrm{g} = 1.17$ eV) and GaAs ($E_\mathrm{g} = 1.52$ eV).

6.3.2 The motion of the chemical potential

We now consider what happens to the chemical potential as a function of T. Setting $n = p$ and equating Equations 6.9 and 6.11 gives

$$\frac{N_\mathrm{c}}{N_\mathrm{v}} = e^{\frac{(2\mu - E_\mathrm{c} - E_\mathrm{v})}{k_\mathrm{B}T}}. \tag{6.13}$$

Now $(N_\mathrm{c}/N_\mathrm{v}) = (m_\mathrm{c}^*/m_\mathrm{hh}^*)^{3/2}$ (see Section 5.2.4), so that the chemical potential is given by

$$\mu = \frac{1}{2}(E_\mathrm{c} + E_\mathrm{v}) + \frac{3}{4}k_\mathrm{B}T \ln\left(\frac{m_\mathrm{hh}^*}{m_\mathrm{c}^*}\right). \tag{6.14}$$

Therefore μ starts off in the middle of the bandgap (*i.e.* exactly half way between the empty and full states) at $T = 0$; as m_hh^* is bigger than m_c^* in most semiconductors, μ usually moves up as T rises.

6.3.3 Intrinsic carrier density

Let us first assume that the only source of electrons in the conduction band and holes in the valence band is the thermal excitation of electrons across the band gap (*i.e.* there are no impurities; see Section 6.3.4 below); such electrons and holes are known as *intrinsic* carriers. In this case (see Equation 6.12)

$$n_\mathrm{i} = n = p = T^{\frac{3}{2}} W^{\frac{1}{2}} e^{-\frac{E_\mathrm{g}}{2k_\mathrm{B}T}}, \tag{6.15}$$

where n_i is the *intrinsic carrier density*. Note that the exponential contains $E_\mathrm{g}/2$ and *not* E_g; this is because the creation of each electron in the conduction band automatically makes a hole in the valence band.

Figure 6.10 shows n_i for Ge, Si and GaAs as a function of T; the exponential dependence on E_g means that relatively small differences in band gap result in n_is which are several orders of magnitude apart.

6.3.4 Impurities and extrinsic carriers

For virtually all practical applications (*i.e.* those at room temperature), the conductivity of semiconductors such as GaAs, Ge and Si is dominated by *extrinsic carriers*, those provided by doping the semiconductor with impurities.

Two types of impurity will concern us, *donors*, which donate electrons, and *acceptors*, which provide holes.

- **Donors** An atom of the semiconductor with valence $\mathcal{V}$ is replaced by a donor atom, with valence $\mathcal{V} + 1$. After the donor has bonded into the host lattice, it will still have one electron left over. Examples of donors include P in Si or Ge, Si on a Ga site in GaAs, and Te on an As site in GaAs.

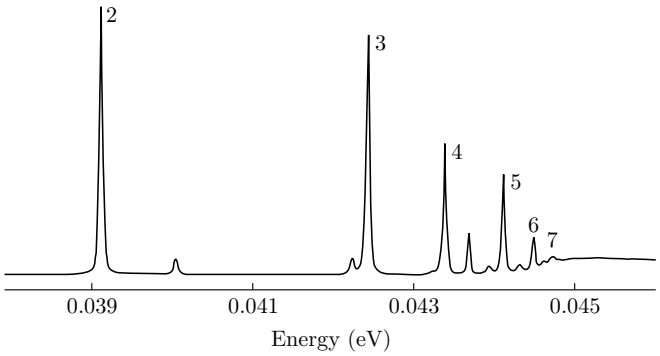


Figure 6.11: Far-infrared absorption due to P impurities (donors) in Si. Hydrogen-like transitions between the $n = 1$ groundstate and higher levels can clearly be seen. (Data courtesy of Prof. R.A. Stradling.)

- **Acceptors** An atom of the semiconductor with valence $\mathcal{V}$ is replaced by an acceptor atom, with valence $\mathcal{V} - 1$. Bonding into the lattice will therefore leave the acceptor short of one electron, *i.e.* with a surplus hole. Typical acceptors include Al and Ga in Si and Ge, Be on a Ga site in GaAs and Si on an As site in GaAs.

The surplus carrier (electron or hole) will orbit the (positively or negatively) charged ionic core of the impurity; again (see Section **??**) the situation is analogous to Hydrogen, with the following alterations (*c.f.* Section **??**)

- the free-electron mass m_e in the Hydrogen is replaced by the effective mass of the carrier;
- the carrier moves through a medium of relative permittivity ϵ_r.

Thus, the energy levels of *e.g.* a donor are given by

$$E(n) = \frac{m_c^*}{m_e}\frac{1}{\epsilon_r^2} \times \frac{13.6 \text{ eV}}{n^2}, \tag{6.16}$$

with n an integer (see Figure 6.11 for a graphic illustration of these energy levels).[5]

As in the exciton model, $n = 1$ represents the ground state (the 1s state) of the donor, and $n = \infty$ represents a free electron in the conduction band (the "ionisation" of the donor). Similar considerations hold for acceptors.

Substituting typical values for ϵ_r ($\epsilon_r \sim 10 - 15$ for many semiconductors) and m_c^* (*e.g.* $m_c^* \approx 0.07 m_e$ in GaAs) into Equation 6.16 gives an impurity binding energy $E(1) \approx 5$ meV. Semiconductors with wide band gaps will tend to have electrons and light holes with heavier effective masses, resulting in larger impurity binding energies, whilst narrow-gap semiconductors will tend to have small impurity binding energies (see Section 6.2.2).

The hydrogen-like approach also yields an effective Bohr radius $a_B^* \sim 10$ nm. The large Bohr radius of the donor implies that the donated electron will spend very little time close to its "parent" ionic core. Therefore the impurity binding energy of a donor will be almost entirely determined by the host semiconductor rather than the parent ion.

6.3.5 Extrinsic carrier density

As long as the numbers of donors and/or acceptors are small enough such that the chemical potential remains in the bandgap (*i.e.* the chance of occupancy of an electron state in the conduction band or a

[5] The perceptive reader will notice that the transitions are *not quite* at the energies predicted by the simple model of Hydrogen. In Ge and Si, a further complication arises from the multi-valley nature of the conduction bands; in Si, for example, there are six impurity states from six minima. These interact so that the final pattern of levels observed is not precisely hydrogenic (the situation in direct-gap semiconductors with a single, zone-centre conduction-band minimum (*e.g.* GaAs) is rather closer to the hydrogenic model). Similarly, the degeneracy of the valence band modifies the simple hydrogenic acceptor levels.

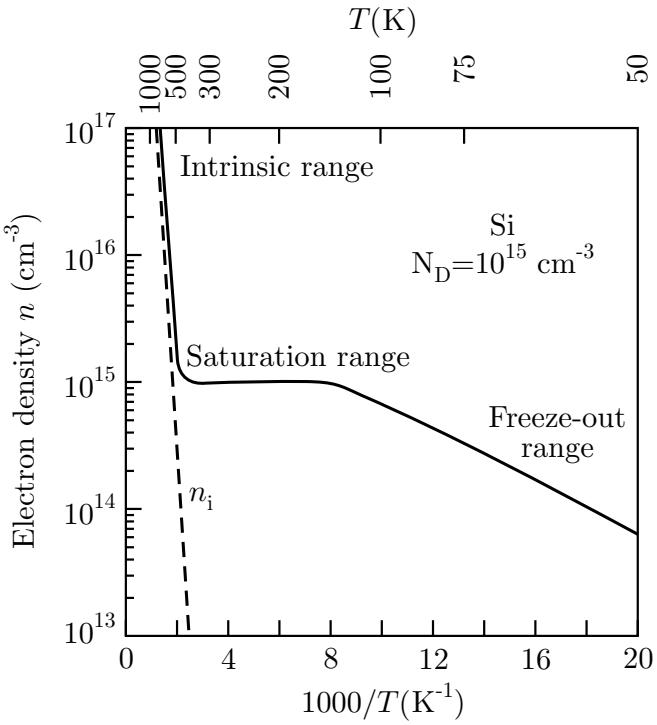


Figure 6.12: Temperature dependence of the electron density in Si with a net donor density of $N_{\rm D}-N_{\rm A} = 10^{15}{\rm cm}^{-3}$.

hole state in the valence band is much less than 1), the derivation of Section 6.3.1 will be applicable, so that the Law of Mass Action will still hold.

In order to find n and p when impurities are present we therefore use Equation 6.12

$$np = T^3 W e^{-\frac{E_{\rm g}}{k_{\rm B}T}}$$

combined with the conservation law

$$n - p = N_{\rm D} - N_{\rm A}, \tag{6.17}$$

where $N_{\rm D}$ is the density of donors and $N_{\rm A}$ is the density of acceptors (both are assumed to only provide one carrier each). The right-hand side of Equation 6.17 shows that the presence of donors can be *compensated* by acceptors; this is known (surprise, surprise!) as *compensation*.

Binding energies of order $5-10$ meV imply that most impurities in semiconductors such as GaAs and Si will be ionised at room temperature.[6] Figure 6.12 shows the temperature dependence of the electron density in the conduction band of Si with a net donor density of $N_{\rm D} - N_{\rm A} = 10^{15}{\rm cm}^{-3}$. At cryogenic temperatures the extrinsic electrons *freeze out* onto the donors; however above about 200 K, all of the donors are ionised (the saturation range) and n is relatively constant. At ~ 500 K the intrinsic contribution to n becomes non-negligible and then rapidly dominates.

A comparison of Figures 6.10 and 6.12 shows that the $n_{\rm i}$ in GaAs and Si are very low indeed at room temperature, much less than practically-achievable values of $N_{\rm D} - N_{\rm A}$; samples of these semiconductors would have to be free of donors to better than about 1 part in $10^{12} - 10^{15}$ (impossible!) for one to see something approaching true intrinsic behaviour.

Finally, we examine the behaviour of the chemical potential in the presence of impurities. To see where μ goes, we repeat the derivation of Section 6.3.2, but using the number density of empty donor states ($N_{\rm D}e^{-(\mu-E_{\rm D})/k_{\rm B}T}$), where $E_{\rm D}$ is the energy of the donor groundstate) in place of the number of

[6] This implies that $(E_{\rm D} - \mu) \gg k_{\rm B}T$ and $(\mu - E_{\rm A}) \gg k_{\rm B}T$, where $E_{\rm D}$ and $E_{\rm A}$ are the energies of the donor and acceptor groundstates respectively. Therefore μ will not only be far from the band edges, it must also be far from the impurity levels.

holes in the valence band.[7] (We are thus considering an n-type semiconductor; however, an equivalent derivation for a p-type semiconductor uses the same ideas.) This gives

$$\mu = \frac{1}{2}(E_{\rm D} + E_{\rm c}) + \frac{1}{2}k_{\rm B}T\ln(\frac{N_{\rm D}}{N_{\rm c}}), \tag{6.18}$$

i.e. μ can either rise or fall with increasing temperature.[8]

6.3.6 Degenerate semiconductors.

Once all of the donors are ionised (the exhaustion regime), $n = N_{\rm D}$; equating Equation 6.9 to $N_{\rm D}$ gives

$$\mu \approx E_{\rm c} + k_{\rm B}T\ln(\frac{N_{\rm D}}{N_{\rm c}}), \tag{6.19}$$

where the $\approx$ re-emphasises that we are approaching a dodgy situtation as far as the approximations used to derive Equation 6.3 are concerned. There are two cases to note.

- $N_{\rm D} < N_{\rm c}$ leads to a chemical potential which is still in the bandgap. The number of electrons in the conduction band is still small compared to the number of available states, so that the quasi–classical statistics used in Equation 6.3 and the derivation that follows it are probably valid. In this case, the semiconductor is *non-degenerate*.
- $N_{\rm D} > N_{\rm c}$ leads to a chemical potential in the conduction band. Therefore we cannot use the quasi-classical approximation for the Fermi-Dirac distribution function. The semiconductor behaves somewhat like a metal and is said to be *degenerate*.

6.3.7 Impurity bands

If a large enough concentration of impurities is placed in a semiconductor, the wavefunctions of the carriers bound to the impurities will start to overlap; the criterion for this for *e.g* donors is $N_{\rm D} \geq\sim 1/a_{\rm B}^{*3}$. The situation becomes rather similar to the tight-binding model, and carriers will be able to move from impurity to impurity. The semiconductor will remain electrically conductive even at very low temperatures; it is another instance of a *degenerate semiconductor*. However, in contrast to the case discussed in the previous section, the chemical potential will depend rather less on the temperature; we will have a situation much more analogous to a (rather dirty) conventional metal.

6.3.8 Is it a semiconductor or an insulator?

Most textbooks state that insulators are just wide-gap semiconductors, the implication being that thermal effects excite a negligible number of electron-hole pairs across the band gap if it is wide (see Equation 6.12).

However, this picture is an over-simplification; as we have seen, the carriers which populate the bands at room temperature come mostly from impurities. Section 6.2.2 stated that carrier effective masses become heavier as the band gap increases; as the effective mass increases, the impurity binding energy will increase (see Equation 6.16). As the gap widens, eventually the impurity will become so deep that it will not be ionised at room temperature.

The distinction between wide-gap semiconductors and insulators is very blurred; *e.g.* studies of doped diamond for use in potential ultra-violet/blue LEDs are carried out, the aim being to find donors and acceptors which are shallow enough not to lead to carrier freeze-out at sensible operating

[7] So as not to complicate the discussion, the expression used for the number of empty donor states is a simplification. Those intending to embark on research should note that the number should be $N_{\rm D}(1 - 1/(1 + g^{-1}e^{(\mu - E_{\rm D})/k_{\rm B}T}))$, where g is the degeneracy of the donor groundstate. To see where g comes from, consider the 1s state of a hydrogenic donor. This is technically doubly degenerate owing to the fact that electrons can have spin up or spin down. Double occupancy of the donor is not energetically favourable; however, the statistical mechanics "knows" that there are two states available at that energy. See *e.g. Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991).

[8] Beware! The derivation of this Equation assumes that $E_{\rm c} - \mu$ and $\mu - E_{\rm D}$ are both much greater than $k_{\rm B}T$. We have already seen that $E_{\rm c} - E_{\rm D} >\sim k_{\rm B}T$ at room temperature; therefore this formula is only valid at much lower temperatures. It is, nevertheless, a useful guide to the low-temperature movement of μ.

temperatures of $\sim$ 300 K. Conversely, much work has been done to dope semiconductors like GaAs with so-called *deep traps*, impurity states which are deliberately made so deep that the electrons and holes will remain frozen on them even at room temperature. Excess impurities will "soak up" any other carriers that are around, giving very high resistivities indeed. Such *semi-insulating* material is often used as a substrate to support devices made from more highly-conductive GaAs and (Ga,Al)As.

6.3.9 A note on photoconductivity.

It has been mentioned that semiconductors are often employed to detect photons. The detection involves *photoconductivity*, which can occur via three mechanisms;

1. the creation of electron-hole pairs through interband (intrinsic) absorption of radiation; typically this occurs at photon energies $\sim E_g \sim 1$ eV or wavelengths $\lambda \sim 1$ μm; later we shall see that by alloying semiconductors such as CdTe with semimetals such as HgTe, materials with band gaps as small as 0.1 eV (corresponding to $\lambda \sim 10$ μm) can be made;

2. the excitation of electrons (or holes) from donor (acceptor) impurities into the conduction (valence) band; typically this occurs at energies ~ 5 meV to 50 meV, corresponding to $\lambda \sim 200 - 20$ μm;

3. free-carrier absorption.

In cases 1 and 2, the carrier numbers are increased; in case 3, the number of carriers remains constant but the conductivity changes because of a change in mobility. The latter mechanism is usually important in cases where the carrier effective mass is low and the band is strongly nonparabolic [9] as the low density of band states (from the small effective mass) results in long relaxation times for the excited carriers. Free-carrier absorption is proportional to the real part of the conductivity $\sigma = \sigma_0/(1 + \mathrm{i}\omega\tau)$, and so increases rapidly at low energies/long wavelengths.

6.4 Reading.

Useful alternative treatments are given in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Chapters 17-19, *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Sections 10.10-10.20 and Chapter 18, and *Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) Chapter 1 and Sections 2.1-2.3, 2.7, 4.4, 5.1-5.3, 9.1-9.2. Optical properties of semiconductors are tackled in depth in *Optical Properties of Solids*, by A.M. Fox (OUP 2000).

Essentials of semiconductor physics by Tom Wenckebach (Wiley, New York, 1999) provides a slightly offbeat, but algebraically rigorous treatment of many of the properties of semiconductors which would be useful for those embarking on research or calculations. A similarly detailed but more traditional approach is given in *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991)and *Fundamentals of semiconductors*, by P. Yu and M. Cardona (Springer, Berlin, 1996). The methods used to calculate the bandstructure of semiconductors have been reviewed by F. Bassani (Chapter 2 of *Semiconductors and Semimetals*, Volume 1, edited by R.K. Willardson and A.C. Beer (Academic Press, New York, 1966)) and E.O. Kane (Chapter 4 of *Handbook on Semiconductors*, volume 1, edited by T.S. Moss (North Holand, Amsterdam, 1982)); the latter reference includes the famous **k.p** method for evaluating the bandstructure close to band extrema (see also *Fundamentals of semiconductors*, by P. Yu and M. Cardona (Springer, Berlin, 1996) Chapter 2.).

A very wide range of properties of semiconductors (band gaps, effective masses *etc.*) can be looked up in the comprehensive Landholt-Börnstein series, *Numerical data and functional relationships in science and technology*, New Series, Group III, Volume 17, edited by O. Madelung (Springer, Berlin, 1982).

Impurities in semiconductors are dealt with comprehensively in the mammoth tome *Doping in III-V semiconductors*, by E.F. Schubert (Cambridge University Press, 1993).

[9] This is the case in the conduction bands of narrow-gap semiconductors such as $Cd_{0.2}Hg_{0.8}Te$ and InSb.

Handout 7

Bandstructure engineering

7.1 Introduction

We shall explore some of the ways in which bands with desired properties can be engineered using what has been termed "chemical architecture". A very simple example of this is the use of *semiconductor alloys*, in which a wide-gap semiconductor and a narrow-gap semiconductor are combined to give a substance with a desired intermediate bandgap. A second example is the *semiconductor superlattice* or *heterostructure*; here, very thin layers of different semiconductors are superimposed.

7.2 Semiconductor alloys

The previous lectures mentioned the fact that direct-gap semiconductors are used in opto-electronic applications. Many of these applications involve the emission of light by recombination of an electron and a hole across the band gap, or absorption of light of the band-gap energy to create an electron-hole pair and hence produce some form of photoconduction;[1] The size of the band gap is therefore of great importance.

Figure 8.1 shows the band gaps and lattice parameters for some of the more common elemental and binary semiconductors. The band gaps available are not generally optimised for practical devices. A desire to improve on this has led to the field of *bandstructure engineering*, where a variety of artificial structures are used to provide bandstructure optimised to a particular application.

The simplest bandstructure engineering involves making *ternary alloys* such as (Ga,Al)As and (Hg,Cd)Te in order to achieve a desired band gap. I list three examples of the uses of such alloys.

$Ga_{1-x}Al_xAs$ is technologically important because GaAs and AlAs form a solid solution over the entire ($0 \leq x \leq 1$) composition range with very little variation of lattice parameter (see Figure 8.1); this means that multilayers of very high quality can be fabricated (see below). A variable direct band gap is obtained for the range $0 \leq x \leq\sim 0.35$; for $x \geq\sim 0.35$ the band-gap is indirect.

The band gap of $Ga_{1-x}In_xAs$ may be adjusted over an energy range which coincides with the low-attenuation region of many optical fibres.

$Hg_{1-x}Cd_xTe$ continues to be of great importance in the fabrication of infrared detectors covering the 10 μm ($x = 0.2$; $E_g \approx 100$ meV) and 5 μm ($x = 0.3$; $E_g \approx 200$ meV) atmospheric windows (regions where the atmosphere has little absorption). The 10 μm region also contains the peak thermal emission of ~ 300 K things, including human beings, so that there are a variety of medical (thermal imaging), meteorological and more sinister applications. Exhausts and jet engines emit well in the 5 μm range, and so the applications of $Hg_{1-x}Cd_xTe$ with $x = 0.3$ can be easily imagined.

[1] See *e.g.* *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991) Chapters 12 and 13; *Optical Properties of Solids* by Mark Fox, (Oxford University Press, Oxford 2001), *Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) Chapter 18.

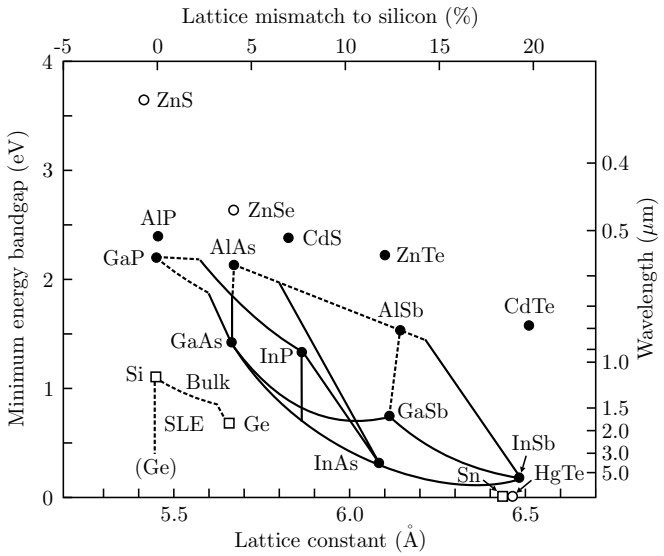


Figure 7.1: Band gap versus lattice parameter for some of the more common semiconductors. The curves indicate commonly-used alloys such as (Ga,Al)As, (Ga,In)As *etc.*; solid curves represent direct gaps and dashed curves indirect gaps.

7.3 Artificial structures

We now turn to a variety of structures which can be grown using techniques developed over the past twenty years or so.

7.3.1 Growth of semiconductor multilayers

The growth of high-quality semiconductor multilayers is described as *epitaxial* (the word is derived from the Greek words *epi* (upon) and *taxis* (arrangement)). The implication is that layers grow on a suitable single-crystal substrate, continuing the crystal structure of that substrate. The layers are thus supposed to be crystallographically well ordered. Two techniques are commonly used to grow epitaxial layers of semiconductors, *Molecular Beam Epitaxy* (MBE) and *Metal-Organic Vapour-Phase Epitaxy* (MOVPE; also known as OMVPE, MOCVD and OMCVD).

- **MBE** Figure 7.2 shows the main components of an MBE machine. The elements which make up the semiconductors to be grown evaporate from *Knudsen cells* at a rate controlled by the cell temperature; cells are also provided for dopants. The evaporated atoms form a beam which travels towards the substrate. Typically the evaporation rates are such that ∼one monolayer is deposited per second. The composition of the layers is controlled by shutters which can swing in front of each cell in ~ 50 ms, cutting off the beam from that cell. The substrate is kept at a well-defined temperature to ensure that the deposited atoms are reasonably mobile, so that they spread out over the substrate in monolayers rather than forming clusters. The whole chamber must be very well evacuated to prevent spurious dopants entering the layers; typical pressures in an MBE machine are $\sim 10^{-11}$ mbar. A number of *in-situ* diagnostic techniques to monitor the growth of the multilayer are provided.

- **MOVPE** Figure 7.3 shows a schematic of an MOVPE machine. The components of the semiconductor to be grown travel to the substrate as gaseous precursors formed by reacting the elements with organic radicals; the precursors are carried along in a stream of hydrogen. Close to the substrate, which is heated by a radiofrequency coil, the precursors react, depositing the semiconductor on the substrate. A typical reaction is

$$\mathrm{Ga(CH)_3 + AsH_3 \rightarrow 3CH_4 + GaAs.} \tag{7.1}$$

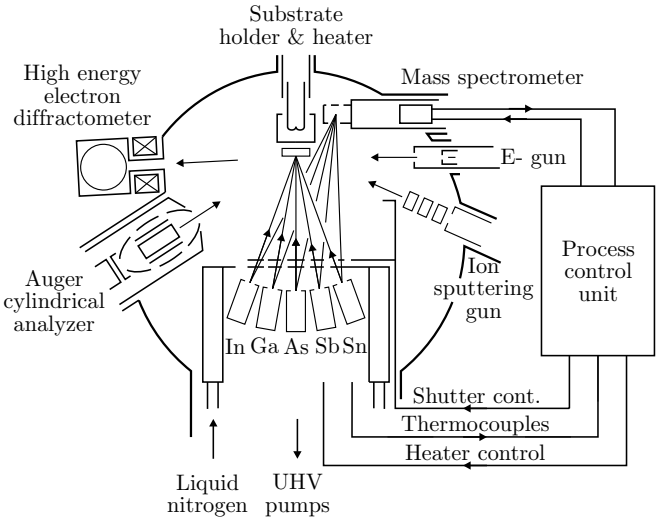


Figure 7.2: Schematic of an MBE machine.

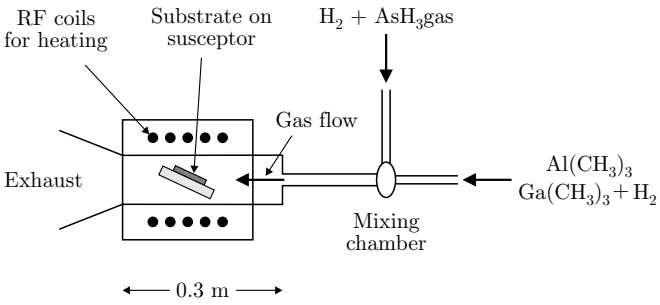


Figure 7.3: Schematic of an MOVPE machine.

Very fast solenoid valves enable the sources of the various components to be switched on and off rapidly.

7.3.2 Substrate and buffer layer

The substrate will be a high quality single crystal, usually of one of the materials to be included in the multilayer. Often the substrate will be *semi-insulating* GaAs (see Section 6.3.8). The first layer to be grown is called the *buffer layer*; it is often the same material as the substrate (but undoped), and is designed to "smooth out" the lumps and bumps of the latter, to provide an atomically-smooth top surface on which to grow the active parts of the multilayer. More recently, *superlattice buffers* consisting of alternating thin ($\sim 1-2$ nm) layers of *e.g.* GaAs and AlAs have sometimes been used for this purpose.

7.3.3 Quantum wells

The simplest multilayer system or *heterostructure* (*hetero* = more than one; *i.e.* the structure is made up of more than one semiconductor) is the *quantum well*. A thick (several hundred nm) layer of a wider gap material such as (Ga,Al)As is grown (the barrier material), followed by a $\sim 2-100$ nm layer of a narrower gap material such as GaAs (the well), followed by another thick layer of wider gap material. Figure 7.4 shows the resulting conduction and valence-band edges; the relative heights of the discontinuities in the conduction- and valence-band edges are instrinsic properties of the two materials involved.

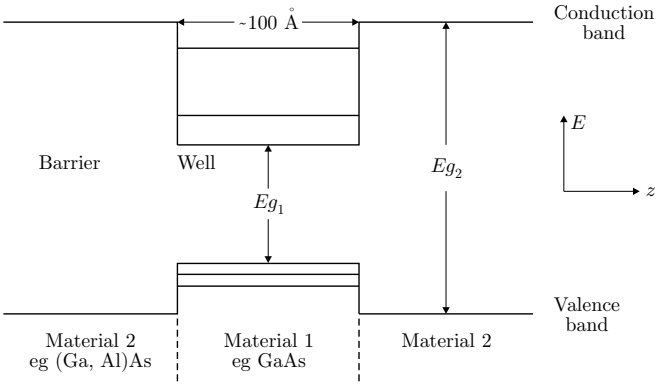


Figure 7.4: Band edges in a quantum well as a function of distance in the growth direction z. The energies of the subbands within the well are shown schematically.

The narrower-gap material forms a one-dimensional potential well in the conduction and valence bands; thus, in the z (growth) direction, the electron and hole levels are bound states of the well, known as *subbands*. The well will contain three sets of subbands,

- the electron subbands,
- the light hole subbands and
- the heavy hole subbands,

each subband within a set being labelled by a quantum number $i = 1,\ 2,\ 3.....$ Motion in the xy plane will be unrestricted; we therefore have a two-dimensional carrier system in the well.

7.3.4 Optical properties of quantum wells

The interband optical absorption or emission of a quantum well will be caused by transitions between hole and electron subbands. The selection rules are determined by the overlap between the electron and hole wavefunctions. Wavefunctions of the same quantum number (or quantum numbers differing by two) will have similar (spatial) shapes and the transition will be strong; wavefunctions whose quantum numbers differ by one will be of opposite symmetry, and so transitions between them will be weak. In summary, the selction rules are[2]

- $\Delta i = 0, 2$: strong, allowed transitions;
- $\Delta i = 1$: weak, "forbidden" transitions.

As there are two sets of hole subbands, strong transitions will occur in pairs, *e.g.* (ith heavy-hole subband to ith electron subband) and (ith light-hole subband to ith electron subband). Figure 7.5 shows the optical absorption of a set of quantum wells of differing well width. The wells are of rather low quality, so that individual transitions are hard to distinguish. However, it is plain that

- the transition energies increase as the wells get narrower, due to the increased subband confinement energy;
- the absorption increases in steps, reflecting the step-like form of the electron-hole joint density of states in two dimensions;
- excitons are very prominent, because the confinement in the well enhances the exciton binding energy by holding the electron and hole closer together than in the bulk semiconductor.

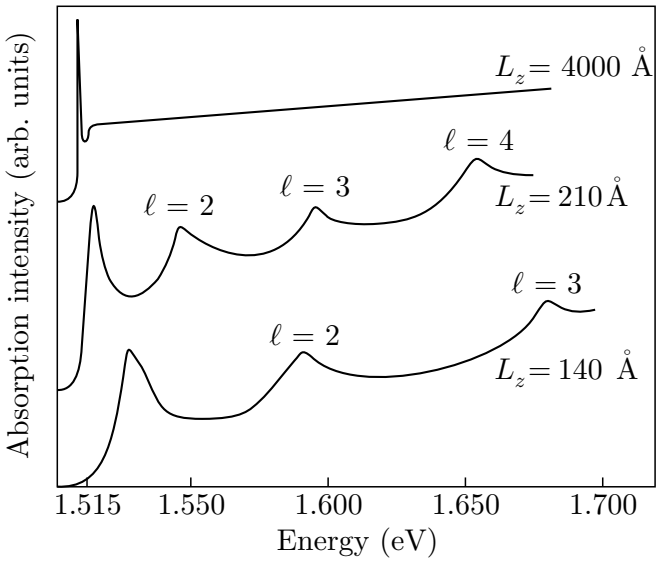


Figure 7.5: Optical absorption of a series of GaAs-(Ga,Al)As quantum wells of differing widths L_z at 4.2 K. (After R. Dingle *et al.*, *Phys. Rev. Lett.* **33**, 827 (1974).)

Figure 7.6 shows the transmission of a set of much higher quality GaAs-(Ga,Al)As quantum wells. Unfortunately the detector used did not give a flat response so that the steps in absorption cannot be seen so clearly; however, the pairs of strong transitions can now be resolved. Note that the excitons involving light holes are almost as strong as excitons involving heavy holes (*c.f.* the bulk case (Figure 6.9), where the latter dominate); this is because quirks of the valence-band bandstructure give the light-hole subbands quite a large in-plane (xy) effective mass, and therefore a large density of states, comparable to that of the heavy holes.

7.3.5 Use of quantum wells in opto-electronics

There are several reasons for the use of quantum wells in opto-electronic applications, *e.g.*

- the energy of the fundamental optical transition can be varied by varying the well width;
- the heavy and light hole degeneracy at the Brillouin-zone centre is broken, removing complications associated with scattering *etc.*;
- all of the transitions are excitonic (*i.e.* sharp features at a well-defined energy, rather than broad edges), even at 300 K;
- the well can be used to hold electrons and holes in close proximity, to encourage more efficient recombination in *e.g.* lasers and LEDs.[3]

Applications include the *Quantum-confined Stark effect modulator*, which is dealt with in detail in *e.g. Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) Section 18.8.

7.3.6 Superlattices

A *superlattice* contains a set of quantum wells which are sufficiently closely spaced for the carriers to tunnel between wells (see Figure 7.7). The situation is analogous to the tight binding or Kronig–Penney models of bandstructure; the subbands will broaden out to form *minibands* with *minigaps* between. One

[2]For a detailed derivation of these selection rules, see M. Fox, *Optical Properties of Solids*, (OUP, 2000).
[3]See *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991) Section 13.2.

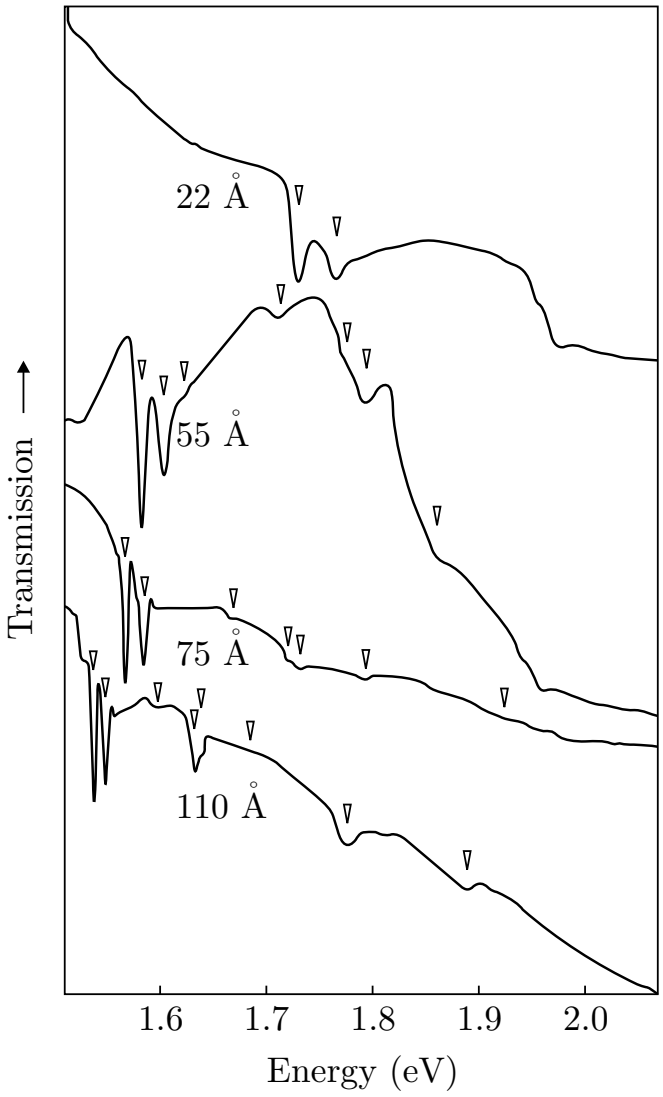


Figure 7.6: Optical absorption of a series of GaAs-(Ga,Al)As quantum wells of differing widths L_z at 55 K. (Data from D.C. Rogers *et al*, *Phys. Rev. B.* **34**, 4002 (1986).)

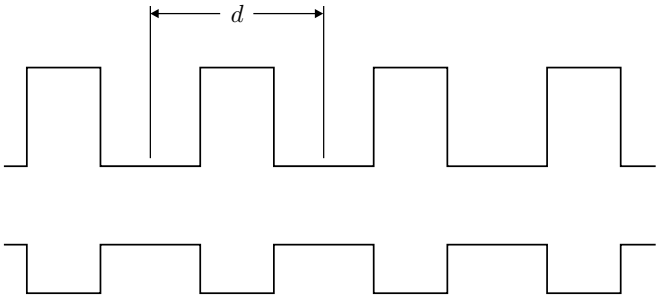


Figure 7.7: Schematic of the band edges in a superlattice.

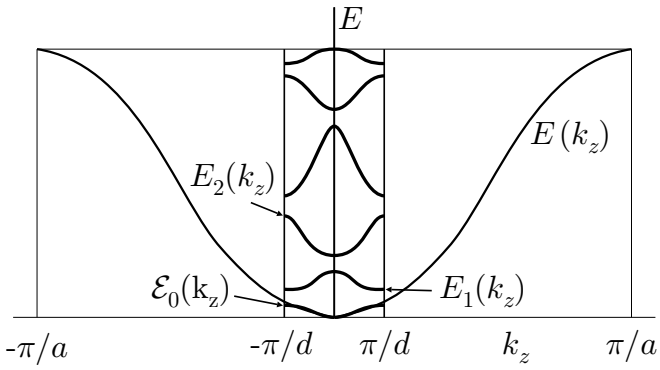


Figure 7.8: Minibands and minigaps in a superlattice (see Figure 7.7). Here, π/a denotes the Brillouin-zone boundary of the underlying lattice; $E(k_z)$ represents the unperturbed, original bulk band. The effect of the superlattice, of period d, is to introduce new "mini Brillouin zones" with boundaries at $\pm\pi/d$. The band is split into *minibands* with dispersion relationships labelled by $E_0(k_z)$, $E_1(k_z)$, $E_2(k_z)$ *etc.*.

can also think of the periodicity of the superlattice introducing a new set of Brillouin-zone boundaries and hence energy gaps (see Figure 7.8).

Superlattices have a variety of applications in *resonant tunnelling structures* and far-infrared detectors; we shall discuss more of the details of these devices in a later lecture.[4]

7.3.7 Heterojunctions and modulation doping

A heterojunction is a single junction between two layers of different semiconductors. A typical example is the GaAs-(Ga,Al)As heterojunction. First, a thick layer of undoped GaAs is grown, followed by a few tens of nm of (Ga,Al)As (the *spacer layer*). After this a section of heavily-doped ($n \sim 10^{18}$ cm^{-3}) (Ga,Al)As is deposited, followed by more undoped (Ga,Al)As (not shown). The final band arrangement is shown in Figure 7.9. The combination of the conduction band offset at the interface and the pinning of the chemical potential inside the GaAs and (Ga,Al)As layers results in a one-dimensional, approximately triangular potential well containing electrons which have "dropped off" the donors in the (Ga,Al)As. As in the case of the quantum well, the electrons' z-direction motion is quantised into subbands; however, the xy motion is unconstrained, so that a degenerate (metallic) two-dimensional electron system results.

The spacer layer separates the ionised donors from the electrons, dramatically reducing the scattering; low-temperature mobilities $\mu_c \sim 10^7 \mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ are possible using this technique, which is known as *modulation doping* (because the dopant concentration in the (Ga,Al)As is modulated).

Applications of heterojunctions include the *High-Electron-Mobility Transistor* (HEMT), an FET with a heterojunction as active layer.[5] Most experiments on the *Quantum Hall Effect* are carried out using heterojunctions; we shall study this effect in detail later.

Other applications of modulation doping include the fabrication of so-called δ-doping layers, in which a very heavy dose of dopant atoms is deposited in a small width ($\sim$ a few monolayers). In effect, this forms a two-dimensional impurity band, a very thin metallic layer within the host semiconductor. Such layers are finding applications as metallic gates and contacts *within* complex, layered semiconductor structures.

7.3.8 The envelope-function approximation

The *envelope-function approximation* is often used to calculate the electronic energy levels in heterostructures, such as GaAs-(Ga,Al)As quantum wells, which are made up from two or more similar semiconductors. Because the two semiconductors involved are very similar from both chemical and

[4] See also See *Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) Chapters 9 and 10, 17-19.

[5] See *Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) Chapter 16.

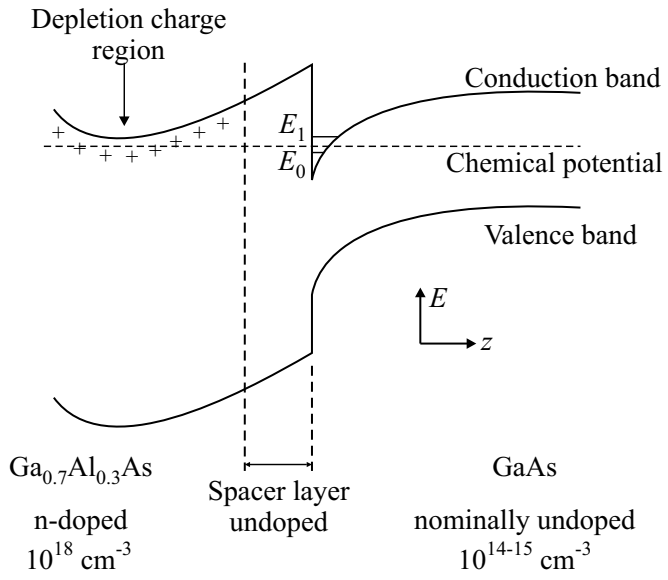


Figure 7.9: The band edges in a modulation-doped GaAs-(Ga,Al)As heterojunction.

crystallographic points of view, it is assumed that the rapidly-oscillating part of the Bloch function (the part which has the periodicity of the lattice) is the same in both materials; only the *envelopes* ϕ of the Bloch functions differ in the two semiconductors.

As an example, consider the GaAs-(Ga,Al)As quantum well shown in Figure 7.4; let a typical envelope function in the GaAs well be $\phi_{\rm A}(z)$ and the corresponding envelope function in the (Ga,Al)As barriers be $\phi_{\rm B}(z)$. The boundary conditions in the Envelope Function Approximation at the interfaces (*i.e.* at $z = \pm\frac{a}{2}$) are

$$\phi_{\rm A} = \phi_{\rm B} \tag{7.2}$$

and

$$\frac{1}{m^*_{\rm A}}\frac{{\rm d}\phi_{\rm A}}{{\rm d}z} = \frac{1}{m^*_{\rm B}}\frac{{\rm d}\phi_{\rm B}}{{\rm d}z}, \tag{7.3}$$

where $m^*_{\rm A}$ and $m^*_{\rm B}$ are the effective masses in the well and barrier respectively. Equation 7.3 ensures that the probability-density flux is conserved.

By considering envelope functions such as $\cos(kz)$ and $\sin(kz)$ in the well and evanescent waves $e^{\pm\kappa z}$ in the barriers, the subband energies can be found. The problems contain an example of this technique, which can easily be extended to more complex structures such as superlattices.

Handout 8

Carbon nanotubes

8.1 Introduction

In the last lecture we started to look at bandstructure engineering and the construction of "artificial" materials. In this section we will take a closer look at one particular nanostructure; the *carbon nanotube*, which is currently the focus of intense research. This system also provides us with an opportunity apply many of the techniques of bandstructure calculation that we have learnt so far.

Carbon nanotubes are single sheets of graphite (called *graphene*) rolled into cylinders. The diameter of the tubes are typically of nanometer dimensions, while the lengths are typically micrometers. This huge aspect ratio leads to unusual electrical transport. Notably some tubes behaving as metals and others as semiconductors.

The bandstructure of graphite was first calculated by Wallace in 1947[1], but it was not until 1991 that multi-walled nanotubes were discovered[2] . We will concentrate on single walled nanotubes which were discovered in 1993.

8.2 Reading

Since carbon nanotubes are currently an area of active research the books on the subject tend to be aimed at a research audience. For those of you who would like to read more about carbon nanotubes I can recommend two books:

1. *Physical properties of carbon nanotubes* by Saito, Dresselhaus and Dresselhaus (Imperial College Press, 1998).

2. *Carbon Nanotubes, basic concepts and physical properties* by Reich, Thomsen and Maultzsch (Wiley, 2004).

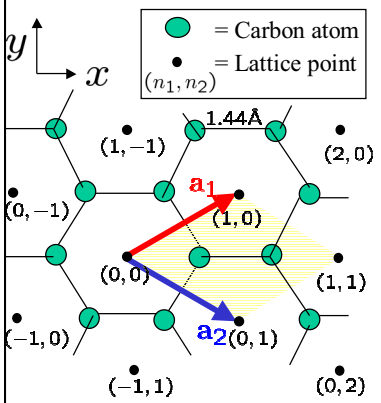
8.3 Lecture slides

The following 7 pages reproduce the slides shown in the lectures.

[1] See P.R. Wallace *Physical Review*, **71**: 622-634 (1947)
[2] S. Iijima, *Nature* **354**: 56-58 (1991)

Graphene and Carbon Nanotubes

Graphene (2D graphite)



Translation vector:

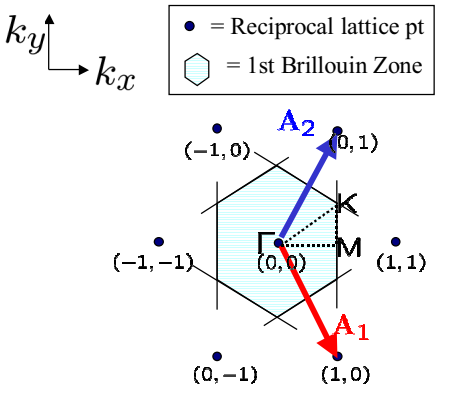
$$\mathbf{T}_{n_1,n_2} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$$

$\mathbf{a}_1 = (\frac{\sqrt{3}a}{2}, \frac{a}{2})$ $\mathbf{a}_2 = (\frac{\sqrt{3}a}{2}, -\frac{a}{2})$

$a = |\mathbf{a}_1| = |\mathbf{a}_2| = 2.46\text{Å}$

2 C atom basis:

$\frac{1}{3}(\mathbf{a}_1 + \mathbf{a}_2)$ and $\frac{2}{3}(\mathbf{a}_1 + \mathbf{a}_2)$



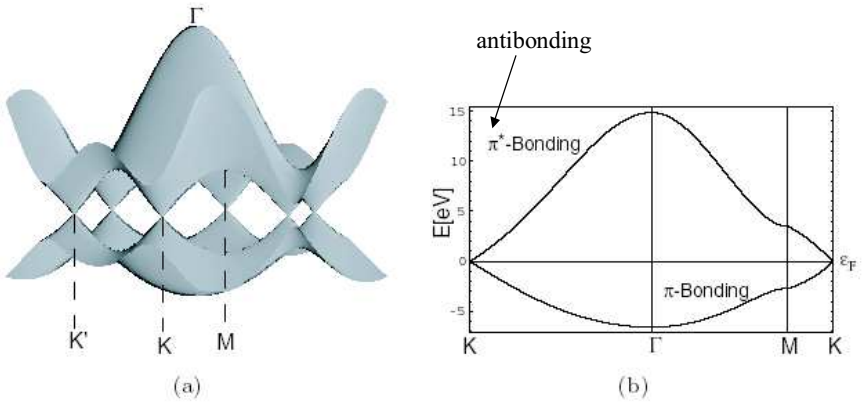
Reciprocal lattice vector:

$$\mathbf{G}_{m_1,m_2} = m_1\mathbf{A}_1 + m_2\mathbf{A}_2$$

$\mathbf{A}_1 = (\frac{2\pi}{\sqrt{3}a}, \frac{2\pi}{a})$ $\mathbf{A}_2 = (\frac{2\pi}{\sqrt{3}a}, -\frac{2\pi}{a})$

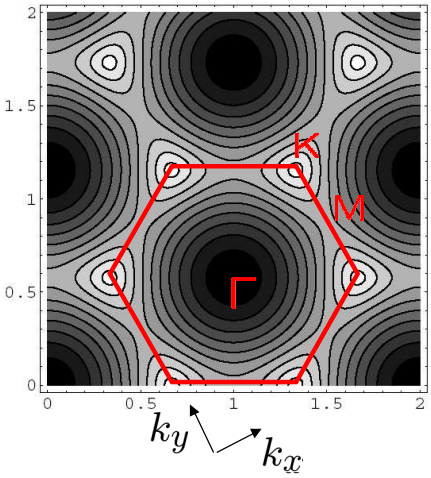
Tight binding Calculation: Graphene

$$E_{\pi \mathbf{k}} = \pm t \left[1 + 4\cos\left(\frac{\sqrt{3}k_x a}{2}\right)\cos\left(\frac{k_y a}{2}\right) + 4\cos^2\left(\frac{k_y a}{2}\right) \right]^{\frac{1}{2}}$$

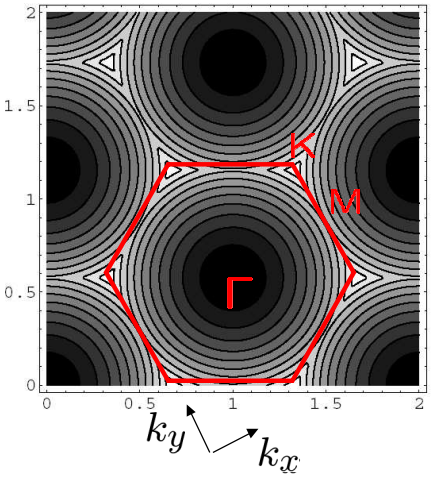


Tight Binding & Nearly Free Electron models applied to Graphene

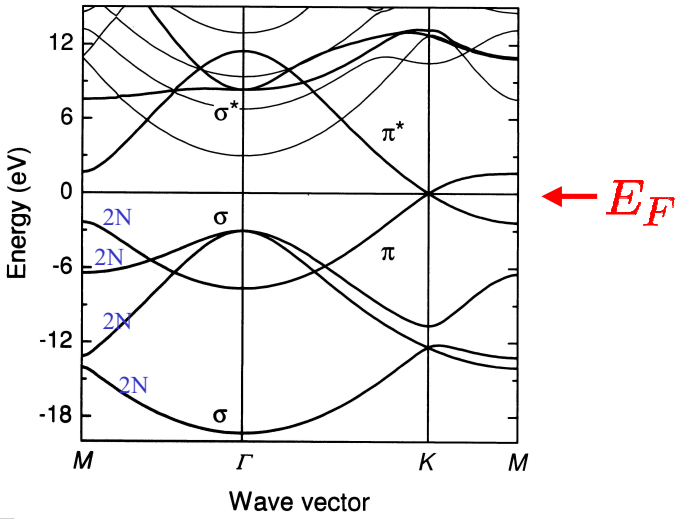
Tight binding Calculation



Nearly Free Electron Calculation

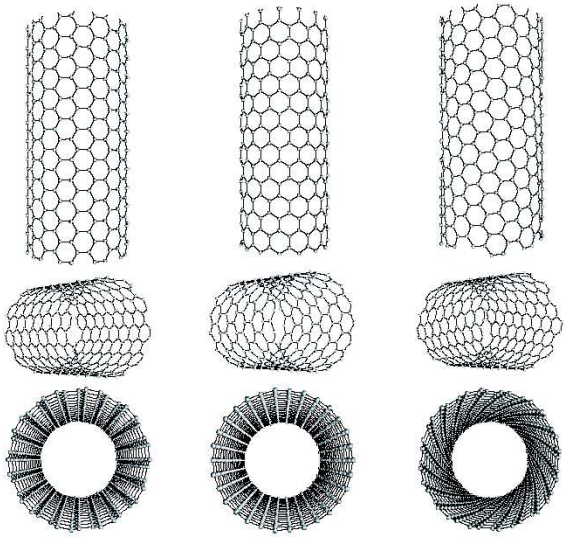


Graphene bandstructure

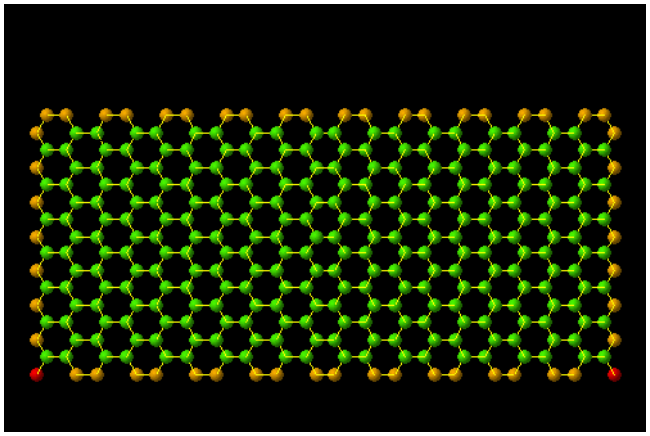


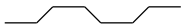
NB, Two C atoms per unit cell (ie 12 electrons, 8 of which are valence)
Non-hybridized (pi) electrons play key role in graphene conductivity

Single walled Carbon nanotubes: Discovered in 1993!



Carbon nanotubes: rolled up graphene!

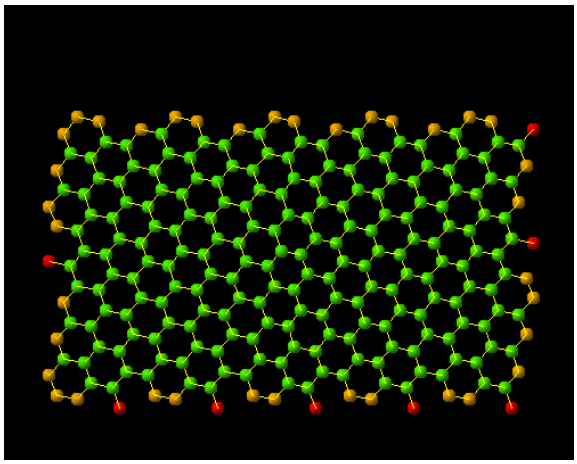


(10,10) Armchair Carbon Nanotubes 

http://www.photon.t.u-tokyo.ac.jp/~maruyama/agallery/agallery.html

Carbon nanotubes

- There are many ways to roll up graphene
- The way that graphene is rolled up drastically changes physical properties



(10,5) Chiral carbon nanotube

http://www.photon.t.u-tokyo.ac.jp/~maruyama/agallery/agallery.html

Carbon Nanotubes (CNTs)

Chiral vector for (n_1, n_2) CNT

$$\mathbf{C}_{CNT} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 = (n_1, n_2)$$

Tube diameter

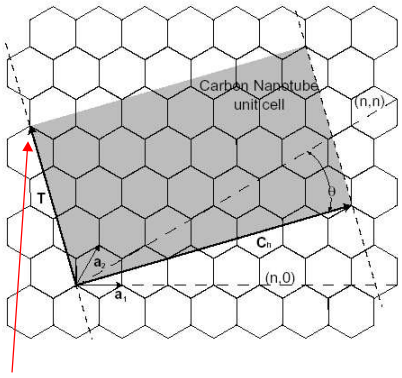
$$d_{CNT} = \frac{|\mathbf{C}_{CNT}|}{\pi} = a\sqrt{n_1^2 + n_2^2 + n_1 n_2}$$

CNT "translation" vector

$$\mathbf{T}_{CNT} = t_1\mathbf{a}_1 + t_2\mathbf{a}_2 = (t_1, t_2)$$

$$t_1 = \frac{2n_2 + n_1}{d_R} \qquad t_2 = -\frac{2n_1 + n_2}{d_R}$$

$d_R =$ Greatest common devisor of $(2n_1 + n_2, n_1 + 2n_2)$



1st lattice point reached!

Number of hexagons in **nanotube** unit cell:

$$N_{CNT} = \frac{2(n_1^2 + n_2^2 + n_1 n_2)}{d_R}$$

Chiral angle (definition)

$$\cos\theta = \frac{\mathbf{C}_{n_1,n_2} \cdot \mathbf{a}_1}{|\mathbf{C}_{n_1,n_2}||\mathbf{a}_1|} = \frac{2n_1 + n_2}{2\sqrt{n_1^2 + n_2^2 + n_1 n_2}}$$

Carbon Nanotubes

Recall Born-von Karman Boundary Conditions for 3D crystal:
Allowed k-states in 1st Brillouin Zone:

$$\mathbf{k} = \frac{m_1}{N_1}\mathbf{A}_1 + \frac{m_2}{N_2}\mathbf{A}_2 + \frac{m_3}{N_3}\mathbf{A}_3$$

- Cyclic BC in directions of $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{a}_3$
- N is very large for typical crystal

But only small number of periods in CNTs, hence we need to consider cyclic boundary conditions carefully!
Reciprocal lattice vectors for CNT unit cell:

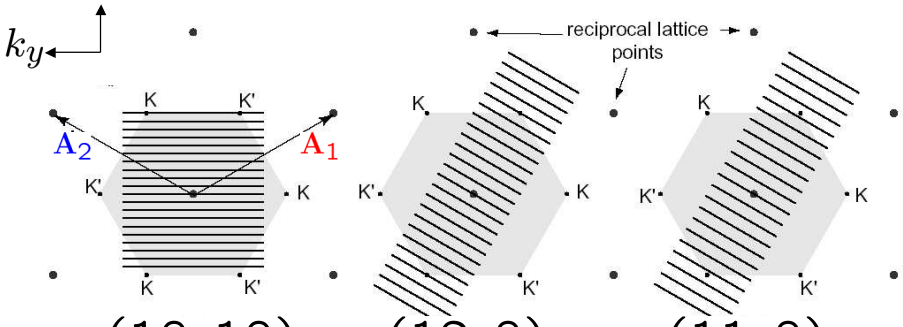
where,

$$\mathbf{K}_1 = \frac{1}{N}(-t_2\mathbf{A}_1 + t_1\mathbf{A}_2) \quad \text{(circumference RLV)}$$

$$\mathbf{K}_2 = \frac{1}{N}(n_2\mathbf{A}_1 - n_1\mathbf{A}_2) \quad \text{(along CNT axis RLV)}$$

If allowed k-state coincides with 1st BZ K point = metallic!

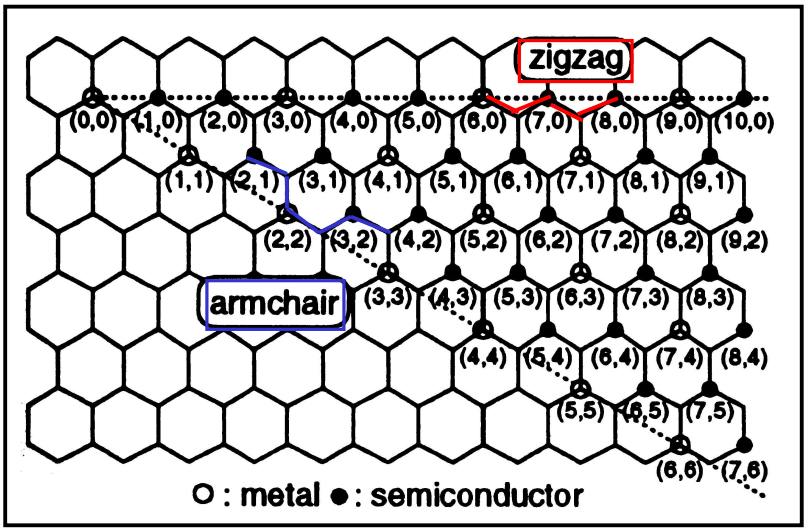
Carbon Nanotubes (examples)



(10, 10)	(12, 0)	(11, 0)
Armchair CNT	Zigzag CNT	Zigzag CNT
$\mathbf{T}_{CNT} = (1, -1)$	$\mathbf{T}_{CNT} = (1, -2)$	$\mathbf{T}_{CNT} = (1, -2)$
$N_{CNT} = 20$	$N_{CNT} = 24$	$N_{CNT} = 22$
$\mathbf{K}_1 = \frac{\mathbf{A}_1}{20} + \frac{\mathbf{A}_2}{20}$	$\mathbf{K}_1 = \frac{2\mathbf{A}_1}{24} + \frac{\mathbf{A}_2}{24}$	$\mathbf{K}_1 = \frac{2\mathbf{A}_1}{22} + \frac{\mathbf{A}_2}{22}$
$\mathbf{K}_2 = \frac{\mathbf{A}_1}{2} - \frac{\mathbf{A}_2}{2}$	$\mathbf{K}_2 = -\frac{\mathbf{A}_2}{2}$	$\mathbf{K}_2 = -\frac{\mathbf{A}_2}{2}$
METAL	METAL	SEMICONDUCTOR

Carbon Nanotubes

Metallic if $2n_1 + n_2$ is a multiple of 3



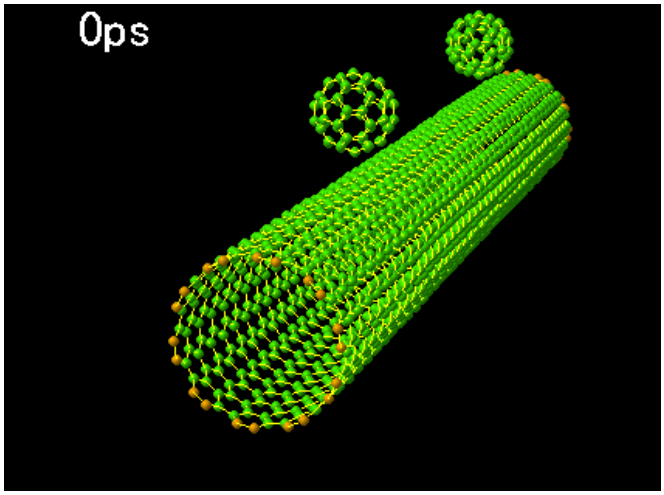

0ps

Figure 8.1 shows the band gaps

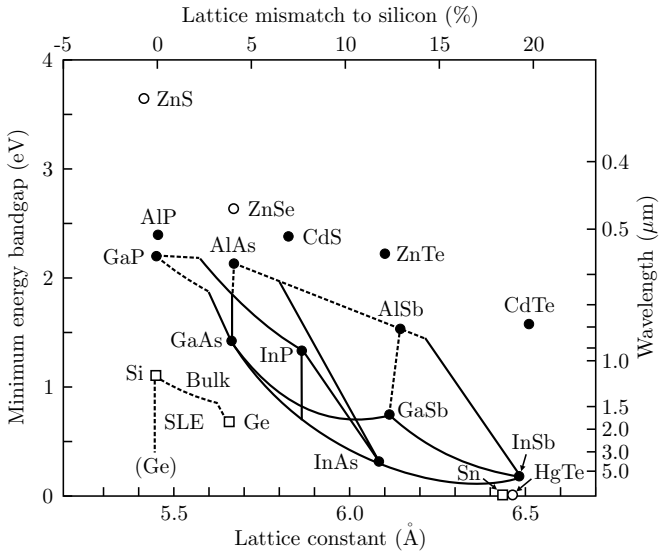


Figure 8.1: Band gap versus lattice parameter for some of the more common semiconductors. The curves indicate commonly-used alloys such as (Ga,Al)As, (Ga,In)As *etc.*; solid curves represent direct gaps and dashed curves indirect gaps.

Handout 9

Measurement of bandstructure

9.1 Introduction

This Lecture will describe a selection of the most common techniques used to measure the bandstructures of solids. Many of these techniques involve the use of magnetic fields to partially quantise the electronic motion; this is known as *Landau quantisation* or *cyclotron motion*. I shall therefore treat the motion of band electrons in some detail before giving brief details of the techniques themselves.

9.2 Lorentz force and orbits

9.2.1 General considerations

Equation 5.9 may be used to calculate the effect of an external magnetic field $\mathbf{B}$ on a band electron, *i.e.*

$$\hbar \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} = -e\mathbf{v} \times \mathbf{B}, \tag{9.1}$$

where the right-hand side is the well-known *Lorentz force*. Equation 9.1 implies that

- the component of $\mathbf{k}$ parallel to $\mathbf{B}$ is constant;
- as $\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}$ is perpendicular to $\mathbf{v}$ (defined property of the $\times$ operation) and as

$$\mathbf{v} = \frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k}) \tag{9.2}$$

(see Equation 5.5), $\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}$ is perpendicular to $\nabla_{\mathbf{k}} E(\mathbf{k})$. This means that the electron path (orbit) is one of constant energy. To see this, consider $\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}.\nabla_{\mathbf{k}} E(\mathbf{k}) = 0$ (because the two are perpendicular); writing this in components gives $\frac{\partial k_x}{\partial t}\frac{\partial E}{\partial k_x} + \frac{\partial k_y}{\partial t}\frac{\partial E}{\partial k_y} + \frac{\partial k_z}{\partial t}\frac{\partial E}{\partial k_z} \equiv \frac{\mathrm{d}E}{\mathrm{d}t}$ from the chain rule. Hence, $\frac{\mathrm{d}E}{\mathrm{d}t} = 0$.

In k-space, the possible electron orbits are therefore described by the intersections of surfaces of constant energy with planes perpendicular to $\mathbf{B}$.

9.2.2 The cyclotron frequency

Consider a section of k-space orbit of constant energy E in a plane perpendicular to $\mathbf{B}$ (see Figure 9.1). The time $t_2 - t_1$ taken to traverse the part of the orbit between $\mathbf{k}_1$ to $\mathbf{k}_2$ is

$$t_2 - t_1 = \int_{t_1}^{t_2} \mathrm{d}t = \int_{\mathbf{k}_1}^{\mathbf{k}_2} \frac{\mathrm{d}k}{|\dot{\mathbf{k}}|}, \tag{9.3}$$

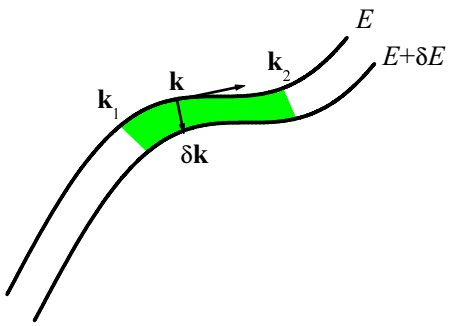


Figure 9.1: Geometrical interpretation of parameters used in the derivation of the cyclotron frequency; the two curves represent constant energy orbits of energy E and $E+\delta E$ in the plane perpendicular to the magnetic field. The electron traverses from $\mathbf{k}_1$ to $\mathbf{k}_2$ on the orbit of energy E; $\delta\mathbf{k}$ is perpendicular to the orbit of energy E and connects it to the orbit of energy $E+\delta E$. The small arrow perpendicular to $\delta\mathbf{k}$ shows the instantaneous direction of dk which runs along the orbit in the integral of Equation 9.4. The shaded area is the area of the k-space plane between orbits E and $E+\delta E$.

where $\dot{\mathbf{k}} = \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}$. Equations 9.1 and 9.2 can be used to obtain $\dot{\mathbf{k}}$, so that Equation 9.3 becomes

$$t_2 - t_1 = \frac{\hbar^2}{eB}\int_{\mathbf{k}_1}^{\mathbf{k}_2} \frac{\mathrm{d}k}{|\nabla_{\mathbf{k}\perp} E|}, \tag{9.4}$$

where $\nabla_{\mathbf{k}\perp} E$ is the component of $\nabla_{\mathbf{k}} E$ perpendicular to the field.

The quantity $\nabla_{\mathbf{k}\perp}$ might seem rather nebulous;[1] however, it has a simple geometrical interpretation. Consider a second orbit in the same plane as the one of energy E defined at the start of this Section, but with energy $E+\delta E$ (see Figure 9.1). Let the vector $\delta\mathbf{k}$ join a point $\mathbf{k}$ on the orbit of energy E to a point on the orbit of energy $E+\delta E$, and let it also be perpendicular to the orbit of energy E at $\mathbf{k}$. If δE is small, then

$$\delta E = \nabla_{\mathbf{k}} E.\delta\mathbf{k} = \nabla_{\mathbf{k}\perp} E.\delta\mathbf{k}. \tag{9.5}$$

As $\nabla_{\mathbf{k}}$ is perpendicular to surfaces of constant energy, then $\nabla_{\mathbf{k}\perp}$ is perpendicular to the orbit of energy E and therefore parallel to $\delta\mathbf{k}$. Equation 9.5 becomes

$$\delta E = |\nabla_{\mathbf{k}\perp} E|\delta k, \tag{9.6}$$

so that Equation 9.4 can be rewritten

$$t_1 - t_2 = \frac{\hbar^2}{eB}\frac{1}{\delta E}\int_{\mathbf{k}_1}^{\mathbf{k}_2} \delta k \mathrm{d}k. \tag{9.7}$$

The integral in Equation 9.7 is the area of the k-space plane between orbits E and $E+\delta E$; therefore if $\delta E \to 0$ then

$$t_2 - t_1 = \frac{\hbar^2}{eB}\frac{\partial A_{1,2}}{\partial E}, \tag{9.8}$$

where $\frac{\partial A_{1,2}}{\partial E}$ is the rate at which the orbit from $\mathbf{k}_1$ to $\mathbf{k}_2$ sweeps ut area in k-space as E increases.

In most cases it is going to be useful to work in terms of closed orbits (cyclotron orbits) in k-space, *i.e.* closed paths where $\mathbf{k}_1 = \mathbf{k}_2$. In this case A becomes the area in k-space of the closed orbit; this will

[1] Or even "nablaous".

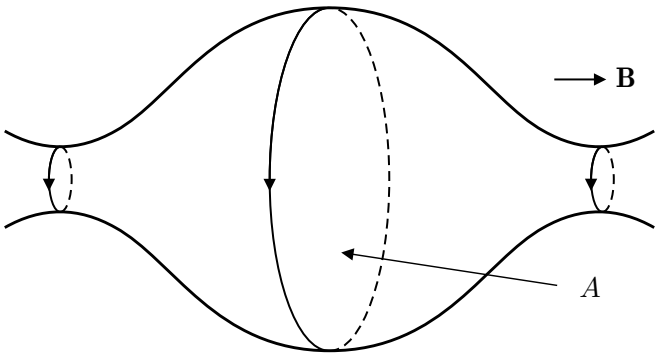


Figure 9.2: Orbits on a Fermi surface section are in planes perpendicular to **B**. Here, A represents the k-space cross-sectional area of the orbit.

depend on E and the component of **k** parallel to **B**, which we denote $k_{||}$. The period τ_c of the closed orbit is

$$\tau_c = \frac{\hbar^2}{eB}\frac{\partial A(E,k_{||})}{\partial E}; \tag{9.9}$$

turning this into an angular frequency ω_c (the *cyclotron frequency*), we obtain

$$\omega_c = \frac{eB}{m^*_{CR}}, \tag{9.10}$$

where

$$m^*_{CR} = \frac{\hbar^2}{2\pi}\frac{\partial A(E,k_{||})}{\partial E}. \tag{9.11}$$

The quantity m^*_{CR} defined in Equation 9.11 is known as the *cyclotron mass*. It is easy to show that for a free electron system this quantity is just m_e, so that Equation 9.10 yields the classical cyclotron frequency $\omega_c = eB/m_e$.[2]. Similarly, in the case of a constant, isotropic effective mass m^*, it can be shown in a straightforward manner that $\omega_c = eB/m^*$.

9.2.3 Orbits on a Fermi surface

Lectures 2–6 showed that the Fermi surface of a metal is a constant–energy surface *par excellence*. Therefore, if the material subjected to the magnetic field is metallic (*i.e.* possesses sections of Fermi surface), then Section 9.2.1 implies that the electrons will perform orbits in k-space about Fermi–surface cross–sections perpendicular to **B** (see Figure 9.2). Equations 9.10 and 9.11 can therefore give information about the cross-sectional areas of a section of Fermi surface. Note that the *sign* of $(\partial A/\partial E)$ allows the carriers on a section of Fermi surface to be identified as hole-like ($(\partial A/\partial E) < 0$) or electron-like ($(\partial A/\partial E) > 0$ in a very appealing manner (see Figure 9.3).

9.3 The introduction of quantum mechanics

9.3.1 Landau levels

A general proof of the quantised motion of a carrier in a magnetic field in an arbitrarily-shaped band is difficult. In this section, I shall use a band defined by the effective–mass tensor of problem set 2, with its minimum centred on $k = 0$ (a good approximation to many of the band extrema in semiconductors) and derive some analytical solutions for the eigenenergies of the electrons (the *Landau levels*). The anisotropic band also yields some useful insights about generalised bandstructure.

[2]See *e.g. Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) page 126.

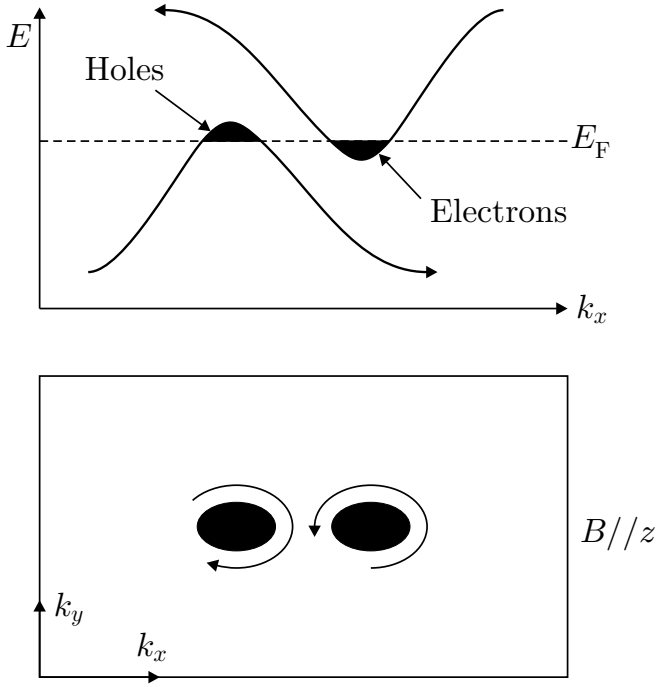


Figure 9.3: The use of the sign of $(\partial A/\partial E)$ to identify a section of Fermi surface as hole-like or electron-like. The top half of the Figure shows a schematic bandstructure with two bands crossing the Fermi energy E_F. E_F crosses the left-hand band near its maximum; the band is almost full, with holes (empty states) at the top of the band surrounded by filled states. E_F crosses the right-hand band near its minimum, so that there are electrons in the bottom of the band surrounded by empty states. The lower half of the Figure shows the corresponding Fermi-surface cross-sections. In the left-hand case, as E_F increases, the area of the section of Fermi surface decreases (hole-like). In the right-hand case, as E_F increases, the area of the section of Fermi surface increases (electron-like). The arrows show the opposite senses of cyclotron motion about the Fermi-surface sections.

For further generality, Section 9.3.2 will derive the Landau quantisation of arbitrarily-shaped bands in the limit that $\hbar\omega_c \ll E_F$.

The Hamiltonian for electrons in the band defined by the effective mass tensor of problem set 2, with its minimum centred on $k = 0$ is

$$\left\{ \frac{\mathcal{P}_x \mathbf{e}_1}{(2m_1)^{\frac{1}{2}}} + \frac{\mathcal{P}_y \mathbf{e}_2}{(2m_2)^{\frac{1}{2}}} + \frac{\mathcal{P}_z \mathbf{e}_3}{(2m_3)^{\frac{1}{2}}} \right\}^2 \psi = E\psi, \tag{9.12}$$

where the $\mathcal{P}$ are one-dimensional momentum operators; *i.e.* all of the effects of the crystalline potential have been absorbed into the bandstructure, which is defined by the effective masses. Substitution of a plane wave solution into Equation 9.12 yields energies of the correct form, *i.e.*

$$E = \frac{\hbar^2 k_x^2}{2m_1} + \frac{\hbar^2 k_y^2}{2m_2} + \frac{\hbar^2 k_z^2}{2m_3}. \tag{9.13}$$

Consider a magnetic field $\mathbf{B}$ directed along the z ($\mathbf{e}_3$) axis. In addition, remember that $\mathbf{B} = \nabla \times \mathbf{A}$, where $\mathbf{A}$ is the magnetic vector potential. One particular $\mathbf{A}$ which gives the correct $\mathbf{B}$ is

$$\mathbf{A} = (0, Bx, 0); \tag{9.14}$$

this choice is known as the *Landau gauge*. For an electron in a magnetic field,

$$\mathcal{P} \rightarrow (\mathcal{P} + e\mathbf{A}). \tag{9.15}$$

Making such a substitution into Equation 9.12 gives

$$\left\{ \frac{\mathcal{P}_x \mathbf{e}_1}{(2m_1)^{\frac{1}{2}}} + \frac{(\mathcal{P}_y + eBx)\mathbf{e}_2}{(2m_2)^{\frac{1}{2}}} + \frac{\mathcal{P}_z \mathbf{e}_3}{(2m_3)^{\frac{1}{2}}} \right\}^2 \psi = E\psi. \tag{9.16}$$

All that we have done is introduce an extra term with x in it; as $[\mathcal{P}_y, x] = [\mathcal{P}_z, x] = 0$.[3] $\mathcal{P}_y$ and $\mathcal{P}_z$ still commute with the Hamiltonian; their associated physical quantities are thus constants of the motion. The operators $\mathcal{P}_y$ and $\mathcal{P}_z$ in Equation 9.16 can therefore be replaced by their constant values $\hbar k_y$ and $\hbar k_z$ respectively. Equation 9.16 becomes

$$\left\{ \frac{\mathcal{P}_x \mathbf{e}_1}{(2m_1)^{\frac{1}{2}}} + \frac{(\hbar k_y + eBx)\mathbf{e}_2}{(2m_2)^{\frac{1}{2}}} + \frac{\hbar k_z \mathbf{e}_3}{(2m_3)^{\frac{1}{2}}} \right\}^2 \psi = E\psi. \tag{9.17}$$

Squaring the bracket, remembering that the Cartesian unit vectors $\mathbf{e}_j$ are mutually perpendicular (*i.e.* $\mathbf{e}_j.\mathbf{e}_l = \delta_{j,l}$), gives

$$\left\{ \frac{\mathcal{P}_x^2}{2m_1} + \frac{(\hbar k_y + eBx)^2}{2m_2} + \frac{\hbar^2 k_z^2}{2m_3} \right\} \psi = E\psi. \tag{9.18}$$

Making the substitutions $E' = E - (\hbar^2 k_z^2/2m_3)$ and $x_0 = -(\hbar k_y/eB)$, and rearranging yields

$$\{ \frac{\mathcal{P}_x^2}{2m_1} + \frac{e^2 B^2}{2m_2}(x - x_0)^2 \} \psi = E'\psi. \tag{9.19}$$

Equation 9.19 looks exactly like the Hamiltonian of a one-dimensional harmonic oscillator, *i.e.*,

$$\{ \frac{\mathcal{P}_x^2}{2m} + \frac{1}{2} m\omega^2 (x - x_0)^2 \} \psi = E'\psi, \tag{9.20}$$

with $\omega = eB/(m_1 m_2)^{\frac{1}{2}}$; thus $E' = (l + \frac{1}{2})\hbar\omega$, with $l = 0, 1, 2, 3.....$ The energy levels of the electron are therefore

$$E(l, B, k_z) = \frac{\hbar^2 k_z^2}{2m_3} + (l + \frac{1}{2})\hbar\omega, \tag{9.21}$$

[3] These are standard commutation relationships; see *e.g.* *Quantum Mechanics*, by Stephen Gasiorowicz (Wiley, New York 1974) pages 51, 141.

with

$$\omega = eB/(m_1 m_2)^{\frac{1}{2}} \equiv \omega_c. \tag{9.22}$$

The equivalence of ω and ω_c, the cyclotron frequency for this geometry (*i.e.* $\omega_c = eB/m^*_{\rm CR}$, with $m^*_{\rm CR} = \frac{\hbar^2}{2\pi}\frac{\partial A(E,k_{||})}{\partial E}$), is left to the reader.

The following points may be deduced from Equation 9.21.

- The energy of the electron's motion in the plane perpendicular to **B** is completely quantised. These quantised levels are known as *Landau levels*.
- The k-space areas of the orbits in the plane perpendicular to **B** are also quantised (this is easy to work out in the present case and is left as an exercise); thus, allowed orbits fall on "Landau tubes" in k-space with quantised cross-sectional area.
- The energy "quantum" for the in-plane motion appears to be $\hbar\times$(the semiclassical cyclotron frequency).
- The motion parallel to **B** is unaffected (*c.f.* Section 9.2.1).

We shall consider the second point in more detail (and more generally) in the following section.

9.3.2 Application of Bohr's correspondence principle to arbitrarily-shaped Fermi surfaces in a magnetic field

In Section 9.4 below we are going to consider the effect of the quantisation of the k-space orbits caused by **B** on the Fermi surfaces of metals; *i.e.* we shall be dealing with Landau levels which cut the Fermi surface. The Fermi surfaces of real metals are in general not as simple as the case dealt with in the previous section. Therefore, in order to treat an arbitrily-shaped band, resulting in an arbitrarily-shaped Fermi surface, we shall use Bohr's correspondence principle, which states in this context that the difference in energy of two adjacent levels is $\hbar$ times the angular frequency of classical motion at the energy of the levels.

The correspondence principle is only valid for levels with very large quantum numbers. However, this is not really a problem, as we are going to be dealing with Landau levels with energies comparable to $E_{\rm F}$; in most metals $E_{\rm F}$ is ∼several eV, whereas the cyclotron energy of free electrons is $\hbar eB/m_{\rm e} \approx 1.16\times10^{-4}$ eV at $B = 1$ T. Laboratory fields are usually ~ 10 T, and so the quantum numbers of Landau levels with energies $\sim E_{\rm F}$ will be $l \sim 10^4$; the correspondence principle should work reasonably well.

The classical frequency at the energy of interest (*i.e.* $E_{\rm F}$) is given by Equation 9.10

$$\omega_c = \frac{eB}{m^*_{\rm CR}}, \tag{9.23}$$

with $m^*_{\rm CR}$ given by Equation 9.11 as follows

$$m^*_{\rm CR} = \frac{\hbar^2}{2\pi}\frac{\partial A(E,k_{||})}{\partial E}. \tag{9.24}$$

Here, A is now specifically *the cross-sectional area of the Fermi surface in a plane perpendicular to* **B**. Thus, the separation in energy of two Landau tubes with energies close to $E_{\rm F}$ is

$$E(l+1,B,k_{||}) - E(l,B,k_{||}) = \frac{\hbar eB}{m^*_{\rm CR}}, \tag{9.25}$$

with $m^*_{\rm CR}$ defined by Equation 9.24. (It is interesting to compare this Equation with Equation 9.21.)

9.3.3 Quantisation of the orbit area

Equation 9.25 may be rewritten as

$$(E(l+1,B,k_{||})-E(l,B,k_{||}))\frac{\partial A(E)}{\partial E}=\frac{2\pi eB}{\hbar}. \tag{9.26}$$

Now as the E in $A(E)$ is of order E_{F}, then the difference between adjacent levels will be $\ll$ the energies of the levels themselves. Thus the approximation

$$\frac{\partial A(E)}{\partial E}=\frac{A(E(l+1,B,k_{||}))-A(E(l,B,k_{||}))}{E(l+1,B,k_{||})-E(l,B,k_{||})} \tag{9.27}$$

may be made, so that Equation 9.26 becomes

$$A(E(l+1,B,k_{||}))-A(E(l,B,k_{||}))=\frac{2\pi eB}{\hbar}. \tag{9.28}$$

This states that classical orbits at adjacent allowed energies and the same $k_{||}$ enclose k-space areas that differ by a fixed amount δA, with

$$\delta A=\frac{2\pi eB}{\hbar}. \tag{9.29}$$

Putting this another way, the k-space area enclosed by an orbit of allowed energy and $k_{||}$ must be given by

$$A(E(l+1,B,k_{||}))=(l+\lambda)\delta A=(l+\lambda)\frac{2\pi eB}{\hbar} \tag{9.30}$$

for large l, where λ is a constant ~ 1.

As in Section 9.3.1, we have found "Landau tubes" which define the allowed k-space orbits; rather than being quasicontinuous, the Fermi surface will now be composed of occupied states lying on Landau tubes (see Figure 9.4. Note that the k-space area of each tube increases with B; as we shall see in the sections below, this means that tubes "pop out" of the Fermi surface as B increases.

9.3.4 The electronic density of states in a magnetic field

The above sections have shown that the energy of the band electron is completely quantised into Landau levels in the plane perpendicular to **B**, whilst the motion parallel to **B** remains unconstrained. This means that the density of states of the electrons is an infinite "ladder" of Landau levels, each with a one-dimensional density of states function, due to the motion parallel to **B**, superimposed.[4] The overall density of states will therefore be of the form $g(E)=\sum_{l=0}^{\infty}C(E-E(l,B))^{-\frac{1}{2}}$, where $E(l,B)$ is the energy of the Landau level (*i.e.* the lowest energy point on the Landau tube) and C is a constant; it is shown schematically in Figure 9.5.

9.4 Quantum oscillatory phenomena

Each time one of the sharp peaks in the electronic density of states moves through the chemical potential μ,[5] there will be a modulation of the density of states at μ. Since almost all of a metal's properties depend on the density of states at μ, we expect this to affect the behaviour of the metal in some way.

Figure 9.6 shows a Landau tube superimposed on constant energy surfaces E and $E+\delta E$. The left-hand part of the figure shows the general case; as the field increases, the tube sweeps outwards, so that some of its states are always crossing the constant energy surfaces. However, in the right-hand part of the figure, the tube reaches an *extremal cross-section* of the constant energy surface E. In this case,

[4] In other words, as far as the density of states is concerned, the magnetic field reduces the effective dimensionality of the electron system by 2.

[5] Remember from earlier lectures that $E_{\mathrm{F}}\equiv\mu(T=0)$. In many books, quantum oscillations are said to result from "Landau levels passing through the Fermi energy". This is only strictly true at $T=0$; the chemical potential is the important energy as far as the electrons are concerned. In metals at low temperatures there will be a very small difference between E_{F} and μ; nevertheless, one should not tolerate sloppy terminology.

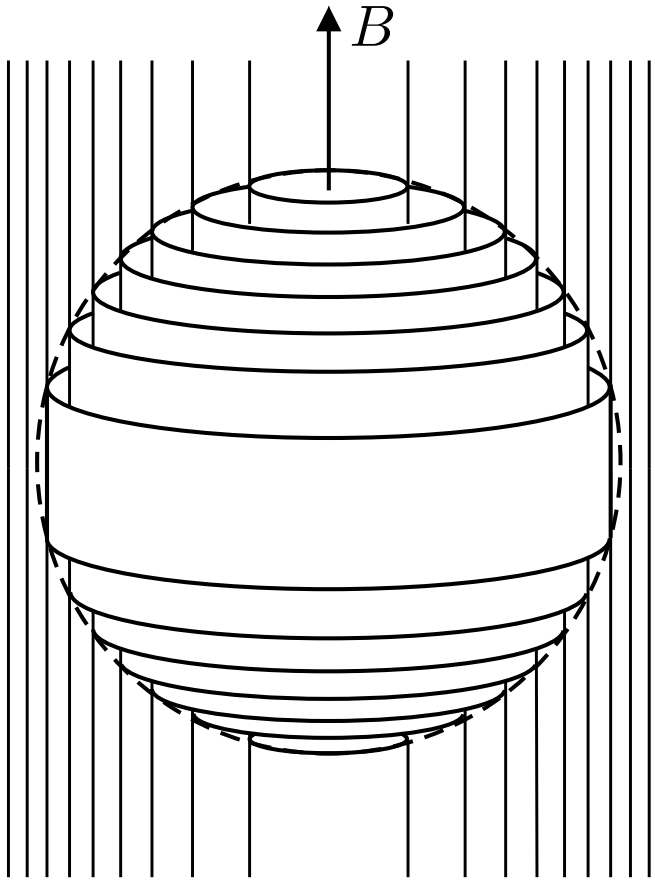


Figure 9.4: Schematic of a Fermi sphere rearranged into Landau tubes.

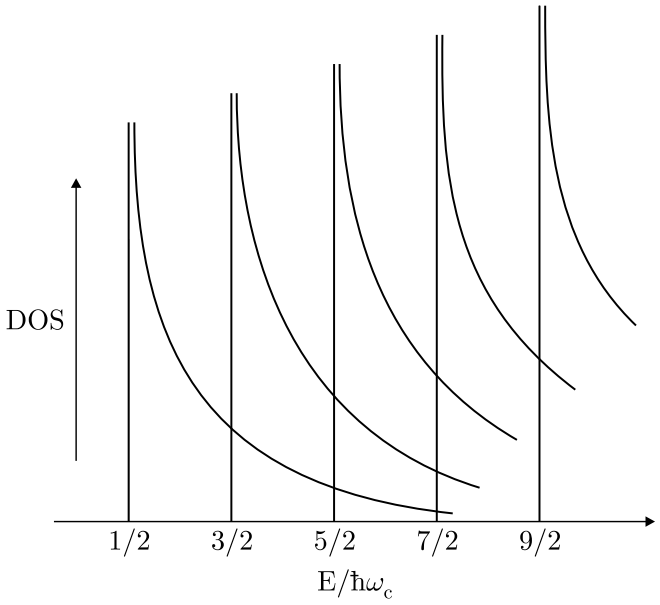


Figure 9.5: Schematic of the electronic density of states in a magnetic field. The "sawtooth" structure results from a one-dimensional density of states (from the motion parallel to **B**) superimposed on the δ-function-like Landau levels.

a much larger area of the cylinder (and therefore a much greater number of states, which lie uniformly in k-space) lies between E and $E+\delta E$. Returning to the case of the Fermi surface, we can therefore say that the maximum effect of the Landau quantisation will occur when a Landau tube (and its associated peak in the density of states) crosses an extremal cross-section of a Fermi surface. By analogy with the discussion of classical orbits about the Fermi surface in Section 9.2.1, such cross-sections are often known as *extremal orbits*.

The Landau tube crosses an extremal orbit/cross-section when

$$(l+\lambda)\frac{2\pi eB}{\hbar} = A_{\text{ext}}, \tag{9.31}$$

where A_{ext} is the k-space area of the extremal cross-section of the Fermi surface in a plane perpendicular to $\mathbf{B}$. Thus, the metal's properties will oscillate as $\mathbf{B}$ changes, with a period given by

$$\Delta\left(\frac{1}{B}\right) = \frac{2\pi e}{\hbar}\frac{1}{A_{\text{ext}}}. \tag{9.32}$$

Such oscillations are known generically as *magnetic quantum oscillations*.

Often it is more convenient to describe a series of quantum oscillations using a frequency known as the *fundamental field*

$$B_{\text{F}} = \frac{1}{\Delta(1/B)} = \frac{\hbar}{2\pi e}A_{\text{ext}}. \tag{9.33}$$

Note that

- the fundamental field of the quantum oscillations is determined solely by the Fermi surface extremal area and fundamental constants;
- several different frequencies (fundamental fields) of quantum oscillation may be simultaneously present for a given orientation of $\mathbf{B}$, corresponding to different possible extremal orbits (see Figure 9.7); a famous example is the simultaneous observation of frequencies from "neck and belly" orbits in Cu and Au;
- a measurement of the observed frequencies as a function of magnetic field orientation allows the Fermi-surface shape to be mapped out;
- open orbits do not give rise to quantum oscillations.[6]

9.4.1 Types of quantum oscillation

As the electronic density of states at E_{F} determines most of a metal's properties, virtually all properties will exhibit quantum oscillations in a magnetic field. Examples include[7]

- oscillations of the magnetisation (the de Haas–van Alphen effect);
- oscillations of the magnetoresistance (the Shubnikov–de Haas effect);
- oscillations of the sample length;
- ocillations of the sample temperature;
- oscillations in the ultrasonic attenuation;
- oscillations in the Peltier effect and thermoelectric voltage;
- oscillations in the thermal conductivity.

[6] However, open orbits do lead to a very interesting quantum phenomenon which has recently been observed in high-frequency experiments; see A. Ardavan *et al., Phys. Rev. B* **60**, 15500 (1999); *Phys. Rev. Lett.* **81**, 713 (1998).

[7] Some pictures of typical data are shown in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) pages 266-268.

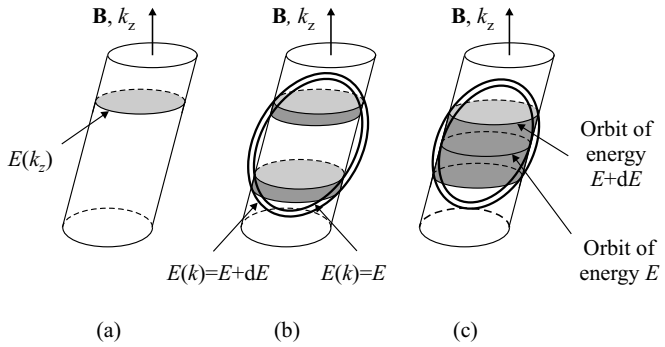


Figure 9.6: Schematic of Landau tubes crossing surfaces of constant energy E and $E + \delta E$. (a) shows the Landau tube on its own; (b) shows the Landau tube intersecting constant energy surfaces E and $E+\delta E$ when the tube does not correspond to an extremal orbit; (c) shows the Landau tube intersecting constant energy surfaces E and $E+\delta E$ when the tube corresponds to an extremal orbit. In (c), a much greater area of the cylinder (and therefore a larger number of states) lies between E and $E + \delta E$ than is the case in (b).

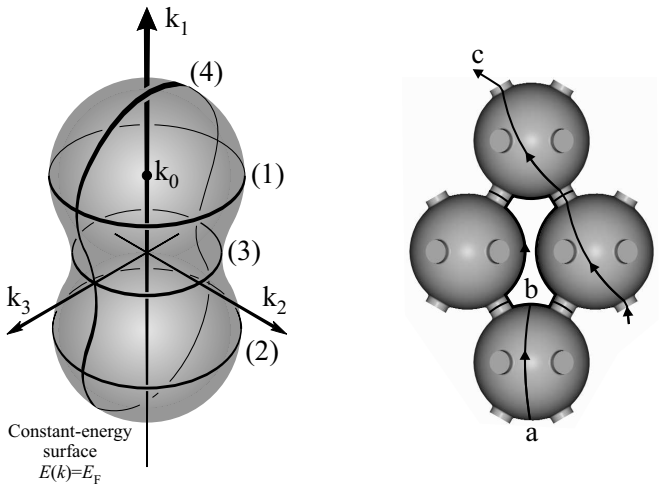


Figure 9.7: Illustration of the types of extremal orbit about Fermi surfaces which can give rise to quantum oscillations. The left-hand figure shows that for **B** parallel to k_1 there will be three extremal orbits ((1), (2) and (3)); as (1) and (2) have the same area, there will be two series of quantum oscillation (*i.e.* two different frequencies will be present). However, for **B** parallel to k_2, there is only one extremal orbit (4); a similar situation would apply if **B** were parallel to k_3. The right-hand figure helps to visualise the different types of orbit possible in metals possessing necks and bellies in their Fermi surfaces, such as Copper, Silver or Gold. The orbit labelled a is an orbit around the "belly" of the Fermi surface. If the magnetic field is applied out of the page, then an extremal "hole-like" orbit (labelled b) will occur around the bone-shaped space in the middle; this is known as the *dog's bone* orbit! In orbit c, electrons run across the bellies and necks in an extended or *open* orbit; *this does not give rise to quantum oscillations*, as there is no closed orbit with which a frequency can be associated. If the magnetic field is applied along one of the neck directions, extremal orbits about both neck and belly are possible, leading to two frequencies of quantum oscillations. Note that different field orientations are required to observe the different cases; the orbits in k-space are *always* in planes perpendicular to **B**.

9.4.2 The de Haas–van Alphen effect

The de Haas-van Alphen effect is perhaps the most significant of the quantum oscillatory phenomena in metals; as the magnetisation M is a thermodynamic function of state;[8] it can be directly related to the density of states and Fermi–Dirac distribution function of the electrons in a metal, *without additional assumptions*. This means that theoretical models for the Fermi surface can be checked in a very rigorous manner.[9]

Figure 9.8 shows a schematic of the types of coil most commonly used to observe the de Haas–van Alphen effect. Each system consists of a coil c which contains the sample and a *compensating coil* c^0; both are placed in a time-varying magnetic field. The coils c and c^0 are connected in series-opposition and carefully balanced so that in the absence of a sample the voltage induced by $\mathrm{d}B/\mathrm{d}t$ in c^0 exactly cancels that in c. When the sample is placed in c, then c and c^0 are no longer balanced; a voltage V is induced, where

$$V = \alpha \frac{\mathrm{d}M}{\mathrm{d}B}\frac{\mathrm{d}B}{\mathrm{d}t}. \tag{9.34}$$

Here α is a parameter depending on the geometry of the coil and sample. The de Haas–van Alphen oscillations occur in $\mathrm{d}M/\mathrm{d}B$ and thus will be visible in V.[10]

The field is provided in two ways.

- For fields of up to $\sim$ 20 T, superconducting magnets are generally used. The magnet consists of two parts, a large outer coil providing a static field and an inner coil which provides a small modulation field (*i.e.* $\mathrm{d}B/\mathrm{d}t$) of mT amplitude at a few tens of Hz.
- Pulsed magnets are used for fields of up to $\sim$ 60 T; in this case the magnetic field rises to its maximum value in a few ms, so that $\mathrm{d}B/\mathrm{d}t$ is large (there is no need for an additional modulation coil).

Some typical de Haas-van Alphen data are shown in Figures 9.9 and 9.10.

The coils are often mounted on a rotation stage so that the sample can be rotated *in situ*, allowing measurements to be made with **B** applied at several orientations to the crystal axes. The coils and sample are mostly mounted in ^{3}He cryostats or dilution refrigerators, as mK temperatures are needed to observe the oscillations in many cases (see below).

9.4.3 Other parameters which can be deduced from quantum oscillations

So far we have concentrated on the information about the Fermi-surface shape provided by the quantum oscillations. However, other valuable information can be deduced. Figure 9.11 shows Shubnikov–de Haas oscillations in an organic solid (see Section 8 for more details of this sort of conductor). Two things are immediately obvious.

1. **The oscillations grow in amplitude as the temperature is lowered.** This occurs because finite temperatures "smear out" the edge of the Fermi-Dirac distribution function which separates filled and empty states (see Figure 1.7). At $T = 0$ the Landau tubes will pop out of a very sharp Fermi surface; at finite T the surface will be "fuzzy". The modulation of the density of states around μ caused by the Landau tubes will therefore be less significant. The thermal smearing implies that $\hbar\omega_\mathrm{c}$ must be greater than $\sim k_\mathrm{B}T$ for oscillations to be observed, *i.e.* low temperatures are needed.

   As the decline in intensity is caused by the Fermi-Dirac distribution function, it can be modelled rather well, and an analytical expression for the temperature dependence of the oscillation amplitude, valid as long as the oscillation amplitude is small, has been derived[11]

$$\text{Amplitude} \propto \frac{\chi}{\sinh\chi}, \tag{9.35}$$

[8] See *e.g. Equilibrium Thermodynamics* by C.J Adkins, Third Edition (Cambridge University Press, Cambridge 1983) page 6 and Appendix.

[9] For enthusiasts, a relatively recent example of this is given in *A numerical model of quantum oscillations in quasi-two-dimensional organic metals in high magnetic fields*, by N. Harrison *et al.*, *Physical Review B*, **54**, 9977 (1996).

[10] de Haas-van Alphen oscillations are also sometimes observed using a *torque magnetometer*; see J.S. Brooks *et al.*, *Rev. Sci. Instrum.* **64**, 3248 (1987).

[11] See *Magnetic oscillations in metals*, by David Shoenberg (Cambridge University Press, Cambridge 1984) Chapter 2.

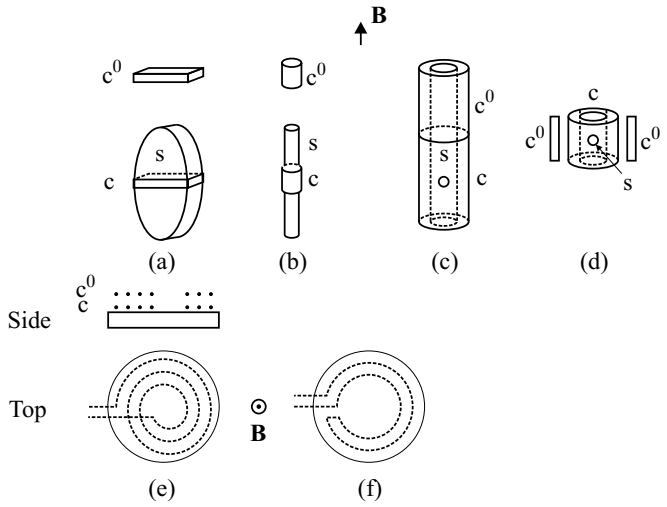


Figure 9.8: Coils for the observation of the de Haas–van Alphen effect. For each set, c is the coil flux-linked to the sample and c^0 is the compensating coil, only weakly flux-linked to the sample: (a) c is wound around the centre of a disc-like sample; (b) c is wound around a sample cut into a cylinder; (c) a small sample inside colinear coils c and c^0; (d) small sample inside coaxial coils c and c^0; (e) side and top views of a coil evaporated down onto a thin sample; (f) as (e) but with only a single annular turn. In both (e) and (f) c^0 has not been shown. (Based on figures in *Magnetic oscillations in metals*, by David Shoenberg (Cambridge University Press, Cambridge 1984).)

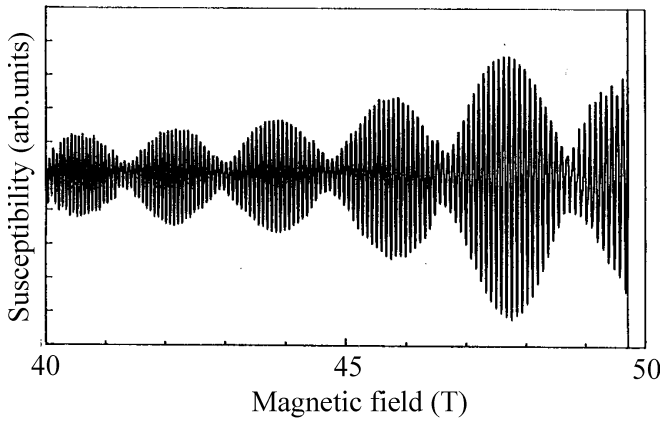


Figure 9.9: de Haas-van Alphen oscillations in Pt at 4.2 K with **B** parallel to [111]. The data have been recorded using a pulsed magnetic field. Note the presence of two frequencies of oscillations due to two extremal orbits about the Fermi surface. The y axis, labelled "susceptibility", represents the voltage V induced in the coil divided by $(\mathrm{d}B/\mathrm{d}t)$, leaving a quantity proportional to the differential susceptibility $(\mathrm{d}M/\mathrm{d}B)$ (see Equation 9.34). (Data from S. Askenazy, Physica B **201**, 26 (1994).)

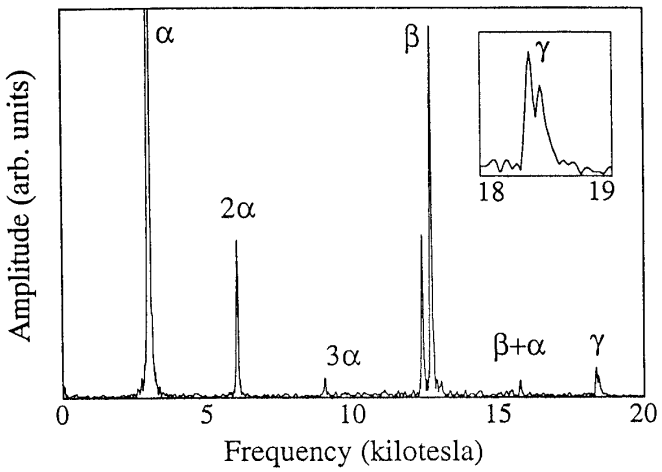


Figure 9.10: Fourier transform of de Haas-van Alphen oscillations of Sr_2RuO_4. Fourier transformation is routinely used to help understand de Haas-van Alphen oscillations. The oscillation data (recorded digitally) are processed so that they are equally spaced in $1/B$; the resulting file is Fourier transformed using a desk-top computer. Each extremal cross-section of the Fermi-surface generates a peak in the Fourier transform, the frequency of which is proportional to the k-space area. Oscillations which are non-sinusoidal in shape generate harmonics in the Fourier transform; frequency mixing effects also occur, resulting in sum frequencies. Three Fermi-surface cross-sections (labelled α, β and γ) are shown, along with some harmonics of α and a sum frequency $\beta+\alpha$. (Data from A.P. McKenzie *et al.*, *J. Phys. Soc. Jpn.* **67**, 385 (1998)).

where $\chi = 14.7 m^*_{\rm CR} T/B$ (with T in Kelvin and B in Tesla) . It is fairly obvious that the parameter χ is proportional to the ratio $k_{\rm B}T/\hbar\omega_{\rm c}$; *i.e.* what we are doing is comparing $k_{\rm B}T$ and $\hbar\omega_{\rm c}$. Fits of the oscillation amplitude to Equation 9.35 can therefore be used to give $m^*_{\rm CR}$.

2. **The oscillations grow in amplitude as B increases.** This occurs because scattering causes the Landau levels to have a finite energy width $\sim \hbar/\tau$. As B increases, $\hbar\omega_{\rm c}$ will increase, so that the broadened Landau levels will become better resolved; theoretical studies have shown that this causes the oscillation amplitude to vary with B as follows (again, the equation is only valid for small oscillations)

$$\text{Amplitude} \;\propto\; e^{-\frac{\pi}{\omega_{\rm c}\tau}}. \tag{9.36}$$

Therefore the scattering rate τ^{-1} can be extracted.

In addition, information about the g-factor of the electrons can be obtained from the oscillations under certain circumstances.[12]

9.4.4 Magnetic breakdown

The frequencies observed in the quantum oscillations correspond to the areas of closed semiclassical orbits about the extremal areas of the Fermi surface. However, under certain circumstances, much higher frequency oscillations can become apparent at high fields, corresponding to larger k-space areas. These are caused by *magnetic breakdown*.

Figure 9.12 shows a simple Fermi surface cross-section with two open semiclassical orbits (incapable of giving quantum oscillations) and a closed orbit (capable of giving quantum oscillations); the arrows show the electron trajectories. At low fields, one frequency of oscillations corresponding to the closed orbit will be seen. At high fields, the electrons have sufficient cyclotron energy to tunnel *in k-space* from one part of the Fermi surface to another. Therefore they can now describe much larger closed k-space

[12]See *Magnetic oscillations in metals*, by David Shoenberg (Cambridge University Press, Cambridge 1984) and J. Singleton, *Reports on Progress in Physics*, **63**, 1111 (2000).

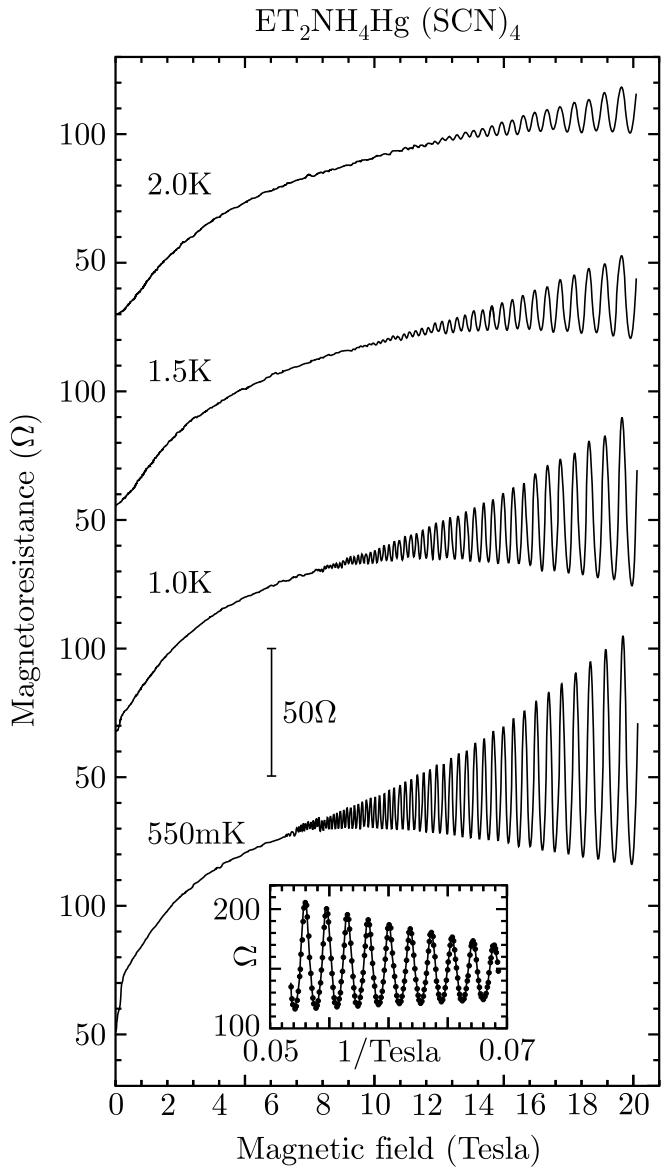


Figure 9.11: Shubnikov–de Haas oscillations in the organic molecular solid α-(BEDT-TTF)$_2$NH$_4$Hg(SCN)$_4$; data for different temperatures have been offset for clarity. The inset shows that the oscillations are periodic in $1/B$. (Data from J. Singleton, *Reports on Progress in Physics*, **63**, 1111 (2000).)

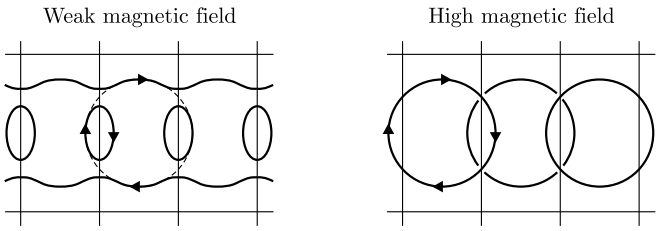


Figure 9.12: Explanation of magnetic breakdown. Left: semiclassical orbits on a simple Fermi surface at low fields; right: magnetic breakdown at high fields.

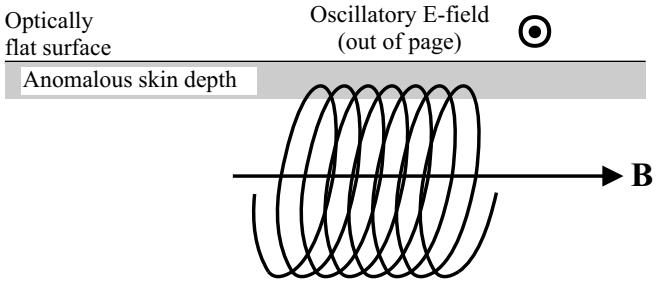


Figure 9.13: Geometry of a cyclotron resonance experiment in a metal. The shaded region indicates the skin depth for the radiation; the helical orbits of the electrons are shown schematically.

orbits, leading to higher-frequency quantum oscillations. In effect, the magnetic field is beginning to break down the arrangement of the bandstructure at zero field.

We shall see an example of magnetic breakdown later.

9.5 Cyclotron resonance

It is possible to make a direct measurement of ω_c using millimetre-waves or far-infrared radiation to excite transitions between Landau levels. Such an experiment is known as *cyclotron resonance*. The conditions for observing cyclotron resonance in metals and semiconductors are rather different, and the two cases are described below.

9.5.1 Cyclotron resonance in metals

The field components of electromagnetic radiation decay with distance z into a conducting material as $\exp(-z/\delta)$, where $\delta = (\frac{1}{2}\sigma\omega\mu_r\mu_0)^{-\frac{1}{2}}$ is the *skin depth* or anomalous skin depth; here σ is the conductivity of the material, $\mu_r\mu_0$ is its permeability and ω is the angular frequency of the radiation. Metals have rather high conductivities and hence small skin depths; typical values for Copper are $\delta \approx 7\times10^{-5}$m at $\omega/2\pi \approx 1$ MHz and $\delta \approx 7\times10^{-7}$m at $\omega/2\pi \approx 10$ GHz. [13] Typical effective masses in metals combined with the magnetic fields readily available in laboratories have meant that frequencies $\omega/2\pi \sim 1-100$ GHz have tended to be applied in cyclotron resonance experiments (see *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) Chapter 8). At such frequencies, radiation cannot penetrate far into the crystal, dictating the geometry of the cyclotron resonance measurement (see Figure 9.13). The magnetic field is applied parallel to a surface of the crystal, which is placed in a region of oscillating electric field in a resonant cavity; **E** of the radiation is arranged to be perpendicular to **B** and parallel to the surface. If the frequency of the radiation $\omega = j\omega_c$, where j is an integer, then the electrons will receive a "kick" from the radiation's electric field every

[13] The skin depth is derived in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) page 236.

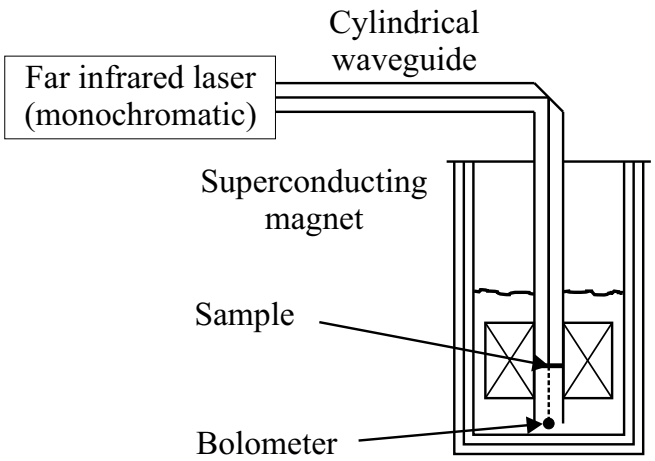


Figure 9.14: Schematic of a cyclotron resonance experiment in a semiconductor.

time they come within a skin depth of the surface; this results in absorption of energy. Usually ω is kept constant and the field is swept, so that absorptions are uniformly spaced in $1/B$. As in the case of the quantum oscillations, low temperatures are required. A magnetic field which is very accurately parallel to the sample's surface is required for a successful measurement. This type of experiment is often referred to as *the Azbel'–Kaner geometry.*

9.5.2 Cyclotron resonance in semiconductors

The number densities of carriers in semiconductor samples are much lower than those in metals, so that the radiation can completely penetrate even large samples. A simple transmission arrangement is usually adopted (see Figure 9.14). The cyclotron resonance is usually recorded by measuring the sample transmission; experiments are carried out either by fixing the magnetic field and varying the energy of the radiation (see Figure 9.15), or by using a fixed-frequency source (such as a far-infrared laser, which can give a number of monchromatic laser lines in the wavelength range 40 μm to 1000 μm) and sweeping the magnetic field (see Figures 9.14 and 9.16).[14] The magnetic field is usually provided by a superconducting magnet ($B = 0- \sim 20$ T). As the whole of the cyclotron orbit experiences the electric field of the radiation (*c.f.* the situation in metals; see above), the quantum-mechanical selection rule for the Landau-level quantum number ($\Delta l = \pm 1$) holds. Thus the resonance condition is

$$\omega = \omega_{\rm c} = \frac{eB}{m^*_{\rm CR}}, \tag{9.37}$$

where ω is the angular frequency of the radiation.

The linewidth of the resonance can give information about the scattering rate τ^{-1}. The scattering induces a frequency uncertainty $\Delta\omega_{\rm c} \sim \tau^{-1}$. If the experiment is a fixed-frequency, swept-field one, this translates to an uncertainty (*i.e.* resonance width) in magnetic field of $\Delta B = \Delta\omega_{\rm c} m^*_{\rm CR}/e \sim m^*_{\rm CR}/e\tau$.[15]

In the case of degenerate semiconductor systems (*e.g.* heterojunctions), free carriers are present even at low temperatures; typical cyclotron resonance data are shown in Figure 9.15. However, in lightly doped samples, the carriers must be excited into the bands by either raising the temperature to cause the impurities to ionise (but not so far as to broaden the Landau levels) or by additionally illuminating the sample with above-band-gap radiation.

[14] An alternative arrangement for semiconductor cyclotron resonance (rather an old-fashioned one) is given in *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) pages 729-734; the description also provides a more rigorous derivation of the linewidth of the cyclotron resonance than the one given here.

[15] Provided that $m^*_{\rm CR} \approx m^*$, the carrier effective mass for linear motion (and this is often true in direct-gap semiconductors like GaAs) it is apparent that $1/\Delta B \sim$ the carrier mobility.

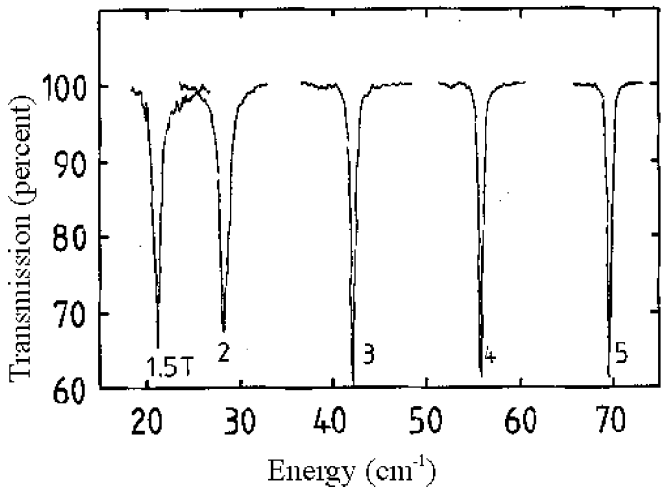


Figure 9.15: Cyclotron resonances in a GaAs-(Ga,Al)As heterojunction at $T = 4.2$ K; the magnetic field has been applied perpendicular to the two-dimensional electron layer, which has an areal carrier density of $\sim 9 \times 10^{10}\text{cm}^{-2}$. The data have been recorded by fixing the magnetic field at 1.5, 2, 3, 4, and 5 T; at each field, a Fourier-transform spectrometer has then been used to record the transmission of the sample as a function of energy. (Data recorded by G. Wiggins, University of Oxford. For further data of this type, see C.J.G.M. Langerak *et al., Phys. Rev. B.* **38** 13133 (1988).)

Material	$m_1 = m_2$	m_3	m_{lCR}	m^*_{hCR}
Ge	0.082	1.64	0.044	0.3
Si	0.19	0.98	0.16	0.3

Table 9.1: Effective masses (in units m_e) in Ge and Si determined using cyclotron resonance. The masses m_1, m_2 and m_3 refer to the components of the effective mass tensors for the conduction band; m_{lCR} and m^*_{hCR} are the cyclotron masses of light and heavy holes respectively.

Figure 9.16 shows an example of the latter technique applied to a single crystal of Ge. In addition to separate resonances caused by light and heavy holes, three electron cyclotron resonances are observed. This occurs becaue the anisotropic conduction-band minima in Ge lie along the [111] axes (see Figure 6.6); in Figure 9.16 the static magnetic field makes three different angles with the four [111] axes. The conduction-band minima in Ge can be described using an effective mass tensor; if one of the principal axes of the co-ordinate system is directed along a [111] axis, then the tensor is diagonal. Under such circumstances (see second problem set), the cyclotron resonance occurs at $\omega = \frac{eB}{m^*}$, with

$$m^* = \left(\frac{B^2 m_1 m_2 m_3}{m_1 B_x^2 + m_2 B_y^2 + m_3 B_z^2} \right)^{\frac{1}{2}},$$

where the components of the magentic field (B_1, B_2, B_3) are defined with respect to the principal axes which make the tensor diagonal. Thus, the cyclotron resonance frequency depends on the orientation of the static magnetic field with respect to the crystal axes. Such angle-dependent cyclotron resonance experiments can be used to determine components of the effective mass tensor.

In the case of Ge and Si, it is found that two of the principal values of the effective mass tensor are equal, so that the constant energy surfaces are ellipsoids with circular cross-sections (see Figure 6.6). The effective masses of Si and Ge determined by cyclotron resonance are shown in Table 9.1.

9.6 Interband magneto-optics in semiconductors

The application of a magnetic field to a semiconductor splits both the valence and conduction bands into Landau levels, and optical transitions can be induced between hole and electron Landau levels.

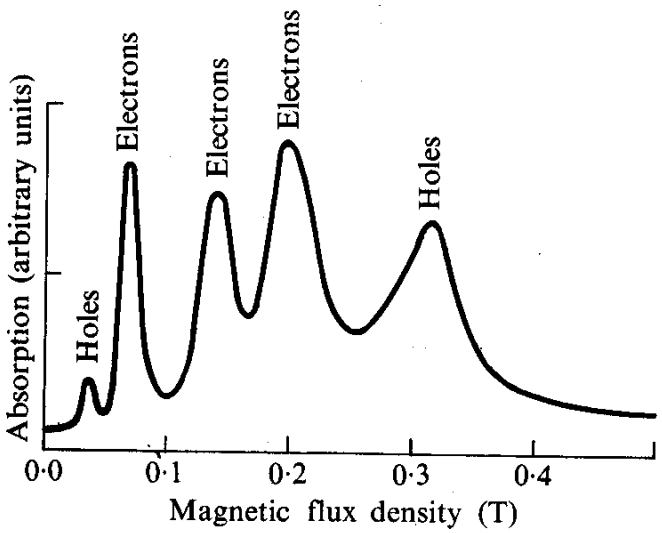


Figure 9.16: Absorption caused by cyclotron resonance of electrons and holes in a single crystal of Ge; the electrons and holes are present because the sample has been illuminated by above-band-gap light. The microwave radiation frequency has been fixed at 24 GHz and the magnetic field, applied in a (110) plane at 60° to a [100] axis, has been swept. Cyclotron resonances due to light (low field) and heavy (high field) holes are visible, as are three electron cyclotron resonances. (Data from Dresselhaus *et al.*, *Phys. Rev.* **98**, 368 (1955).)

In reasonably simple systems, the selection rule for the Landau level index is $\Delta l = 0$; more subtle selection rules apply in materials with complex bandstructures.[16] Taking the former, simpler case, the transitions therefore have the energies

$$E(l,B) = E_g + \frac{\hbar eB}{m^*_{\rm hCR}}(l+\frac{1}{2}) + \frac{\hbar eB}{m^*_{\rm cCR}}(l+\frac{1}{2}), \tag{9.38}$$

where $l = 0, 1, 2, 3..$, and $m^*_{\rm cCR}$ and $m^*_{\rm hCR}$ are the cyclotron masses of the electron and the hole (usually the heavy holes dominate, because of their large density of states) respectively. Such a case is dealt with in the problem sets.

Illustrative experimental data are shown in Figures 9.17 and 9.18. In Figure 9.17, the transmission of a thin film of PbTe has been measured using fixed-frequency radiation of energy $h\nu = 0.2077$ eV as the magnetic field is swept. Transitions between valence-band Landau levels and conduction-band Landau levels are observed as dips in the transmission; note that the dips get further apart and better resolved as the field increases. Equation 9.38 gives a qualitative understanding of these observations; as the field increases, the inter-Landau-level transitions will successively pass through the radiation energy $h\nu$, resulting in a spectrum which looks roughly periodic in $1/B$. PbTe has complex, anisotropic conduction and valence-band extrema, so that the conduction and valence-band Landau level energies vary as the orientation of the magnetic field changes; this is illustrated by the rapid change in the spectra as the sample is tilted in the field, showing how a comprehensive study of this kind can map out the band shapes.

In systems which involve excitonic effects, there are usually complications to be taken into account. Figure 9.18(a) shows the transmission of a GaAs-(Ga,Al)As quantum well at four fixed fields as a function of energy. At zero field, only the excitonic transitions from the hole subbands to the conduction-band subbands are observed, but as the field increases, transitions between the heavy-hole Landau

[16] This is a consequence of the angular momentum carried by the phonon. The valence bands of many semiconductors are based on atomic orbitals with total angular momentum quantum number $J = 3/2$, whereas the conduction band is based on orbitals with $J = 1/2$. Thus the one unit of angular momentum carried by the photon is required to promote an electron across the band gap, leaving nothing to spare to change the Landau-level quantum number l. By contrast, in intraband magneto-optical transitions (cyclotron resonance), the photon's angular momentum is used to change l by one ($\Delta l = \pm 1$).

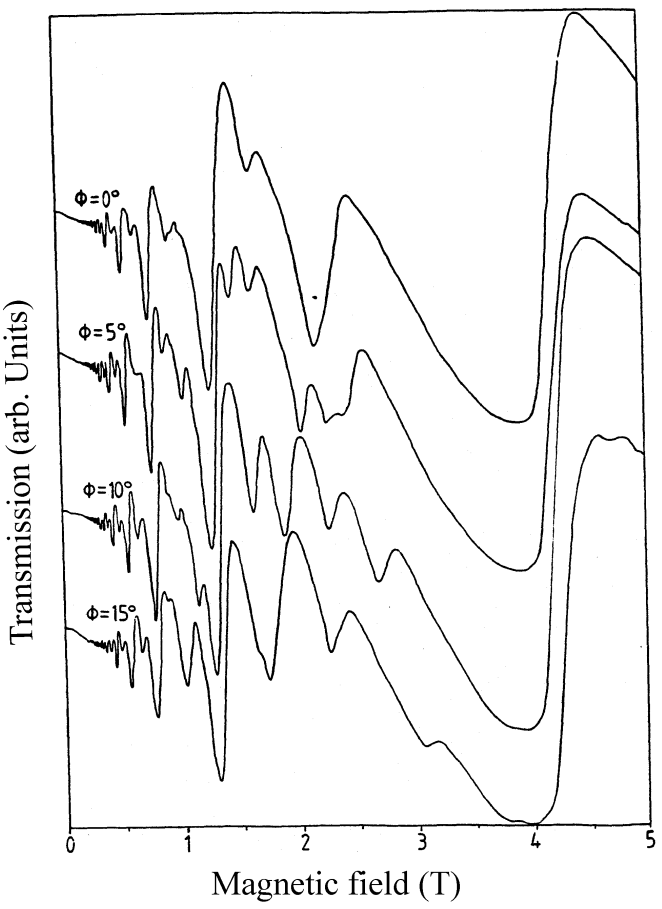


Figure 9.17: Transmission of PbTe at a fixed photon energy of $h\nu = 0.2077$ eV as a function of magnetic field ($T = 1.9$ K). Data for several orientations of the sample in the external field, which is at all times perpendcular to [111], are shown; ϕ is the angle between the magnetic field and $[\bar{1}10]$. (Data from J. Singleton *et al., J. Phys. C: Solid State Phys.* **19**, 77 (1986).)

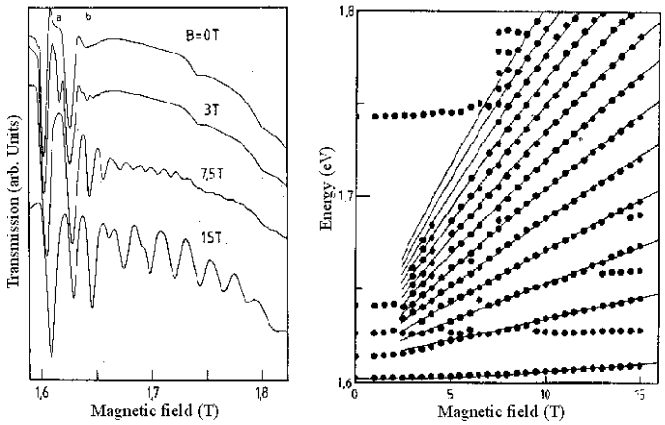


Figure 9.18: (a) Transmission of a 55 ÅGaAs-(Ga,Al)As quantum well at various fixed magnetic fields ($T = 55$ K). The magnetic field has been applied perpendicular to the plane of the quantum well. (b) Resulting fan diagram of transition energies versus magnetic field; points are data, and the lines are a theoretical fit. (From D.C. Rogers *et al., Phys. Rev. B.* **34**, 4002-4009 (1986).)

levels and the conduction-band Landau levels become visible. Figure 9.18 shows a "fan diagram" of the transition energies versus magnetic field for the same sample. Note the following:

- the Landau-level transitions no longer extrapolate back to a point (*c.f.* Equation 9.38), but instead evolve directly from the various states of the exciton (this is particularly notable for the lowest ($l = 0$ and $l = 1$) transitions); there is a one-to-one correspondence between the hydrogen-like levels of the exciton and the Landau level transitions (1s evolves into $l = 0$ transition, 2s evolves into $l = 1$ transition *etc.*;
- the field dependences of the transition energies deviate from straight lines at high energies, curving downwards; this is due to the *non-parabolicity* of the bands as one moves away from the band extrema.

Data such as these can be used to extract the effective masses of the electron and heavy-hole hole subbands, parameters which describe the nonparabolicity and the exciton binding energy.

9.7 Reading

A more detailed treatment of the basic ideas involved in magnetic quantum oscillations is given in *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapter 14. Some case studies are given in *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001) Chapter 8. Simpler descriptions are in *Electrons in Metals and Semiconductors*, by R.G. Chambers (Chapman and Hall, London 1990) Chapter 7, and *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapters 8 and 9.

For enthusiasts, a splendid and detailed review of magnetic quantum oscillations in metals is found in *Magnetic oscillations in metals*, by David Shoenberg (Cambridge University Press, Cambridge 1984).

Handout 10

Transport of heat and electricity in metals and semiconductors

10.1 Thermal and electrical conductivity of metals

10.1.1 The "Kinetic theory" of electron transport

We now apply our knowledge of bandstructure, Fermi surfaces and electron statistics to the problem of transport of heat and electricity by electrons. We are going to look initially at a metal with isotropic bands characterised by an effective mass m^*. This is in effect the Sommerfeld model with the additional refinement of an adjustable m^* which characterises the effect of the periodic lattice potential on the electrons; the Fermi surface will be spherical. (More complex Fermi surfaces will be mentioned later.) In Lecture 1 we used the Relaxation Time Approximation to derive the expression

$$\sigma = \frac{ne^2\tau}{m^*} \tag{10.1}$$

for the electrical conductivity σ of such a metal. Here n is the number density of electrons and τ^{-1} is the scattering rate. In addition, we used Kinetic Theory and the Sommerfeld model to obtain the expression

$$\kappa = \frac{1}{6}n\pi^2 k_{\mathrm{B}}^2 \frac{T}{E_{\mathrm{F}}} v_{\mathrm{F}}^2 \tau \tag{10.2}$$

for the electronic thermal conductivity κ. Here v_{F} is the Fermi velocity and E_{F} is the Fermi energy. Equations 10.1 and 10.2 were combined to give the Wiedemann–Franz Ratio or Lorenz number

$$L \equiv \frac{\kappa}{\sigma T} = \frac{\pi^2}{3}\frac{k_{\mathrm{B}}^2}{e^2} \equiv L_0 \tag{10.3}$$

and it was found that this constant was in good agreement with many experimental data at room temperature. However, in many cases, experimental values of L differed markedly (by orders of magnitude!) from L_0 at lower temperatures.

The problem has arisen because two scattering times τ (which are not necessarily the same) have been cancelled in Equation 10.3. Acknowledging this fact we rewrite Equations 10.1 and 10.2 in the following form:

$$\sigma = \frac{ne^2\tau_\sigma}{m^*} \tag{10.4}$$

and

$$\kappa = \frac{1}{6}n\pi^2 k_{\mathrm{B}}^2 \frac{T}{E_{\mathrm{F}}} v_{\mathrm{F}}^2 \tau_\kappa, \tag{10.5}$$

i.e. we have distinguished the two scattering times.

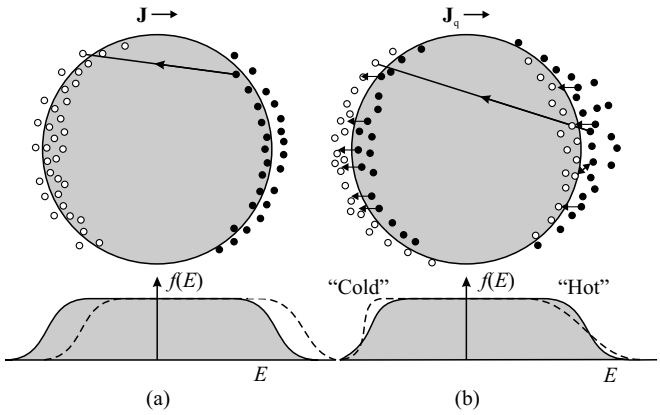


Figure 10.1: Top: representation of the alteration of the Fermi surface caused by electrical conduction (a) (current density $\mathbf{J}$) and thermal conduction (b) (heat flux $\mathbf{J}_q$); black circles represent filled states and white circles empty states. In electrical conduction, the Fermi surface is displaced by a tiny amount corresponding to the drift velocity caused by the electric field; thus, an excess of electrons travel towards the right. In thermal conduction, "hotter" electrons head right and "cooler" electrons head left. The effects of these changes are represented in the lower half of the Figures by plotting the Fermi-Dirac distribution functions for left-heading and right-heading electrons (dotted lines) compared to the equilibrium situation (grey shading). The long arrows indicate phonon scattering events of the kind encountered at room temperature, which lead to degradation of both the electrical and thermal currents. The short arrows are the phonon scattering processes encountered at low temperatures; note that these can degrade the thermal current both by "warming up" "cold" electrons and "cooling" "hot" electrons.

10.1.2 What do τ_σ and τ_κ represent?

In all of the following, we are going to be talking about the electron velocity $\mathbf{v}$. Previously, we have seen that $\mathbf{v} = (1/\hbar)\nabla_{\mathbf{k}} E(\mathbf{k})$; this means that electron velocities are always perpendicular to surfaces of constant energy. As always, the electrons at the Fermi surface are going to see most of the action; their velocities may be imagined as vectors normal to the Fermi surface, as the Fermi surface is *the* constant energy surface *par excellence*.[1]

Consider first electrical conduction. In the absence of an electric field, there are as many electrons travelling in any one direction as in the opposite direction, so that no net current flows. When an electric field is applied, electrons at the Fermi surface acquire a small amount of extra velocity (the drift velocity derived in the Relaxation-Time Approximation). The effect is shown in Figure 10.1a; electrons on the right-hand side of the Fermi surface (with initial velocities parallel to the direction of the drift velocity) move up into slightly higher energy (velocity) states. Other electrons can then fill the states vacated. The net effect is to displace the Fermi surface by a very small amount $\sim m^* v_\mathrm{d}/\hbar$, where v_d is the drift velocity (typically $v_\mathrm{d} \sim 10^{-3}\mathrm{ms}^{-1}$, whereas $v_\mathrm{F} \sim 0.01c$).

In order to visualise the equivalent modification of the Fermi surface in the case of thermal conductivity, it is useful to recall the Kinetic Theory derivation of κ. Thermal conductivity occurs in gases because the gas molecules travelling in one direction have a higher characteristic thermal energy (or effective temperature) than those travelling in the opposite direction. The equivalent situation for electrons is represented in Figure 10.1b; the electrons travelling leftwards are "cooler" and therefore have a less smeared out Fermi-Dirac distribution function, whereas those going rightwards (in the direction of heat flow) are "hotter", possessing a more broadened Fermi-Dirac distribution function characteristic of a higher temperature.

We can now identify what τ_κ and τ_σ represent.

- τ_σ represents the characteristic time to randomise/get rid of an electron's excess of forward velocity (or deficiency of backward velocity). A single scattering event of the type shown by an arrow in

[1] In the case of the simple bandstructure defined at the beginning of this discussion, $\mathbf{v}$ and $\mathbf{k}$ *will* be parallel; however, it is worth remembering that this will not usually be the case for bands in general.

Figure 10.1a would be sufficient for this purpose; an electron heading rightwards (in the direction of the current flow) has been scattered into an empty state on the opposite side of the Fermi surface with almost opposite momentum and velocity.

- τ_κ represents the characteristic time to randomise/get rid of an electron's excess/deficiency of thermal energy. This can be achieved in a variety of ways shown in Figure 10.1b; *e.g.* an electron can lose $\sim k_\mathrm{B}T$ of energy, and drop into an empty state close by in k-space (vertical process) or it can be shot from the "hot" side of the Fermi surface to the "cold" side by a scattering event involving large momentum change (horizontal process).

10.1.3 Matthiessen's rule

In the following discussion, we are going to assume that the electronic scattering rates are additive, *i.e.*

$$\frac{1}{\tau} = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3} + ..., \tag{10.6}$$

where the τ_j^{-1} are scattering rates due to different processes (*e.g.* collisions with or emission/absorption of phonons, scattering from impurities *etc.*). Equation 10.6 implies that the scattering process with the shortest τ_j will dominate, allowing us to predict regions of temperature in which we can ignore all forms of scattering but one; *e.g.* it is reasonable to assume that scattering of electrons at high temperatures will be almost entirely due to phonons, because there are a lot of phonons around.

Equation 10.6 is known as *Matthiessen's rule*; it must be admitted that it is only a crude guide as to what to expect in the presence of more than one scattering mechanism. It fails most spectacularly when

- the outcome of one scattering process influences the outcome of another;
- one or more τ_j is a function of $\mathbf{k}$.

In the latter case, the conductivities κ and σ will involve total τs due to all processes averaged over all $\mathbf{k}$, whereas Equation 10.6 implies the summation of reciprocals of each τ_j individually averaged over all $\mathbf{k}$; these two operations are very unlikely to lead to the same result.

Having stated this *caveat*, we shall first think about highish temperatures, where electron–phonon scattering events predominantly determine τ_κ and τ_σ.

10.1.4 Emission and absorption of phonons

When discussing the idea of bands, we have seen that electrons are only scattered when something disturbs the periodicity of the crystal. We are first going to consider the role of phonons, which may be pictured as propagating local distortions of the crystal. Such distortions may scatter an electron; two processes must be considered.

1. **Elastic processes** Both the phonon and electron change wavevector and energy, constrained by conservation of energy and momentum.
2. **Inelastic processes** The phonon may be emitted or absorbed by an electron, causing the electron's wavevector and energy to change.

Our derivations of bandstructure also made a hidden assumption, that the positions of the ions were not affected by the presence of mobile electrons. This is of course somewhat unrealistic; electrons and ions are highly charged, and the passage of an electron will result in the distortion of the lattice around it. The electron can be scattered by this; in wave-mechanical terms it has emitted a phonon, causing its energy and momentum to change. This is another inelastic process.

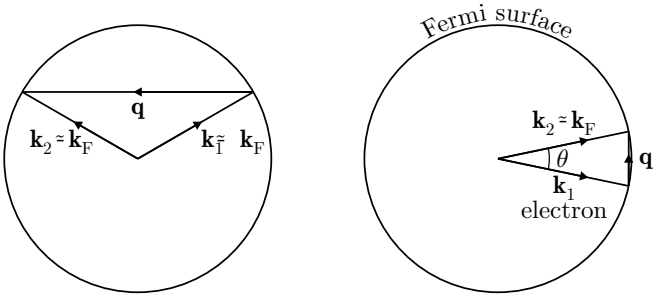


Figure 10.2: Geometry of electron-phonon scattering events at $T \sim \theta_D$ (left) and $T \ll \theta_D$ (right). In each case $\mathbf{k}_1$ and $\mathbf{k}_2$ are the initial and final electron wavevectors (both of magnitude $\approx k_F$) and $\mathbf{q}$ is the phonon wavevector.

10.1.5 What is the characteristic energy of the phonons involved?

Phonons behave as massless bosons, that is to say they can be created and destroyed in a similar way to photons.[2] Phonons therefore have a "black-body" type of energy distribution, with a peak at an energy $\sim k_B T$, *i.e.* the characteristic phonon energy at a temperature T is $\hbar\omega \sim k_B T$. We therefore expect that when an electron scatters from or absorbs a phonon, that phonon will typically have an energy $\sim k_B T$.

But what about the emission of phonons? The form of the Fermi–Dirac distribution function for the electrons means that the only empty states below the Fermi energy have energies $E_F - (\sim k_B T)$. Similarly, the energy of the most energetic electrons will be roughly $E_F + (\sim k_B T)$. Therefore, an electron can only emit a phonon with energy $\sim k_B T$, *i.e.* even though the electron distorts the lattice at low temperatures, it cannot emit energetic phonons because there are no accessible final states for such a process!

The probability of emitting a phonon of energy $\sim k_B T$ will have a very similar temperature dependence to the probability of absorbing such a phonon. This is easy to see when one considers that the latter depends on the number of phonons around with such an energy (*i.e.* on the phonon density of states at an energy $\hbar\omega \sim k_B T$) whereas the former depends on the density of available final phonon states (*i.e.* the phonon density of states at an energy $\hbar\omega \sim k_B T$).

Therefore, in all that follows, I am just going to talk about elastic phonon scattering processes and inelastic phonon scattering processes where the latter means *either* emission *or* absorption of phonons; *i.e.* I shall assume that the probability of all processes involving phonons will follow a similar temperature dependence.

10.1.6 Electron–phonon scattering at room temperature

The Debye temperatures θ_D of most metals are less than or of the order of room temperature. Now θ_D is roughly the energy of the most energetic phonons in the metal, so that phonons with energy $\hbar\omega \sim k_B T$ will have wavevectors $q \approx$ (half the width of the Brillouin zone) $\sim k_F$, where k_F is the Fermi wavevector. Thus one phonon scattering event (inelastic or elastic) will scatter the electron to the opposite side of the Fermi surface (see Figure 10.2). Thus $\tau_\sigma^{-1} \approx \tau_\kappa^{-1} \propto$ (number of phonons with $\hbar\omega \sim k_B T$) $\propto T$.

10.1.7 Electron-phonon scattering at $T \ll \theta_D$

In this case, phonons with energy $\hbar\omega \sim k_B T$ will have energies $\ll k_B \theta_D$ and therefore $q \ll$ (Brillouin zone size), *i.e.* $q \ll k_F$. Thus, one *inelastic* phonon scattering event will be able to change the electron's energy by $\sim k_B T$; hence $\tau_\kappa^{-1} \propto$ (number of phonons with $\hbar\omega \sim k_B T$) $\propto T^3$.[3] However, one phonon scattering event (elastic or inelastic) will be unable to knock the electron to the other side of the Fermi

[2] See Statistical Mechanics books such as *e.g.* *Statistical Physics*, by Tony Guenault (Routledge, London 1988) page 124 .

[3] The T^3 power is well known from the Debye heat capacity derivation.

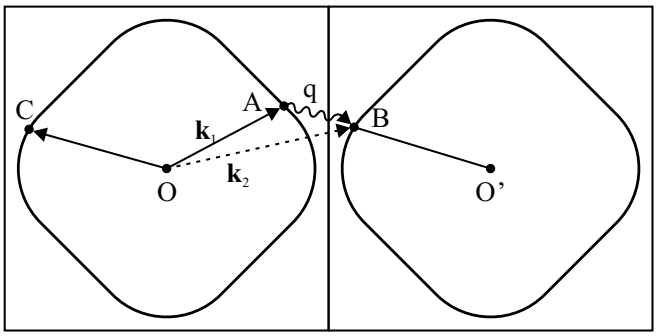


Figure 10.3: Geometry of electron-phonon umklapp scattering. A phonon with $|\mathbf{q}| < k_{\rm F}$ is able to scatter an electron of wavevector $\mathbf{k}$ with a positive velocity from the first Brillouin zone (A) to a state with a negative velocity (B) in the second Brillouin zone. The state at B is equivalent to that at C in the first Brillouin zone.

surface (see Figure 10.2) and so $\tau_\sigma \gg \tau_\kappa$. This is the reason for the failure of the Wiedemann–Franz ratio at low temperatures.

In order to take account of the fact that many, many scattering events are required before the excess forward velocity of the electron is thoroughly randomised, the scattering rate contains a weighting factor $(1-\cos\theta)$, where θ is the scattering angle (see Figure 10.2). Now as θ is small,

$$1-\cos\theta \approx 1-(1-\frac{\theta^2}{2}) = \frac{\theta^2}{2} \approx \frac{q^2}{2k_{\rm F}^2} \approx \frac{\omega^2}{2k_{\rm F}^2 v_\phi^2} \quad \propto \quad T^2, \tag{10.7}$$

where we have used the fact that the dispersion relationship for low energy phonons is close to $\omega = v_\phi q$, with v_ϕ the speed of sound. Therefore we have $\tau_\sigma \propto T^{-5}$ and $\tau_\kappa \propto T^{-3}$.

10.1.8 Departures from the low temperature $\sigma \propto T^{-5}$ dependence.

The $\sigma \propto T^{-5}$ temperature dependence is rarely if ever obeyed exactly. There are two main causes involving electron-phonon processes which contribute to this.

- **The periodicity of k-space** often allows phonons with small q to scatter electrons at the Fermi surface into empty states with energy $\sim E_{\rm F}$ in an adjacent Brillouin zone; these states may have a velocity which is almost opposite to that of the initial state. This is shown schematically in Figure 10.3; the process is known as *electron umklapp scattering.*

- **Complicated Fermi surfaces** may have lobes, lozenges, ellipsoids *etc.* (plus their replicas, due to k-space periodicity, from other zones) all over the Brillouin zone. This means that phonons with short q can cause scattering of electrons between Fermi surface sections with very different characteristic velocities.

Both of these effects give a scattering rate which is roughly exponential, $\tau^{-1} \propto e^{-\theta_{\rm F}/T}$, where $\theta_{\rm F}$ is a characteristic temperature depending on the Fermi-surface geometry.

10.1.9 Very low temperatures and/or very dirty metals

In the case of very low temperatures, the phonon scattering becomes negligible and scattering of electrons by impurities and defects becomes dominant. Impurities have a different ionic core from the host metal, and therefore will often appear to be charged with respect to the background. The scattering of electrons by impurities is therefore like Rutherford scattering, with electrons being deflected through large angles. One "event" therefore degrades the transport of heat and electricity equivalently, so that $\tau_\kappa = \tau_\sigma =$ constant and the Wiedemann–Franz ratio again holds.

Temperature (scatterer)	Scattering times	κ σ	W-F ratio
Very low (impurities)	$\tau_\kappa \approx \tau_\sigma$ $\sim$ const	$\kappa \propto T$, $\sigma \sim$ const	L_0
$T \sim \theta_D/10$ (phonons)	$\tau_\kappa \propto T^{-3}$, $\tau_\sigma \propto T^{-5} \rightarrow e^{-\theta_F/T}$	$\kappa \propto T^{-2}$, $\sigma \propto T^{-5} \rightarrow e^{-\theta_F/T}$	$< L_0$
$T >\sim \theta_D$ (phonons)	$\tau_\kappa \approx \tau_\sigma$ $\propto T^{-1}$	$\kappa =$const, $\sigma \propto T^{-1}$	L_0

Table 10.1: Summary of the temperature dependences of scattering times and electrical and thermal conductivities.

10.1.10 Summary

A summary of the temperature dependences of scattering times and electrical and thermal conductivities is given in Table 1. Typical electrical resistivity and thermal conductivity data are shown in Figures 10.4 and 10.5. Note that for reasonably pure metals with reasonably simple bandstructures (*e.g.* alkali and noble metals), electrical resistivity data are all of a similar form, and when normalised to the value at $T = \theta_R$ lie roughly on the same curve when plotted against T/θ_R (see Figure 10.4); here θ_R is a characteristic temperature similar (but not identical) to θ_D.

10.1.11 Electron–electron scattering

In metals with simple Fermi surfaces, electron–electron scattering is relatively unimportant. Initial and final states for both electrons must have energies close to E_F and wavevectors close in magnitude to k_F; in addition, energy and momentum must be conserved. The combination of these two requirements makes electron-electron scattering quite unlikely (and actually completely forbidden at $T = 0$).

However, electron-electron scattering becomes more important when

- the Fermi surface is complicated, so that there the conservation of energy and momentum becomes easy for a wider variety of possible scattering processes and/or
- the density of states at the Fermi energy is very large (because the effective mass is large), bumping up the number of initial and final states (*e.g.* transition metals, heavy fermion compounds).

Reasonably simple arguments[4] show that electron-electron scattering leads to $\tau^{-1} \propto T^2$.

Some typical data for transition metals are shown in Figure 10.6. At low temperatures, the resistivities tend towards the T^2 dependence expected for electron-electron scattering.

10.2 Electrical conductivity of semiconductors

10.2.1 Temperature dependence of the carrier densities

Over wide ranges of temperature, the dominant contribution to the temperature dependence of the electrical conductivity is the rapidly varying number of free carriers. Recalling the results of Section 6.3.4, in order to find n and p when impurities are present we use the law of Mass Action (Equation 6.12)

$$np = T^3 W e^{-\frac{E_g}{k_B T}}$$

combined with the conservation law (Equation 6.17)

$$n - p = N_D - N_A,$$

where N_D is the density of donors and N_A is the density of acceptors (both are assumed to only provide one carrier each).

[4]See *e.g. Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) page 347.

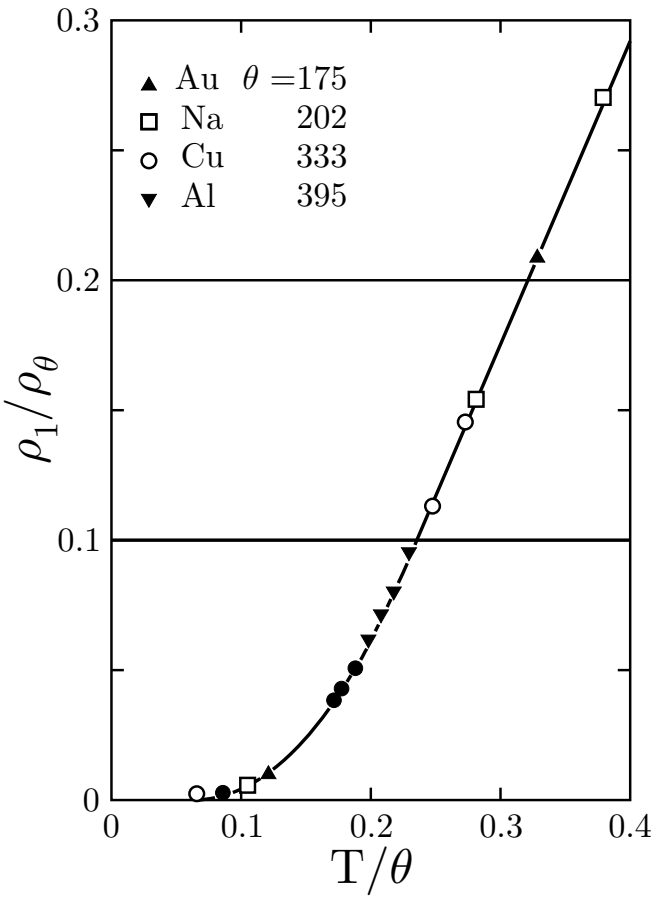


Figure 10.4: Normalised electrical resistivity data for several metals with reasonably simple Fermi surfaces plotted as a function of the normalised temperature $T/\theta_{\rm R}$. $\theta_{\rm R}$ is shown in Kelvin for each metal at the top of the Figure.

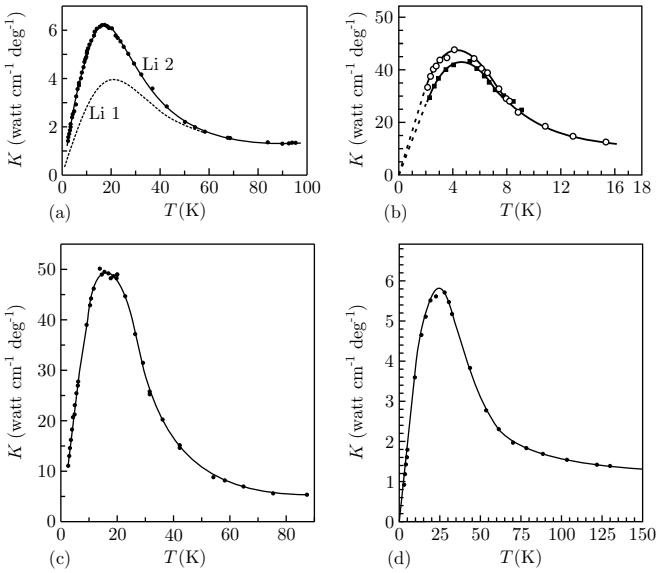


Figure 10.5: Thermal conductivity data for (a) Li (sample Li 1 is of lower purity than sample Li 2); (b) Na (again showing the effects of dirtier (lower curve) and cleaner (upper curve) samples); (C) Cu; (d) Cr.

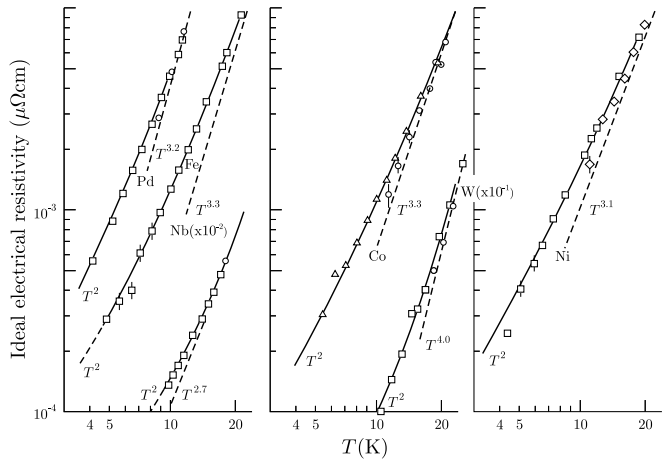


Figure 10.6: Resistivities of transition metals, showing the approach to T^2 at low temperatures.

The electrical conductivity of a semiconductor consists of a sum of contributions from all carrier types. The simple model of Section 6.3.4, which is a reasonable approximation for many semiconductors, has just two types, electrons and heavy holes. Therefore, the conductivity will contain just two contributions, *i.e.*

$$\sigma = ne\mu_{\rm c} + pe\mu_{\rm hh}, \tag{10.8}$$

where $\mu_{\rm c}$ and $\mu_{\rm hh}$ are the electron and hole mobilities respectively. The mobilities are defined as the drift velocity of the carrier per unit electric field; the Relaxation-Time Approximation (see the derivation of Equation 1.6) can be used to give

$$\mu_{\rm c} = \frac{e\tau_{\rm c}}{m^*_{\rm c}} \tag{10.9}$$

and

$$\mu_{\rm hh} = \frac{e\tau_{\rm hh}}{m^*_{\rm hh}}, \tag{10.10}$$

where $\tau_{\rm c}^{-1}$ and $\tau_{\rm hh}^{-1}$ are the scattering rates for the electrons in the conduction band and the heavy holes in the valence band respectively.

The temperature dependence of the electrical conductivity is therefore determined by convolutions of the temperature dependences of n and $\tau_{\rm c}$ and p and $\tau_{\rm hh}$.

When $n \geq p$ (n-type or intrinsic semiconductors) the holes can be completely ignored in most cases; this is because in the majority of semiconductors, the electrons have a much smaller effective mass than the holes, resulting in a smaller density of states and hence a longer τ. Therefore the electron mobility is often much, much greater than that of the holes.

10.2.2 The temperature dependence of the mobility

There are two important sources of scattering.

1. **Impurities** ($T \ll \theta_{\rm D}$). Charged impurity scattering is rather like Rutherford scattering; hence the scattering cross-section varies as E^{-2}. In the nondegenerate case, $E \sim k_{\rm B}T$, so that the cross-section is proportional to T^{-2}; hence the mean-free path is proportional to T^2. The carrier speed is proportional to $E^{\frac{1}{2}}$, *i.e.* $T^{\frac{1}{2}}$. Therefore

$$\tau \propto \frac{T^{-\frac{1}{2}}}{T^{-2}} = T^{\frac{3}{2}}. \tag{10.11}$$

   Note that this contrasts greatly with the situation in metals ($\tau \sim$ constant for impurity scattering), where all of the action goes on at or close to the Fermi surface; the carriers which scatter from the impurities in metals (see Section 10.1.9) have virtually constant (*i.e.* temperature-independent) energies. In non-degenerate semiconductors, the free carriers have a quasi-Boltzmann-like energy distribution, so that the average energy of the carriers varies with temperature.

2. **Phonons** ($T \sim \theta_{\rm D}$). The number of phonons at such temperatures will be proportional to T (see Section 10.1.6), leading to a mean-free-path proportional to T^{-1}. As before, the speed is proportional to $T^{\frac{1}{2}}$. Therefore

$$\tau \propto \frac{T^{-1}}{T^{\frac{1}{2}}} = T^{-\frac{3}{2}}. \tag{10.12}$$

10.3 Reading

There are a large number of similar treatments of this topic *e.g.* (in increasing order of complexity) *Electricity and Magnetism*, by B.I. Bleaney and B. Bleaney, revised third/fourth editions (Oxford University Press, Oxford) Chapters 11, 17; *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Sections 9.1-9.14; *Introduction to Solid State Physics*, by Charles Kittel, seventh edition (Wiley, New York 1996) Chapters 6 and 7; *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapters 1-3, 29. See also sections 4.23-4.27 of *Low temperature solid state physics*, by H.M. Rosenberg (OUP 1963).

Handout 11

Magnetoresistance in three-dimensional systems

11.1 Introduction

Magnetoresistance is a general term for the changes in the components of the resistivity and conductivity tensors of materials caused by the application of magnetic field. We are going to treat the magnetoresistance of metals in a quite general and simple manner. First, however, the Hall effect in a system with more than one type of carrier will be described, as it helps to illuminate the more general discussion of metals that will follow, and gives a clue as to the origins of magnetoresistance.

11.2 Hall effect with more than one type of carrier

11.2.1 General considerations

We consider the Hall effect with two or more carrier types present (*e.g.* electrons and holes). The geometry of a Hall effect measurement is shown in Figure 11.1; the magnetic field $\mathbf{B}$ is applied parallel to the z direction (*i.e.* $\mathbf{B} = (0, 0, B)$), whilst the current I is driven through the sample in the x direction. The electric field $\mathbf{E}$ is assumed to be $\mathbf{E} = (E_x, E_y, 0)$; we assume that any effect is going to occur in the plane perpendicular to $\mathbf{B}$ because of the nature of the Lorentz force (see Equation 11.1 below). Voltage measuring contacts are provided on the sample so that E_x and E_y can be deduced (see Figure 11.1).

We assume that the drift velocity $\mathbf{v}$ of each species of carrier can be treated using the Relaxation Time Approximation, *i.e.*

$$m^*\{\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \frac{\mathbf{v}}{\tau}\} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}, \tag{11.1}$$

where q is the charge of the carrier, m^* is its effective mass (assumed isotropic and energy-independent) and τ^{-1} is its relaxation (scattering) rate. Note that all changes in $\mathbf{v}$ occur in the plane perpendicular to $\mathbf{B}$; therefore it is sufficient to split Equation 11.1 into x and y components to give

$$m^*\{\frac{\mathrm{d}v_x}{\mathrm{d}t} + \frac{v_x}{\tau}\} = qE_x + qv_yB \tag{11.2}$$

and

$$m^*\{\frac{\mathrm{d}v_y}{\mathrm{d}t} + \frac{v_y}{\tau}\} = qE_y - qv_xB. \tag{11.3}$$

The Hall effect represents a steady state of the system, *i.e.* $\mathrm{d}v_x/\mathrm{d}t = \mathrm{d}v_y/\mathrm{d}t = 0$. Substituting Equation 11.2 into Equation 11.3 with $\mathrm{d}v_x/\mathrm{d}t = \mathrm{d}v_y/\mathrm{d}t = 0$ gives

$$\frac{m^*v_y}{\tau} = qE_y - \frac{qB\tau}{m^*}\{qE_x + qv_yB\}, \tag{11.4}$$

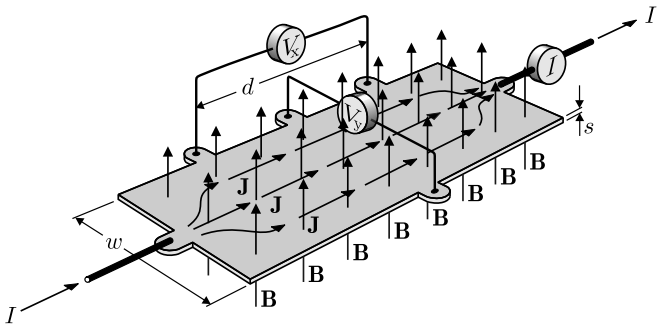


Figure 11.1: Geometry of a Hall effect measurement on a sample of thickness s and width w. On entering the sample, the current I becomes a current density $\mathbf{J}$ of average magnitude I/ws. The magnetic field (flux density) is uniform within the sample. The positions of voltmeters for measuring $E_x = V_x/d$ and $E_y = V_y/w$ are shown symbolically.

which can be rearranged to give

$$v_y\{\frac{m^*}{\tau} + \frac{q^2B^2\tau}{m^*}\} = qE_y - \frac{q^2B\tau}{m^*}E_x. \tag{11.5}$$

Dividing through by m^*/τ and making the identification $eB/m^* \equiv \omega_c$ (*i.e.* the cyclotron frequency) gives

$$v_y\{1 + \omega_c^2\tau^2\} = \frac{q\tau}{m^*}\{E_y - \omega_c\tau E_x\}. \tag{11.6}$$

Hall effect experiments are usually quite deliberately carried out at low magnetic fields, such that $\omega_c\tau \ll 1$, implying that terms $\sim \omega_c^2\tau^2$ can be neglected. Therefore Equation 11.6 becomes

$$v_y = \frac{q\tau}{m^*}\{E_y - \omega_c\tau E_x\} = \frac{q\tau}{m^*}\{E_y - \frac{qB\tau}{m^*}E_x\}. \tag{11.7}$$

We now consider an arbitrary number of carrier types, with each type being labelled by the integer j; the jth carrier type has effective mass m_j^*, charge q_j, scattering rate τ_j and number density n_j. For each carrier type, Equation 11.7 therefore becomes

$$v_{y,j} = \frac{q_j\tau_j}{m_j^*}\{E_y - \frac{q_jB\tau_j}{m_j^*}E_x\}. \tag{11.8}$$

Now the net transverse current must be zero, as there is nowhere for it to go (see Figure 11.1). Therefore

$$\sum_j n_j v_{y,j} q_j = 0. \tag{11.9}$$

Equations 11.8 and 11.9 can be used to derive the Hall coefficient for an arbitrary number of carrier types. We shall use them to treat the simple case of electrons and heavy holes in a semiconductor; as usual, the light holes, with their relatively feeble density of states compared to that of the heavy holes, will be ignored.

11.2.2 Hall effect in the presence of electrons and holes

In the case of electrons in the conduction band (with effective mass m_c^*, scattering rate τ_c^{-1}, charge q_c, density n) and heavy holes in the valence band (with effective mass m_{hh}^*, scattering rate τ_{hh}^{-1}, charge q_{hh}, density p), Equations 11.8 and 11.9 combine to give

$$\frac{nq_c^2\tau_c}{m_c^*}\{E_y - \frac{q_cB\tau_c}{m_c^*}E_x\} + \frac{pq_{hh}^2\tau_{hh}}{m_{hh}^*}\{E_y - \frac{q_{hh}B\tau_{hh}}{m_{hh}^*}E_x\} = 0. \tag{11.10}$$

Equation 11.10 can be rearranged to give

$$E_y\{n\mu_c + p\mu_{hh}\} = E_x\{p\mu_{hh}^2 - n\mu_c^2\}B, \tag{11.11}$$

where $\mu_{hh} = |q_{hh}\tau_{hh}/m_{hh}^*|$ is the heavy hole mobility and where $\mu_c = |q_c\tau_c/m_c^*|$ is the electron mobility, and I have substituted $q_{hh} \equiv +e$ and $q_c \equiv -e$. Now

$$E_x = \frac{J_x}{\sigma} = \frac{J_x}{|e|(n\mu_c + p\mu_{hh})}, \tag{11.12}$$

where J_x is the current density in the x direction. Combining Equations 11.11 and 11.12 gives

$$R_H \equiv \frac{E_y}{J_x B} = \frac{1}{|e|}\frac{(p\mu_{hh}^2 - n\mu_c^2)}{(n\mu_c + p\mu_{hh})^2}. \tag{11.13}$$

This treatment is explored in more depth in the Problems.[1]

11.2.3 A clue about the origins of magnetoresistance

Equations 11.9 and 11.10 show that, although no *net* current flows in the y direction, the currents carried in the y direction by a particular type of carrier may (*i.e.* probably will) be non-zero. Carriers flowing in the y direction will experience a Lorentz force caused by $\mathbf{B}$ in the *negative x direction* (you can satisfy yourself that this will always be the case). This backflow of carriers will act to change the apparent resistivity E_x/J_x, *i.e.* cause magnetoresistance.[2]

In the following Section we shall explore this idea more formally and in a very general manner. We shall see that the presence of more than one "carrier type" (and here the term is used very imprecisely) is necessary for magnetoresistance to be observed.

In order to treat all of the different contributions to the conductivity and resistivity in a sanitary fashion, we shall introduce the idea of conductivity and resistivity tensors.

11.3 Magnetoresistance in metals

11.3.1 The absence of magnetoresistance in the Sommerfeld model of metals

We consider first of all a metal with a simple spherical Fermi surface and isotropic, energy-independent effective mass. As in Section 11.2.1, the magnetic field $\mathbf{B}$ will be parallel to z (see Figure 11.1), and we shall use the same symbols (effective mass m^*, scattering rate τ^{-1} and electronic charge $-e$). Applying Equation 11.1, we have

$$m^*\{\frac{d\mathbf{v}}{dt} + \frac{\mathbf{v}}{\tau}\} = -e\mathbf{E} - e\mathbf{v}\times\mathbf{B}. \tag{11.14}$$

Two things may be deduced from this equation.

- The motion of the electrons in the direction parallel to $\mathbf{B}$ is unaffected. Therefore there will be *no longitudinal magnetoresistance*; in this context, the longitudinal resistivity is measured in the direction parallel to the field $\mathbf{B}$ (*i.e.* both the applied current density and the measured electric field are parallel to $\mathbf{B}$).

- There may well be *transverse magnetoresistance*. Here *transverse* resistivity means that measured in the direction perpendicular to the field $\mathbf{B}$ (*i.e.* both the applied current density and the measured electric field are in the plane perpendicular to $\mathbf{B}$).

[1] Some excellent illustrative data are shown in Figure 4.3 of *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991).

[2] Those who are unconvinced by this hand-waving argument should go back to Equation 11.6 and repeat the above derivation for $\omega_c\tau \gg 1$. They will find that $((p/\mu_{hh}) + (n/\mu_c))E_y = (p-n)E_xB$ (*i.e* the Hall field is zero if $n = p$). Putting $n = p$ yields $J_x = ((p/\mu_{hh}) + (n/\mu_c))eE_x/B^2$, *i.e.* $\rho \propto B^2$. This is the reason for the very large magnetoresistance in compensated semimetals (equal number of holes and electrons at Fermi surface) such as Bi.

Figure 11.2: Geometrical interpretation of the components of current density J_x, J_y and caused by electric field component E_x and magnetic field $(0, 0, B)$; **J** is the total current density.

Let us look at the second point in more detail. To simplify matters, we shall initially consider an electric field directed only along the x direction (*i.e.* $\mathbf{E} = (E_x, 0, 0)$); our tactic will be to deduce the components of the current density $\mathbf{J} = (J_x, J_y, 0)$ that flow in response to **B** and **E**.

As in the previous Section, we are dealing with a steady state of the system, *i.e.* $\mathrm{d}v_x/\mathrm{d}t = \mathrm{d}v_y/\mathrm{d}t = 0$. Taking $\mathbf{B} = (0, 0, B)$ and $\mathbf{E} = (E_x, 0, 0)$ as defined above, we rewrite Equation 11.2 and Equation 11.3 in the form

$$v_{\mathrm{d},x} = -\frac{e\tau}{m^*}\{E_x + v_{\mathrm{d},y}B\} \tag{11.15}$$

and

$$v_{\mathrm{d},y} = \frac{e\tau}{m^*}v_{\mathrm{d},x}B, \tag{11.16}$$

where the subscript "d" emphasises the fact that we are dealing with a drift velocity.

Equations 11.15 and 11.16 show that the magnetic field has made the conductivity anisotropic; it has become a tensor, rather than a scalar. In order to work out the components of the *conductivity tensor*, we look at the current densities $J_x = -nev_{\mathrm{d},x}$ and $J_y = -nev_{\mathrm{d},y}$. Substituting these into Equations 11.15 and 11.16 yields, after some rearrangement

$$J_x = \sigma_{xx}E_x \text{ and } J_y = \sigma_{yx}E_x,$$

where

$$\sigma_{xx} = \frac{\sigma_0}{1 + \omega_{\mathrm{c}}^2\tau^2} \tag{11.17}$$

and

$$\sigma_{yx} = \frac{\sigma_0\omega_{\mathrm{c}}\tau}{1 + \omega_{\mathrm{c}}^2\tau^2}. \tag{11.18}$$

Here $\sigma_0 = ne^2\tau/m^*$ is the zero-field conductivity in the Sommerfeld model and $\omega_{\mathrm{c}} = eB/m^*$ is the cyclotron frequency.

The conductivity tensor shows that, in a magnetic field, the total current density **J** no longer flows parallel to the applied E-field, E_x; instead, it now contains both x and y components. Figure 11.2 gives a geometrical interpretation of **J** and the components of current density J_x, J_y caused by electric field component E_x and magnetic field $(0, 0, B)$.

Equation 11.17 shows that as $B \to \infty$, $\sigma_{xx} \propto B^{-2}$. We might therefore expect to see some magnetoresistance. However, most experiments (see Figure 11.1) measure *voltages* dropped in the x and y directions between pairs of contacts, rather than measuring the x and y components of the current density. In such experiments the current is forced to go along the x direction, so that $\mathbf{J} \equiv J_x\mathbf{e}_1$; in contrast, the electric field will have components in both x and y directions (see Figure 11.3). Therefore we want the components

$$\rho_{xx} \equiv \frac{E_x}{J_x} \text{ and } \rho_{yx} \equiv \frac{E_y}{J_x} \tag{11.19}$$

of the *resistivity* tensor, rather than the conductivity.

The general conductivity tensor is

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{yx} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}. \tag{11.20}$$

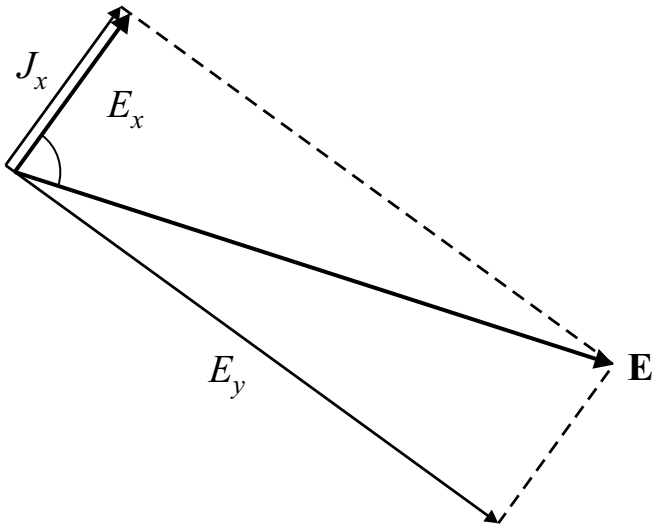


Figure 11.3: Geometrical interpretation of the components of electric field E_x, E_y and the total field **E** caused by current density component J_x and magnetic field $(0, 0, B)$.

The derivations which start at Equations 11.15 and 11.16 can be repeated with $\mathbf{E} = (0, E_y, 0)$ to yield $\sigma_{xy} = -\sigma_{yx}$ and $\sigma_{yy} = \sigma_{xx}$,[3] so that we have

$$\sigma = \begin{pmatrix} \sigma_{xx} & -\sigma_{xy} \\ \sigma_{xy} & \sigma_{xx} \end{pmatrix} = \frac{\sigma_0}{1 + \omega_c^2 \tau^2} \begin{pmatrix} 1 & \omega_c \tau \\ -\omega_c \tau & 1 \end{pmatrix}. \tag{11.21}$$

This tensor can then be inverted using standard methods to give the resistivity tensor

$$\rho = \begin{pmatrix} \rho_{xx} & \rho_{yx} \\ \rho_{xy} & \rho_{yy} \end{pmatrix} = \frac{1}{\sigma_0} \begin{pmatrix} 1 & -\omega_c \tau \\ \omega_c \tau & 1 \end{pmatrix}. \tag{11.22}$$

The components of interest in the experimental arrangement shown in Figure 11.1 are

$$\rho_{xx} = \rho_0 \text{ and } \rho_{yx} = -\rho_0 \omega_c \tau = -\frac{B}{ne}, \tag{11.23}$$

where $\rho_0 = 1/\sigma_0$ (see Figure 11.3). Therefore we get no magnetoresistance in the diagonal components of the resistivity tensor and the familiar Hall effect for one carrier in the off-diagonal components.

11.3.2 The presence of magnetoresistance in real metals

Almost all real metals exhibit some form of magnetoresistance, and so we must try to find out what is wrong with the approach above. In the above derivation we assumed that all carriers had the same value of m^* and τ. However in a real metal we could have

- electrons with different values of m^* (*e.g.* from anisotropic bands);
- electrons with different values of τ (*e.g.* some parts of the Fermi surface may have higher scattering probabilities than others);
- a combination of both.

We therefore split the current density **J** into several components

$$\mathbf{J}_j = \sigma_{xx,j} E_x \mathbf{e}_1 + \sigma_{yx,j} E_y \mathbf{e}_2 \tag{11.24}$$

[3] This fact can also be deduced using symmetry considerations.

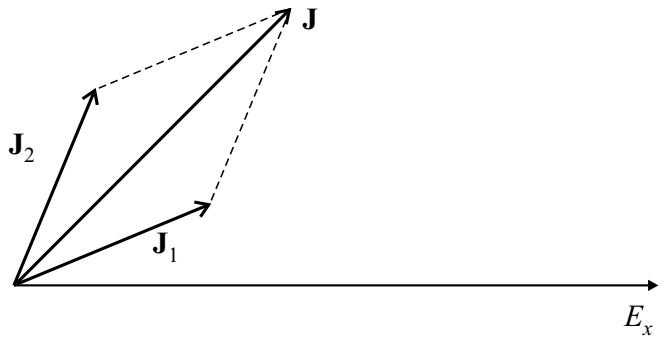


Figure 11.4: Geometrical interpretation of the components of current density $\mathbf{J}_1$ and $\mathbf{J}_2$ due to two different species of carrier caused by electric field component E_x and magnetic field $(0, 0, B)$.

where the index j indicates a contribution from the jth type of carrier. Each type of carrier will have a different density n_j and/or effective mass m_j^* and/or scattering rate τ_j^{-1}. Hence, the components of the conductivity tensor $\sigma_{xx,j}$ and $\sigma_{yx,j}$ will differ for each type of carrier, resulting in $\mathbf{J}_j$s which do not in general point in the same direction as each other.

The total current $\mathbf{J}$ is just the sum of all of the components

$$\mathbf{J} = \sum_j \mathbf{J}_j. \tag{11.25}$$

In order to see what happens, we take a very simple case of just two carrier types, $j = 1, 2$. Equation 11.24 shows that, barring some very unlikely coincidence, $\mathbf{J}_1$ and $\mathbf{J}_2$ will be in different directions. This situation is illustrated in Figure 11.4; the application of the magnetic field means that $\mathbf{J}_1$ and $\mathbf{J}_2$ are no longer parallel, so that

$$|\mathbf{J}| \leq |\mathbf{J}_1 + \mathbf{J}_2|, \tag{11.26}$$

i.e. the resistivity increases with increasing magnetic field. We therefore have magnetoresistance.

We note in passing that, as above, $\sigma_{xx} \propto B^{-2}$ as $B \to \infty$ (see Equation 11.17 and the paragraph following it). This will be important in the discussion of the following section.

11.3.3 The use of magnetoresistance in finding the Fermi surface shape

We consider first a *closed* section of Fermi surface, about which a carrier can perform closed orbits under the influence of a magnetic field (see Figure 11.5). As $B \to \infty$, $\omega_c\tau \to \infty$, so that an electron will tend to make many circuits of the Fermi surface before scattering. Therefore, the velocity of the electron in the plane perpendicular to $\mathbf{B}$ will average to zero; this is the reason why σ_{xx} and σ_{yy} both vary as B^{-2} in very high fields. Using the conductivity tensor components, the current densities can be written

$$J_x = \sigma_{xx}E_x + \frac{1}{RB}E_y \tag{11.27}$$

and

$$J_y = -\frac{1}{RB}E_x + \sigma_{yy}E_y \tag{11.28}$$

where R is the Hall coefficient, and the off-diagional tensor components have been written $\sigma_{xy} = 1/RB$ and $\sigma_{yx} = -1/RB$. Eliminating E_y gives

$$E_x = \frac{1}{\sigma_{xx} + (R^2B^2\sigma_{yy})^{-1}}J_x + \frac{RB}{1 + R^2B^2\sigma_{xx}\sigma_{yy}}J_y. \tag{11.29}$$

As mentioned above, σ_{xx} and σ_{yy} both vary as B^{-2} in very high fields, so that as $B \to \infty$, $(B^2R^2\sigma_{yy})^{-1} \gg \sigma_{xx}$. Therefore, $\rho_{xx} = E_x/J_x$ tends to a constant at high fields, *i.e.* it *saturates*.

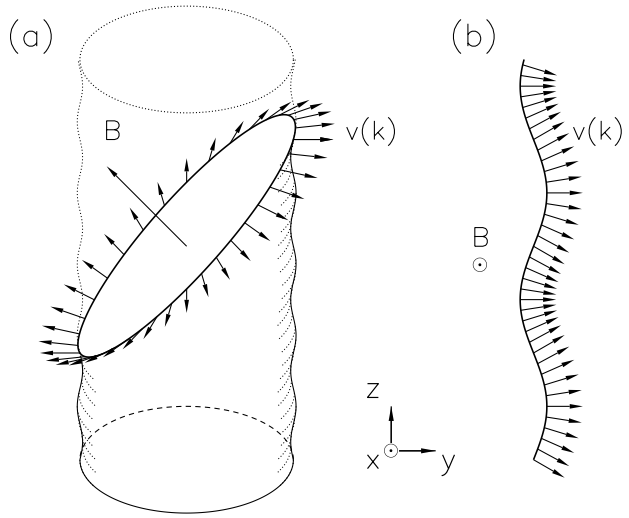


Figure 11.5: (a) Schematic of electron motion on a closed section of Fermi surface in a magnetic field. The arrows indicate the velocities of an electron following a closed orbit about the Fermi surface in a plane perpendicular to the magnetic field **B**. (b) An open orbit on the Fermi surface. In an in-plane magnetic field, electrons will be driven across the Fermi surface, so that their velocities (shown by arrows) will rock from side to side.

We now turn to an *open* section of Fermi surface (see Figure 11.5) about which an electron cannot perform closed orbits under the influence of a magnetic field. In this case, even as $B \to \infty$, the average value of v_y remains finite; therefore $\sigma_{yy} \to C$, a constant. Substituting $\sigma_{yy} = C$ and $\sigma_{xx} = AB^{-2}$, where A is another constant, into Equation 11.29 yields

$$\rho_{xx} = \frac{E_x}{J_x} = \frac{1}{AB^{-2} + (R^2B^2C)^{-1}} \propto B^2, \tag{11.30}$$

i.e. $\rho_{xx} \propto B^2$ as $B \to \infty$.

We therefore have two distinct results

- **closed orbits** produce a ρ_{xx} which saturates as $B \to \infty$;
- **open orbits** produce a ρ_{xx} proportional to B^2 as $B \to \infty$.

This has been used to great effect in elucidating the Fermi surface of metals such as Copper, where changing the orientation of the magnetic field can produce open or closed orbits about the Fermi surface (see Figure 11.6).

11.4 The magnetophonon effect

Oscillations can be observed in the resistivity of both bulk and two-dimensional semiconductors at elevated temperatures ~ 100 K; this is known as *the magnetophonon effect* or *magnetophonon resonance*. The effect is caused by resonant inter-Landau-level scattering of electrons by long-wavelength longitudinal optic (LO) phonons; such phonons are very effective scatterers of electrons. (Why? think about the type of polarisation field that they produce.) As such phonons have virtually zero wavevector, the transition is "vertical", *i.e.* between almost identical points in k-space in the initial and final Landau levels involved. By conservation of energy, the condition for the magnetophonon effect to occur is therefore

$$j\omega_c = \omega_{LO}, \tag{11.31}$$

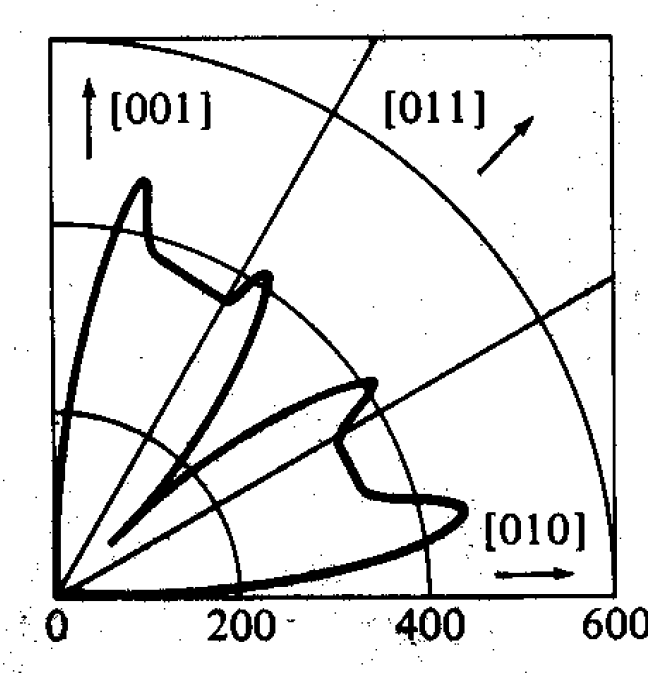


Figure 11.6: Magnetoresistance of Copper at a temperature of 4.2 K and a fixed magnetic field of 1.8 T; the current has been applied in the [100] direction (perpendicular to the plane of the page) and the magnetic field has been rotated from the [001] direction to the [010] direction. The magnetoresistance has been plotted radially as $(\rho(B) - \rho(B = 0))/\rho(B = 0)$. (Data from J.R. Klauder and J.E. Kunzler, *The Fermi Surface*, edited by W. Harrison (Wiley, New York, 1960.)

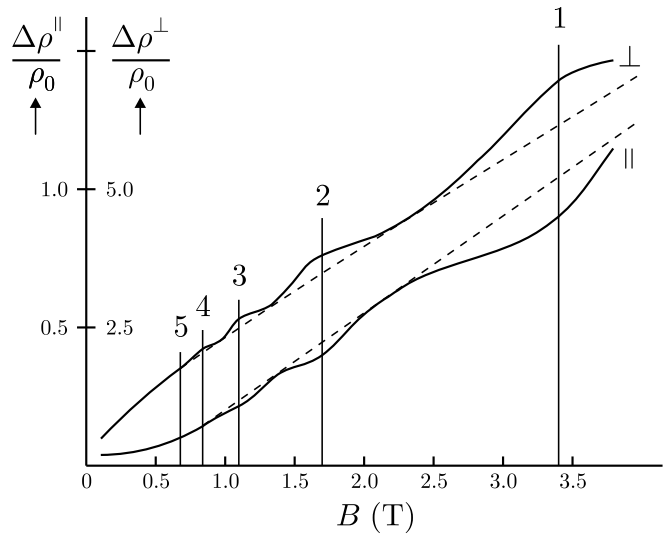


Figure 11.7: Magnetophonon resonances in the longitudinal and transverse resistvities of InSb at 90 K.

where $\omega_{\rm LO}$ is the phonon frequency, *i.e.* an integer number j of Landau level spacings matches an LO phonon energy. This leads to oscillations in the resistivity periodic in $1/B$; if the phonon frequency is known, the effective mass can be deduced from Equation 11.31.

The conditions for magnetophonon resonance to be observed are

- the temperature should be low enough for the Landau levels to be resolved;
- the temperature should be high enough for a substantial population of LO phonons.

In practice, 70-100 K seems to be a good compromise. Figure 11.7 shows magnetophonon resonances in InSb at 90 K.

11.5 Reading

Some useful general reading on magnetoresistance is contained in *Electrons in Metals and Semiconductors*, by R.G. Chambers (Chapman and Hall, London 1990) Chapters 1, 2 and 11, *Semiconductor Physics*, by K. Seeger (Springer, Berlin 1991) Chapter 9 (hard), *Solid State Physics*, by N.W Ashcroft and N.D. Mermin (Holt, Rinehart and Winston, New York 1976) Chapters 12, 13 and 15 (hard).

Handout 12

Magnetoresistance in two-dimensional systems and the quantum Hall effect

12.1 Introduction: two dimensional systems

Lectures 9 and 11 described the effects of a magnetic field on three-dimensional carrier systems; in particular, Lecture 9 discussed Landau quantisation. We shall now look at two-dimensional semiconductor systems in an analogous manner. The two-dimensionality removes one degree of freedom, so that at high magnetic fields the Landau quantisation results in a density of states consisting entirely of discrete levels, rather than a series of one-dimensional density-of-state functions. This means that the oscillatory phenomena seen in high magnetic fields are often far more extreme in two-dimensional systems than they are in the three-dimensional equivalent.

The reduced dimensionality also results in some other novel phenomena; the quantum Hall effect is a general low-temperature property of two-dimensional or quasi-two-dimensional systems and has been observed in semiconductor structures (*e.g.* Si MOSFETs, GaAs-(Ga,Al)As heterojunctions and quantum wells containing two-dimensional holes or electrons) and, very recently, in some organic molecular conductors. Most of the research on the quantum Hall effect has been carried out using modulation-doped GaAs-(Ga,Al)As heterojunctions, as these can provide an exceptionally clean two-dimensional electron system with a very low scattering rate (see Lecture 8). We shall therefore confine our discussion to electrons in GaAs-(Ga,Al)As heterojunctions, although similar principles will apply to any of the other systems mentioned above. The effective mass of the electrons in the heterojunction will be labelled m^*, and their areal carrier density N_s.

In a heterojunction, the motion of the electrons in the growth (z) direction is confined by the approximately triangular potential well at the interface; in this direction, the electrons are in bound states of the well known as *subbands*. We let the jth subband have energy E_j. If a magnetic field is applied in the z direction, then the electrons' motion in the xy plane will also become quantised into Landau levels, with energy $(l + \frac{1}{2})\hbar\omega_\mathrm{c}$. Therefore, the total energy of the electron becomes

$$E(B, l, j) = (l + \frac{1}{2})\hbar\omega_\mathrm{c} + E_j, \tag{12.1}$$

i.e. the system has become completely quantised into a series of discrete levels.

(We remark at this stage that the magnetic field has reduced the effective dimensionality of the electron system by two; its density of states is now like that of *e.g.* an atom, with no electronic dispersion whatsoever. This is a general property of a magnetic field; in the case of a three-dimensional system, the electronic density of states becomes a series of density of states functions for one-dimensional motion.)

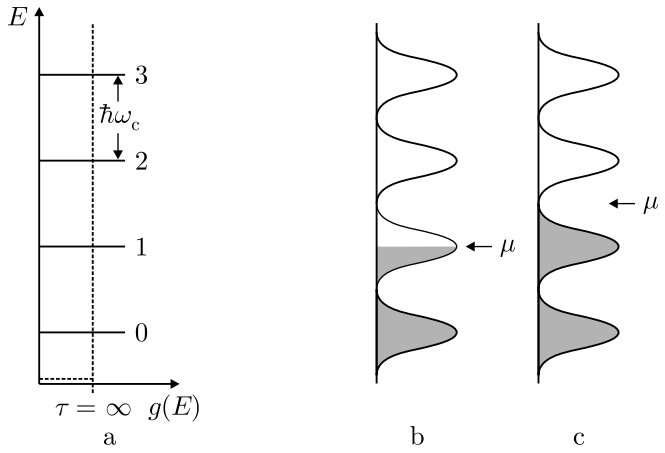


Figure 12.1: The Landau level density of states with zero scattering (left) and in the presence of broadening due to scattering (right). The shaded portions indicate states filled with electrons. Two positions of the chemical potential (Fermi energy) are shown.

12.2 Two-dimensional Landau-level density of states

Let us assume for the time being that only the lowest subband is populated with electrons. The density of states, which is a series of delta functions separated by $\hbar\omega_c$, is shown in Figure 12.1a. In a real sample, these levels will not be perfectly sharp, but instead broadened due to the scattering of electrons by imperfections at the surface of the layer and the presence of impurities. If the typical relaxation time between scattering events is τ (*i.e.* the lifetime of a particular quantum-mechanical state is τ), the Uncertainty Principle leads to broadening of the levels by an amount $\delta E = \hbar/\tau$. The effects of the magnetic field will therefore only be important if the energy separation of the Landau levels, $\hbar\omega_c$, is larger than this broadening (*i.e.* $\omega_c\tau \gg 1$). We shall concentrate on this high-field regime of well-separated Landau levels.

Since the number of electronic states at zero field has to equal the number at finite field, the number of electrons which it is possible to put into each Landau level may be calculated from the number of free electron wavevector states with energies between adjacent Landau levels. The wavevector k_l with zero field energy corresponding to the lth Landau level is given by

$$\frac{\hbar^2 k_l^2}{2m^*} = (l + \frac{1}{2})\hbar\omega_c. \tag{12.2}$$

We divide the area of wavevector-space between k_l and k_{l+1} by the well-known degeneracy of two-dimensional phase-space ($N = 2(2\pi)^{-2}A_rA_k$, where A_r is the area in r-space, A_k is the area in k-space and the factor 2 takes into account spin) to get the number of states per unit area per Landau level

$$\frac{N}{A} = 2\frac{\pi k_{n+1}^2 - \pi k_n^2}{(2\pi)^2} = \frac{2eB}{h}. \tag{12.3}$$

Each Landau level can hold at most this number of electrons before it is necessary to start to fill the next level.

The electrical properties of the sample will depend very strongly upon how the Landau levels are occupied. Let us assume first of all that the temperature is absolute zero, and that the highest occupied Landau level is half full (Figure 12.1b), *i.e.* the chemical potential μ is in the middle of a Landau level. Under such conditions, the system looks like a metal; there are lots of empty states which are just above μ, and so the system will have a high electrical conductivity.

Next, consider what happens when the highest filled Landau level is completely filled (Figure 12.1c). In this case, μ will be in the gap between the highest occupied and lowest unoccupied Landau level; there will be no empty states available at energies close to μ This situation is analogous to that of an

insulator; the conductivity of the sample will obviously be a minimum under such circumstances, and for well-separated Landau levels will be effectively zero.

Finite temperatures will cause the Fermi–Dirac distribution function to become "smeared out" close to μ; however, as long as $\hbar\omega_c >\sim k_B T$, the general picture remains true that the conductivity oscillates dramatically as the field is swept. The conductivity will be very small whenever we have a completely filled highest Landau level, *i.e* when

$$N_s = j\frac{2eB}{h}, \tag{12.4}$$

where j is an integer and N_s is the number of electrons per unit area. The maximum conductivity case occurs when

$$N_s = (j + \frac{1}{2})\frac{2eB}{h}. \tag{12.5}$$

The conductivity will therefore oscillate as a function of magnetic field, and will be periodic in $1/B$, with a periodicity given by

$$\Delta(1/B) = \frac{2e}{hN_s}. \tag{12.6}$$

12.2.1 Resistivity and conductivity tensors for a two-dimensional system

The experimental geometry for the quantum Hall effect is usually exactly the same as that for the conventional Hall effect (see Figure 11.1). Thus, it is the components of the resistivity tensor that we measure, and not the conductivity. We can relate the current densities J_x and J_y (in two dimensions current densities will have the units *current per unit length*) and electric fields E_x and E_y by the expressions

$$J_x = \sigma_{xx}E_x + \sigma_{xy}E_y \tag{12.7}$$

and

$$J_y = -\sigma_{xy}E_x + \sigma_{xx}E_y \tag{12.8}$$

since $\sigma_{xx} = \sigma_{yy}$ and $\sigma_{xy} = -\sigma_{yx}$ for a homgeneous isotropic substance. If, as usual, we allow a current to flow along only one direction (x) (see Figure 11.1), then $J_y = 0$ and we can say that $E_y/E_x = \sigma_{xy}/\sigma_{xx}$. Therefore we can define the resistivities

$$\rho_{xx} \equiv \frac{E_x}{J_x} = \frac{\sigma_{xx}}{\sigma_{xx}^2 + \sigma_{xy}^2} \tag{12.9}$$

and

$$\rho_{xy} \equiv \frac{E_y}{J_x} = \frac{\sigma_{xy}}{\sigma_{xx}^2 + \sigma_{xy}^2}. \tag{12.10}$$

Note that the units of ρ_{xx} are known as Ω per square; a dimensionless geometrical factor (sample length/sample width) must be introduced to work out the resistance of a particular sample.

We use the Relaxation-Time Approximation to work out the motion of the electrons, *i.e.*

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{e}{m^*}\mathbf{E} - \frac{e}{m}\mathbf{v}\times\mathbf{B} - \frac{\mathbf{v}}{\tau}. \tag{12.11}$$

In steady state, the LHS vanishes, and the two components of the equation read

$$v_x = -\frac{e\tau}{m^*}E_x - \omega_c\tau v_y \tag{12.12}$$

and

$$v_y = -\frac{e\tau}{m^*}E_y + \omega_c\tau v_x, \tag{12.13}$$

where we have written $\omega_c = eB/m^*$. If we impose the condition $v_y = 0$ (no current in the y direction) we have

$$\frac{E_y}{E_x} = -\omega_c\tau. \tag{12.14}$$

Writing $J_x = -eN_s v_x$ Equations 12.12 and 12.14 can be combined to give

$$R_H \equiv \frac{E_y}{J_x B} = -\frac{1}{N_s e}. \tag{12.15}$$

R_H is of course the *Hall coefficient*. Referring to Figure 11.1, we see that for a two-dimensional system, $J_x = I/w$ and $E_y = V_y/w$, where w is the width of the sample, *i.e.*

$$R_H = \frac{V_y}{IB}. \tag{12.16}$$

As in the case of the three-dimensional systems discussed previously, the low-field Hall coefficient can be used to determine N_s.

In high magnetic fields, $\omega_c \tau \gg 1$. Under these conditions, Equation 12.14 indicates that $|\sigma_{xy}| \gg |\sigma_{xx}|$, so that Equations 12.9 and 12.10 give

$$\rho_{xx} \approx \frac{\sigma_{xx}}{\sigma_{xy}^2} \tag{12.17}$$

and

$$\rho_{xy} \approx \frac{1}{\sigma_{xy}} = R_H B. \tag{12.18}$$

We are therefore left with the rather unexpected result that the resistivity is proportional to the conductivity. This is due to the very important influence of the Hall field, and the way in which resistivity and conductivity have been defined; the resistivity is defined for zero Hall current, and the conductivity for zero Hall field.

At the values of magnetic field corresponding to a completely filled highest Landau level, the "insulating" case, it is the *conductivity* σ_{xx} which goes to zero, and hence the resistivity ρ_{xx} also tends to zero. Physically speaking, what has happened is that, as the probability of scattering goes to zero, an electron (for $E_y = 0$) will go around a cyclotron orbit with a drift in a direction perpendicular to the electric and magnetic fields, and no current will flow along the direction of the applied electric field. For the resistivity (with $J_y = 0$) the Hall field balances out the Lorenz force, and so a current can flow with very small scattering probability, giving a vanishing resistivity. Experimentally, ρ_{xx} can become immeasurably small.

12.3 Quantisation of the Hall resistivity

Some very interesting things happen when we look at the Hall voltage. Experimentally, at the magnetic field values where ρ_{xx} goes to zero, the Hall voltage becomes independent of field, and a *plateau* appears. When ρ_{xx} increases again, then the Hall voltage jumps up to the next plateau. The most interesting feature of this behaviour is the *value* of the Hall resistivity on the plateaux. We may calculate this very easily.

It was mentioned above that ρ_{xx} goes to zero when a whole number of Landau levels are completely filled. At this point the density of electrons is (see Equation 12.4)

$$N_s = j\frac{2eB}{h} \tag{12.19}$$

And therefore the Hall resistivity is given by

$$\rho_{xy} = R_H B = \frac{B}{N_s e} = \frac{1}{j}\frac{h}{2e^2}. \tag{12.20}$$

As we have seen above (see Equation 12.16), for a truly two-dimensional system the width of the sample cancels out (*i.e.* ρ_{xy} is measured in Ω), so that

$$\rho_{xy} \equiv \frac{V_y}{I} = \frac{1}{j}\frac{h}{2e^2} = \frac{12906.5}{j}\ \Omega. \tag{12.21}$$

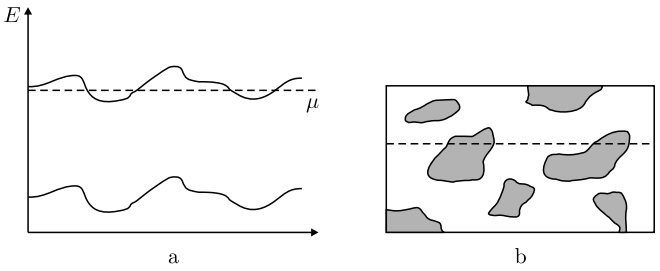


Figure 12.2: (a) Spatial variation of Landau level energies caused by disorder (impurities *etc.*). μ indicates the chemical potential. (b) Plan of the resulting "puddles" of electrons.

The remarkable feature of the experiment is that this value of Hall resistivity, determined only by the fundamental constants h and e, is obtained extremely accurately (see also Section 12.3.2).

However, even though our simple derivation has predicted the value of the quantised resistance, it only gives this value *at one point*, the field at which an integer number of Landau levels are completely filled. In experiments, the plateaux *extend over finite widths of magnetic field.* In order to see why this occurs, we must introduce *disorder*.

12.3.1 Localised and extended states

Figure 12.1 showed broadening of the Landau levels due to scattering of the electrons. One way to look at this is to say that the imperfections in the sample cause a potential energy variation through space, and that the magnitude of this variation is equal to the broadening of the levels. This means that the energy of a Landau level moves up and down as we move through the sample. This is shown in Figure 12.2a.

Let us imagine varying the field. Starting at the field at which one Landau level is filled, the field is lowered. This lowers the Landau level degeneracy so that some electrons must go into the next Landau level. These electrons fill up the lowest states, in the minima of the potential fluctuations. The next step is to draw a two-dimensional map of the occupied states. This is shown in Figure 12.2b. Provided that the field does not change too much, then the occupied states will be in isolated islands, which are not connected to each other. Because they are isolated, these electrons will not play any part in conduction, and the resistivity and Hall voltage from the sample will remain constant, independent of whether these states are occupied or not. Hence we can see the Hall plateaux and zero σ_{xx} (and ρ_{xx}) over a range of field.

We therefore have two classes of state, *extended states* at the centres of Landau levels, which allow the electrons to move through the heterojunction and *localised states*, in the "tails" of the Landau levels, which strand the electrons in isolated puddles. When the μ is in the localised states between Landau level centres, $\sigma_{xx} = \rho_{xx} = 0$ and ρ_{xy} is quantised; when μ is in the extended states close to the Landau level centres, σ_{xx} and hence ρ_{xx} are finite.

12.3.2 A further refinement– spin splitting

When experimental data are examined, the Hall plateaux are actually found at resistances

$$\rho_{xy} \equiv \frac{V_y}{I} = \frac{1}{\nu}\frac{h}{e^2} = \frac{25812.8}{\nu}\ \Omega, \tag{12.22}$$

(*c.f.* Equation 12.21) where ν is an integer. The problem is that we have ignored the spin-splitting of the Landau levels in the simple treatment above. The Landau level energies are in fact

$$E(l,B) = \frac{\hbar eB}{m^*}(l+\frac{1}{2}) \pm \frac{1}{2}g^*\mu_{\rm B}B, \tag{12.23}$$

where $\mu_{\rm B}$ is the Bohr magneton, the $\pm$ takes account of spin-up and spin-down electrons and g^* is the effective g-factor. The resulting Landau-level density of states is shown in Figure 12.3; we now have

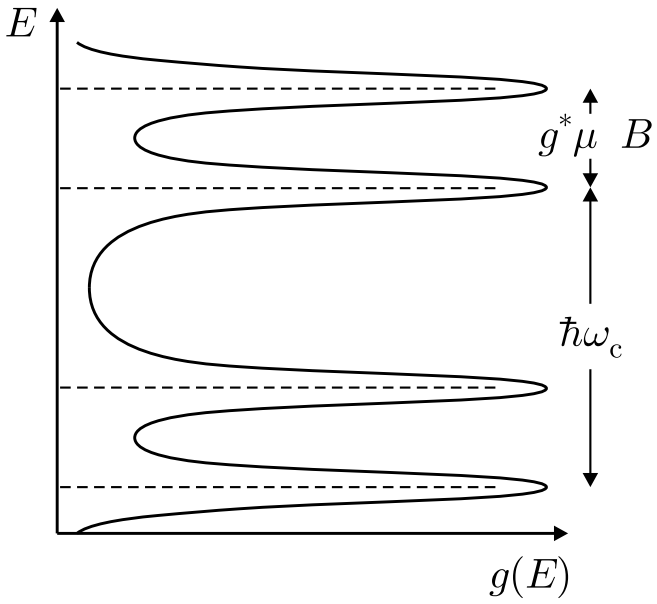


Figure 12.3: Two-dimensional Landau levels in the presence of spin splitting.

twice as many levels as before, each with half the degeneracy, thus explaining the quantised values in Equation 12.22.

The parameter ν is known as the *filling factor*; it is the number of filled (spin-split) half Landau levels, and is given by

$$\nu = \frac{N_s h}{eB}. \tag{12.24}$$

The spin–splitting is much smaller than $\hbar\omega_c$ (*e.g.* $g^* \sim -0.4$ in GaAs), and so Hall plateaux corresponding to *odd* ν (*i.e.* those where μ is in a spin gap rather than a cyclotron gap) tend to be resolved only at higher fields.

12.4 Summary

The experimental data in Figure 12.4 summarise the above discussion very nicely. The following points may be noted.

- At integer ν, the chemical potential μ is in the localised states (*i.e.* away from the centres of the levels). Therefore $\sigma_{xx} = \rho_{xx} = 0$ and $\rho_{xy} = h/\nu e^2$.
- Between integer ν, μ is in the extended states. Therefore σ_{xx} and ρ_{xx} become finite and ρ_{xy} changes.
- As the temperature rises, the plateaux and regions of zero ρ_{xx} become narrower. This is because the Fermi–Dirac distribution function becomes smeared out on either side of μ, allowing some electrons to access extended states even when μ is amongst the localised states.
- Hall plateaux and ρ_{xx} minima with odd values of ν disappear at low fields. This is because odd ν corresponds to μ being in an energy gap $\sim g^*\mu_B B$, much smaller than $\hbar\omega_c$. By contrast, even ν features correspond to μ being in a gap $\sim \hbar\omega_c$.

The Quantum Hall Effect was first discovered in 1980 by von Klitzing *et al.*. Von Klitzing was awarded the 1985 Nobel Physics Prize for its discovery. To date the value of the quantized Hall resistance is known to an accuracy of one part in $\sim 10^8$ and has become the international standard for resistance. It is also the most accurate way to determine the atomic fine-structure constant.[1]

[1]The atomic fine structure constant is $\alpha = e^2/(4\pi\epsilon_0\hbar c)$; α is linked to h/e^2 only by defined constants.

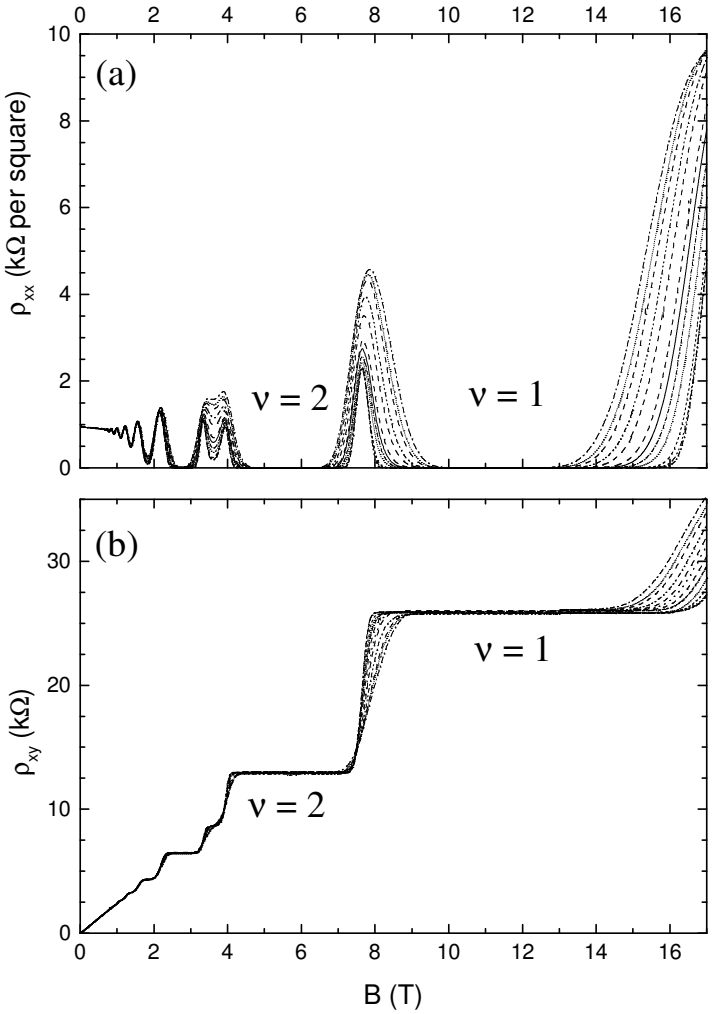


Figure 12.4: Experimental ρ_{xx} and ρ_{xy} data for a GaAs-(Ga,Al)As heterojunction. Data for temperatures between 30 mK and 1.5 K are shown. (Data courtesy of Phil Gee).

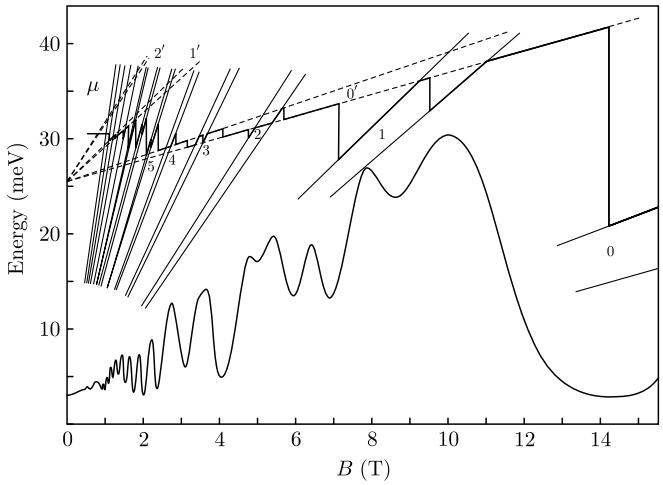


Figure 12.5: ρ_{xx} for a (Ga,In)As-(Al,In)As heterojunction with two populated subbands, plus a schematic representation of the motion of the chemical potential (Fermi energy). (Data from J.C. Portal *et al.*, *Solid State Commun.* **43**, 907 (1982).)

12.5 More than one subband populated

Figure 12.5 shows ρ_{xx} for a (Ga,In)As-(Al,In)As heterojunction with two populated subbands, plus a schematic representation of the motion of the chemical potential (Fermi energy). At low fields, the levels are poorly resolved, and the effects of spin-splitting are not important; two distinct periodicities are observed, given by

$$\Delta(1/B)_1 = \frac{2e}{hN_{s1}}, \tag{12.25}$$

and

$$\Delta(1/B)_2 = \frac{2e}{hN_{s2}}, \tag{12.26}$$

where N_{s1} and N_{s2} are the populations of the two subbands.[2]

At intermediate fields ($\sim$ 10 T), the presence of the second subband complicates the structure of the ρ_{xx} oscillations. However, at high enough fields, the levels will become widely separated in energy and the quantum Hall effect will re-emerge.

12.6 Reading

Simple treatments of the quantum Hall effect are found in *Solid State Physics*, by G. Burns (Academic Press, Boston, 1995) Chapter 18 and *Low-Dimensional Semiconductor Structures*, by M.J. Kelly, (Clarendon Press, Oxford 1995) (spread through the book). This material is a summary of *Band theory and electronic properties of solids*, by John Singleton (Oxford University Press, 2001), Chapter 10, which also describes the *fractional quantum Hall effect*. A good recent review is given in *Fractional quantum Hall effects*, by M. Heiblum and A. Stern, *Physics World*, **13**, no. 3, page 37 (2000)

[2]This result is a straightforward consequence of the direct proportionality of the cross-sectional area of each subband's two-dimensional Fermi surface to the number of carriers it contains. The discussion of the Shubnikov-de Haas effect in Lecture 9 reminds us that the Fermi-surface area is directly related to the frequency of the oscillations.