

## 6. Metals

In a metallic crystal, such as magnesium or iron, we need to take account of the fact that some electrons are free to move around the crystal and are not necessarily bound to specific atoms as they are in an insulator. There is a background sea of mobile electrons. To a first approximation we can treat these electrons as an ideal gas, though we must be careful to take into account the quantum nature of the electrons.

Consider a cubic crystal of size  $L$ , so the volume is  $V = L^3$ . Impose periodic boundary conditions on the electron wave-functions,

$$\Psi(x + L, y, z) = \Psi(x, y, z),$$

and similarly for the  $y$  and  $z$  directions. Assume any wave-function can be expanded in a basis of plane-waves,

$$\psi_{\mathbf{k}}(\mathbf{r}, t) = e^{i(\mathbf{k}\cdot\mathbf{r} - \omega_{\mathbf{k}}t)}.$$

$\psi_{\mathbf{k}}$  is periodic, with period  $L$ , if  $k_x = 0, \pm\frac{2\pi}{L}, \pm\frac{4\pi}{L}, \dots$ , with similar conditions on  $k_y$  and  $k_z$ . Unlike the case of phonons there is no upper limit on  $k_i$  being imposed here.

In the absence of any interactions the Schrödinger equation for  $\psi_{\mathbf{k}}$  is

$$i\hbar\frac{\partial\psi_{\mathbf{k}}}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi_{\mathbf{k}}$$

which gives

$$\hbar\omega_{\mathbf{k}} = \frac{\hbar^2}{2m}\mathbf{k}\cdot\mathbf{k} \quad \Rightarrow \quad \varepsilon_{\mathbf{k}} = \hbar\omega_{\mathbf{k}} = \frac{\hbar^2}{2m}k^2.$$

The momentum is

$$\hat{\mathbf{p}}\psi_{\mathbf{k}} = -i\hbar\nabla\psi_{\mathbf{k}} = \hbar\mathbf{k}\psi_{\mathbf{k}} \quad \Rightarrow \quad \mathbf{p} = \hbar\mathbf{k},$$

as usual. The quantum nature of the electrons, together with the periodic boundary conditions implies that there is one quantum state, one wave-vector, for every volume  $\left(\frac{2\pi}{L}\right)^3$  in  $k$ -space. If there are  $N$  mobile electrons<sup>9</sup> in the volume  $V$  each must occupy a separate quantum state, because of the exclusion principle. One might expect the electrons quantum states to fill a sphere in  $k$ -space of volume  $N\left(\frac{2\pi}{L}\right)^3 = N\left(\frac{8\pi^3}{V}\right)$ , except there is an extra factor of 2 due to the fact that electrons have spin-1/2 and therefore have two spin states, spin up and spin down, for each value of  $\mathbf{k}$ , so they actually fill a sphere with half this volume,  $N\left(\frac{4\pi^3}{V}\right)$ . For finite  $N$  the distribution of quantum states in  $k$ -space is discrete but in the limit of large  $N$  it can be approximated with a smooth continuous distribution within a sphere of volume

$$\frac{4\pi}{3}k_F^3 = N\left(\frac{4\pi^3}{V}\right)$$

---

<sup>9</sup>  $N$  might not be the same as the number of primitive cells in the crystal, there could be more than one mobile electron per primitive cell, *e.g.* Mg. For a monovalent metal with a monatomic basis, *e.g.* Na, K,  $N=\mathcal{N}$ .

and radius

$$k_F = \left( 3\pi^2 \frac{N}{V} \right)^{\frac{1}{3}}.$$

This is called the **Fermi sphere** and  $k_F$  is the **Fermi wave-number**,  $p_F := \hbar k_F$  is the **Fermi momentum**. The energy of a state with wave-number  $k_F$  is called the **Fermi energy**, for non-interacting electrons the Fermi energy is

$$\varepsilon_F = \frac{p_F^2}{2m} = \frac{\hbar^2 k_F^2}{2m} = \frac{\hbar^2}{2m} \left( \frac{3\pi^2 N}{V} \right)^{\frac{2}{3}} = \frac{1}{2} m v_F^2, \quad (28)$$

where

$$v_F = \frac{p_F}{m} = \frac{\hbar k_F}{m} = \frac{\hbar}{m} \left( \frac{3\pi^2 N}{V} \right)^{\frac{1}{3}}$$

is the **Fermi velocity**. In a simple cubic crystal with a monatomic, monovalent basis with lattice spacing  $5\text{\AA}$  for example,  $\frac{N}{V} = \left( \frac{1}{5 \times 10^{-10}} \right)^3 m^{-3} = 8 \times 10^{27} m^{-3}$  giving  $v_F = 7 \times 10^5 m/s \approx 2 \times 10^{-3} c$ .

When interactions between electrons and the lattice are included the lattice structure will distort the Fermi surface away from a sphere to a shape with less symmetry than a sphere but which reflects the symmetry of the underlying lattice.

At zero temperature all quantum states with  $0 \leq k \leq k_F$  are filled and all quantum states with  $k > k_F$  are empty. At finite temperature thermal fluctuations can kick an electron with  $k < k_F$  into a quantum state with  $k > k_F$  provided  $\epsilon(k) \lesssim \epsilon(k_F) + k_B T$ , leaving a state with  $k < k_F$  empty. Such an empty state is called a **hole**.

From (28) the Fermi energy depends on  $N$ , conversely  $N$  depends on the Fermi energy. We can determine the density of states for electrons at any energy from

$$\varepsilon = \frac{\hbar^2}{2m} \left( \frac{3\pi^2 N_\varepsilon}{V} \right)^{\frac{2}{3}} \quad \Rightarrow \quad N_\varepsilon = \left( \frac{2m\varepsilon}{\hbar^2} \right)^{\frac{3}{2}} \frac{V}{3\pi^2}.$$

$N_\varepsilon$  is the total number of quantum states available to an electron with energy less than  $\varepsilon$ . The density of states is then

$$\mathcal{D}(\varepsilon) = \frac{dN_\varepsilon}{d\varepsilon} = \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \varepsilon^{\frac{1}{2}} = \frac{3}{2} \frac{N_\varepsilon}{\varepsilon}. \quad (29)$$

## Fermi-Dirac distribution function

In order to understand the physics of electrons in metals it is important to know the probability distribution for the number of electrons expected to have energy  $\varepsilon$  at any given temperature. For a gas of photons, Planck proposed that the photons corresponding to light with frequency  $\nu = \frac{\omega}{2\pi}$  could only carry energy which is an integral multiple of  $\varepsilon = \hbar\omega$ ,  $E = n\varepsilon$ , with  $n = 1, 2, 3, \dots$  and this notion generalises to other particles, such

as electrons. One subtlety is that for particles that are not simply free or non-interacting, such as electrons in a solid or particles in a non-ideal gas, the energy can be shifted by a constant, called the **chemical potential**  $\mu$ , so  $\varepsilon \rightarrow \varepsilon - \mu$ . Essentially  $\mu$  is the amount of energy needed to add one more particle to a system of  $N$  particles, if  $\mu < 0$  then particles are attracted to the system and  $|\mu|$  is a binding energy, if  $\mu > 0$  then particles are pushed away from the system (chemical potentials are covered in more depth in the statistical mechanics module).

The Boltzmann distribution for system of particles with allowed energy levels  $E_n$  in thermal equilibrium with a heat bath at temperature  $T$  gives the probability of energy level  $\varepsilon$  being occupied,

$$P(E_n) = \frac{e^{-\frac{n(\varepsilon-\mu)}{k_B T}}}{Z},$$

with  $Z = \sum_{n=0}^{\infty} e^{-\frac{n(\varepsilon-\mu)}{k_B T}}$  chosen<sup>10</sup> that the total probability  $\sum_{n=0}^{\infty} P(\varepsilon) = 1$ .

Let  $y = e^{-\frac{\varepsilon-\mu}{k_B T}}$  then for bosons, such as photons,  $Z = \sum_{n=0}^{\infty} y^n = \frac{1}{1-y}$  giving the distribution function for bosons with angular frequency  $\omega$ ,

$$f_B(\varepsilon) := P(\varepsilon) = \frac{y}{1-y} = \frac{1}{e^{\frac{\varepsilon-\mu}{k_B T}} - 1}, \quad (30)$$

provided  $\varepsilon > \mu$ .  $f_B(\varepsilon)$  represents the probability of finding a boson with energy  $\varepsilon$  when the ambient temperature is  $T$  — it is called the **Bose-Einstein** distribution.

For a system of fermionic particles, such as electrons, the basic principle is the same, except that it must be remembered that each individual term for a given  $n$  in the sum for  $Z$  in the Bose-Einstein distribution corresponds physically to  $n$  bosons occupying the same quantum state. As many bosons as one wishes can go into the same energy state, but the Pauli exclusion states that *at most* one Fermion can occupy each quantum state, so for Fermions  $n = 0$  or  $1$  only,  $n \geq 2$  is not allowed. This means  $Z = \sum_{n=0}^1 y^n = 1 + y$ , leading to the **Fermi-Dirac** distribution function for fermions,  $f_F(\varepsilon) := P(\varepsilon) = \frac{y}{y+1}$  or

$$f_F(\varepsilon) = \frac{1}{e^{\frac{\varepsilon-\mu}{k_B T}} + 1}. \quad (31)$$

The difference in sign in the denominators of (30) and (31) is the source a great difference between the behaviour of Fermions and Bosons at low temperatures.

If  $\mu < 0$  and  $\frac{|\mu|}{k_B T} \gg \frac{\varepsilon}{k_B T}$  the 1 in the denominator is irrelevant and both the Bose-Einstein and the Fermi-Dirac distributions look the same. This is the case in a classical gas where both reduce to the Boltzmann distribution

$$f(\varepsilon) \approx e^{\frac{\mu}{k_B T}} e^{-\frac{\varepsilon}{k_B T}} = \frac{e^{-\frac{\varepsilon}{k_B T}}}{Z},$$

---

<sup>10</sup>  $Z$  is called the **partition function** for the system.

with  $Z = e^{-\frac{\mu}{k_B T}} = \sum e^{-\frac{\epsilon}{k_B T}}$ . We can calculate  $\mu$  assuming that the gas particles are free, so the allowed energies,  $\epsilon = \frac{1}{2}mv^2$ , are just their kinetic energies at speed  $v$  and, in the classical limit, replace the sum with an integral. Then using (29), with the factor 2 arising from spin states of an electron removed so  $V \rightarrow \frac{V}{2}$ ,

$$\begin{aligned} Z &= \int_{\epsilon} e^{-\frac{\epsilon}{k_B T}} \frac{dN_{\epsilon}}{d\epsilon} d\epsilon = \frac{V}{4\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^{\infty} \epsilon^{\frac{1}{2}} e^{-\frac{\epsilon}{k_B T}} d\epsilon \\ &= \frac{V}{4\pi^2} \left( \frac{2mk_B T}{\hbar^2} \right)^{\frac{3}{2}} \int_0^{\infty} u^{\frac{1}{2}} e^{-u} du \quad \text{with} \quad u = \frac{\epsilon}{k_B T} \\ &= \frac{V}{4\pi^2} \left( \frac{2mk_B T}{\hbar^2} \right)^{\frac{3}{2}} \frac{\sqrt{\pi}}{2} = V \left( \frac{mk_B T}{2\pi\hbar^2} \right)^{\frac{3}{2}} = e^{-\frac{\mu}{k_B T}}. \end{aligned}$$

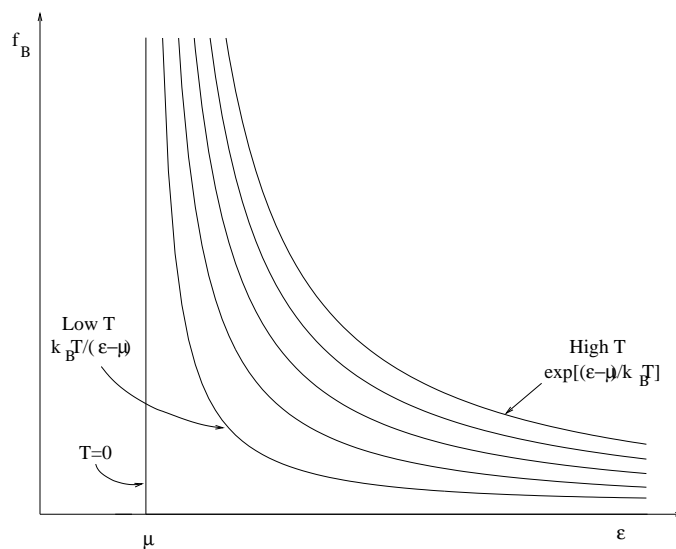
hence

$$\mu = -\frac{3}{2}k_B T \ln(k_B T) + \text{const}$$

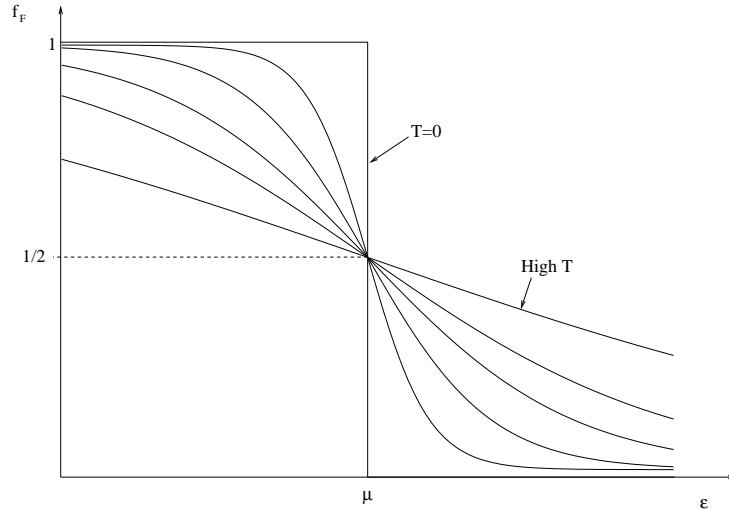
is negative.<sup>11</sup>

The Boltzmann distribution is the large  $T$ , low density, limit of the distribution — the classical limit.

When quantum effects are important the 1 in the denominator of equations (30) and (31) cannot be ignored — roughly speaking this happens when the particle density becomes large enough for their wave-packets to overlap significantly, that is when the separation between the particles becomes of the order of, or less than, their de Broglie wave-length. The Bose-Einstein and Fermi-Dirac distributions, for fixed  $\mu$  and different  $T$ , are plotted below:

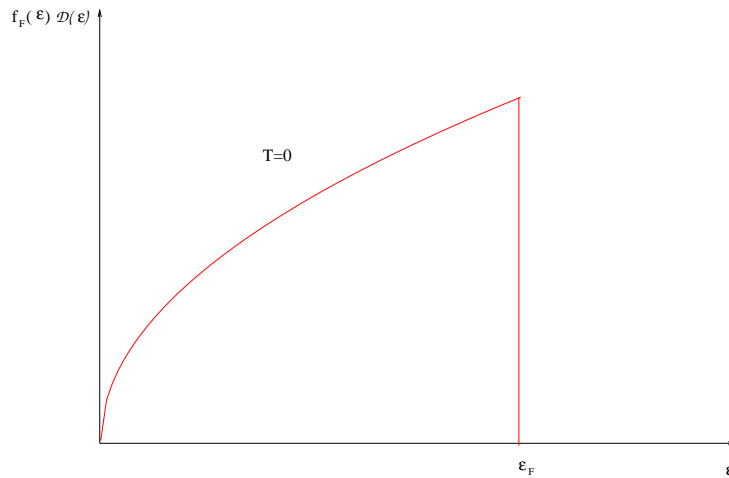


<sup>11</sup> At first sight it may come as a surprise that  $\mu$  is negative for a gas of free particles, an ideal gas, when  $\mu$  is interpreted as the energy to add a particle to the system. This is because  $\mu$  is the energy to add a particle *keeping the temperature and the pressure constant* and this requires the volume to increase — the reduction in energy following from this volume increase is the reason for  $\mu < 0$ .



At  $T = 0$  the Bose-Einstein distribution has no states with  $\varepsilon > \mu$  occupied, the only occupied states are those with  $\varepsilon = \mu$ . All particles occupy the *same* quantum state, a situation known as Bose-Einstein condensation. This phenomenon occurs in superfluids and superconductors.

At  $T = 0$  the Fermi-Dirac distribution has every quantum state with  $\varepsilon < \mu$  occupied,  $f_F(\varepsilon) = 1$  for  $\varepsilon < \mu$ , and every state with  $\varepsilon > \mu$  unoccupied,  $f_F(\varepsilon) = 0$  for  $\varepsilon > \mu$ .  $\varepsilon = \mu = \varepsilon_F$  is the Fermi surface.<sup>12</sup>

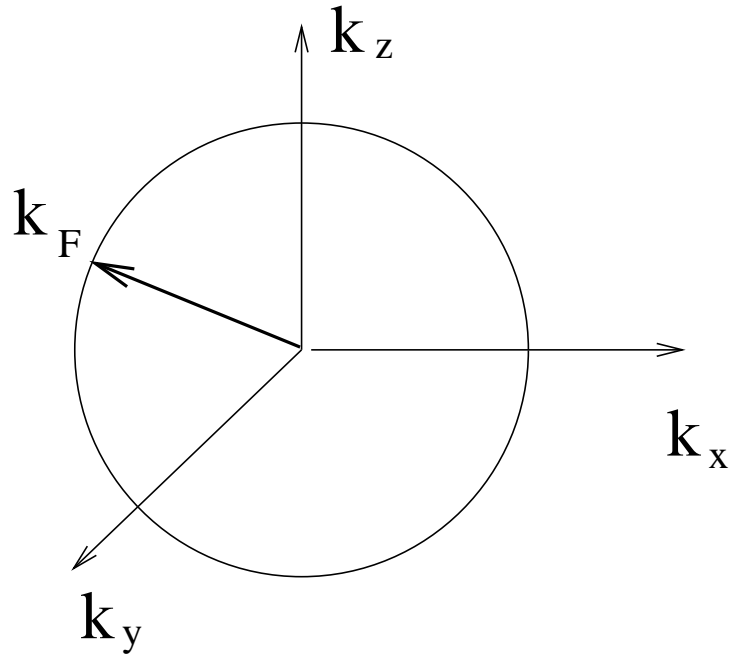


At  $T = 0$  the product  $f_F(\varepsilon)\mathcal{D}(\varepsilon)$  cuts off at  $\varepsilon = \varepsilon_F$  and drops to zero.

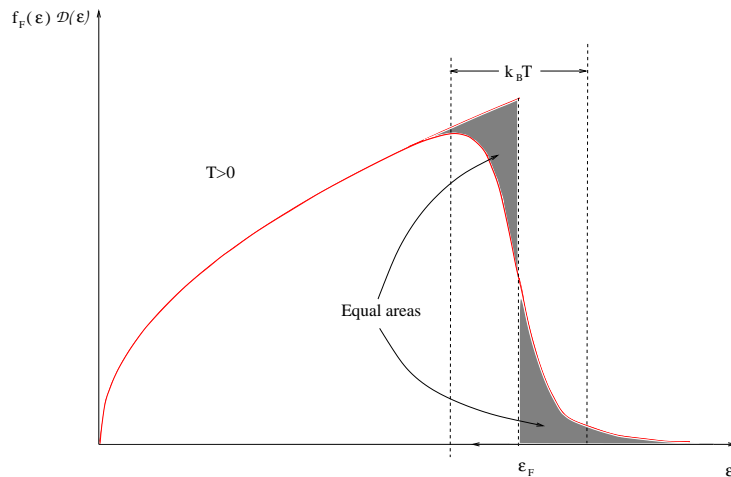
The distribution of quantum states in  $k$ -space at  $T = 0$  is the interior of a solid ball of radius  $k_F$ :

---

<sup>12</sup> Actually the plot of  $f_F(\varepsilon)$  above is produced assuming  $\mu$  is independent of  $T$ . As we have seen this is not a good assumption for a classical ideal gas (the classical case is best visualised by the high  $T$  curves in  $f_B$ ), but is often a good assumption for low  $T$ , in particular when  $k_B T \ll \mu$ . As we shall see for a real Fermi gas  $\mu$  has a slight  $T$  dependence near  $\varepsilon_F$  and the curves for  $f_F(\varepsilon)$  at different  $T$  do not all cross at exactly the same point.



For  $T > 0$  some electrons are thermally excited above  $\varepsilon$ , leaving unoccupied states below  $\varepsilon$ . The excited and unoccupied states lie in a band of width  $k_B T$  around  $\varepsilon_F$ .



The above graph of  $f_F(\varepsilon)\mathcal{D}(\varepsilon)$  as a function of  $\varepsilon$  shows the number of electrons with energy  $\varepsilon$ : the total area under the curve is  $N$ , the number of electrons in the crystal. The chemical potential can be determined as a function of  $T$  and  $N$  by the condition

$$N = \int_0^{\infty} f_F(\varepsilon)\mathcal{D}(\varepsilon)d\varepsilon = \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{\varepsilon^{\frac{1}{2}}}{\left( e^{\frac{\varepsilon-\mu}{k_B T}} + 1 \right)} d\varepsilon. \quad (32)$$

## Heat capacity of a free electron gas

Classically, an ideal gas of  $N$  particles has heat capacity<sup>13</sup>

$$C_V = \frac{3}{2}Nk_B, \quad (33)$$

and specific heat

$$c_V = \frac{C_V}{V} = \frac{3}{2} \frac{N}{V} k_B = \frac{3}{2} n_e k_B, \quad (34)$$

where  $n_e = \frac{N}{V}$  is the number of electrons per unit volume.<sup>14</sup> The observed  $c_V$  in metals is only about  $\sim 1\%$  of (34), a phenomenon which perplexed nineteenth century physicists but which, from a more modern perspective, can be qualitatively understood as being due to the Pauli exclusion principle. Only those electrons within a distance  $\sim k_B T$  of the Fermi surface are free to contribute to the specific heat, electrons any deeper than  $k_B T$  below the Fermi surface have their dynamics ‘frozen’ by the exclusion principle: there are no unoccupied energy states nearby and so these electrons have no degrees of freedom to contribute to the specific heat.

To get a quantitative expression for the specific heat due to free electrons in a metal we first need the internal energy

$$U(T) = \int_0^\infty \varepsilon f_F(\varepsilon) \mathcal{D}(\varepsilon) d\varepsilon \quad (35).$$

The heat capacity is then  $\left. \frac{\partial U}{\partial T} \right|_V$  but one subtlety is that we need to calculate  $U(T)$  at constant  $N$ , while the right hand side of (35) is a function of  $\mu$  which in turn depends on  $T$  and  $N$  through (32). It is not possible to perform the integral (35) analytically so we resort to an approximation, but it is a very good approximation. Define the *Fermi temperature*,  $T_F$  by

$$k_B T_F = \varepsilon_F.$$

Typically  $T_F \approx 50,000^\circ K$  is a very large temperature and we shall expand  $U(T)$  as a function  $\frac{T}{T_F}$  — this is known as a **Sommerfeld expansion**. To see how the Sommerfeld expansion works first start with  $N$  rather than  $U$ , (32),

$$N = \int_0^\infty f_F(\varepsilon) \mathcal{D}(\varepsilon) d\varepsilon = \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty \varepsilon^{\frac{1}{2}} f_F(\varepsilon) d\varepsilon,$$

---

<sup>13</sup> This can be understood in terms of the equipartition theorem: the energy of a free particle is distributed equally among the three degrees of freedom associated with the three components of momenta. The equipartition theorem of thermodynamics states that each degree of freedom receives an amount of energy  $\frac{1}{2}k_B T$ , so the energy of each particle is  $\frac{3}{2}k_B T$ . The total energy in a gas consisting of  $N$  particles is  $\frac{3}{2}Nk_B T$  and the heat capacity is  $\frac{3}{2}Nk_B$ .

<sup>14</sup> The specific heat is a better representation of the intrinsic characteristics of the metal as it is independent of the size of the crystal, the heat capacity itself depends on how big the crystal is.

from (29). The most important region of the integral is around  $\varepsilon = \mu$ , where the integrand fall off steeply (see last figure), and we can focus on this region by integrating by parts:

$$N = \frac{2}{3} \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \left[ \varepsilon^{\frac{3}{2}} f_F(\varepsilon) \right]_0^\infty - \frac{2}{3} \frac{V}{2\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty \varepsilon^{\frac{3}{2}} \frac{df_F(\varepsilon)}{d\varepsilon} d\varepsilon.$$

The first term vanishes at both limits while, at least for  $\frac{T}{T_F} \ll 1$ , the second is dominated by the region around  $\varepsilon \approx \mu$ , where  $-\frac{df_F(\varepsilon)}{d\varepsilon}$  is large. Write

$$-\frac{df_F(\varepsilon)}{d\varepsilon} = -\frac{d}{d\varepsilon} \left( e^{(\varepsilon-\mu)/k_B T} + 1 \right)^{-1} = \frac{1}{k_B T} \frac{e^x}{(e^x + 1)^2},$$

where  $x := \frac{\varepsilon-\mu}{k_B T}$ . Now

$$N = \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_0^\infty \varepsilon^{\frac{3}{2}} \frac{e^x}{(e^x + 1)^2} \frac{d\varepsilon}{k_B T} = \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_{-\frac{\mu}{k_B T}}^\infty \varepsilon^{\frac{3}{2}} \frac{e^x}{(e^x + 1)^2} dx,$$

where the integration variable has been changed from  $\varepsilon$  to  $x$  in the last expression. Now we do two things: first observe that for  $x < -\frac{\mu}{k_B T}$ , that is  $\varepsilon < 0$ , the integrand is exponentially suppressed<sup>15</sup> so we can extend the lower limit of integration down to  $-\infty$  and this has a negligible effect on the integral; second we Taylor expand  $\varepsilon^{\frac{3}{2}}$  about  $\varepsilon = \mu$ ,

$$\begin{aligned} \varepsilon^{\frac{3}{2}} &= \mu^{\frac{3}{2}} + (\varepsilon - \mu) \left. \frac{d}{d\varepsilon} \varepsilon^{\frac{3}{2}} \right|_{\varepsilon=\mu} + \frac{1}{2} (\varepsilon - \mu)^2 \left. \frac{d^2}{d\varepsilon^2} \varepsilon^{\frac{3}{2}} \right|_{\varepsilon=\mu} + \dots \\ &= \mu^{\frac{3}{2}} + \frac{3}{2} (\varepsilon - \mu) \mu^{\frac{1}{2}} + \frac{3}{8} (\varepsilon - \mu)^2 \mu^{-\frac{1}{2}} + \dots \end{aligned}$$

This allow us to re-write

$$\begin{aligned} N &= \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_{-\infty}^\infty \varepsilon^{\frac{3}{2}} \frac{e^x}{(e^x + 1)^2} dx \\ &= \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \int_{-\infty}^\infty \frac{e^x}{(e^x + 1)^2} \left( \mu^{\frac{3}{2}} + \frac{3}{2} k_B T x \mu^{\frac{1}{2}} + \frac{3}{8} (k_B T x)^2 \mu^{-\frac{1}{2}} + \dots \right) dx, \end{aligned}$$

since  $\varepsilon - \mu = k_B T x$ .

Each individual integral over  $x$  on the right hand side can now be evaluated:

$$\begin{aligned} \int_{-\infty}^\infty \frac{e^x}{(e^x + 1)^2} dx &= \int_{-\infty}^\infty -\frac{d}{dx} \left( \frac{1}{e^x + 1} \right) dx = - \left[ \frac{1}{e^x + 1} \right]_{-\infty}^\infty = 1, \\ \int_{-\infty}^\infty \frac{x e^x}{(e^x + 1)^2} dx &= \int_{-\infty}^\infty \frac{x}{(e^x + 1)(1 + e^{-x})} dx = 0, \\ \int_{-\infty}^\infty \frac{x^2 e^x}{(e^x + 1)^2} dx &= \int_{-\infty}^\infty \frac{x^2}{(e^x + 1)(e^{-x} + 1)} dx = 2 \int_0^\infty \frac{x^2}{(e^x + 1)(e^{-x} + 1)} dx \\ &= -2 \int_0^\infty x^2 \frac{d}{dx} \left( \frac{1}{e^x + 1} \right) dx = 4 \int_0^\infty \frac{x}{e^x + 1} dx = \frac{\pi^2}{3}. \end{aligned}$$

<sup>15</sup> Putting in some numbers, with  $\mu \approx \varepsilon_F$ ,  $x < -\frac{\mu}{k_B T} \Rightarrow x < -\frac{\varepsilon_F}{k_B T} \approx -100$  and  $e^{-x}$  is tiny.

The second integral vanishes because the integrand is an odd function of  $x$  and the third is left as an exercise.<sup>16</sup> We now have

$$N = \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \mu^{\frac{3}{2}} + \frac{V}{3\pi^2} \left( \frac{2m}{\hbar^2} \right)^{\frac{3}{2}} \frac{\pi^2}{8} (k_B T)^2 \mu^{-\frac{1}{2}} + \dots$$

In terms of the Fermi energy (28) this is

$$\begin{aligned} N &= N \left( \frac{\mu}{\varepsilon_F} \right)^{\frac{3}{2}} + N \frac{\pi^2}{8} \frac{(k_B T)^2}{\varepsilon_F^{\frac{3}{2}} \mu^{\frac{1}{2}}} + \dots \\ \Rightarrow \quad 1 &= \left( \frac{\mu}{\varepsilon_F} \right)^{\frac{3}{2}} + \frac{\pi^2}{8} \left( \frac{k_B T}{\varepsilon_F} \right)^2 \left( \frac{\varepsilon_F}{\mu} \right)^{\frac{1}{2}} + \dots, \end{aligned}$$

where the dots indicate terms of order  $\left( \frac{k_B T}{\varepsilon_F} \right)^4$  and higher. This implies that  $\frac{\mu}{\varepsilon_F} = 1 + o\left( \frac{k_B T}{\varepsilon_F} \right)^2$ , so we can replace the  $\left( \frac{\varepsilon_F}{\mu} \right)^{\frac{1}{2}}$  factor in the second term on the right hand side above with unity, absorbing the difference into the dots, giving

$$\frac{\mu}{\varepsilon_F} = \left[ 1 - \frac{\pi^2}{8} \left( \frac{k_B T}{\varepsilon_F} \right)^2 \right]^{\frac{2}{3}} + \dots = 1 - \frac{2}{3} \frac{\pi^2}{8} \left( \frac{k_B T}{\varepsilon_F} \right)^2 + \dots,$$

giving finally

$$\frac{\mu}{\varepsilon_F} = 1 - \frac{\pi^2}{12} \left( \frac{k_B T}{\varepsilon_F} \right)^2 + o\left( \frac{k_B T}{\varepsilon_F} \right)^4. \quad (36)$$

This an extremely good approximation: remember  $\varepsilon_F = k_B T_F$  typically corresponds to a temperature  $T_F \approx 50,000^\circ K$  so, even at  $T = 500^\circ K$ ,  $\frac{k_B T}{\varepsilon_F}$  is only of order  $10^{-2}$  and  $\left( \frac{k_B T}{\varepsilon_F} \right)^2$  is of order  $10^{-4}$ , so  $\mu = \varepsilon_F$  to one part in 10,000 and the quadratic approximation to  $\mu(T)$  in (36) is accurate to one part in one hundred million.

The integral in (35) can be evaluated using the same techniques. The calculation is left as an exercise and the answer is

$$U = \frac{3}{5} N \varepsilon_F \left( \frac{\mu}{\varepsilon_F} \right)^{\frac{5}{2}} + \frac{3\pi^2}{8} N \varepsilon_F \left( \frac{k_B T}{\varepsilon_F} \right)^2 \left( \frac{\mu}{\varepsilon_F} \right)^{\frac{1}{2}} + \dots$$

Using the expansion (36) then gives

$$U = \frac{3}{5} N \varepsilon_F + \frac{\pi^2}{4} N \varepsilon_F \left( \frac{k_B T}{\varepsilon_F} \right)^2 + \dots$$

---

<sup>16</sup> Hint: write  $\frac{1}{e^x+1} = \frac{e^x}{1+e^{-x}} = e^x \sum_{n=0}^{\infty} (-1)^n e^{-nx}$  and use Gamma functions,  $\Gamma(n) = \int_0^{\infty} x^{(n-1)} e^{-x} dx = (n-1)!$ .

The electronic specific heat is now given by

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = \frac{N\pi^2 k_B^2 T}{2\varepsilon_F} + o\left( \frac{k_B T}{\varepsilon_F} \right)^3.$$

This explains why electronic heat capacities are only about 1% of the classical value (33), they are reduced by the factor  $\frac{\pi^2 k_B T}{3\varepsilon_F} \approx 10^{-2}$  by the exclusion principle. The specific heat is

$$c_V = \frac{N}{V} \frac{\pi^2 k_B^2 T}{2\varepsilon_F} + o\left( \frac{k_B T}{\varepsilon_F} \right)^3 = \frac{m}{\hbar^2} \frac{(n_e)^{\frac{1}{3}} k_B^2 T}{(3\pi^2)^{\frac{2}{3}}} + o\left( \frac{k_B T}{\varepsilon_F} \right)^3, \quad (37)$$

a formula that is accurate to one part in 10,000 for  $T \approx 500^\circ K$ .

The full specific heat of a metallic crystal includes the phonon contribution from the lattice,  $\propto T^3$  at low temperatures and  $3nk_B$  at high temperatures (the Dulong-Petit value).<sup>17</sup> At very low temperatures the electronic specific heat dominates.

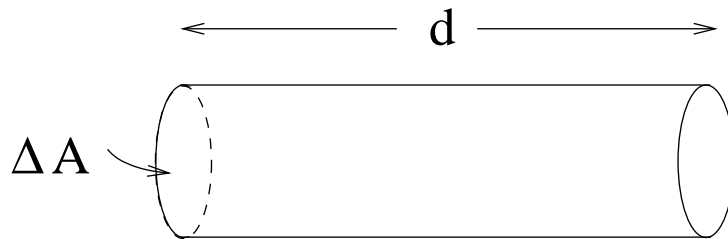
This calculation also has implications for the thermal conductivity  $\kappa$ , since  $\kappa = \frac{c_v v_F^2 \tau}{3}$  and  $c_v$  is reduced by a factor of 1/100 from its naive classical value.

## Electrical conductivity

Consider a segment of metal wire of length  $d$  and constant cross-sectional area  $\Delta A$ . The electrical resistance,  $R$ , depends on the geometry, doubling the length doubles the resistance and thick wires have less resistance than thin wires:  $R$  is proportional to the length and inversely proportional to  $\Delta A$ ,

$$R = \frac{d}{\Delta A} \rho,$$

where  $\rho$  is called the resistivity of the metal (it has dimensions of *Ohms*  $\times$  *length*).  $R$  is not intrinsic to the material but  $\rho$  is, for this reason  $\rho$  is more fundamental than  $R$ .



A voltage  $V$  applied along the wire will generate a current  $I$  proportional to  $\Delta A$ ,

$$I = j\Delta A$$

---

<sup>17</sup> Remember that  $n = \frac{N}{V}$  is the number of primitive lattice cells per unit volume in the crystal, which is not the same as  $n_e = \frac{N}{V}$ , the number of electrons per unit volume, except for crystals with a monatomic basis of monovalent metals.

where  $j$  is called the current density, the current per unit area in the wire. This voltage will give rise to an electric field of magnitude  $E = \frac{V}{d}$ , so Ohm's law  $V = IR$  can be written

$$Ed = (j\Delta A) \left( \frac{\rho d}{\Delta A} \right) \Leftrightarrow E = \rho j.$$

Physicists prefer to work with the conductivity, defined to be the inverse of the resistivity,  $\sigma = 1/\rho$ , and write  $j = \sigma E$ . This should really be written as a vector equation, currents have a direction associated with them so  $j$  is a vector, as is  $E$ . For an isotropic conductor

$$\mathbf{j} = \sigma \mathbf{E}.$$

It is a difficult, but important, task to calculate  $\sigma$  for any given material, metal or semi-conductor, from a knowledge of the microscopic structure at the atomic level. To get some understanding of the underlying physics, consider the force acting on a charge  $-e$ , such as an electron, due to an applied electric field  $\mathbf{E}$ : Newton's second law implies

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = -e\mathbf{E}.$$

If  $\mathbf{E}$  is constant this immediately integrates to

$$\mathbf{p}(t) - \mathbf{p}(0) = -e\mathbf{E}t,$$

where  $\mathbf{p} = m\mathbf{v}$  is the electron's momentum. In a perfectly pure crystal the electron would accelerate indefinitely but in a real crystal this acceleration will be impeded by collisions with phonons (at room temperature) or imperfections in the crystal (more important at low temperatures, when phonons are scarce). Suppose the electron is stopped in its tracks by these collisions and denote the average time between collisions by  $\tau$ , then the average electron velocity will be  $\mathbf{v} = \mathbf{a}\tau = -\frac{e}{m}\mathbf{E}\tau$ . If the electron number density is  $n$  then the current density is proportional to  $n$ ,

$$\mathbf{j} = -en\mathbf{v} = \frac{n_e e^2 \tau}{m} \mathbf{E} = \sigma \mathbf{E},$$

from which we derive the conductivity in terms of  $\tau$ ,

$$\sigma = \frac{n_e e^2 \tau}{m}. \tag{38}$$

This is called the **Drude** formula. It is useful in that it gives an intuitive understanding of what affects the conductivity, many rapid collisions means  $\tau$  is small and the conductivity is low, very few collisions means  $\tau$  is a long time and the conductivity is high.  $\tau$  is in

fact a function of temperature, at room temperature  $\tau \propto 1/T$ . The conductivity is also proportional to the electron density, which is reasonable — more electron means more current. A quantity which reflects how easily the electrons can move, without reference to the number of electrons is the **mobility**,  $\mu$ , defined by<sup>18</sup>

$$\mathbf{v} = \mu \mathbf{E} \quad \Leftrightarrow \quad \mu = -\frac{\sigma}{n_e e}.$$

The minus sign is because  $e < 0$  for an electron.

### Wiedemann-Franz law

We can now derive a relation between the electrical and thermal conductivities in a metal, which follows from the fact that it is mobile electrons that are responsible for both. Using our expression for thermal conductivity (24) and the electron specific heat (37),  $c_v = \frac{\pi^2}{2} \frac{k_B^2}{\varepsilon_F} T n_e$ , where the Fermi energy  $\varepsilon_F$  is related to the Fermi velocity  $v_F$  by  $\varepsilon_F = \frac{1}{2} m v_F^2$ , gives  $\kappa = \frac{c_v v_F^2 \tau}{3} = \frac{\pi^2 n}{3m} k_B T \tau$ , we have

$$\frac{\kappa}{\sigma} = \frac{\pi^2 k_B^2}{3e^2} T.$$

The ratio of thermal to electrical conductivities in a metal is proportional to the absolute temperature of the crystal, with a co-efficient given by known physical constants. Historically this relation, known as the Wiedemann-Franz law, was discovered before it was understood why specific heats in metals were so low, 1% of their expected classical value — the thermal conductivity is also low for the same reason, Fermi blocking.

### Hall effect

The Hall effect occurs in very thin slabs of conducting material placed in a transverse magnetic field when a current is passed through the slab. Discovered in 1879 it was a very important step in understanding the nature of electric currents.

Consider the effect on the analysis above of including a magnetic field. The force on the electron is given by the Lorentz force law,

$$\mathbf{F} = m \dot{\mathbf{v}} = -e(\mathbf{E} + (\mathbf{v} \times \mathbf{B})).$$

For a constant magnetic field in the  $z$ -direction,  $\mathbf{B} = B \hat{\mathbf{z}}$ , and a constant electric field perpendicular to  $\mathbf{B}$ ,  $\mathbf{E} = E \hat{\mathbf{x}}$ , this gives

$$\begin{aligned} m \dot{v}_x &= -eE - e v_y B \\ m \dot{v}_y &= e v_x B \\ m \dot{v}_z &= 0. \end{aligned}$$

---

<sup>18</sup> Not to be confused with the chemical potential.

Combining the first two of these equations implies

$$\ddot{v}_x = -\frac{e^2 B^2}{m^2} v_x,$$

which is the harmonic oscillator equation for  $v_x$  with frequency  $\omega_c = \frac{eB}{m}$ , called the **cyclotron frequency**. This is a characteristic frequency for a charged particle moving in a magnetic field. Solving this equation gives

$$v_x(t) = v_0 \cos(\omega_c t)$$

where  $v_0$  is a constant and, using this in the  $v_y$  equation above, gives

$$\dot{v}_y(t) = v_0 \omega_c \cos(\omega_c t) \quad \Rightarrow \quad v_y(t) = v_0 \sin(\omega_c t) + a,$$

with  $a$  a constant of integration. Then putting this form of  $v_y(t)$  into the equation for  $\dot{v}_x$  above yields

$$m\dot{v}_x = -m\omega_c v_0 \sin(\omega_c t) = -eE - eB(v_0 \sin(\omega_c t) + a),$$

which fixes  $a$  to be the ratio  $a = -\frac{E}{B}$ . Finally  $\dot{v}_z = 0$  implies that  $v_z$  is a constant and the motion of the electron is given by

$$\mathbf{v}(t) = v_0 (\cos(\omega_c t) \hat{\mathbf{x}} + \sin(\omega_c t) \hat{\mathbf{y}}) - \frac{E}{B} \hat{\mathbf{y}} + v_z \hat{\mathbf{z}}$$

which integrates to

$$\mathbf{r}(t) = \frac{v_0}{\omega_c} (\sin(\omega_c t) \hat{\mathbf{x}} - \cos(\omega_c t) \hat{\mathbf{y}}) - \frac{E}{B} t \hat{\mathbf{y}} + v_z t \hat{\mathbf{z}},$$

where we have chosen the origin so that  $\mathbf{r}(0) = \mathbf{0}$ . If otherwise unhindered the electron performs a circular motion in the  $x - y$  plane superimposed on a constant drift in the direction  $v_z \hat{\mathbf{z}} - \frac{E}{B} \hat{\mathbf{y}}$ .

We can now include the effect of collisions by adding a term  $\mathbf{F}_c = -\frac{m}{\tau} \mathbf{v}$  to the Lorentz force,

$$m\ddot{\mathbf{v}} = -\frac{m}{\tau} \mathbf{v} - e(\mathbf{E} + (\mathbf{v} \times \mathbf{B})).$$

Again with  $\mathbf{B} = B\hat{\mathbf{z}}$  in the  $z$ -direction, but with a more general  $\mathbf{E}$ , this can be written in components as

$$\begin{aligned} m\dot{v}_x &= -e(E_x + v_y B) - \frac{mv_x}{\tau} \\ m\dot{v}_y &= -e(E_y - v_x B) - \frac{mv_y}{\tau} \\ m\dot{v}_z &= -eE_z - \frac{mv_z}{\tau}. \end{aligned}$$

Rather than solving these equations in complete generality, we just look for a steady state solution with  $\dot{\mathbf{v}} = 0$ , which will be sufficient for a discussion of DC currents. Such a solution is

$$\begin{aligned}v_x &= -\frac{\tau e}{m}E_x - \omega_c \tau v_y \\v_y &= -\frac{\tau e}{m}E_y + \omega_c \tau v_x \\v_z &= -\frac{\tau e}{m}E_z.\end{aligned}$$

Now focus on a thin slab of conducting material in the  $x - y$  plane. The electrons are confined to the plane, so we shall assume  $v_z = 0$  and  $E_z = 0$ . Consider a rectangular slab with its edges aligned in the  $x$  and  $y$  directions. If a current is passed through the slab in the  $x$ -direction, then  $v_y = 0$  as well and

$$E_x = -\frac{mv_x}{\tau e}, \quad E_y = \omega_c \frac{m}{e}v_x = -\omega_c \tau E_x$$

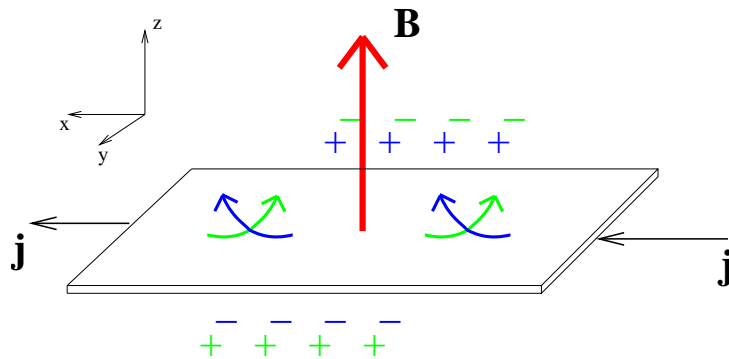
so

$$E_y = -\frac{eB}{m}\tau E_x.$$

The current density is the charge passing unit area in unit time, which is  $-e$  times the number of electrons passing unit area in unit time. The latter is  $n_e \mathbf{v}$ , where  $n_e$  is the number of electrons per unit volume, so

$$\mathbf{j} = -en_e \mathbf{v} = \frac{n_e e^2 \tau}{m} E_x \hat{\mathbf{x}}.$$

From this we see that the conductivity  $\sigma = \frac{n_e e^2 \tau}{m}$  is just the same as in the  $B = 0$  case above, but now we no longer have the vector relation  $\mathbf{j} \propto \mathbf{E}$ , because  $E_y \neq 0$  and there is no component of the current in the  $y$ -direction. What is happening here is illustrated below, where the green arrows indicate what the electron trajectories would be if  $E_y$  were zero:



As the electrons are forced through the slab by  $E_x$  the Lorentz force pushes them in the  $-y$  direction and their trajectories are bend toward the edge of the sample, but they cannot escape the confines of the slab so a negative charge builds along the top edge of the slab. At the same time electrons are depleted from the lower edge. This charge generates

a voltage in the  $y$ -direction which builds up until there is a Coulomb force in the positive  $y$ -direction ( $E_y \neq 0$ ) which exactly cancels the Lorentz force due to the magnetic field and the electrons just move in straight lines in the  $x$ -direction. This voltage is called the **Hall voltage**. Note that the Hall voltage would have the opposite sign if the particles carrying the current had a positive electric charge (blue arrows above). By measuring the sign of the Hall voltage we can tell that electrons in metals carry a negative charge.

We have

$$j_x = \frac{n_e e^2 \tau}{m} E_x = -\frac{n_e e}{B} E_y \quad \Rightarrow \quad E_y = R_H B j_x$$

where

$$R_h = -\frac{1}{n_e e}$$

is intrinsic to the material and is called the **Hall co-efficient**. The sign of the Hall co-efficient depends on the sign of the electric charge on the charge carriers (it is negative for electrons) and we can also obtain  $n_e$  directly by measuring  $R_H$ . For example silver has  $R_H = -9 \times 10^{-11} \text{ m}^3 \text{ C}^{-1}$  from which  $n_e = 7 \times 10^{28} \text{ m}^{-3}$ . Some materials do have positive Hall co-efficients, for example aluminium has  $R_H = 1.0 \times 10^{-10} \text{ m}^3 \text{ C}^{-1}$  and so behaves as though the current is being carried by *positive* charges. This apparently anomalous behaviour is explained below in terms of *energy bands*.

## Energy Bands and Bloch's Theorem

The crystalline structure of metals puts strong constraints on the form of the electron wave-functions. Again, for simplicity, we consider one-dimension first. Model the motion of an electron moving in a one-dimensional crystal by demanding that the electron moves in a potential that is periodic with period  $a$ ,  $U(x) = U(x + a)$ . Denote the electron wave-function by  $\psi(x)$ . The electron density  $n(x) \propto |\psi(x)|^2$  is also periodic,  $n(x) = n(x + a)$ , so  $|\psi(x)|^2 = |\psi(x + a)|^2$ , but this does not necessarily mean that  $\psi(x)$  itself is periodic with period  $a$ , only that it is periodic up to a phase  $\psi(x + a) = e^{i\phi} \psi(x)$ . However we can impose periodic boundary conditions in  $\psi$  over the whole crystal,  $\psi(x + \mathcal{N}a) = \psi(x)$  where  $\mathcal{N}a$  is the size of the crystal (for large  $\mathcal{N}$  the boundary conditions do not affect the behaviour in the interior of the crystal very much).

We now argue that  $\phi$  is independent of position and can only have a discrete set of possible values. The Schrödinger equation for a state with energy  $E$  is

$$-\frac{\hbar^2}{2m} \psi''(x) + U(x) \psi(x) = E \psi(x), \quad (39)$$

and lattice periodicity requires

$$\begin{aligned} &-\frac{\hbar^2}{2m} \psi''(x + a) + U(x + a) \psi(x + a) = E \psi(x + a), \\ \Rightarrow &-\frac{\hbar^2}{2m} \psi''(x + a) + U(x) \psi(x + a) = E \psi(x + a). \end{aligned}$$

In particular  $\psi(x)$  and  $\psi(x+a)$  satisfy the same second order ODE with the same boundary conditions, hence they are linearly dependent (in general the real and imaginary parts of  $\psi(x)$  are linearly independent). Thus  $\psi(x+a) = c\psi(x)$  with  $c$  a (possibly complex) constant. Since  $\psi(x+\mathcal{N}a) = \psi(x)$  we conclude that  $c^\mathcal{N} = 1$ , hence  $c = e^{i\phi} = e^{\frac{2\pi ij}{\mathcal{N}}}$  where  $j = 0, \dots, \mathcal{N}-1$  is an integer with  $\mathcal{N}$  possible values (or, if  $\mathcal{N}$  is even, we can use  $j = -\frac{\mathcal{N}}{2}, \dots, \frac{\mathcal{N}}{2}$ ). The integer  $j$  labels different solutions of the Schrödinger equation: let  $k = \frac{2\pi j}{\mathcal{N}a}$  then we denote the eigenfunction associate with any particular choice of  $j$  by  $\psi_k(x)$ . These eigenfunctions have the property that

$$\psi_k(x+a) = e^{ik a} \psi_k(x). \quad (40)$$

$k$  is a wave-vector and if  $j = -\frac{\mathcal{N}}{2}, \dots, \frac{\mathcal{N}}{2}$  then  $-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}$  and  $k$  takes  $\mathcal{N}$  discrete values in the first Brillouin zone of the crystal. For large  $\mathcal{N}$  the allowed values of  $k$  are crowded very close together and approximate a continuum as  $\mathcal{N} \rightarrow \infty$ , but they never stray outside the first Brillouin zone.

An example of a function with the property (40) is  $e^{ikx}$ , though this is not a solution of Schrödinger's equation above unless  $U = 0$ , in which case  $E = \frac{\hbar^2 k^2}{2m}$  and  $\hbar k$  has the physical interpretation of being the electron's momentum.

It is not possible to solve the Schrödinger equation (39) in closed form for a general periodic potential  $U(x)$ . Nevertheless the assumed periodicity of the potential allows a simplification and we shall prove a theorem, called **Bloch's theorem**, that the eigenfunctions can always be written in the form<sup>19</sup>

$$\boxed{\psi_k(x) = e^{ikx} B_k(x)}. \quad (41)$$

where  $k$  lies in the first Brillouin zone and  $B_k(x) = B_k(x+a)$  is a periodic function. This automatically satisfies (40).

The proof of Bloch's theorem is instructive as it sheds light on the general structure of the energy spectrum of electrons in a crystal. Let  $\psi(x)$  be an eigenfunction of the Schrödinger equation with energy  $E$ ,

$$\hat{H}\psi(x) = E\psi(x),$$

where the Hamiltonian operator is  $\hat{H} = -\frac{\hbar^2}{2m} \frac{d}{dx^2} + U(x)$ . Any  $\psi(x)$  can be expanded as a sum of plane waves

$$\psi(x) = \sum_q b(q) e^{iqx} = \sum_{s=-\infty}^{\infty} b_s e^{\frac{2\pi i s}{\mathcal{N}a} x}, \quad (42)$$

where the co-efficients  $b(q) = b_s$  are complex constants. With periodic boundary conditions  $\psi(x+\mathcal{N}a) = \psi(x)$ , the allowed values of  $q = \frac{2\pi s}{\mathcal{N}a}$  are necessarily discrete with  $s$  taking on

---

<sup>19</sup> Strictly speaking the one-dimensional version quoted here is **Floquet's theorem**. Bloch proved the three-dimensional version quoted later.

all integer values between  $-\infty$  and  $\infty$ . Unlike phonons there is no bound on the electron momentum, electrons can have wavelengths much shorter than the lattice spacing.

In contrast the potential  $U(x)$  is periodic with period  $a$ ,  $U(x+a) = U(x)$ , so its Fourier decomposition involves only reciprocal lattice vectors  $G = \frac{2\pi h}{a}$ ,

$$U(x) = \sum_{G'} \tilde{U}(G') e^{iG'x} = \sum_{h'=-\infty}^{\infty} \tilde{U}_{h'} e^{\frac{2\pi i h'}{a} x}.$$

Let  $E^{(0)}(q) = \frac{\hbar^2 q^2}{2m}$  be the energy of a free electron with wave-vector  $q$  (*i.e.* the energy for  $U = 0$ ).

The Schrödinger's equation can be written

$$\begin{aligned} & \sum_q E^{(0)}(q) b(q) e^{iqx} + \sum_{G', q'} \tilde{U}(G') b(q') e^{i(q'+G')x} = E \sum_q b(q) e^{iqx} \\ \Rightarrow & \sum_q (E^{(0)}(q) - E) b(q) e^{iqx} + \sum_{G', q} \tilde{U}(G') b(q - G') e^{iqx} = 0, \quad \text{where } q = q' + G', \\ \Rightarrow & (E^{(0)}(q) - E) b(q) + \sum_{G'} \tilde{U}(G') b(q - G') = 0. \end{aligned} \quad (43)$$

Now  $\tilde{U}(G')$  are given complex numbers, they are determined by  $U(x)$ , so this is a set of linear equations for the unknown numbers  $b(q)$ . Up till now  $q = \frac{2\pi s}{Na}$  and  $s$  could have any integral value but it is often convenient to add a reciprocal lattice vector to  $q$  so as to force it into the first Brillouin zone: for any  $q$  choose  $G = \frac{2\pi h}{a}$  so that  $k = q - G$  lies in the first Brillouin zone (for a given  $q$  this is a unique decomposition, there is only one  $k$  in the first Brillouin zone and one reciprocal lattice vector  $G$  that can satisfy this). To emphasise the distinction between the wavevector  $k$ , which lies in the first Brillouin zone, and  $G$ , which is a reciprocal lattice vector, we write  $b(k + G) = b_k(G)$ . With this notation (43) reads

$$\{E^{(0)}(k + G) - E\} b_k(G) + \sum_{G'} \tilde{U}(G') b_k(G - G') = 0.$$

Then, with  $G'' = G - G'$ , we have

$$\boxed{\{E^{(0)}(k + G) - E\} b_k(G) + \sum_{G''} \tilde{U}(G - G'') b_k(G'') = 0.} \quad (44)$$

This is called the **central equation**, though it is actually a set of coupled linear equations for the  $b_k(G)$ . Alternatively it can be written, using  $G = \frac{2\pi h}{a}$  and the notation  $\tilde{U}(G) = \tilde{U}_h$ ,<sup>20</sup>

$$\left\{ E^{(0)} \left( k + \frac{2\pi h}{a} \right) - E \right\} b_k(h) + \sum_{h''} \tilde{U}_{h-h''} b_k(h'') = 0. \quad (45)$$

---

<sup>20</sup> With a slight abuse of notation we also write  $b_k(G) = b_k(h)$  where  $G = \frac{2\pi h}{a}$ .

The sum in (45) is over the integers and this form shows that the equation can be written as an infinite matrix equation

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & E^{(0)}(k + \frac{2\pi}{a}) - E(k) & \tilde{U}_1 & \tilde{U}_2 & \cdots & \cdots \\ \cdots & \tilde{U}_{-1} & E^{(0)}(k) - E(k) & \tilde{U}_1 & \cdots & \cdots \\ \cdots & \tilde{U}_{-2} & \tilde{U}_{-1} & E^{(0)}(k - \frac{2\pi}{a}) - E(k) & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots \end{pmatrix} \begin{pmatrix} \vdots \\ b_k(1) \\ b_k(0) \\ b_k(-1) \\ \vdots \end{pmatrix} = 0. \quad (46)$$

Remember  $\tilde{U}_{-h} = \tilde{U}_h^*$  for a real potential. For simplicity we have chosen to shift the potential by a constant so that  $U_0 = 0$ , this simply adds an overall constant to the energy and does not change any physics. In principle this is an infinite matrix equation, but in practice we can cut  $|h|$  off at some large but finite value with negligible error. Physically, if the momentum of the electron gets too large it will escape from the crystal anyway and the whole description breaks down.

The eigenvalues are determined by requiring that the matrix has zero determinant,

so as to ensure that a non-zero solution, a non-zero eigenvector  $\begin{pmatrix} \vdots \\ b_k(1) \\ b_k(0) \\ b_k(-1) \\ \vdots \end{pmatrix}$ , exists. For a

given potential, and hence given Fourier co-efficients  $\tilde{U}_h$ , calculating the energy eigenvalues and their associated eigenvectors still involves calculating the determinant of a very large matrix. Fortunately it is often the case that many of the off-diagonal components are very small, in real situations it is usually the case that  $|\tilde{U}_1| \gg |\tilde{U}_2| \gg |\tilde{U}_3| \gg \cdots$ , but before looking at some explicitly solvable cases we pause to prove Bloch's theorem.

The structure of (46) shows that there are energy eigenfunctions associated with any given value of  $k$  (indeed there are many). From (42) and the above analysis we can write an energy eigenfunction associated with any particular value of  $k$  as

$$\psi_k(x) = \sum_{h=-\infty}^{\infty} b_k(h) e^{i(k + \frac{2\pi h}{a})x} = e^{ikx} \sum_{h=-\infty}^{\infty} b_k(h) e^{\frac{2\pi i h}{a}x} := e^{ikx} B_k(x),$$

where the **Bloch function**  $B_k(x) = \sum_{h=-\infty}^{\infty} b_k(h) e^{\frac{2\pi i h}{a}x}$  is, by construction, periodic  $B_k(x + a) = B_k(x)$ . This is Bloch's theorem, (41).

Of course, since  $B_k(x)$  is periodic, it has a Fourier expansion

$$B_k(x) = \sum_G \tilde{B}_k(G) e^{iGx},$$

with  $G = \frac{2\pi h}{a}$  and we see that the wave-function co-efficients,  $b_k(G)$  are the Fourier co-efficients of the Bloch function.

Note that  $\hbar k$  is *not* the electron momentum, that would correspond to the eigenvalue of the operator  $\hat{p} = -i\hbar \frac{d}{dx}$  and  $\psi_k(x)$  is not an eigenfunction of  $\hat{p}$  in general (unless  $U(x) = 0$ ). Indeed  $-i\hbar \frac{d\psi_k(x)}{dx} \neq \hbar k \psi_k(x)$  unless  $B_k(x)$  is independent of  $x$ . In a crystal electron energy eigenfunctions do not have a specific momentum, a crystal does not have translational invariance under infinitesimal translations, it is only invariant under finite lattice transformations, so momentum is not conserved. The Hamiltonian  $\hat{H}$  does not commute with the momentum operator  $\hat{p}$ , so these two operators cannot be simultaneously diagonalised — an eigenstate of the Hamiltonian cannot simultaneously be an eigenstate of momentum, it must be a linear combination of different momenta states. Nevertheless  $\hbar k$  is a momentum of sorts, it is called the **crystal momentum**, but it is not equal to the electron's momentum in general. We shall give a physical interpretation of the crystal momentum later.

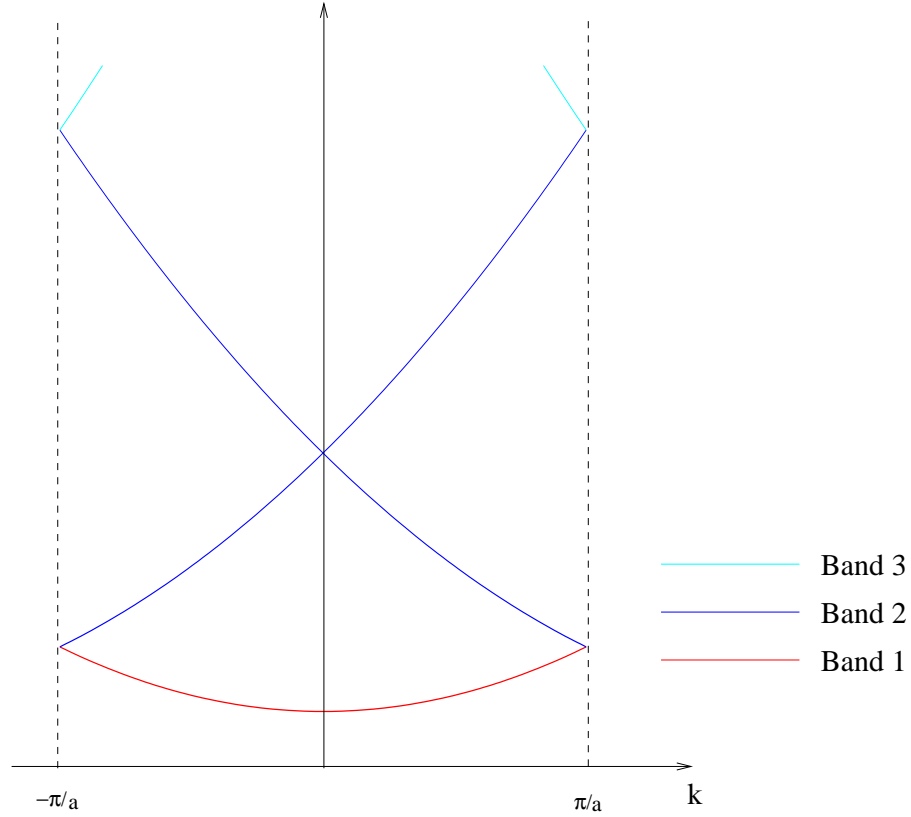
## Free Electron Approximation

To get some feeling for the behaviour of the solution of (45) consider the crudest possible approximation, just setting all the  $\tilde{U}_h$  to zero. That is  $U(x) = 0$  in the Schrödinger equation and we are dealing with free electrons. To solve (45) choose a specific value of  $h$  and  $k$  and set

$$E = E^{(0)} \left( k + \frac{2\pi h}{a} \right)^2 = \frac{\hbar^2}{2m} \left( k + \frac{2\pi h}{a} \right)^2,$$

and the eigenvectors for a specific choice of  $h$  are  $b_k(h) \neq 0$  with all other  $b_k$  zero. Remember that  $k$  is restricted to lie in the first Brillouin zone  $-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}$ . The energy spectrum is shown below, the energy is a multivalued function of  $k$  and there are different bands of energy labelled by  $n = |h|$ : band one is  $n = 0$ , band two is  $n = \pm 1$ , band three is  $n = \pm 2$  etc. The allowed energy in each band is a function of  $k$  and the band is identified with a subscript called the band index,  $E_n(k)$ .

The energy spectrum is the same as that of a single parabola,  $E = \frac{\hbar q^2}{2m}$  with  $-\infty < q < \infty$ , but it is represented as different pieces all with  $k = q + \frac{2\pi h}{a}$  and  $h$  chosen so that  $-\frac{\pi}{a} \leq k \leq \frac{\pi}{a}$ . This trick, of knocking  $q$  into the first Brillouin zone by adding a reciprocal lattice vector to it, is called the **reduced zone scheme**. It is completely equivalent to a single energy band with  $E = \frac{\hbar q^2}{2m}$  and  $-\infty < q < \infty$ , which is called the **extended zone scheme**. It is a matter of taste which description is used, though the reduced zone scheme is often more convenient.



In three-dimensions the story is essentially the same, but the notation gets a little messier. Solutions of the Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{x}) + U(\mathbf{x})\psi(\mathbf{x}) = E\psi(\mathbf{x}), \quad (47)$$

can be expanded in plane waves as

$$\psi(x) = \sum_{\mathbf{q}} b(\mathbf{q})e^{i\mathbf{q}\cdot\mathbf{x}},$$

with constant co-efficients  $b(\mathbf{q})$ , and the sum is over three-dimensional wave-vectors  $\mathbf{q}$ . We can impose periodic boundary conditions in three dimensions,  $\psi(\mathbf{x} + \mathcal{N}_i\mathbf{a}_i) = \psi(\mathbf{x})$ , where  $i = 1, 2, 3$  and  $\mathcal{N}_i$  are three large integers and  $\mathbf{a}_i$  are primitive lattice vectors — this is a natural generalisation of periodic boundary conditions in one-dimension and are known as Born - von Karman boundary conditions. In a crystal of volume  $V = \mathcal{N}_1\mathcal{N}_2\mathcal{N}_3V_c = \mathcal{N}V_c$ , where  $V_c$  is the volume of a primitive cell. These boundary conditions make the allowed values of  $\mathbf{q}$  a discrete set, though the allowed values get more and more dense and closer together as  $\mathcal{N} \rightarrow \infty$ .

Let  $\psi(\mathbf{x})$  be an eigenfunction of the Schrödinger equation (47) with energy  $E$ . The symmetries of the lattice dictate that  $U(\mathbf{x})$  is periodic in all three primitive lattice vectors,  $U(\mathbf{x}) = U(\mathbf{x} + \mathbf{a}_1) = U(\mathbf{x} + \mathbf{a}_2) = U(\mathbf{x} + \mathbf{a}_3)$ , so it can be decomposed into Fourier modes

$$U(\mathbf{x}) = \sum_{\mathbf{G}'} \tilde{U}(\mathbf{G}')e^{i\mathbf{G}'\cdot\mathbf{x}}$$

where the sum is over all reciprocal lattice vectors and the Fourier co-efficients,  $\tilde{U}(\mathbf{G}')$ , are complex numbers. Let  $E^{(0)}(\mathbf{q}) = \frac{\hbar^2 \mathbf{q} \cdot \mathbf{q}}{2m}$  be the energy of a free electron with wave-vector  $\mathbf{q}$  (*i.e.* the energy for  $U = 0$ ).

The Schrödinger equation (47) can be written

$$\begin{aligned} & \sum_{\mathbf{q}} E^{(0)}(\mathbf{q})b(\mathbf{q})e^{i\mathbf{q} \cdot \mathbf{x}} + \sum_{\mathbf{G}', \mathbf{q}'} \tilde{U}(\mathbf{G}')b(\mathbf{q}')e^{i(\mathbf{q}'+\mathbf{G}') \cdot \mathbf{x}} = E \sum_{\mathbf{q}} b(\mathbf{q})e^{i\mathbf{q} \cdot \mathbf{x}} \\ \Rightarrow & \sum_{\mathbf{q}} (E^{(0)}(\mathbf{q}) - E)b(\mathbf{q})e^{i\mathbf{q} \cdot \mathbf{x}} + \sum_{\mathbf{G}', \mathbf{q}} \tilde{U}(\mathbf{G}')b(\mathbf{q} - \mathbf{G}')e^{i\mathbf{q} \cdot \mathbf{x}}, = 0 \quad \text{where } \mathbf{q} = \mathbf{q}' + \mathbf{G}', \\ & \Rightarrow (E^{(0)}(\mathbf{q}) - E)b(\mathbf{q}) + \sum_{\mathbf{G}'} \tilde{U}(\mathbf{G}')b(\mathbf{q} - \mathbf{G}') = 0. \end{aligned} \quad (48)$$

Now  $\tilde{U}(\mathbf{G}')$  are given complex numbers, they are determined by  $U(\mathbf{x})$ , so this is a set of linear equations for the unknown co-efficients  $b(\mathbf{q})$ . Up till now  $\mathbf{q}$  could be arbitrarily large but, as in the one-dimensional case, it is often convenient to add to it a reciprocal lattice vector so as to force it into the first Brillouin zone: for any  $\mathbf{q}$  in our discrete set choose  $\mathbf{G}$  so that  $\mathbf{k} = \mathbf{q} - \mathbf{G}$  lies in the first Brillouin zone (for a given  $\mathbf{q}$  this is a unique decomposition, there is only one  $\mathbf{k}$  in the first Brillouin zone and one reciprocal lattice vector  $\mathbf{G}$  that can satisfy this). Again to emphasise the distinction between the wave-vector  $\mathbf{k}$ , which is a wave-vector in the first Brillouin zone, and  $\mathbf{G}$ , which is a reciprocal lattice vector, we write  $b_{\mathbf{k}+\mathbf{G}} = b_{\mathbf{k}}(\mathbf{G})$  and write (48) as

$$\left\{ E^{(0)}(\mathbf{k} + \mathbf{G}) - E \right\} b_{\mathbf{k}}(\mathbf{G}) + \sum_{\mathbf{G}'} \tilde{U}(\mathbf{G}')b_{\mathbf{k}}(\mathbf{G} - \mathbf{G}') = 0.$$

Then, with  $\mathbf{G}'' = \mathbf{G} - \mathbf{G}'$ , we have

$$\boxed{\left\{ (E^{(0)}(\mathbf{k} + \mathbf{G}) - E) \right\} b_{\mathbf{k}}(\mathbf{G}) + \sum_{\mathbf{G}''} \tilde{U}(\mathbf{G} - \mathbf{G}'')b_{\mathbf{k}}(\mathbf{G}'') = 0.} \quad (49)$$

This is the three-dimensional **central equation**, the three-dimensional analogue of (44).

The situation in three-dimensions is more complicated, even for free electrons. Then the reciprocal lattice vectors are labelled by three integers  $h$ ,  $k$  and  $l$  (there is a clash of notation here,  $k$  is an integer in the notation  $(hkl)$  denoting a reciprocal lattice vector — this should not be confused with the length of the wave-vector  $|\mathbf{k}|$ , hopefully it will be clear which is meant from the context). For a simple cubic lattice, for example,

$$\mathbf{G}_{hkl} = \frac{2\pi}{a}(h\hat{\mathbf{x}} + k\hat{\mathbf{y}} + l\hat{\mathbf{z}})$$

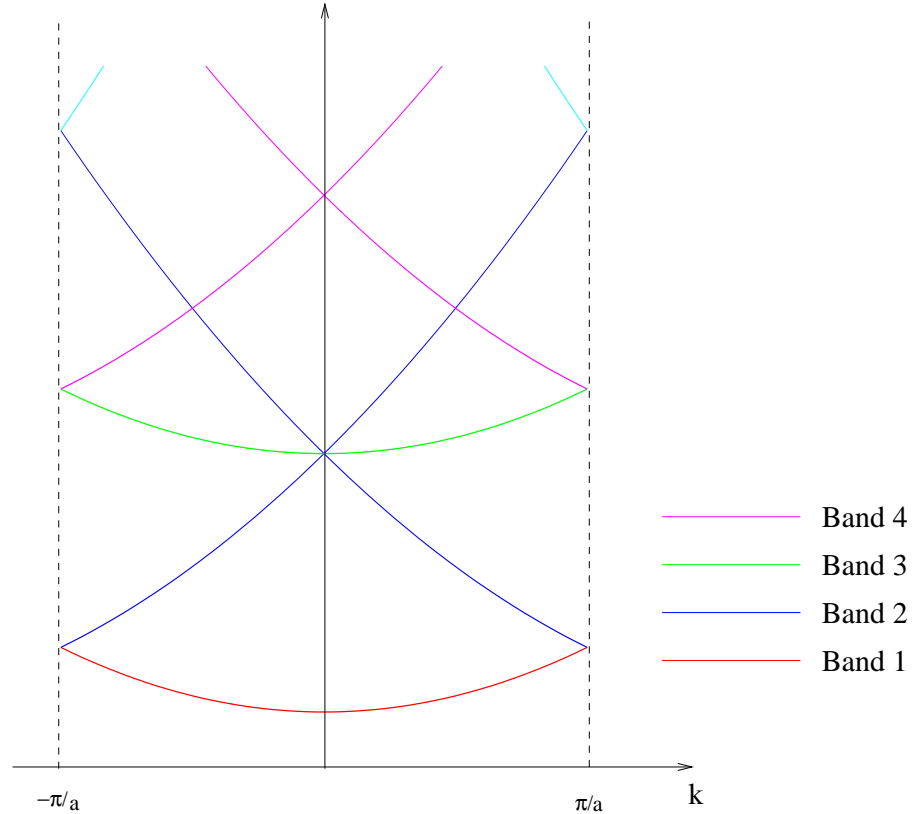
and for a free electron,  $U(\mathbf{x}) = 0$ , the allowed energies are

$$\begin{aligned} E(\mathbf{k}) &= E_{\mathbf{G}}(\mathbf{k}) = \frac{\hbar^2}{2m}(\mathbf{k} + \mathbf{G})^2 \\ &= \frac{\hbar^2}{2m} \left\{ (k_x + G_x)^2 + (k_y + G_y)^2 + (k_z + G_z)^2 \right\} \\ &= \frac{\hbar^2}{2m} \left\{ \left( k_x + \frac{2\pi h}{a} \right)^2 + \left( k_y + \frac{2\pi k}{a} \right)^2 + \left( k_z + \frac{2\pi l}{a} \right)^2 \right\}. \end{aligned}$$

The first few energy bands are listed in the table below, for wave-vectors with  $k_y = k_z = 0$ :

hkl	$\frac{2m}{\hbar^2} E_{hkl}(0, 0, 0)$	$\frac{2m}{\hbar^2} E_{hkl}(k_x, 0, 0)$
000	0	$k_x^2$
100, $\bar{1}00$	$\left(\frac{2\pi}{a}\right)^2$	$\left(k_x + \frac{2\pi}{a}\right)^2$
010, $0\bar{1}0$ , 001, $00\bar{1}$	$\left(\frac{2\pi}{a}\right)^2$	$k_x^2 + \left(\frac{2\pi}{a}\right)^2$

These energy bands are shown graphically below, in the reduced zone scheme,



You can amuse yourself by including  $k_y$  and  $k_z$  or by constructing the energy band structure for free electrons in other crystal structures such as FCC or BCC.

Now we shall go beyond the free electron approximation and include the effects of a non-zero potential  $U(\mathbf{r})$ . For simplicity consider a one-dimensional crystal with central equation (46) and focus on the  $2 \times 2$  sub-matrix obtained by restricting to  $h = 0$  and  $h = -1$ ,

$$\begin{pmatrix} E^{(0)}(k) - E & \tilde{U}_1 \\ \tilde{U}_{-1} & E^{(0)}\left(k - \frac{2\pi}{a}\right) - E \end{pmatrix} \begin{pmatrix} b_k(0) \\ b_k(-1) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (50).$$

For a non-trivial solution of the central equation, with  $b_k(0)$  and  $b_k(-1)$  not both zero, the determinant of the  $2 \times 2$  matrix must vanish, giving

$$E^2 - \left\{ E^{(0)}(k) + E^{(0)}\left(k - \frac{2\pi}{a}\right) \right\} E + E^{(0)}(k)E^{(0)}\left(k - \frac{2\pi}{a}\right) - |\tilde{U}_1|^2 = 0$$

(remember  $\tilde{U}_{-1} = \tilde{U}_1^*$  for a real potential). Solving for the eigenvalues  $E$  gives  $E(k)$ , a dispersion relation. Using  $E^{(0)}(k) = \frac{\hbar^2}{2m}k^2$ , there are two possibilities,

$$E(k) = \frac{\hbar^2}{m} \left\{ \frac{k^2}{2} + \frac{\pi}{a} \left( \frac{\pi}{a} - k \right) \pm \frac{\pi}{a} \sqrt{\left( k - \frac{\pi}{a} \right)^2 + \left( \frac{ma}{\pi\hbar^2} \right)^2 |\tilde{U}_1|^2} \right\}. \quad (51)$$

A consequence of this  $2 \times 2$  approximation (50) to the full central equation (46) is that the latter is clearly symmetric under  $k \rightarrow -k$ , while (50) is not. The true energy should be even an function of  $k$ , but (51) is not. We shall remedy this defect later but, for the moment, just concentrate on  $0 \leq k \leq \frac{\pi}{a}$ .

If  $0 < |\tilde{U}_1| \ll \frac{\pi^2\hbar^2}{a^2m}$ , then the second term under the square root in (51) is small relative to the first, unless  $k$  is close to  $\frac{\pi}{a}$ , and the two roots are

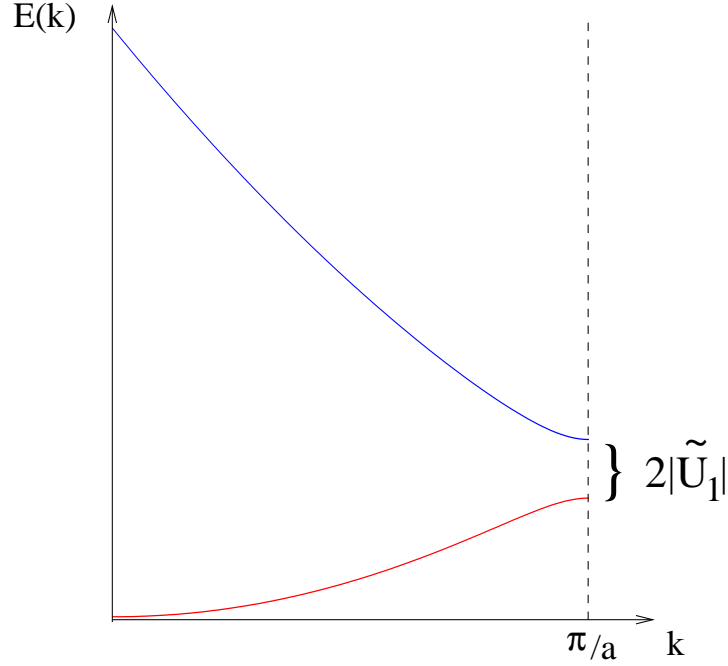
$$E(k) = \begin{cases} \frac{\hbar^2 k^2}{2m} \\ \frac{\hbar^2}{2m} \left( k - \frac{2\pi}{a} \right)^2 + o\left( \frac{|\tilde{U}_1|}{\pi} \right)^2, \end{cases} \quad \text{for } k \ll \frac{\pi}{a}.$$

We see that  $\tilde{U}_1$  doesn't make much difference for  $k \ll \frac{\pi}{a}$ , but it has a significant effect when  $k$  is on or near the first Brillouin zone boundary. In particular for  $k = \frac{\pi}{a}$ , equation (51) gives

$$E\left(\frac{\pi}{a}\right) = \frac{\hbar^2}{2m} \left(\frac{\pi}{a}\right)^2 \pm |\tilde{U}_1|$$

a 'gap' has opened up in the energy spectrum, at  $k = \frac{\pi}{a}$  of magnitude  $2|\tilde{U}_1|$ . This is called a **band gap** and is a generic feature of solutions of (46), the periodic potential  $U$  causes such gaps to open up in the spectrum.

The two solutions (51) are plotted below,



Now focus on the  $2 \times 2$  sub-matrix of the central equation (46) obtained by restricting to  $h = +1$  and  $h = -1$ .

$$\begin{pmatrix} \frac{\hbar^2}{2m} \left(k + \frac{2\pi}{a}\right)^2 - E & \tilde{U}_2 \\ \tilde{U}_{-2} & \frac{\hbar^2}{2m} \left(k - \frac{2\pi}{a}\right)^2 - E \end{pmatrix} \begin{pmatrix} b_k(1) \\ b_k(-1) \end{pmatrix} = 0. \quad (52)$$

Again demanding a non-zero solution for  $b_k(1)$  and  $b_k(-1)$  requires that the determinant of the associated  $2 \times 2$  matrix vanishes,

$$\begin{vmatrix} \frac{\hbar^2}{2m} \left(k + \frac{2\pi}{a}\right)^2 - E & \tilde{U}_2 \\ \tilde{U}_{-2} & \frac{\hbar^2}{2m} \left(k - \frac{2\pi}{a}\right)^2 - E \end{vmatrix} = 0.$$

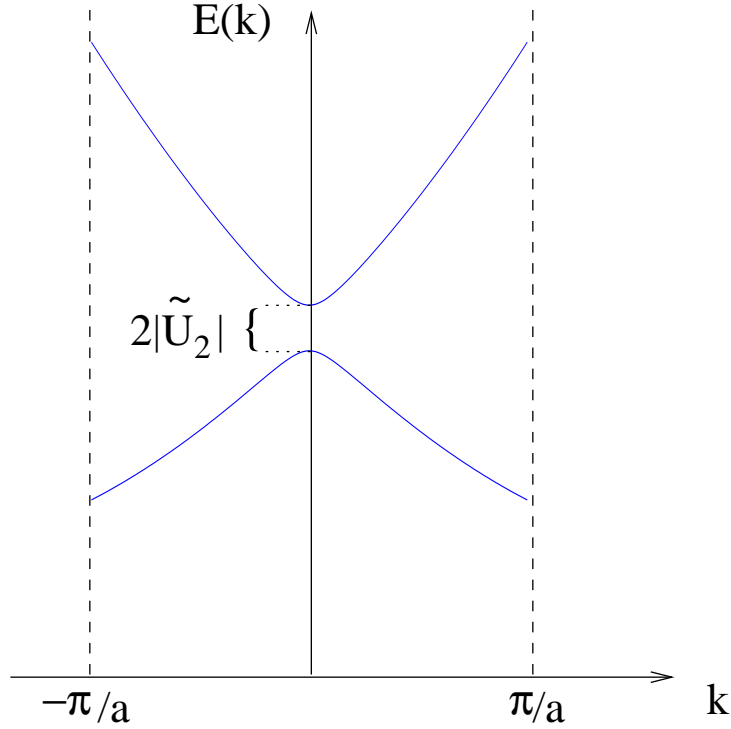
The solutions are

$$E(k) = \frac{\hbar^2}{m} \left\{ \frac{k^2}{2} + \frac{2\pi^2}{a^2} \pm \frac{4\pi}{a} \sqrt{k^2 + \left(\frac{ma}{2\pi\hbar^2}\right)^2 |\tilde{U}_2|^2} \right\}.$$

Now there is a gap between the two bands at  $k = 0$ ,

$$E(0) = \frac{2\pi^2\hbar^2}{ma^2} \pm |\tilde{U}_2|.$$

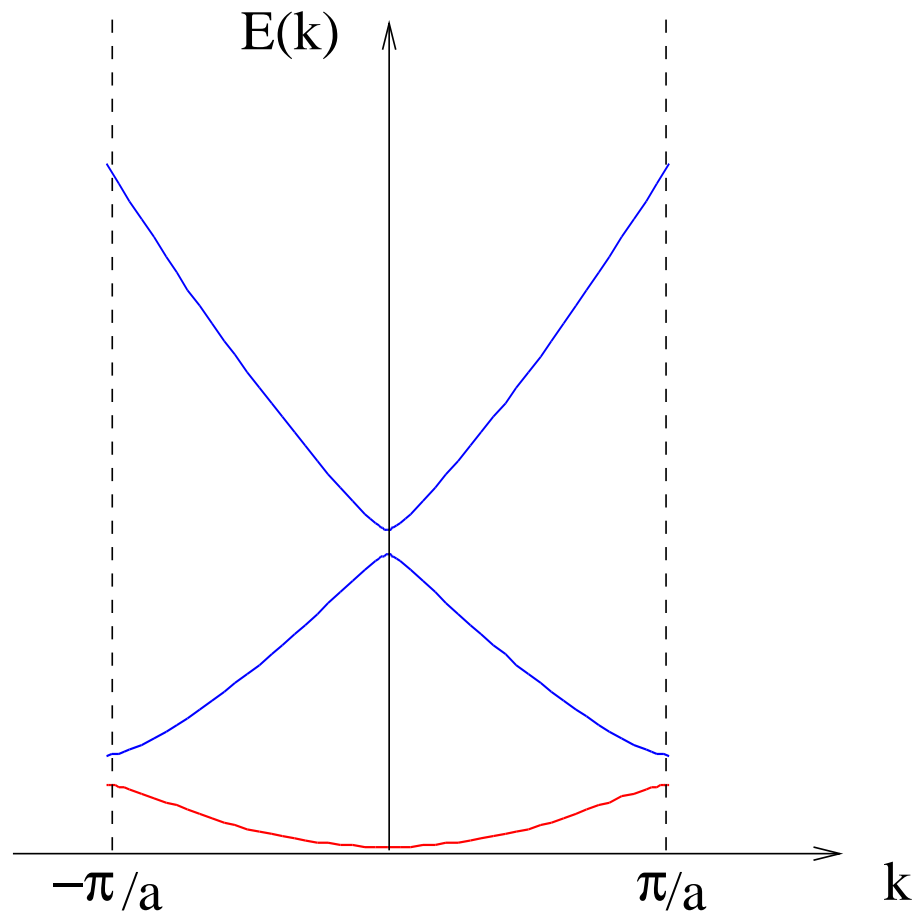
The structure of these two bands is shown below



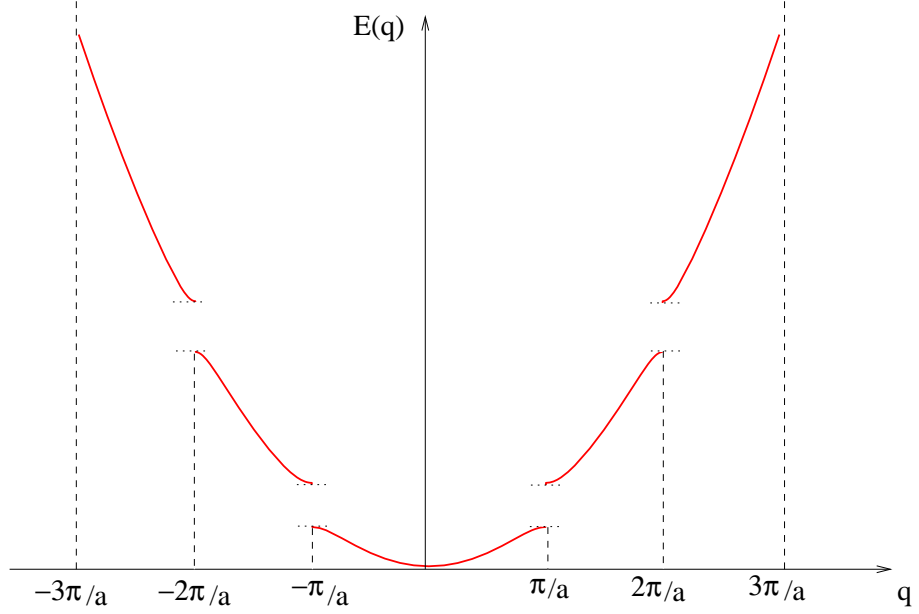
The above results can be combined by focusing on the  $3 \times 3$  sub-matrix of the central equation (46) obtained by restricting to  $h = +1, 0$  and  $-1$ . We now demand that the determinant of the matrix

$$\begin{pmatrix} \frac{\hbar^2}{2m} \left(k + \frac{2\pi}{a}\right)^2 - E & \tilde{U}_1 & \tilde{U}_2 \\ \tilde{U}_1^* & \frac{\hbar^2}{2m} k^2 - E & \tilde{U}_1 \\ \tilde{U}_2^* & \tilde{U}_1^* & \frac{\hbar^2}{2m} \left(k - \frac{2\pi}{a}\right)^2 - E \end{pmatrix}$$

must vanish. This gives a cubic equation for  $E(k)$ , which can be solved analytically to find the three roots, all of which are even functions  $E(k) = E(-k)$ , but the explicit expressions are not very illuminating. It is more instructive to plot the three functions, using Maple or Mathematica for example. Typically  $|\tilde{U}_2|$  and  $|\tilde{U}_1| \ll \frac{\pi^2 \hbar^2}{a^2 m}$  and the spectrum is shown below. There are three bands exhibiting the band gaps found above and, comparing with the figure on page 77, we see that  $\tilde{U}_2$  has split the second energy band of the free spectrum (the blue band on page 77) into two bands here.



In the extended zone scheme the above spectrum looks like a piecewise parabola with jumps at the Brillouin zone boundaries. The gaps can be thought of as being due to reflection of electrons off the zone boundaries.



Sometimes it is convenient to continue the reduced zone scheme to wavevectors outside the first Brillouin zone to get an infinite number of copies of the reduced zone scheme with period  $\frac{2\pi}{a}$  — this is called the **periodic zone scheme**.

To gain a deeper understanding of the band structure for electrons in metals consider the energy eigenfunctions in the  $2 \times 2$  matrix equation (50). For simplicity, assume that  $\tilde{U}_1$  is real,  $\tilde{U}_1^* = \tilde{U}_1$ , and negative so that  $U(x)$  is negative for  $x = na$ , corresponding to attractive ion cores at  $x = na$ . We can determine the eigenvectors  $\begin{pmatrix} b_k(0) \\ b_k(-1) \end{pmatrix}$  by putting the eigenvalues (51) into (50). At the zone boundary,  $k = \frac{\pi}{a}$ , denote the eigenvalues by  $E_{\pm}$ , with  $E_+ > E_-$ ,

$$E_{\pm} = \frac{\hbar}{2m} \left( \frac{\pi}{a} \right)^2 \pm |\tilde{U}_1| \quad \Rightarrow \quad E^{(0)} \left( \frac{\pi}{a} \right) - E_{\pm} = \mp |\tilde{U}_1|$$

and the eigenvectors are determined by

$$\begin{pmatrix} \mp |\tilde{U}_1| & -|\tilde{U}_1| \\ -|\tilde{U}_1| & \mp |\tilde{U}_1| \end{pmatrix} \begin{pmatrix} b_{\frac{\pi}{a}}(0) \\ b_{\frac{\pi}{a}}(-1) \end{pmatrix} = 0.$$

The solutions are

$$b_{\frac{\pi}{a}}(0) = \mp b_{\frac{\pi}{a}}(-1)$$

which determines the wave-function  $\psi_k(x) = \sum_G b_k(G) e^{i(k+G)x}$  at  $k = \frac{\pi}{a}$ . In this approximation there are only two terms in the sum

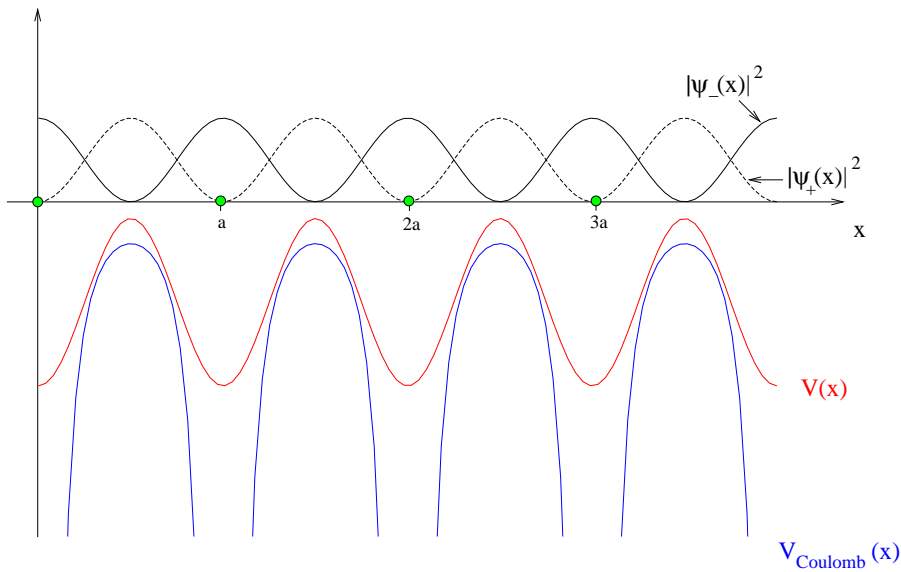
$$\psi_{\frac{\pi}{a}}(x) = \sum_{h=-1}^0 b_{\frac{\pi}{a}}(h) e^{i\left(\frac{\pi}{a} + \frac{2\pi h}{a}\right)x} = b_{\frac{\pi}{a}}(0) \left( e^{\frac{i\pi}{a}} \mp e^{-\frac{i\pi}{a}} \right) := \psi_{\pm}(x).$$

Hence the energies and associated wave-functions are

$$E_- = E^{(0)} - |\tilde{U}_1|, \quad \psi_-(x) \propto \cos\left(\frac{\pi x}{a}\right)$$

$$E_+ = E^{(0)} + |\tilde{U}_1|, \quad \psi_+(x) \propto \sin\left(\frac{\pi x}{a}\right).$$

The energy eigenstates form standing waves as the electrons are reflected off the zone boundaries. The higher energy state,  $E_+$ , has the electron wave-function concentrated mid-way between the positive ions, at  $x = (n + \frac{1}{2})a$ , while the lower energy state,  $E_-$ , has the electron wave-function concentrated at the positive ions, at  $x = na$ , where  $n$  is an integer. This is because the negatively charged electrons are attracted to the positively charged atomic cores (green in the figure below).



In this illustrative example the attractive Coulomb potential (blue above) is modelled by a periodic potential  $V(x) = -V_0 \cos\left(\frac{2\pi x}{a}\right)$  with  $\tilde{U}_i = 0$  for  $i \neq \pm 1$  and  $\tilde{U}_1 = -\frac{V_0}{2}$  (red). The lower energy state electron wave-function (solid black line) is attracted to the potential minima at  $x = na$  and hence peaks there. The higher energy state wave-function (dotted black line) has a minimum at the potential minima and a maximum at the potential maxima.

In three-dimensions the electron wave-function is periodic in three directions,  $\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathcal{N}_1 \mathbf{a}_1) = \psi(\mathbf{r} + \mathcal{N}_2 \mathbf{a}_2) = \psi(\mathbf{r} + \mathcal{N}_3 \mathbf{a}_3)$  and the details are somewhat more complicated, but the basic concepts are the same. There can be different band structures in different directions.

### Fermi surfaces for metals

The dynamics of mobile electrons in crystals is greatly affected by the fact that they are *fermions* and fermions must obey the Pauli exclusion principle — no two fermions can occupy the same quantum state.

Again consider a one-dimensional crystal with length  $L = \mathcal{N}a$  and  $\mathcal{N}$  sites. The allowed wavevectors in the first Brillouin zone are  $k = \frac{2\pi j}{\mathcal{N}a}$ , with  $j = \pm 1 \dots, \pm \frac{\mathcal{N}}{2}$  (assume  $\mathcal{N}$  is even, when  $\mathcal{N}$  is very large this doesn't matter) and they form a discrete set with spacing  $\Delta k = \frac{2\pi}{L} = \frac{2\pi}{\mathcal{N}a}$  between allowed states. For free electrons the de Broglie relation between momentum and wavevector in quantum mechanics,  $p = \hbar k$ , implies that there are also  $\mathcal{N}$  allowed momentum states  $p = \frac{2\pi\hbar j}{\mathcal{N}a}$  with one momentum state in each interval  $\Delta p = \frac{2\pi\hbar}{\mathcal{N}a}$  in momentum space. Since electrons have spin one-half it has two spin states and each momentum state can accommodate at most two electrons, one spin-up and one spin-down. If the energy is an even function of momentum,  $E(-p) = E(p)$ , and is independent of spin then there are four quantum states for each allowed value of the energy,  $\pm p$ , spin-up and spin-down.

In a monovalent metal crystal with a monatomic basis (*e.g.* Na or K) only the single electron in the outer electronic shell of the atom is mobile in the metal, so there is one mobile electron per atom or one mobile electron per lattice site. In a divalent metal there would two mobile electrons per atom and, if the basis is monatomic, two mobile electrons per lattice site. Consider a monatomic crystal of a monovalent metal, so there are  $\mathcal{N}$  mobile electrons. Imagine starting with no mobile electrons and adding them to the metal one by one. The lowest energy, with four quantum states, is filled first, once that is full, the fourth electron has to go into the next available energy state, that is the next energy level above the lowest one, because it is excluded from the lowest energy by the Pauli exclusion principle. Continue in this way until all  $\mathcal{N}$  electrons have been used up and the lowest  $\frac{\mathcal{N}}{4}$  energy states are full. The electrons in the topmost filled energy state will have a momentum  $|p_F| = \hbar k_F$  with  $\frac{k_F}{\Delta k} = \frac{\mathcal{N}}{4}$  where  $\Delta k = \frac{2\pi}{\mathcal{N}a}$ , so

$$k_F = \frac{\pi}{2a}.$$

The Fermi wavevector,  $k_F$ , is half-way to the first Brillouin zone boundary (this factor of two is because of spin degeneracy). The Fermi momentum associated with  $k_F$  is

$$p_F := \hbar k_F = \frac{\pi\hbar}{2a},$$

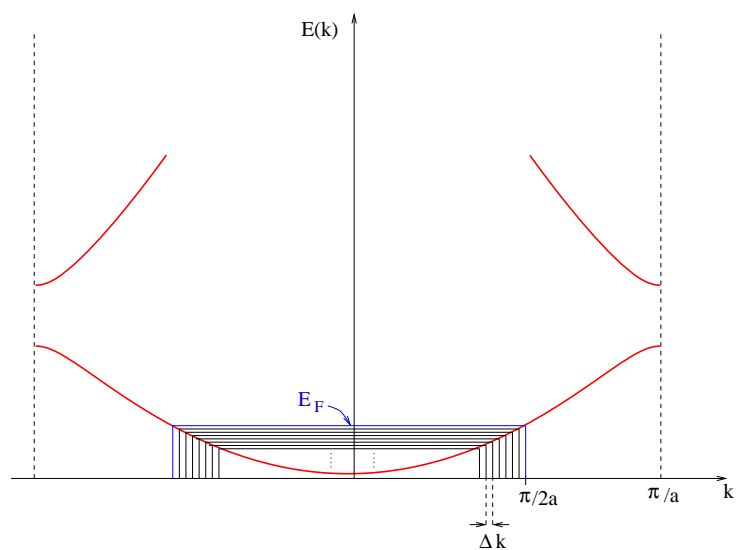
and the Fermi energy is  $E_F = E(p_F)$ . For free electrons, for example,

$$E_F = \frac{p_F^2}{2m} = \frac{\hbar^2\pi^2}{8ma^2} \quad (53)$$

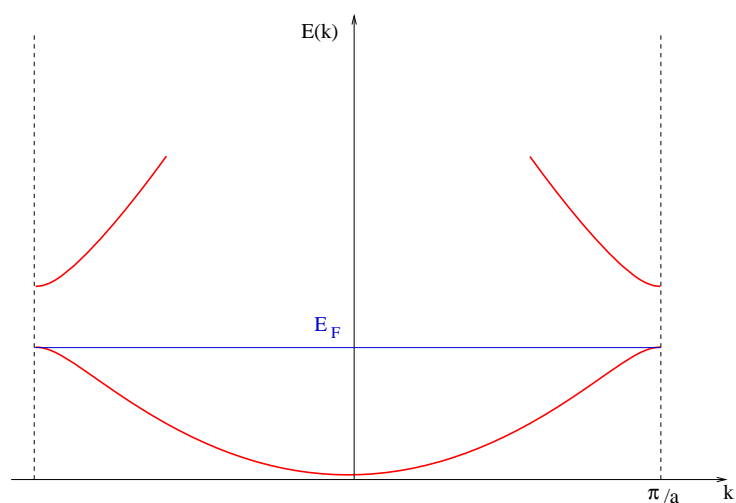
but the concept of the Fermi momentum and the Fermi energy is valid for any dispersion relation.

The energy levels are sketched below. For any finite  $\mathcal{N}$  the allowed  $k$ -values are always discrete but as  $\mathcal{N}$  increases the allowed states crowd closer and closer together as  $\Delta k$  gets smaller and smaller. As the number of atoms,  $\mathcal{N}$ , increases the number of mobile electrons also increases in just the right way so that the maximum wavenumber,  $k_F$ , and the top filled energy level,  $E_F$ , are independent of  $\mathcal{N}$ . We can even let  $\mathcal{N} \rightarrow \infty$ , giving a continuum of states but still with the same Fermi momentum  $p_F = \frac{\hbar\pi}{2a}$  and Fermi energy  $E_F = E(p_F)$ .

The Fermi momentum and Fermi energy are intrinsic to the microscopic crystal structure and are independent of the crystal size — this is an important point, if this were not the case they would not be such useful concepts.



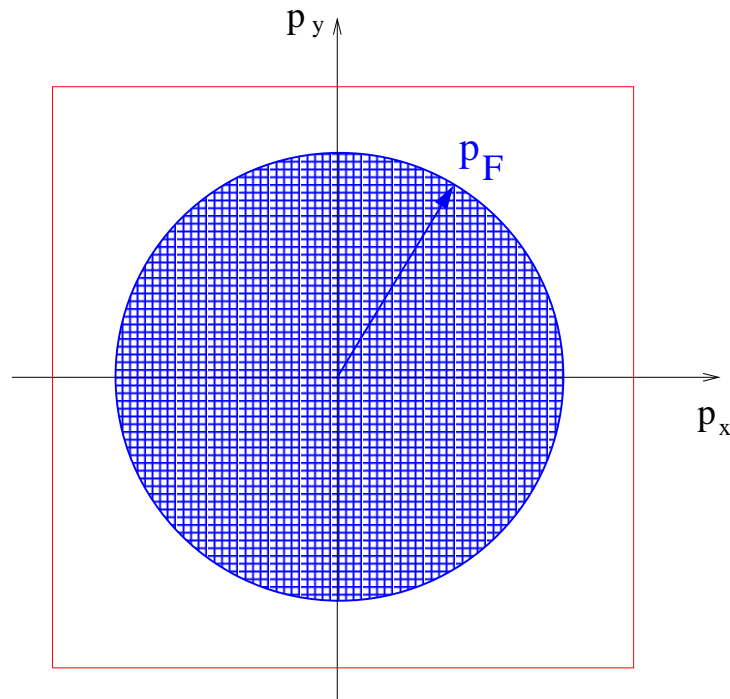
The above sketch is specific to a monovalent, monatomic basis. For a metal with either a divalent monatomic basis or a monovalent diatomic basis there are  $2\mathcal{N}$  electrons in a crystal with  $\mathcal{N}$  cells, but the same number of momentum states, so the Fermi wavevector reaches all the way to edge of the first Brillouin zone,  $k_F = \frac{\pi}{a}$ ,



In two dimensions there is a whole grid of allowed points in two-dimensional  $k$ -space,

spanned by the two components of the wavevector,  $k_x$  and  $k_y$ . Consider a two-dimensional square lattice with lattice spacing  $a$ ,  $\mathcal{N}_1$  cells in the  $x$ -direction and  $\mathcal{N}_2$  cells in the  $y$ -direction, so the total number of cells in the crystal is  $\mathcal{N} = \mathcal{N}_1\mathcal{N}_2$  and the area of the crystal is  $\mathcal{N}a^2$ . For simplicity we shall choose  $\mathcal{N}_1 = \mathcal{N}_2$ , so  $\mathcal{N} = \mathcal{N}_1^2$ , but the changes for  $\mathcal{N}_1 \neq \mathcal{N}_2$  are pretty straightforward. The spacing of allowed states in both the  $k_x$ -direction and the  $k_y$ -direction is  $\frac{2\pi}{\mathcal{N}_1 a}$ , so there is one state per area  $(\frac{2\pi}{\mathcal{N}_1 a})^2 = (\frac{4\pi^2}{\mathcal{N}a^2}) := \Delta^2 k$  in  $k$ -space. If the dispersion relation is rotationally symmetric and  $E(p)$  depends only on the magnitude of the momentum  $\mathbf{p}$  and not its direction, then the quantum states will fill up a disc in two-dimensional momentum space. For example for free fermions  $E(p) = \frac{p^2}{2m}$  and the filled states form a disc of radius  $p_F$  in momentum space. The radius of this disc is independent of  $\mathcal{N}$  and, when  $\mathcal{N}$  is very large we can think of the distribution of states as a continuum.

For a monovalent monatomic basis the area of this disc (the Fermi disc) in the space of wavevectors is one-half the area of the first Brillouin zone<sup>21</sup> which is  $\frac{1}{2} \frac{4\pi^2}{a^2} = \frac{2\pi^2}{a^2}$ . Hence the Fermi disc has radius  $k_F$  given by  $\pi k_F^2 = \frac{2\pi^2}{a^2}$ , or  $k_F = \frac{\sqrt{2\pi}}{a}$ . The reciprocal lattice for a simple cubic lattice with lattice spacing  $a$  is another simple cubic lattice with lattice spacing  $\frac{2\pi}{a}$  and the first Brillouin zone is a square with  $-\frac{\pi}{a} \leq k_x \leq \frac{\pi}{a}$  and  $-\frac{\pi}{a} \leq k_y \leq \frac{\pi}{a}$ . Since  $\frac{\sqrt{2\pi}}{a} = 0.7979 \frac{\pi}{a} < \frac{\pi}{a}$  the Fermi disc lies entirely within the first Brillouin zone.



Note that the number of states with energy  $E_F$  is the circumference of the disc divided by the average separation between two states in  $k$ -space, which we can take to be  $\sqrt{\Delta^2 k}$ , and

---

<sup>21</sup> Again the factor of  $\frac{1}{2}$  is due to electron spin — there are two quantum states for every allowed momentum state.

again multiply by two for spin. For large  $\mathcal{N}$

$$2 \left( \frac{2\pi k_F}{\sqrt{\Delta^2 k}} \right) = 2 \left\{ \frac{\frac{(2\pi)^{2/3}}{a}}{\left(\frac{2\pi}{\sqrt{\mathcal{N}a}}\right)} \right\} = 2\sqrt{2\pi\mathcal{N}},$$

and so grows like  $\sqrt{\mathcal{N}} = \frac{L}{a}$ , linearly with the size of the crystal. This is in marked contrast to the one-dimensional case where the number of states with energy  $E_F$  is always only four, two spin states for  $k = \pm \frac{\pi}{a}$ .

For a divalent metal there are twice as many mobile electrons for the same number of lattice cells and the area of the Fermi disc is doubled, the radius therefore increases by  $\sqrt{2}$  to  $k_F = \frac{2\sqrt{\pi}}{a} = 1.128\frac{\pi}{a} > \frac{\pi}{a}$  and the Fermi disc extends outside of the first Brillouin zone into the second zone.

In three dimensions consider a crystal of volume  $V$  and a simple cubic lattice structure with lattice spacing  $a$ . For simplicity we assume the crystal is a cube of size  $L$ , with edges aligned with the crystal axes, so  $V = L^3$  and the number of primitive cells is  $\mathcal{N} = \frac{L^3}{a^3}$  (for large  $\mathcal{N}$  the overall shape of the crystal is not important, it is only  $\mathcal{N}$  that matters). There is one momentum state per volume  $\Delta^3 k = \left(\frac{2\pi}{L}\right)^3 = \frac{8\pi^3}{V}$  in wavevector space and, for  $\mathcal{N}$  large we picture the allowed states as sequentially filling a sphere in  $k$ -space of volume  $\frac{4\pi}{3}k_F^3$ , called the **Fermi sphere**, with

$$\mathcal{N} = 2 \frac{\left(\frac{4\pi}{3}k_F^3\right)}{\Delta^3 k} \quad \Rightarrow \quad \frac{4\pi k_F^3}{3} = 4\pi^3 \frac{\mathcal{N}}{V} \quad (54)$$

(again the factor of 2 is for spin) giving  $k_F = (3\pi^2 n_c)^{\frac{1}{3}} = \frac{(3\pi^2)^{\frac{1}{3}}}{a} \approx 0.985\left(\frac{\pi}{a}\right) < \frac{\pi}{a}$  where  $n_c := \frac{\mathcal{N}}{V} = \frac{1}{a^3}$  is the number of primitive cells per unit volume. For a free electron dispersion relation,  $E = \frac{p^2}{2m}$ , this gives the Fermi energy

$$E_F = \frac{\hbar^2}{2m} (3\pi^2 n_c)^{\frac{2}{3}} \quad (55)$$

Since  $k_F = 0.985\left(\frac{\pi}{a}\right)$  the Fermi surface does not extend as far as the edge of the first Brillouin zone at  $k = \frac{\pi}{a}$ , but it almost does and any slight distortion of it can easily send part of the Fermi surface into the second Brillouin zone.

The number of states with energy  $E_F$  in this case is twice the area of the Fermi sphere divided by the average area of a single state in  $k$ -space when projected onto the sphere, which we can take to be  $(\Delta^3 k)^{2/3}$  giving

$$2 \left\{ \frac{\left(\frac{4\pi k_F^2}{3}\right)}{\Delta^3 k} \right\} = 2 \frac{\left\{ \frac{4\pi}{3} \left(\frac{3\pi^2}{a^3}\right)^{2/3} \right\}}{\left(\frac{8\pi^3}{\mathcal{N}a^3}\right)^{2/3}} = 2 \left(\frac{2\pi}{3}\right)^{\frac{1}{3}} \mathcal{N}^{\frac{2}{3}} = 2 \left(\frac{2\pi}{3}\right)^{\frac{1}{3}} \left(\frac{L}{a}\right)^2,$$

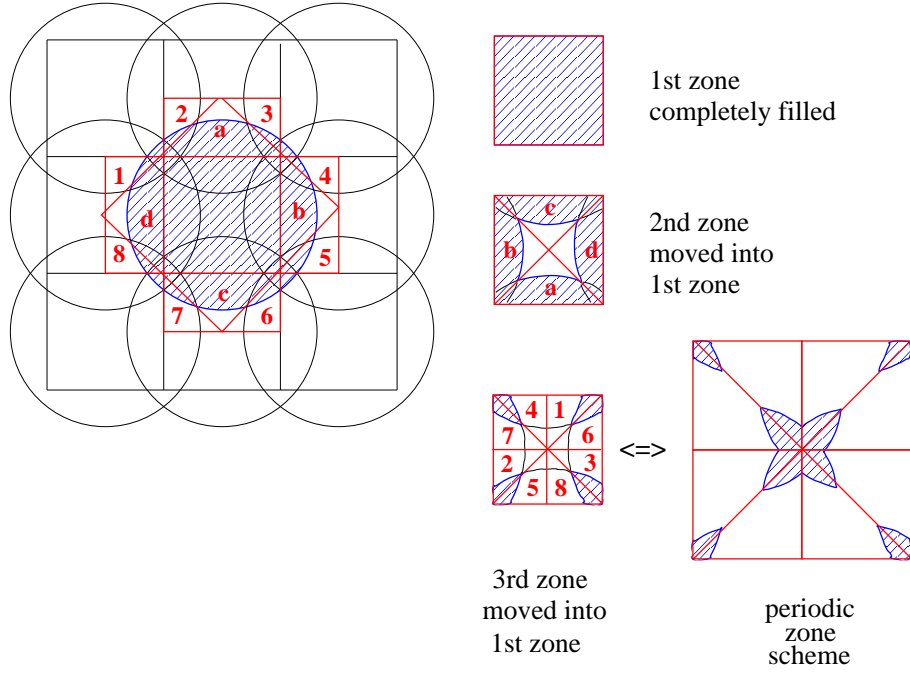
quadratically with the size of the crystal.

For a divalent metal with a monatomic basis, such as magnesium for example, there are two mobile electrons for every primitive cell which doubles the density of electrons and doubles the volume of the Fermi sphere (54), increases  $k_F$  by a factor  $2^{\frac{1}{3}}$  to  $k_F = (6\pi^2 n_c)^{\frac{1}{3}}$  and increases the Fermi energy by a factor  $2^{\frac{2}{3}}$ .

Since the boundary of the first Brillouin zone for a simple cubic lattice with lattice spacing lies at  $k = \frac{\pi}{a}$ , the Fermi surface of a monovalent metal  $k_F = \frac{(3\pi^2)^{\frac{1}{3}}}{a} \approx \frac{3.09}{a} < \frac{\pi}{a}$  lies inside the first Brillouin zone. For a divalent metal  $k_F = \frac{(6\pi^2)^{\frac{1}{3}}}{a} \approx \frac{3.90}{a} > \frac{\pi}{a}$  and the Fermi surface extends into the second Brillouin zone.

The concept of the Fermi surface is of central importance in understanding the dynamics of electrons in metals. A crystal with inter-atomic spacing  $a = 4 \text{ \AA} = 4 \times 10^{-10} \text{ m}$  has Fermi energy  $E_F = 1.5 \times 10^{-17} \text{ J} \approx 94 \text{ eV}$ , so an electron deep within a Fermi sphere has no empty quantum states near it, its energy must change by at least  $\approx 90 \text{ eV}$  in order for it to change quantum state. Any interaction with neighbouring electrons, phonons, photons or anything else leaves it completely unaffected unless the energy transfer is of order  $90 \text{ eV}$ . This is a very large energy, for example the thermal energy of an electron at room temperature is of order  $k_B T = 4 \times 10^{-21} \text{ J} \approx 0.02 \text{ eV} \ll E_F$ . This means that an electron deep within the Fermi sphere is essentially frozen out of all the dynamics. Only electrons near the Fermi surface, in a thin shell of thickness  $k_B T$ , or about 1% of the radius of the sphere in momentum space, can be thermally excited out of their filled energy state into an available empty energy state nearby, so only about 1% of all electrons are available to transport quantities such as electric current or heat energy.

The shape of the Fermi surface is very important in understanding transport properties of electrons in metals and it is often useful to visualise it in the reduced zone scheme. For a two-dimensional crystal with a monatomic basis of divalent metal atoms  $k_F = \frac{2\sqrt{\pi}}{a} > \frac{\pi}{a}$  and the Fermi surface extends into the second, the third and even the fourth Brillouin zone.



In the above picture the Fermi sphere is broken up into pieces lying in different Brillouin zones which are then moved around by reciprocal lattice vector translation to re-assemble the pieces into a single shape in each zone. The four pieces in the second zone are labelled  $a$ ,  $b$ ,  $c$  and  $d$  and the eight pieces in the third zone are labeled  $1, \dots, 8$ . In the reduced zone scheme there is a ‘hole’ of empty states in the middle of the second Brillouin zone states and a star-shaped region of filled states in the third zone. The pictures are drawn assuming that the Fermi surface is a perfect circle, which is a consequence of using the free particle relation between energy and momentum (53), when the periodic potential energy is taken into account and band gaps open up the general shape doesn’t change much except that the sharp edges in these pictures are replaced by more rounded edges.

The shape of the Fermi surface can be explored experimentally using magnetic fields. An electron moving in a magnetic field experience the Lorentz force  $\mathbf{F} = -e(\mathbf{v} \times \mathbf{B})$  so  $\mathbf{F} \cdot \mathbf{v} = 0$  and the electron’s energy does not change, it moves on a surface of constant energy in momentum space. In particular electrons can move around on the Fermi surface, but cannot leave it.

For a given dispersion relation (not necessarily quadratic)  $E(\mathbf{k})$  the group velocity of the particles,  $\mathbf{v}^g$ , has components

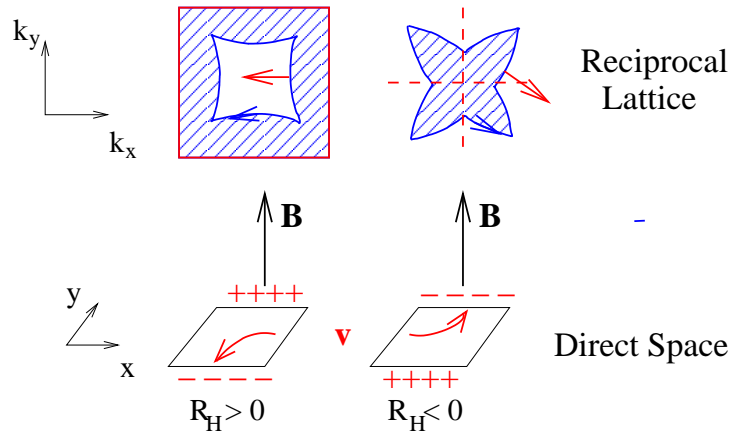
$$v_i^g = \frac{1}{\hbar} \frac{\partial E}{\partial k_i} \quad \text{so} \quad \mathbf{v}^g = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}),$$

where  $\nabla_{\mathbf{k}}$  denotes the gradient operator in  $\mathbf{k}$ -space. The Lorentz force is then

$$\mathbf{F} = \dot{\mathbf{p}} = \hbar \dot{\mathbf{k}} = -e\mathbf{v} \times \mathbf{B} = -\frac{e}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \times \mathbf{B},$$

so the force is in a direction perpendicular to both  $\mathbf{B}$  and  $\nabla_{\mathbf{k}} E(\mathbf{k})$  in  $\mathbf{k}$ -space. Now the gradient operator  $\nabla_{\mathbf{k}} E(\mathbf{k})$  returns a vector that is normal to surfaces of constant energy.

This direction is indicated by the red arrow in the figures below — it points in opposite directions for electrons in a Brillouin zone with a hole in the middle compared to Brillouin zones with a solid island in the middle, and the resulting forces in the direct lattice point in opposite directions. In the former case the electrons circulate around the hole in a clockwise direction (left-hand figure below, blue arrow), in the latter case they circulate anti-clockwise (right-hand figure below, blue arrow). Assuming the dispersion relation  $E(\mathbf{p})$  depends only on  $p^2 = \mathbf{p} \cdot \mathbf{p}$ , as is the case for cubic lattices by symmetry, then  $\mathbf{v}$ , and hence  $\mathbf{p}$ , point in the same direction in the direct lattice as  $\nabla_{\mathbf{k}}E(k)$  in the reciprocal lattice and the resultant electron motion is shown in real space in the second line of figures below, for the two cases. We see that the two possibilities result in different signs for the Hall co-efficients.

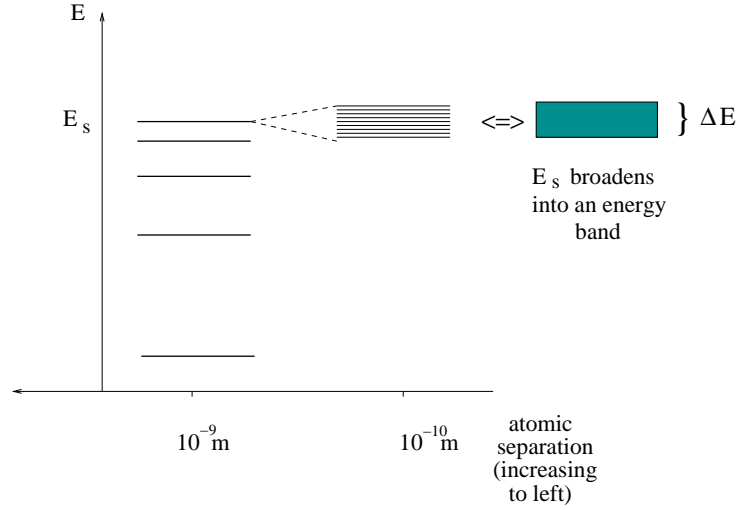


The overall sign of the Hall co-efficient depends on how many electrons lie on the part of the Fermi surface in the second Brillouin zone and how many lie on the part of the Fermi surface in the third zone. Magnesium actually has a negative Hall co-efficient, indicating that electrons win over holes, but aluminium (valence 3), has a positive Hall co-efficient, indicating that there are more electrons orbiting around a hole in the Fermi surface of aluminium than around an island. This explains the mystery of the positive Hall co-efficients described on page 73.

### Calculation of energy bands — tight binding method

So far we have discovered and analysed the band structure of mobile electrons in metals by considering perturbations of free electrons. We started with free electrons and then modelled the effects of the crystal lattice by introducing a small periodic potential  $U(\mathbf{r})$ . We can gain further insights by going to the opposite extreme and starting with the electrons tightly bound in  $\mathcal{N}$  neutral atoms in free space and then trying to account for perturbations on this model when these atoms are brought into close proximity to each other in a crystal. Suppose the electrons in the outer shell of the free atom are in an  $s$ -wave orbital with wave-function  $\phi(\mathbf{r})$  and an energy  $E_s$  which is non-degenerate. Then this orbital is  $\mathcal{N}$ -fold degenerate in a system of  $\mathcal{N}$  atoms, because there are  $\mathcal{N}$  such orbitals each of which has the same energy. When a perturbation is switched on energy states generally tend to have their degeneracy lifted and they split into a family of  $\mathcal{N}$  very close energy levels. Suppose we perturb the energy levels of the free atoms by bringing

the atoms so close together that the electrons in one atom start to feel forces due to the electrons in its neighbour — their energy levels are split as shown in the diagram below,



In  $\mathcal{N}$  is very large, of the order of Avogadro's number  $10^{23}$  in crystals a few millimetres across, then the split energy levels are so close they almost form a continuum, or a band, with a thickness  $\Delta E$ .

In the **tight binding approximation** we use the free neutral atom orbital wave-functions for an electron at position  $\mathbf{r}$  relative to the centre of the atom,  $\phi(\mathbf{r})$ , as a basis for wave-functions of the mobile electrons in the metal,

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_j c_{\mathbf{k},j} \phi(\mathbf{r} - \mathbf{r}_j),$$

where the sum is over all the atoms in the crystal (we assume a monatomic basis),  $\mathbf{r}_j$  is the position of atom  $j$  (lattice sites) and the  $c_{\mathbf{k},j}$  are constants. While the  $\psi(\mathbf{r} - \mathbf{r}_j)$  are strongly localised around  $\mathbf{r} = \mathbf{r}_j$  the electron wave-function,  $\psi_{\mathbf{k}}(\mathbf{r})$ , extends throughout the whole crystal.

By Bloch's theorem we have

$$\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{L}) = e^{i\mathbf{k}\cdot\mathbf{L}} \psi_{\mathbf{k}}(\mathbf{r})$$

for any lattice vector  $\mathbf{L}$ . This will be true if the  $c_{\mathbf{k},j}$  are of the form  $c_{\mathbf{k},j} = \frac{1}{\sqrt{\mathcal{N}}} e^{i\mathbf{k}\cdot\mathbf{r}_j}$ , since then

$$\begin{aligned} \psi_{\mathbf{k}}(\mathbf{r} + \mathbf{L}) &= \frac{1}{\sqrt{\mathcal{N}}} \sum_j e^{i\mathbf{k}\cdot\mathbf{r}_j} \phi(\mathbf{r} + \mathbf{L} - \mathbf{r}_j) \\ &= \frac{1}{\sqrt{\mathcal{N}}} \sum_j e^{i\mathbf{k}\cdot(\mathbf{r}_j + \mathbf{L})} \phi(\mathbf{r} - \mathbf{r}_j) \\ &= e^{i\mathbf{k}\cdot\mathbf{L}} \psi_{\mathbf{k}}(\mathbf{r}). \end{aligned}$$

The energy associated with this wave-function is,

with  $\phi_j := \phi(\mathbf{r} - \mathbf{r}_j)$ ,

$$\begin{aligned} E(\mathbf{k}) &= \langle \psi_{\mathbf{k}} | \hat{H} | \psi_{\mathbf{k}} \rangle = \frac{1}{\mathcal{N}} \sum_{j,j'} e^{i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_{j'})} \langle \phi_{j'} | \hat{H} | \phi_j \rangle \\ &= \sum_j e^{i\mathbf{k} \cdot (\mathbf{r}_j - \mathbf{r}_0)} \int dV \phi^*(\mathbf{r} - \mathbf{r}_j) H(\mathbf{r}) \phi(\mathbf{r} - \mathbf{r}_0), \end{aligned} \quad (56)$$

where the integral is over the whole crystal and  $\langle \phi_j | \hat{H} | \phi_i \rangle = \int dV \phi^*(\mathbf{r} - \mathbf{r}_j) H(\mathbf{r}) \phi_i(\mathbf{r} - \mathbf{r}_i)$ . Assuming only nearest neighbour atoms have a non-zero overlap of their wave-functions define

$$\alpha := - \int dV \phi^*(\mathbf{r} - \mathbf{r}_0) H(\mathbf{r}) \phi(\mathbf{r} - \mathbf{r}_0),$$

where  $\alpha$  is positive since it is close to the energy of the orbital of an electron when the atoms are free and the electron is bound to the atom, and

$$\int dV \phi^*(\mathbf{r} - \mathbf{r}_j) H(\mathbf{r}) \phi(\mathbf{r} - \mathbf{r}_0) := \gamma$$

if  $\mathbf{r}_j$  and  $\mathbf{r}_0$  are nearest neighbours and all other overlaps are zero. With this approximation the energy eigenvalues (56) are

$$E(\mathbf{k}) = -\alpha - \gamma \sum_j e^{i\mathbf{k} \cdot \mathbf{r}_j} \quad (57)$$

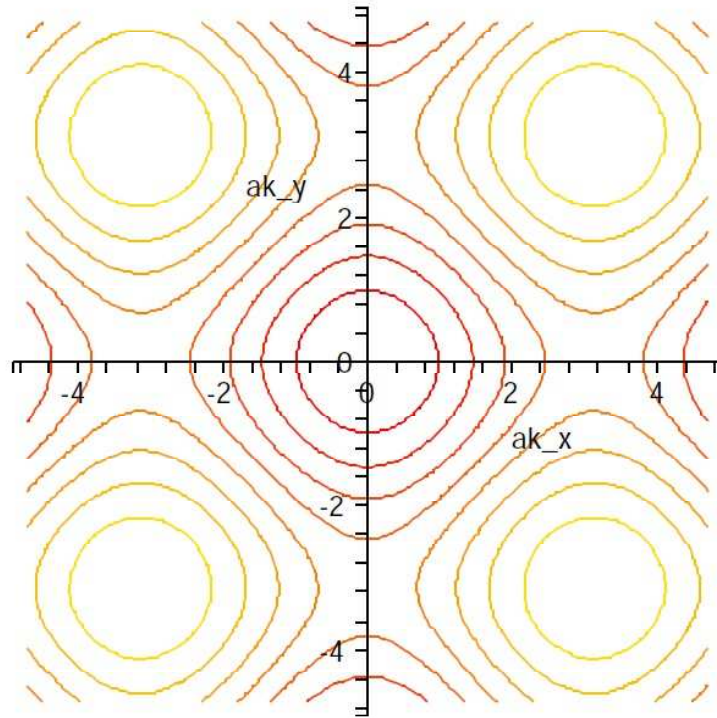
where the sum is only over lattice sites that are nearest neighbours to a reference site at  $\mathbf{r}_0$ , which can be taken to be the origin. For example in a simple cubic lattice each lattice site has six nearest neighbours, at the origin these are

$$(\pm a, 0), \quad (0, \pm a, 0) \quad \text{and} \quad (0, 0, \pm a)$$

and the energy (57) is

$$E\mathbf{k} = -\alpha - 2\gamma(\cos(k_x a) + \cos(k_y a) + \cos(k_z a)).$$

A contour plot of  $E(\mathbf{k})$  in the  $k_x - k_y$  plane, with  $k_z = 0$ , is shown below and we see that lines of equal energy are distorted from the free electron result,  $E(\mathbf{k}) = \frac{\hbar^2}{2m}(k_x^2 + k_y^2)$ , for which they would be circles.



In three dimensions the perfect spheres of the free electron approximations are distorted into curved cubical shapes. It was shown above that a spherical Fermi surface for a monatomic basis of monovalent atoms does not extend as far as the edge of the first Brillouin zone, but when it is distorted by including interactions between electronic orbitals in the tight binding approximation it can, as shown below

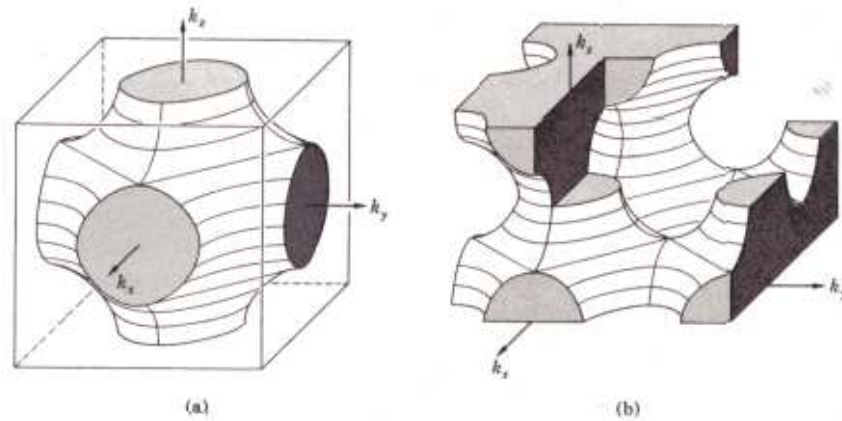
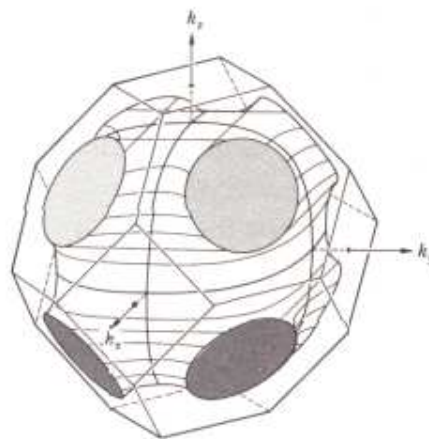


Figure 15 Constant energy surface in the Brillouin zone of a simple cubic lattice, for the assumed energy band  $\epsilon_{\mathbf{k}} = -\alpha - 2\gamma(\cos k_x a + \cos k_y a + \cos k_z a)$ . (a) Constant energy surface  $\epsilon = -\alpha$ . The filled volume contains one electron per primitive cell. (b) The same surface exhibited in the periodic zone scheme. The connectivity of the orbits is clearly shown. Can you find electron, hole, and open orbits for motion in a magnetic field  $B\hat{z}$ ? (A. Sommerfeld and H. A. Bethe.)

Copper is a valence one metal with a face centred cubic lattice and the Fermi surface looks like this, in relation to a Wigner-Seitz cell of the BCC reciprocal lattice,



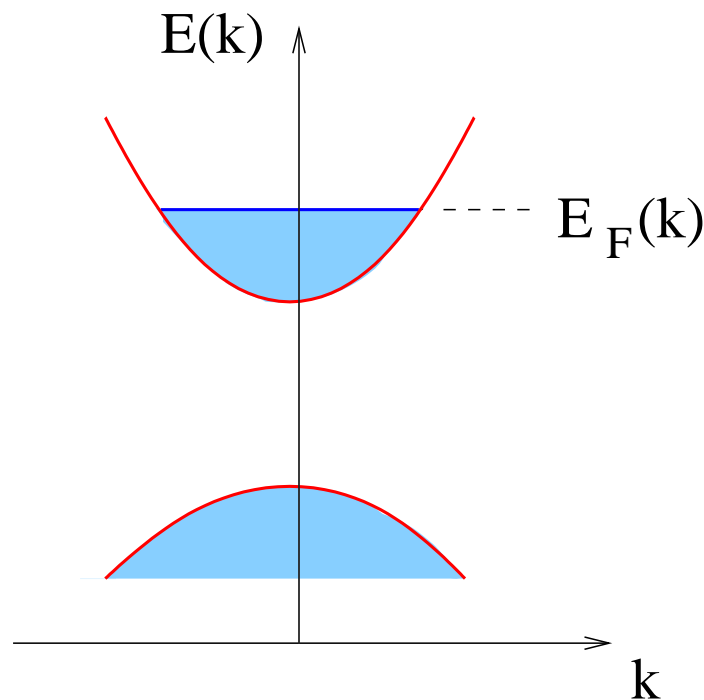
## 7. Semi-conductors

Broadly speaking conductors are materials that conduct electricity (*e.g.* metals) while insulators do not. More quantitatively conductors have small resistivities (large conductivities) at room temperature: for example silver has  $\rho = 1.59 \times 10^{-8} \Omega m$  and copper has  $\rho = 1.68 \times 10^{-8} \Omega m$ . At room temperature the main contributor to resistivity is scattering of electrons off phonons and as the temperature decreases the number of phonons decreases and the resistivity goes down like  $\rho \propto T$ , but at very low temperatures electrons will scatter off impurities in the metal and the resistivity tends to a constant value,  $\rho \rightarrow \rho_0$  as  $T \rightarrow 0$ , since the impurity density is independent of  $T$ .

Insulators typically have very high resistivities,  $\rho \approx 10^{20} \Omega m$  unless applied voltages get large enough to cause electrical breakdown.

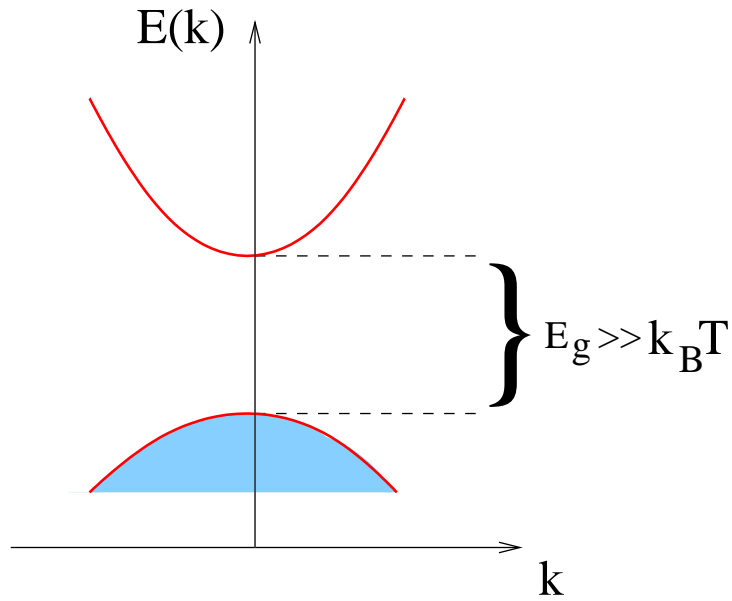
Semi-conductors are materials that have resistivities intermediate between metals and insulators, with a wide range of values  $\rho = 10^{-3} \sim 10^9 \Omega m$ . Furthermore the resistivity in these materials is very sensitive to temperature and impurity density, the resistivity *increases* as the temperature goes down — the opposite behaviour to metals! Examples of semi-conducting materials are silicon and germanium (both valence four elements).

The behaviour of these different types of materials can be understood in terms of their Fermi surfaces and the size of their band gaps.

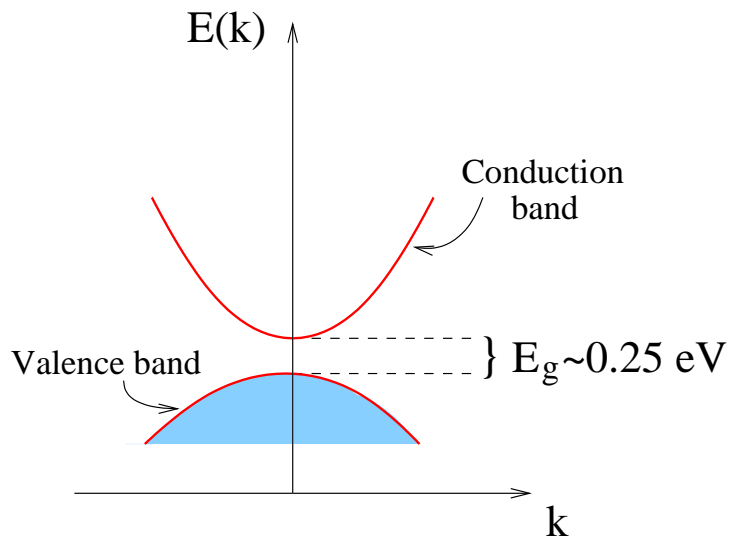


In a metal the Fermi surface is not near a band gap, there are empty states available arbitrarily close to the Fermi surface, as shown above, and there is a large number of states available with the same  $E_F$ . Electrons with energy near  $E_F$  can move easily when an electric field is applied: indeed we would expect the resistivity to be zero were it not for the fact that all real metals inevitably have a certain amount of impurity present, either foreign atoms or imperfect crystal structure, and these impede the electron's progress giving a finite resistivity  $\rho_0$ , even at  $T = 0$ .

At finite temperature there is a strip of width  $k_B T$  just above the Fermi surface which electrons can scatter into by thermal excitations, leaving behind an empty state just below the Fermi surface. Thus there is a strip with width of a few  $k_B T$  at the Fermi surface in which electrons can scatter off phonons and this scattering also contributes to the resistivity. The width of the strip is  $\propto T$  and hence the resistivity is  $\propto T$  as mentioned above.



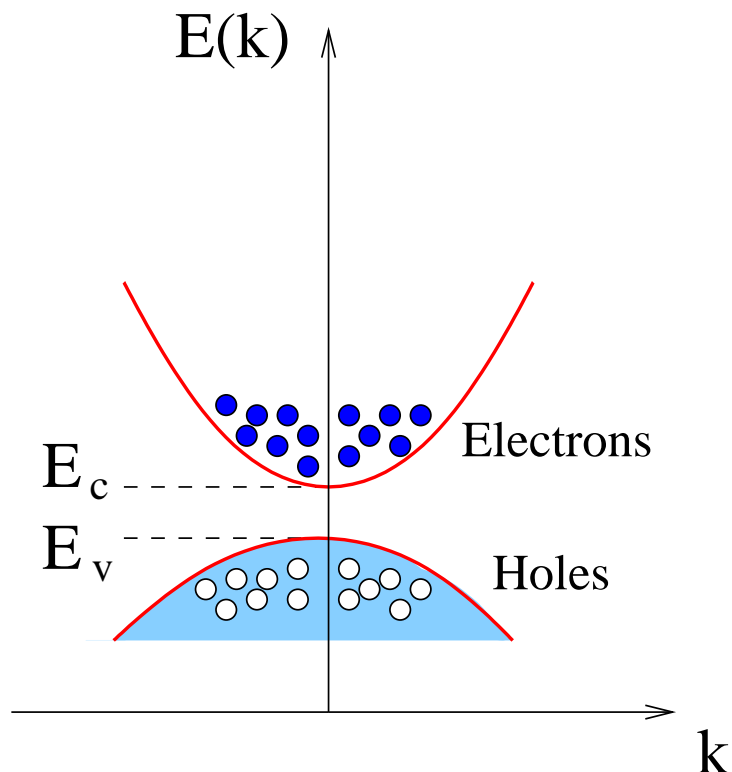
In an insulator, shown above, the Fermi energy coincides with the top of an energy band and the gap above,  $E_g$ , is much larger than the thermal energy  $k_B T$  for any realistic temperature (*i.e.* below the melting point). There are no nearby empty states available for an electron to move into — the electrons cannot move and the resistivity is essentially infinite. Neither can the electrons be thermally excited into the band above because  $E_g \gg k_B T$ . At room temperature  $k_B T \approx 0.025 \text{ eV}$  while typical band gaps in insulators are a few  $\text{eV}$ .



In a semi-conductor the Fermi energy again coincides with the top of an energy band but now the band gap,  $E_g$ , is only a few times the thermal energy  $k_B T$ . At zero temperature semi-conductors are insulators but at room temperature electrons can be thermally excited into the band above  $E_F$  and then there are nearby empty states available and they can move under the influence of an external electric field and carry a current. The higher the temperature the more nearby empty states there are and the lower the resistivity. The

upper band is called the **conduction band**, because this is the energy band in which electrons can carry current, while the lower band is called the **valence band**, because this is the band that is exactly filled by virtue of the valence of the material.

When electrons in a semi-conductor are thermally excited into the conduction band they leave behind empty states in the valence band which are called **holes**.



A hole is the absence of an electron. If an electric field is applied to make the electrons in the conduction band move to the left a rightward moving current is generated. When a hole is present the same electric field makes the electron in the valence band which is just to the right of the hole jump into the hole on its left, then the next electron to the right does the same, and so on. The net effect is that the hole moves to the right, contribute to a rightward current — the hole behaves for all practice purposes like a positively charged particle.

Denoting the energy at the bottom of the conduction band by  $E_c$  and at the top of the valence band by  $E_v$ , so the energy gap is  $E_g = E_c - E_v$ , we can Taylor expand  $E(k)$  around  $E_c$ , assuming  $E(k) = E(-k)$ , for small  $k$

$$E(k) = E_c + \frac{\hbar^2 k^2}{2m_e} + o(k^4).$$

The parameter  $m_e$  appearing in the second term of the expansion behaves like a mass for a free particle. In it is natural to *define*

$$m_e := \frac{2}{\hbar^2} \left. \frac{dE}{d(k^2)} \right|_{E_c}$$

as being the mass of the electrons in the conduction band, it is called the **effective mass** of the electrons. This depends on the dispersion relation and is often very different to the mass of an isolated free electron. In GaAs, for example, the effective mass  $m_e = 0.066m_{free}$  is a little less the 7% of the free electron mass.

Similarly if we Taylor expand around the top of the valence band

$$E(k) = E_v - \frac{\hbar^2 k^2}{2m_h} + o(k^4),$$

where we have defined

$$m_h := - \left. \frac{2}{\hbar^2} \frac{dE}{d(k^2)} \right|_{E_v}.$$

$m_h$  behaves like an effective mass for holes. Holes and electrons can have different masses in different materials, again in GaAs, for example,  $m_h = 0.082m_{free}$ . Notice that the mass of the charge carriers affects the Drude result for the conductivity (38), the contributions of electrons and holes to the conductivity can be different even when their densities are the same.

For small  $k$  the density of states for electrons has the free electron form (29), with the substitution of the effective mass and  $\varepsilon \rightarrow E - E_c$ . We also define the density of states for holes,

$$\mathcal{D}_e(E) = \frac{V}{2\pi^2} \frac{(2m_e)^{\frac{3}{2}}}{\hbar^3} (E - E_c)^{\frac{1}{2}}$$

$$\mathcal{D}_h(E) = \frac{V}{2\pi^2} \frac{(2m_h)^{\frac{3}{2}}}{\hbar^3} (E - E_v)^{\frac{1}{2}}.$$

Provided the number of electrons in the conduction band is not too large there will be many more quantum states available at the bottom of the band than there are electrons and we can evaluate the number of electrons in the conduction band with a given energy, as a function of temperature, using Maxwell-Boltzmann statistics

$$f_e = \frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) + 1} \approx e^{-\frac{(E-\mu)}{k_B T}}, \quad E > E_c.$$

The electron concentration is then

$$n_e = \frac{N_e}{V} = \frac{1}{V} \int_{E_c}^{\infty} \mathcal{D}_e(E) f_e(E) dE = \frac{1}{2\pi^2} \left( \frac{2m_e}{\hbar^2} \right)^{\frac{3}{2}} e^{\frac{\mu}{k_B T}} \int_{E_c}^{\infty} \sqrt{E - E_c} \exp\left(-\frac{E}{k_B T}\right) dE,$$

where  $N_e$  is the total number of electrons. The integral can be evaluated analytically,

$$\int_{E_c}^{\infty} \sqrt{E - E_c} \exp\left(-\frac{E}{k_B T}\right) dE = (k_B T)^{\frac{3}{2}} e^{-\frac{E_c}{k_B T}} \Gamma\left(\frac{3}{2}\right) = (k_B T)^{\frac{3}{2}} e^{-\frac{E_c}{k_B T}} \left(\frac{\sqrt{\pi}}{2}\right),$$

and

$$n_e = 2 \left( \frac{m_e k_B T}{2\pi \hbar^2} \right)^{\frac{3}{2}} \exp\left(\frac{\mu - E_c}{k_B T}\right).$$

The thermal distribution of holes can be determined by recalling that  $f_e$  is the probability of finding an electron with energy  $E$  when the temperature is  $T$ . For  $E < E_v$ , every state is either a hole or is filled with an electron so, if  $f_h$  is the distribution of holes,  $f_e + f_h = 1$  with probability one. Hence

$$f_h = 1 - \frac{1}{\exp\left(\frac{E-\mu}{k_B T}\right) + 1} = \frac{1}{\exp\left(\frac{\mu-E}{k_B T}\right) + 1} \approx e^{\frac{(E-\mu)}{k_B T}}, \quad E < E_v.$$

The hole concentration (conventionally denoted  $p_h$ , to remind us that holes carry a positive charge) is

$$\begin{aligned} p_h &= \frac{N_e}{V} = \frac{1}{V} \int_{-\infty}^{E_v} \mathcal{D}_h(E) f_h(E) dE = \frac{1}{2\pi^2} \left(\frac{2m_h}{\hbar^2}\right)^{\frac{3}{2}} e^{\frac{\mu}{k_B T}} \int_{-\infty}^{E_v} \sqrt{E_v - E} \exp\left(\frac{E}{k_B T}\right) dE \\ &= 2 \left(\frac{m_h k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(\frac{E_v - \mu}{k_B T}\right). \end{aligned}$$

The chemical potential disappears from the product

$$n_e p_h = 4(m_e m_h)^{\frac{3}{2}} \left(\frac{k_B T}{2\pi\hbar^2}\right)^3 \exp\left(\frac{-E_g}{k_B T}\right),$$

where  $E_g = E_c - E_v$  is the band gap. This is known as the **law of mass action**.<sup>22</sup>

In the semi-conductors described so far  $N_e = N_p$ , because every electron gives rise to a hole, so  $n_e = p_n$  and

$$n_e = p_h = 2(m_e m_h)^{\frac{3}{4}} \left(\frac{k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}} \exp\left(\frac{-E_g}{2k_B T}\right). \quad (58)$$

Both electrons and holes contribute to the conductivity and the Drude formula (38) gives

$$\sigma = \frac{n_e e^2 \tau_e}{m_e} + \frac{p_h e^2 \tau_h}{m_h} \quad (59)$$

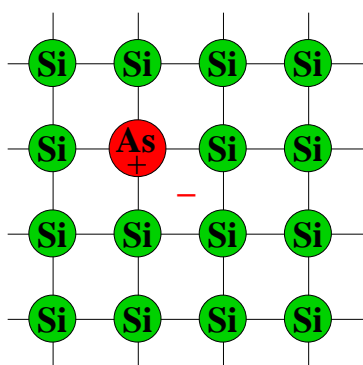
(electrons and holes can have different scattering times  $\tau_e \neq \tau_h$ , just as they can have different masses, because their dynamics can be different). The dominant temperature dependence here is the exponential behaviour in (58) and, since  $E_g > 0$  by definition, this explains why the conductivity goes up as the temperature goes up — the higher the temperature the more electrons are excited into the conduction band, increasing  $n_e$ , and at the same time more holes are created, increasing  $p_h$ . The exponential dependence on

---

<sup>22</sup> In analogy with chemical reactions where the same equation relates the density of two constituent parts of a compound molecule which forms from its constituents with release of energy  $\Delta E = E_g$ .

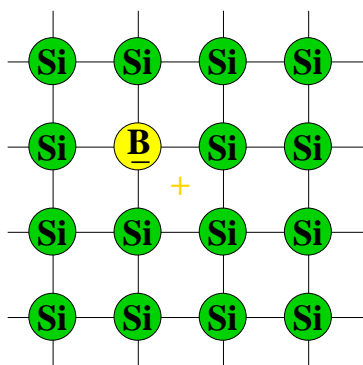
the temperature also explains the conductivity is a very sensitive function of temperature in a semi-conductor.

Semi-conductors with  $n_e = p_h$  are called **intrinsic semi-conductors**, but it is also possible to arrange for materials with  $n_e \neq p_h$  by deliberately adding impurities, a procedure called **doping**, resulting in **doped semi-conductors**. For example silicon (valence IV) is a semi-conductor. If we replace a silicon atom in a crystal of silicon with an arsenic atom then, since arsenic lies in the column just to the right of silicon in the periodic table and hence has valence V, the arsenic atom has one more electron in its outermost orbital than the silicon atom it replaced had and this electron becomes mobile in the crystal and contributes to the conductivity so  $N_e = N_p + 1$ . The arsenic atom is called a **donor**, because it donates an electron to become a mobile charge carrier.



If we replace a number of silicon atoms with arsenic atoms, but too many so that the silicon crystal retains its integrity as a crystal, then more generally  $N_e > N_p$ . By varying the concentration of arsenic we can control the conductivity quite carefully.

Similarly we could replace some silicon atoms with boron atoms, which are in the column immediately to the left of silicon in the periodic table and hence have valence III, then the boron atoms have one electron less in their outer shell than silicon atoms. Mobile electrons from silicon in the crystal then tend to get attracted to the boron and lose their mobility, thus reducing  $N_e$ , so that  $N_e < N_p$ , effectively increasing the number of holes. The boron atoms are called **acceptors**.



Semi-conductors doped with donors are called *n*-type semi-conductors while those doped with acceptors are called *p*-type.