

Chapter 1

Macroscopic Current Flow

In order to be able to gain any insight into current flow at the nanoscale, we must first consider what happens at macroscopic scales and then see the effect of reducing the dimensions to the nanoscale. The behaviour of conventional electrical circuits can be understood in quite simple terms, whereas at the nanoscale, there are a number of subtle effects which can only be understood within the framework of quantum mechanics, as we shall see in this chapter and in Chapter 2. In between these disparate regimes, we have *mesoscopic* transport, which we will briefly consider in Chapter 3.

On the basis of his detailed experimental observations in the 1820s, Georg Ohm formulated his famous Law [1] which states the following:

For a constant temperature, the current flowing through a conductor is directly proportional to the potential difference between its ends. Or, $V = IR$.

This is illustrated in Fig. 1.1. The constant of proportionality between voltage and current is known as resistance, R . The resistance in turn depends on the geometry of the conductor and a material constant, resistivity (ρ) as $R = \rho l/A$, where l and A are the length and cross-sectional area of the conductor, respectively. As we will see later, when any of the dimensions of a conductor are at the nanoscale, resistivity itself becomes dependant on geometry. For now however, let us consider Ohm's law in more detail, from a purely classical (i.e. non-quantum) standpoint — Drude's model of electronic conduction.

We will then introduce the necessary corrections which must be made to be consistent with quantum mechanics. We would like to explain the following points:

1. What is electric current?
2. Why and how does current depend on voltage?
3. What are typical values of resistance/resistivity of conductors?
4. What is the effect of changing temperature?

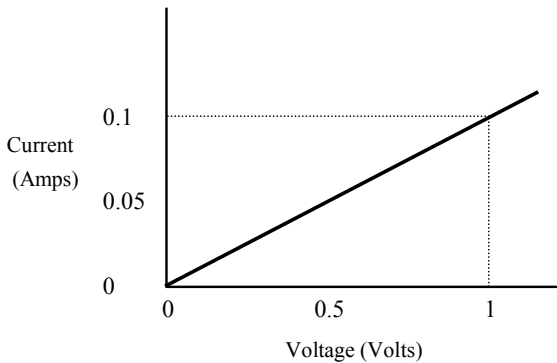


Fig. 1.1. Relationship between current and voltage for a conductor. In this case the resistance is 10Ω .

1.1. The Classical (Drude) Model of Electronic Conduction and Ohm's Law

To answer the above questions, we need to consider what happens when we apply a voltage across a conductor. The voltage causes there to be a local electric field \mathbf{E} , within the conductor. This electric field induces a force on the electrons (of charge, $e = -1.6 \times 10^{-19} \text{ C}$) of strength $e\mathbf{E}$, in a direction opposite to that of the applied field. On this basis, we would therefore expect that electrons will continue to accelerate as they traverse the conductor. The flow of charge carriers within a conductor is known as a current, and the magnitude of the current is the amount of charge (in Coulombs) passing

a point in the conductor per second. A current of 1 A corresponds to 3.25×10^{18} electrons passing a point per second.

In reality, we know that electrons do not accelerate indefinitely as they flow through a conductor, but they drift along at a finite speed (called the drift velocity, typically 10^6 m.s^{-1}). This drift velocity is somewhat analogous to the terminal velocity experienced by falling objects (which continually lose momentum to air molecules), and is due to the electrons colliding with impurities, lattice imperfections and lattice vibrations (phonons) within the conductor. The average distance and time between collisions are the *mean free path*, ℓ and *mean free time*, τ respectively and are of the order 10–50 nm and 10^{-14} s for a metal at room temperature [2]. Given that the force on an electron is $e\mathbf{E}$, and force is mass times acceleration, we find that the average velocity (*acceleration* \times *time*) of the electrons immediately before a collision is $\mathbf{v} = e\mathbf{E}\tau/m$. If we have n electrons per unit volume, then it follows that the current density, $\mathbf{J} = nev = ne^2\tau\mathbf{E}/m$. This is essentially a statement of Ohm's law, as it predicts that the current density is proportional to the electric field, which will be proportional to the applied voltage. The constant of proportionality between current density and electric field is the conductivity, σ ; i.e. $\mathbf{J} = \sigma\mathbf{E}$, where $\sigma = ne^2\tau/m$. The electrical resistivity, ρ is $1/\sigma$, so $\rho = m/ne^2\tau$.

To investigate the effect on the current of changing temperature, we need only look at the above formula. As we increase temperature, we cause the atoms in the conductor to move more vigorously. This has the effect of reducing τ and ℓ , and consequently increases the resistivity. In practice, as we increase the current flowing through a conductor, we also increase the number of collisions between the electrons and entities within the conductor, which has the effect of heating the conductor. This is the principle behind the operation of electric filament heaters and light bulbs. There comes a point when this heating causes the resistance to increase, and the current–voltage characteristics become non-linear, instead following the curve shown in Fig. 1.2 (also see Chapter 5).

Thus far, whilst we have deduced Ohm's law, which is an experimental fact, we have not considered the quantum nature of

materials. In order to gain a deeper insight into conduction, we must now consider the free-electron model of electronic conduction.

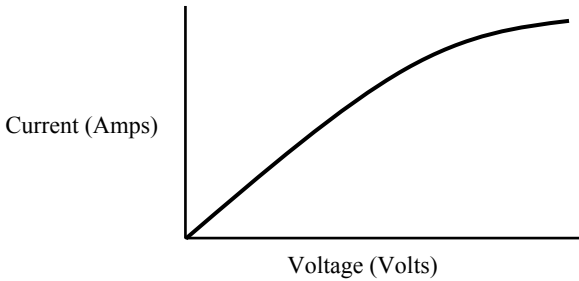


Fig. 1.2. Relationship between current and voltage for a conductor, where the current induces heating.

1.2. The Quantum (Free-Electron) Model of Electronic Conduction

In the Drude model which we have just discussed, we have implicitly assumed the following:

- The electrons do not interact with each other or the crystal lattice;
- The electrons can have *any* velocity and hence, energy;
- All electrons within a conductor contribute to conduction;
- The electron gas follows Maxwell–Boltzmann statistics.

The first assumption can be justified for most metals, where the atomic potential and electronic interactions are effectively screened out due to the high electron density. For semiconductors however, the situation is altogether different, leading to the formation of a band-gap.

The concept of the electrons inside a conductor not interacting with anything is analogous to a gas, hence the name *free-electron gas*. The second assumption is not viable, as quantum mechanics tells us that the energy of an electron in a conductor will have certain, discrete values. Finally, electrons are Fermions, and are known not to follow

Maxwell–Boltzmann statistics, but rather Fermi–Dirac statistics. These facts ultimately lead to the failure of the Drude model. The Drude model simply cannot explain a large number of phenomena, e.g. the specific heat of materials, the resistivity of materials, and many other points which we will visit later. To see where this discreteness comes from, we need to turn to a quantum description of the electrons in a conductor. Consider a cubic-shaped conductor, of side, L , as shown in Fig. 1.3. The time-independent free-particle Schrödinger equation is

$$-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \psi_k(x, y, z) = E_k \psi_k(x, y, z). \quad (1.1)$$

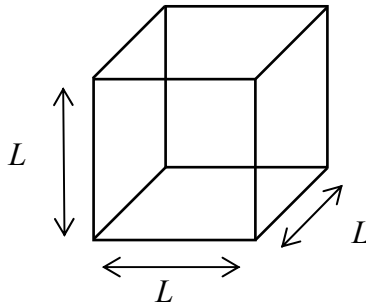


Fig. 1.3. Cube of conductor, side L , used for calculating the number of electron states.

The solution to this is

$$\psi_k(x, y, z) = A \sin(l\pi x/L) \sin(m\pi y/L) \sin(n\pi z/L), \quad (1.2)$$

where l , m , and n are integers. For generality, we must satisfy the periodic boundary condition (known as the *Born von Karman* boundary condition) that $\psi_k(x+L, y, z) = \psi_k(x, y, z)$, and similarly for y and z . This condition can only be met when l , m and n are all *even* numbers, so instead of having $k_x = l\pi/L$, we actually have $k_x = 2l\pi/L$, and similarly for k_y and k_z .

The energy of a given wavefunction (state) is therefore $E_k = \hbar^2 k^2 / 2m$ (i.e. Energy = $(1/2)mv^2$), giving a parabolic E_k - k (dispersion) relationship, as illustrated in Fig. 1.4.

Substituting for the values of k from above, we get

$$E_k = 4\pi^2 \hbar^2 (l^2 + m^2 + n^2) / 2mL^2, \quad (1.3)$$

which clearly has discrete values — a defining characteristic of quantum systems.

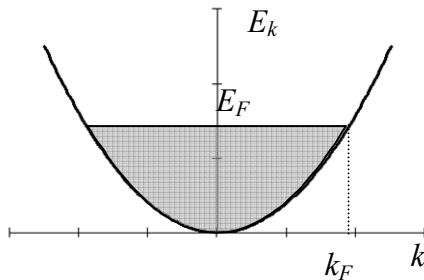


Fig. 1.4. Energy (E_k)–momentum (k) dispersion relationship according to the free-electron model.

As an example, let us consider two cubic pieces of conductor with different dimensions; one has a side of 5 nm, the other is 1 mm. The spacing between the first and second levels is given by

$$\Delta E_k = 3\hbar^2 / 2mL^2. \quad (1.4)$$

For the 5 nm conductor, this corresponds to 0.179 eV, whereas for the 1 mm conductor it is 4.5×10^{-12} eV. Given that electrons have a thermal energy of $k_B T$, which is around 25 meV at room temperature, we will not notice energy quantization in macroscopic conductors, but once any of the dimensions of the conductor are of the order 10 nm, it will have a marked effect. Figure 1.5 is a plot of the $\Delta E_k / k_B T$ as a function of size, where we have highlighted the transition between quantum, mesoscopic and macroscopic regimes. The boundaries between these regions will

shift to larger feature sizes as the temperature is reduced, and if we use semiconductors, as the electrons then have a lower effective mass.

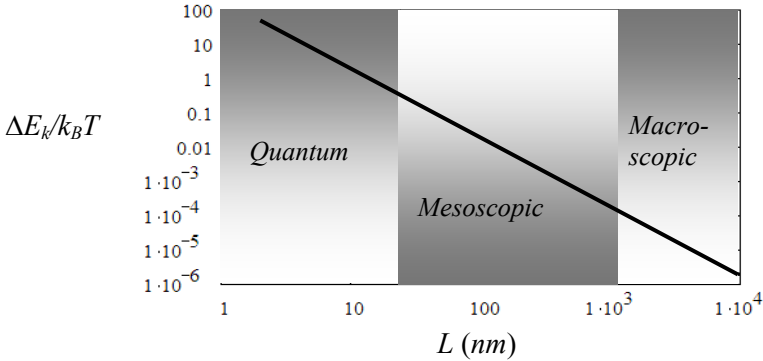


Fig. 1.5. Plot of $\Delta E_k/k_B T$ versus conductor size, for conductors ranging between 1 nm and 10 μm across.

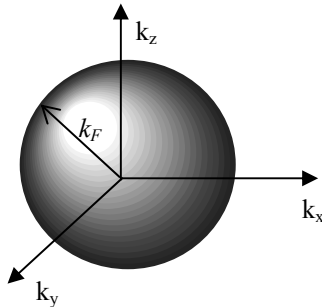


Fig. 1.6. Fermi sphere. All the k states up to k_F are filled.

As there are a finite number of electrons within a conductor, all the energy states up to a particular value (the *Fermi energy*, E_F) will be filled, as shown in Fig. 1.4. If we consider a single electron travelling through the conductor, it will have a wave-function of the form $\psi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$. It is instructive to think of the occupied states as being represented by points within a sphere in k -space. The surface of the sphere, where $\mathbf{k} = \mathbf{k}_F$ is known as the *Fermi surface*, and represents states at the Fermi energy, as illustrated in Fig. 1.6.

The energy of electrons at the Fermi surface is just $E_F = \hbar^2 k_F^2 / 2m$. In order to calculate the value of k_F , we note that there are only two points for each volume element $(2\pi/L)^3$ (the factor of 2 comes from two possible values of spin). The number of electrons in the conductor, N , is the ratio of the total volume of the Fermi sphere to the volume per state, i.e.

$$N = 2 \frac{4\pi k_F^3 / 3}{(2\pi/L)^3} = \frac{V k_F^3}{3\pi^2}, \quad (1.5)$$

where V is the volume of the conductor. It follows that

$$k_F = \left(\frac{3\pi^2 N}{V} \right)^{1/3} \Rightarrow E_F = \frac{\hbar^2}{2m} \left(\frac{3\pi^2 N}{V} \right)^{2/3}. \quad (1.6)$$

As an example, if we consider gold, which has a Fermi energy of 5.51 eV, we find that the electron concentration ($n = N/V$) is 5.93×10^{28} electrons.m⁻³, which corresponds to approximately one electron per atom. An extremely important quantity in solid-state physics, and in nanoscience is the number of states per unit energy within the energy interval dE , per unit volume, termed the *density of states*, $\mathcal{D}(E)$ of a material. This is essentially $(1/V)dN/dE$.

In 3D, the volume of a shell, radius k , thickness $dk = 4\pi k^2 dk$

$$\text{Volume per } k\text{-state} = \left(\frac{2\pi}{L} \right)^3$$

$$\Rightarrow \text{Density of states} = 2 \left(\frac{4\pi k^2 dk}{(2\pi/L)^3} \right) = \frac{L^3 k^2 dk}{\pi^2}$$

$$\begin{aligned}
 E = \frac{\hbar^2 k^2}{2m} &\Rightarrow k = \sqrt{\frac{2mE}{\hbar^2}} \Rightarrow dk = \left(\frac{2mE}{\hbar^2}\right)^{-1/2} \frac{m}{\hbar^2} dE \\
 &\Rightarrow \mathcal{D}(E)_{3D} = \frac{E^{1/2}}{2\pi^2} \left(\frac{2m}{\hbar^2}\right)^{3/2}.
 \end{aligned} \tag{1.7}$$

In *two* dimensions (where there is confinement in one direction — a *quantum well*), the Fermi sphere is replaced by a Fermi circle, and instead of calculating the number of k -states in a shell of radius k , thickness dk , we need to calculate the number in an annular ring of radius k , thickness dk , which is just $2\pi k dk$. The area per state in 2D is $(2\pi/L)^2$. Thus, the 2D density of states, taking into account spin degeneracy is $k dk L^2/\pi$, or in terms of number of states per unit area, is $k dk/\pi$. In terms of energy, this is

$$\begin{aligned}
 \mathcal{D}(E)_{2D} &= k \frac{dk}{\pi} = \sqrt{\frac{2mE}{\hbar^2}} \left(\frac{2mE}{\hbar^2}\right)^{-1/2} \frac{m}{\hbar^2} dE \\
 &= \frac{m}{\pi \hbar^2} dE.
 \end{aligned} \tag{1.8}$$

Now, this is the density of states for each value of bound state energy within the quantum well. The total density of states, taking into account all of the bound states is:

$$\mathcal{D}(E)_{2D} = \frac{m}{\pi \hbar^2} \sum_i H(E - E_i) dE, \tag{1.9}$$

where $H(E - E_i)$ is the Heaviside step function.

In *one* dimension, i.e. a *quantum wire*, the density of states is $2Ldk/\pi$, or per unit of space is $2dk/\pi$. This gives:

$$\begin{aligned} \mathcal{D}(E)_{1D} &= 2 \frac{dk}{\pi} = \frac{2}{\pi} \left(\frac{2mE}{\hbar^2} \right)^{-1/2} \frac{m}{\hbar^2} dE \\ &= \frac{1}{\pi} \left(\frac{m}{\hbar^2} \right)^{1/2} \frac{1}{E^{1/2}} dE. \end{aligned} \quad (1.10)$$

These have the form of peaks at each of the quantized energy levels, and are known as *van hove singularities*.

A consequence of this is that there is a *very* strong dimensional dependence on the electrical properties of conductors, as we shall see later.

Thus far, we have neglected the effect of temperature, and our analysis is only valid for $T=0$ K. As we increase temperature, the kinetic energy of the electrons within the conductor will increase, with the effect of smearing out the Fermi surface. This happens as some of the electrons at and just below the Fermi energy will go into slightly higher energy states just above the Fermi energy, depopulating their original states. The net result of this is that, around the Fermi energy, the probability of a state being occupied is no longer a simple step function (1 below E_F , 0 above E_F), but is determined by the *Fermi–Dirac distribution*, $f(E)$. The details of the derivation of this function have been reported in numerous solid-state and statistical mechanics textbooks, and we are only concerned with the result, which is

$$f(E) = \frac{1}{e^{\left(\frac{E-\mu}{k_B T}\right)} + 1}, \quad (1.11)$$

where μ is the chemical potential of the conductor: the energy of the highest occupied state, which is E_F at absolute zero. The product of $f(E)\mathcal{D}(E)$ is the density of *filled* states, and is shown in Fig. 1.7.

From a quantum mechanical standpoint, how does conduction occur? We have already ascertained that the electrons within the conductor experience a force $e\mathbf{E}$. Force is rate of change of momentum,

which in quantum mechanical terms is the rate of change of $\hbar\mathbf{k}$, i.e. $e\mathbf{E} = d(\hbar\mathbf{k})/dt$. Thus, the change in the k -vector in a time, t is $e\mathbf{E}t/\hbar$. If we assume that after each collision between the electron and the conductor, the velocity goes to zero, then the velocity gained in one mean free path is $\hbar\Delta\mathbf{k}/m = e\mathbf{E}\tau/m$, the same result we obtained earlier. If we apply an electric field in the $-x$ direction, the Fermi sphere moves in the $+x$ direction, eventually maintaining a steady position due to scattering events. This is illustrated in Fig. 1.8.

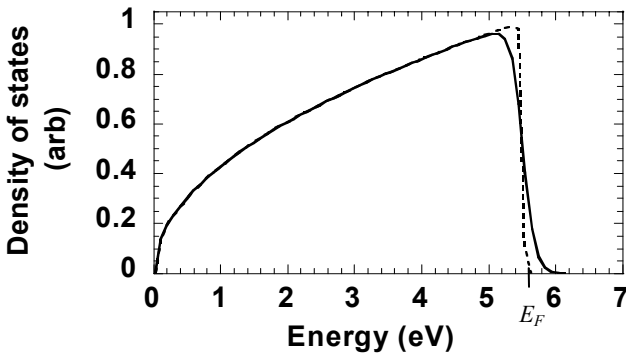


Fig. 1.7. Density of occupied states for gold ($E_F = 5.51$ eV). The solid and dotted curves are for $T = 1000$ K and 20 K, respectively.

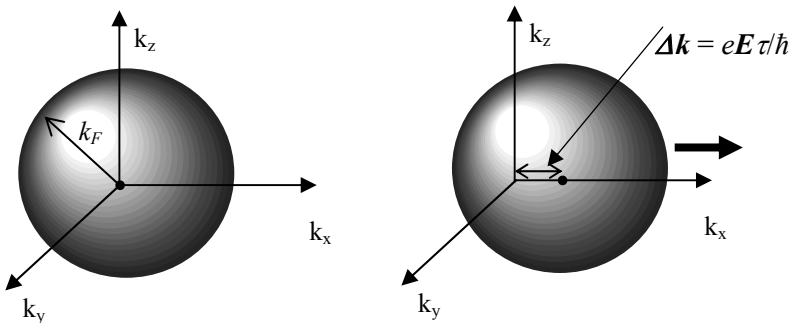


Fig. 1.8. Fermi sphere in the presence of an electric field.

We can better appreciate the subtleties of electronic transport if we look at a cross-section through the Fermi sphere, as shown in Fig. 1.9.

In summary, when we apply a voltage across a conductor, the electrons acquire some momentum opposite to the direction of the electric field. For many of the electrons within the conductor, this additional momentum is less than the random momentum they already have ($\hbar k$), and the only electrons which contribute to conduction are those which are near the leading edge of the Fermi surface, which is the shaded region shown in Fig. 1.9.

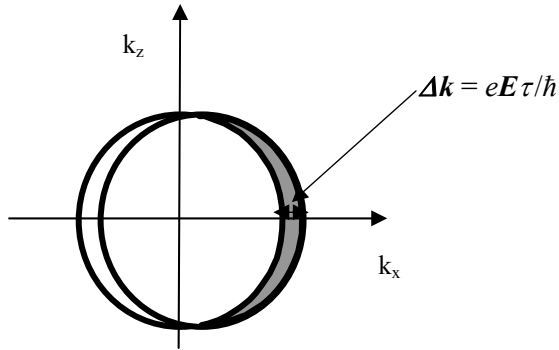


Fig. 1.9. Slice through the Fermi sphere in the presence of an electric field. The dark shaded region shows the new states around the Fermi energy which have been populated due to the field. These are the electrons contributing to conduction. The thickness of the line describing the circumference of the Fermi sphere is $k_B T$.

Therefore, for most situations, we might expect the conduction of electricity through a conductor to be determined by the density of states at the Fermi energy, $\mathcal{D}(E_F)$.

Combining the equations above for $\mathcal{D}(E)$ and E_F , we find that $\mathcal{D}(E_F) = 3n/(2E_F)$. In Table 1.1, we have listed the calculated n , E_F and $\mathcal{D}(E_F)$, as well as experimental values for the conductivity for some of the most technologically relevant metals.

The disparity between the calculated and observed trends in $\mathcal{D}(E_F)$ versus σ illustrates that conduction is not just determined by the

number of electrons within the conductor, but there are other factors at play.

To gain an understanding of this, we need to consider a more detailed version of conduction through materials — the *nearly-free-electron model*. Only then will we be able to appreciate the dominant factor in determining the electrical resistance of a conductor — scattering. This will then lead us onto phonons and the effect of temperature on electrical resistance.

Table 1.1. Calculated conduction parameters of selected metals.

Metal	Number of electrons/m ³	Fermi energy, eV	$\mathcal{D}(E_F)$, arb. units	σ (measured) $10^6 \Omega^{-1} \text{cm}^{-1}$
Au	5.90×10^{28}	5.51	1.07	0.45
Ag	5.85×10^{28}	5.48	1.07	0.63
Cu	8.45×10^{28}	7	1.21	0.59
Al	18.06×10^{28}	11.63	1.55	0.38

1.3. The Nearly-Free-Electron Model of Electronic Conduction and Band Structure

In crystalline materials, the electrons are in a periodic potential. The form of the electrical potential in a crystal is such that there are potential wells centred on the atomic cores. Each of these wells will have discrete allowed energy levels of a form similar to those of Hydrogen. Due to the proximity of the atoms to each other, the tails of the potentials overlap and modify the overall potential. This coupling causes the energy levels to shift and split, and for a number N of atoms, we will have N energy states, corresponding to $2N$ possible electron states (two comes from the two spin states of an electron).

Our aim is to find the form of the E - k relationship for electrons in such a potential, and see how it compares to the free electron case. Then we will be in a better position to understand the origin of the differences between metals, semiconductors and insulators.

There are a number of methods which may be used to investigate the nearly-free-electron model, including the Kronig-Penney model [3], which is an approximate method. We are going to use a more general method, based on solving Schrödinger's equation for an arbitrary periodic potential. Our starting point is to look at the basic form of the potential, as shown in Fig. 1.10.

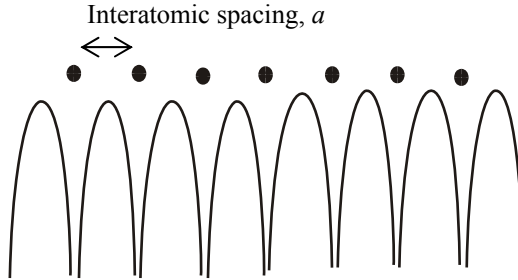


Fig. 1.10. Periodic potential as seen by an electron in a crystal. The black dots represent atoms.

The steps we will follow are:

1. Define potential as seen by an electron;
2. Expand as a Fourier series;
3. Solution to Schrödinger's equation contains information about periodic potential super-imposed on free-electron wave-functions (which are simple plane waves);
4. This periodic part to the solution can also be expressed as a Fourier series;
5. Insert all of the above into Schrödinger's equation and solve.

The reciprocal lattice vectors are given by \mathbf{G} , and the lattice spacing is a . As we are dealing with a periodic system, it is useful to use the Fourier expansion of the crystal potential, i.e.

$$V(x) = \sum_p V_p e^{iG_p x}, \quad (1.12)$$

where the Fourier coefficients are given by

$$V_p = \frac{1}{a} \int_0^a V(x) e^{-iG_p x} dx \quad (1.13)$$

and $p = 0, \pm 1, \pm 2, \dots$ and $G_p = 2\pi p/a$.

The general solution to the Schrödinger equation with a periodic potential is $\psi(x) = e^{ikx}u(x)$. This is a plane wave modulated by the function $u(x)$, where $u(x)$ is a periodic function with the periodicity of the lattice, i.e. $u(x)$ represents the influence of the crystal potential. This is known as **Bloch's theorem** [4], and $u(x)$ as a **Bloch function**. In Fig. 1.11, we show the typical form of the wave-functions for the free- and nearly-free-electron models, and we include the approximate lattice potential for reference.

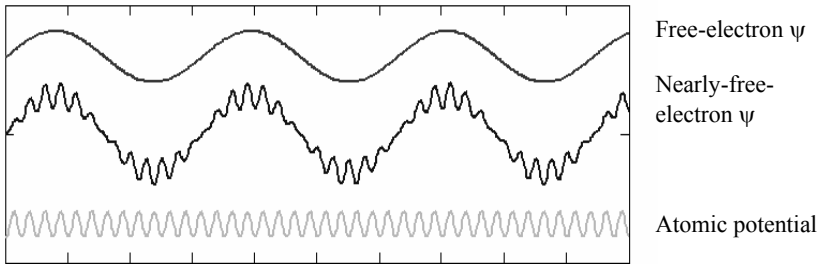


Fig. 1.11. Approximate form of single-particle wave-functions for the free- and nearly-free-electron models, and the atomic potential we are using in the nearly-free-electron model.

In the same way as we expanded the potential as a Fourier series, we can now do the same for $u(x)$, to obtain:

$$u(x) = \sum_n C_n e^{iG_n x}, \quad (1.14)$$

where $n = 0, \pm 1, \pm 2, \dots$ and $G_n = 2\pi n/a$.

That gives for the total expansion of the wave-function:

$$\psi(x) = \sum_n C_n e^{i(k+G_n)x}. \quad (1.15)$$

If we now insert the Fourier expansions of both $\psi(x)$ and $V(x)$ into Schrödinger's equation, we end up with a set of simultaneous equations in the unknown C_n . Note that the coefficients V_p are known, as the form of the crystal potential is assumed initially. There are an infinite number of terms, so to make the problem manageable, we will artificially truncate the series and consider only the leading-order terms given by $n = 0, \pm 1$. This is justified for weak potentials such as those found in metals, i.e. we write $V(x)$ as:

$$V(x) = V_0 + V_1 e^{iG_1 x} + V_{-1} e^{iG_{-1} x} = V_0 + 2V_1 \cos(G_1 x). \quad (1.16)$$

If we continue along the same lines, we can assume without any loss of generality that the electronic wave-function also only contains leading-order terms, i.e.

$$\psi(x) = [C_0 + C_1 e^{iG_1 x} + C_{-1} e^{iG_{-1} x}] e^{ikx}. \quad (1.17)$$

Substituting the above two equations into Schrödinger's equation we find

$$\begin{aligned} & \left(-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V_0 + V_1 e^{iG_1 x} + V_{-1} e^{iG_{-1} x} \right) [C_0 + C_1 e^{iG_1 x} + C_{-1} e^{iG_{-1} x}] e^{ikx} \\ & = E (C_0 + C_1 e^{iG_1 x} + C_{-1} e^{iG_{-1} x}) e^{ikx}. \end{aligned} \quad (1.18)$$

If we just consider a region where C_0 and C_{-1} dominate (i.e. we are only considering electrons travelling in the positive x -direction, with positive k), we are left with the relationships (noting that $G_{-1} = -G_1$ etc.):

$$\begin{pmatrix} -\frac{\hbar^2 k^2}{2m} C_0 + V_0 C_0 + C_0 V_1 e^{iG_1 x} + C_0 V_{-1} e^{iG_{-1} x} \\ -\frac{\hbar^2 C_{-1} e^{iG_{-1} x} (k + G_{-1})^2}{2m} + V_0 C_{-1} e^{iG_{-1} x} + V_1 C_{-1} + V_{-1} C_{-1} e^{2iG_{-1} x} \end{pmatrix} = EC_0 + EC_{-1} e^{iG_{-1} x}. \quad (1.19)$$

Collecting terms in $e^{iG_{-1} x}$, we find that:

$$C_0 V_{-1} = C_{-1} \left(\frac{\hbar^2}{2m} (k + G_{-1})^2 + E - V_0 \right). \quad (1.20)$$

Terms without any exponent give:

$$C_{-1} V_1 = C_0 \left(\frac{\hbar^2 k^2}{2m} + E - V_0 \right). \quad (1.21)$$

For a non-trivial solution, both ratios for C_{-1}/C_0 should be equal, i.e.

$$\left(\frac{\hbar^2 k^2}{2m} + E - V_0 \right) \left(\frac{\hbar^2 (k + G_{-1})^2}{2m} + E - V_0 \right) = V_1 V_{-1} = |V_1|^2. \quad (1.22)$$

This gives us a solution for positive values of k , and we can follow similar steps as above to include C_0 and C_1 to have solutions with negative values of k . A plot of E versus k is in Fig. 1.12. This is known as a *dispersion relation*, or a *band diagram*.

Points of interest to note relative to free-electron case:

- The energy is shifted up by the amount V_0 , where V_0 is the spatial average of $V(x)$.
- At certain values of k , i.e. at $k = -G_{-1}/2 = \pi/a$, gaps appear in the dispersion relationship, which define the *Brillouin Zone* boundary, and this corresponds to Bragg reflection. At this value of k , there are two possible values of energy:

$$E = V_0 + \hbar^2 (\pi/a)^2 \pm |V_1|. \quad (1.23)$$

We can interpret this as the lower and higher energies being the valence and conduction-band edges, respectively. The separation in energy, i.e. the *band-gap* is $2|V_1|$, which is just twice the first term in the Fourier series expansion of the crystal potential, and the Schrödinger equation has no wave-like solutions in this gap.

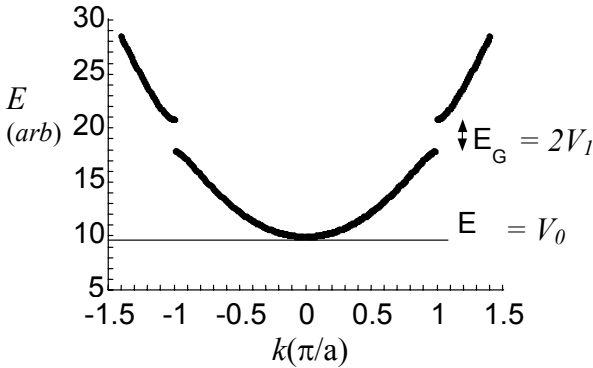


Fig. 1.12. E_k - k (dispersion) relationship according to the nearly-free-electron model.

The wave-functions at the valence- and conduction-band edges are:

$$\psi_{c,v}(x) = C_0 \left(1 \pm e^{\frac{2i\pi x}{a}} \right) e^{\frac{i\pi x}{a}}, \quad (1.24)$$

which can be written as:

$$\psi_v(x) = 2C_0 \cos\left(\frac{\pi x}{a}\right) \quad \text{and} \quad \psi_c(x) = 2iC_0 \sin\left(\frac{\pi x}{a}\right). \quad (1.25)$$

These are standing waves which means electrons at these energies do not travel through the crystal, and hence do not contribute to conduction.

For each term we add to the Fourier expansion of the crystal potential, there will be a band-gap, at successive multiples of $k = \pi/a$. Real materials tend to have many bands, as the atomic potential is not a simple sine function. Band structure is also a function of the direction we are looking in, as the interatomic spacing will appear different in different directions. A proper calculation would involve all three dimensions as well as the correct form of the atomic potential. The band-gap will depend very strongly on the interatomic distance — the closer together the atoms are, the larger will be the band-gap, as the overall interaction will be stronger. This is evident by looking at the band-gaps of diamond and silicon, which are both formed from group IV elements, C and Si, the only difference being the atomic size. In diamond, the band-gap is over 5 eV, whereas in Si, it is around 1.14 eV. This is directly related to the interatomic spacing which is 235 pm in Si and 142 pm in diamond.

The representation of E versus k in Fig. 1.12 is known as the *extended-zone scheme*. A more-commonly used representation is the *reduced-zone scheme*, as illustrated in Fig. 1.13. Here, the bands are all shifted laterally to enable them to fit within the first Brillouin zone.

Looking at Fig. 1.13, the bands are symmetric about $k = 0$; i.e. the conduction band minimum occurs at the same k -value (in this case, 0) as the valence band maximum. This is known as a *direct-gap* material. There are also *indirect-gap* materials, of which silicon is the most common example. Once the band-gap is too large for electrons to be thermally promoted from the valence-band to the conduction-band, the material is said to be insulating. This occurs once the band-gap is above around 3 eV (this also explains why many insulators are either white or transparent — they do not absorb much light).

So far, we have seen that classical mechanics gets it partly right, that is, it predicts the same value as quantum mechanics does for the conductivity of materials. By employing the nearly-free-electron model, we have demonstrated that we create a band-gap simply by placing electrons in a crystalline material, i.e. one with a regular (periodic)

atomic spacing. The band-gap depends on the interatomic spacing and on the atomic potential.

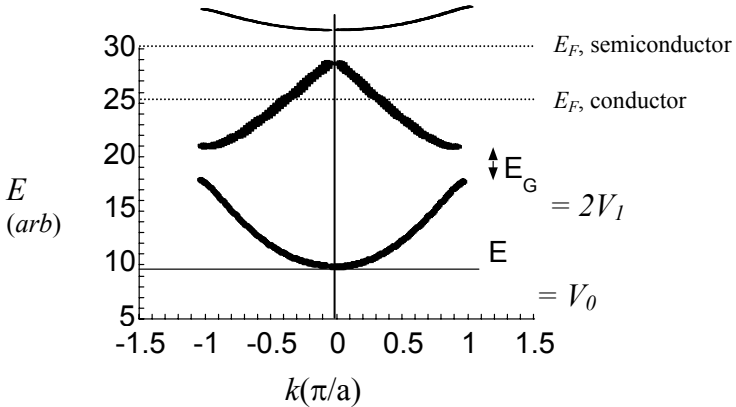


Fig. 1.13. E_k - k (dispersion) relationship according to the nearly-free-electron model, drawn in the reduced-zone scheme, in this case with three bands. The horizontal lines indicate the position of the Fermi energy in a typical conductor and semiconductor.

On a more subtle note, we have also found that the wave-function for an electron in a crystalline material is of the form

$$\psi(x) = \sum_n C_n e^{i(k+G_n)x}, \quad (1.26)$$

which is simply a linear superposition of plane waves, and an example is shown in Fig. 1.11. This wave-function has no damping, which means that an electron should be able to traverse any arbitrarily large distance in the material without a change in its wave-function. Another way of putting this is to say that *an electron should encounter no resistance whilst travelling through a crystal!* This is clearly at odds with what we observe experimentally, so the only reasonable conclusion we can draw is that real materials must not be perfect crystals. In fact, *anything* which disturbs the periodicity of a crystal will give rise to electrical resistance, by scattering the electrons. There are two dominant sources of scattering:

- Defects: a missing or substituted atom will locally distort the atomic potential
- Phonons: lattice vibrations which cause the atoms to jiggle around about their mean positions, continually distorting the atomic potential and destroying the periodicity.

The resistance due to defects is temperature-independent (known as residual resistance, and is sample-dependent), whereas that due to phonons has a very strong temperature dependence, becoming approximately linear above around 100 K. When dealing with conventional resistors, the widely-used assumption is that resistance scales linearly with temperature [5], i.e.

$$R = R_0 + \alpha \Delta T, \quad (1.27)$$

where α is the temperature coefficient of resistance of the material. In the case where the power dissipation in the resistor is large enough to cause heating, then $\Delta T \propto \text{Power} \propto I^2$ which means $R = R_0/(1 - \alpha\beta I^2)$, where β is the proportionality factor between temperature and power. This relationship, whilst empirical is extremely useful, and we will encounter it again when we are looking into transport in nanowires. For now, we will consider how resistance *should* vary with temperature, by considering the effect of phonons. First though, we will briefly introduce the concept of electron *effective mass*.

1.4. Effective Mass

When we apply either a voltage or a magnetic field to a conductor, the electrons will move due to the Lorentz force. In fact, we have already implicitly assumed that the force on an electron due to an applied electric field is given by $\mathbf{F} = e\mathbf{E}$. From the relationships $\mathbf{v} = \hbar\mathbf{k}/m$ and energy, $E = \hbar^2\mathbf{k}^2/2m + V$, we find that the electron velocity, $\mathbf{v} = (1/\hbar)dE/d\mathbf{k}$. From classical mechanics, we have that the power dissipated by the electron in moving in an electric field \mathbf{E} at velocity

\mathbf{v} is given by $dE/dt = e\mathbf{E}\mathbf{v}$. Combining these two relationships, we find that

$$dE = \left(\frac{e\mathbf{E}}{\hbar} \right) \frac{dE}{d\mathbf{k}} dt \quad (1.28)$$

and

$$\frac{d\mathbf{v}}{dt} = \left(\frac{1}{\hbar} \right) \frac{d^2E}{d\mathbf{k}^2} \frac{d\mathbf{k}}{dt}. \quad (1.29)$$

From Eq. (1.28), we have that $\frac{d\mathbf{k}}{dt} = \left(\frac{e\mathbf{E}}{\hbar} \right)$, which, upon substitution into Eq. (1.29) gives the following:

$$\frac{d\mathbf{v}}{dt} = \left(\frac{e\mathbf{E}}{\hbar^2} \right) \frac{d^2E}{d\mathbf{k}^2}. \quad (1.30)$$

Comparing this to Newton's second law, i.e. that Force = mass*acceleration, i.e. $\mathbf{F} = m d\mathbf{v}/dt$, we can re-write the above equation as

$$\begin{aligned} \mathbf{F} = m \frac{d\mathbf{v}}{dt} = e\mathbf{E} \Rightarrow m &= \frac{e\mathbf{E}}{\frac{d\mathbf{v}}{dt}} \\ \Rightarrow m^* &= \hbar^2 \left(\frac{d^2E}{d\mathbf{k}^2} \right)^{-1}. \end{aligned} \quad (1.31)$$

The star on the m denotes the fact that this is an *effective* mass, i.e. it is the mass that an electron appears to have within a material, be it a conductor or a semiconductor. The effective mass within conductors is generally approximately equal to the free-electron mass, but it can be significantly different inside a semiconductor. Within the free-electron approximation, the $E-\mathbf{k}$ dispersion relationship is quadratic, so the effective mass is a constant. However, as we have already seen in the nearly-free-electron approximation, most materials have a more complex band structure, so the concept of effective mass is less meaningful.

However, in the vicinity of band maxima and minima, the bands are approximately quadratic, so an effective mass may be assigned. As the atomic spacing within many crystals is direction-dependant, so is the band structure and therefore so is the effective mass, so we usually have an effective mass tensor. Thankfully, in many cases, the average value of effective mass over all directions can be used. Effective mass can be measured by a variety of means, but the most common are cyclotron resonance and angle-resolved photoemission (which, incidentally is how band structure is measured for materials today).

By considering band structure and values of the effective mass of electrons in different semiconductor materials, we can begin to understand why certain materials are so widely used by the semiconductor industry. In Table 1.2, we have written the effective masses of electrons and holes in Si, Ge, and GaAs.

Table 1.2. Effective carrier masses in semiconductor materials.

Material	Electron effective mass as a proportion of free-electron mass	Hole effective mass as a proportion of free-electron mass
Si	0.36	0.81
GaAs	0.067	0.45
Ge	0.55	0.37

To appreciate the significance of these values, remember from earlier that the maximum velocity an electron reaches within a conductor under the application of an electric field E is

$$\mathbf{v} = \frac{e\tau\mathbf{E}}{m} = \mu\mathbf{E}, \quad (1.32)$$

where μ is called the *mobility*. The higher the mobility, the faster the electrons travel and ultimately the faster the device/circuit can operate. Clearly, the mobility is inversely proportional to the electron's effective mass, so high-speed devices (the most pervasive example being mobile phones and networks) tend to use GaAs and alloys of GaAs rather than

Si. Also, GaAs is a direct-gap semiconductor unlike Si, so it has applications in lasers and photodetectors. This is the main reason why GaAs and GaAlAs are used in quantum well, wire and dot devices, which we will analyse in the next chapter. Si also has applications in photonic devices, in the form of strained silicon and polysilicon (abundance of surface states lower the band-gap sufficiently to render it effectively direct-gap).

An important question then is — why is silicon still the material of choice in most semiconductor devices and in microprocessors where speed is so crucial. The answer, not surprisingly comes down to economics — Si is one of the most abundant materials in the earth’s crust (27%), it is cheap and easy to process (GaAs and its derivatives are extremely expensive to deposit and process), and its oxide, SiO_2 is one of the best insulators known — it can easily be grown, has few defects, and grows epitaxially on Si.

1.5. The Origins of Electrical Resistance

We have just seen that the nearly-free-electron model of conduction predicts that a perfect conductor should have no resistance, and that the resistance we observe is due to scattering. We will explore the quantum nature of scattering from defects in the next chapter, but first we will investigate scattering from phonons, which is the dominant source of electrical resistance. In Fig. 1.14, we have summarised the main point: an electron travels through the crystal for a certain distance

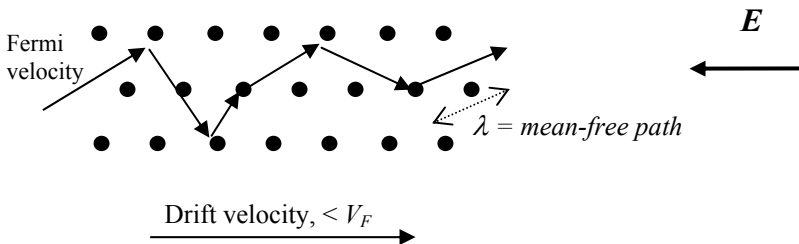


Fig. 1.14. Electron–phonon scattering. The dots represent atoms. On average, an electron will travel a distance λ before being scattered.

(the mean free path, on average) before being scattered from either a defect or a phonon. *After the scattering event, both the electron and phonon have their direction and momentum changed.* Between collisions, the electron travels at the Fermi velocity, and the transport is said to be “Ballistic”.

To investigate the effect of phonons, we must first understand some basic points about them. We will start with the fact that the atoms in a crystal are bound together by a potential of the approximate form:

$$U(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]. \quad (1.33)$$

This is known as the *Lennard-Jones 6-12 potential* [6] for inert gases.

This potential has a minimum at $r = \sigma$, which is the equilibrium bond-length, with values between 2 and 3 Å for most materials.

The force which binds atoms together in a crystal is $-dU/dr$, which is plotted in Fig. 1.15.

For this analysis, we are just considering a 1D crystal, and are assuming the atoms are all identical.

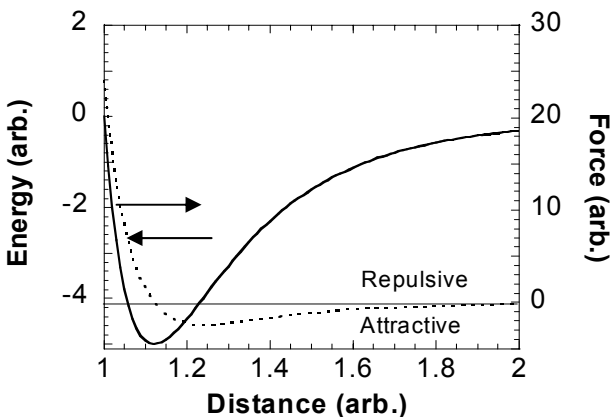


Fig. 1.15. Form of inter-atomic energy and forces. Around the equilibrium position and in contact ($F = 0$), the force depends linearly on distance for small displacements.

For small displacements about their equilibrium positions, however, this force is approximately linear, with the result that we can consider the atoms as being held together by springs. To understand the nature of phonons then, we must combine Schrödinger's equation with Hooke's law, i.e. we have an atom of mass m suspended from a spring of stiffness κ , which has $x = 0$ as its equilibrium position, as is illustrated in Fig. 1.16. The force on the particle, $F = -\kappa x$.

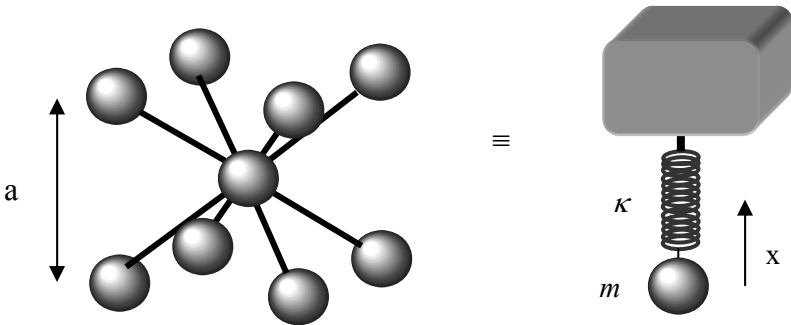


Fig. 1.16. Quantum simple harmonic oscillator. Atoms in a crystal (left) can be considered to be held together by springs (right).

According to classical mechanics, this atom on a spring will oscillate at the natural frequency $\omega_c = (\kappa/m)^{1/2}$.

The potential energy of the system is V , where

$$V = 1/2\kappa x^2 = 1/2m\omega_c^2 x^2. \quad (1.34)$$

Schrödinger's equation for this system is

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + \frac{1}{2}m\omega_c^2 x^2 \psi = E\psi. \quad (1.35)$$

This problem can be simplified if we employ a change of variables from x to y , where $y = (m\omega_c/\hbar)^{1/2} x$, and define $\alpha = 2E/\hbar\omega_c$.

Schrödinger's equation is now:

$$\frac{d^2\psi}{dy^2} + (\alpha - y^2)\psi = 0. \quad (1.36)$$

The solution of which is known to be $\psi(y) = F(y)\exp(-y^2/2)$ (To see where this comes from, look at the asymptotic solution when $y \gg a$, which is of the form $y = y^n \exp(-y^2/2)$.)

Substituting this form of $\psi(y)$ into the above equation, we find

$$F'' - 2yF' + (\alpha - 1)F = 0. \quad (1.37)$$

We should now assume a power series solution for $F(y)$;

$$F = \sum_{p=0}^{\infty} a_p y^p. \quad (1.38)$$

From which it can be seen that:

$$F' = \sum_{p=0}^{\infty} p a_p y^{p-1} \quad \text{and} \quad F'' = \sum_{p=0}^{\infty} p(p-1) a_p y^{p-2}. \quad (1.39)$$

An important point here is that y can never have a negative power, as that would lead to an infinity at $y = 0$ which would be unphysical (ψ must always be finite, as $|\psi(x)|^2$ represents the probability of the particle being located at position x). The first two terms of F'' therefore must equal 0, so we must put $p = p + 2$ in the above expression for F'' .

Substituting for F , F' and F'' into Schrödinger's equation leads to the following:

$$\sum_{p=0}^{\infty} [(p+2)(p+1)a_{p+2} - (2p+1-\alpha)a_p] y^p = 0. \quad (1.40)$$

For a non-trivial solution, the coefficient of each power of y must vanish, leading to the following recursion relationship:

$$\frac{a_{p+2}}{a_p} = \frac{(2p+1-\alpha)}{[(p+1)(p+2)]}. \quad (1.41)$$

However, the resulting power series tends to infinity with increasing y (the limit of a_{p+2}/a_p tends to $1/p$, the sum of which is infinity), so we must truncate the power series.

The solution can actually be re-written as two power series, each containing all even or odd powers of y . Using the recursion relation, all coefficients can be expressed in terms of either a_0 or a_1 . Then we need to choose some value for p , say, n , such that $2p+1-\alpha=0$. That power series will end there, and we need to neglect the other power series. Both of these conditions lead to the following:

$$\begin{aligned}\alpha &= 2n+1 \text{ for } n=0, 1, 2, \dots, \\ a_1 &= 0 \text{ for } n \text{ even, } a_0 = 0 \text{ for } n \text{ odd.}\end{aligned}$$

From our definition of $\alpha = 2E/\hbar\omega_c$ we have for the energy eigenstates of the quantum simple harmonic oscillator:

$$E_n = (n+1/2)\hbar\omega_c. \quad (1.42)$$

That is, discrete, equally spaced energy levels, with a ground state or *zero-point energy* of $(1/2)\hbar\omega$. Each energy level corresponds to a phonon mode.

What is the consequence of the zero-point Energy? It means that according to quantum mechanics, a harmonic oscillator can never be completely at rest, because then we would know it is momentum (zero) and position precisely, which goes against Heisenberg's Uncertainty principle. It means that even at absolute zero, the atoms in a material will still be jiggling around by a very small amount.

Normalising (i.e. using the condition that $\int_{-\infty}^{\infty} |\psi(x)|^2 dx = 1$), we find the first three wave functions are

$$\left. \begin{aligned}\psi_0(x) &= (m\omega_c/\pi\hbar)^{1/4} \exp(-m\omega_c x^2/2\hbar), \\ \psi_1(x) &= (4\pi)^{1/4} x/4(m\omega_c/\pi\hbar)^{3/4} \exp(-m\omega_c x^2/2\hbar), \\ \psi_2(x) &= (m\omega_c/4\pi\hbar)^{1/4} (2m\omega_c/\hbar)x^2 - 1) \exp(-m\omega_c x^2/\hbar).\end{aligned}\right\} \quad (1.43)$$

In the ground state, the most probable position is in the centre, whereas for higher levels, the probability oscillates quickly, as shown in Fig. 1.17. Comparing to the conventional (classical) simple harmonic oscillator: the most probable position is always at the extremes (the velocity is lowest there). This is just another example of a difference between classical and quantum descriptions of the same system: they often disagree. However, when looking at highly-excited states of the quantum harmonic oscillator, the envelope of the probability distribution approaches that expected classically. This is quite a common occurrence: the quantum mechanical description of systems tends to converge towards the classical one for highly excited states.

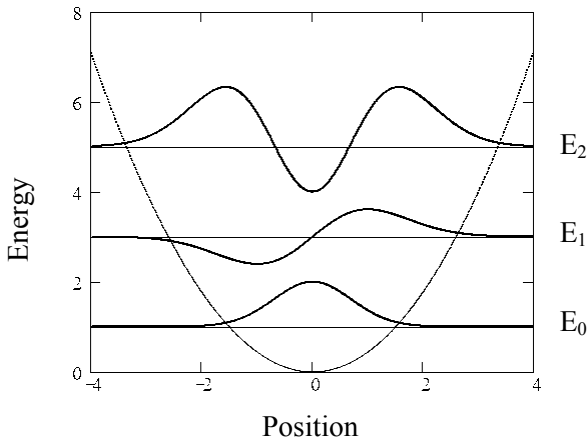


Fig. 1.17. First three levels of the quantum simple harmonic oscillator, with associated wave-functions. The oscillator potential is shown as a dotted curve.

How many phonons are present in a lattice? That is, if we assume a constant electron–phonon coupling strength and probability, then the temperature dependence of this contribution to resistance can be calculated by knowing the number of phonons.

The probability of occupation of the n th energy level due to thermal excitation is $e^{(-E_n/k_B T)}$ (*cf. Boltzmann factor*).

Therefore, we can write the average internal energy of the crystal as:

$$\langle E \rangle = \frac{\sum_{n=0}^{\infty} E_n e^{-E_n/k_B T}}{\sum_{n=0}^{\infty} e^{-E_n/k_B T}}. \quad (1.44)$$

Given that the spacing between energy levels is $\hbar\omega_c$ and the energy per oscillator is $\hbar\omega_c$ we can write the phonon occupation number $\langle n \rangle$ as $\langle E \rangle / \hbar\omega_c$, i.e.

$$\langle n \rangle = \frac{1}{e^{\hbar\omega_c/k_B T} - 1}. \quad (1.45)$$

This is plotted in Fig. 1.18, from which we can see that the number of phonons does indeed increase approximately linearly around and above room temperature, justifying the common usage of the expression $R = R_0 + \alpha\Delta T$ when dealing with macroscopic resistors.

In fact, if one looks at resistance versus temperature curves for most metals, their overall shape is almost exactly the same as that shown in Fig. 1.18, except for an offset known as *residual resistance*, dependent on the number of defects in the material, which is sample-dependent.

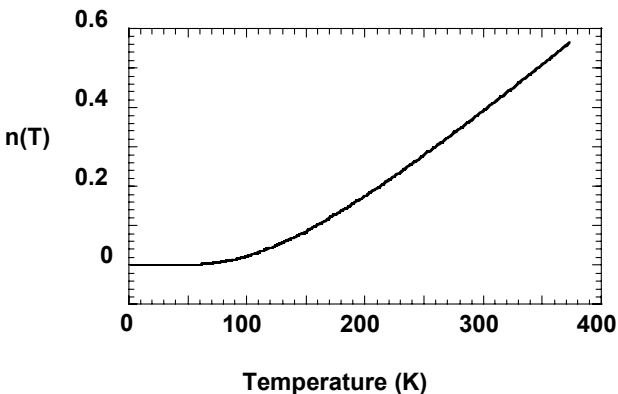


Fig. 1.18. Calculated phonon occupation number as a function of temperature.

In summary, we have seen that electrical resistance is due to the non-ideal nature of real materials, and that we can gain a basic understanding of electronic transport in conductors using quantum mechanics. A key feature of the quantum universe is the idea of quantisation, or “things being discrete”, i.e. coming in small packets.

1.6. Size Effects on Electrical Resistance

The question which we are now able to address, and which is a key topic in this text, is “*how does the resistance change as we make a conductor smaller?*”? What we have seen already, particularly through Fig. 1.5, is that quantum size effects tend to go unnoticed for dimensions above around 10 nm. Assuming that Ohm’s law were to hold true for all dimensions, the resistance of a gold wire as a function of its diameter *should* vary as shown in Fig. 1.19.

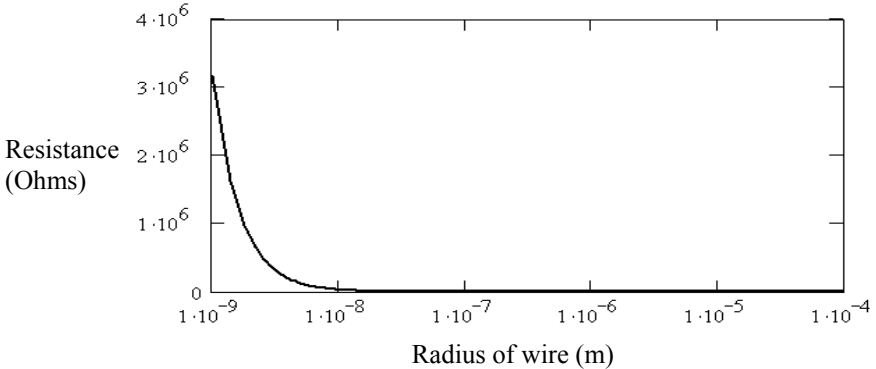


Fig. 1.19. Calculated resistance of a gold wire of length 1 mm, as its radius varies from 1 nm to 0.1 mm, using $R = \rho l/A$.

Indeed, this holds true for dimensions down to around 50–60 nm, but then begins to change dramatically, initially due to mesoscopic effects, and finally due to quantum effects. Resistance versus size in the range 1–100 nm is illustrated in Fig. 1.20.

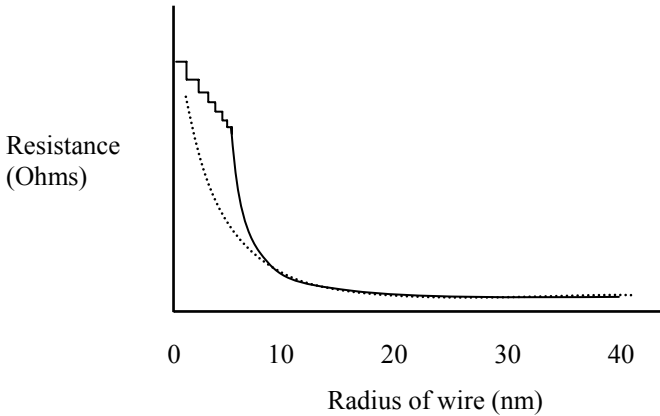


Fig. 1.20. Dependence of resistance of a wire as its radius varies from 1 nm to 40 nm. The dotted curve is that expected from the expression $R = \rho l/A$.

So far, we have restricted ourselves to looking at bulk materials and wires. In real devices of course it is the entity to which the wires are connected that is important. In the vast majority of cases, this entity is a transistor in some form or other. What is so special about the nanoscale is that a wire is no longer just a passive component which carries current, but many intriguing things start to happen, and it becomes just as important as the “transistor”. We will now take a brief look at semiconductor transistors, and later chapters are dedicated to wires and molecular transistors.

1.7. Overview of Transistors

Since its humble beginnings in the early 20th century in the guise of the vacuum tube, the transistor has become the single most important electronic component on the planet. It is at the heart of operation of most electrical devices from energy-saving light bulbs to computers, mobile phones and televisions. In the first 40 years of its development, very little changed in the appearance and operation of the vacuum tube. It was only in the 1940s that all changed with the advent of the solid-state transistor by Bardeen, Brattain and Schokley at Bell

Labs. [7]. The first transistor was several cm across, whereas the transistors now in high-end computer microprocessors are around $1\ \mu\text{m}$ across, the active region being less than 100 nm across.

Transistors broadly fall into two classes: bipolar junction transistors (BJTs) and field-effect transistors (FETs). The former are typically used in high-current, high-power applications, whereas the latter are used in low-current precision devices, most notably microprocessors. We will therefore concentrate on FETs as they are by far the most technologically relevant. All transistors are three-terminal devices where the current between two of the terminals is controlled by a small current (BJT) or voltage (FET) into or at the third, controlling terminal. In a BJT, the terminals are called the Base (controlling terminal), Collector and Emitter, whereas in the FET, the current flowing between the Drain and Source is controlled by the voltage at the Gate. The basic principle of operation of a FET is illustrated in Fig. 1.21. The current flows in a *channel* between the drain and source, and the length of this channel is commonly termed the *gate length*. The gate region is adjacent to this channel and an electrode is patterned on top of this and is electrically isolated from it by a thin layer of insulator, the *gate oxide*.

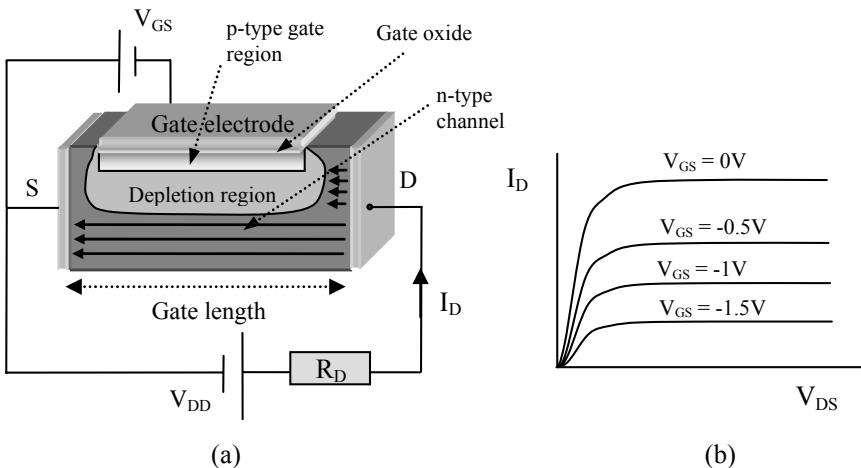


Fig. 1.21. (a) Schematic of *n*-channel FET (*S* and *D* are the source and drain, respectively, *G* refers to the gate); (b) current–voltage characteristics of FET.

The width of the channel is controlled via the gate voltage; if the gate-source junction is reverse-biased then as V_{GS} increases by an amount ΔV_{GS} , the depletion region (shown in Fig. 1.21) will spread out into the channel, increasing the channel's resistance and hence decreasing the drain current, I_D . The resistor R_D is chosen such that the resulting change in voltage across it, i.e. $\Delta I_D R_D > \Delta V_{GS}$. In this way, the circuit has *amplification* — a small input signal produces a large output signal. The entire device is typically made from silicon which is doped in one of two possible configurations: the channel is n -type and the gate is p -type or vice-versa. The gate oxide is SiO_2 , which has excellent electrical characteristics. We have illustrated a device with an n -type channel in Fig. 1.21(a), as this is the most commonly found one. In Fig. 1.21(b), we show the typically observed electrical characteristics of such an FET.

The speed of operation of a computer depends inversely on the size of the transistors within the microprocessor, and on the number of transistors. Since the late 1960s, the gate length has decreased by several orders of magnitude, and the number of transistors per microchip has increased by a factor of 2 around every 18 months. This trend is famously known as “Moore’s Law”, after Gordon Moore, the co-founder of Intel who first pronounced it back in the late 1960s [8]. This is illustrated in Figs. 1.22 and 1.23 which show how the number of transistors per microchip and the gate length and operation speed are evolving. In fact, since around 2000, the rate of miniaturisation has increased, despite early predictions to the contrary!

The benchmark of one billion transistors per microchip has just been surpassed in 2005/06 with Intel’s Montecito chip, weighing in at 1.72 billion transistors. For comparison, the number of connections in the human brain is around 10^{15} — a further six orders of magnitude more!

The dimensions of transistors are approaching that at which quantum effects will become noticeable. In fact, this is already happening, as the gate oxide is typically 0.8 nm thick, so current can leak between the gate and the channel, via a process known as *quantum tunneling*, which we will look at in more detail in the next chapter. The

gate length is below 50 nm made using 90 nm process technology, 35 nm using 65 nm technology, and < 25 nm using 45 nm technology, which is due to come on-line in 2007 [9]. Prototype transistors with gate lengths in the range 10–15 nm have already been produced and have been shown to have excellent performance. IBM have fabricated a proof-of principle transistor with a gate length of 6 nm [10].

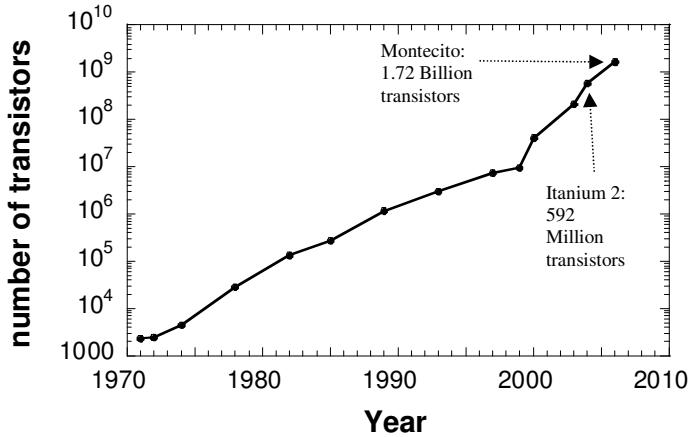


Fig. 1.22. Chronological development of the number of transistors per microchip (Moore’s law).

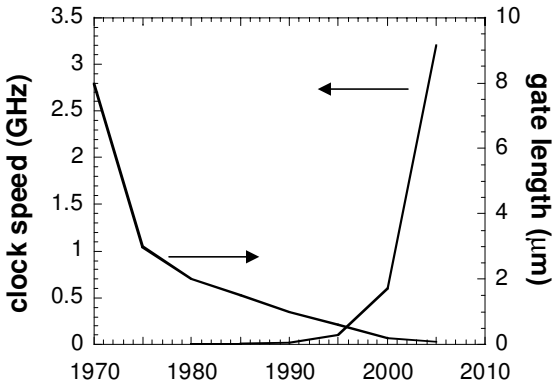


Fig. 1.23. Chronological development of the gate length and speed of transistors.

1.8. Surface Effects

The famous physicist Wolfgang Pauli once said that “*God made solids but surfaces were the work of the devil*”. This statement was prompted by the fact that the properties of surfaces are very different to those of the bulk. The reason for this is of course that while atoms in the bulk are surrounded by other atoms on all sides, i.e. they are fully co-ordinated and in stable positions, those on the surface are not. This broken symmetry leads to surface atoms being highly reactive, and in response to the lack of co-ordination, on many surfaces the atoms re-arrange their positions relative to those of bulk atoms to lower their energy. The resulting surface reconstructions have different electronic states to those found in the bulk [11]. Fortunately, this is something which we can take advantage of, and is one of the reasons why nanostructures have drawn so much interest, because by virtue of their size, they are more surface than anything else. Pauli’s statement has been superseded by E. W. Plummer who remarked that “*surfaces are the playground of solid state physics*”.

Earlier, we saw that the wave-functions of electrons in a periodic potential are of the form $\psi(x) = e^{ikx}u(x)$. If we now consider the case of a 3D semi-infinite solid, as shown in Fig. 1.24, the wave-functions are:

$$\psi_k(r) = e^{ik \cdot r} u_k(r). \quad (1.46)$$

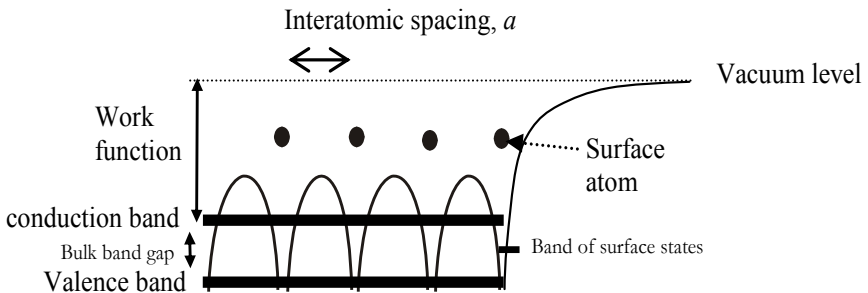


Fig. 1.24. Lattice potential near a surface. The break in symmetry caused by the surface induces the formation of surface states known as *Tamm* states.

Now, we have implicitly assumed in our previous analysis that the k vector is real. If it were to be imaginary, then $\psi_k(r)$ could diverge to infinity within the crystal. One possibility is to make the component of k parallel to the surface (k_{\parallel}) real whilst making k_{\perp} imaginary, i.e.

$$\psi_k^s(r) = e^{ik_{\parallel}r_{\parallel}} u_{k_{\parallel}}(r_{\parallel}) e^{-k_{\perp}r_{\perp}}. \quad (1.47)$$

This results in wave-functions which decay exponentially away from the surface both into the bulk and into free space (these wave-functions are bound to the surface and are called *surface states*). One therefore needs to find wave-functions within the bulk which show this property of growing exponentially towards the surface, as is illustrated in Fig. 1.25.

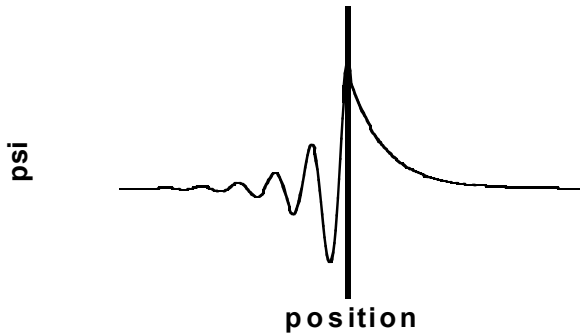


Fig. 1.25. Surface state wave-function. The vertical line indicates the position of the surface. The envelope of the wave-function (ψ) decays exponentially away from the surface.

As one might expect, for any given value of k_{\parallel} there is a finite spectrum of allowed k_{\perp} values. Thus, the spectrum of surface state energy levels is discrete. One point to note is that surface states are non-degenerate with bulk states, as then it would be possible for an electron in a surface state to couple to the bulk and there would no longer be any confinement of electrons on the surface. Therefore, *surface state energy levels exist within the bulk band-gaps*. Surface states broadly fall into two classes, *Shockley* and *Tamm* states. Basically, Tamm states are due to the broken periodicity of a crystal due to the presence of a surface,

whereas Shockley states arise due to the re-arrangement of surface atoms. The origin of both types of state are illustrated in Figs. 1.24 and 1.26. Therefore, whilst both tend to co-exist, the relative amount of the two types of state depends on the nature of the surface atomic arrangement.

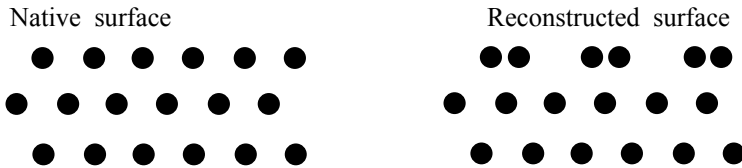


Fig. 1.26. Re-arrangement of surface atoms (*surface reconstruction*) induces the formation of *Shockley* surface states.

Surface states can be probed in a number of different ways which are beyond the scope of this book, but briefly they are based on performing energy spectroscopy on electrons which have been ejected from the surface, either using photons — XPS (X-ray photoelectron spectroscopy), electrons (Auger spectroscopy), or ions. Surface reconstructions may be observed locally in real-space using scanning-probe microscopy, as we will see in Chapter 4, or globally in reciprocal space using low energy electron diffraction (leed).

References for Chapter 1

1. G. Ohm, *Die galvanische Kette: mathematisch bearbeitet* (*The Galvanic Circuit Investigated Mathematically*) (Riemann, Berlin, 1827).
2. N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College Publishing, Philadelphia, 1976).
3. S. M. McMurry, *Quantum Mechanics* (Addison-Wesley, 1993).
4. F. Bloch, *Z. Physik* **52**, 555–600 (1928).
5. P. W. Horowitz and W. Hill, *The Art of Electronics* (Cambridge University Press, 1980).

6. J. E. Lennard-Jones, *Proc. Phys. Soc.* **43**, 461–482 (1931).
7. The junction transistor, US Patent #02569347.
8. G. E. Moore, *Electronics* **38** (1965).
9. Source: Intel website.
10. Source: IBM website.
11. An excellent introduction to surface science is A. Zangwil, *Physics at Surfaces* (Cambridge University Press, Cambridge, 1988).